UNIVERSITY OF CALIFORNIA SAN DIEGO

Identifying Markers and Mechanisms of Atherosclerosis in Endothelial Cells Under Shear Stress Using High-Throughput Sequencing Data

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Nassim Elias Ajami

Committee in charge:

Professor Shankar Subramaniam, Chair
Professor Shu Chien, Co-Chair
Professor Terry Gasterlaand
Professor Nathan Lewis
Professor John Shyy

2018

The Dissertation of Nassim Elias Ajami is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2018

DEDICATION

To my mother, Sabah Khalil Ajami, to my father, Elias Nassim Ajami – and to my future descendants, who may have some variation of “Elias”, “Nassim”, or “Ajami” in their name (and if they don’t, I’m sure one of them found their way to this thesis and is laughing to themselves right now).

EPIGRAPH

If you want to go fast, go alone; but if you want to go far, go together.

Unattributed African Proverb

TABLE OF CONTENTS

SIGNATURE PAGE……iii
DEDICATION....iv
EPIGRAPH....v
TABLE OF CONTENTS....vi
LIST OF FIGURES....viii
LIST OF TABLES....xi
ACKNOWLEDGEMENTS....xii
VITA....xiii
ABSTRACT OF THE DISSERTATION....xiv
Chapter 1: General Introduction to Atherosclerosis, Endothelial Cells, and the Application of "omics" Data....1
Atherosclerosis, Endothelial Cells, and Hemodynamic Forces....2
"Omics" data and atherosclerosis....5
References....9
Chapter 2: Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress....11
Introduction....12
Methods....12
Results....15
Discussion....21
Acknowledgments....25
References....44
Chapter 3: Novel Mechanisms of Endothelial Response to Shear Derived From "Omics" Data Analysis....49
Introduction....50
Methods – "Omics" Data Generation and Processing"....50
Methods – LINC00520 Validation Experiments....53
Results – *in silico* Analysis....59
Results – Validation of LINC00520 (LEENE) as a mechanosensitive lnc-eRNA relevant to endothelial homeostasis.....66
Discussion Of LEENE Study....73
Acknowledgments....76
References....114

Chapter 4: Development and Application of a Web-Based “omics” Data Analysis and Integration Platform 121
Introduction 122
Implementation - *Methods* 123
Implementation - Structure and Interface 124
Example Analysis 129
Discussion 131
Acknowledgements 131
References 136
Chapter 5: Challenges, Conclusions, and Future Directions 137
Challenges – Biological 138
Challenges – Technical and Systemic 140
Conclusions 143
References 145

LIST OF FIGURES

Figure 2.1: Details of Cluster Analysis of TFs in TF-to-Gene Network. 26

Figure 2.2: Number of differentially expressed genes. 27

Figure 2.3: G1-to-S transition pathway. 28

Figure 2.4: Cluster analysis of functional pathway enrichment. 29

Figure 2.5: Reconstructed pathway of oxidative stress and superoxide metabolism. 30

Figure 2.6: AP-1 family of genes and the antioxidant product NQO1. 31

Figure 2.7: NF-κB and NF-κB target genes. 32

Figure 2.8: Endothelial-mesenchymal transition (EndoMT) and related pathways. 33

Figure 2.9: Largest contiguous portion of TF-to-target network for TFs that are downregulated in OS vs PS. 34

Figure 2.10: Largest contiguous portion of TF-to-target network for TFs that are upregulated in OS vs PS. 35

Figure 2.11: Longest paths in TF to gene networks. 36

Figure 2.12: Relative differences in gene expression in ECs exposed to OS or PS over time. 37

Figure 3.1: Visualization of the 27 possible point-wise states for a given gene's expression across three pairwise conditions and three possible states of regulations. 78

Figure 3.2: Quality control of CRISPR-Cas9 gene editing system targeting LEENE promoter/enhancer (ED) (a and b) and coding region (CD) (c). 79

Figure 3.3: Proposed mechanosensitive mechanism of regulation of endothelial morphology. 80

Figure 3.4: LINC00520 region with visualized histone ChIP-seq and 4C datasets. 81

