THE UNIVERSITY OF CHICAGO

GENETICALLY ENCODED MECHANOTRANSDUCTION RESPONSE IN ENDOTHELIAL CELLS MEDIATES HUMAN CARDIOVASCULAR TRAITS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PATHOLOGY

BY
MATTHEW DANIEL KRAUSE

CHICAGO, ILLINOIS

MARCH 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to thank my advisor Dr. Yun Fang for the opportunity to do gene regulatory research. I am particularly grateful for his support in allowing me to explore and develop new techniques to enrich my research from transfection of endothelial cells, CRISPR, NGS, and bioinformatics. I thank you for your patience during my training and for the opportunities to perform and present my research. Your love for science will stick with me into my career and the rest of my life.

I am very thankful to the members of the Fang lab. David Wu has been a strong mentor about science, medicine, and life and has been an excellent resource for advice and support. Ru-Ting Huang has been a stalwart member of the lab keeping things running smoothly and always knows where all the reagents are. Ru-Ting and I also chatted as bench neighbors and she shared her experience and wisdom doing science and research. I appreciate Heng Duong for our extensive conversations that could totally make my day. I appreciate Heng for reminding me that accomplishments need to be celebrated and to be generous with others. Devin Harrison and Jin Li were wonderful to work with, teach, as well as learn from.

I want to thank the members of my committee Kate Reardon, Ivan Moskowitz, and Anna Di Rienzo. Your mentorship and advice helped me to pursue novel, rigorous scientific research. I thank Rajiv and Alex of the Moskowitz lab for the fun times we had and for helping with ATAC-seq. I want to thank the Di Rienzo lab, particularly Olivia for being an awesome peer that I had the pleasure of working and learning with. I want to thank Dave Witonsky for his help in data science and bioinformatics that I will carry into my career.

I want to thank my program chairs Steve Meredith and Matthew Brady. Steve is a wonderful person to learn from and I appreciate his perspectives on science and faith. Matthew has been a beacon of support since I interviewed and always had a smiling face to check in on me as a I progressed through graduate school, which I most surely needed.

I want to acknowledge my friends in my cohort: Alex, Anastasia, Kyle, Logan, and Sriram have been there since day one of graduate school. You are all wonderful, successful, and smart people that will go very far in life and I am honored to be among you. I cherish the times we spent on and off campus exploring the city and playing board games. I want to thank Casey and Tiffany for their friendship and that our dogs could bond and play during the days we worked from home. I want to thank Lee for being a good friend at a time I really needed one.

I want to thank my family for rearing me in an environment where I could be challenged thrive intellectually. I am grateful to my dad who was the first in his family to go to college and get as far as his doctorate. You have always been an inspiration to me, and I am proud to follow in your footsteps in earning my doctorate. I want to thank my mother for being someone I can talk to. I want to thank my grandparents for their legacy of work and love. I want to thank my sister Megan who

I have to thank my dog, Bisou, for her unconditional love, her need to walk every day that got me back into shape, wet kisses, and the pure joy she brings to my life.

I want to give my biggest thanks to my wife Sofia. You have been here all the way and I love you so much. You have worked selflessly to support me, to make our house a home, and I am proud to stand by and watch you succeed. We made it!

ABSTRACT OF THE DISSERTATION

Endothelial cell morphology and gene expression are driven by shear stress. Disturbed flow causes constitutive endothelial activation resulting in chronic, non-resolving, and pathological inflammation underlying human CAD. Human CAD is a complex disease that is strongly influenced by heredity. It is unclear how genetic variation can mediate shear stress response in human endothelium. In this work I demonstrate that mechanosensitive enhancers transcriptionally regulate many flow-sensitive genes in human endothelial cells using fine mapping strategies, epigenomic profiling of histone proteins, chromatin accessibility, gene expression, genome editing, and other functional assays. First, I found that a human polymorphism associated human cardiovascular disease predisposition, rs17114036, to be located within an endothelial-specific enhancer that is selectively activated by unidirectional flow. CRISPR-Cas9-based genome editing functionally demonstrates that removing this rs17114036-containing region decreases PLPP3 under unidirectional flow and results in a pro-inflammatory, athero-relevant phenotype. Next, to systematically investigate all genomic flow-sensitive cis-regulatory elements, I performed differential bioinformatics analyses on RNA and ATAC-seq data from endothelial cells subjected to unidirectional and disturbed flow to identify thousands of genes and nucleosome-free clusters that are significantly regulated by flow. Lastly, I integrated this data with promoter capture Hi-C from endothelial cells to directly link regulatory elements with its target gene to probe the gene-regulatory changes that occur in response to flow. I further discovered that many of these elements harbor hundreds of polymorphisms from the GWAS catalog, associating mechanosensitive enhancers with many human cardiovascular phenotypes. Taken together this work finds networks of genes, sets of gene regulatory elements, and human polymorphisms related to human cardiovascular traits.

These results demonstrate that human genetic variation mediates endothelial mechanotransduction mechanisms and that these thousands of gene-regulatory elements influence human cardiovascular disease predisposition and traits.

Chapter 1: Introduction

1.1 Atherosclerosis in Coronary Artery Disease and Ischemic Stroke

Atherosclerosis is the progressive thickening of the intimal layer of blood vessels and is an underlying cause of coronary artery disease (CAD) and ischemic stroke (IS) (Pasternak *et al*., 2004). Atherosclerotic plaques preferentially develop in large elastic arteries and affect the flow of blood, oxygen exchange, and metabolite delivery to tissues (Hansson *et al*., 2005). This pathogenic constriction of blood vessels due to these plaques trigger many human diseases including CAD, aortic disease, peripheral artery disease, and ischemic stroke (Libby *et al*., 2005). The rupture of atherosclerotic lesions and thrombosis leads to heart attacks and strokes, which are the main causes of death among adults in the world (Tunick *et al*., 2000; Benjamin *et al*., 2018).

The initiatory steps of atherosclerosis are triggered by endothelial dysfunction, which is a pro-inflammatory state that leads to cellular signaling to attract leukocytes that adhere to and migrate into the endothelial layer (Kasikara *et al*., 2018). Notably, monocytes transmigrate and develop into macrophages. The macrophages phagocytose oxidized low-density lipoprotein (LDL) accumulated in the vessel walls as the result of hypercholesterolemia. Macrophages loaded with lipid, termed foam cells, are present in the early fatty streaks, many of which progress to more advanced atherosclerotic lesions (Kasikara *et al*., 2018). These foam cells along with other cells in the innate (e.g. dendritic cells) and adaptive immune system (e.g. T cells) produce the chronic pathogenic and non-resolving inflammation that promotes the pathogenesis of atherosclerosis.

The early stages of atherosclerotic disease are asymptomatic. More progressed atherosclerosis is manifested in symptoms such as breathing trouble and chest pains. Rupture of

the atherosclerotic plaque and erosion of the intimal layer of endothelial cells lead to acute CAD (Fuster *et al*., 1992; Hansson *et al*., 2005). This plaque rupture leads to local clotting of blood and thrombosis formation within a blood vessel resulting in vessel occlusion and ischemia (Cort *et al*., 2002; Kasikara *et al*., 2018). These conditions result in the most severe CAD symptoms including myocardial infarction and sudden cardiac death (Libby *et al*., 2005). The majority of the plaques form in predictable locations such as arterial bifurcations, curvatures and branches (Davies *et al*., 2013). Ischemic stroke is when unstable atherosclerotic plaques block or rupture within the arteries supplying oxygen to the brain (Bonita *et al*., 1992). In the majority of stroke cases the plaques occur preferentially at the carotid sinus in the bifurcation, where atherogenesis progresses and leads to symptoms (Touzé *et al*., 2005).

1.2 Endothelial Mechanotransduction in Atherosclerosis

The majority of the atherosclerotic plaques form in predictable locations in the vasculature such as arterial bifurcations, curvatures and branches (Davies *et al*., 2013). These atherosclerotic susceptible regions of the vasculature are chronically subjected to high fluid turbulence that creates disturbed flow patterns due to atypical vascular geometry (Chiu and Chen, 2011; Davies *et al*., 2013; Gimbrone and Garcia-Cardena, 2016). The frictional force between flowing blood and endothelial cells is termed shear stress (Zhou *et al*., 2014). Shear stress is strongly determined by flow velocity and therefore, by vessel geometry and shape (Bryan *et al*., 2014). This interaction between flowing fluid within the confines of the blood vessels creates complex hemodynamic patterns or waveforms. As stated by Poiseuille's Law, high shear stress is generated by unidirectional flow with high velocity, whereas in vessels that bifurcate and otherwise twist and turn result in disturbed flow featuring oscillation, flow reversal,

and low average shear stress (Wang *et al.*, 2015). These waveforms initiate responses from the vasculature to dictate structure, growth, and remodeling (Vermot *et al.*, 2009; Davies *et al.*, 2013). The hemodynamic forces generated by blood flow have significant consequences for the endothelial cells that directly interface with the blood. Endothelial mechanotransduction is the process by which endothelial cells sense and convert physical force, in this case shear stress, into a biological signaling. As a consequence of different flow patterns arising from the circulation, not all regions of the vasculature are equally susceptible to lesion formation. Blood vessels subjected to high shear stress and unidirectional flow, such as the ascending aorta, are protected against atherogenesis as seen by quiescent endothelial cells that form a strong barrier and appear elongated in the direction of flow. Vessels subjected to disturbed flow, such as the coronary bifurcations, carotid sinus, and aortic arch, are more susceptible to atherogenesis. The endothelial cells at these sites are pro-inflammatory, have compromised barrier integrity, and exhibit a "cobble-stone" morphology. The structure and function of the vessel wall are thus greatly dictated by applied hemodynamic forces, as well as pathogenic remodeling observed in atherosclerosis (Davies *et al.*, 2005).

A number of genes and proteins have been previously demonstrated to contribute to mechanosensing mechanisms in endothelial cells from in vivo models to cultured cells subjected to differing flow waveforms (Dai *et al.*, 2004; Civelek *et al.*, 2011; Qiao *et al.*, 2016; Khan *et al.*, 2018; Niu *et al.*, 2019). Mechanical forces have been shown by these and other studies to regulate gene expression in a number of contexts such as substrate stiffness, stretch, and shear stress. There are believed to be mechanosensors such as proteins in the cytoskeletons or plasma membranes that respond to tension or other forces and initiate an intracellular response (Noguchi and Jo 2011). In vivo and in vitro investigations have established the response of a number of

mechanosensitive genes in vascular endothelium (Dai *et al*., 2004; Qiao *et al*., 2016). Among these genes is Phospholipid Phosphatase 3 (PLPP3 or PPAP2B) (Wu *et al*., 2015). Hemodynamic flows dynamically modulate endothelial PLPP3 expression through transcriptional and post-transcriptional mechanisms through Krüpple-like factor 2 (KLF2) and microRNA-92a (miR-92a) respectively. Vascular integrity is promoted though unidirectional flow-induced upregulation of PLPP3 which also promotes a quiescent endothelial phenotype (Wu *et al*., 2015). This mechanism is conserved in mammals from mice to pigs and is present in humans. Mechanosensitive genes such as PLPP3 are critical for driving endothelial functions and understanding the genetic basis of atherosclerosis growth as well as restriction.

1.3 Mechanosensitive Transcription Factors in Endothelium

Transcription factors have been proposed to be critical to endothelial flow sensing (Niu *et al*., 2019). Transcription factors KLF2 and KLF4 are among the most studied mechanosensitive transcription factors in endothelial cells (Dekker *et al*., 2005; Dekker *et al*., 2006; Villareal *et al*., 2010; Sungwung *et al*., 2017; He *et al*., 2019). KLF2 and KLF4 are direct regulators of Nitric Oxide Synthase 3 (NOS3), which is critical for vascular tone (Zhou *et al*., 2012). Knockout mouse models have suggested redundant roles for KLF2 and KLF4 as they share many downstream target genes (Sangwung *et al*., 2017). These two factors are also protective against atherosclerosis and inhibit pro-inflammatory gene expression through competition of histone acetyltransferase p300 (p300) with nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) subunits in athero-prone smooth muscle cells (Atkins *et al*., 2008; Sweet *et al*., 2018). Another important target of KLF2 includes PLPP3, which has been demonstrated to be induced

by unidirectional flow in vascular endothelium in athero-protected endothelium and associated with CAD implicated by GWAS (Wu *et al*., 2015).

Disturbed flow preferentially activates distinct networks of genes, governed by canonical pro-inflammatory transcription factors such as Activator Protein 1 (AP-1) and NFκB (Niu *et al*., 2019). AP-1 and NFκB transcription factors have been known for decades to be activated by disturbed flow to promote endothelial activation (Davies, *et al*. 1994). These factors regulate many leukocyte adhesion molecules including chemokine (C-C motif) ligand 2 (CCL2), E-selectin (SELE), and vascular cell adhesion molecule 1 (VCAM1) (Nogochi *et al*., 2011). Unidirectional flow can protect against tumor necrosis factor alpha (TNFα)-induced inflammation by repressing AP-1 and NFκB (Partridge, *et al*., 2007). Disturbed flow induces hypoxia-inducible factor 1-alpha (HIF-1α), which has been shown to regulate endothelial metabolism under disturbed flow both in vitro as well as in vivo in mouse, pig, and human models (Wu *et al*., 2017). HIF-1α also regulates many inflammatory processes in vascular endothelium and promotes atherogenesis (Feng *et al*., 2017). Subjecting endothelial cells to disturbed flow induces HIF-1α expression, which is also seen in human and animal tissues (Wu *et al*., 2017). HIF-1α up-regulates many enzymes in the glycolysis pathway. As a result, endothelial cells are highly glycolytic and disturbed flow is critical for the metabolic reprogramming via modulation of gene expression via HIF-1α.

Chromatin immuno-precipitation with whole genome sequencing (ChIP-seq) is used to study DNA-protein interactions, notably transcription factors and histone post-translational modifications. ChIP-seq utilizes antibodies against DNA binding proteins to enrich and detect sequences bound by the protein of interest via PCR and genome sequencing. Endothelial-specific enhancers have been identified using ChIP-seq data for enrichment of histone modifications of

H3K27ac and H3K4me2 present at regulatory chromatin (Linnemann *et al.*, 2011; Hogan *et al.*, 2017). These enhancers have been shown to be critical for regulating endothelial identity genes as well as activating transcription for genes induced by extracellular signaling molecules such as cytokines and lipids (Linnemann *et al.*, 2011; Hogan *et al.*, 2017). ChIP-seq data of AP-1 and GATA-binding factor 2 (GATA2) transcription factors in human umbilical vein endothelial cells (HUVEC) co-occupy many sites in the endothelial genome and regulate many endothelial-specific cis-regulatory elements, such as enhancers (Linnemann *et al.*, 2011). ETS Related Gene (ERG), JUN, GATA2, and other transcription factors occupy sites in the endothelial genome near many endothelial-specific genes (Hogan *et al.*, 2017). It is interesting to note that in these two independent reports AP-1 and GATA proteins are highly enriched at many endothelial-specific enhancers and regulate key endothelial genes (Linnemann *et al.*, 2011; Hogan *et al.*, 2017). NFκB, interferon regulatory factor 1 (IRF1), and nuclear factor erythroid 2-related factor 2 (NRF2) are signal-dependent transcription factors that respond to TNF-α, Interleukin 1β (IL1β), and Oxidized 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphocholine (OxPAPC) treatments, respectively (Hogan *et al.*, 2017). This suggests a model whereby sets of transcription factors cooperate to establish endothelial cell-specific enhancers, such as ERG, or bind de novo enhancers that are activated in response to a stimulus, such as NFκB in response to TNF-α signaling (Hogan *et al.*, 2017). ERG has also been shown to establish many strong enhancers in endothelial cells that regulate endothelial-specific genes (Kalna *et al.*, 2019). Induction of endothelial activation using TNF-α has been shown to drive de novo enhancer activation by NFκB to regulate many pro-inflammatory genes that drive atherogenesis (Brown *et al.*, 2014). These studies have validated that a small number of transcription factors, including AP-1 and GATA proteins, are among the most important for endothelial-specific gene regulation

from lineage determination, metabolism, and inflammatory processes. The main limitation of these ChIP-seq experiments is that they were performed in cultured endothelium under static conditions, which is not physiological. These studies have also been limited in that some factors including KLF2, KLF4, and HIF-1α lack suitable antibodies for ChIP-seq experiments. Nonetheless the importance of these transcription factors in mediating key endothelial functions via cis-regulatory elements has been demonstrated, validated, and reproduced through various approaches.

1.4 Genetics in Atherosclerosis

Coronary artery disease heredity estimates vary from 40-60% suggesting a complicated interplay between genetic and lifestyle risk factors (Watkins and Farrall, 2006; Khera and Katherisan, 2017; Inouye *et al*., 2018). Genetic variants have a range of penetrance from determinant to a milder risk factor for a resulting phenotype. Mendelian diseases have a high penetrance generally caused by a change in the structure or function of a protein (McCarthy *et al*., 2008). A prime example is familial hypercholesterolemia. Mutations in the LDL receptor impair the ability of the cell to endocytose and metabolize extracellular lipids from the blood and results in extremely high levels of circulating LDL and precocious CAD, especially in patients with two mutant alleles (Brown and Goldstein, 1974). The first instance of a hereditary CAD identified was caused by a 5 kilobase deletion of the LDL receptor, which eliminated several exons of the transcript resulting in a non-functional protein, high circulating LDL, and severe CAD symptoms in human patients (Lehrman *et al*., 1985). In these cases rare variants were identified in patients with rare disease through linkage analysis, but the same method cannot be

applied to detect common variants that affect common CAD symptoms affecting millions of individuals.

GWAS are well suited for the investigation of complex disease as these studies involve recruitment of large cohorts split into cases (affected) and controls (unaffected) (Tam *et al*., 2019). GWAS with large cohorts allow for the statistical power to detect rare variants with small effect sizes (Tam *et al*., 2019). GWAS of coronary artery disease have identified significantly associated loci that are present in common, non-familial CAD (Samani *et al*., 2007; Schunkert *et al*., 2011; Dichgans *et al*., 2014; Nikpay *et al*., 2015; Howson *et al*., 2017; Van der Harst *et al*., 2018) The number of significant loci has grown from 9 to 161 in just over a decade (Samani *et al*., 2007; Van der Harst *et al*., 2018). In order to accurately correlate cardiovascular phenotypes with genotype, it is necessary to measure and asses the subject's cardiovascular health. A person's cardiovascular disease fitness and history are based on a number of definitions of CAD such as: self-reported CAD events, myocardial infarction, strokes, transluminal balloon angioplasty, and other therapeutic operations (van der Harst and Verweij, 2018). The summation of these factors and events thus defined the CAD-individuals for CAD GWAS. Persons placed in the control group had no personal or familial history of any CAD-related events or procedures that applied to individuals in the disease population and were therefore used for comparison. The GWAS analysis then probes for variants that are over-represented or under-represented in each population to find polymorphisms that associate with disease predisposition. Each individual is genotyped for millions of single nucleotide polymorphisms (SNPs) and thus even rare variants can be detected. A statistical test is then performed to find enrichment of polymorphisms in either the case or control population. The significant results of a GWAS typically must pass the genome-wide significance threshold of 5e-8 to account for multiple hypothesis testing (Tam *et*

al., 2019). This approach has identified thousands of significant of SNPs in hundreds of human diseases and traits (Visscher *et al*., 2017).

SNPs are nucleotides that vary between human populations around the world. SNPs represent only about 1% of the genome, and the majority reside in non-coding regions. Natural genetic variation has been proposed to be a major contributor to complex disease risk (Visscher *et al*., 2017). With the majority of SNPs residing in non-coding regions that means only a small fraction of the polymorphisms change coding genes or proteins. Genes and their expression are tightly controlled in order to ensure appropriate spatio-temporal localization of a protein within a cell, tissue, or organ. There have been efforts to correlate SNPs with gene expression, which are known as expression quantitative trait loci (eQTLs). Systematic approaches to look at expression of genes across human tissues suggest that many eQTLs are specific to certain tissues (Aguet *et al*., 2017).The genetic contribution to common diseases is believed to be driven more by gene regulation, rather than changes to the coding regions of genes themselves (Visscher *et al*., 2017, Liu *et al*., 2019). Due to the function of genes, perturbations of their expression patterns or quantities have consequences for cellular and organismal phenotypes (Visscher *et al*., 2017, Liu *et al*., 2019). SNPs are thus important markers of genomic regions that influence human gene expression, diseases, and traits (Tam *et al*., 2019).

The challenges and limitations of a GWAS are to biologically and mechanistically validate associated loci. This is challenging as the majority of SNPs are non-coding and have a small individual effect size on the phenotype (Tam *et al*., 2019). The majority of SNPs reside in regions of the non-coding genome of open chromatin (Maurano *et al*., 2012). Additionally, many SNPs are also eQTLs (Nicolae *et al*., 2010). These two lines of evidence thus suggest that causal variants in complex human disease lie in gene regulatory elements that affect the level of gene

expression as opposed to changes in the genes themselves seen in monogenic Mendelian diseases.

The first case to demonstrate that interpretation of GWAS results can be difficult can be seen in genetic studies of obesity. The human gene fat mass and obesity-associated protein (FTO) harbors the most significantly identified SNPs associated with obesity within intron 1 (Smemo *et al.*, 2014). Due to these findings, FTO was targeted for understanding human obesity (Fischer *et al.*, 2009; Church *et al.*, 2010). However, definitively and mechanistically linking an enhancer to its target gene requires both gene expression and chromatin conformation data to observe a direct interaction between regulatory elements and corresponding changes in expression of the target gene. Combining chromatin conformation capture (3C) and RNA-seq approaches led to identification of the distal gene Iroquois homeobox 3 (IRX3), but not FTO, to be the causal gene related to inherited obesity via a gene enhancer identified from GWAS (Smemo *et al.*, 2014). This was an important study for appreciating the need to consider gene expression and regulation within the context of a 3D genome.

To date, 161 loci have been identified from CAD GWAS, yet there remain few functional studies mechanistically linked these variants to a human phenotype (van Der Harst *et al.*, 2018). Genotypes at the 1p13 locus were identified from GWAS with variants significantly associating with myocardial infarction and LDL levels (Samani *et al.*, 2007; Schunkert *et al.*, 2011). These variants correlated with expression of the gene Sortilin 1 (SORT1) (Musunuru *et al.*, 2010). This expression difference was specific to the liver. Modulation of SORT1 in vitro and in vivo affected LDL levels by altering hepatic secretion of very low-density lipoprotein (VLDL) in humans. Gene reporter experiments with DNA sequences from the 1p13 locus were significantly affected by the alleles harbored. The differences in DNA sequence had different affinities for a

transcription factor CCAAT-enhancer-binding protein (C/EBP), which could directly up-regulate SORT1 expression. This study delineated a functional genomic mechanism by which genotype affected liver gene expression through a transcription factor that had consequences on lipid metabolism and CAD risk (Musunuru *et al*., 2010).

Endothelial-specific enhancers have been identified from CAD GWAS that are affected by genotype. The single most significantly associated locus with CAD is found at the 9p21 locus (McPherson *et al*., 2007; Schunkert *et al*., 2011; Dichgans *et al*., 2014). This locus is also associated with diabetes, a known risk factor for CAD. Using transcriptomic data, 9 different enhancers were identified in the 9p21 locus specifically in endothelial cells (Harismendy *et al*., 2011). One enhancer in particular harbored 33 GWAS SNPs associated with CAD risk. Variants in this enhancer altered binding patterns for transcription factor signal transducer and activator of transcription 1 (STAT1). STAT1 binding was induced by treatment with interferon gamma (IFNγ). Changes in variants and sequence at this enhancer altered expression of a number of genes including cyclin-dependent kinase inhibitor 2A and B (CDKN2A/B), interferon alpha 21 (IFNA21), and S-methyl-5'-thioadenosine phosphorylase (MTAP) (Harismendy *et al*., 2011).The researchers further performed a chromatin conformation capture assay that was able to detect physical interaction between the identified enhancer and the promoters of the downstream genes (Harismendy *et al*., 2011). This study provided a genetically encoded mechanism by which interferon induction of STAT1 regulated inflammatory pathways relevant to CAD pathogenesis through direct, physical between an enhancer and multiple target genes (Harismendy *et al*., 2011). The study added a novel layer of evidence by showing physical contact between promoters and enhancers in addition to gene reporter assays and eQTL. This highlighted the importance of considering chromatin conformation in enhancer function, activity, and regulation.

A more recent example of a long-range interaction between a gene and its regulatory element is seen in endothelin 1 (EDN1) regulation in aorta (Gupta *et al*., 2017). The SNP rs9349379 is associated with several phenotypes including CAD, migraine headache, cervical artery dissection, fibromuscular dysplasia, and hypertension (Gupta *et al*., 2017). This SNP lies more than 600 kb away from EDN1 but was causally linked to EDN1 expression in vascular endothelium using human transcriptomic data, chromatin marks, chromosome conformation, and gene editing. Use of gene editing technologies allows for investigation of gene expression and regulation within a cellular context, which is an improvement over plasmid-based reporter assays.

GWAS have identified chromosome 1p32.2 as one of the most strongly associated loci with susceptibility to CAD and IS (Schunkert *et al*., 2011; Dichgans *et al*., 2014). Within this locus resides phospholipid phosphatase 3 (PLPP3). PLPP3 is a phosphatase that hydrolyzes the phosphate group on lysophosphatidic acid (LPA) (Panchatcharam *et al*., 2014; Wu *et al*., 2015). LPA is an endogenous bio-active lipid. LPA is a ligand for LPA receptors (G-protein coupled receptors) present on many cell types including endothelial cells (Yung *et al*., 2014). These receptors have varying affinities for LPA and different G proteins. Downstream of LPA signaling, induced by LPA receptors, in endothelial cells are pathways that control inflammation, cell morphology, and leukocyte adhesion (Panchatcharam *et al*., 2014; Yung *et al*., 2014; Wu *et al*., 2015). The inflammatory pathway leads to the production of cytokines IL6 and IL8 as well as chemokines such as CCL2 in endothelial cells. The inflammatory response also leads to increased cell adhesion markers such as E-selection, P-selectin, and VCAM1 (Wu *et al*., 2015). Since PLPP3 inhibits this signaling, it decreases endothelial inflammation and interaction with circulating blood cells (Wu *et al*., 2015).

PLPP3 has been previously shown to be mechanosensitive in endothelial cells (Wu *et al.*, 2015). At the transcript and protein levels in mouse, pig, and human models, PLPP3 is induced in human aortic endothelial cells (HAEC) in response to unidirectional flow and in the descending thoracic aorta from human and animal tissues. PLPP3 has also been shown to be required for normal development, particularly in vascular development, and promoting endothelial integrity (Panchatcharam *et al.*, 2014). In addition, PLPP3 is a major regulator contributing to the quiescent and elongated phenotype of vascular endothelium under athero-protective unidirectional flow (Wu *et al.*, 2015). Moreover, eQTL mapping showed that CAD-protective allele at 1p32.2 is associated with increased PLPP3 expression in an endothelium-specific manner (Wu *et al.*, 2015). However, whether genetic variants and mechanosensing mechanisms converge on PLPP3 expression is unclear. In addition, causal SNPs at locus 1p32.2 remain unknown.

1.5 Enhancer regulation of gene expression in endothelial cells

The genetic code in humans consists of billions of base-paired nucleotides comprised of regions that code for proteins and regulatory elements. The genetic code is packaged and condensed into chromatin by histone proteins (Heintzman *et al.*, 2009; Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011). The histone proteins and other chemical modifications affect the expression of genes at a particular locus (Heintzman *et al.*, 2009; Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011). Nucleosome-free chromatin is open to polymerases and allows for active transcription (Klemm *et al*, 2019). Nucleosome-dense chromatin is sterically hindered and closed to polymerase enzymes. Transcription factors are key mediators to establishing regulatory chromatin. Many genes are constitutively expressed at loci that are open. Other genes are dependent on signals and cellular context (Heinz *et al*, 2015). Induction of these signal-

dependent genes is often mediated by transcription factor proteins and accompanying changes in the chromatin itself.

These sequences are bound by transcription factors and chromatin remodeling proteins to directly control the expression of a target gene (Spitz *et al*., 2012). Cis-regulatory elements, often called enhancers, control gene expression in response to lineage or signal dependent transcription factors (Heinz *et al*., 2015). An enhancer regulates a gene independent of its proximity to a target gene, unlike a gene promoter (Musunuru *et al*., 2010; Harsimendy *et al*., 2011; Smemo *et al*., 2014). It is proposed that promoter-enhancer interactions are where transcription factors and the pre-initiation complex of RNA polymerase may converge and drive gene expression (Schoenfelder *et al*., 2019). In many cases an enhancer is critical for the spatio-temporal expression of a gene and changes in enhancers can have consequences for human health, including CAD (Montefiori *et al*., 2018).

The main characteristics of an enhancer are a lack of nucleosomes, transcription factor binding, and the presence of post-translational modifications on key residues of histone proteins (Heintzman *et al*., 2009; Creyghton *et al*., 2010; Rada-Iglesias *et al*., 2011). Histones harbor many chemical modifications on key amino acid residues (Creyghton *et al*., 2010). Among the most relevant to enhancers are Lysine 4 and 27 on histone H3 (H3K4me2 and H3K27ac) (Heintzman *et al*., 2009; Creyghton *et al*., 2010; Rada-Iglesias *et al*., 2011). ChIP-seq using antibodies against histone protein modifications and coupled with whole genome sequencing reveals a landscape of active chromatin. Enrichment of H3K27ac and H3K4me2 for example correlates with open chromatin and enhancer activity. These histone modifications are enriched at sites of regulatory chromatin, but their mechanistic role is unclear. It is presumed to involved

protein-DNA interactions and accessibility of transcription factors and regulatory proteins (Heinz *et al.*, 2015).

Nucleosome-free regions are also present at active regulatory elements. Assay of Transposase Accessible Chromatin (ATAC-seq) is a method that directly detects open chromatin (Buenrostro *et al.*, 2013). The transposase enzyme fragments nucleosome-free chromatin and adds in adapter sequences that can be amplified using PCR. The results of these experiments, like ChIP-seq, profile the genome for open chromatin and active regulatory elements (Shoenfelder *et al.*, 2019). ChIP-seq and ATAC-seq experiments have identified thousands of enhancers across many cell types (Encode Consortium 2012; Ernst *et al.*, 2011; Calo and Wysocka 2013). Profiling one cell type from an individual, such as immortalized lymphoblastoid cell lines, is insufficient for understanding the chromatin landscape within an entire organism. While the epigenetic state may be conserved near constitutive, house-keeping genes, there are also many states that are cell type specific for many lineage and signal-dependent genes (Zentener *et al.*, 2011; Heinz *et al.*, 2015).

Due to the organization of the genome, cis regulatory elements, such as enhancers, may be kilobases away. Yet, due to the folding of the genome, may in fact lie in 3D proximity and come into contact with a gene promoter (Boney and Cavalli *et al.*, 2016). Enhancers may be proximal or distal to its target gene (Misfud *et al.*, 2015). Enhancers are critical for the spatial and temporal expression of genes and drive gene expression through direct interaction of proteins to gene promoters and transcription start sites (Romanoski *et al.*, 2010).

A number of genomic sequencing methods have been identified to investigate enhancer structure, function, and regulation (Misfud *et al.*, 2015). Techniques that determine interactions between regions of chromatin are performed by chromatin conformation capture assays (Dekker

et al., 2002; Zhao *et al*., 2006; Dostie *et al*, 2006). These experiments rely on ligation of proximal sequences that have been crosslinked. Detection of these interactions are then assessed by PCR for sequencing methods. An improvement of these methods is designed to detect promoter enhancer interactions. For instance, promoter-capture Hi-C relies on a library of biotinylated probes that match promoter regions of the human genome (Misfud *et al*., 2015). These probes are fixed and then ligated to interacting regions. Paired end sequencing then determines the sequence and direction of the interacting fragments (Misfud *et al*., 2015).

Stretch enhancers also known as super enhancers or nucleosome-free clusters exist in the genome as large stretches of regulatory chromatin (Whyte *et al*., 2013). Super enhancers often regulate multiple genes as opposed to just one (Harsimendy *et al*., 2011; Hnisz *et al*., 2013). Super enhancers are computationally defined as a region with high normalized tag density in a multi-kilobase region (Whyte *et al*., 2013). Transcription factors are strongly present at super enhancers (Hnisz *et al*., 2013, Whyte *et al*., 2013). For instance, NFκB regulates many strong enhancers that drive pro-inflammatory gene expression in endothelial cells (Brown *et al*., 2014). ERG is another transcription factor that establishes super enhancer activity and drives expression of many endothelial identity genes (Kalna *et al*., 2019).

Given the key role of cis-regulatory elements in controlling the spatiotemporal gene expression it is imperative to systematically identify the biochemical mediators, targets, and pathways that they govern. With the approaches available to identify enhancers, it can be appreciated the need to perform these experiments in specific cellular contexts to identify cis-regulatory elements that are selectively activated in response to chemical and physical stimuli.

1.6 Aims of this dissertation to understand genetic mechanisms governing flow-sensitive genes and enhancers in human endothelial cells

CAD is a 21st century epidemic. There is a great need to better understand the genetic risk factors that mediate CAD. Shear stress is a known risk factor for atherosclerosis and regulates gene expression in endothelial cells. It is unclear how genetic variation can mediate shear stress response in human endothelium. PLPP3 is a gene at the interface of CAD genetics and endothelial mechanotransduction, being shown to be flow-sensitive and significantly associated with CAD (Wu *et al*., 2015). The genetic mechanism by which PLPP3 is differentially expressed between human genotypes is unknown. In chapter 2 of this dissertation I present work delineating a pathway where human alleles differentially regulate PLPP3 under shear stress, which affects important endothelial cell phenotypes relevant to human atherosclerosis. The results of this study are important for identifying a genetically encoded mechanosensitive region of the genome that provides a layer of gene regulation in endothelial mechanotransduction.

Enhancers are an important layer of gene regulation at the transcriptional level (Heinz *et al*., 2010; Romanoski *et al*., 2015). Genetic investigation of complex diseases has found many non-coding genetic variants lie within regulatory elements such as enhancers that mediate human phenotypes through. Through the advent of whole genome sequencing it is possible to profile the entire genome and transcriptome. In chapter 3, I present data identifying genes and regions of chromatin that are differentially changed in response to shear stress in human endothelial cells. The results of these experiments are important for understanding regulation of the genome and epigenome in shear stress, genetically encoded mechanisms by which they are regulated, the direct targets of non-coding regulatory elements, and the myriad of human phenotypes that are affected.

CHAPTER 2: Genetic Variant at Coronary Artery Disease and Ischemic Stroke Locus 1p32.2 Regulates Endothelial Responses to Hemodynamics

2.1 Abstract

Biomechanical cues dynamically control major cellular processes, but whether genetic variants actively participate in mechanosensing mechanisms remains unexplored. Vascular homeostasis is tightly regulated by hemodynamics. Exposure to disturbed blood flow at arterial sites of branching and bifurcation causes constitutive activation of vascular endothelium contributing to atherosclerosis, the major cause of coronary artery disease (CAD) and ischemic stroke (IS). Conversely, unidirectional flow promotes quiescent endothelium. Genome-wide association studies (GWAS) have identified chromosome 1p32.2 as strongly associated with CAD/IS; however, the causal mechanism related to this locus remains unknown. Using statistical analyses, assay of transposase accessible chromatin with whole-genome sequencing (ATAC-seq), H3K27ac/H3K4me2 ChIP with whole-genome sequencing (ChIP-seq), and CRISPR interference in human aortic endothelial cells (HAECs), my results demonstrate that rs17114036, a common noncoding polymorphism at 1p32.2, is located in an endothelial enhancer dynamically regulated by hemodynamics. CRISPR-Cas9-based genome editing shows that rs17114036-containing region promotes endothelial quiescence under unidirectional shear stress by regulating phospholipid phosphatase 3 (PLPP3). Chromatin accessibility quantitative trait locus (caQTL) mapping using HAECs from 56 donors, allelic imbalance assay from 7 donors, and luciferase assays demonstrate that CAD/IS-protective allele at rs17114036 in PLPP3 intron 5 confers increased endothelial enhancer activity. ChIP-PCR and luciferase assays show that CAD/IS-protective allele at rs17114036 creates a binding site for transcription factor Krüppel-like factor 2 (KLF2), which increases the enhancer activity under unidirectional flow. These

results demonstrate that a human SNP contributes to critical endothelial mechanotransduction mechanisms and suggest that human haplotypes and related cis-regulatory elements provide a previously unappreciated layer of regulatory control in cellular mechanosensing mechanisms.

2.2 Introduction

Mechanical stimuli regulate major cellular functions and play critical roles in the pathogenesis of diverse human diseases (Jaalouk and Lammerding, 2009). This is especially important in the vasculature, where endothelial cells are activated by local disturbed flow in arterial regions prone to atherosclerosis (Davies *et al*., 2013; Hahn and Schwartz, 2009; Zhou *et al*., 2014; Gimbrone and García-Cardeña, 2016), the major cause of coronary artery disease (CAD) and ischemic stroke (IS). The role of biomechanical forces on the noncoding and regulatory regions of the human genome is unexplored. Recent studies demonstrated that the noncoding, nontranscribed human genome is enriched in cis-regulatory elements (Encode Consortium, 2012). In particular, enhancers are distinct genomic regions that contain binding sites for sequence-specific transcription factors (Ong and Corces, 2011). Enhancers spatially and temporally control gene expression with cell type- and cell state-specific patterns (Heinz *et al*., 2015). Notably, top-associated human disease-associated SNPs are frequently located within enhancers that explicitly activate genes in disease-relevant cell types (Ernst *et al*., 2011). The nature of mechanosensitive enhancers and their biological roles in vascular functions have not been identified.

Atherosclerotic disease is the leading cause of morbidity and mortality worldwide. Genome-wide association studies (GWAS) identified chromosome 1p32.2 as one of the most strongly associated loci with susceptibility to CAD and IS (Schunkert *et al*., 2011; Deloukas *et*

al., 2013; Dichgans *et al*., 2014). One candidate gene in this locus is phospholipid phosphatase 3 (PLPP3; also known as phosphatidic acid phosphatase-type 2B), which inhibits endothelial inflammation and promotes monolayer integrity by hydrolyzing lysophosphatidic acid (LPA) that activates endothelium (Wu *et al*., 2015; Panchatcharam *et al*., 2014). My recent study demonstrated that PLPP3 expression is significantly increased in vascular endothelium by unidirectional flow in vitro and in vivo (Wu C *et al*., 2015). Moreover, expression quantitative trait locus (eQTL) mapping showed that CAD-protective allele at 1p32.2 is associated with increased PLPP3 expression in an endothelium-specific manner (Wu C *et al*., 2015). However, whether genetic variants and mechanosensing mechanisms converge on PLPP3 expression is unclear. In addition, causal SNPs at locus 1p32.2 remain unknown.

Using statistical analyses, assay of transposase accessible chromatin with whole-genome sequencing (ATAC-seq), H3K27ac ChIP with whole-genome sequencing (ChIP-seq), H3K4me2 ChIP-seq, luciferase assays, and CRISPR-based approaches, I report that rs17114036-containing genomic region at 1p32.2 causatively promotes endothelial expression of PLPP3 and governs the athero-resistant endothelial phenotype under unidirectional shear stress by functioning as a mechanosensitive endothelial enhancer. Using human aortic endothelial cells (HAECs) isolated from a cohort of human subjects, I performed transcriptome analyses and chromatin accessibility quantitative trait locus (caQTL) mapping showing nucleotide-specific epigenetic and transcriptomic effects of rs17114036 in humans. Allelic imbalance (AI) assays, ChIP-PCR, and luciferase assays collectively demonstrate that, due to a single base pair change, the CAD/IS-protective allele at rs17114036 confers increased activity of an endothelial intronic enhancer that is dynamically activated by unidirectional blood flow and transcription factor Krüppel-like factor 2 (KLF2). This report elucidates underlying molecular mechanisms related to CAD/IS locus

1p32.2 and linking human disease-associated genetic variants to critical mechanotransduction mechanisms. The molecular insights suggest that human genetic variants provide a layer of molecular control by which cells convert physical stimuli into biological signaling via tissue-specific enhancers.

2.3 Results

2.3.1 Bayesian Refinement and Conditional and Joint Multiple-SNP Analyses Predict That rs17114036 and rs2184104 Are Possible Causal SNPs Located in CAD/IS Locus 1p32.2.

rs17114036 is the tag SNP used in most CAD/IS GWAS (Schunkert *et al*., 2011; Deloukas *et al*., 2013; Dichgans *et al*., 2014) and in eQTL mapping (Wu C *et al*., 2015); however, there are 44 common SNPs in high linkage disequilibrium (LD; r2 > 0.8) with rs17114036, and any of these SNPs could conceivably be a causal variant. To refine the list of possible causal SNPs at the 1p32.2 locus, I conducted two statistical analyses. First, I used a Bayesian statistical approach to assign posterior probabilities and credible sets of SNPs that refine the association signals of GWAS-detected loci (Maller *et al*., 2015). Second, I applied conditional and joint association analyses using summary-level statistics of GWAS data to predict causal variants (Yang *et al*., 2016). I chose to apply these two methods because they rely upon additional, relevant information while eliminating the problematic assumption that rs17114036 regulates PLPP3 due to its proximity. The combination of these analyses does not rule out a single variant, as there may be more than one with significant relevance (Table 2.1). This method also tests the independent associations of these variants, which is essential for disentangling and illuminating the genetic and biological significance. Using Bayes' theorem in the cohort of 45 SNPs at 1p32.2, I identified 15 SNPs with >95% posterior probability to be

causal (Figure 2.1). Using the approximate conditional and joint association analyses, I identified seven 1p32.2-associated SNPs to be possible causal (Figure 2.1). Only two SNPs, rs17114036 and rs2184104, were predicted to be causal by both methods (Figure 2.1).

2.3.2 CAD/IS-Associated SNP rs17114036 Is Located in an Enhancer Element in HAECs.

Both rs17114036 and rs2184104 are located in noncoding regions. To probe the regulatory functions of these two SNPs in vascular endothelium, I performed ATAC-seq as well as H3K27ac and H3K4me2 ChIP-seq in HAECs. ATAC-seq is a high-throughput, genome-wide method to define chromatin accessibility that correlates with precise measures of transcription factor binding (Buenrostro *et al*., 2013). The combination of H3K27ac and H3K4me2 ChIP-seq marks was used to identify active, regulatory chromatin. The intersection of these three methods will give a very rigorous set of possible enhancers within the endothelial genome. The use of both ATAC and ChIP-seq approaches will also lend further evidence of reproducible results as one assay relies on enzymatic treatment of cell nuclei and the other uses an antibody for enrichment of DNA sequences at chromatin with the modified histone. At rs17114036 there is evidence of open chromatin, H3K27ac, and H3K4me2 (Figure 2.2). At rs2184104 there were none of these markers (Figure 2.2). Other SNPs that were bioinformatically filtered lie in a mix of possible enhancers and in areas with little to no signal from ATAC or ChIP-seq (Figure 2.2).

Since the human PLPP3 gene is expressed from the minus strand in the annotated human genome, I use alleles in the minus strand at rs17114036 and rs2184104 in this work. It is important to note that, because CAD/IS risk alleles at rs17114036 (T) and rs2184104 (A) are major alleles in all ethnic groups (70–99% frequency) (Abecasis *et al*., 2012), my experiments, unless specified otherwise, were conducted in HAEC lines from donors who carry major alleles

at rs17114036 and rs2184104. As demonstrated, rs17114036 in the intron 5 of the PLPP3 resides in an enhancer-like element [chr1:56962213–56963412, University of California, Santa Cruz (UCSC) version hg19] identified by ATAC-seq and H3K27ac/H3K4me2 ChIP-seq in HAECs. Encyclopedia of DNA Elements (ENCODE) also reported a DNase hypersensitive site and an H3K27ac/H3K4me1 peak in an ∼1-kb region enclosing rs17114036 in human umbilical vein endothelial cells (HUVECs) (Figure 2.3). Notably, this region does not exhibit enhancer-like marks in other ENCODE cell types, such as K562, GM12878, and NHEK cells (Figure 2.3). In contrast, the other putative causal SNP, rs2184104, is located ∼120 kb downstream of the PLPP3 transcription start site at a location that lacks enhancer-like features. ENCODE data also signify an inactive chromatin domain surrounding rs2184104 (Figure 2.3). Figure 2.2 shows the ATAC-seq and H3K27ac/H3K4me2 tracks in HAECs at 1p32.2 locus. The enhancer activity of chr1:56962213–56963412 was experimentally demonstrated by a luciferase reporter assay (Figure 2.4). Plasmid transfection was first validated in HAECs using electroporation of pmax green fluorescent protein (pmaxGFP)-expressing constructs (Figure 2.5). A 1,200-bp DNA sequence corresponding to human chr1:56962213–56963412 was cloned upstream of firefly luciferase that was driven by a minimal promoter (Figure 2.4). Reporter assays demonstrated that insertion of this putative enhancer region with major allele T at rs17114036 significantly increased the luciferase activity in HAECs (Figure 2.6). No significant enhancer activity was implicated when rs2184104-containing region (chr1:56911623–56912823) was cloned into the same reporter vector compared with the rs17114036-containing region (Figure 2.7). I further cloned the rs17114036-containing region into a luciferase vector that contains the human PLPP3 promoter. Endogenous human PLPP3 promoter led to a 7.9-fold higher luciferase activity in HAECs compared with the vector with minimal promoter (Figure 2.6). Moreover, insertion of

chr1:56962213–56963412 resulted in a 2.14-fold increase in luciferase activity compared with the vector with only PLPP3 promoter (Figure 2.6). Minimal enhancer activities chr1:56962213–56963412 were detected when the constructs were expressed in the nonendothelial cell line HEK 293 (Figure 2.8). ATAC-seq, H3K27ac, H3K4me2 ChIP-seq, and luciferase reporter assays suggest that chr1:56962213–56963412 functions as an enhancer in HAECs.

2.3.3 CRISPR-Based Approaches Identified rs17114036-Containing Region as a cis-Regulatory Element for Endothelial PLPP3 Expression.

To determine the causal role of rs17114036-containing genomic locus in regulating endothelial PLPP3 expression, the bacterial CRISPR-Cas9 system was used to selectively delete an ~66-bp genomic region (chr1:56962783–56962849) enclosing rs17114036 in HAECs using a pair of guide RNAs (Figure 2.9) (Ran *et al*., 2013). I preassembled the Cas9–guide RNA ribonucleoprotein (RNP) complex by incubating guide RNAs with recombinant Streptococcus pyogenes Cas9 followed by the delivery to cells using cationic liposome transfection reagents (Figure 2.9). To improve efficiency of the CRISPR-Cas9–mediated deletion, cells were reverse transfected with the RNP complex four times before flow cytometry to sort single cells (Figure 2.10). Immortalized HAECs (carrying major alleles at rs17114036) with high proliferating capacity were used for single-cell clonal isolation to select a genetically identical cell line. Among 459 HAEC colonies that I grew to confluency, PCR assays detected 17 lines with ~66-bp genetic deletion in PLPP3 intron 5 enclosing rs17114036 (Figure 2.11). DNA deletion was further confirmed by cloning and Sanger sequencing (Figure 2.12). Endothelial PLPP3 expression is significantly reduced in the genome-edited cells compared with teloHAECs that underwent CRISPR-Cas9 treatment and single-cell clonal isolation but showed no sign of

deletion at chr1:56962783–56962849 (Figure 2.13). In addition, deletion of this putative enhancer in PLPP3 intron 5 resulted in an increase of LPA-induced E-selectin expression (Figure 2.14) and leukocyte adhesion (Figure 2.15), in agreement with the anti-inflammatory/adhesive role of endothelial PLPP3 (Wu *et al*., 2015; Panchatcharam *et al*., 2014). Moreover, trans-endothelial electrical resistance detected increased monolayer permeability in rs17114036-deleted HAECs (Figure 2.16), consistent with PLPP3's role in maintaining endothelial monolayer integrity (Wu C *et al*., 2015; Panchatcharam *et al*., 2014). CRISPR interference was recently developed to suppress the activity of cis-regulatory elements without mutating the DNA sequence (Qi *et al*., 2013). Here, I showed that rs17114036-targeted (Figure 2.17) but not rs2184104-targeted (Figure 2.18) guide RNAs significantly reduce PLPP3 mRNA expression in HAECs. These results demonstrate that rs17114036-containing region causatively regulates PLPP3 expression and endothelial functions.

2.3.4 Unidirectional Flow Increases the Enhancer Activity in Vascular Endothelium.

Given the critical role of hemodynamics in controlling endothelial PLPP3 transcription (Wu *et al*., 2015), I tested whether shear stresses regulate the enhancer activity. ATAC-seq and H3K27ac ChIP-seq were conducted in HAECs subjected to "athero-protective" unidirectional flow representing wall shear stress in human distal internal carotid artery or "athero-susceptible" flow mimicking hemodynamics in human carotid sinus (Figure 2.19). ATAC-seq captured increased open chromatin in HAECs under unidirectional flow compared with cells under disturbed flow. H3K27ac ChIP-seq indicated an increased histone acetylation in HAECs under unidirectional flow (Figure 2.20). In addition, genetic deletion of the rs17114036-containing region by CRISPR-Cas9 significantly impaired unidirectional flow-induced PLPP3 expression in

HAECs (Figure 2.21). Moreover, histone acetylation at chr1:56962213–56963412, measured by H3K27ac ChIP-PCR, was increased in control HAEC but not in rs17114036-(biallelic) deleted cells when subjected to 24 h of unidirectional flow (Figure 2.22). These results collectively demonstrate that enhancer activity at this locus is dynamically activated by the athero-protective unidirectional flow to regulate PLPP3 in human endothelial cells.

2.3.5 CAD/IS-Protective Allele C at rs17114036 Confers a Higher Enhancer Activity

GWAS have linked the minor allele C at rs17114036 at 1p32.2 to reduced CAD/IS susceptibility (Schunkert *et al.*, 2011; Deloukas *et al.*, 2013; Dichgans *et al.*, 2014), and eQTL mapping described increased PLPP3 expression in HAECs with minor allele C (Wu *et al.*, 2015). Here, I investigated the genotype-dependent effect of rs17114036 on the enhancer activity of chr1:56962213–56963412 by ATAC-seq and luciferase assays. In addition to HAEC lines carrying major (risk) allele T at rs17114036, I conducted ATAC-seq in HAECs isolated from donors who are heterozygous (T/C; ∼20% of Europeans) at rs17114036, allowing us to perform caQTL mapping (van de Geijn *et al.*, 2015). caQTL was recently developed to detect between-individual signaling in cis-regulatory elements as a function of genetic variants (Figures 2.23 and 2.24) I performed ATAC-seq analysis in four HAEC lines heterozygous at rs17114036 under 24-h unidirectional flow to perform open chromatin AI analysis. In all four selected HAEC lines heterozygous at rs17114036, unidirectional flow increases ATAC-seq peaks in the proposed enhancer region in PLPP3 intron 5, in agreement with increased ATAC-seq reads in rs17114036-containing region (Figure 2.24). ATAC-seq detected significantly increased numbers of reads corresponding to rs17114036-containing region in HAEC lines that contain one CAD-protective allele (T/C) compared with HAECs from donors homozygous of CAD risk allele (T/T),

supporting increased chromatin accessibility associated with C allele at rs17114036 (Figure 2.25). In addition, RNA-seq analysis in these cells demonstrates that there is a strong correlation between enhanced chromatin accessibility in rs17114036-containing region and increased mRNA levels of PLPP3 in HAECs, further suggesting that chr1:56962213–56963412 functions as an enhancer in promoting endothelial PLPP3 transcription (Figure 2.27). In addition to the ATAC-seq experiments in HAECs homozygous at rs17114036 under flow, AI analysis showed an enrichment of ATAC reads from the chromosome harboring the protective C allele (Figure 2.29). In contrast, ATAC-seq detected no AI at rs6421497, a common SNP in high linkage disequilibrium (LD) with rs17114036 (Figure 2.26). Moreover, ATAC-seq experiments in HAEC lines heterozygous at rs17114036 further allow us to determine whether the chromosome with C at rs17114036 exhibits higher chromatin accessibility at chr1:56962213–56963412 compared with the chromosome with T allele. This is achieved by AI analysis, which assigns next generation sequencing reads overlapping heterozygous sites to one chromosome or the other for allele-specific signals (Figure 2.28) (van de Geijn *et al*., 2015). ATAC-seq detected reads enriched from the C-containing chromosome compared with that from T allele in HAECs heterozygous at rs1711403, further supporting the increased chromosome accessibility associated with C allele at rs17114036. Lastly, luciferase assays were conducted to support the genotype-dependent enhancer activity of chr1:56962213–56963412. Replacement of T allele with C allele led to a much higher luciferase activity (~5.2-fold, C vs. T) in endothelium (Figure 2.30). Polymorphisms (A or G) at rs2184104 had no effect of the chr1:56911623–56912823 in the luciferase assay (Figure 2.31). Taken together, these results demonstrate that CAD-protective allele C at rs17114036 confers a higher enhancer activity of chr1:56962213–56963412 to promote PLPP3 expression in endothelial cells.

2.3.6 CAD/IS-Protective C Allele at rs17114036 Promotes Flow-Induced, KLF2-Mediated Enhancer Activity

I further examined whether the genetic variants at rs17114036 modulate the flow-induced enhancer activity of chr1:56962213–56963412. Luciferase assays detected an increase of the enhancer activity of chr1:56962213–56963412 (with protective C allele at rs17114036) in cells under 18-h unidirectional flow compared with disturbed flow (Figure 2.32), while no significant increase of the enhancer activity by unidirectional flow was detected with the risk T allele (Figure 2.33). The protective allele C at rs17114036 creates a CACC box that is a binding site for KLF2, which mediates the flow sensitivity of a cohort of endothelial genes, including PLPP3 (Wu *et al*., 2015, Dekker *et al*., 2006; SenBanerjee *et al*., 2004; Parmar *et al*., 2006; Dekker *et al*., 2005). I then tested whether KLF2 dynamically activates this rs17114036-containing enhancer and if rs17114036 alleles impact KLF2-mediated enhancer activity. The affinity of KLF2 to the rs17114036-containing locus was determined by KLF2 ChIP-PCR assays in HAECs carrying a protective allele at rs17114036 (T/C), showing a physical binding of KLF2 to the rs17114036-containing region and the CACC site in the PLPP3 promoter (Figure 2.34). Enhancer activities of chr1:56962213–56963412 were further determined in HAECs as a function of KLF2 expression. Constructs of enhancer (chr1:56962213–56963412) and PLPP3 promoter were co-transfected with KLF2-overexpressing plasmids. Luciferase assays detected a 2.9-fold increase of luciferase activity in the T allele-containing construct as the result of KLF2 overexpression (Figure 2.35). Moreover, KLF2 overexpression led to a 4.7-fold increase of luciferase activity when T allele was substituted by the protective allele C at rs17114036 (Figure 2.35). Collectively, KLF2 ChIP-PCR and luciferase assays demonstrate that CAD/IS-protective

allele C at rs17114036 confers a higher KLF2-dependent enhancer activity of chr1:56962213–56963412 in vascular endothelium (Figure 2.36).

2.4 Discussion

Although it is proposed that genetic and environmental factors jointly influence the risk of most common human diseases, the interplay between genetic predisposition and biomechanical cues at the molecular level is poorly understood. The biology underlying the majority of CAD and IS GWAS loci remains to be elucidated (Nurnberg *et al*., 2016). Most of the CAD and IS SNPs reside in the noncoding genome. Gupta *et al*. 2017 recently reported that the noncoding common variant at rs9349379, implicated in CAD by GWAS, regulates endothelin 1 expression in endothelium. Atherosclerotic lesions preferentially develop in elastic arteries where vascular endothelial cells are activated by local disturbed flow (Davies *et al*., 2013; Hahn and Schwartz, 2009; Zhou *et al*., 2014; Gimbrone and García-Cardeña, 2016). As of now, it remains unknown whether disease-associated genetic variants contribute to mechanosensing mechanisms by which cells sense and convert biomechanical stimuli to biological signaling. My results here elucidate the convergence of CAD/IS genetic predisposition and mechanotransduction mechanisms in endothelial PLPP3 expression. Statistical analyses, whole-genome chromatin accessibility/enhancer marks, CRISPR interference (CRISPRi), genome editing, enhancer assays, caQTL mapping, and AI assay collectively demonstrate that CAD/IS locus 1p32.2 harbors a mechanosensitive endothelial enhancer that regulates PLPP3 expression. Moreover, CAD/IS-protective allele at rs17114036 confers an increased enhancer activity that is dynamically regulated by unidirectional flow and transcription factor KLF2 (Figure 2.36).

Dysregulation of mechanosensing mechanisms contributes to the etiology of a wide range of human diseases in cardiovascular, pulmonary, orthopedic, muscular, and reproductive systems (Jaalouk and Lammerding, 2009). The genetic basis of these complex human diseases has been strongly suggested by GWAS, but the interplay between genetic variants and mechanosensing mechanisms has not been investigated. My data provide a line of evidence supporting the genetic regulation of mechanotransduction mechanisms in complex human diseases and suggest an underappreciated role of genetic predisposition in cellular mechanosensing processes.

Transcriptional enhancers orchestrate the majority of cell type-specific patterns of gene expression and play key roles in development, evolution, and disease (Ong and Corces, 2011), which are tightly regulated by mechanical cues (Jaalouk and Lammerding, 2009). My data provide molecular evidence that the noncoding genome actively participates in cellular mechanotransduction mechanisms that are influenced by human genetic variances. In addition to the flow regulation of the specific locus 1p32.2, my results provide a dataset to systematically determine the mechanosensitive chromatin accessibility and putative enhancer regions at the whole-genome scale in endothelial cells. It is important to note that most of the epigenome studies including ENCODE were conducted in cells without physiological or pathophysiological mechanical stimuli, such as HUVECs under static (no flow) conditions (Encode *et al*., 2012). Since major endothelial functions are tightly and dynamically regulated by hemodynamics flow, my whole-genome epigenome profiling in HAECs under athero-relevant flows may benefit future studies to investigate mechanical regulation of the noncoding genome in vascular cells.

Mechanosensitive transcription factors have been proposed as major regulators to determine endothelial functions relevant to atherogenesis. For instance, nuclear factor-κB and HIF-1α mediate gene sets associated with proinflammatory, procoagulant, and glycolytic

endothelial phenotypes under disturbed flow (Lan *et al*., 1994; Khachigian *et al*., 1995; Wu *et al*., 2017; Feng *et al*., 2017), while KLFs and nuclear factor erythroid 2-like 2 regulate gene networks promoting the quiescent endothelial phenotype under unidirectional flow (SenBanerjee *et al*., 2004; Parmar *et al*., 2006; Dekker *et al*., 2005, Huang *et al*., 2017; Fang and Davies, 2017; Li *et al*., 2017; Zhou *et al*., 2012). However, the interaction between flow-sensitive transcription factors and disease-associated genetic predisposition in vascular functions has not been suggested. My results demonstrate that a genetic variant can influence important endothelial functions via a noncoding enhancer region recognized by the mechanosensitive transcription factor KLF2. These results are consistent with emerging evidence showing that top-scoring disease-associated SNPs are frequently located within enhancers explicitly active in disease-relevant cell types (Ernst *et al*., 2011, Hogan *et al*., 2017). Moreover, the data suggest that disease-associated genetic variants, via modulation of transcription factor binding, may regulate the enhancer activities dynamically responding to biomechanical cues that are instrumental to key cellular processes.

GWAS related to atherosclerotic diseases have suggested previously unsuspected loci, genes, and biology involved in lipoprotein metabolism (Nurnberg *et al*., 2016), resulting in the development of new cholesterol-lowering therapies (Sabatine *et al*., 2017). Despite that dyslipidemia is a major systemic risk factor of CAD and IS, atherosclerotic lesions largely initiate and develop at arterial regions of atypical vascular geometry associated with disturbed flow. Previous studies demonstrated that cellular mechanotransduction mechanisms, particularly endothelial responses to hemodynamics, causatively contribute to the focal nature of atherosclerotic lesions (Davies *et al*., 2013; Hahn and Schwartz, 2009; Zhou *et al*., 2014; Gimbrone and García-Cardeña, 2016; Nam *et al*., 2009). My studies here demonstrate that

genetic variants contribute to not only interindividual variation in plasma lipid concentrations (Willer *et al*., 2013) but also, endothelial responses to blood flow. Indeed, genetic variants at rs17114036 predict CAD susceptibility independent of traditional systemic risk factors, such as cholesterol and diabetes mellitus (Schunkert *et al*., 2011; Deloukas *et al*., 2013). Recent GWAS identified 15 new CAD risk loci near genes of key functions in endothelial, smooth muscle, and white blood cells (Howson *et al*., 2017), further highlighting the potential importance of genetic contribution to the arterial wall-specific mechanisms in atherogenesis. My results indicate that CAD genetic predisposition and disturbed flow converge to inhibit endothelial PLPP3 expression and that restoration of endothelial PLPP3 in athero-susceptible regions may provide an attractive approach for future arterial wall-based atherosclerosis therapy complementary to current pharmacological treatments targeting systemic risk factors.

My studies demonstrate that the latest human genetics approaches, such as caQTL mapping (Kumasaka *et al*., 2016), AI assay (van de Geijn *et al*., 2015), CRISPRi (Qi *et al*., 2013), and CRISPR-based assays (Ran *et al*., 2013), are powerful tools to investigate possible genetic contributions to cellular mechanotransduction. Miao *et al*. (Miao *et al*., 2018) recently applied CRISPR-Cas9 to achieve high efficiency of a 10-kb deletion of an enhancer region in bulk HUVECs. Here, I expanded the applications of CRISPR-based techniques to investigate key vascular functions. Isogenic adult aortic endothelial lines subjected to CRISPR-based deletion were successfully selected to determine the causal role of an ~66-bp genomic region in regulating endothelial PLPP3. Nevertheless, one limitation is that I was unable to replace this human SNP at rs17114036 in adult aortic endothelial cells, although I have tried various methods to promote homology-directed repair. This will be the subject of a future study. Nevertheless, caQTL mapping (Kumasaka *et al*., 2016) and AI assay (van de Geijn *et al*., 2015) provide

complementary approaches detecting at the single-nucleotide resolution that CAD/IS-protective allele at rs17114036 confers a higher enhancer activity at the PLPP3 intron 5.

Cellular mechanotransduction is required for physiological control of tissue homeostasis, while abnormal cell response to mechanical forces promotes pathologies of numerous human diseases. Although investigations, such as GWAS, have suggested the genetic basis of complex human diseases, the interplay between genetic predispositions of mechanosensing mechanisms remains virtually unknown. My results identified that CAD-associated genetic variant at rs17114036 interacts with hemodynamics in concert to regulate endothelial PLPP3 expression and consequently, key vascular functions. Moreover, my experiments provide evidence supporting the regulatory role of the noncoding, non-transcribed genome in mechanotransduction mechanisms. In summary, this study demonstrates that human haplotypes and related cis-regulatory elements provide an important layer of molecular control by which cells convert physical stimuli into biological signaling.

2.5 Materials and Methods

2.5.1 Cell culture

Human Aortic Endothelial Cells (HAECs) were procured from Lonza (Allendale, NJ) (CC-2535). HAEC lines that are heterozygous at rs17114036 were from aortic explants of heart transplant donors in the UCLA transplant program. For CRISPR/Cas9-mediated genome editing, I used teloHAECs (ATCC #CRL-4052) that express hTERT and large-T antigen. teloHAECs were used due to their ability to form colonies, which was necessary for generating isogenic cell lines with deletions. Cells were grown in EGM-2 medium supplemented with SingleQuots from Lonza (CC-3156 & CC-4176) and Antibiotic-Antimycotic from Gibco (Grand Island, NY)

(15240062) in a 37°C incubator with 5% CO_2. HAECs were used from passages 6-10. teloHAECs, and CRISPR-edited teloHAECs were used from passages 10-15. THP-1 cells were maintained using RPMI 1640 medium (Gibco) containing 10% FBS (Biowest).

2.5.2 H3K27ac and H3K4me2 chromatin immuno-precipitation with whole genome sequencing (ChIP-seq)

HAECs were washed three times with warm PBS and then trypsinized. Cells were pelleted at 3000 x g for 5 min before being fixed at room temperature with 1% paraformaldehyde in PBS for 10 min and quenched with 125 mM glycine. 1 million cells were used for each ChIP-seq. Cell lysates were sonicated using BioRuptor Pico (Diagenode, Belgium), and then immunoprecipitated using antibodies against H3K27ac (Active Motif, Carlsbad, CA, #39135) or H3K4me2 (EMD Millipore, Billerica, MA, #07–030), bound to a 2:1 mixture of Protein A Dynabeads (Invitrogen #10002D) and Protein G Dynabeads (Invitrogen #10004D). Following immunopreciptation, crosslinking was reversed and libraries were prepared using the same method described previously (Hogan *et al.*, 2017) for RNA-seq beginning with dsDNA end repair and excluding UDG. For each sample condition, an input library was also created using an aliquot of sonicated cell lysate that had not undergone immunoprecipitation. These samples were sequenced on an Illumina HiSeq 4000 and used to normalize ChIP-seq results.

2.5.3 Chromatin accessibility quantitative trait locus (caQTL) mapping and Allelic Imbalance

ATAC-seq was performed as previously described (Buenrostro *et al.*, 2013) using Tn5 transposase (Illumina, San Diego CA). Libraries were sequenced on an Illumina HiSeq 4000 according to manufacturer's specifications by the Genomics Core Facility at the University of

Chicago. The reads were aligned to the UCSC hg19 genome using Bowtie2 (Langmead and Salzberg, 2012).

Chromatin accessibility quantitative trait locus (caQTL) mapping was performed to test for association between genotype at rs17114036 and chromatin accessibility measured by ATAC-seq. I pulled genotypes for HAEC donors from my previous study (Romanoski *et al*., 2010) and imputed linked alleles using IMPUTE2 and SHAPEIT as I published previously (Hogan *et al*., 2017). Association testing between ATAC-seq tags at the rs17114036 enhancer and genotype were performed using the Combined Haplotype Test in WASP (van de Geijn *et al*., 2015). To perform allelic imbalance (AI) analysis that assigns next generation sequencing reads overlapping heterozygous sites to one chromosome or the other, I quantified ATAC-seq tags at the rs17114036 enhancer using HOMER's annotatePeaks function to plot the log2 normalized tags in this region.

To test whether alleles at SNP rs17114036 associated with differences in chromatin accessibility, I used the Combined Haplotype Test (CHT) in the software suite WASP (van de Geijn *et al*., 2015). The CHT test provides one statistic from jointly modeling two components of association between alleles and sequencing reads. The two components of the model are visualized separately in Figures 2.28 and 2.29 but result in one p-value (4.21e-04). The first component of the model takes into account the allelic imbalance at phased heterozygous SNPs (Figure 2.29), whereas the second component measures the total read depth in the target region versus the diploid genotype of the individual (Fig. 2.28). Therefore, the p-value of 4.21e-04 reflects the CHT statistic derived from a joint test of data shown in Figures 2.28 and 2.29.

2.5.4 RNA-seq

Total RNA was isolated from HAECs using the Quick-RNA Micro Prep kit from ZymoResearch (#R1051), including optional DNase I treatment. mRNA was selected through poly-A isolation using Oligo d(T)25 beads (New England BioLabs #S1419S), fragmented, and cDNA was synthesized with the SuperScript III First-Strand Synthesis System (Invitrogen # 18080051). Second strand synthesis was performed using DNA Polymerase I (Enzymatics #P7050L). To generate sequencing libraries, DNA ends were repaired with T4 DNA Polymerase (Enzymatics #P7080L). Six-mer barcoded adapters (BIOO Scientific NEXTflex #514104) were ligated with T4 DNA Ligase (Enzymatics #L-6030-HC- L) and samples were treated with Uracil DNA Glycosylase (UDG) (Enzymatics #G5010L). Libraries were then amplified less than 14 cycles by PCR (Phusion Hot Start II #F549L) and purified (Zymo #D5205) for high-throughput sequencing at the University of Chicago.

2.5.5 Normalization of High-throughput Sequencing Data

Reads from ATAC-seq and ChIP-seq were mapped to the hg19 build of the human genome with Bowtie2 (Langmead and Salzberg, 2012) and RNA-seq reads were mapped with STAR (Dobin *et al.*, 2013). For RNA-seq, ATAC-seq, and ChIP-seq mapped reads were normalized to 10 million tags per experiment and PCR duplicates were removed in HOMER (Heinz *et al.*, 2010). The ATAC-seq and ChIP-seq aligned files were then converted to BAM format and sorted using SAMtools (Li *et al.*, 2009). The files were then filtered to remove unmapped and mitochondrial reads using SAMtools (Li *et al.*, 2009). Peak calling was performed using MACS2 (Zhang *et al.*, 2008). The bedgraph files from MACS2 were converted into BigWig files using HOMER for visualization on the genome browser (Heinz *et al.*, 2010). The images in this manuscript had the peaks smoothed to 3 pixels for aesthetics. Using the RNA-

seq datasets, PLPP3 expression was quantified using the 'analyzeRepeats' function and is expressed as log2 normalized tag counts. Sequencing data generated in this study is available under Gene Expression Omnibus #GSE112340 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112340; token: srqrygwoxhajbgx).

2.5.6 Dual luciferase assay

HAECs and teloHAECs were electroporated using the Neon system (ThermoFisher Scientific). Briefly, the cells were re-suspended in resuspension buffer to have around 40,000 cells per µl. 1.2 µg of eGFP expression plasmid (pmaxGFP) was first used to assess transfection efficiency. For dual luciferase reporter assays, 1 µg of pGL4.23 plasmids carrying firefly luciferase inserted with chr1:56962213-56963412 (T or C at rs17114036, UCSC VERSION hg19) or chr1:56911623-56912823 (A or G at rs2184104, UCSC VERSION hg19) and PLPP3 promoter or minimal promoter at 1 µg/µl and 200 ng of pRL-TK (control reporters) at 1 µg/µl were added per 8.8 µl of cells. The cells were electroporated at 1200 V, for 1 pulse and then were immediately plated onto 12-well culture plates. Media was changed after 4-6 hours to remove dead cells. The peak expression of plasmids was shown at 24 hours post-electroporation and was chosen as the end-point for most luciferase experiments. Cells were collected by adding passive lysis buffer and then put at -80°C overnight. Luminescence was measured by Dual-Luciferase Reporter Assay System (Promega) according to manufacturer's instructions using a Cytation3 plate reader (BioTek).

2.5.7 CRISPR Cas9-mediated deletion of enhancer in teloHAECs

The CRISPR reagents were adapted from the Alt-R system from IDT (IDT, Coralville, IA). The guide RNAs were designed using an online tool at http://crispr.mit.edu/ to minimize off targeting effects using two guides to create a ~66 bp deletion. The guide RNAs were made by

annealing the tracrRNA to the sgRNA. Cas9-guide RNA ribonucleoprotein (RNP) complex by incubating guide RNAs with recombinant S. pyogenes Cas9, followed by the delivery to cells using Lipofectamine RNAiMAX (Thermo). For each successive treatment the reagent amounts were scaled relative to the size of the destination vessel to compensate for the number of cells in the reaction. The volumes for each part of the reaction was increased 4x when treating cells from the 96-well to a 6-well, and 16x when moving from the 6-well to the T-75 flask.

After four treatments with the CRISPR RNP, treated cells were trypsinized to a single cell suspension. Using a FACS AriaII flow cytometer single cells were plated onto 96 well plates containing EGM-2 media with 10% FBS for a total of 576 possible clones to screen. The cells were incubated for 20 days prior to grow to confluence. The cells were then split, and half of each clone was placed onto two new plates to continue growing clones, and to collect DNA for screening. To confirm the CRISPR Cas9-mediated deletion, DNA for screening was collected using 50 µl QuickExtract solution (Epicentre, Madison WI). Mutations were detected using SYBR green (Roche, Indianapolis, IN) and CRISPR Genotyping Screening primers flanking the deletion region. Positive clones were checked with another PCR using Platinum Taq and CRISPR Genotyping TA Cloning primers flanking the deletion. The PCR products for clones were purified using Qiagen PCR clean-up kit (Qiagen, Venlo, Netherlands). The PCR products were ligated into the pGEMT-easy linear vector, cloned, and then submitted for Sanger sequencing to confirm deletions of the target.

(Guide RNA sequences: 5’-TAGTGATATCAACCATTTGACGG-3’, 5’-TGACTTCAGCTCTTGCTGATAGG-3’)

(CRISPR Genotyping Screening F: 5’- TCCTCCACGTTTAGTTGCCA -3’, R: 5’-AAGGAATCCAGGGTGTAACCG -3’)

(CRISPR Genotyping TA Cloning, F: 5’- GGACGCTGGGAATGAGTGAT-3’, R: 5’-ATTGCCCATATCTGCAACCC-3’)

2.5.8 CRISPR Interference (CRISPRi)

A fusion protein of catalytically dead Cas9 (dCas9) fused to KRAB repressor protein was expressed in HAEC using in vitro transcripts. In a 6-well plate, 100 ng of fusion protein and 100 ng of sgRNA were diluted in 50 µl of opti-MEM. 1.5 µl of Messenger MAX was diluted into 50 µl of opti-MEM. The in vitro transcripts and Messenger MAX dilutions were combined and incubated for 10 minutes at RT. The complex was then added to the cells, which were incubated for 8 hours prior to cell lysis, RNA collection, and analysis via qPCR. Non-targeting control guide RNA was purchased from IDT.

(Guide RNA sequences targeted to rs17114036: 5’-GTTGATATCACTAAGTTTTCAGG-3’, 5’-CAAGAGCTGAAGTCAGGCAGTGG-3’)

(Guide RNA sequence targeted to rs2184104: 5’-GGACGACTGCAAACACCAGA-3’)

2.5.9 Leukocyte adhesion assay

Monolayers of teloHAECs were activated using lysophosphatidic acid (LPA) for 3 hours. THP-1 cells at 5-fold number of teloHAECs were pelleted and resuspended in serum-free RPMI and then labeled with 5 µM Calcein AM dye (ThermoFisher) for 30 minutes at 37 °C. The cells were then pelleted and resuspended in serum-free RPMI. Labeled THP-1 cells were then added to the teloHAECs and incubated at 37 °C for 1 hour, with gentle rocking every 30 minutes. The cells were then washed extensively 5 times with warm DPBS. Fluorescence of the Calcien AM dye was then measured on a Cytation 3 (Biotek, Winooski, VT) device in area scanning mode, with gain of 80, and excitation at 492 nm, and emission 550 nm.

2.5.10 Measurement of Transendothelial Electrical Resistance

Cell permeability was evaluated by measuring transendothelial electrical resistance (TER) across CRISPR/Cas9-edited clones on 8 well electrode arrays 8W10E+ (Applied Biophysics) by an electrical cell-substrate impedance sensing system Model 1600R (Applied Biophysics).