Figure 3.5: Co-regulation of LEENE and eNOS. 82

Figure 3.6: Pathway enrichment analysis from 1177 genes that are differentially regulated between OS and PS and interact with LEENE genomic loci. 83

Figure 3.7: KLF2 and KLF4 transcriptionally regulate LEENE. 84

Figure 3.8: LEENE RNA is nucleus-localized and its DNA lies in enhancer region interacting with eNOS promoter. 85

Figure 3.9: Subcellular quantitation for lncRNAs DANCR, Casc7, and TUG1. 86

Figure 3.10: Transcription of LEENE neighboring genes. 87

Figure 3.11: Heatmap of flow-regulated 81 genes in LEENE interactome. 88

Figure 3.12: Hi-C analysis of LEENE-eNOS interaction map in non-EC cell types. 89

Figure 3.13: Histone ChIP-seq tracks of eNOS promoter/enhancer region from four different cell types. 90

Figure 3.14: Gene editing of LEENE locus influences eNOS transcription. 91

Figure 3.15: Quality control of CRISPR-Cas9 gene editing. 92

Figure 3.16: LEENE RNA regulates eNOS expression and EC function. 93

Figure 3.17: Effect of LEENE knock-down on eNOS mRNA level in HAoECs. 94

Figure 3.18: Effect of LEENE knock-down on pro-inflammatory gene expression. 95

Figure 3.19: Densitometry quantification of Western Blot experiments. 96

Figure 3.20: Overexpression of LEENE using adenovirus. 97

Figure 3.21: LEENE RNA promotes RNA Pol II binding and eNOS transcription. 98

Figure 3.22: IgG isotype control for the RIP assay. 99

Figure 3.23: Quality and negative controls of LEENE ChIRP assay. 100

Figure 3.24: Effect of LEENE depletion on KLF4/Med1 binding to eNOS promoter. 101

Figure 3.25: LEENE homolog in mouse. 102

Figure 3.26: LEENE enhancer is prerequisite to LEENE RNA-eNOS locus association and LEENE RNA is not sufficient to enhance eNOS level. 103

Figure 3.27: Schematic illustration of LEENE–eNOS regulatory mechanism 104

Figure 3.28: Effect of LEENE knock-down or overexpression on Tm mRNA level. 105

Figure 3.29: Hi-C analysis of LEENE-Tm interaction map in HUVECs. 106

Figure 3.30: Detailed gene information of 4C-seq circoplot. 107

Figure 4.1: Visual representation of the D-ChIP analysis method..... 132

Figure 4.2: Summary of analysis pipeline and location of the website where each analysis step occurs. 133

Figure 4.3: D-ChIP systems architecture..... 134

Figure 4.4: Screenshots of D-ChIP's analysis pages..... 135

LIST OF TABLES

Table 1.1: All “omics” data sets generated by collaborating labs. 8

Table 2.1: Top enriched pathways through GSEA: The top 10 up-regulated enriched Reactome pathways.. 38

Table 2.2: Top enriched pathways through GSEA: The top 10 down-regulated enriched Reactome pathways. 39

Table 2.3: Top enriched pathways through GSEA: The top 10 up-regulated enriched KEGG pathways. 40

Table 2.4: Top enriched pathways through GSEA: The top 10 down-regulated enriched KEGG pathways. 41

Table 2.5: Log2 fold change of select DE genes in OS vs PS across time. 42

Table 2.6: Select functional pathways pertaining to relevant endothelial functions under shear, and listings of the number of their participating genes in the TRANSFAC-derived network that is displayed in Figure 2.10. 43

Table 3.1: sgRNA sequences. 108

Table 3.2: LEENE LNA-GapmeRs sequences. 109

Table 3.3: LEENE ChIRP probe sequences. 110

Table 3.4: Human primer sequences. 111

Table 3.5: Mouse primer sequences. 112

Table 3.6: List of PS-Distinct and OS-Distinct genes. 113

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Shankar Subramaniam for his support as the chair of my committee. I would also like to acknowledge the Subramaniam Lab, particularly Dr. Shakti Gupta and Dr. Mano Maurya, for their support and advice over the course of my PhD.