2.5.11 Chromatin Immunoprecipitation PCR

Two CRISPR clones, one non-deletion clone and one bi-allelic deletion clone were chosen for ChIP PCR experiments. The CRISPR clones were grown under static or subjected to 24 hours of unidirectional flow. Cell crosslinking and chromatin immunoprecipitation were performed as previously described (Hogan *et al*., 2017). 3 µl of H3K27ac antibody (Active Motif, Carlsbad, CA, #39135) was added to chromatin and incubated for 2 hours for IP. The enhancer was detected by qPCR (F: 5'-GGAGTTCATGTTGGCTGATCT-3', R 5'-TCACAAAGGAATCCAGGGTGT-3'). A total of 4 IPs were performed for each genotype grown under static and UF conditions.

HAECs heterozygous at rs17114036 were transfected with in vitro transcripts of KLF2 with HA tag (Huang *et al*., 2017) for 6 hours before cross-linking, digestion, and immunoprecipitation according to the Pierce Agarose ChIP Kit. Briefly, cells were cross-linked with 1% formaldehyde in growth media (EGM2). After a 10-minute incubation, glycine was added to 125 mM final concentration and, after 5 minutes, the solution was aspirated. The cells were washed with ice-cold PBS twice, before being scraped in 1mL PBS and 10 µL protease inhibitors (Halt Cocktail, Pierce). The cells were pelleted at 3000 x g for 5 min before undergoing MNase digestion at 10 U/µL for 15 min. After recovery of digested chromatin, the solution was immunoprecipitated with anti-HA antibody at 3 µg/ 2 million cells, and with normal

rabbit IgG as control overnight. After binding to agarose beads, the immunoprecipitate was eluted before DNA clean up and then purified DNA was detected by qPCR with PLPP3 enhancer (F: 5'-AGACTAAGACGACGCTCTCC-3', R: 5'- GTGGCACCTACATCATGTTGT -3') and PLPP3 promoter primers (F: 5'-TTGCTAACCCTCACAGAGCA-3', R: 5'-ATCCTGTGACTCTGTGCCTC-3') as well as negative control primers (F: 5'-ATGTGGCCAGAGTGAAACCA-3', R: 5'-TCTACACCCAACAGCCTTCT-3').

2.5.12 mRNA quantitative real-time PCR

Total RNA was isolated using Direct-zol RNA MiniPrep (Zymo Research). mRNA was reverse-transcribed into cDNA using High Capacity cDNA Reverse Transcription kit (Life Technologies). cDNA quantification was performed on LightCycler 480 II (Roche) using SYBR Green I Master for mRNA. Absolute quantification of each gene of interest was normalized to the geometric mean of beta-actin, ubiquitin B, and GAPDH. The following primer sequences were used (IDT, Coralville, IA):

PPAP2B, 5’-CAGCGCCATGCAAAACTACA-3’ 5’-AAAACCCTCGGTGGTAAGGC-3’

SELE, 5’-CCGAGCGAGGCTACATGAAT-3’ 5’-GCCACATTGGAGCCTTTTGG-3’

GAPDH 5’- TGCACCACCAACTGCTTAGC-3’ 5’- GGCATGGACTGTGGTCATGAG-3’

ACTB 5’- TCCCTGGAGAAGAGCTACGA-3’ 5’- AGGAAGGAAGGCTGGAAGAG-3’

UBB 5’- ATTTAGGGGCGGTTGGCTTT-3’ 5’- TGCATTTTGACCTGTTAGCGG-3’

2.5.13 Reagents and antibodies

Anti-HA antibody was from Abcam (ab9110). Lysophosphatidic acid (LPA) was obtained from Santa Cruz Biotechnology (Cat# 22556-62-3). High molecular weight dextran was

obtained from Sigma (Cat# 31392-50G). H3K27ac antibody was purchased from Active Motif (Cat# 39135).

2.5.14 Application of athero-relevant flows

Hemodynamic forces were applied to cultured HAECs in a 6 well plate using a 1° tapered stainless-steel cone held by a computerized stepper motor UMD-17 (Arcus Technology, Livermore CA). The cone rotation recapitulated the shear stress waveform mimicking the disturbed flow at the athero-susceptible human carotid artery or the shear stress profile mimicking the unidirectional flow at the athero-resistant distal internal carotid artery (Dai *et al.*, 2004). The flow device was placed in a 37°C incubator with 5% CO2. HAECs at 100% confluence, maintained in EGM2- medium containing 4% dextran (Sigma-Aldrich, St. Louis, MO, 31392) in 6-well plates, were subjected to unidirectional flow or disturbed flow for 24 hours before being processed.

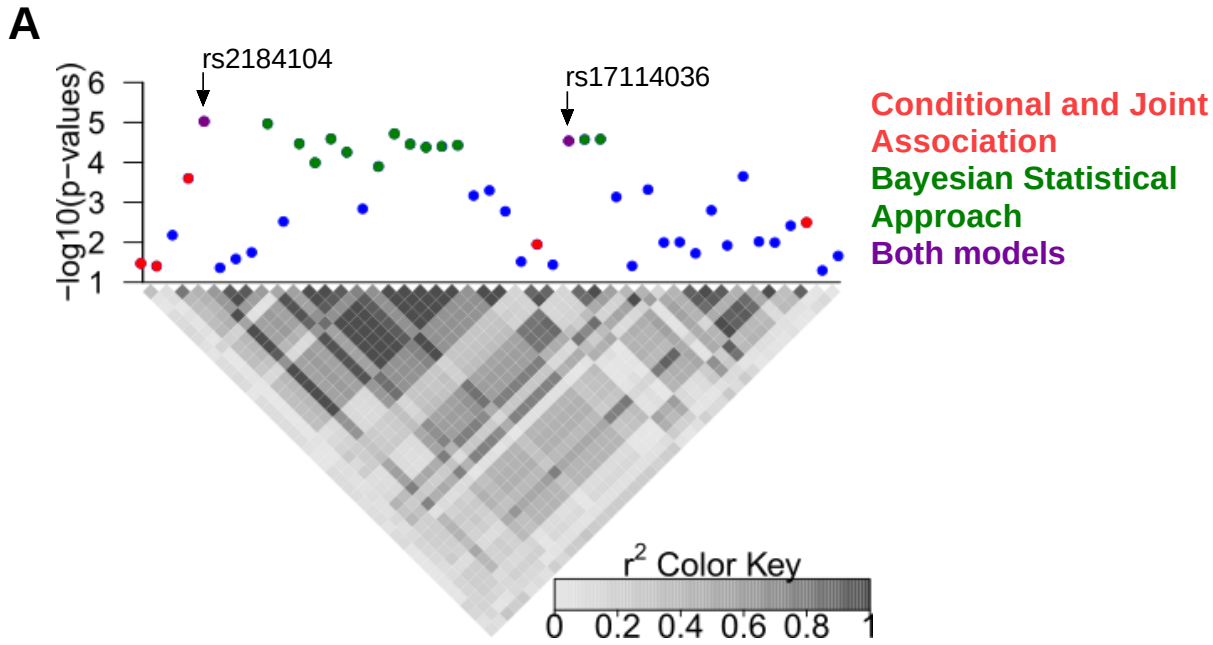

Figure 2.1: Fine mapping strategies at 1p32.2 locus

In silico prediction of causal SNPs in the CAD locus 1p32.2. Heatmap displays the association (−log10P) and LD pattern of a total of 45 common SNPs in the 1p32.2 locus. Green circles indicate SNPs predicted by the Bayesian Statistical Approach. Red circles indicate SNPs predicted by the Approximate Conditional and Joint Association Analysis. Purple circles indicate SNPs predicted by both statistical analyses.

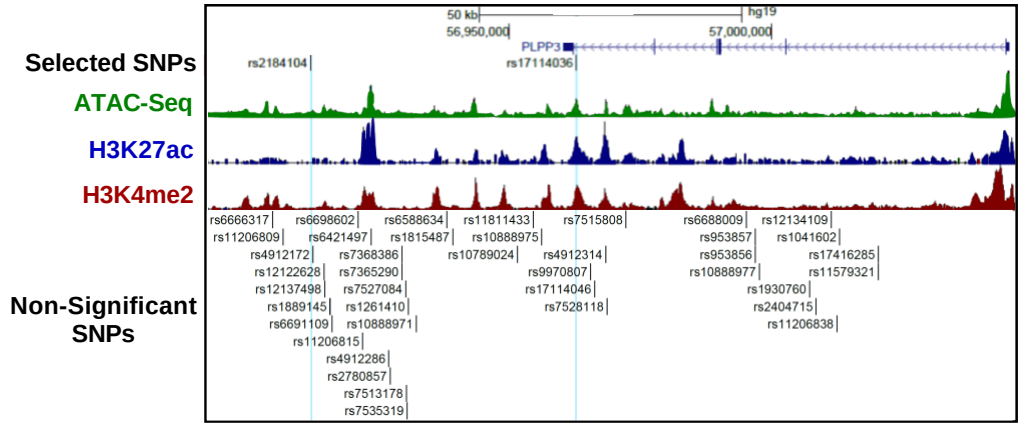

Figure 2.2: Genome browser image of chromatin state at PLPP3 locus with SNPs within high LD of rs17114036

Chromatin accessibility and canonical enhancer marks along with common SNPs in CAD locus 1p32.2 in human aortic endothelial cells. All common SNPs at CAD locus 1p32.2 tested for predicted causality are plotted next to the ATAC-seq, H3K27ac ChIP-seq, and H3K4me2 ChIP-seq data performed in HAEC.

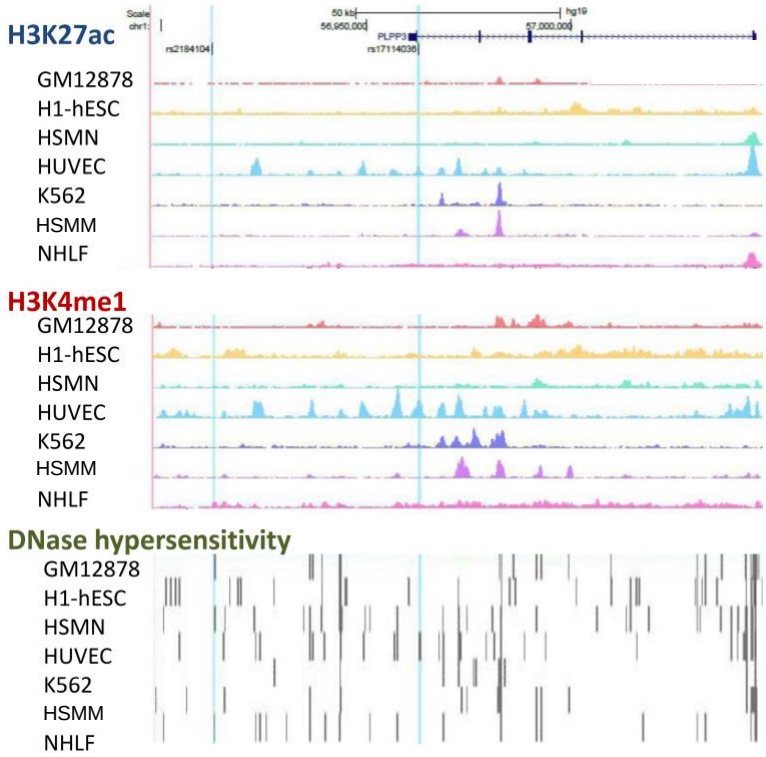

Figure 2.3: Chromatin and histone modifications cross cell lines

Encyclopedia of DNA Elements (ENCODE) studies indicate rs17114036 is located in an enhancer-like element in Human Umbilical Vein Endothelial Cells (HUVEC). H3K27ac ChIP-seq, H3K4me1 ChIP-seq, DNase hypersensitivity data sets curated by the ENCODE project in GM12878, h1-ESC, HSMM, HUVEC, K562, NHEK, and NHLF are plotted at the PLPP3 locus alongside the predicted causative CAD SNPs rs17114036 and rs2184104.

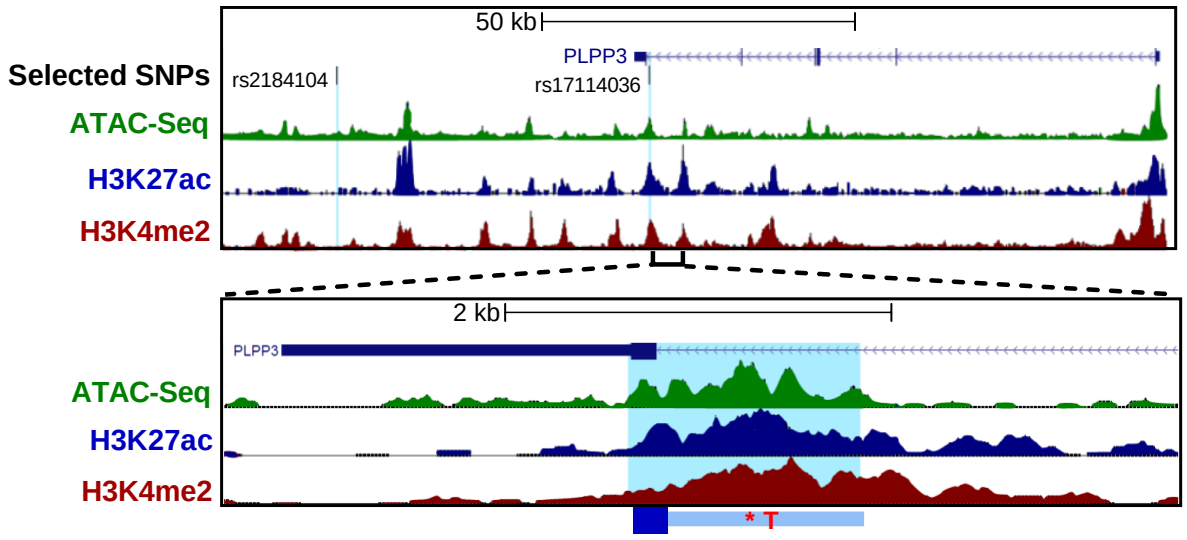

Figure 2.4: Genome browser image of chromatin state at PLPP3 locus

Chromatin accessibility and canonical enhancer marks in chr1:56962213–56963412 region enclosing rs17114036 in HAEC. ATAC-seq and H3K27ac/H3K4me2 ChIP-seq collectively identified an enhancer-like region in chr1:56962213–56963412 in HAECs. All sequencing experiments were performed in duplicate, and the merged tracks are shown. Zoomed in region shows strong presence of enhancer markers that was cloned for further experimentation near rs17114036. There is a lack of signal at rs2184104, suggesting no presence of open, regulatory chromatin.

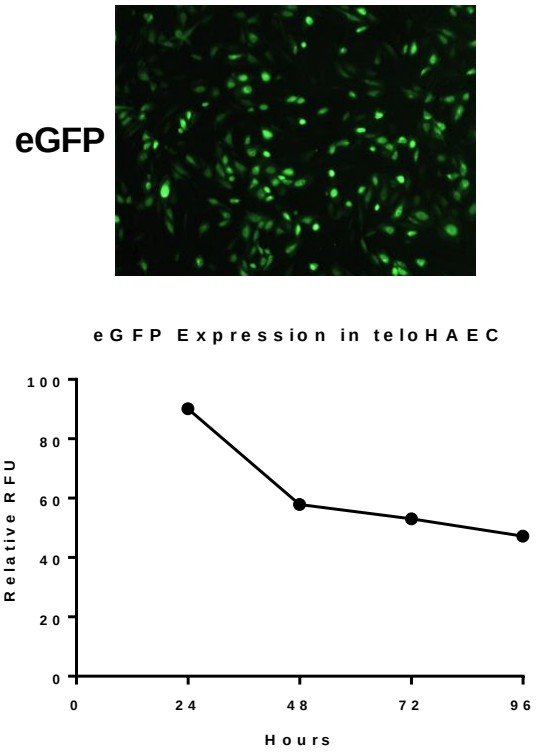

Figure 2.5: Electroporation of plasmids results in high and stable transfection efficiency in teloHAEC

Electroporation results in high transfection efficiency of plasmids in teloHAEC. teloHAEC were transfected with 1.2 µg of eGFP plasmid via electroporation using the Neon® Transfection system (Thermofisher). The cells were visualized under a fluorescent microscope 24 hours post electroporation. GFP measurements were recorded at 24, 48, 72, and 96 hours post-transfection.

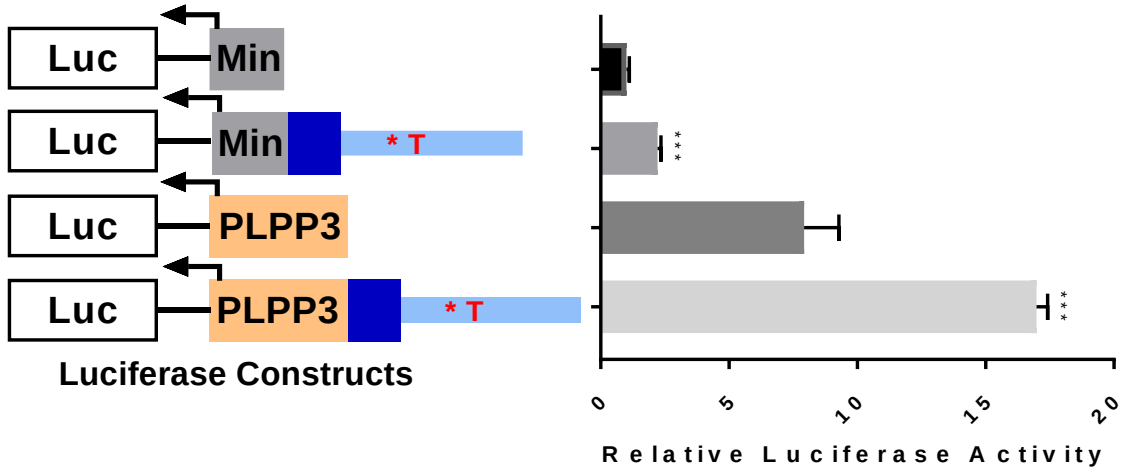

Figure 2.6: Enhancer activity of chr1:56962213–56963412 in vascular endothelium

DNA sequences of chr1:56962213–56963412 were cloned into luciferase reporters (firefly luciferase construct pGL4) that contain a minimal promoter or human PLPP3 promoter. The red asterisks denote the relative position of rs17114036 in the luciferase construct. Dual luciferase reporter assays were conducted in HAECs 24 h after the electroporation-based transfection (using pRL-TK plasmid carrying Renilla luciferase as transfection controls) in HAECs, detecting increased firefly luciferase as the result of insertion of chr1:56962213–56963412. Data represent mean ± SEM. ***P < 0.0005 as determined by Student's t test.

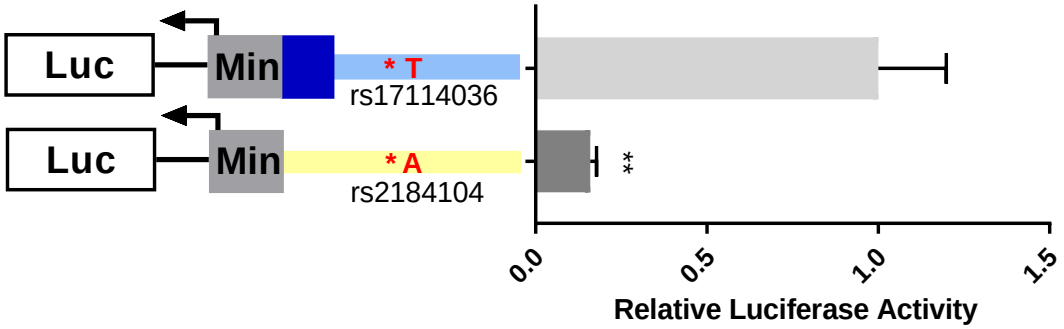

Figure 2.7: Enhancer activity of DNA sequence near rs2184104 in HAEC

Dual luciferase reporter assays show a 6.3-fold difference in luciferase reporter activity in 1.2 kb regions surrounding predicted, causative SNPs rs17114036 and rs2184104. Equal sized genomic regions with either rs17114036 or rs2184104 in the middle were cloned into pGL4.23 luciferase reporter vectors, electroporated into teloHAEC along with pRLTK internal control, and incubated for 24 hours prior to lysate collection and luciferase activity measurement. **p<0.005 as determined by Student's t-test. Data represent mean ± SEM.

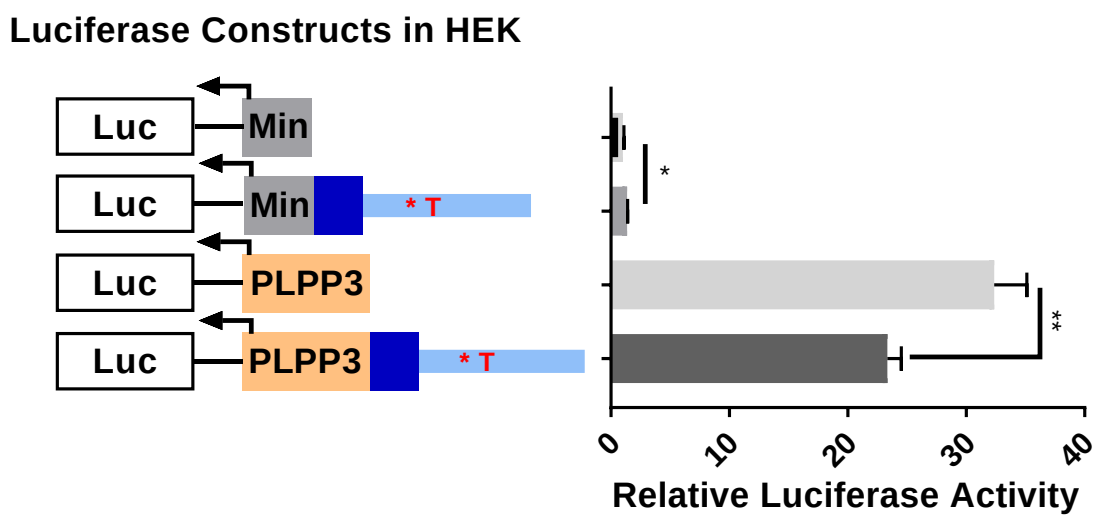

Figure 2.8: Gene reporter activity minimal in HEK cells

Dual luciferase reporter assays demonstrate minimal enhancer activity of chr1:56962213-56963412 in embryonic kidney cells 293 (HEK-293). Luciferase reporters constructs with and without the enhancer as well as with minimal promoter or human PLPP3 promoter were transfected into HEK-293 using lipofectamine and incubated 24 hours before measuring luciferase activity. n=5. *p<0.05, **p<0.005 as determined by Student's t-test. Data represent mean ± SEM.

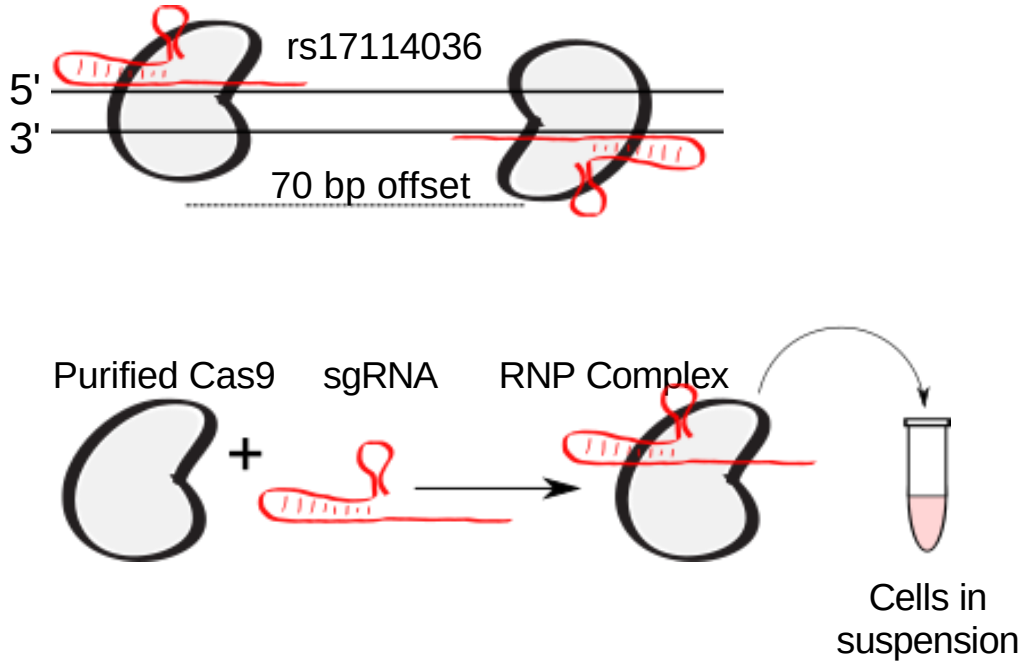

Figure 2.9: CRISPR-Based Genome Engineering using two guide RNAs to create a genomic deletion delivered via RNP transfection into teloHAEC

RNA RNP complex that contains recombinant S. pyogenes Cas9 and two sgRNAs flanking rs17114036. Two guide RNAs were designed to flank rs17114036 and excise the sequence. The RNP complex is created by combining purified Cas9 protein and synthetic guide RNA. The complex is reverse transfected into cells.

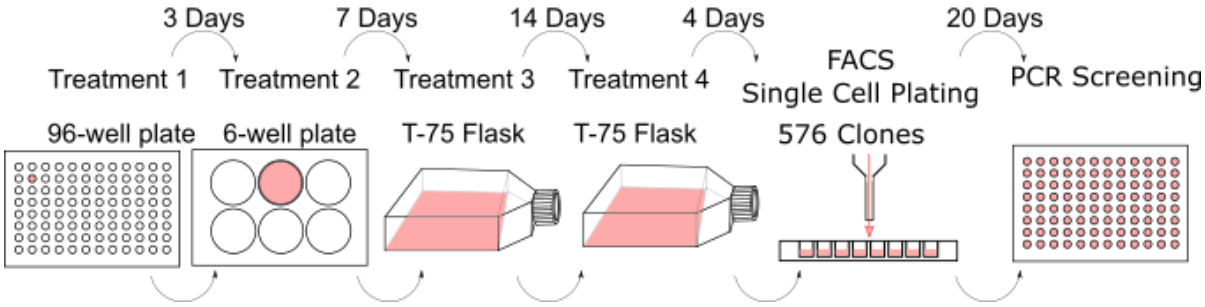

Figure 2.10: Timeline of RNP transfection and clonal selection in teloHAEC

Experimental overview of CRISPR-Cas9–mediated genomic deletion in HAECs. TeloHAECs were treated with RNP complex four times before single-cell sorting and isogenic clone selection. Cells were subjected to several treatments to increase likelihood of bi-allelic deletions, which would result in the complete removal of the enhancer. FACS was used to plate single cells to increase chances of successful isogenic clonal generation. Clones were screened by PCR and gel electrophoresis.

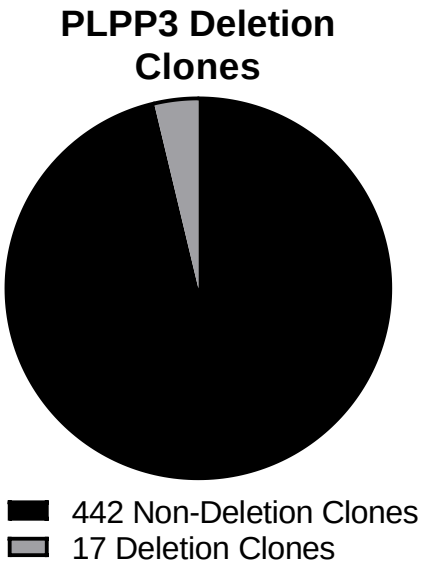

Figure 2.11: CRISPR Deletion efficiency

A total of 576 single cells were grown for clonal selection. 459 clones were viable after 20 days of incubation. 17 clones were detected to have either a mono or bi-allelic deletion by PCR screening, resulting in a 3.7% clonal deletion efficiency.

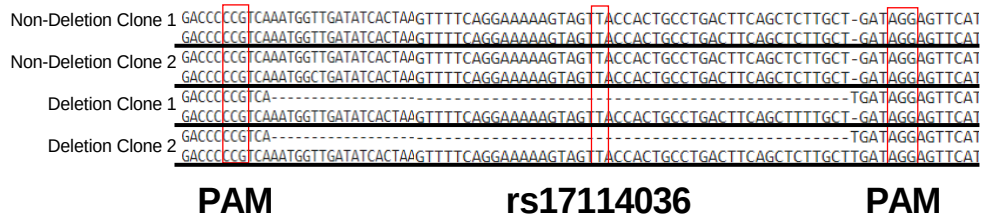

Figure 2.12: Confirmation of successful CRISPR/Cas9-based genomic deletion by Sanger sequencing

Bacterial CRISPR-associated protein-9 nuclease (CRISPR/Cas9) system results in successful genomic deletion at site of interest in teloHAEC. The PCR products were cloned (TA cloning) and sequenced to confirm deletion of the targeted region surrounding rs17114036. Protospacer adjacent motif (PAM) sequences, which are recognized and cleaved by Cas9, and rs17114036 are highlighted for reference.

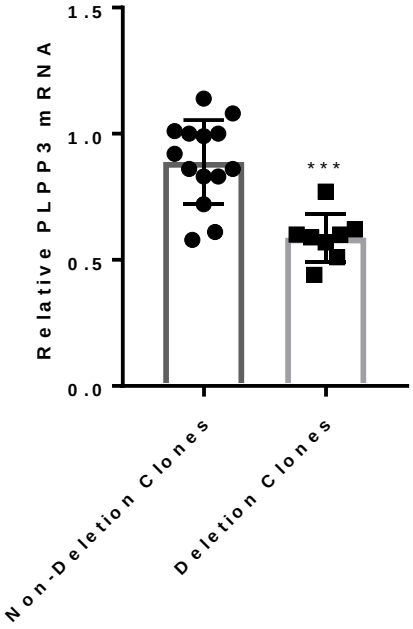

Figure 2.13: Down-regulation of PLPP3 transcript in deletion clones

Unstimulated CRISPR clones were grown under standard conditions. Total RNA was collected for quantification analysis. PLPP3 mRNA was measured using qPCR absolute quantification with a standard curve. One dot represents one clone. Data represent mean ± SEM. ***P < 0.0005 as determined by Student's t test.

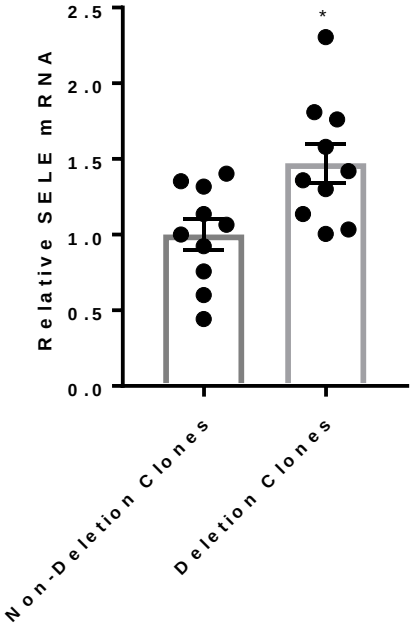

Figure 2.14: SELE mRNA increased in deletion clones

Deletion clones were treated with 10 μM of LPA for three hours. After 3 hours the total RNA was collected for quantitative analysis. SELE, a known downstream gene activated by LPA signaling was quantified using absolute quantification with a standard curve. One dot represents one clone. Data represent mean ± SEM. *P < 0.05 as determined by Student's t test

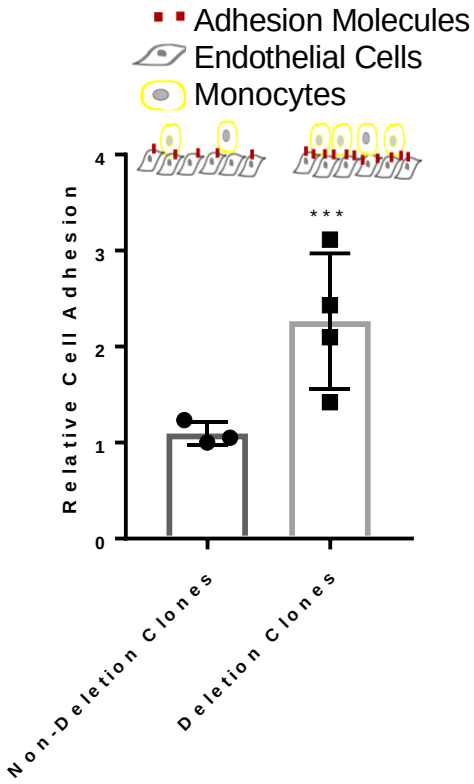

Figure 2.15: Increased monocyte adhesion in deletion clones as proxy of increased endothelial inflammation

Clones harboring deletion at enhancer were co-cultured with labeled THP-1 monocyte cells. After extended washing the adherent monocytes were detected using a plate reader. Increased signal detects greater cell adhesion. Experiment was performed in biological triplicate. One dot represents one clone. Data represent mean ± SEM. ***P < 0.0005 as determined by Student's t test.

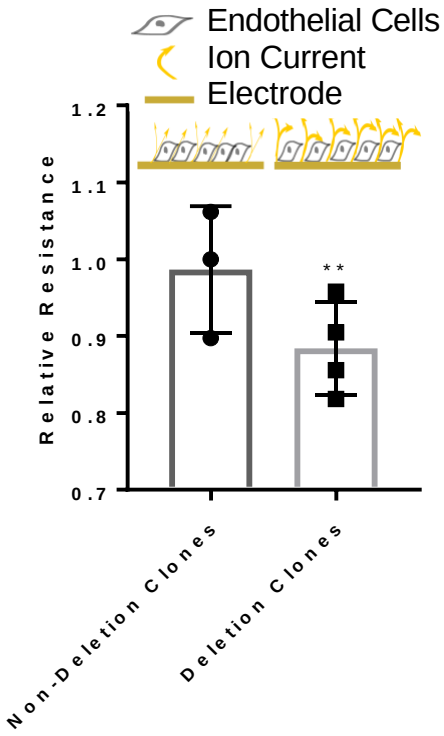

Figure 2.16: Increased monolayer permeability in deletion clones indicates increased endothelial dysfunction

Control and deletion clones were plated on an electrical array to measure resistance over time. The average resistance was taken for each clone over three hours. Experiment was performed in biological triplicate. One dot represents one clone. Data represent mean ± SEM. **P < 0.005 as determined by Student's t test.

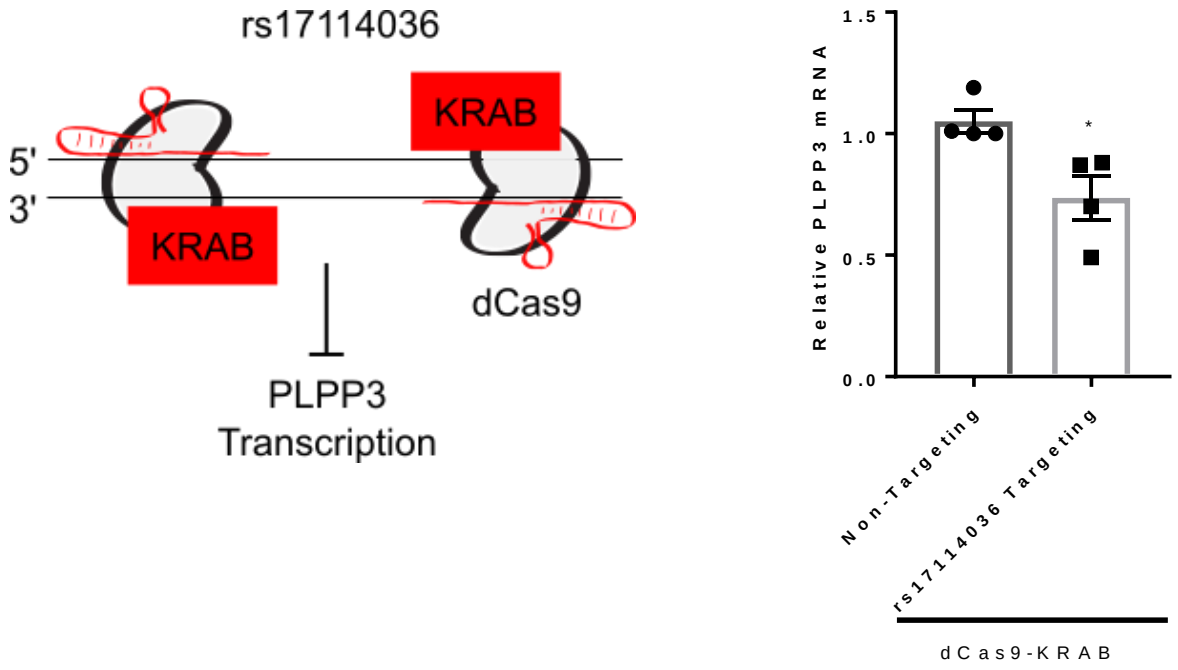

Figure 2.17: CRISPR Interference at rs17114036 sufficient to down-regulate PLPP3 mRNA expression

CRISPRi uses a fusion protein of catalytically dead Cas9 bound to KRAB repressor domain to hinder targeted to the genome by a guide RNA to gene transcription. Dead Cas9/KRAB transcripts and sgRNAs were transfected into HAEC and incubated for 8 hours prior to cell lysis, sample collection, and analysis via qPCR. One dot represents one biological sample. Data represent mean ± SEM. *P < 0.05 as determined by Student's t test.