I would also like to acknowledge my friends and colleagues within the Bioinformatics and Systems Biology program. Specifically Jeffrey Yuan, Rob Foreman, Ben Kellman, Anthony Aylward, and Shamim Mollah, whose perspectives, counsel, patience, and senses of humor were essential to surviving the trials and tribulations of graduate school – particularly our first year together. I would also like to acknowledge Bill Greenwald, with whom I shared many excellent conversations over coffee during our tenures in graduate school.

Chapters 1, 2, and 3 are a modified presentation of “Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress" as it appears in PNAS 2017 by Nassim E Ajami, Shakti Gupta, Mano R Maurya, Phu Nguyen, Julie Yi-Shuan Li, John Y-J Shyy, Zhen Chen, Shu Chien, and Shankar Subramaniam. The dissertation author was the primary author of this material.

Chapter 3 is a modified presentation of “Enhancer-Associated Long Non-Coding RNA LEENE Regulates Endothelial Nitric Oxide Synthase and Endothelial Function" as it appears in Nature Communications 2018 by Yifei Miao, Nassim E Ajami, Tse-Shun Huang, Feng-Mao Lin, Chih-Hong Lou, Yun-Ting Wang, Shuai Li, Jian Kang, Hannah Munkacsi, Mano R Maurya, Shakti Gupta, Shu Chien, Shankar Subramaniam, and Zhen Chen. The dissertation author was a co-primary author of this material.

Chapter 4 is a modified presentation of material currently being prepared for submission for publication as “D-ChIP: An accessible online suite for the comparison of broad-peaked ChIP-seq data sets" by Nassim E Ajami, Mano R Maurya, Shakti Gupta, and Shankar Subramaniam. The dissertation author is the primary author of this material.

VITA

2012 Bachelor of Arts, University of Illinois Urbana-Champaign

2018 Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Identifying Markers and Mechanisms of Atherosclerosis in Endothelial Cells Under Shear Stress Using High-Throughput Sequencing Data

by

Nassim Elias Ajami

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2018

Professor Shankar Subramaniam, Chair

Professor Shu Chien, Co-Chair

Atherosclerotic cardiovascular disease is the top cause of mortality worldwide and has imposed a multi-billion-dollar burden on health care and economic growth. Vascular endothelial cells form the lining of the interior walls of blood vessels and are in direct contact with blood circulating through the vasculature. These endothelial cells exhibit physiological and pathophysiological responses to hemodynamic forces, including the formation of

atherosclerotic lesions. The exact mechanisms of the dynamic response to hemodynamic forces by endothelial cells remain unclear. The advent of high-throughput sequencing technologies, along with the advent and growth of bioinformatics as a discipline, has brought with them an unprecedented ability to generate rich biological datasets and interrogate biological mechanisms on a systems level.

In this dissertation, I take a bioinformatics and systems biology approach to understanding endothelial response to hemodynamic forces. In chapter 1, I provide an overview on atherosclerosis, endothelial cells, hemodynamic forces, and the *in vitro* techniques used to study them. In chapter 2, I describe a systems biology analysis of time-series RNA-seq data that describes a dynamical map of endothelial response and a reconstruction of differences in transcriptional regulation across two different forms of hemodynamic forces. In chapter 3, I describe the study and characterization of several individual genes identified in my RNA-seq analysis, including LEENE, a long non-coding RNA whose relevance to endothelial homeostasis and shear stress was discovered through analysis of high-throughput sequencing data. In chapter 4, I describe a web suite constructed for the explicit purpose of data integration, along with integrated analysis of histone ChIP-seq data and RNA-seq data of endothelial cells exposed to shear stress. Chapter 5 concludes this dissertation with a discussion on the future directions and challenges of systems biology in application to atherosclerosis and endothelial cells.