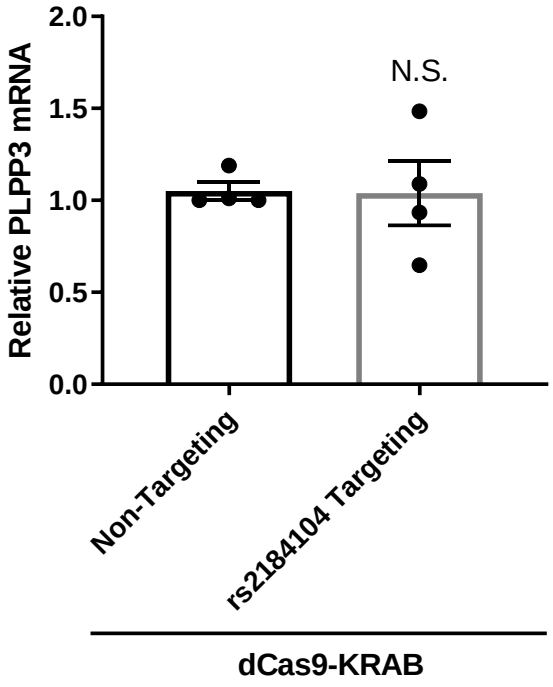

Figure 2.18: CRISPR interference on rs2184104 has no significant effect on PLPP3 mRNA expression

CRISPR interference (CRISPRi) using a sgRNA to guide a catalytically dead Cas9 (dCas9) fused to KRAB transcriptional repressor was targeted to the genomic sequence near rs2184104. Dead Cas9/KRAB transcripts and sgRNAs were transfected into HAEC and incubated for 8 hours prior to cell lysis, sample collection, and analysis via qPCR. One dot represents on biological sample. N.S. $p>0.05$ as determined by Student's t-test. Data represent mean ± SEM.

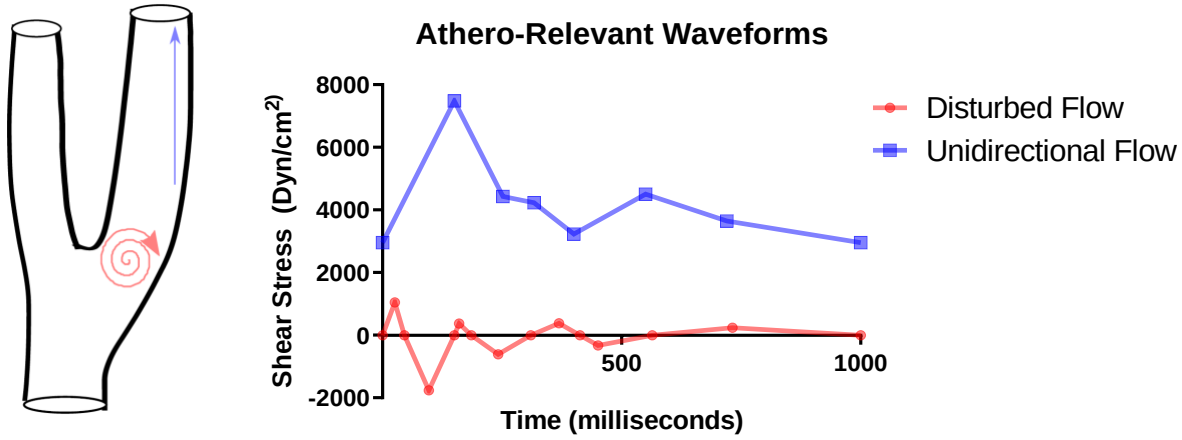

Figure 2.19: Shear stress programs used in experiments

Plot of time and shear stress. In the vasculature, endothelial cells are subjected to differing blood flow patterns, which affect endothelial health, function, morphology, and gene expression. The two lines correspond to the programs used to control the motors to recapitulate physiological hemodynamic flow. The blue line for unidirectional flow subjects the cells to high shear stress. The red line for disturbed flow subjects to low and oscillatory shear stress.

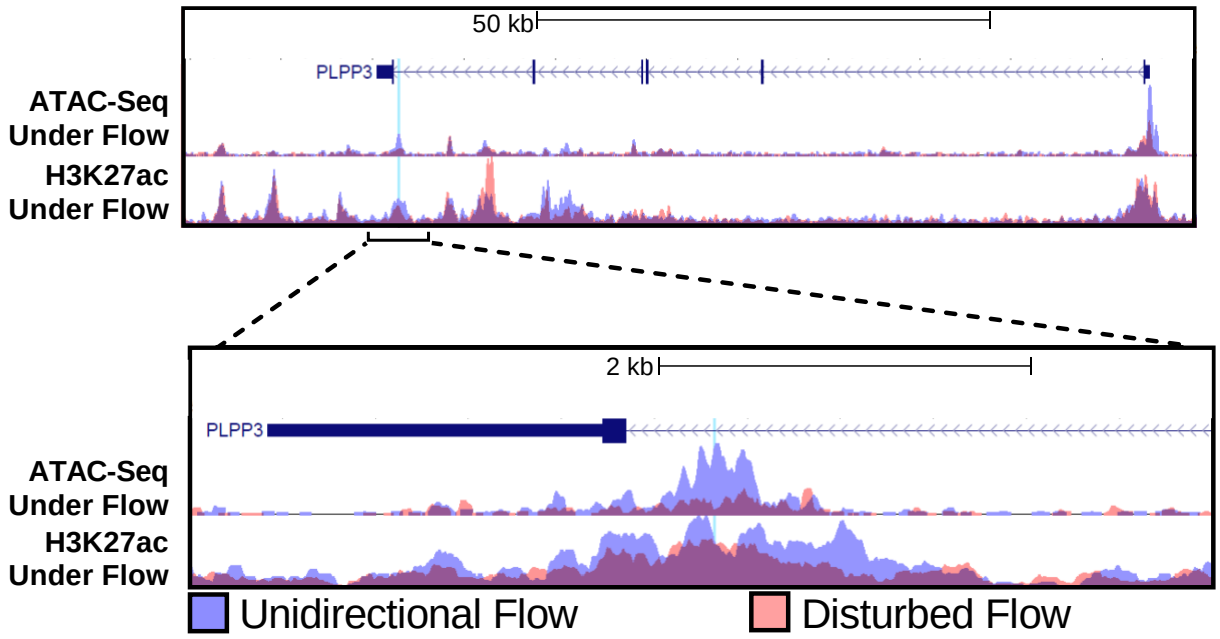

Figure 2.20: Unidirectional flow increases chromatin accessibility at enhancer

Increased chromatin accessibility and H3K27ac mark at chr1:56962213–56963412 in HAECs subjected to 24-h athero-protective unidirectional flow compared with cells under 24-h athero-susceptible disturbed flow. The PLPP3 locus is shown and zoomed in to demonstrate details around the enhancer region of interest, with rs17114036 highlighted with a vertical blue line.

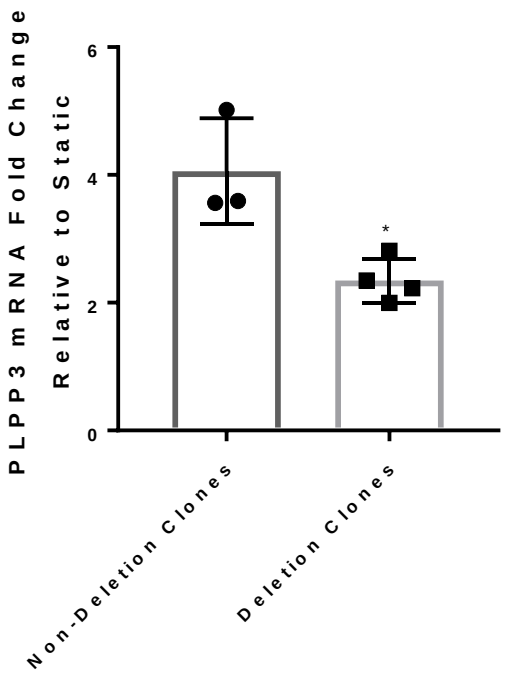

Figure 2.21: PLPP3 mRNA mechanosensitivity decreased under unidirectional flow in deletion clones

Reduced UF-induced PLPP3 expression in HAECs with genomic deletion at rs17114036-containing genomic locus. Control and genome-edited (~66-bp deletion) HAECs were subjected to 24-h UF. The y axis represents the fold change of PLPP3 mRNA quantities between the static conditions and UF for each individual clone. Nondeletion clones, n = 3; deletion clones, n = 4. Data represent mean ± SEM. *P < 0.05 as determined by Student's t test.

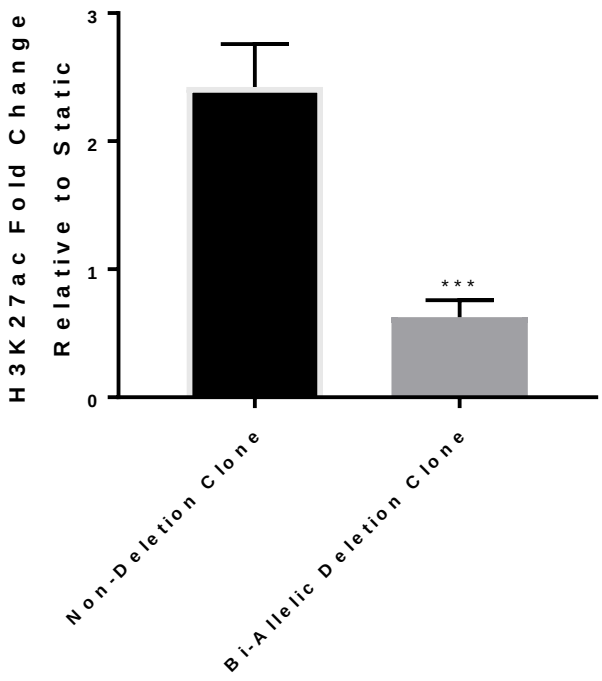

Figure 2.22: Deletion clones exhibit decreased histone acetylation under unidirectional flow at enhancer

H3K27ac ChIP-PCR performed in two CRISPR clones; one non-deletion clone and one bi-allelic deletion clone harboring the 66-bp deletion near rs17114036. The cells were subjected to static or UF conditions before cross-linking and ChIP. PCR primers were designed to amplify a region within the enhancer but not overlapping the deleted region. Data are shown as fold change of the UF-treated samples compared with static conditions. Data represent mean ± SEM. ***P < 0.0005 as determined by Student's t test.

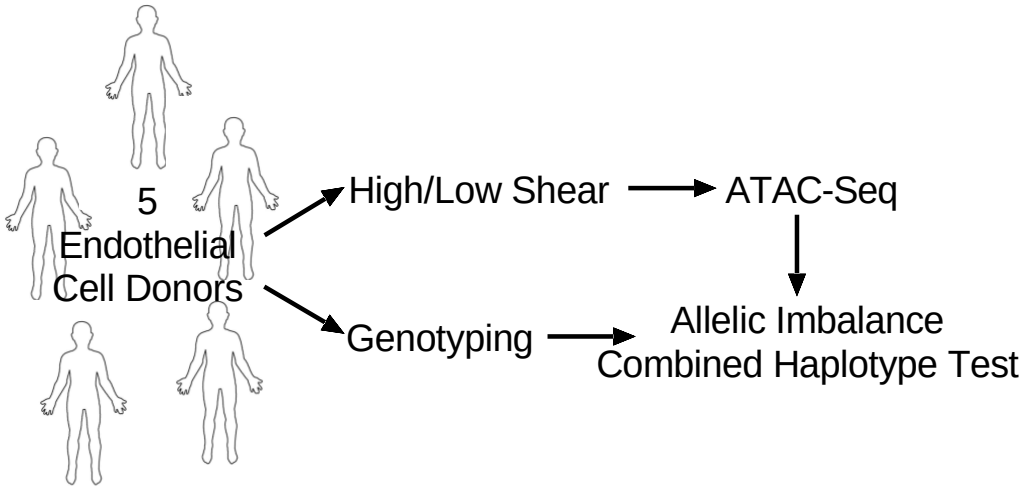

Figure 2.23: Overview of inter-individual chromatin accessibility response to flow

Endothelial cells from 5 unrelated, healthy donors were cultured and subjected to 24-hours of high or low shear stress. Genotyping data from these donors were acquired to allow for allelic imbalance and CHT to observe genetic response to flow treatment.

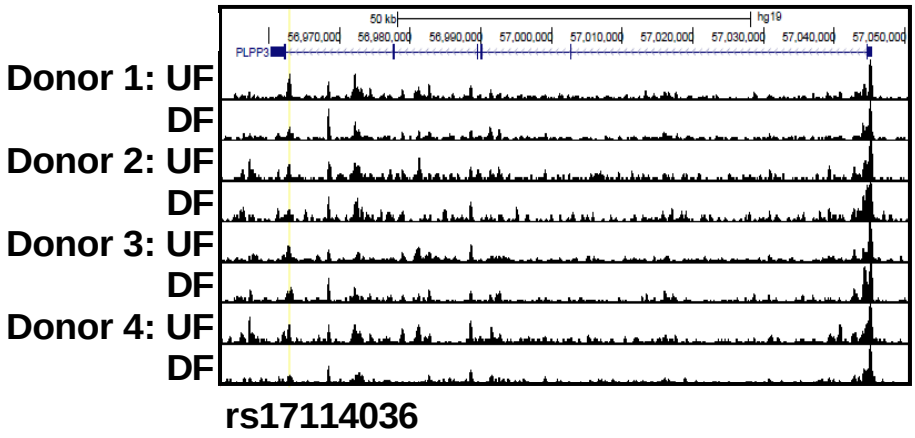

Figure 2.24. Genome browser view of ATAC-seq at PLPP3 locus from 5 donor cell lines subjected to unidirectional or disturbed flow

Unidirectional flow (UF) increases chromatin accessibility in PLPP3 intron 5 in HAECs when compared with cells under disturbed flow (DF). ATAC-seq was conducted in HAEC lines from 4 individual donors heterozygous (T/C) at rs17114036 under 24-hr athero-protective UF and athero-susceptible DF.

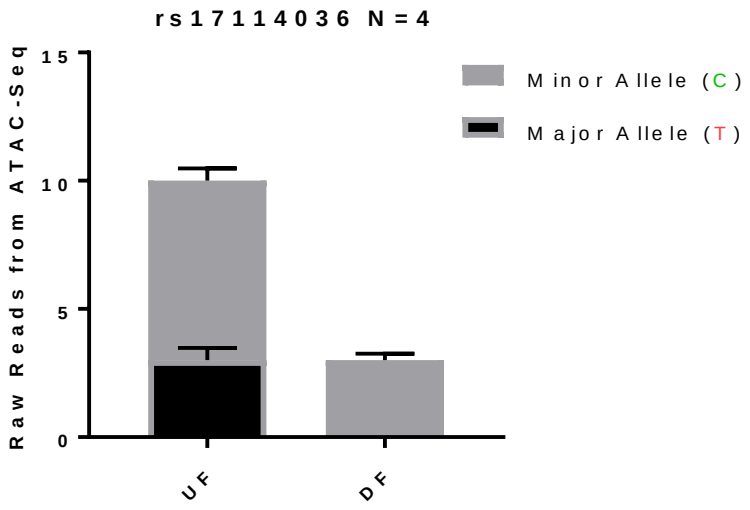

Figure 2.25: Unidirectional flow increases chromatin accessibility at rs17114036

Increased ATAC-seq reads in rs17114036-containing region in HAECs under UF compared with DF. ATAC-seq was conducted in HAEC lines heterozygous at rs17114036 under 24-h athero-relevant flows, detecting increased ATAC reads in cells under UF and higher reads from the C allele-containing chromosome compared with T allele.

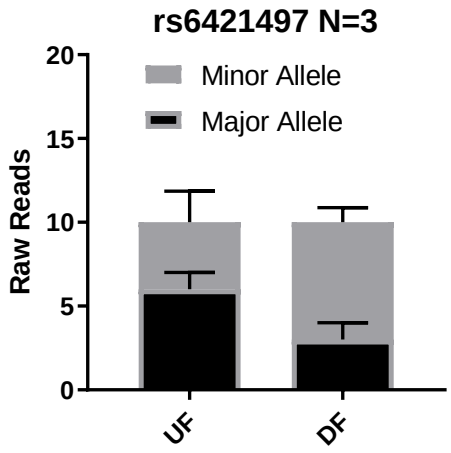

Figure 2.26: Unidirectional flow does not increase chromatin accessibility at rs6421497

Allelic imbalance analyses show unidirectional flow does not increase chromatin accessibility at locus related to the minor allele at rs6421497 in human aortic endothelium. SNP rs6421497 was not predicted to be causal but resides within a peak from ATAC-seq data set in at least 3 donors. The number of ATAC-seq reads detected under unidirectional and disturbed flow are reported and the proportion harboring either the major or minor allele.

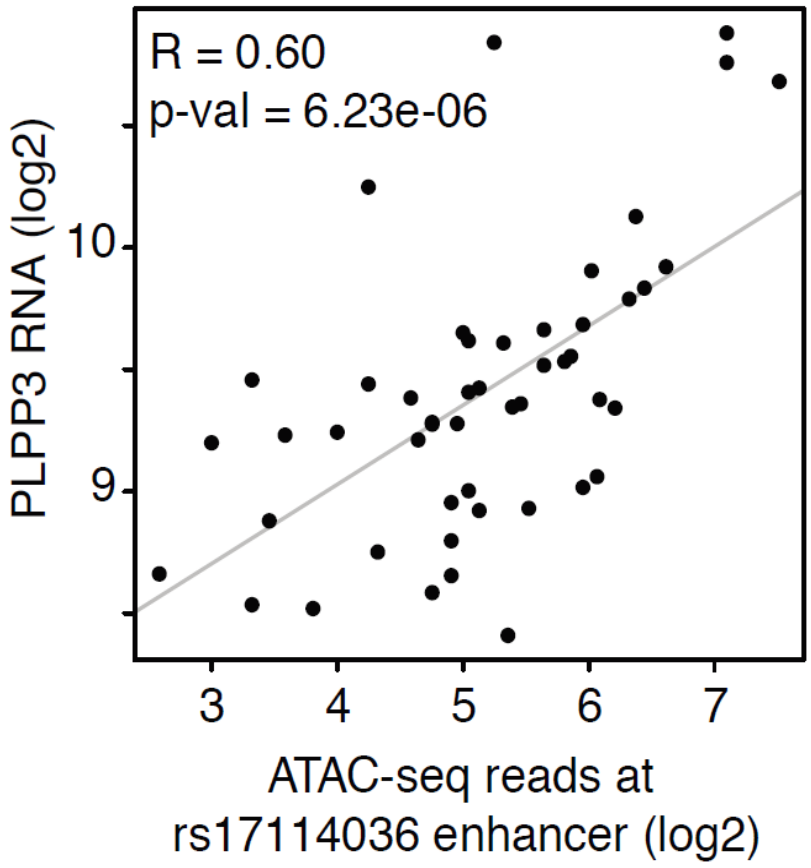

Figure 2.27: Chromatin accessibility at rs17114036 positively correlates with PLPP3 expression in human aortic cells

A positive correlation (R = 0.6, P value = 6.23e-06) between ATAC-seq reads at chr1:56962213–56963412 and PLPP3 mRNA detected by RNA-seq in 56 HAEC lines.

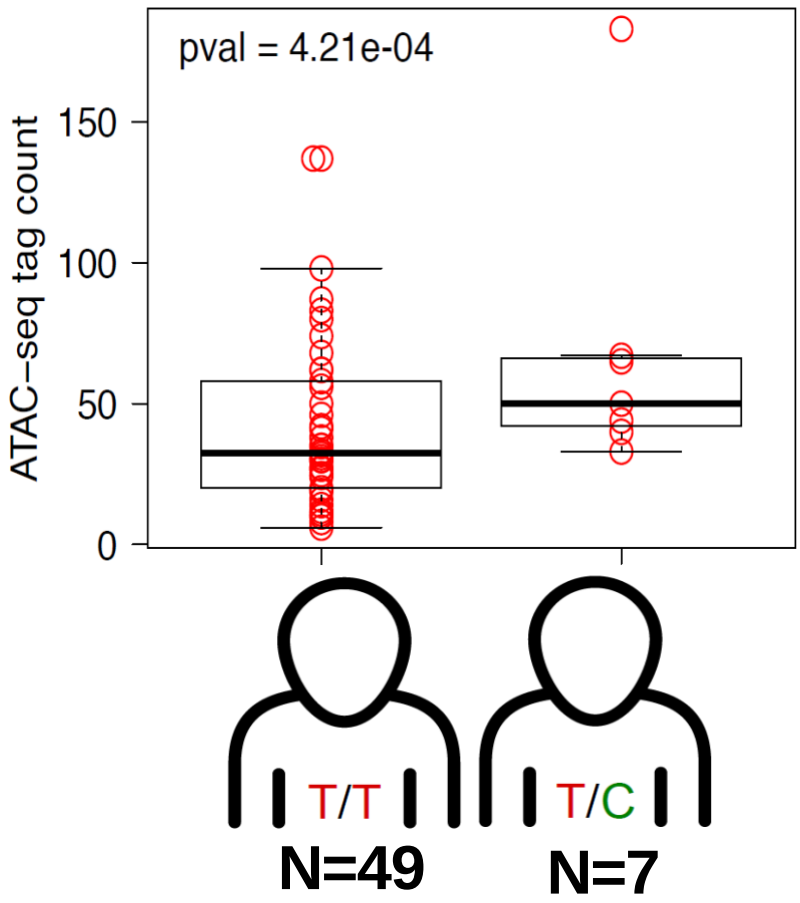

Figure 2.28: Genotype correlates with increased chromatin accessibility at rs17114036 in human aortic endothelial cells

Increased ATAC-seq reads in chr1:56962213–56963412 region from HAECs isolated from people heterogeneous (T/C) at rs17114036 compared with HAECs from people heterozygous (T/T) at rs17114036.

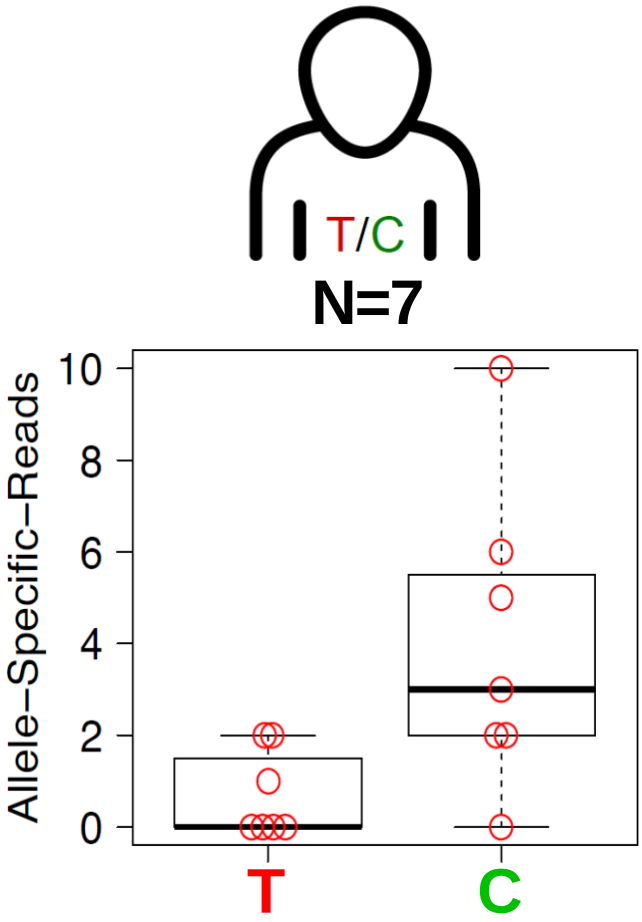

Figure 2.29: Heterozygous individuals harbor increased accessible chromatin signal at rs17114036 minor allele

Increased ATAC-seq reads at rs17114036-containing genomic locus from C (rs17114036)-containing chromosome compared with T-containing chromosome in HAEC lines heterozygous (T/C) at rs17114036.

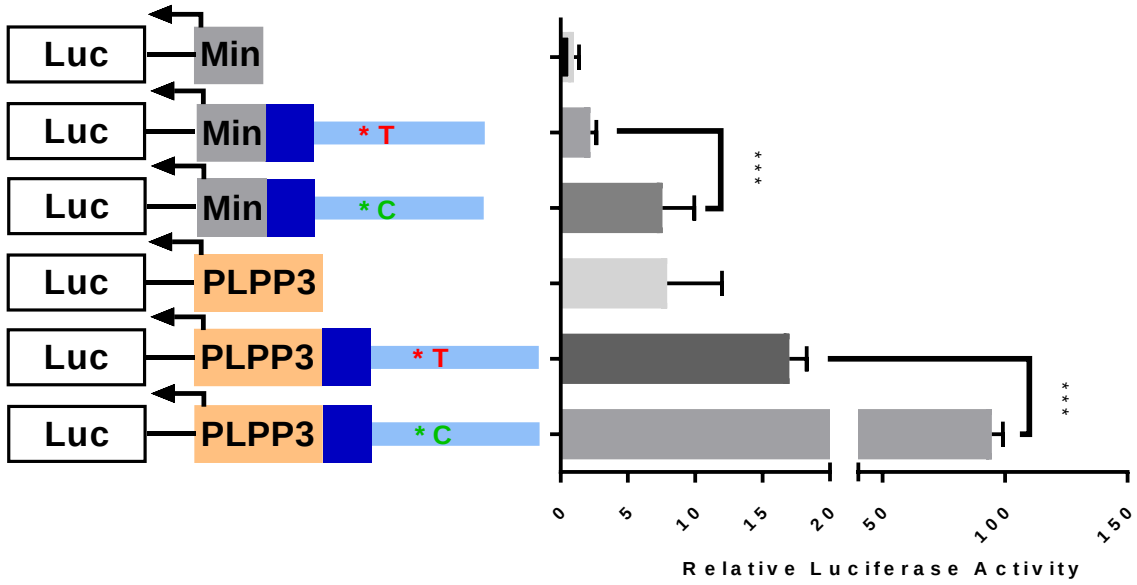

Figure 2.30: Single nucleotide substitution of C to T as rs17114036 increases gene reporter with and without PLPP3 promoter

Increased enhancer activity of chr1:56962213–56963412 with C allele at rs17114036 compared with T allele. Dual luciferase reporter assays were conducted in teloHAEC. The red and green asterisks denote the relative position of rs17114036 polymorphisms in the luciferase construct. n = 4–6. Data represent mean ± SEM. ***P < 0.0005 as determined by Student's t test.

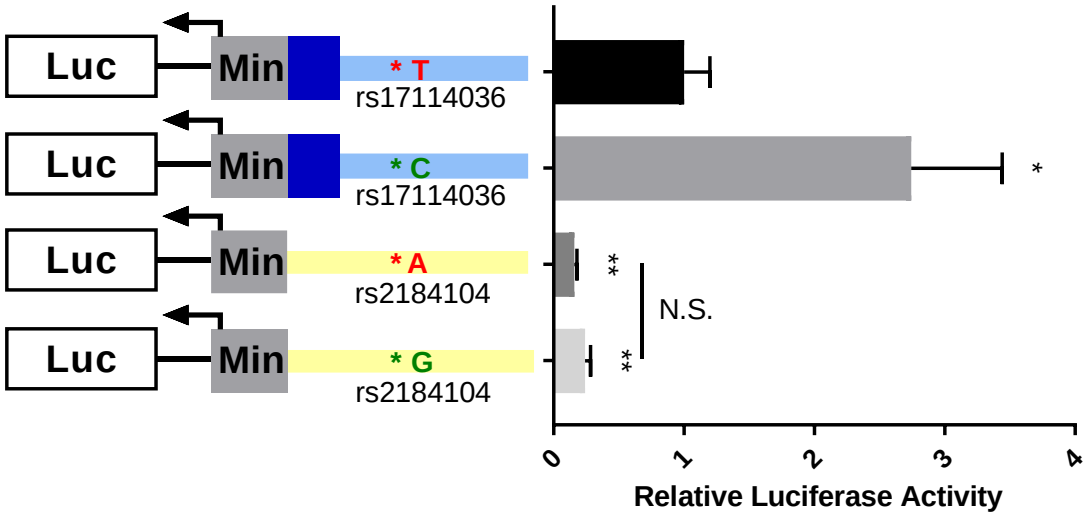

Figure 2.31: Allele-specific gene reporter activity observed between rs17114036 major and minor alleles, but not rs2184104

Comparison of luciferase reporter activity of predicted causative SNPs rs17114036 and rs2184104 reference and alternative alleles. Genomic regions around rs17114036 and rs2184104 were cloned and confirmed to harbor the reference alleles. Site-directed mutagenesis changed the reference to alternative alleles at these SNPs. Luciferase plasmids were electroporated into teloHAEC along with pRLTK internal control and incubated for 24 hours prior to lysate collection and luciferase activity measurement. *p<0.05, **p<0.005 as determined by Student's t-test. Data represent mean ± SEM.

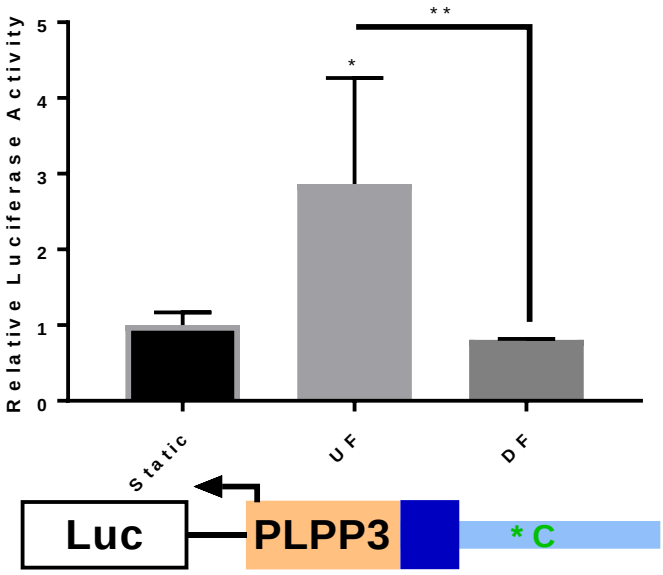

Figure 2.32: Unidirectional flow sufficient to increase activity of gene reporter containing PLPP3 promoter and enhancer with rs17114036 minor allele

Increased enhancer activity of chr1:56962213–56963412 (with C allele at rs17114036) under unidirectional flow (UF) but not disturbed flow (DF). Experiment was performed in biological triplicate and technical triplicate. Data represent mean ± SEM. *P < 0.05 as determined by two-way ANOVA; **P < 0.005 as determined by two-way ANOVA. The green asterisk denotes the relative position of rs17114036 in the luciferase construct.

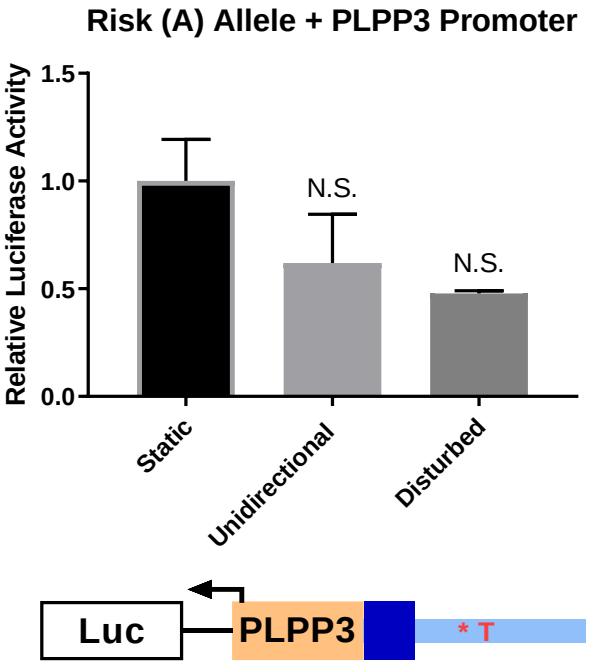

Figure 2.33: Unidirectional flow insufficient to increase activity of gene reporter containing PLPP3 promoter and enhancer with rs17114036 major allele

Cells electroporated with luciferase reporter construct harboring the PLPP3 promoter and the chr1:56962213-56963412 region (with T allele at rs17114036) under static, unidirectional flow, disturbed flow. Cells were electroporated and then left to settle for 6 hours prior to 18 hours of static, UF, or DF conditions. 24 hours post-transfection the samples were collected. Experiment was performed in biological triplicate and technical triplicate. N.S. $p>0.05$, as determined by two-way ANOVA. Data represent mean ± SEM.

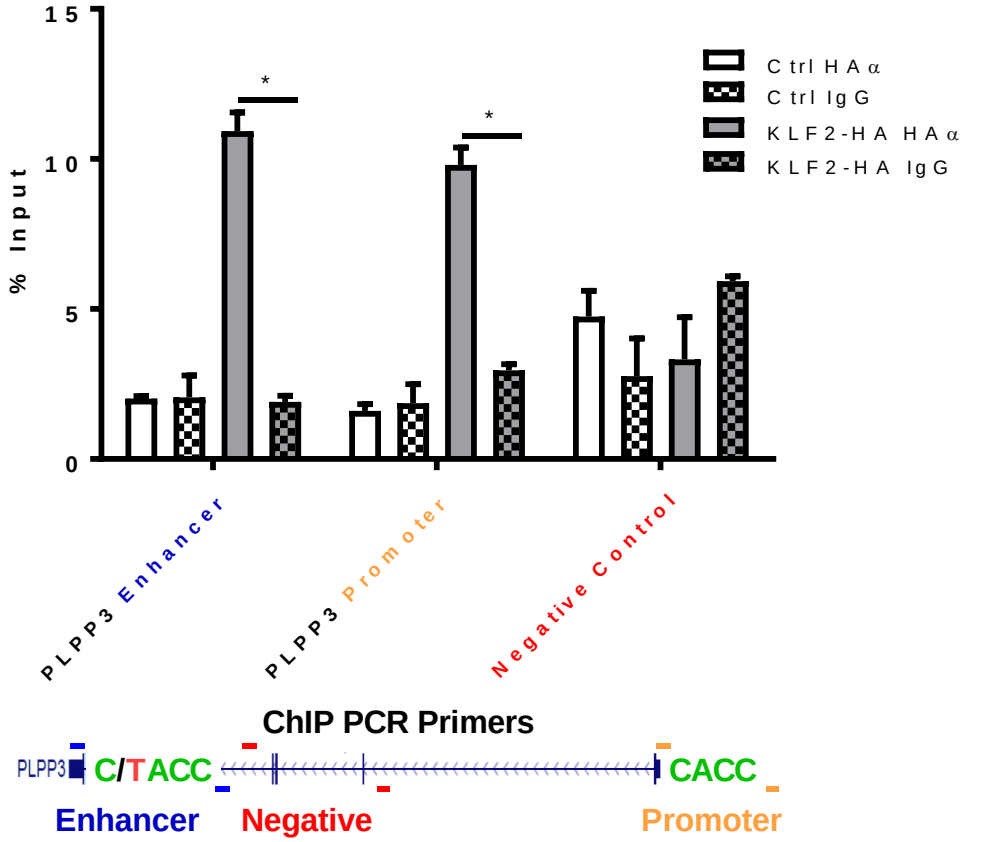

Figure 2.34: KLF2 binding at rs17114036 detected by ChIP using HA epitope tag

KLF2 affinity to CACC sites in human PLPP3 promoter and intron 5. ChIP-qPCR was performed with either a control IgG antibody or the antibody against HA followed by qPCR using primers detecting CACC sites in PLPP3 promoter or rs17114036-enclosing region from control HAECs (Ctrl) or HAECs transfected with KLF2 transcripts with HA tag. Primers that detect a site ~200 bp from the CACC at rs17114036 were used as a negative control. n = 4.

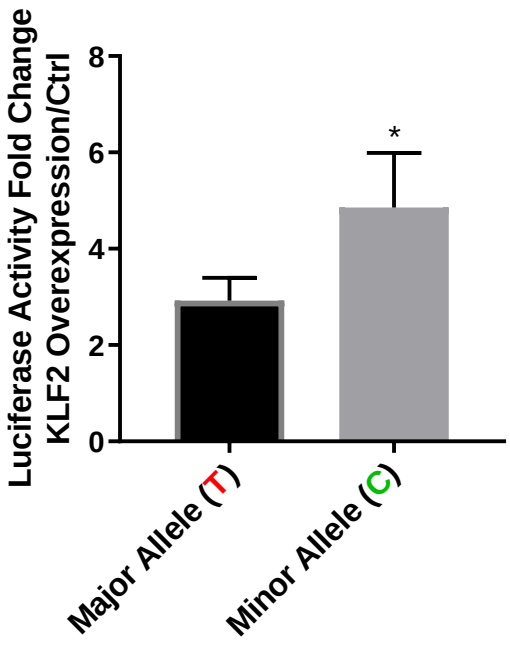

Figure 2.35: KLF2 over-expression significantly increases reporter activity

Increased enhancer activity of chr1:56962213–56963412 by KLF2 overexpression. Dual luciferase reporter assays were conducted in HAEC transfected with luciferase constructs containing the PLPP3 promoter and enhancer with either the major (T) or minor (C) allele at s17114036 along with KLF2-overexpressing or control plasmids. KLF2 overexpression resulted in a 2.9-fold increased luciferase activity in HAECs transfected with T allele-containing construct and a 4.7-fold increase in cells transfected with C allele-containing construct. n = 3. Data represent mean ± SEM. *P < 0.05 as determined by Student's t test.