Chapter 1: General Introduction to Atherosclerosis, Endothelial Cells, and the Application of “omics” Data

Atherosclerosis, Endothelial Cells, and Hemodynamic Forces

Atherosclerosis is the buildup of plaques in arteries over long periods of time. These plaques consist of lipids, cellular waste products, and other materials. The formation of plaques begins in childhood, beginning as so-called “fatty streaks”. Over time, these streaks can develop into intermediate lesions and slowly progress to fibrous, complex plaques. These plaques restrict arterial blood flow by narrowing the arterial lumen, impairing the ability of the body to supply oxygen to cells and placing greater burden on the heart’s ability to pump blood. In advanced stages of atherosclerosis, the plaques can rupture, which leads to the formation of blood clots (thrombosis), which can lead to heart attacks, strokes, and other vascular complications. Although the initial fatty streaks are observed in childhood, it is not until decades later that these plaques reach a stage of development where they are capable of causing vascular complications [Ross, 1993].

The manifestation of atherosclerosis has been observed in ancient populations, including individuals from ancient Egypt and ancient American civilizations [Thompson et al, 2013]. It would not be until the 17th century when William Harvey would describe the circulation of blood flow, and not until the 18th century when Rudolf Virchow would form the modern concept of the pathogenesis of plaques in arterial walls [Slijkhuis et al, 2009]. In the contemporary 21st century, atherosclerotic cardiovascular disease is described as the leading cause of mortality worldwide, with its prevalence linked to lifestyle factors and higher population BMI [Barquera et al, 2015].

The mechanisms of atherosclerotic plaque formation are complex and include diverse cellular elements such as platelets, macrophages, and monocytes [Baeyens et al, 2016; Gimbrone et al, 2013; Gimbrone et al, 2016]. Arguably the most crucial cellular elements are vascular endothelial cells (ECs), which line the interior arterial wall as a monolayer (known as the endothelium) and, by the nature of their physiological location, comprise the interface

where atherosclerotic plaques may form. The role of endothelial dysfunction in atherosclerosis has been hypothesized and studied since the 1970s [Ross, 1993]. Endothelial cells in healthy individuals perform a variety of functions. Their membrane proteins allow for the adhesion of leukocytes, which can then pass through the endothelium and into other tissue. Endothelial cells can control blood pressure through the processes of vasodilation and vasoconstriction via nitric oxide production. The endothelium is also involved in angiogenesis – the creation of new blood vessels, and in thrombosis – the process of blood clotting [Chiu et al, 2011].

In addition to being in direct contact with circulating blood, endothelial cells exhibit physiological responses to the forces exerted by blood flow. These forces are referred to as hemodynamic forces, and are determined by the composition of the blood and the geometry of the blood vessel in which the blood flows. Generally, in straight portions of the vasculature, the blood flow is pulsatile – with a positive mean flow rate and with a significant net direction. In bent and curved portions of the vasculature, the blood flow is disturbed – with an irregular pattern of flow that includes oscillatory, reciprocating flow with little mean flow rate. Pulsatile blood flow exerts pulsatile shear stress (PS) on the endothelium, whereas disturbed flow is associated with oscillatory shear stress (OS) on the endothelium. Endothelial cells exposed to PS exhibit an atheroprotective endothelial phenotype, with atherosclerotic plaques rarely developing in these regions over time. Endothelial cells exposed to OS exhibit an atherogenic endothelial phenotype, with plaques frequently forming in these regions. This association between atherosclerosis and shear stress has been explicitly established by previous studies, which measured the waveforms of flow analyses in human arterial geometries that are considered “susceptible” or “resistant” to plaque formation [Baeyens et al, 2016; Chiu et al, 2011; Chistiakov et al, 2016; Dai et al, 2004].

There are several well-defined molecular mechanisms for endothelial responses to different forms of shear stress. ECs feature several mechanosensors that include VE-

cadherin, integrins, and ion channels [Zhou et al, 2014; Gerhold et al, 2016]. The upregulation of Krüppel-like factor 2 (KLF2) is a well-established hallmark of endothelial response to PS [Dekker RJ et al, 2006; Parmar et al, 2006], along with endothelial nitric oxide synthase (eNOS) upregulation and activity [Hecker et al, 1994]. Compared to laminar or pulsatile flow, disturbed flow impairs endothelial homeostasis through many major processes. Disturbed flow increases the permeability of the endothelium, allowing for a greater number of macromolecules to enter the extracellular matrix and thereby allowing for more opportunities for plaque to build and rupture. Disturbed flow induces higher levels of reactive oxygen species via the upregulation of NADPH oxidase (NOX) and dysfunction of mitochondrial respiration chain. Disturbed flow causes a higher rate of endothelial proliferation and turnover in part due to mTOR activation. Disturbed flow also elicits a pro-inflammatory response through the activation of NF-κB and AP-1 transcription factors (TFs), thereby regulating adhesion molecules such as VCAM-1 and E-selectin and cytokines such as MCP-1. Endothelial cells under disturbed flow also have a nonuniform morphology, whereas under laminar flow they tend to be elongated in the direction of flow [Chiu et al, 2011; Chistiakov et al, 2016].

The mechanisms of EC response to distinct shear stresses have been inferred primarily from signaling and transcriptional measurements [Zhang et al, 2012; Simmons et al, 2016; Qiao et al, 2016; Ohura et al, 2003; Sangwung et al, 2017]. The means by which endothelial response to shear is studied in our collaborating labs *in vitro* is through an apparatus known the parallel-plate flow chamber. This device consists of a polycarbonate base plate, two gaskets, and a glass slide with EC monolayer. The base plate has inlet and outlet ports where fluid flows in and out of the device, passing over and exerting fluid shear stress on the endothelial cells. Cultured medium is used as fluid and flow is induced using an external pump [Chiu et al, 2011]. Depending on the settings of the pump, our collaborating

labs can simulate pulsatile shear stress of 12 ± 4 dyn/cm^2 and oscillatory shear stress of 0.5 ± 4 dyn/cm^2. These flow rates can be consistently maintained for periods of 24 hours. Although these flow patterns are imperfect models of the observed blood flow patterns *in vivo*, studies using this in vitro setup have successfully replicated key phenotypic features of endothelial shear response. Compared to endothelial cells exposed to OS and PS *in vivo*, endothelial cells exposed to OS and PS in a flow chamber have similar relative morphologies, similar relative rates of proliferation and migration, similar permeability, and similar levels of gene expression for key genes, including inflammation-related genes such as NF-κB [Chiu et al, 2011].

"Omics" data and atherosclerosis

High-throughput sequencing "omics" technologies have become a common component in biological experiments. Ever since the first sequence of a human genome was completed in 2003, the paradigm of biological research has dramatically shifted towards generating "omics" data for biological systems of interest in order to generate hypotheses for later validation. Different "omics" data sets provide insight on different components of cellular systems and therefore generate different hypotheses. RNA-seq data sets are generated in order to study the gene expression of samples across conditions [Conesa et al, 2016]. ChIP-seq data sets provide insight on transcription factor binding sites and histone modifications [Mundade et al, 2014]. Chromosome conformation capture-based studies, such as 4C and hi-C, offer a glimpse into chromatin-chromatin interactions and chromosomal architecture [Dekker J et al, 2013]. These data sets, when integrated, can offer valuable systems-level insights to physiology and disease [Sun et al, 2016].

Many studies have generated and used "omics" data to study the pathogenesis of atherosclerosis and endothelial dysfunction. A review by R. Simmons, S. Kumar, and H. Jo covers several publications that investigated methylomic and transcriptomic data sets [Simmons et al, 2016]. Although prior studies have examined transcriptional regulation from

RNA-seq data of ECs under shear stress, the majority of these studies only look at a single time point post-shear [Qiao et al, 2016; McCormick et al, 2017; Maleszewska et al, 2016]. The inherent limitation to taking a single temporal “snapshot” of transcriptional regulation is the inability to examine how the expression levels of different genes change relative to each other over time.

The ability to study co-expression over time allows for the creation of hypotheses that define causal mechanisms. For example, if gene A is differentially expressed immediately before (or at the same time as) gene B, then gene A may have a role in the regulation of gene B. This hypothesis is strengthened if gene A is a transcription factor, or if gene A has been shown to regulate gene B in a previous study, even in a different cellular context. A 2012 study generated time-series transcriptional profiles of porcine ECs exposed to shear using microarray techniques, and provided an initial insight into temporal mechanisms [Zhang et al, 2012]. However, time-series RNA-seq studies that study endothelial response to shear stress are rare, particularly because of the complex experimental design and the prohibitive cost of generating a time-series “omics” data set of sufficient temporal resolution. Furthermore, while the promise of integrating disparate “omics” data sets in order to generate comprehensive mechanistic hypotheses remains lucrative, many studies focus on only one type of “omics” data set, again likely due to prohibitive costs and complex experimental design.