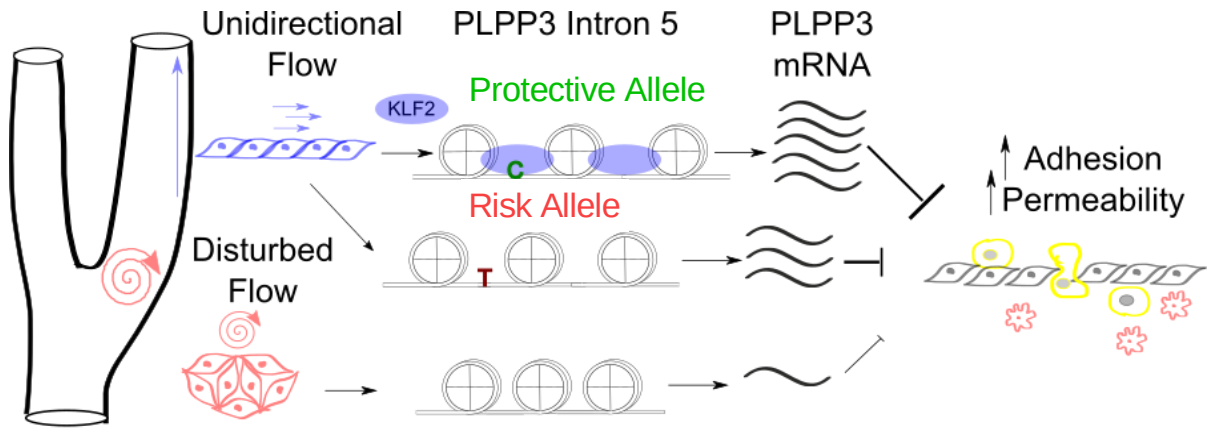

Figure 2.36: Model of a gene-by-environment interaction connecting hemodynamic flow, chromatin accessibility, PLPP3 expression, and endothelial function

The interplay between hemodynamic forces, chromatin landscapes at PLPP3 intron 5, and rs17114036 at the molecular level in regulating endothelial PLPP3 expression and vascular functions.

Table 2.1: Results from the bioinformatic screening for predicted, causal variants. The results of each test are either "pass" or "fail" highlighted in green or red respectively for each SNP.

SNP ID	Conditional and Joint Association	Bayesian Statistical Approach
rs4926709	Pass	Fail
rs778125	Pass	Fail
rs6666317	Fail	Fail
rs11206809	Pass	Fail
rs2184104	Pass	Pass
rs4912172	Fail	Fail
rs12122628	Fail	Fail
rs12137498	Fail	Fail
rs1889145	Fail	Pass
rs6691109	Fail	Fail
rs6698602	Fail	Pass
rs11206815	Fail	Pass
rs6421497	Fail	Pass
rs4912286	Fail	Pass
rs2780857	Fail	Fail
rs7368386	Fail	Pass
rs7365290	Fail	Pass
rs7527084	Fail	Pass
rs7513178	Fail	Pass
rs7535319	Fail	Pass
rs1261410	Fail	Pass
rs10888971	Fail	Fail
rs6588634	Fail	Fail
rs1815487	Fail	Fail
rs10789024	Fail	Fail
rs11811433	Pass	Fail
rs10888975	Fail	Fail
rs17114036	Pass	Pass
rs9970807	Fail	Pass
rs17114046	Fail	Pass
rs4912314	Fail	Fail
rs7528118	Fail	Fail
rs7515808	Fail	Fail
rs6688009	Fail	Fail
rs953857	Fail	Fail
rs953856	Fail	Fail
rs10888977	Fail	Fail
rs1930760	Fail	Fail
rs2404715	Fail	Fail
rs12134109	Fail	Fail
rs11206838	Fail	Fail
rs1041602	Fail	Fail
rs17416285	Pass	Fail
rs11579321	Fail	Fail
rs736109	Fail	Fail

CHAPTER 3: Genome-Wide Identification of Endothelial Mechanosensitive transcriptome and cis-Regulatory Elements

3.1 Abstract

Mechanotransduction is the biological process in which cells convert physical force to a biochemical response. Here, experiments were performed to identify whole-genome transcriptome (Wu *et al*., 2017) and cis-regulatory elements (Krause *et al*., 2018) in human endothelium subjected to arthero-prone disturbed or athero-protective unidirectional flows. I have identified 1,423 genes to be differentially regulated by these athero-relevant flow waveforms (Wu *et al*., 2017). To investigate flow-sensitive cis-regulatory elements, I performed bioinformatics analysis on ATAC-seq results to identify nucleosome-free clusters with high densities of open, possibly active chromatin. The vast majority, 86% of nucleosome-free clusters, lie within close proximity of a gene promoter. These flow-sensitive cis-regulatory elements are enriched for binding sites for known mechanosensitive and endothelial-enriched transcription factors such as ERG, JUN, KLF2, NFκB. In addition to the rs17114036 described in Chapter 2, more than 300 SNPs from the GWAS catalog are present in these mechano-sensitive cis-regulatory elements, suggesting possible contribution of genetic variants in regulating endothelial mechanotransduction via affecting functions of cis-regulatory elements. This study systematically identifies endothelial mechanosensitive genes, cis-regulatory elements related to transcription factor binding, and genetic variants implicated in complex human disease.

3.2 Introduction

Identification of causal SNPs and causal genes implicated by GWAS remains challenging (Smemo *et al*., 2014). The majority of GWAS SNPs are located in the noncoding genome (Maurano *et al*., 2012). Emerging evidence suggest that the majority of these loci affect the

expression of a gene through cis-regulatory elements in the genome that have one or more consequential targets (Wright *et al*., 2010; Cowper-Sallari *et al*., 2011; Smemo *et al*., 2014; Musunuru *et al*., 2010; Claussnitzer *et al*., 2015; Montefiori *et al*., 2018). Regulatory chromatin is impossible to identify through DNA sequence alone, which proves a challenge to identify its function and targets (Smemo *et al*., 2014).

Chromosome conformation techniques rely on proximity ligation to piece together interacting fragments of the genome. The first of these techniques, Chromosome Conformation Capture (3C), used crosslinked chromatin cleaved by restriction enzymes and then ligated together using ligase to piece together proximal sequences of DNA, which were then detected by PCR (Dekker *et al*. 2002). 3C assays have since evolved with the advent of whole genome, paired-end sequencing into Hi-C techniques that map genome-wide chromatin interactions (Liberman-Aiden *et al*., 2009). Hi-C has been performed in an array of human cell types from stem cells, immune cells, and primary cells (Dixon *et al*., 2012; Dixon *et al*., 2015; Rao *et al*., 2014; Jin *et al*., 2013; Schmitt *et al*., 2016). Contacts within the genome are generally consistent across cell types, with a few exceptions seen at lineage-specific genes (Schmitt *et al*., 2016). A limitation of these studies is that the resolution is limited by sequencing depth, many interactions are not between promoters and regulatory elements, and the resolution is > 40kb (Montefiori *et al*., 2018).

In order to directly detect more functional and consequential interactions with greater resolution, Hi-C was modified to incorporate labeled probes matching genomic promoters. This new technique, promoter-capture Hi-C, enriches for promoter-enhancer interactions at higher resolution (Javeirre *et al*., 2016). Promoter-capture datasets have been performed in 17 different blood cell types, demonstrating that promoter-enhancer interaction patterns are cell type specific.

This approach has also linked non-coding GWAS variants with putative target genes (Javierre *et al*., 2016). A similar study looked at autoimmune diseases in primary human cells and found the majority of contacts made between enhancers were with distal, rather than proximal promoters (Mauger *et al*., 2017).

More than 4000 GWAS have been published and 90,000 SNPs have been associated with one or more complex traits (NHGRI GWAS catalog, https://www.ebi.ac.uk/gwas/). The majority of these SNPs lie in non-coding regions. In this study I integrate ATAC-seq and RNA-seq from endothelial cells subjected to unidirectional, and disturbed flow, with over 100,000 promoter-capture contacts to discover possible causal flow-sensitive genes regulating a complex human disease, namely atherosclerosis.

3.3 Results

3.3.1 RNA-sequencing identifies 1,432 unique differentially expressed genes under flow

HAEC were subjected to 24 hr treatments of unidirectional or disturbed flow (Figure 3.1) followed by RNA-seq to identify whole-genome transcriptome changes. This approach identified 1,432 unique differentially expressed genes. Of these 1,432 genes 742 were up-regulated in unidirectional flow, while under disturbed flow 690 different genes were induced (Figure 3.2). Many of the differentially expressed genes are involved in stress response, hemostasis, and blood vessel development suggesting many endothelial-enriched genes are affected (Figure 3.3). Specifically, gene sets associated with cell cycle, hemostasis, and chromatin organization are enriched in mRNAs induced by unidirectional flow (Figure 3.4). Unidirectional flow-induced genes are also implicated in hematopoiesis and blood related pathways (Figure 3.4). The genes up-regulated by disturbed flow are implicated in translation, hypoxia, and glycolysis, which

supports previous work done in our lab that showed disturbed flow induced HIF-1α to increase glycolysis in endothelial cells (Figure 3.5) (Wu *et al*., 2017).

3.3.2 ATAC-sequencing identifies 2,473 unique differentially accessible chromatin regions regulated by athero-relevant flows

ATAC-seq was performed on HAEC subjected to 24 hr treatments of unidirectional or disturbed flow (Figure 3.1). Quantification and differential analysis by HOMER (Heinz *et al*., 2010) identified 2,473 chromatin accessible regions, of which 2,433 were increased under unidirectional flow. Near these 2,433 peaks are 1990 unique, proximal genes. These peaks are shown on a heatmap of normalized reads across the differential peaks where there is a greater quantity visible in the unidirectional flow treated samples (Figure 3.6). The data suggest no obvious difference in the quantification between static and disturbed (Figure 3.6). 90% of these peaks are located in an intergenic region more than 5 kb from a gene transcription start site (TSS) (Figure 3.7). Metascape analysis of the closest gene near one of 2,433 differential ATAC-seq peaks showed strong enrichment for genes involved in cytoskeleton and morphology (Figure 3.8). This is important as it supports previous data and observations that flow regulates endothelial morphology (Dai *et al*., 2004; Davies *et al*., 2013).

Integration of flow-sensitive mRNAs and gene promoters with close proximity to flow-sensitive cis-regulatory elements identified 198 genes that are regulated by flow putatively via mechanosensitive cis-regulatory elements (Figure 3.9). Metascape analysis of these 198 genes that are differentially expressed and lie near a differential ATAC-seq peak implicated that these genes are involved in key endothelial function such as blood vessel morphogenesis and actin filament-based process (Figure 3.10). Figure 3.11 illustrates the RNA-seq and ATAC-seq peaks

of three representative genes, KLF2, KLF4, and CD34, all of which have key regulatory roles in endothelial functions mediated by hemodynamics (Linnemann *et al*., 2011; Hogan *et al*., 2017; Sangwung *et al*., 2017). This is encouraging data that shows that this method is reliable for validating previous observations and will allow for unbiased discovery of novel targets and genes that are regulated by flow. Identification of these genes is important to better understanding flow-regulation of endothelial cell function.

3.3.3 Nucleosome-free clusters in endothelial cells

Flow-sensitive ATAC-seq peaks indicate global epigenome changes in ECs responding to flow. Yet these peaks are of varying sizes. Using HOMER to analyze the ATAC-seq results I identified 657 nucleosome-free clusters (Figure 3.12) (Heinz *et al*., 2010; Whyte *et al*., 2013). To visualize these nucleosome-free clusters I created a heatmap of the normalized reads in 10 kb windows, corresponding to the average nucleosome-free cluster sequence length, in 50 bp increments by each sample (Figure 3.13). Similar to the differential chromatin accessibility sites, 85% of these nucleosome-free clusters lie in intergenic regions more than 5 kb from the nearest TSS (Figure 3.14). Metascape analysis of the closest gene near one of 657 nucleosome-free cluster ATAC-seq peaks showed enrichment of nearby genes involved in cytoskeleton and cell adhesion, similar to what was seen in the differential ATAC peaks (Figure 3.15).

I then identified the genes that were both differentially expressed from RNA-seq by applications of flows and were near a nucleosome-free cluster region regulated by hemodynamics, resulting a list of 94 genes (Figure 3.16). Metascape analysis of these 94 showed enrichments of cell adhesion and morphology regulation. (Figure 3.17).

3.3.4 ERG, KLF, JUN, and NFKB transcription factors bind many flow-sensitive regions in endothelial genome

Following the identification of 2,433 differential chromatin accessibility sites and 657 nucleosome-free clusters from ATAC-seq, I wanted to unbiasedly find important transcription factors present at these sites. I performed motif analysis using HOMER, which detects over-represented transcription factor binding motifs. The most enriched binding sites at differential peaks consisted of ETS, AP-1, KLF, SOX, and GATA transcription factor families (Figure 3.18). This is largely in line with previous literature regarding the transcription factors present at endothelial-specific enhancers (Hogan *et al*., 2017). I overlapped available transcription factor binding ChIP-seq data for ERG, JUN, NFκB, NRF2, and IRF1 transcription factors that have been performed in HAEC (Hogan *et al*., 2017) The predicted proteins JUN and ERG have binding sites in many of the differential peaks, as well as NFκB (Figure 3.19).

Motif analysis was also conducted on the nucleosome-free cluster regions. The most significant transcription factor families represented were KLF and ETS (Figure 3.20). I overlapped available transcription factor binding ChIP-seq data with nucleosome-free clusters (Figure 3.21) (Hogan *et al*., 2017). This data validates the previous literature that have shown the roles that these factors play in endothelial cell specific gene expression and enhancer activity (Linnemann *et al*., 2011; Hogan *et al*., 2017; Kalna *et al*., 2019). Identification of these factors further highlights their role in endothelial gene regulation at the transcriptional level.

3.3.5 GWAS loci in flow-sensitive cis-regulatory elements

In order to link these regulatory elements with a human trait or disease, I looked at the intersection of all the SNPs identified from GWAS studies with my sets of differential peaks and

nucleosome-free clusters. Roughly 3.9% of differential peaks contain a human GWAS polymorphism (Figure 3.22 and Table 3.1). 33.8%, or about one third of nucleosome-free clusters overlap with a human GWAS SNP (Figure 3.22 and Table 3.2). This represents a nearly nine-fold enrichment of GWAS SNPs in nucleosome-free clusters compared to differential peaks. Among the loci identified from this analysis was rs17114036, which lies within a differentially accessible chromatin region, and was linked to PLPP3 expression (Krause *et al*., 2018). This method is then successful in identifying interesting SNPs that lie within flow-sensitive regions of the genome with relevance to human traits and diseases such as CAD. The diversity of other phenotypes from obesity, hematopoiesis, and even cancer suggests a role of endothelial cells and the vasculature in a myriad of complex human disease that merit further investigation.

3.3.6 Promoter capture Hi-C in endothelial cells

Many genomic annotation tools rely on sequence proximity to assign cis-regulatory elements such as enhancers to the nearest gene (Heinz *et al*., 2010). Knowing that the genome exists in 3D, we collaborated with the Nobrega and Di Rienzo laboratories to obtain promoter-capture Hi-C data from endothelial cells (Javierre *et al*., 2016; Montefiori *et al*., 2018) The results of this analysis are two sets of coordinates. The first correspond to the gene promoter from the probe library, and the other set are for the detected interacting fragment (Figure 3.23). To simplify analysis, I turned the coordinates for the promoter into the name of the gene with which it is associated. I then took the intersection of differential chromatin regions and nucleosome-free clusters with these capture sites to find how many peaks and cis-regulatory elements made physical contact with at least 1 gene promoter. Of the 2,433 differential peaks

only 14.8% of peaks contact a promoter (Figure 3.24). In the nucleosome-free clusters, 86.2%, or nearly every nucleosome-free cluster lies in physical proximity with a gene (Figure 3.24). As previously shown, these stretches are more enriched than other sites in the genome to regulate neighboring or distal genes (Whyte *et al*., 2013). Physical proximity to a gene TSS is evidence of direct regulation between an enhancer and a target gene (Heinz *et al*., 2015). Investigation of prioritized flow-sensitive nucleosome-free clusters and the genes with which they contact would provide direct, transcriptional regulation and lead to better understanding of how flow regulates and activates gene enhancers.

Two interesting loci were detected; vascular endothelial growth factors receptor 1 (FLT1), which is an important growth factor receptor in endothelial cells (Figure 3.25), and Golgi SNP receptor complex member 2 (GOSR2) (Figure 3.26). These genes stand out as they contain flow-sensitive enhancers in direct contact with the gene promoters, and harbor GWAS SNPs for human cardiovascular traits.

3.4 Discussion

3.4.1 2% of chromatin accessible regions are differentially induced by flow

My results demonstrated that a small fraction of the genome that is nucleosome-free in endothelial cells is regulated by athero-relevant hemodynamics. While I observed differentially expressed genes in both unidirectional and disturbed flow, nearly all of the nucleosome-free clusters are changed under unidirectional flow and not disturbed flow. It would seem likely that the cell's epigenetic state is very similar between static and disturbed flow, which was proposed by whole-genome transcriptome analyses (Qiao *et al*., 2016). There are also more than 2,400 differentially accessible regions and only 1,690 differentially expressed genes. Therefore, it is important for future studies to further investigate the interplay between cis-regulatory elements

and gene expression by integrating the RNA-seq and ATAC-seq results with GWAS SNPs, eQTLs, and promoter-capture sites in order to identify functional regulatory chromatin.

3.4.2 Majority of flow-sensitive genes are indirect transcriptional targets

Differential RNA-seq identified 1,432 genes with statistically significant change in response to flow waveforms in vitro. Of these 1,432 genes, only 198 had a change in chromatin accessibility near the locus, and 94 were near a flow-sensitive nucleosome-free clusters. While transcriptional activation does not require a change in chromatin accessibility, it is nonetheless evidence of de novo enhancer activation (Heinz *et al*., 2015; Hogan *et al*., 2017). It seems likely that many of the direct targets, such as transcription factors KLF2 and KLF4 may in turn activate genes further downstream, independent of enhancer activation and drive transcription through a promoter or other already active enhancers. KLF2 and KLF4 have many redundant roles and targets and thus it is reasonable to speculate they regulate many of the same flow-sensitive genes (Sangwung *et al*., 2017). Other factors such as ERG and AP-1 may regulate other, more distinct sets of mechanosensitive genes from KLF target genes (Hogan *et al*., 2017; Kalna *et al*., 2019).

Future experiments such as ChIP-seq and CRISPR/Cas9-based genome editing will be informative to investigate the subsets of mechanosensitive genes regulated by specific transcription factors. ChIP-seq using an antibody against the factor of interest would identify the binding sites of the factor of interest but this information would lack important data on the corresponding change in gene expression or enhancer activity. Knockout or mutation of transcription factors followed by differential ATAC and RNA-seq compared to a non-deletion control would give sets of differential genes and accessible chromatin sites. This would provide consequential evidence of the genes and enhancers regulated by the factor of interest, and when coupled with ChIP-seq data would give a genome-wide view of mechanosensitive genes and the

factors that regulate them. In particular, genomic deletion or mutation of the factor of interest would circumvent the limitation of ChIP-seq requiring an adequate antibody, which currently precludes many experiments. Epitope tagging of the factor of interest allows for ChIP-seq experiments to be performed. However, this approach is still fraught with limitations such as that the function of the protein may be compromised by the tag, over-expression systems may not represent physiological conditions, and genetically encoded tags may not be feasible in some experimental models such as primary HAEC. Differential ATAC and RNA-seq with mutant or knockout of the factors is much more feasible but assumes that mutation or deletion of the factor is not lethal to the cells. KLF2 is an important mechanosensitive transcription factor but is embryonically lethal in knockout mouse models (Lee *et al.*, 2006). RNAi methods knocking down KLF2 are possible, suggesting that endothelial cells in vitro can survive with 60% knockdown of KLF2. ChIP-seq and gene editing methods together would provide consequential evidence of the genes and enhancers regulated by a factor of interest and give a genome-wide view of mechanosensitive genes, enhancers, and the factors that regulate them.

3.4.3 Nearly all nucleosome-free clusters contacts at least one gene promoter

Differential chromatin accessible regions overlap with promoter capture sites only represent 14.8% of cases, while 86.2% of nucleosome-free clusters are in contact with at least one gene promoter. ATAC-seq is a powerful technique to detect regions of open chromatin and potential enhancers in the genome but is limited in drawing conclusions linking an enhancer with its target gene (Buenrostro *et al.*, 2013). Promoter-capture Hi-C is well-suited for finding sequences that are in direct contact with gene promoters, but the data can be difficult to interpret (Javierre *et al.*, 2016). Merging the data from ATAC and promoter-capture Hi-C yielded possible enhancers and the promoter with which it is in contact. These methods can be further supported

using fluorescent in situ hybridization (FISH) combined with super resolution microscopy (Fields *et al*., 2019). Such an experiment would provide visual evidence of the proximity between a gene promoter and an enhancer. Granted, nucleosome-free clusters are simply bigger than the differential peaks, which may partially account for the increased odds that a promoter is in contact. Additional statistical analysis can give greater confidence that nucleosome-free clusters are more likely to be functional by taking into account for example, the size of the open chromatin. Use of a random control set of peaks would serve as evidence that it is not simply by chance that these enhancers and peaks co-localize with gene promoters (Montefiori *et al*., 2018).

3.4.4 One in three nucleosome-free clusters co-localizes with GWAS SNPs

Using the entire GWAS catalog I found that 33.8% of nucleosome-free clusters harbor one or more SNPs associated with a human trait, including traits of cardiovascular disease. It is important to note that this strategy was successful in validating rs17114036 residing in a differential peak as reported in chapter 2 (Krause *et al*., 2018). This suggests that this systematic approach identified hundreds of loci worth investigating with important biological questions and mechanisms to explore further. GOSR2 and FLT1 are two such examples of genes that are likely regulated by flow-sensitive enhancers and have important consequences for endothelial cell function and human health.

PLPP3, GOSR2, and FLT1 are examples of genes associated with CAD and have evidence of being regulated by flow. PLPP3 has been shown to be a critical mediator of endothelial inflammation and morphology and associated with CAD (Wu *et al*., 2015; Krause *et al*., 2018). FLT1 is named for being an important receptor for endothelial growth factors and important for endothelial cell-cell interactions (Fong *et al*., 1995; Ferrara *et al*., 2003). It is then

no surprise to find a SNP at the FLT1 locus that associates with human coronary heart disease. Further investigation is needed to uncover the mechanism by which this variant affects FLT1 and resulting cellular phenotypes. GOSR2 has not been studied within the context of endothelial biology nor in mechanotransduction yet holds promise of new biology as the locus is strongly associated with CAD, blood, pressure, aortic structure and other cardiovascular-related phenotypes.

3.5 Materials and Methods

3.5.1 Tissue culture of human aortic endothelial cells

Human aortic endothelial cells (HAECs) were procured from Lonza (Allendale, NJ) (CC-2535). Other HAEC lines were from aortic explants of heart transplant donors in the UCLA transplant program. Cells were grown in EGM-2 medium supplemented with SingleQuots from Lonza (CC-3156 & CC-4176) and Antibiotic-Antimycotic from Gibco (Grand Island, NY) (15240062) in a 37°C incubator with 5% CO_2. HAECs were used from passages 6-10.

3.5.2 RNA-seq for mechanosensitive genes

The experiment was performed in biological duplicate for cells treated under each flow condition. Total RNA was collected for cDNA library generation and subjected to whole genome sequencing for paired-end sequencing. The raw data was quality control checked to ensure high quality sequencing data. The sequencing reads were aligned to the hg19 reference transcriptome using HISAT2 (Pertea *et al*. 2016). The transcripts were assembled and quantified using Stringtie. The transcript and gene-level expression data were tested for differential significance using Ballgown. This pipeline identified 1,432 unique differentially expressed genes. To

visualize the data on the UCSC genome browser I used to HOMER to create a multiwig hub to plot the overlapped normalized data by treatment.

3.5.3 ATAC-seq for Mechanosensitive Chromatin Accessibility

The cell nuclei were collected and Tagmented using Tn5 transposase enzyme. The DNA libraries were submitted for whole genome sequencing. The raw data were aligned using Bowtie2 to the UCSC hg19 reference genome (Langmead *et al*., 2012). SAMtools suite was used to remove unmapped reads or reads that mapped to the mitochondria. Peak calling for enrichment of signal over background was performed using HOMER (Heinz *et al*., 2010). Differential chromatin accessibility sites between cells subjected to unidirectional or disturbed flow was performed using DESeq2. To visualize the data on the UCSC genome browser I used to HOMER to create a multiwig hub to plot the overlapped normalized data by treatment.

3.5.4 Quantification and heatmap visualization

To visualize the differential ATAC-seq peaks I created a heatmap of the normalized reads in 2 kb windows, corresponding to about the average peak size, in 50 bp increments by each sample (Figure 3.6).

3.5.5 Nucleosome-free cluster analysis

In order to identify a cohort of peaks I performed nucleosome-free cluster analysis using HOMER. This analysis uses normalized reads counts in large genomic regions. These regions are then ranked and plotted against normalized tag density. From this plot a cutoff is applied to call anything where the slope is greater than 1 is a nucleosome-free cluster.

3.5.6 Motif analysis

I performed motif analysis using HOMER, which detects over-represented transcription factor binding motifs in differential peaks and nucleosome-free clusters from ATAC-seq (Heinz *et al*., 2010). Analysis was performed using all human genomic promoter sequences as background.

3.5.7 Gene ontology analysis

Gene ontology analysis using Metascape was performed for general pathways that are over-represented in the lists of differentially expressed genes (Zhou *et al*., 2019). Gene ontology analysis using Metascape was performed for general pathways that are over-represented in the lists of differentially expressed genes. Metascape was accessed online (http://metascape.org/gp/index.html#/main/step1) and default parameters were used for analysis.

3.5.8 Promoter-capture Hi-C and enhancer-promoter interaction analysis

Promoter-capture Hi-C was a collaborative effort between the Fang, Nobrega, and Di Rienzo laboratories. Promoter-capture Hi-C was performed on cultured HAEC as previously described (Montefiori *et al*., 2018). The resulting data was provided as two sets of coordinates – the promoter region and the interacting fragment detected from paired-end sequencing. I used the refseq to assign the promoter probes to its corresponding gene. From this, I "named" every interaction with a gene name and the coordinates of the interacting fragment/enhancer and converted the data into a BED file.

I integrated this data with the results obtained from ATAC-seq. I found

interacting fragments that overlapped with a flow-sensitive chromatin region or a nucleosome-free cluster and the promoter of the gene. I used BEDtools to find overlaps between peaks from ATAC-seq and the fragments detected by promoter-capture Hi-C (Quinlan and Hall, 2010).

3.5.9 Genomic annotation of ATAC peaks

ATAC-seq peaks were categorized as being within a promoter, intragenic region, or intergenic region using HOMER (Heinz *et al*., 2010). Promoters were defined to be within 5 kb of a transcription start site. Intragenic peaks reside within a coding region. Intragenic peaks are outside of a coding region and 5 kb away from a transcription start site.

3.5.10 Transcription factor binding analysis

ChIP-seq for ERG, JUN, p65, NRF2, and IRF1 performed in HAEC was acquired from the Hogan *et al*. 2017 publication. The ChIP-seq peaks were arranged in a BED file format along with the ATAC-seq peaks. Overlaps between the datasets were determined using BEDtools (Quinlan and Hall, 2010).

3.5.11 GWAS analysis

The GWAS catalog was acquired from the UK biobank website (https://www.ebi.ac.uk/gwas/). The entire collection of GWAS data and phenotypes was included for an unbiased analysis. The coordinates of all GWAS SNPs were put into a BED file format. Overlaps between the datasets were determined using BEDtools for both differential chromatin regions and nucleosome-free clusters (Quinlan and Hall, 2010).

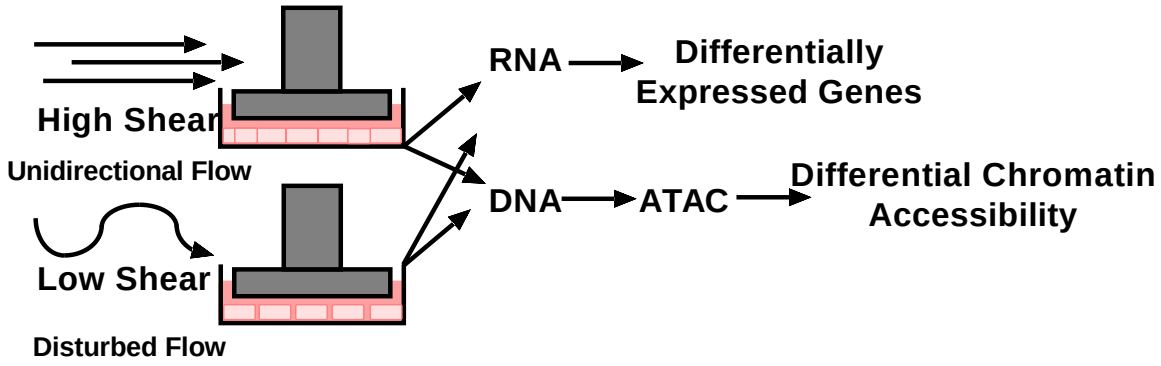

Figure 3.1: Experimental overview to investigate differential gene expression and chromatin accessibility under flow in human aortic endothelial cells

Cultured HAEC are grown in 6-well plates. A stainless-steel cone attached to an electric motor recapitulates physiological flow waveforms. The experiment consists of subjected HAEC to unidirectional or disturbed flow for 24 hours. After 24 hours, total RNA is collected and subjected to RNA-sequencing. Cell nuclei are collected and treated with Tn5 transpose for ATAC-seq.

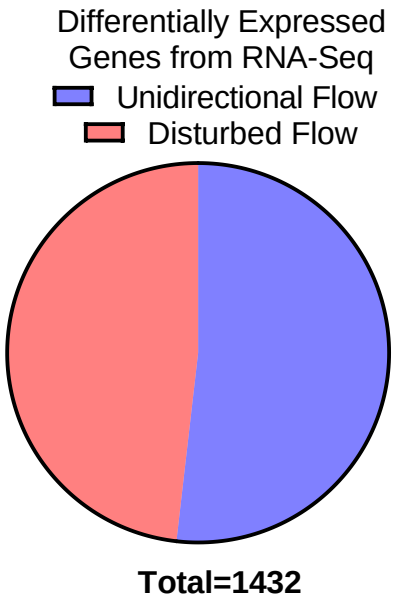

Figure 3.2: Total differentially expressed genes from RNA-seq

Cells were subjected to 24 hours of unidirectional or disturbed flow. Following RNA collection, library preparation, and sequencing the data were analyzed using the New Tuxedo Suite software pipeline. A total of 1432 genes were shown to be differentially expressed by FDR < 0.05. 742 unique genes were up-regulated and enriched under unidirectional flow. 690 unique genes were up-regulated and enriched under disturbed flow.

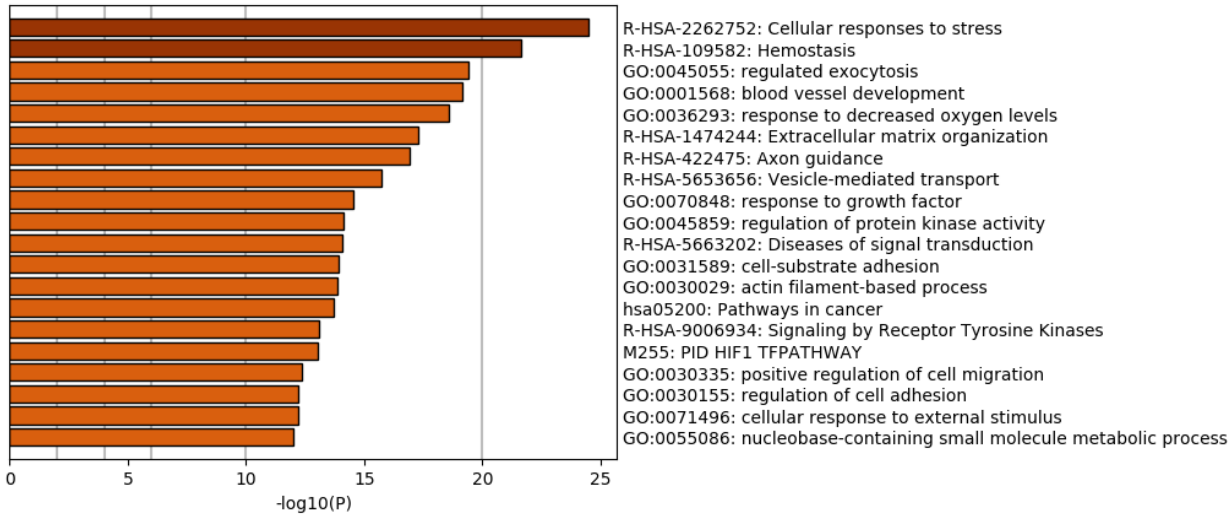

Figure 3.3: Enriched gene ontology terms for all differentially expressed genes by RNA-seq

All 1432 unique genes differentially expressed under both unidirectional and disturbed flow were entered into Metascape gene ontology analysis. The top gene sets are represented with the P-value plotted on the horizontal axis (-log10).

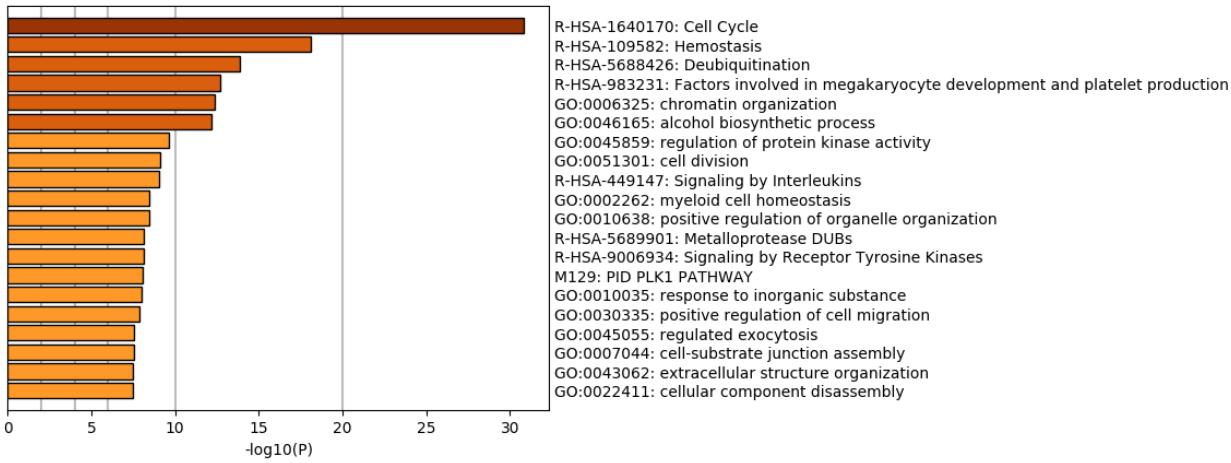

Figure 3.4: Enriched gene ontology terms for all differentially expressed genes by RNA-seq up-regulated by unidirectional flow

All 742 unique genes differentially expressed under unidirectional flow were entered into Metascape gene ontology analysis. The top gene sets are represented with the P-value plotted on the horizontal axis (-log10).

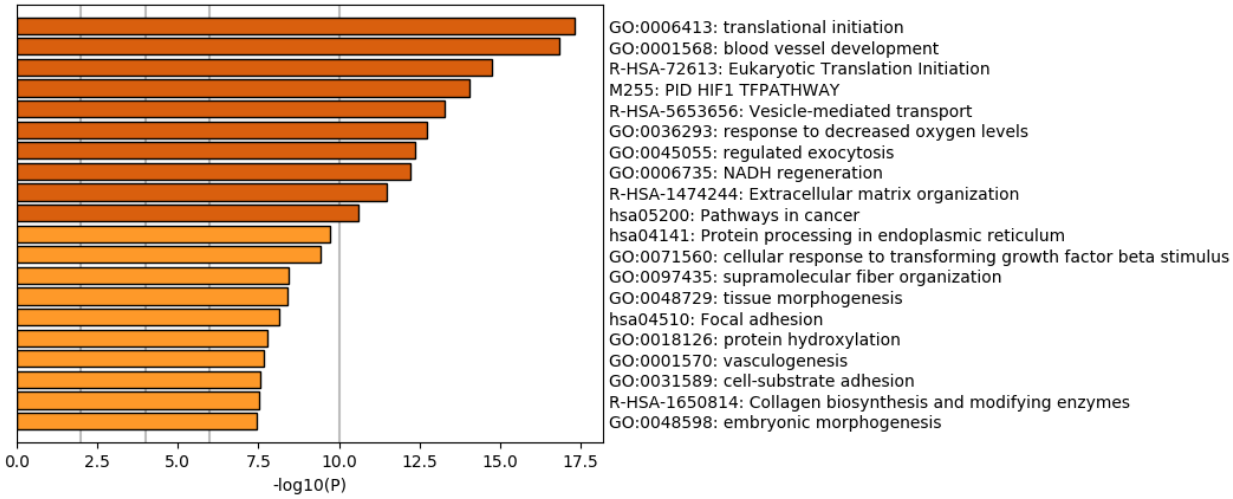

Figure 3.5: Enriched gene ontology terms for all differentially expressed genes by RNA-seq up-regulated by disturbed flow

All 690 unique genes differentially expressed under disturbed flow were entered into Metascape gene ontology analysis. The top gene sets are represented with the P-value plotted on the horizontal axis (-log10).

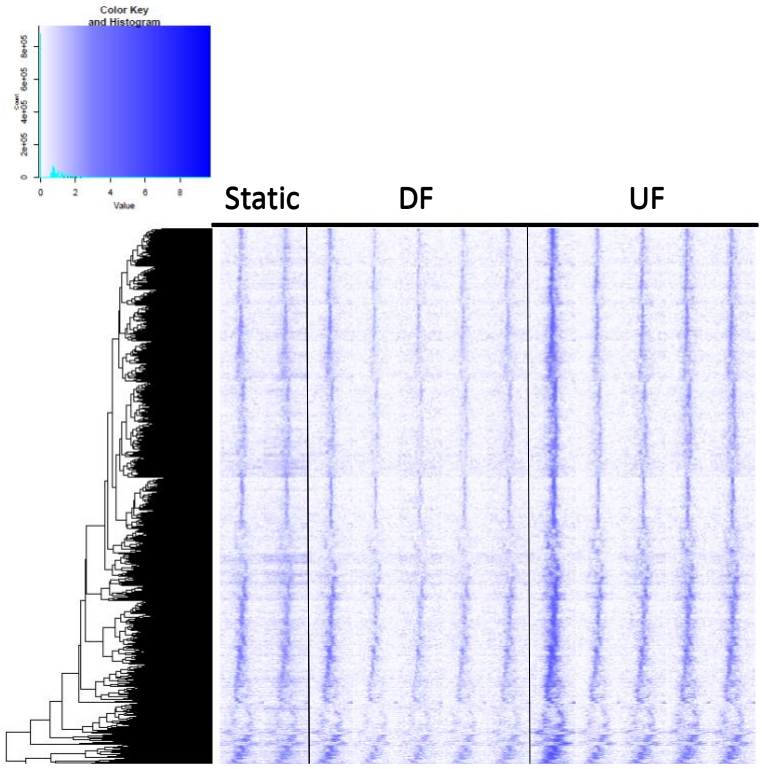

Figure 3.6: Heatmap of ATAC-seq peaks that were statistically significant

Heatmap was generated by quantifying normalized tag counts across a 2 kb region in 50 bp increments centered on 2,433 differential chromatin accessible regions induced by unidirectional flow. The UF samples show greater normalized quantification. A dendrogram groups more similar peaks together by hierarchical clustering.

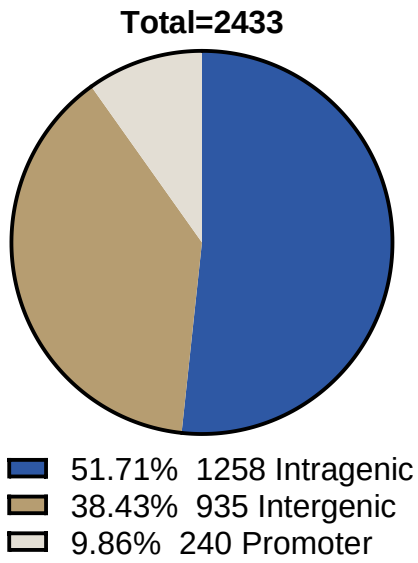

Figure 3.7: Annotation of differential ATAC-seq peaks by genomic location

Peaks from ATAC-seq analysis were assigned categorically to genomic features using refseq. Intragenic peaks were defined to lie within the coding region of any gene within the genome. Promoters were defined to be any peak within 5 kb of a transcription start site. Intergenic peaks were defined as any chromatin accessible region outside of a coding region and greater than 5 kb from a transcription start site.

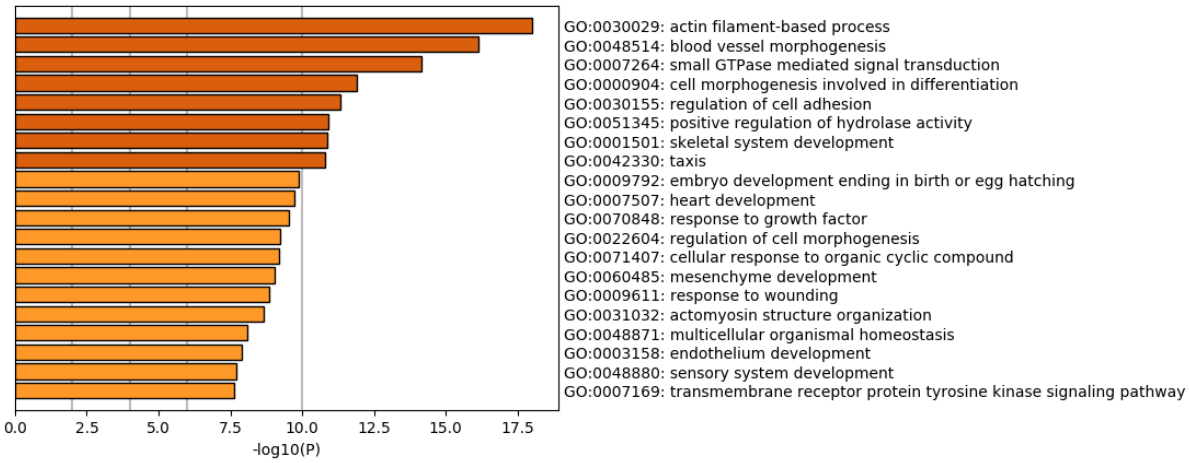

Figure 3.8: Enriched gene ontology terms for genes near differential chromatin accessible regions

Metascape analysis of the closest gene near one of 2,433 differential ATAC-seq peaks. Annotation was performed using HOMER to identify the closest gene to an accessible chromatin region induced by unidirectional flow. The gene ontology terms are listed with –log10 of the P-values for statistical significance.

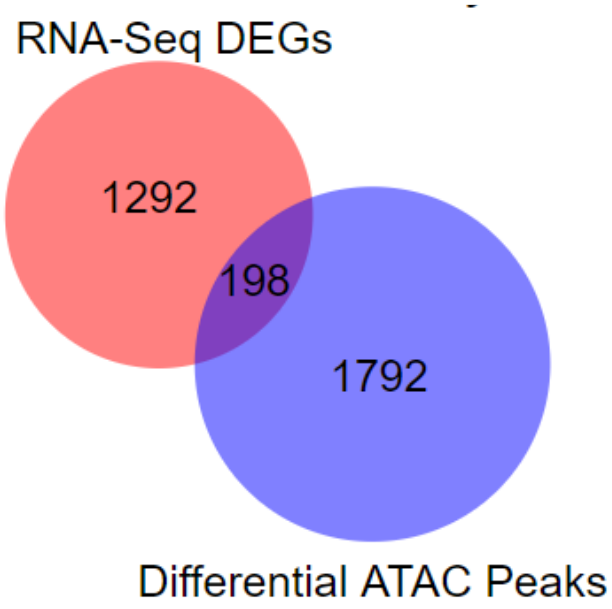

Figure 3.9: Intersection of genes with significant changes in gene expression and chromatin accessibility

Overlap of genes identified as differentially expressed from RNA-seq analysis, and annotated to be near a differential chromatin accessible site from ATAC-seq. The intersection of these two circles indicates a gene is both differentially expressed and is the nearest gene to at least one differential peak. These 198 genes are most likely to be directly, transcriptionally-regulated by shear stress.

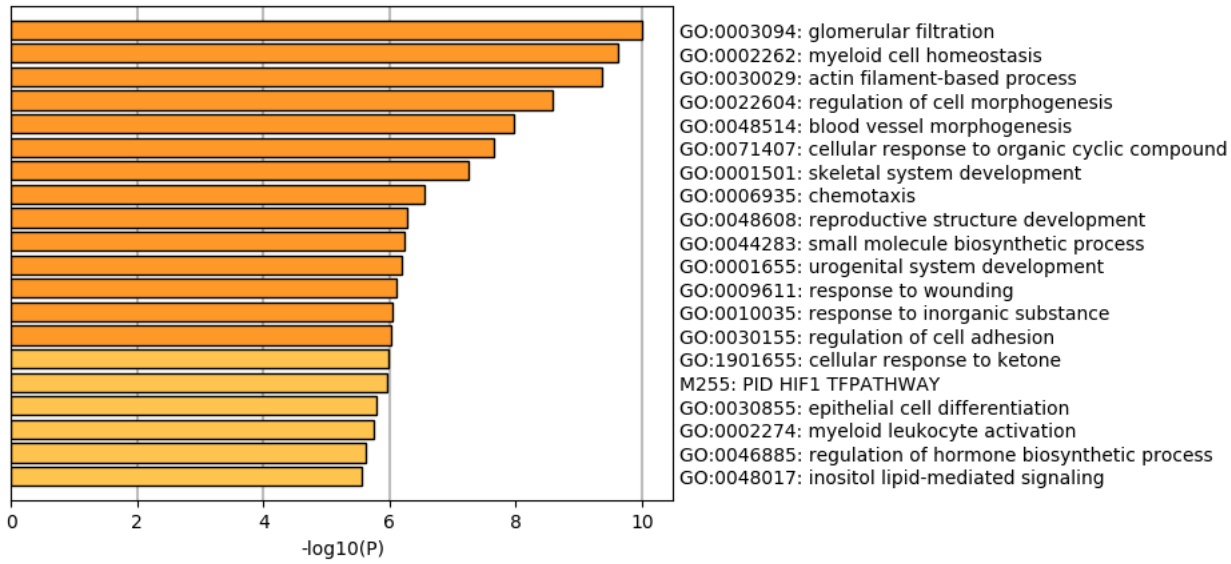

Figure 3.10: Gene ontology analysis of 198 genes that are differentially expressed with a nearby flow-sensitive nucleosome-free region

Metascape analysis of the intersection between 1,432 differentially expressed ones from RNA-seq and a gene near one of 2,433 differential ATAC-seq peaks. Intersection was performed by looking for genes appearing in both datasets. Annotation was performed using HOMER to identify the closest gene to an accessible chromatin region induced by unidirectional flow. The gene ontology terms are listed with –log10 of the P-values for statistical significance.

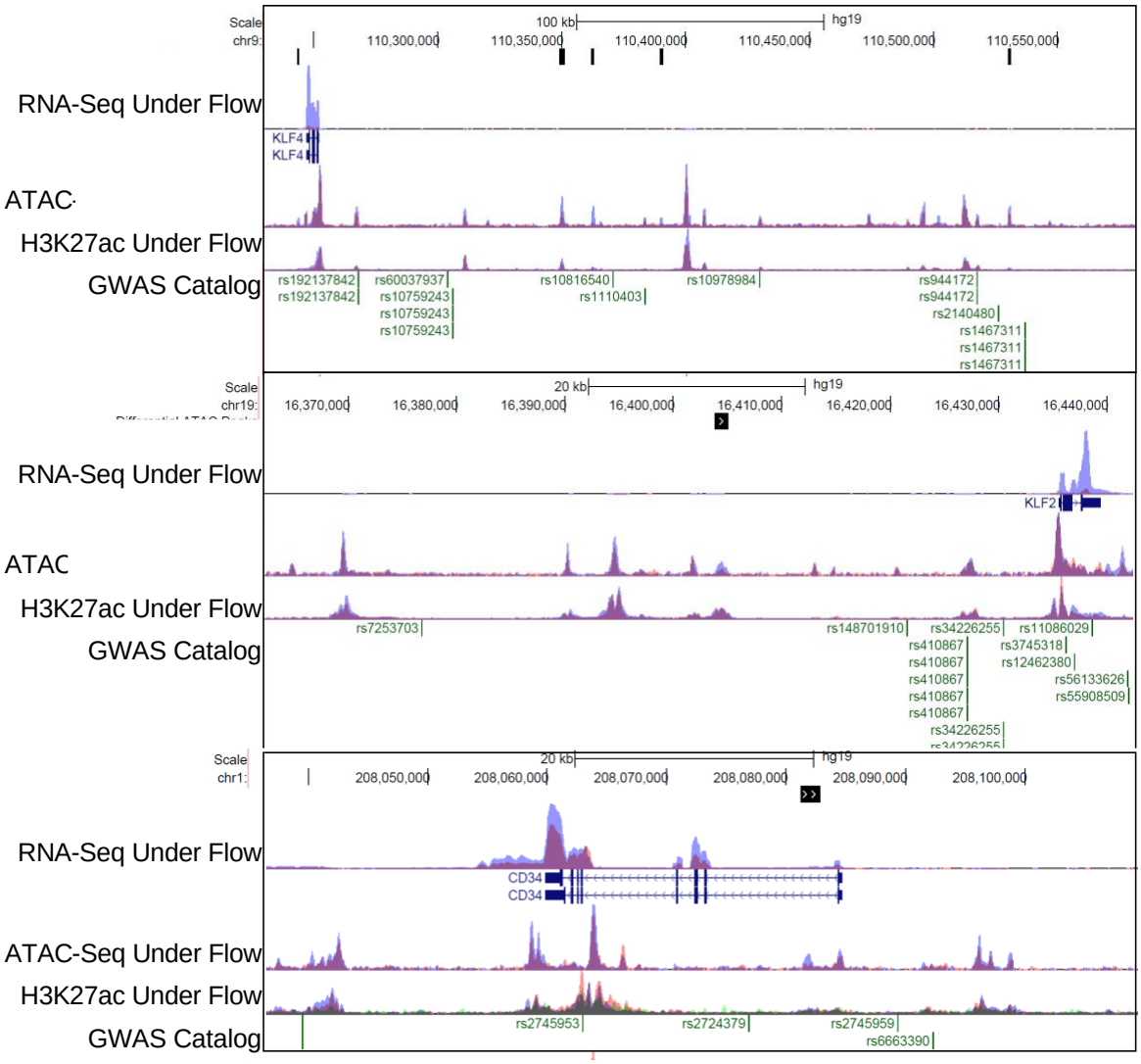

Figure 3.11: Representative genome browser images of genes that are significantly regulated by unidirectional flow with presence of increased chromatin accessibility

Representative genome browser images of possible, direct targets of shear stress. CD34, KLF2, and KLF4 are known to be critical endothelial genes. These genes are differentially expressed by RNA-seq and are proximal at least one differential peak from ATAC-seq. Also plotted are important SNPs identified from genome-wide association studies near these loci. Differential peaks are marked by black boxes.

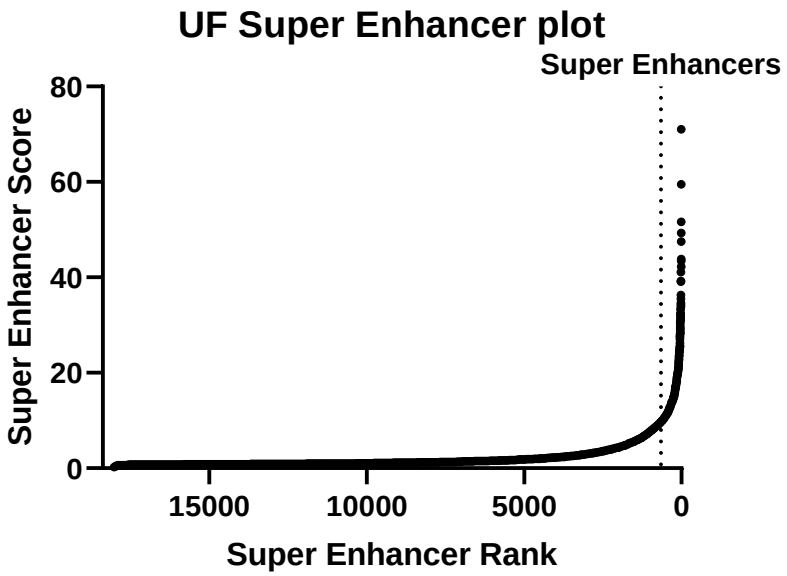

Figure 3.12: Nucleosome-Free Cluster plot

Nucleosome-free cluster analysis was performed using HOMER. Normalized tag counts were quantified and stitched together in 10 kb segments. The segments were than ranked from highest to lowest and normalized to the highest score. All points passed the point where the slope is greater than one was defined as a nucleosome-free cluster. This analysis identified 657 nucleosome-free clusters in endothelial cells under unidirectional flow

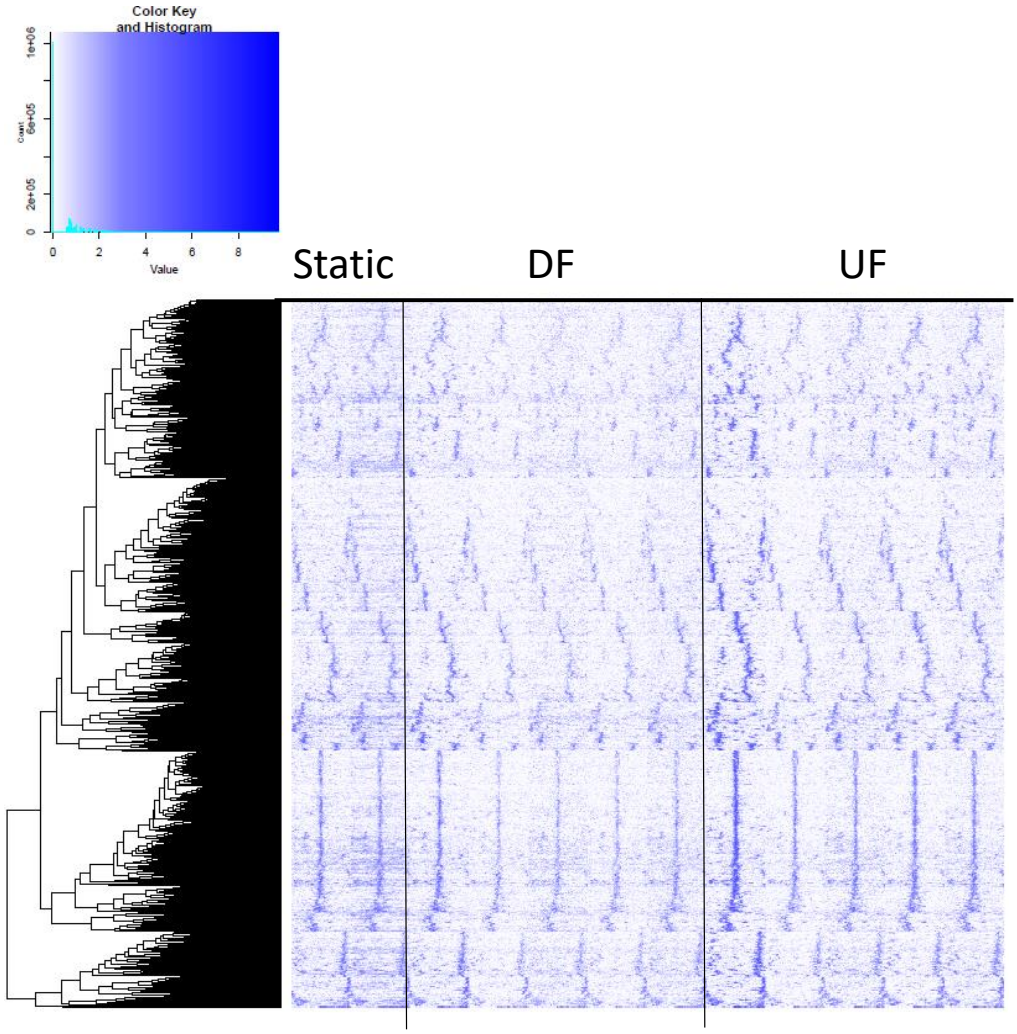

Figure 3.13: Heatmap of Nucleosome-free clusters identified from ATAC-seq in human aortic under unidirectional flow

Heatmap was generated by quantifying normalized tag counts across a 10 kb region in 50 bp increments centered on nucleosome-free cluster regions from ATAC-seq. The UF samples show greater normalized quantification. A dendrogram groups more similar peaks together by hierarchical clustering, which shows different patterns of chromatin accessibility in nucleosome-free cluster regions.

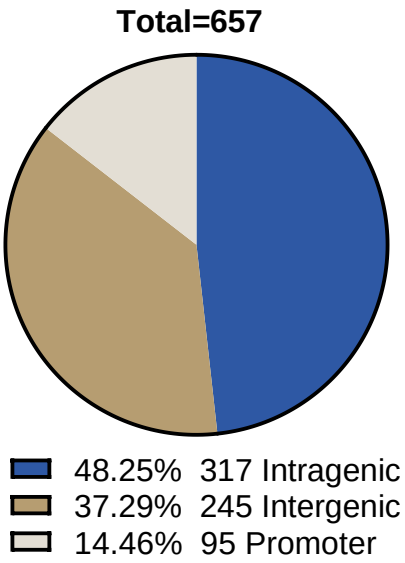

Figure 3.14: Annotation of nucleosome-free clusters identified from ATAC-seq peaks by genomic location

Nucleosome-free clusters from ATAC-seq analysis were assigned categorically to genomic features using refseq. Intragenic peaks were defined to lie within the coding region of any gene within the genome. Promoters were defined to be any peak within 5 kb of a transcription start site. Intergenic peaks were defined as any chromatin accessible region outside of a coding region and greater than 5 kb from a transcription start site.

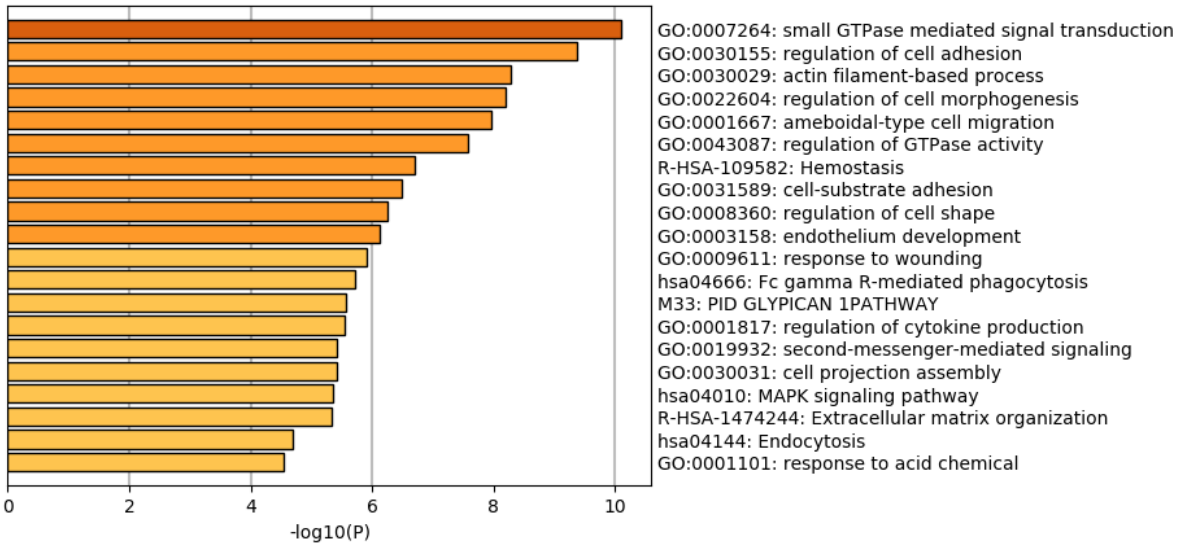

Figure 3.15: Enriched gene ontology terms for genes near nucleosome-free clusters in human aortic endothelial cells subjected to unidirectional flow

Metascape analysis of genes near nucleosome-free Clusters identified from ATAC-seq. Annotation was performed using HOMER to identify the closest gene to an accessible chromatin region induced by unidirectional flow. The gene ontology terms are listed with –log10 of the P-values for statistical significance.

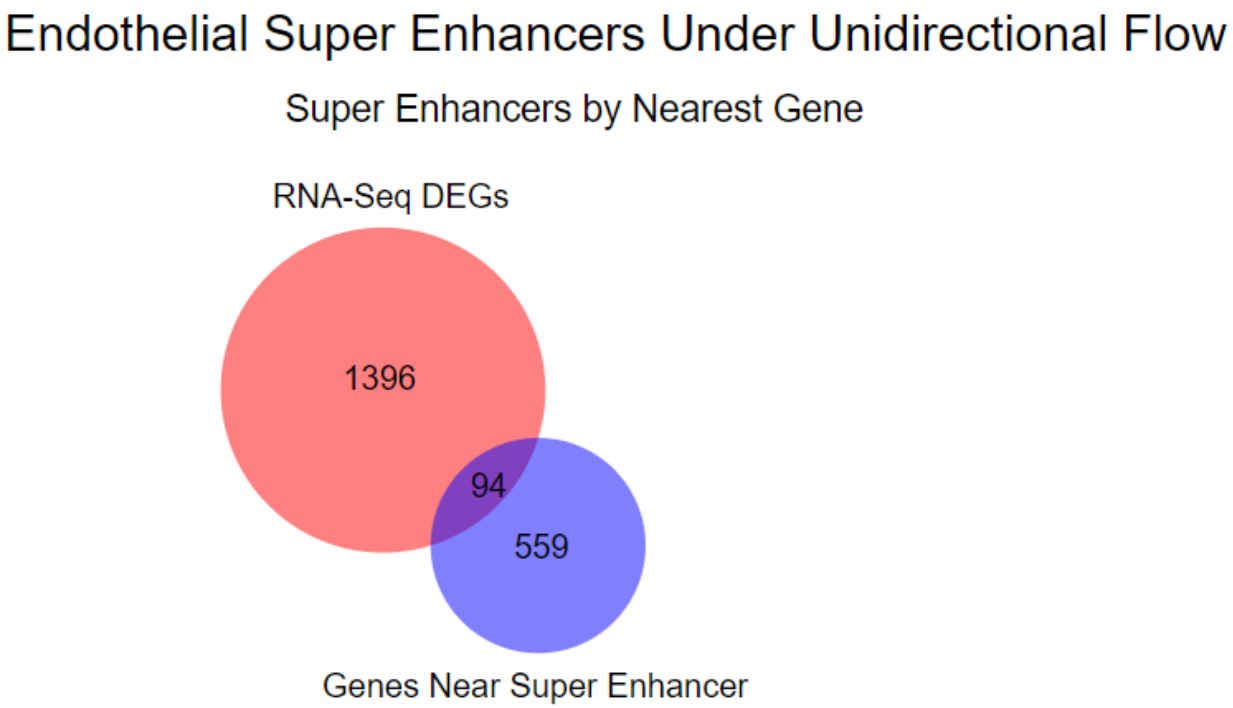

Figure 3.16: Intersection of genes with significant changes in gene expression and chromatin accessibility

Overlap of genes identified as differentially expressed from RNA-seq analysis, and annotated to be near a nucleosome-free cluster from ATAC-seq. The intersection of these two circles indicates a gene is both differentially expressed and is the nearest gene to at least one differential peak. These 94 genes are most likely to be directly, transcriptionally regulated by shear stress via a nucleosome-free cluster.

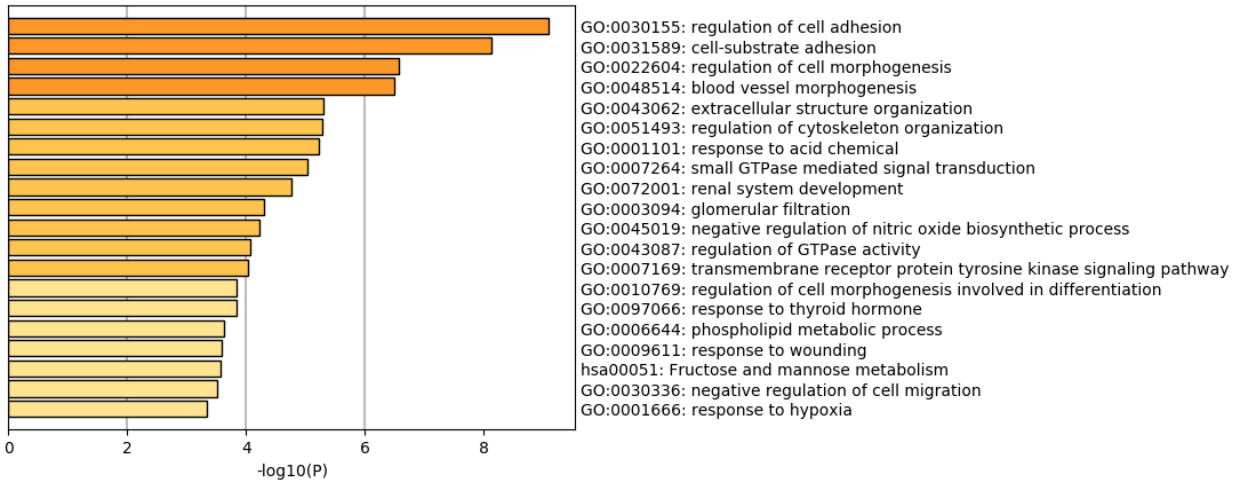

Figure 3.17: Metascape of overlapping 94 genes near a nucleosome-free cluster Identified by ATAC-seq and Flow-Sensitive from RNA-seq

Metascape analysis of the closest gene near one of 198 genes that are differentially expressed and lie near a nucleosome-free cluster ATAC-seq peak were enriched with genes regulated cell adhesion and morphology. These genes are mechanosensitive in response to flow and lie near predicted regulatory chromatin in endothelial cells.

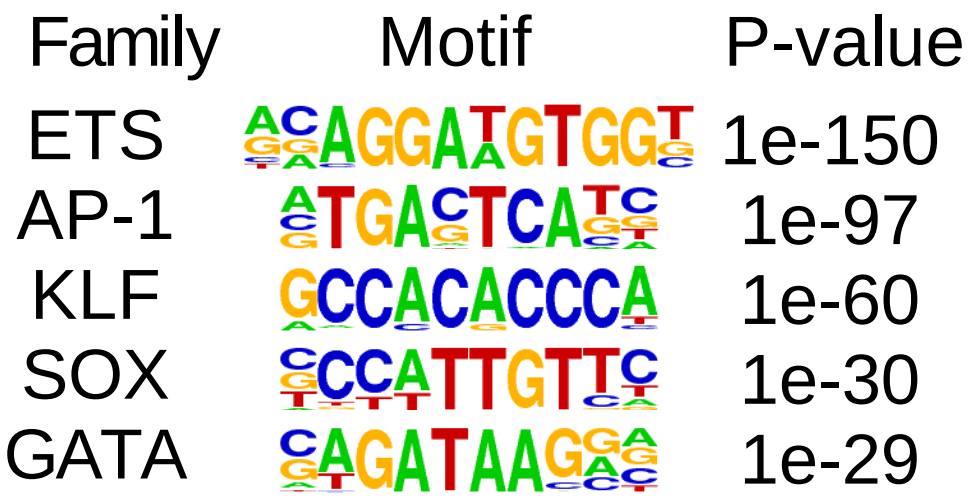

Figure 3.18: Motif analysis of differential peaks

The 2,433 differential peaks from ATAC-seq were analyzed for relevant DNA sequences. Over-representation of transcription factor consensus sequences were determined using HOMER. The top 5 most prevalent transcription factor families are presented with the identified motif and statistical significance.

Transcription Factor	Diff peaks	Transcription factor peaks	Intersected peaks	%Peaks intersect
ERG	2,433	5,940	635	26.10%
JUN	2,433	50,814	1,766	72.59%
NFκB	2,433	5,980	684	28.11%
NRF2	2,433	198	108	4.44%
IRF1	2,433	1,881	102	4.19%

Figure 3.19: Transcription factors bound to differential peaks from ATAC-seq in human aortic endothelial cells

ChIP-seq data of transcription factors gives the binding sites of regulatory proteins. ERG, JUN, NFκB, NRF2, and IRF1 have been immunoprecipitated from chromatin in human aortic endothelial cells. The genomic locations of these binding sites were overlapped with the differential peak locations using BEDtools. The proportion of peaks that overlap with at least one transcription factor binding site are reported as well the total number.

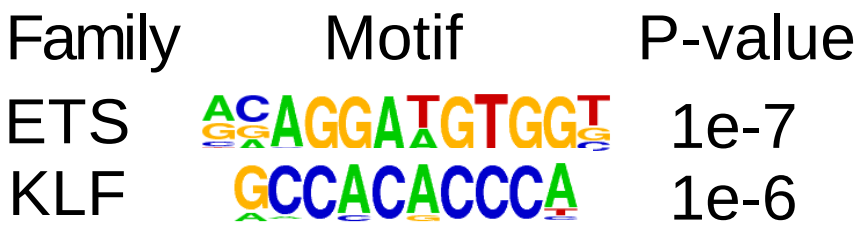

Figure 3.20: Motif analysis of nucleosome-free cluster regions

Over-representation of transcription factor consensus sequences were determined using HOMER. The top 2 most prevalent transcription factor families are presented with the identified motif and statistical significance.

Transcription Factor	Nucleosome-Free Clusters	Transcription factor peaks	Intersected peaks	%Peaks intersect
ERG	657	5,940	367	55.86%
JUN	657	50,814	608	92.54%
NFκB	657	5,980	393	59.82%
NRF2	657	198	93	14.16%
IRF1	657	1,881	90	13.70%

Figure 3.21: Transcription factors bound to nucleosome-free clusters in human endothelial cells from ATAC-seq data

ChIP-seq data of transcription factors gives the binding sites of regulatory proteins. ERG, JUN, NFκB, NRF2, and IRF1 have been immunoprecipitated from chromatin in human aortic endothelial cells. The genomic locations of these binding sites were overlapped with the nucleosome-free cluster locations using BEDtools. The proportion of nucleosome-free clusters that overlap with at least one transcription factor binding site are reported as well the total number.

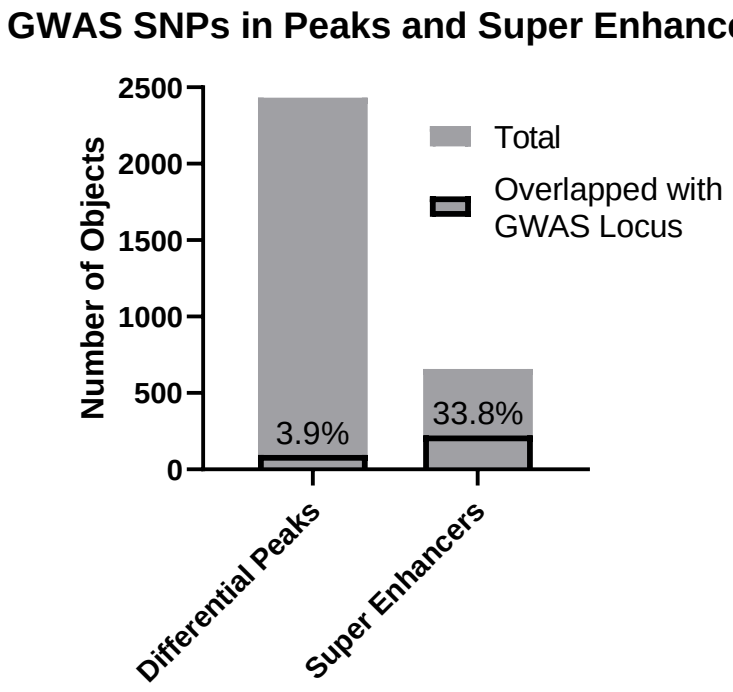

Figure 3.22: Overlap of differential peaks and nucleosome-free clusters with significant GWAS SNPs

The 2,433 differential peaks and 657 nucleosome-free clusters were overlapped with the GWAS catalog using BEDtools. The proportion of peaks containing a GWAS are reported. The overlap suggests a relation between a gene regulatory element and a human phenotype or disease.

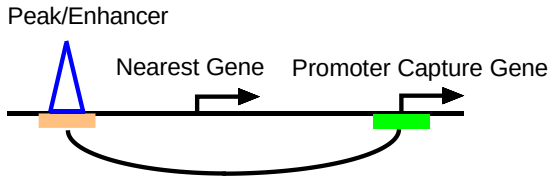

Figure 3.23: Promoter-capture HiC

Peak and enhancer have thus far been assigned using gene proximity based on 2D coordinates. The nucleus and genome are organized in 3D. Promoter-Capture Hi-C uses proximity ligation to join interacting DNA sequences. A library of probes that correspond to human gene promoters are pulled down in the assay and deep sequencing allows for detection of the interacting sequence. Thus, promoter-enhancer interactions can be assigned based on spatial proximity and not just sequence proximity and allows for higher confidence of direct interactions by regulatory elements in the genome.

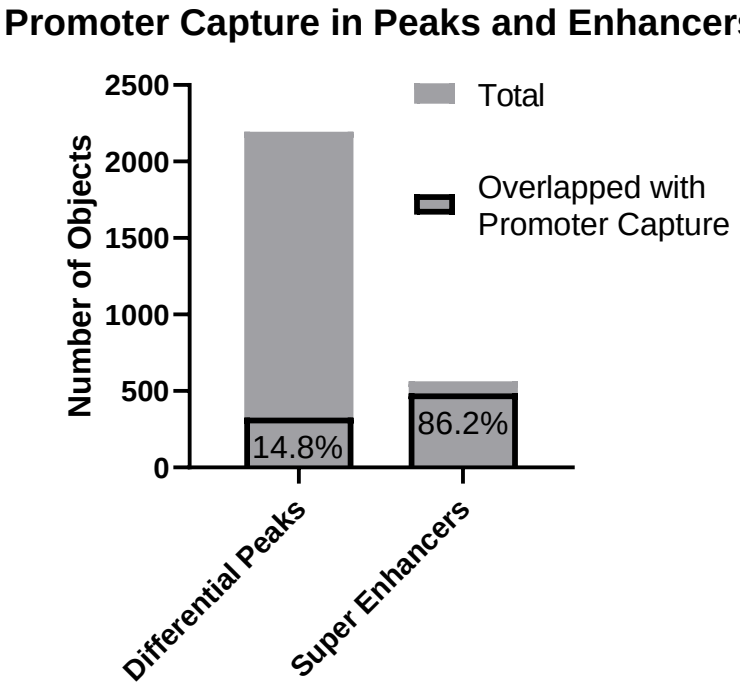

Figure 3.24: Overlap of differential peaks and nucleosome-free clusters with promoter-capture sites

Promoter-Capture Hi-C uses probes to detect interactions between gene promoters and regulatory sequences. All peaks or nucleosome-free clusters already within a promoter were excluded. Regulatory sequences were analyzed using BEDtools for sequence overlap with the 2,194 differential peaks and 563 nucleosome-free clusters with the detected promoter captures from HiC. Overlap suggests that a peak or enhancer is in direct contact with a gene promoter. Nearly all nucleosome-free clusters come in contact with a gene promoter.