Our collaborating labs have generated “omics” data sets on human umbilical vein endothelial cells (HUVECs) exposed to oscillatory shear (OS) and pulsatile shear (PS) using a parallel-plate flow chamber experimental design. These data sets, along with other, non-sheared EC data sets, are outlined in Table 1.1. The ten time point RNA-seq data offers an opportunity to generate hypotheses on a level of resolution previously unseen with sheared endothelial cells. The ChIP-seq data sets and 4C data sets alongside the RNA-seq data sets, all generated by the same collaborating labs with the same experimental design for shear

application, allows for the unique opportunity for data integration and complex hypothesis formulation.

Chapter 1 is a modified presentation of “Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress" as it appears in PNAS 2017 by Nassim E Ajami, Shakti Gupta, Mano R Maurya, Phu Nguyen, Julie Yi-Shuan Li, John Y-J Shyy, Zhen Chen, Shu Chien, and Shankar Subramaniam. The dissertation author was the primary author of this material.

Table 1.1: All "omics" data sets generated by collaborating labs.

Data Type	Experiment Details	Conditions
RNA-seq	Time-series shear 10 time points	OS, PS, "low flow" (~ST)
RNA-seq	Time-series shear 4 time points	OS, PS, ST
RNA-seq	Overexpression	KLF2, SREBP2, AMPK, AMPK + SIRT1
RNA-seq	Shear, 16 hours	OS, PS
ChIP-seq	H3K27ac, 24 hrs H3K27me3, 24 hrs H3K4me, 24 hrs	OS, PS
4C	KLF2, NOS3 loci – 24 hrs	OS, PS

References

Baeyens N, Bandyopadhyay C, Coon BG, Yun S, Schwartz MA. Endothelial fluid shear stress sensing in vascular health and disease. J Clin Invest. 2016 Mar 1;126(3):821-8.

Barquera S, Pedroza-Tobías A, Medina C, Hernández-Barrera L, Bibbins-Domingo K, Lozano R, Moran AE. Global Overview of the Epidemiology of Atherosclerotic Cardiovascular Disease. Arch Med Res. 2015 Jul;46(5):328-38.

Chistiakov DA, Orekhov AN, Bobryshev YV (2016) Effects of shear stress on endothelial cells: go with the flow. Acta physiologica.

Chiu JJ & Chien S (2011) Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. Physiological reviews 91(1):327-387.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016 Jan 26;17:13.

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. Proc Natl Acad Sci U S A. 2004 Oct 12;101(41):14871-6.

Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet. 2013 Jun;14(6):390-403.

Dekker RJ, Boon RA, Rondaij MG, Kragt A, Volger OL, Elderkamp YW, Meijers JC, Voorberg J, Pannekoek H, Horrevoets AJ. (2006) KLF2 provokes a gene expression pattern that establishes functional quiescent differentiation of the endothelium. Blood 107(11):4354-4363.

Gerhold KA, Schwartz MA (2016) Ion Channels in Endothelial Responses to Fluid Shear Stress. Physiology 31(5):359-369.

Gimbrone MA Jr, Garcia-Cardena G (2016) Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. Circulation research 118(4):620-636.

Gimbrone MA Jr, Garcia-Cardena G. (2013) Vascular endothelium, hemodynamics, and the pathobiology of atherosclerosis. Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology 22(1):9-15.

Hecker M, Mulsch A, Bassenge E, Forstermann U, & Busse R. (1994) Subcellular localization and characterization of nitric oxide synthase(s) in endothelial cells: physiological implications. The Biochemical journal 299 (Pt 1):247-252.

Maleszewska M, Vanchin B, Harmsen MC, & Krenning G. (2016) The decrease in histone methyltransferase EZH2 in response to fluid shear stress alters endothelial gene expression and promotes quiescence. Angiogenesis 19(1):9-24.