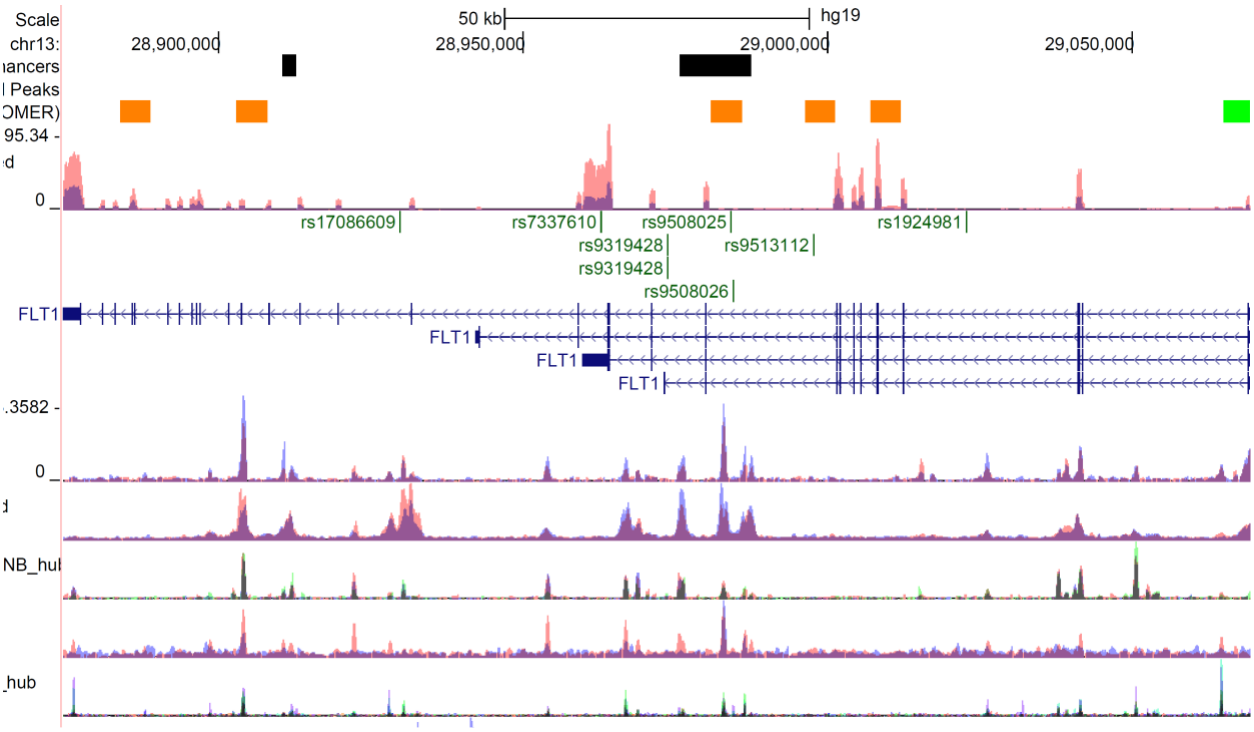

Figure 3.25: Genome browser image of FLT1 locus

Combining together several data sets shows the co-localization of a nucleosome-free cluster, promoter-capture, and GWAS SNP for coronary artery disease near FLT1. Promoter-capture Hi-C shows a direct interaction between a nucleosome-free cluster and the FLT1 promoter, suggesting a direct regulatory function. There are also peaks for ERG, JUN, and NFκB transcription factors suggesting potential regulatory mechanisms. Functional validation of this nucleosome-free cluster may have interesting insights into regulation of FLT1 and human coronary artery disease traits.

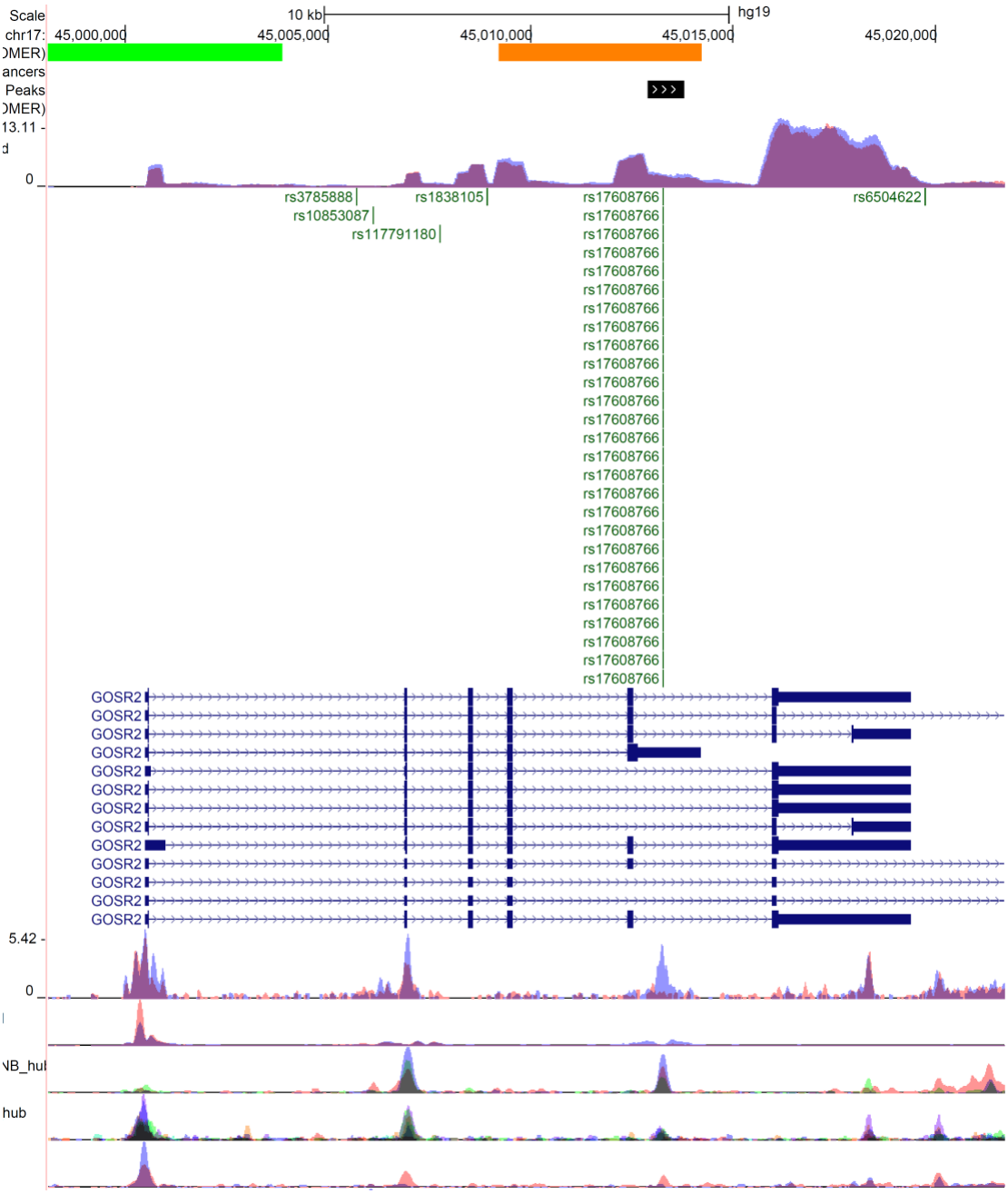

Figure 3.26: Genome browser image of GOSR2 locus

Combining together several data sets shows the co-localization of a differential chromatin accessible region, promoter-capture, and GWAS SNPs for many cardiovascular traits near GOSR2. Promoter-capture Hi-C shows a direct interaction between a nucleosome-free region and the GOSR2 promoter, suggesting a direct regulatory function. There are also peaks for ERG, JUN, and NFκB transcription factors suggesting potential regulatory mechanisms. Functional validation of this nucleosome-free cluster may have interesting insights into regulation of GOSR2 and human cardiovascular traits

Table 3.1: List of SNPs from GWAS Catalog that reside within a differential chromatin region from ATAC-seq

SNP	Reported Gene	Trait
rs829402	RAP1GAP	High light scatter reticulocyte count
rs6588505	LOC388630	Obesity-related traits
rs17114036	PPAP2B	Coronary artery disease or large artery stroke
rs17114036	PPAP2B	Coronary artery disease or ischemic stroke
rs17114036	PPAP2B	Coronary artery disease
rs17114036	PPAP2B	Coronary heart disease
rs17412403	LEPR, RP4-630A11.3	Blood protein levels
rs75494728	LRRC8D, ZNF326	Obsessive-compulsive symptoms
rs6702619	PALMD	Aortic root size
rs6702619	PALMD	Calcific aortic valve stenosis
rs11143145 2	ADORA3	Energy intake
rs12025416	CD58	Multiple sclerosis
rs11204682	ENSA	Non-albumin protein levels
rs11204682	ENSA	Albumin-globulin ratio
rs11204682	ENSA	Sum eosinophil basophil counts
rs11204682	ENSA	Neutrophil percentage of granulocytes
rs11204682	ENSA	Eosinophil percentage of white cells
rs11204682	ENSA	Eosinophil percentage of granulocytes
rs11204682	ENSA	Eosinophil counts
rs1387389	PBX1	Breast cancer (prognosis)
rs78346539	DNM3	Platelet distribution width
rs78346539	DNM3	Mean platelet volume
rs12023563	NR	Biliary atresia
rs6701752	RP5-940F7.2	White blood cell count (basophil)
rs6701752	RP5-940F7.2	Neutrophil percentage of granulocytes
rs6701752	RP5-940F7.2	Basophil percentage of white cells
rs6701752	RP5-940F7.2	Basophil percentage of granulocytes
rs2802728	SDCCAG8	Toenail selenium levels
rs2476491	IL2RA	Juvenile idiopathic arthritis (oligoarticular or rheumatoid factor-negative polyarticular)
rs12722497	IL2RA	Interleukin-2 receptor antagonist levels
rs12722497	NR	Strep throat
rs10795763	IL2RA	Systemic lupus erythematosus
rs2986971	intergenic	Non-alcoholic fatty liver disease histology (lobular)
rs10761779	REEP3	Liver enzyme levels
rs7919758	Intergenic	Self-reported math ability (MTAG)
rs4295981	NR	Blood protein levels
rs2403597	MICALCL, PARVA	Lung adenocarcinoma
rs12282742	NR	Bipolar disorder and schizophrenia

SNP	Reported Gene	Trait
rs4988321	LRP5	Heel bone mineral density
rs4988321	LRP5	Heel bone mineral density
rs4988321		Heel bone mineral density
rs4988321	LRP5	Heel bone mineral density
rs2111398	GPRC5A	Cutaneous malignant melanoma
rs7134108	Intergenic	Cognitive impairment test score
rs11689679 2	intergenic	Reticulocyte fraction of red cells
rs556429	SMAD9	Total body bone mineral density
rs11303659 5	CLDN10-AS1	Iron status biomarkers (ferritin levels)
rs1330950	LINCOO443	Response to anti-retroviral therapy (ddI/d4T) in HIV-1 infection (Grade 2 peripheral neuropathy)
rs36059107	NR	Schizophrenia
rs3783637	GCH1	Obesity-related traits
rs3783637	GCH1	Rheumatoid arthritis
rs3783637	GCH1	Obesity-related traits
rs11158026	GCH1	Parkinson's disease
rs11158026	GCH1	Parkinson's disease
rs11158026	GCH1	Parkinson's disease
rs12591650	RORA	Subcutaneous adipose tissue
rs7178424	C2CD4A	Height
rs12592898	CTSH	Blood protein levels
rs3784526	CHSY1	Monocyte percentage of white cells
rs3784526	CHSY1	Granulocyte percentage of myeloid white cells
rs2072327	ALDH3A1	Intraocular pressure
rs17608766	GOSR2	QRS duration
rs17608766	GOSR2	Blood pressure
rs17608766	GOSR2	Blood pressure
rs17608766	GOSR2	Systolic blood pressure
rs17608766	GOSR2	QRS duration
rs17608766	GOSR2	QRS duration
rs17608766	GOSR2	Systolic blood pressure
rs17608766	GOSR2	Systolic blood pressure
rs17608766	NR	QRS complex (Cornell)
rs17608766	GOSR2	Mean arterial pressure
rs17608766	GOSR2	Coronary artery disease
rs17608766	GOSR2	Coronary artery disease
rs17608766	GOSR2	Systolic blood pressure
rs17608766	GOSR2	Pulse pressure
rs17608766	GOSR2	Aortic root size
rs17608766	NR	QRS duration
rs1815823	NR	Neuroticism

SNP	Reported Gene	Trait
rs29943	KCTD15	Body mass index in physically active individuals
rs29943	KCTD15	Body mass index
rs29943	KCTD15	Body mass index
rs2965162	NR	Alzheimer's disease or family history of Alzheimer's disease
rs11083766	MARK4, AC006126.3	Platelet distribution width
rs11083766	MARK4, AC006126.3	Platelet count
rs11667509	EXOC3L2	Mean platelet volume
rs4666014	RBKS	Hematocrit
rs1529102	RPL23AP37, LOC100128607	Breast size
rs116165775	ETAA1	Lung cancer in ever smokers
rs7601485	Intergenic	Educational attainment (MTAG)
rs62182887	TMEM194B, NAB1, MFSD6, INPP1	Daytime sleep phenotypes
rs17572109	ARPC2	Platelet distribution width
rs17572109	ARPC2	Platelet count
rs17572109	ARPC2	Mean platelet volume
rs2382817	NR	Ulcerative colitis
rs2382817	CXCR2, CXCR1, GPBAR1, AAMP, RUFY4, ARPC2, PNKD, TMBIM1	Inflammatory bowel disease
rs2382817	ARPC2, TMBIM1, CTDSP1, SLC11A1, CXCR2, CXCR1, PNKD	Inflammatory bowel disease
rs2382817	NR	Crohn's disease
rs430086	MACROD2	Percentage gas trapping
rs2295888	MYH7B, EDEM2, TRPC4AP, PROCR	Prothrombin time
rs6782029	VGLL4	Anorexia nervosa
rs12633551	ATG7	HDL cholesterol levels
rs35199932	NR	Maximum cranial width
rs2643826	LOC105377005, SLC4A7	Systolic blood pressure x alcohol consumption interaction (2df test)
rs2643826	LOC105377005, SLC4A7	Systolic blood pressure x alcohol consumption interaction (2df test)
rs2643826	LOC105377005, SLC4A7	Mean arterial pressure x alcohol consumption interaction (2df test)
rs2643826	LOC105377005, SLC4A7	Mean arterial pressure x alcohol consumption (light vs heavy) interaction (2df test)
rs2643826	LOC105377005, SLC4A7	Diastolic blood pressure x alcohol consumption interaction (2df test)

SNP	Reported Gene	Trait
rs2643826	LOC105377005, SLC4A7	Diastolic blood pressure x alcohol consumption interaction (2df test)
rs13077017	DNASE1L3, FLNB	Eating disorders (purging via substances)
rs11719450	intergenic	Hemoglobin concentration
rs34116314	NR	Blood protein levels
rs62271373		Heel bone mineral density
rs967367	ST6GAL1	Albumin-globulin ratio
rs5001409	ST6GAL1	Non-albumin protein levels
rs6771736	LPP	Glaucoma (primary open-angle)
rs17003427	NR	Circulating odd-numbered chain saturated fatty acid levels (C23:0)
rs12644284	TRIM2	Multiple sclerosis
rs735949	ACSL1	Type 2 diabetes
rs7713972	Intergenic	Blood protein levels
rs2199161	MAP1B	Attention deficit hyperactivity disorder
rs147399060		Heel bone mineral density
rs72792324	DIAPH1	Mean platelet volume
rs72813183	MXD3	Obesity-related traits
rs72813183	MXD3	Obesity-related traits
rs10479542	RUFY1	Lung cancer
rs11242704	RP11-157J24.1	Response to hepatitis C treatment
rs2012011	GPR116	Monocyte count
rs7757421	NR	Vaginal discharge (itching)
rs71575922	SYNE1	Uterine fibroids
rs71575922	SYNE1	Endometriosis
rs71575922	SYNE1	Endometriosis
rs926657	C6orf99, RSPH3, TAGAP, FNDC1, SOD2, PNLDC1, MAS1, IGF2R	Lipoprotein (a) levels
rs926657	TAGAP	Lipoprotein (a) - cholesterol levels
rs4719697	NR	Adolescent idiopathic scoliosis
rs875971	CRCP	Aortic root size
rs342275	FLJ36031, PIK3CG	Platelet count
rs4265116	Intergenic	Highest math class taken (MTAG)
rs7784447	MDFIC	Obesity
rs41748	MET	Resting heart rate
rs41748	MET	Resting heart rate
rs13226190	GIMAP6	Fibrinogen levels
rs202280	intergenic	Overall survival in serous epithelial ovarian cancer treated with paclitaxel and cisplatin
rs1487241		Heel bone mineral density
rs1487241	PCAT1	Heel bone mineral density
rs1487241	PCAT1	Heel bone mineral density
rs1487241	PCAT1	Heel bone mineral density

SNP	Reported Gene	Trait
rs10968457	KIAA0020	Obesity-related traits
rs28558845	GLIS3	Systolic blood pressure
rs10511652	SH3GL2, ADAMTSL1	Spherical equivalent or myopia (age of diagnosis)
rs10511652	SH3GL2, ADAMTSL1	Myopia
rs334354	NR	Myopia (age of diagnosis)
SNP	Reported Gene	Trait
rs3818638	NDUFA8	Obesity-related traits
rs10818894	DENND1A	Adverse response to chemotherapy in breast cancer (alopecia) (cyclophosphamide+doxorubicin+/-5FU)
rs2519796	VAV2	Hemoglobin concentration
rs2519796	VAV2	Hematocrit

Table 3.2: List of SNPs from GWAS Catalog that reside within a nucleosome-free cluster region from ATAC-seq

SNP	Reported Gene	Trait
rs79280654	EFHD2, RP3-467K16.4, FHAD1, CELA2A	Alcoholic chronic pancreatitis
rs71636778	ARID1A	Monocyte percentage of white cells
rs71636778	ARID1A	Granulocyte percentage of myeloid white cells
rs3737801	FGR	Systolic blood pressure
rs3753366	Intergenic	Colorectal cancer
rs12750249	NR	IgG glycosylation
rs17396055	ARHGAP29	Systolic blood pressure
rs2660304	DPYD, MIR137, MIR137HG, MIR2682	Schizophrenia
rs2660304	NR	Schizophrenia
rs6702619	PALMD	Aortic root size
rs6702619	PALMD	Calcific aortic valve stenosis
rs12025416	CD58	Multiple sclerosis
rs12045807	CTSS, CTSK, ARNT	Congenital left-sided heart lesions
rs6657658	THEM4, S100A10	Urinary metabolite ratios in chronic kidney disease
rs6587640	S100A10	Urinary metabolite concentrations in chronic kidney disease
rs859362	TNR	Corneal astigmatism
rs12125543	SOX13	Itch intensity from mosquito bite
rs3795579	SOX13	Itch intensity from mosquito bite adjusted by bite size
rs3795579	SOX13	Itch intensity from mosquito bite adjusted by bite size
rs112725747	NR	Skin pigmentation
rs2483058	SRGAP2	Cholesterol and Triglycerides
rs12023563	NR	Biliary atresia
rs113711540	TGFB2	Intraocular pressure
rs2799098		Heel bone mineral density
rs17047703	TGFB2	Intraocular pressure
rs4846476	TGFB2	Central corneal thickness
rs628839	TGFB2, LYPLAL1	Rosacea symptom severity
rs73103335	LYPLAL1-AS1	Intraocular pressure
rs140084787	CAPN2	Febrile seizures (MMR vaccine-related)
rs11583244	SRP9	Endometrial cancer
rs9782955	LYST	Systemic lupus erythematosus
rs9782955	LYST	Systemic lupus erythematosus
rs6701752	RP5-940F7.2	White blood cell count (basophil)
rs6701752	RP5-940F7.2	Neutrophil percentage of granulocytes

SNP	Reported Gene	Trait
rs6701752	RP5-940F7.2	Basophil percentage of white cells
rs6701752	RP5-940F7.2	Basophil percentage of granulocytes
rs9970896	RP5-940F7.2	Monocyte percentage of white cells
rs4745982	HK1	Glycated hemoglobin levels
rs10823343	HK1	Hemoglobin A1c levels
rs10823343	HK1	Glycated hemoglobin levels
rs16926246	HK1	Hemoglobin levels
rs16926246	HK1	Glycated hemoglobin levels
rs16926246	HK1	Hemoglobin
rs16926246	HK1	Hematocrit
rs17476364	HK1	Hematocrit
rs17476364	HK1	Hemoglobin concentration
rs17476364	HK1	High light scatter reticulocyte count
rs17476364	HK1	High light scatter reticulocyte percentage of red cells
rs17476364	HK1	Mean corpuscular hemoglobin
rs17476364	HK1	Red blood cell count
rs17476364	HK1	Mean corpuscular volume
rs17476364	HK1	Reticulocyte fraction of red cells
rs17476364	HK1	Red cell distribution width
rs17476364	HK1	Reticulocyte count
rs72805692	HK1	Immature fraction of reticulocytes
rs10159477	HK1	Red blood cell traits
rs7072268	HK1	Glycated hemoglobin levels
rs11596587	HK1	Ischemic stroke (cardioembolic)
rs72821233	Intergenic	Cognitive performance
rs4295981	NR	Blood protein levels
rs55710213	SCD	White blood cell count (basophil)
rs6585202	TCF7L2	Breast cancer
rs2362515	NR	Childhood ear infection
rs3781458	NR	Male-pattern baldness
rs3781458	FAM53B	Male-pattern baldness
rs3781454	FAM53B, RP11-12J10.3	White blood cell count
rs11025109	NR	Antipsychotic drug-induced QTc interval change in schizophrenia
rs831636	CD59	Blood protein levels
rs11039436	intergenic	White blood cell count
rs11039436	intergenic	Hematocrit
rs6592362	intergenic	Inflammatory bowel disease
rs2111398	GPRC5A	Cutaneous malignant melanoma
rs3741578	ZBTB39	Asthma
rs1021469	Intergenic	Self-reported math ability (MTAG)
rs6538452	NR	Adolescent idiopathic scoliosis
rs7134108	Intergenic	Cognitive impairment test score

SNP	Reported Gene	Trait
rs9508025	FLT1	Coronary heart disease
rs9508026	NR	Myopia (age of diagnosis)
rs563538	SMAD9	Spontaneous preterm birth (preterm birth)
rs140971918	NR	Prostate cancer
rs61989804	intergenic	Midgestational circulating levels of PCBs (fetal genetic effect)
rs76734026	SPTB	Pediatric areal bone mineral density (radius)
rs117672662	ACTN1	Mean platelet volume
rs117672662	ACTN1	Platelet distribution width
rs117672662	ACTN1	Platelet count
rs76745467	ACTN1	Cognitive decline rate in late mild cognitive impairment
rs76745467	ACTN1	Cognitive decline rate in late mild cognitive impairment
rs4902681	Intergenic	Educational attainment (MTAG)
rs12434701		Response to paliperidone in schizophrenia (positive Marder score)
rs59721556	intergenic	Red cell distribution width
rs59721556		Response to paliperidone in schizophrenia (positive Marder score)
rs59721556		Response to paliperidone in schizophrenia (PANSS score)
rs59721556		Response to paliperidone in schizophrenia (Multivariate)
rs59721556		Response to paliperidone in schizophrenia (CGI-S score)
rs59721556		Response to paliperidone in schizophrenia (CGI-S score)
rs4900069	C14orf159	Resting heart rate
rs1885073	NR	Adolescent idiopathic scoliosis
rs117540572	DUT	Blood protein levels
rs17352842	FBN1	Hepcidin/ferritin ratio
rs11070836	TNFAIP8L3, CYP19A1	Airway wall thickness
rs11070836	TNFAIP8L3	Airway wall thickness
rs11070836	TNFAIP8L3	Airway wall thickness
rs11071657	C2CD4B	Fasting blood glucose
rs55703462	RP11-69G7.1	HDL cholesterol
rs2652834	LACTB	HDL cholesterol
rs2652834	LACTB	HDL cholesterol
rs4984390	MCTP2	Drug-induced liver injury (flucloxacillin)
rs11637339	LRRC28, MEF2A	Cervical cancer
rs28479400	intergenic	Neuritic plaque
rs2173714	NR	General cognitive ability
rs3790087	WWP2	Diverticular disease

SNP	Reported Gene	Trait
rs7342694	PLCG2	Coronary artery disease
rs7199941	PLCG2	Coronary artery disease
rs7199941	PLCG2	Coronary artery disease
rs850526	Intergenic	Primary sclerosing cholangitis
rs627386	NR	IgG glycosylation
rs8074980	intergenic	Hippocampal sclerosis
rs8074980	CUEDC1	Multiple sclerosis--Brain Glutamate Levels
rs9892152	PECAM1	Coronary artery disease
rs9900062	PLEKHM1P	QT interval
rs7229639	SMAD7	Colorectal cancer
rs7229639	NR	Colorectal cancer
rs7229639	SMAD7	Colorectal cancer
rs7229639	SMAD7	Colorectal cancer
rs4939567	NR	Colorectal cancer
rs11874392	NR	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	SMAD7	Colorectal cancer
rs4939827	NR	Colorectal cancer
rs4939827	NR	Colorectal cancer
rs12953717	SMAD7	Colorectal cancer
rs12953717	SMAD7	Colorectal cancer
rs7226855	SMAD7	Colorectal cancer
rs2337106		Heel bone mineral density
rs2337106	SMAD7	Glomerular filtration rate
rs2337106	SMAD7	Creatinine levels
rs2337106	SMAD7	Red blood cell count
rs2337106	SMAD7	Hematocrit
rs2878889	SMAD7	Hemoglobin concentration
rs35715456	SMAD7	Parental longevity (at least one long-lived parent)
rs11083766	MARK4, AC006126.3	Platelet distribution width
rs11083766	MARK4, AC006126.3	Platelet count
rs6716724	E2F6	Sudden cardiac arrest
rs6709385	NR	Adolescent idiopathic scoliosis
rs1534422	MIR4262, TRIB2	Multiple sclerosis
rs1534422	TRIB2	Graves' disease
rs1534422	TRIB2	Autoimmune thyroid diseases (Graves disease or Hashimoto's thyroiditis)
rs1534422	intergenic	Type 1 diabetes
rs62124708	Intergenic	Self-reported math ability (MTAG)

SNP	Reported Gene	Trait
rs478222	EFR3B, 3NCOA1, C2orf79, CENPO, ADCY3, DNAJC27, POMC, DNMT3A	Type 1 diabetes
rs1530016	EFR3B	Body mass index in non-asthmatics
rs6740838	AFF3, LINC01104	Non-albumin protein levels
rs10185424	NR	Crohn's disease
rs10185424	NR	Ulcerative colitis
rs10179654	IL1RL1	Blood protein levels
rs12470864	IL1RL2, IL18R1	Allergic disease (asthma, hay fever or eczema)
rs12470864	NR	Blood protein levels
rs6723108	TMEM163	Type 2 diabetes
rs13397529		Educational attainment (years of education)
rs16861137	NR	Moyamoya disease
rs78585132	NR	Facial morphology (factor 18)
rs17572109	ARPC2	Platelet distribution width
rs17572109	ARPC2	Platelet count
rs17572109	ARPC2	Mean platelet volume
rs6084946	NR	IgG glycosylation
rs3746428	EDEM2	HDL cholesterol
rs2295888	MYH7B, EDEM2, TRPC4AP, PROCR	Prothrombin time
rs6065325	NR	Depressive symptoms (MTAG)
rs6102324	NR	Left ventricle wall thickness
rs127430	APCDD1L, STX16	Cardiovascular risk factors (age interaction)
rs6005551	Intergenic	Facial depth
rs134041	MN1	Pulse pressure
rs17041333	NR	3-hydroxy-1-methylpropylmercapturic acid levels in smokers
rs1027327	LMCD1/Dyxin	Narcolepsy
rs171408	LMCD1	Mitral valve prolapse
rs164958	NR	Pediatric areal bone mineral density (radius)
rs4955158	TGFBR2, GADL1	Night sleep phenotypes
rs905456	NR	Positive affect
rs17029445	NR	Adolescent idiopathic scoliosis
rs9311171	CTDSPL	Prostate cancer
rs7625806	DLEC1	Carotid intima media thickness
rs111231532	TTC21A, XIRP1, CX3CR1, CSRNP1	Cognitive decline rate in late mild cognitive impairment
rs111231532	TTC21A, CSRNP1, XIRP1, CX3CR1	Cognitive decline rate in late mild cognitive impairment
rs73088024	NR	General cognitive ability
rs17005647	FRMD4B	Atrial fibrillation
rs1483890	FRMD4B	Resting heart rate

SNP	Reported Gene	Trait
rs35515217	RUVBL1	Post bronchodilator FEV1/FVC ratio
rs35515217	RUVBL1	Post bronchodilator FEV1
rs36069551	RUVBL1	Post bronchodilator FEV1
rs11714052	RUVBL1	Post bronchodilator FEV1/FVC ratio
rs11714052	RUVBL1	Post bronchodilator FEV1
rs11709377	RUVBL1	Post bronchodilator FEV1
rs36059338	RUVBL1	Post bronchodilator FEV1
rs34228187	RUVBL1	Post bronchodilator FEV1
rs34562493	RUVBL1	Post bronchodilator FEV1
rs6787614	RUVBL1	Post bronchodilator FEV1
rs11720239	RUVBL1	Post bronchodilator FEV1
rs34709844	RUVBL1	Post bronchodilator FEV1
rs11715661	RUVBL1	Post bronchodilator FEV1
rs55777628	RUVBL1	Post bronchodilator FEV1
rs11719079	RUVBL1	Post bronchodilator FEV1
rs67674204	RUVBL1	Post bronchodilator FEV1
rs34955225	RUVBL1	Post bronchodilator FEV1
rs35437840	RUVBL1	Post bronchodilator FEV1
rs35804055	RUVBL1	Post bronchodilator FEV1
rs13082038	RUVBL1	Post bronchodilator FEV1
rs13062252	RUVBL1	Post bronchodilator FEV1
rs60387553	RUVBL1	Post bronchodilator FEV1
rs35421223	RUVBL1	Post bronchodilator FEV1/FVC ratio
rs35421223	RUVBL1	Post bronchodilator FEV1
rs33998546	RUVBL1	Post bronchodilator FEV1
rs17787940	WWTR1	Preeclampsia
rs6799682	intergenic	Anxiety disorder
rs17776120	NR	IgG glycosylation
rs17776120	NR	IgG glycosylation
rs3821819	NR	IgG glycosylation
rs967367	ST6GAL1	Albumin-globulin ratio
rs9855415	LOC647323	Carotid intima media thickness
rs4687530	NR	Reaction time
rs41530644	NR	Adolescent idiopathic scoliosis
rs114518130	IGFBP7	Brain volume in infants (grey matter)
rs35769988	IGFBP7	Alcohol consumption (maxi-drinks)
rs10049992	NR	IgG glycosylation
rs1718849	IGFBP7	Blood protein levels
rs1718845	IGFBP7	Diastolic blood pressure
rs11434868		Heel bone mineral density
rs340630	AFF1	Systemic lupus erythematosus
rs2452591	PDLIM5	Lewy body disease
rs17054392	PALLD	Response to iloperidone treatment (QT prolongation)

SNP	Reported Gene	Trait
rs7696431	PALLD	Coronary artery disease
rs7696431	PALLD	Coronary artery disease
rs7696431	PALLD	Coronary artery disease
rs6880621	intergenic	Red blood cell count
rs972761	intergenic	Hemoglobin concentration
rs972761	intergenic	Hematocrit
rs7714709	C1QTNF3-AMACR, RAI14	Glomerular filtration rate
rs7714709	C1QTNF3-AMACR, RAI14	Creatinine levels
rs11742570	NR	Crohn's disease
rs11742570	PTGER4	Crohn's disease
rs11742570	PTGER4	Crohn's disease
rs11742570	PTGER4	Inflammatory bowel disease
rs11742570	NR	Ulcerative colitis
rs11742570	Intergenic	Crohn's disease
rs11742570	Intergenic	Inflammatory bowel disease
rs11742570	Intergenic	Ulcerative colitis
rs6451493	PTGER4	Ulcerative colitis
rs11952819	ZNF366	Spherical equivalent or myopia (age of diagnosis)
rs341338	NR	DNA methylation variation (age effect)
rs12521868	c5orf56	Diastolic blood pressure
rs12521868	C5orf56	Diastolic blood pressure
rs12521868	IL3, IRF1, SLC22A5, SLC22A4	Crohn's disease
rs6866614	C5orf56	Perceived unattractiveness to mosquitoes
rs7713065	C5orf56	Lung function (FEV1/FVC)
rs11745587	C5orf56	Asthma
rs6894249	C5orf56	LDL cholesterol
rs6894249	IRF1, C5orf56	Juvenile idiopathic arthritis (oligoarticular or rheumatoid factor-negative polyarticular)
rs6894249	SLC22A5, IRF1	Asthma
rs3749833	C5orf56	Allergic disease (asthma, hay fever or eczema)
rs2522056	C5orf56	Lymphocyte counts
rs2522056	IRF1	Fibrinogen
rs72792324	DIAPH1	Mean platelet volume
rs979455	CCDC69	Obesity-related traits
rs9501753	intergenic	Response to anti-retroviral therapy (ddI/d4T) in HIV-1 infection (Grade 3 peripheral neuropathy)
rs6916016	LOC285766, DUSP22	Non-albumin protein levels
rs17138114	KU-MEL-3, PECI, ECI2	Obesity-related traits
rs6942227	CMAHP	Bipolar disorder lithium response (categorical) or schizophrenia

SNP	Reported Gene	Trait
rs12209477	GLO1	Blood protein levels
rs12191082	GPR116	Mean platelet volume
rs2012011	GPR116	Monocyte count
rs2817782	intergenic	Electroencephalogram traits
rs1080261	NR	Plasma omega-6 polyunsaturated fatty acid levels (arachidonic acid)
rs577721086	RSPO3	Bone mineral density (spine)
rs17057678	ARHGAP18	Antineutrophil cytoplasmic antibody-associated vasculitis
rs3799385	NR	Adolescent idiopathic scoliosis
rs9364687	NR	Body mass index
rs9364687	CAHM	Body mass index
rs9364687	CAHM	Body mass index
rs9364687	CAHM	Body mass index
rs4708620		Heel bone mineral density
rs4708620	KIF25-AS1	Heel bone mineral density
rs4507656	RAPGEF5	Diastolic blood pressure
rs2237457	DDC, GRB10, LOC100128600	Schizophrenia (treatment resistant)
rs723527	EGFR	Glioblastoma (age-stratified)
rs723527	EGFR	Glioblastoma (age-stratified)
rs723527	EGFR	Glioblastoma
rs59060240	NR	Glioma
rs59060240	NR	Glioblastoma
rs7785013	NR	Adolescent idiopathic scoliosis
rs2158836	DLD, SLC26A3, LAMB1	Ulcerative colitis
rs2072209	LAMB1	IgG glycosylation
rs2072209	LAMB1	IgG glycosylation
rs2072209	LAMB1	IgG glycosylation
rs2072209	LAMB1	IgG glycosylation
rs38855	MET	Triglyceride levels
rs38855	MET	Triglycerides
rs10237377	PARP12	Coronary artery disease
rs10237377	PARP12	Coronary artery disease
rs10155912	PARP12	Vitiligo
rs13226190	GIMAP6	Fibrinogen levels
rs2220321	FAM84B	Obesity-related traits
rs7011138	PCAT1	Intracranial aneurysm
rs10974465	GLIS3, SLC1A1	Squamous cell lung carcinoma
rs1780159	C9orf52	Calcium levels
rs2282335	TJP2	Renal sinus fat
rs11142400		Heel bone mineral density
rs28537805	Intergenic	Highest math class taken
rs963265	intergenic	Migraine with aura

SNP	Reported Gene	Trait
rs13293114	RP11-439K3.1	Breast cancer and/or colorectal cancer
rs12339094	NR	Smoking quantity
rs4837181	ENG	Blood protein levels

CHAPTER 4: Summary and Conclusions

The purpose of the experiments reported in this dissertation was to systematically investigate the contribution of human genetic variants to endothelial mechanoresponses to hemodynamics, a previously unrecognized layer of molecular controls. To accomplish this, I performed whole-genome transcriptome profiling, chromatin accessibility, and chromatin interactions to probe mechanosensitive cis-regulatory elements, followed by experimental validations such as genome editing as well as molecular and phenotypic analyses. The list of these putative mechanosensitive cis-regulatory elements was interrogated with human disease GWAS results; the intersections of these datasets resulted in a cohort of cis-regulatory elements, which respond to hemodynamics, regulate endothelial gene expression, and could causatively contribute to the inter-individual susceptibility to human traits and diseases. This data is significant in understanding the direct transcriptional regulation of genes by blood flow in endothelial cells that dictate endothelial cell phenotypes with relevance to complex human diseases. Particularly, one of these mechanosensitive cis-regulatory elements, chr1:56962213–56963412, was experimentally validated for its major role in regulating key endothelial responses to hemodynamics as a function of genetic predisposition to CAD (Krause *et al.*, 2018).

4.1 Endothelial PLPP3 expression in humans is regulated by human genetic variation and is relevant to CAD

In the work presented I have delineated the molecular mechanisms in which a human genetic variant regulates key endothelial mechanotransduction responses related to CAD susceptibility. Endothelial mechanotransduction is essential to control vascular heath and disease, as seen by disturbed-flow promoting atherogenesis in human arteries (Chiu and Chen, 2011; Davies *et al.*, 2013; Gimbrone and Garcia-Cardena, 2016). Genetic linkage and GWAS

have identified inheritable risk factors to complex human diseases, and yet the intersection between genetic predisposition, mechanotransduction mechanisms, and human disease heritability is enigmatic (Watkins and Farrall, 2006; Khera and Katherisan, 2017; Inouye *et al*., 2018). This study demonstrated that the genetic variant at rs17114036 controls key endothelial mechanosensing mechanisms by regulating endothelial gene expression of PLPP3 which hydrolyzes the vascular endothelium activating lipid lysophosphatidic acid (LPA) (Krause *et al*., 2018). Moreover, these results support a previously unappreciated role of genetic variants and related cis-regulatory elements in mediating endothelial mechanoresponse to disease-associated hemodynamic forces.

Much of the human genetic studies related to CAD have focused on the roles of lipids and lipoproteins, due in part to their established roles in the pathogenesis or prevention of atherosclerosis (Brown and Goldstein, 1974; Lehrman *et al*., 1985). Indeed, the most widely used CAD therapies aim to lower lipid levels and have shown a benefit to disease outcomes (Navarese *et al*., 2015). CAD GWAS have been performed to understand the polygenic structure of such a complex disease with strong genetic and environmental influences (Samani *et al*., 2007; Schunkert *et al*., 2011; Dichgans *et al*., 2014; Nikpay *et al*., 2015; Howson *et al*., 2017; Van der Harst *et al*., 2018). To date, GWAS have uncovered 161 loci that are associated with CAD susceptibility and protection as both are critical to understanding the complex interplay of biology and environment that ultimately dictate disease progression and severity. Nonetheless, the phenomenon still stands that atherosclerosis is a focal disease as dictated by disturbed flow hemodynamics. This study shows that blood flow regulates the epigenome and transcriptome differently under disturbed and unidirectional flow in cultured human endothelial cells. These

results show human genetics contribute to flow-mediated endothelial functions which can be intrinsically distinct between individuals.

Genetic variation at rs17114036 is in fact a risk factor for CAD, independent of common behavior and lifestyle factors such as diabetes and lipid levels (Schunkert *et al*., 2011; Dichgans *et al*., 2014; Krause *et al*., 2018). The results of this study suggest that PLPP3, which is regulated by atherosclerosis-associated hemodynamics as well the genetic variants, may provide a perpendicular therapeutic avenue to current treatments to promote endothelial health and attenuate atherogenic mechanisms.

Although it is proposed that interactions between genetics, lifestyle, and environmental factors contribute to complex disease risk, the underlying biology linking associated variants to phenotypes is a major challenge for the field (Musunuru *et al*., 2011; Harsimendy *et al*., 2011; Smemo *et al*., 2014; Gupta *et al*., 2017). There are instances where this has been done successfully. EDN1 and the 9p21 loci are associated with CAD and have been investigated to mechanistically link genetic variations to changes or perturbation in endothelial gene regulation (Harsimendy *et al*., 2011; Gupta *et al*., 2017). These studies are all limited in that none of them take into account the contribution of hemodynamic forces which drive the focal nature of atherosclerosis. The studies presented in this work establish an important role of CAD predisposition in endothelial mechanotransduction mechanisms responsible for the development of atherosclerosis in arterial areas exposed to disturbed flow. Using both genetics and reductionist approaches, our results point to a flow-sensitive endothelial enhancer located in the PLPP3 intron 5 that confers the CAD predisposition associated with rs17114036 identified by GWAS. There are also many opportunities from the results of these experiments to repeat the

experimental framework and study hundreds of different mechanosensitive genes and regulatory elements that mediate human traits and disease susceptibility.

4.2 A causal nucleotide(s) in the CAD loci 1p32.2

While these experiments have shown that deletion of a ~70bp region near rs17114036 significantly reduces the enhancer activity and consequently decreases endothelial PLPP3 expression, it remains to be explored whether rs17114036 solely contributes to the CAD susceptibility associated with chromosome 1p32.2. Due to the rarity of the minor allele it is extremely difficult to obtain primary cells from donors with homozygous genotype for the minor allele. Even with CRISPR technologies, it remains technically challenging to replace a single nucleotide of interest in adult endothelial cells. It is possible that the use of a base-converting enzymes that can be targeted using dCas9 can change a single nucleotide in a specific DNA sequence (Komor *et al*., 2016). Yet this still remains a barrier as the sequence change required would be from T to C while current enzymes only convert the opposite reaction (Komor *et al*., 2016). It may also be feasible to perform the homology directed repair (HDR) within induced pluripotent stem cells (iPSCs) that can be differentiated into adult endothelial cells, which has been done previously (Gupta *et al*., 2017). This still harbors a limitation as the epigenetic state of the iPSC may not exactly match an adult human endothelial cell type, which can be a critical flaw into investigating the epigenetic role of this locus in gene. Furthermore, regulation of this enhancer and/or PLPP3 itself may be altered during the differentiation process, which highlights the need to validate results and observations with primary cell models whenever possible.

4.3 Additional cis-regulatory mechanisms associated with CAD locus 1p32.2

Although our studies demonstrated that PLPP3 is the causal gene contributing to the etiology of atherosclerosis associated with the CAD locus 1p32.2 and rs17114036 is a major mediator of disease risk, a few open questions remained (Krause *et al*., 2018). For example, CAD is a chronic and progressive disease that develops over the lifetime of an individual and overt symptoms do not appear for decades (Chiu and Chen, 2011; Davies *et al*., 2013; Gimbrone and Garcia-Cardena, 2016). It would be interesting to stratify the chromatin accessibility and gene expression at this locus, and indeed many others, and observe changes that may occur over time. It has been demonstrated that PLPP3 expression is dictated by genotype (Wu *et al*., 2015; Krause *et al*., 2018). One hypothesis is that there is a negative correlation of chromatin accessibility with respect to age suggesting that this enhancer activity and subsequent PLPP3 expression may coincide with onset of CAD symptoms. This would be an interesting find and further suggestive of the protective role of endothelial PLPP3 against atherogenesis (Panchatcharam *et al*., 2014; Wu *et al*., 2015) Taken a step further this would be greater evidence of PLPP3 being a potential therapy target to alleviate human CAD. Nevertheless, the current data suggests that should the enhancer activity around rs17114036 remain constitutive throughout life, activation of this enhancer by unidirectional flow may still prove to be beneficial to human health regardless of the aging process.

Our data suggest that the genotype at rs17114036-containing region alone explained ~40% of PLPP3 expression in human aortic endothelium (Figure 2.13). PLPP3 expression, like for many genes, is tightly controlled within the cell as perturbations result in cellular phenotype changes. The mechanism presented in this work is thus among the most significant pathway regulating PLPP3, especially in response to blood flow. It is still within the realm of possibility

that in addition to the rs17114036-containaing region, there are other important regulatory elements such as the endogenous promoter or other enhancers that control PLPP3 expression. CRISPR/Cas9-mediated 70bp deletion resulted in the reduction of endothelial PLPP3 expression and decreased its sensitivity to hemodynamics but did not completely ablate gene expression. Indeed, the promoter of PLPP3 which contains KLF2 binding sites appears to also be differentially regulated under unidirectional flow (Wu *et al*., 2015; Krause *et al*., 2018). Our luciferase assays suggested that human PLPP3 promoter regulates PLPP3 expression collaboratively with the rs17114036-enclosing enhancer (Figure 2.6). In addition, our ATAC-seq data has identified near PLPP3 additional open chromatin regions that are induced by unidirectional flow These may also harbor regulatory enhancer regions contributing to PLPP3 mechanosensitivity. Future experiments can apply gene reporter experiments and genome-editing at these loci for identification functional validation.

4.4 Allelic Distribution and Selection Signals at rs17114036

It is interesting to note that the minor allele (CAD protective allele) at rs17114036 is present in other mammals and the major allele (CAD risk allele) is only found in humans (UCSC genome browser). One can speculate the fitness conferred by this sequence variant was in fact reversed in other animals and in early human history. Since endothelial PLPP3 plays important roles in vascular development and host immunity, it is possible that increased vasculogenesis and inflammation, as the result of decreased endothelial PLPP3 caused by the CAD risk allele at rs17114036, were beneficial during developmental stages or protective against infection (Bai *et al*., 2013; Yung *et al*., 2014). It is unlikely that CAD would be a major driver of adaptation. CAD typically affects the elderly who have already reproduced and passed their genes to their

offspring. In vivo models of PLPP3 polymorphisms in studies of early development and infectious diseases may aid in investigating these hypotheses in the future.

The allele frequency of rs17114036 varies rather subtly in different human populations around the world (http://popgen.uchicago.edu/ggv/). This suggests that any adaptation or selective pressure is not limited to any particular environmental need. Looking at early human genomes, it is possible to identify signals of selection by methods such as a Selective Sweep Scan, which detects non-random changes in allele-frequency with respect to time (Nielsen *et al*., 2005). In both Neanderthals and Denisovan genomes, there are positive indications of selection at rs17114036. As stated previously, one can only speculate as to what selective pressure was driving this adaptation; whether it is in normal development or host immunity.

4.5 Additional causal SNPs in the CAD locus 1p32.2

In this work, Bayesian Refinement and Conditional and Joint Multiple-SNP Analyses were conducted to predict the causal SNPs in the CAD locus 1p32.2 (Maller *et al*., 2015; Yang *et al*., 2016; Figure 2.1). These results narrowed down a cohort of 45 SNPs to 2 (rs17114036 and rs2184104). While our studies demonstrated rs17114036 is the causal SNP in this locus by regulating key endothelial functions, we cannot rule out the possibility that rs2184104 harbors a regulatory element in another CAD relevant cell type. This would be suggested should a cis-regulatory element enclosing rs2184104 be indicated by datasets such as ENCODE, ATAC-seq and ChIP-seq in other cell types/tissues. That work could largely follow the model of this study where the enhancer is identified and then mechanistically studied to assess functional significance using reporter assays, genome-editing, and correlations from human samples.

4.6 Increasing gene-editing efficiency in primary human endothelial cells

While this study has shown that CRISPR-based genome editing is possible in endothelial cells, the procedure remains extremely labor intensive. This protocol can be improved in the future to achieve higher editing efficiency and clonal generation. One simple solution would be to further optimize culture and transfection conditions, which are the critical first steps to delivering the CRISPR/Cas9 RNP to the cells (Jinek *et al*., 2012). Another improvement could be in the clonal selection step. The protocol presented in this work used a flow sorting to plate single cells with no selection marker. One can optimize the procedure to enrich for CRISPR/Cas9 RNP transfected cells. For example, the sgRNA can be labeled with a fluorescent marker that can be detected by FACS. Additionally, use of a live cell maker such as 7-AAV, which intercalates into dead cells, will further allow for selection of only live cells (Bak and Porteus 2017).

4.7 Additional GWAS SNPs located in mechanosensitive cis-regulatory elements

This study has identified 303 disease-associated SNPs in mechanosensitive cis-regulatory elements in endothelial cells. The identified SNPs are associated with many complex disease traits such as CAD, hematopoiesis, diabetes, and inflammatory diseases (Table 3.1). In the future, these SNPs can be functionally validated as has been done with PLPP3 in this work. Probing the role of SNPs in regulating endothelial functions could reveal possible new causal variants and regulatory pathways contributing to inter-individual variation in disease susceptibility. Identification of causal variants can enable more accurate assessment of disease risk and allow for earlier interventions for patients at higher risk. Identification of unifying

mechanisms and pathways could lead to targeted therapies that can address an individual's risk to CAD.

4.8 Flow-sensitive transcription factors in the regulation of mechanosensitive cis-regulatory elements

These investigations have identified over-represented transcription factor binding motifs in mechanosensitive cis-regulatory elements in endothelial cells. Nevertheless, whether the motif analyses faithfully describe the binding of these transcription factors and their direct downstream targets remains to be experimentally determined. For example, the motif analyses predict that ERG and JUN would each bind to roughly 50% of flow-sensitive enhancers (Figures 3.19 and 3.21). ERG and JUN are known to be important for establishing endothelial-specific genes but whether these transcription factors interact with flow-sensitive cis-regulatory elements remains to be elucidated (Linneman *et al*., 2011; Hogan *et al*., 2017). In addition to the motif predictions, my analyses using published ERG and JUN ChIP-seq results identified 55.86% and 92.54% of the mechanosensitive cis-regulatory elements are bound by ERG and/or JUN, further suggesting that ERG and JUN could contribute to the chromatin accessibility at endothelial-specific genes as previously shown, and at the flow-sensitive regulatory regions identified in this study (Hogan *et al*., 2017; Figures 3.19 and 3.21). This assertion would be greatly bolstered by studies directly showing differentially accessible chromatin regions using ATAC-seq with either knockdown or genomic deletion of these factors to directly observe whether or not these factors contribute to changes in chromatin accessibility or are capable of binding in these regions. Our results also implicated transcription factors KLF2 and KLF4 in controlling and/or regulating the activity of these mechanosensitive cis-regulatory elements, consistent with the role of KLF2 and KLF4 in

mediating key endothelial responses to hemodynamics (Dekker *et al*., 2005; Dekker *et al*., 2006; Villareal *et al*., 2010; Sungwung *et al*., 2017; He *et al*., 2019). ChIP-seq results showing transcription factor binding for KLF2 and KLF4 would be especially valuable as there is no such data yet in endothelial cells under well-define flow conditions.

RNA-and ATAC-seq profile mRNA expression and chromatin accessibility, respectively. The combination of these two approaches is ideal to fully investigate direct transcriptional changes mediated by cis-regulatory elements. Experiments knocking down via RNAi or knocking out with CRISPR-based deletion of transcription factors of interest would be valuable in the future to identify the subset of genes and regulatory elements controlled by flow-sensitive transcription factors in vascular endothelium. The fields of vascular biology and functional genomics would also benefit from differential transcriptome profiling in endothelial cells subjected to unidirectional flow combined with modulation or mutation of JUN, ERG, KLF2, and KLF4, which could uncover networks of flow-sensitive genes and the factors that regulate them.

4.9 Necessity of patho-physiological flow in endothelial biology studies

The data in this work demonstrated that hemodynamic forces are major regulators of endothelial transcriptome and epigenome (Wu *et al*., 2017; Krause *et al*., 2018), emphasizing the importance of incorporating mechanical forces into investigations of endothelial functions (Wu *et al*., 2017). For instance, endothelium under static conditions strongly resembles the cells under disturbed flow but not the quiescent endothelial cells typically exposed to unidirectional flow in vivo. Studies using endothelial cells under static conditions then can be problematic and difficult to interpret as disturbed flow is a disease state, albeit a physiological one (Qiao *et al*., 2016).

Culturing endothelial cells under defined flow conditions would be important for studying endothelial cells and the roles of transcription factors/cis-regulatory elements under physiological conditions.

4.10 Future studies to identify additional cis-regulatory elements at the interface of genetic predisposition and endothelial mechanosensing mechanisms

By using genomic and transcriptome-wide approaches with ATAC-, ChIP-, and RNA-seq on endothelial cells subjected to well-defined flow conditions, I have systematically identified a cohort of genes and cis-regulatory elements regulated by hemodynamics. With quite a few candidate genes and cis-regulatory elements to choose from for future study, there are additional criteria which can be applied to identify the most relevant loci with potential biological and medical significance. Investigating differential peaks and enhancers near GWAS loci, as done with the PLPP3 experiments, is valuable for identifying genes and regulatory elements with implicated relevance to human diseases and traits. Two examples of such a region are found near FLT1 and GOSR2, as described in Chapter 3. Future studies integrating chromatin associability, ChIP-seq, allelic imbalance, promoter capture Hi-C, genome-editing, endothelial functional assays, as well as animal models can be employed to determine the functional role of these cis-regulatory elements at the interface of genetic predisposition and endothelial mechanosensing mechanisms related to complex human diseases.

Works Cited

Abecasis GR, et al.; 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65. doi: 10.1038/nature11632. PMID: 23128226

Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016 Oct 14;17(11):661-678. doi: 10.1038/nrg.2016.112. PMID: 27739532

Brown JD, Lin CY, Duan Q, Griffin G, Federation A, Paranal RM, Bair S, Newton G, Lichtman A, Kung A, Yang T, Wang H, Luscinskas FW, Croce K, Bradner JE, Plutzky J. NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis. Mol Cell. 2014 Oct 23;56(2):219-231. doi: 10.1016/j.molcel.2014.08.024. Epub 2014 Sep 25. PMID: 25263595

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213-8. doi: 10.1038/nmeth.2688. Epub 2013 Oct 6. PMID: 24097267

Chiang TW, le Sage C, Larrieu D, Demir M, Jackson SP. CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. Sci Rep. 2016 Apr 15;6:24356. doi: 10.1038/srep24356. PMID: 27079678

Civelek M, Manduchi E, Riley RJ, Stoeckert CJ Jr, Davies PF. Coronary artery endothelial transcriptome in vivo: identification of endoplasmic reticulum stress and enhanced reactive oxygen species by gene connectivity network analysis. Circ Cardiovasc Genet. 2011 Jun;4(3):243-52. doi: 10.1161/CIRCGENETICS.110.958926. Epub 2011 Apr 14. PMID: 21493819

Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014 Jan;15(1):34-48. doi: 10.1038/nrg3575. Epub 2013 Dec 3. PMID: 24296534

Consortium EP; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616

Dai G, Kaazempur-Mofrad MR, Natarajan S, Zhang Y, Vaughn S, Blackman BR, Kamm RD, García-Cardeña G, Gimbrone MA Jr. Distinct endothelial phenotypes evoked by arterial waveforms derived from atherosclerosis-susceptible and -resistant regions of human vasculature. Proceedings of the National Academy of Sciences U S A. 2004 Oct 12;101(41):14871-6. Epub 2004 Oct 4. PMID: 15466704

Davies PF, Civelek M, Fang Y, Fleming I (2013) The atherosusceptible endothelium: Endothelial phenotypes in complex haemodynamic shear stress regions in vivo. Cardiovascular Research. 2013 Jul 15;99(2):315-27. doi: 10.1093/cvr/cvt101. Epub 2013 Apr 25. PMID: 23619421

Dekker RJ, van Thienen JV, Rohlena J, de Jager SC, Elderkamp YW, Seppen J, de Vries CJ, Biessen EA, van Berkel TJ, Pannekoek H, Horrevoets AJ. Endothelial KLF2 links local arterial shear stress levels to the expression of vascular tone-regulating genes. American Journal of Pathology. 2005 Aug;167(2):609-18. PMID: 16049344

Dekker RJ, Boon RA, Rondaij MG, Kragt A, Volger OL, Elderkamp YW, Meijers JC, Voorberg J, Pannekoek H, Horrevoets AJ. KLF2 provokes a gene expression pattern that establishes functional quiescent differentiation of the endothelium. Blood. 2006 Jun 1;107(11):4354-63. Epub 2006 Feb 2. PMID: 16455954

Deloukas P, et al.; CARDIoGRAMplusC4D Consortium; DIAGRAM Consortium; CARDIOGENICS Consortium; MuTHER Consortium; Wellcome Trust Case Control Consortium (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genetics. 2013 Jan;45(1):25-33. doi: 10.1038/ng.2480. Epub 2012 Dec 2. PMID: 23202125

Dichgans M, et al.; METASTROKE Consortium; CARDIoGRAM Consortium; C4D Consortium; International Stroke Genetics Consortium (2014) Shared genetic susceptibility to ischemic stroke and coronary artery disease: A genome-wide analysis of common variants. Stroke. 2014 Jan;45(1):24-36. doi: 10.1161/STROKEAHA.113.002707. Epub 2013 Nov 21. PMID: 24262325

Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. Cell. 2006 Aug 25;126(4):677-89. PMID: 16923388

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011 May 5;473(7345):43-9. doi: 10.1038/nature09906. Epub 2011 Mar 23. PMID: 21441907

Fang Y, Davies PF. Site-specific microRNA-92a regulation of Kruppel-like factors 4 and 2 in atherosusceptible endothelium. Arteriosclerosis Thrombosis Vascular Biology. 2012 Apr;32(4):979-87. doi: 10.1161/ATVBAHA.111.244053. Epub 2012 Jan 19. PMID: 22267480

Feng S, Bowden N, Fragiadaki M, Souilhol C, Hsiao S, Mahmoud M, Allen S, Pirri D, Ayllon BT, Akhtar S, Thompson AAR, Jo H, Weber C, Ridger V, Schober A, Evans PC. Mechanical Activation of Hypoxia-Inducible Factor 1α Drives Endothelial Dysfunction at Atheroprone Sites. Arteriosclerosis Thrombosis Vascular Biology. 2017 Nov;37(11):2087-2101. doi: 10.1161/ATVBAHA.117.309249. Epub 2017 Sep 7. PMID: 28882872

Gimbrone MA, Jr, García-Cardeña G (2016) Endothelial cell dysfunction and the pathobiology of atherosclerosis. Circ Res. 2016 Feb 19;118(4):620-36. doi: 10.1161/CIRCRESAHA.115.306301. PMID: 26892962

Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, Emdin CA, Hilvering CRE, Bianchi V, Mueller C, Khera AV, Ryan RJH, Engreitz JM, Issner R, Shoresh N, Epstein CB, de Laat W, Brown JD, Schnabel RB, Bernstein BE, Kathiresan S. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. Cell. 2017 Jul 27;170(3):522-533.e15. doi: 10.1016/j.cell.2017.06.049. PMID: 28753427

Hahn C, Schwartz MA (2009) Mechanotransduction in vascular physiology and atherogenesis. Nature Reviews Molecular Cell Biology. 2009 Jan;10(1):53-62. doi: 10.1038/nrm2596. PMID: 19197332

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature. 2011 Feb 10;470(7333):264-8. doi: 10.1038/nature09753. PMID: 21307941

He M, Huang TS, Li S, Hong HC, Chen Z, Martin M, Zhou X, Huang HY, Su SH, Zhang J, Wang WT, Kang J, Huang HD, Zhang J, Chien S, Shyy JY. Atheroprotective Flow Upregulates ITPR3 (Inositol 1,4,5-Trisphosphate Receptor 3) in Vascular Endothelium via KLF4 (Krüppel-Like Factor 4)-Mediated Histone Modifications. Arteriosclerosis Thrombosis Vascular Biology. 2019 May;39(5):902-914. doi: 10.1161/ATVBAHA.118.312301. PMID: 30917677

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PMID: 20513432

Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3):144-54. doi: 10.1038/nrm3949. Epub 2015 Feb 4. PMID: 25650801

Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, Glass CK, Romanoski CE. Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. Elife. 2017 Jun 6;6. pii: e22536. doi: 10.7554/eLife.22536. PMID: 28585919

Howson JMM, et al.; CARDIoGRAMplusC4D; Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nature Genetics. 2017 Jul;49(7):1113-1119. doi: 10.1038/ng.3874. Epub 2017 May 22. PMID: 28530674

Huang RT, Wu D, Meliton A, Oh MJ, Krause M, Lloyd JA, Nigdelioglu R, Hamanaka RB, Jain MK, Birukova A, Kress JP, Birukov KG, Mutlu GM, Fang Y. Experimental Lung Injury Reduces Krüppel-like Factor 2 to Increase Endothelial Permeability via Regulation of RAPGEF3-Rac1 Signaling. Am J Respir Crit Care Med. 2017 Mar 1;195(5):639-651. doi: 10.1164/rccm.201604-0668OC. PMID: 27855271

Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S, Brozynska M, Wang T, Ye S, Webb TR, Rutter MK, Tzoulaki I, Patel RS, Loos RJF, Keavney B, Hemingway H, Thompson J, Watkins H, Deloukas P, Di Angelantonio E, Butterworth AS, Danesh J, Samani NJ; UK Biobank CardioMetabolic Consortium CHD Working Group. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 2018 Oct 16;72(16):1883-1893. doi: 10.1016/j.jacc.2018.07.079. PMID: 30309464

Jaalouk DE1, Lammerding J. Mechanotransduction gone awry. Nature Reviews Molecular Cell Biology. 2009 Jan;10(1):63-73. doi: 10.1038/nrm2597. PMID: 19197333

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816-21. doi: 10.1126/science.1225829. Epub 2012 Jun 28. PMID: 22745249

Kalna V, Yang Y, Peghaire CR, Frudd K, Hannah R, Shah AV, Osuna Almagro L, Boyle JJ, Göttgens B, Ferrer J, Randi AM, Birdsey GM. The Transcription Factor ERG Regulates Super-Enhancers Associated with an Endothelial-Specific Gene Expression Program. Circulation Research. 2019 Apr 26;124(9):1337-1349. doi: 10.1161/CIRCRESAHA.118.313788. PMID: 30892142

Khachigian LM, Resnick N, Gimbrone MA Jr, Collins T. Nuclear factor-kappa B interacts functionally with the platelet-derived growth factor B-chain shear-stress response element in vascular endothelial cells exposed to fluid shear stress. Journal of Clinical Investigation. 1995 Aug;96(2):1169-75. PMID: 7635955

Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nature Reviews Genetics. 2017 Jun;18(6):331-344. doi: 10.1038/nrg.2016.160. Epub 2017 Mar 13. PMID: 28286336

Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics. 2019 Apr;20(4):207-220. doi: 10.1038/s41576-018-0089-8. PMID: 30675018

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19;533(7603):420-4. doi: 10.1038/nature17946. Epub 2016 Apr 20. PMID: 27096365

Krause MD, Huang RT, Wu D, Shentu TP, Harrison DL, Whalen MB, Stolze LK, Di Rienzo A3, Moskowitz IP, Civelek M, Romanoski CE, Fang Y. Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. PNAS. 2018 Nov 27;115(48):E11349-E11358. doi: 10.1073/pnas.1810568115. Epub 2018 Nov 14. PMID: 30429326

Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet. 2016 Feb;48(2):206-13. doi: 10.1038/ng.3467. Epub 2015 Dec 14. PMID: 26656845

Lan Q, Mercurius KO, Davies PF. Stimulation of transcription factors NF kappa B and AP1 in endothelial cells subjected to shear stress. Biochem Biophys Res Commun. 1994 Jun 15;201(2):950-6. PMID: 8003036

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923. PMID: 22388286

Le HQ, Ghatak S, Yeung CY, Tellkamp F, Günschmann C, Dieterich C, Yeroslaviz A, Habermann B, Pombo A, Niessen CM, Wickström SA. Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. Nature Cell Biology. 2016 Aug;18(8):864-75. doi: 10.1038/ncb3387. Epub 2016 Jul 11. PMID: 27398909

Li Z, Martin M, Zhang J, Huang HY, Bai L, Zhang J, Kang J, He M, Li J, Maurya MR, Gupta S, Zhou G, Sangwung P, Xu YJ, Lei T, Huang HD, Jain M, Jain MK, Subramaniam S, Shyy JY. Krüppel-Like Factor 4 Regulation of Cholesterol-25-Hydroxylase and Liver X Receptor Mitigates Atherosclerosis Susceptibility. Circulation. 2017 Oct 3;136(14):1315-1330. doi: 10.1161/CIRCULATIONAHA.117.027462. Epub 2017 Aug 9. PMID: 28794002

Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014 Dec 15;3:e04766. doi: 10.7554/eLife.04766. PMID: 25497837

Linnemann AK, O'Geen H, Keles S, Farnham PJ, Bresnick EH. Genetic framework for GATA factor function in vascular biology. PNAS 2011 Aug 16;108(33):13641-6. doi: 10.1073/pnas.1108440108. Epub 2011 Aug 1. PMID: 21808000

Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 2019 May 2;177(4):1022-1034.e6. doi: 10.1016/j.cell.2019.04.014. PMID: 31051098

Maher B. Personal genomes: The case of the missing heritability. Nature. 2008 Nov 6;456(7218):18-21. doi: 10.1038/456018a. PMID: 18987709

Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, Howson JM, Auton A, Myers S, Morris A, Pirinen M, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Craddock N, Hurles M, Ouwehand W, Parkes M, Rahman N, Duncanson A, Todd JA, Kwiatkowski DP, Samani NJ, Gough SC, McCarthy MI, Deloukas P, Donnelly P. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nature Genet 2012 Dec;44(12):1294-301. doi: 10.1038/ng.2435. Epub 2012 Oct 28. PMID: 23104008

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E,

Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53. doi: 10.1038/nature08494. PMID: 19812666

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008 May;9(5):356-69. doi: 10.1038/nrg2344. PMID: 18398418

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nature Genetics. 2015 Jun;47(6):598-606. doi: 10.1038/ng.3286. Epub 2015 May 4. PMID: 25938943

Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, Bozek G, Moskowitz IP, McNally EM, Nóbrega MA. A promoter interaction map for cardiovascular disease genetics. Elife. 2018 Jul 10;7. pii: e35788. doi: 10.7554/eLife.35788. PMID: 29988018

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Koteliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010 Aug 5;466(7307):714-9. doi: 10.1038/nature09266. PMID: 20686566

Nam D, Ni CW, Rezvan A, Suo J, Budzyn K, Llanos A, Harrison D, Giddens D, Jo H. Partial carotid ligation is a model of acutely induced disturbed flow, leading to rapid endothelial dysfunction and atherosclerosis. Am J Physiol Heart Circ Physiol. 2009 Oct;297(4):H1535-43. doi: 10.1152/ajpheart.00510.2009. Epub 2009 Aug 14. PMID: 19684185

Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ,

Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature Genetics. 2015 Oct;47(10):1121-1130. doi: 10.1038/ng.3396. Epub 2015 Sep 7. PMID: 26343387

Nurnberg ST, Zhang H, Hand NJ, Bauer RC, Saleheen D, Reilly MP, Rader DJ. From Loci to Biology: Functional Genomics of Genome-Wide Association for Coronary Disease. Circ Res. 2016 Feb 19;118(4):586-606. doi: 10.1161/CIRCRESAHA.115.306464. PMID: 26892960

O'Donnell CJ, Nabel EG. Genomics of cardiovascular disease. N Engl J Med. 2011 Dec 1;365(22):2098-109. doi: 10.1056/NEJMra1105239. PMID: 22129254

Panchatcharam M1, Salous AK, Brandon J, Miriyala S, Wheeler J, Patil P, Sunkara M, Morris AJ, Escalante-Alcalde D, Smyth SS. Mice with targeted inactivation of ppap2b in endothelial and hematopoietic cells display enhanced vascular inflammation and permeability. Arterioscler Thromb Vasc Biol. 2014 Apr;34(4):837-45. doi: 10.1161/ATVBAHA.113.302335. Epub 2014 Feb 6. PMID: 24504738\

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016 Sep;11(9):1650-67. doi: 10.1038/nprot.2016.095. Epub 2016 Aug 11. PMID: 27560171

Ong CT and Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011 Apr;12(4):283-93. doi: 10.1038/nrg2957. Epub 2011 Mar 1. PMID: 21358745

Parmar KM, Larman HB, Dai G, Zhang Y, Wang ET, Moorthy SN, Kratz JR, Lin Z, Jain MK, Gimbrone MA Jr, García-Cardeña G. Integration of flow-dependent endothelial phenotypes by Kruppel-like factor 2. J Clin Invest. 2006 Jan;116(1):49-58. Epub 2005 Dec 8. PMID: 16341264

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013 Feb 28;152(5):1173-83. doi: 10.1016/j.cell.2013.02.022. PMID: 23452860

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature Protocols. 2013 Nov;8(11):2281-2308. doi: 10.1038/nprot.2013.143. Epub 2013 Oct 24. PMID: 24157548

Romanoski CE, Lee S, Kim MJ, Ingram-Drake L, Plaisier CL, Yordanova R, Tilford C, Guan B, He A, Gargalovic PS, Kirchgessner TG, Berliner JA, Lusis AJ. Systems genetics analysis of gene-by-environment interactions in human cells. Am J Hum Genet. 2010 Mar 12;86(3):399-410. doi: 10.1016/j.ajhg.2010.02.002. Epub 2010 Feb 18. PMID: 20170901

Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, Wang H, Liu T, Wasserman SM, Sever PS, Pedersen TR; FOURIER Steering Committee and Investigators. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. New England Journal of Medicine. 2017 May 4;376(18):1713-1722. doi: 10.1056/NEJMoa1615664. Epub 2017 Mar 17. PMID: 28304224

Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nature Reviews Genetics. 2019 May 13. doi: 10.1038/s41576-019-0128-0. PMID: 31086298

Schumann K, Lin S, Boyer E, Simeonov DR, Subramaniam M, Gate RE, Haliburton GE, Ye CJ, Bluestone JA, Doudna JA, Marson A. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. PNAS. 2015 Aug 18;112(33):10437-42. doi: 10.1073/pnas.1512503112. Epub 2015 Jul 27. PMID: 26216948

Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I; Cardiogenics, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Mühleisen TW, Muhlestein JB, Münzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nöthen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schäfer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, März W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J; CARDIoGRAM Consortium, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature Genetics. 2011 Mar 6;43(4):333-8. doi: 10.1038/ng.784. PMID: 21378990

SenBanerjee S, Lin Z, Atkins GB, Greif DM, Rao RM, Kumar A, Feinberg MW, Chen Z, Simon DI, Luscinskas FW, Michel TM, Gimbrone MA Jr, García-Cardeña G, Jain MK. KLF2 Is a novel transcriptional regulator of endothelial proinflammatory activation. J Exp Med. 2004 May 17;199(10):1305-15. Epub 2004 May 10. PMID: 15136591

Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB, Lisbon E, Gutierrez M, Webb C, Wu R, Du Y, Kranz T, Gasparino E, Swergold GD. Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. N Engl J Med. 2012 Mar 22;366(12):1108-18. doi: 10.1056/NEJMoa1105803. PMID: 22435370

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. 2019 May 8. doi: 10.1038/s41576-019-0127-1. [Epub ahead of print] PMID: 31068683

van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nature Methods. 2015 Nov;12(11):1061-3. doi: 10.1038/nmeth.3582. Epub 2015 Sep 14. PMID: 26366987

van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res. 2018 Feb 2;122(3):433-443. doi: 10.1161/CIRCRESAHA.117.312086. Epub 2017 Dec 6. PMID: 29212778

Watkins H, Farrall M. Genetic susceptibility to coronary artery disease: from promise to progress. Nat Rev Genet. 2006 Mar;7(3):163-73. PMID: 16462853

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013 Apr 11;153(2):307-19. doi: 10.1016/j.cell.2013.03.035. PMID: 23582322

Willer CJ, et al. Discovery and refinement of loci associated with lipid levels. Nature Genetics. 2013 Nov;45(11):1274-1283. doi: 10.1038/ng.2797. Epub 2013 Oct 6. PMID: 24097068

Wu C, Huang RT, Kuo CH, Kumar S, Kim CW, Lin YC, Chen YJ, Birukova A, Birukov KG, Dulin NO, Civelek M, Lusis AJ, Loyer X, Tedgui A, Dai G, Jo H, Fang Y. Mechanosensitive PPAP2B Regulates Endothelial Responses to Atherorelevant Hemodynamic Forces. Circ Res. 2015 Jul 31;117(4):e41-e53. doi: 10.1161/CIRCRESAHA.117.306457. Epub 2015 Jun 1. PMID: 26034042

Wu D, Huang RT, Hamanaka RB, Krause M, Oh MJ, Kuo CH, Nigdelioglu R, Meliton AY, Witt L, Dai G, Civelek M, Prabhakar NR, Fang Y, Mutlu GM. HIF-1α is required for disturbed flow-induced metabolic reprogramming in human and porcine vascular endothelium Elife. 2017 May 30;6. pii: e25217. doi: 10.7554/eLife.25217. PMID: 28556776

Yang J, Ferreira T, Morris AP, Medland SE; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden PA, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics. 2012 Mar 18;44(4):369-75, S1-3. doi: 10.1038/ng.2213. PMID: 22426310

Zhou G, Hamik A, Nayak L, Tian H, Shi H, Lu Y, Sharma N, Liao X, Hale A, Boerboom L, Feaver RE, Gao H, Desai A, Schmaier A, Gerson SL, Wang Y, Atkins GB, Blackman BR, Simon DI, Jain MK. Endothelial Kruppel-like factor 4 protects against atherothrombosis in mice. J Clin Invest. 2012 Dec;122(12):4727-31. doi: 10.1172/JCI66056. Epub 2012 Nov 19. PMID: 23160196

Zhou J, Li YS, Chien S. Shear stress-initiated signaling and its regulation of endothelial function. Arterioscler Thromb Vasc Biol. 2014 Oct;34(10):2191-8. doi: 10.1161/ATVBAHA.114.303422. Epub 2014 May 29. PMID: 24876354

Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen ZY, Liu DR. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol. 2015 Jan;33(1):73-80. doi: 10.1038/nbt.3081. Epub 2014 Oct 30. PMID: 25357182