McCormick ME, Manduchi E, Witschey WRT, Gorman RC, Gorman JH 3rd, Jiang YZ, Stoeckert CJ Jr, Barker AJ, Yoon S, Markl M, Davies PF. (2017) Spatial phenotyping of the

endocardial endothelium as a function of intracardiac hemodynamic shear stress. Journal of biomechanics 50:11-19.

Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle. 2014;13(18):2847-52.

Ohura N, Yamamoto K, Ichioka S, Sokabe T, Nakatsuka H, Baba A, Shibata M, Nakatsuka T, Harii K, Wada Y, Kohro T, Kodama T, Ando J. (2003) Global analysis of shear stress-responsive genes in vascular endothelial cells. Journal of atherosclerosis and thrombosis 10(5):304-313.

Parmar KM, Larman HB, Dai G, Zhang Y, Wang ET, Moorthy SN, Kratz JR, Lin Z, Jain MK, Gimbrone MA Jr, García-Cardeña G. (2006) Integration of flow-dependent endothelial phenotypes by Kruppel-like factor 2. The Journal of clinical investigation 116(1):49-58.

Qiao C, Meng F, Jang I, Jo H, Chen YE, Zhang J. (2016) Deep transcriptomic profiling reveals the similarity between endothelial cells cultured under static and oscillatory shear stress conditions. Physiological genomics 48(9):660-666.

Ross R. (1993) The pathogenesis of atherosclerosis: a perspective for the 1990s. Nature. Apr 29;362(6423):801-9.

Sangwung P, Zhou G, Nayak L, Chan ER, Kumar S, Kang DW, Zhang R, Liao X, Lu Y, Sugi K, Fujioka H, Shi H, Lapping SD, Ghosh CC, Higgins SJ, Parikh SM, Jo H, Jain MK. (2017) KLF2 and KLF4 control endothelial identity and vascular integrity. JCI insight 2(4):e91700.

Simmons RD, Kumar S, & Jo H (2016) The role of endothelial mechanosensitive genes in atherosclerosis and omics approaches. Archives of biochemistry and biophysics 591:111-131.

Slijkhuis W, Mali W, Appelman Y. A historical perspective towards a non-invasive treatment for patients with atherosclerosis. Neth Heart J. 2009 Apr;17(4):140-4.

Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Adv Genet. 2016;93:147-90.

Thompson RC, Allam AH, Lombardi GP, Wann LS, Sutherland ML, Sutherland JD, Soliman MA, Frohlich B, Mininberg DT, Monge JM, Vallodolid CM, Cox SL, Abd el-Maksoud G, Badr I, Miyamoto MI, el-Halim Nur el-Din A, Narula J, Finch CE, Thomas GS. Atherosclerosis across 4000 years of human history: the Horus study of four ancient populations. Lancet. 2013 Apr 6;381(9873):1211-22.

Zhang J & Friedman MH (2012) Adaptive response of vascular endothelial cells to an acute increase in shear stress magnitude. American journal of physiology. Heart and circulatory physiology 302(4):H983-991.

Zhou J, Li YS, Chien S (2014) Shear stress-initiated signaling and its regulation of endothelial function. Arteriosclerosis, thrombosis, and vascular biology 34(10):2191-2198.

Chapter 2: Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress

Introduction

In this chapter, we explore the dynamics using RNA-seq measurements at several distinct time points followed by temporal longitudinal analysis of the mechanisms of response. We present a dynamical map of endothelial response, as well as reconstruct the differences in transcriptional regulation, across OS and PS conditions. The analysis of temporally longitudinal data shows the evolution of cellular response to stress, implicating genes representing several cellular and tissue functions including oxidative stress, inflammation, and cell cycle. This study provides the first detailed temporally longitudinal experimental study and systems model of endothelial responses to shear stress.

Methods

Cell Culture and Shear Stress Experiments. Human umbilical vein endothelial cells (HUVECs) were cultured in medium M199 (Gibco) supplemented with 15% FBS (Omega), 3 ng/mL β-EC growth factor (Sigma), 4 U/mL heparin (Sigma), and 100 U/mL penicillin-streptomycin [Chen et al, 2010]. Culture conditions of human umbilical vein endothelial cells (HUVECs) and shear stress experiments for OS, PS, and low flow (static) conditions were performed as previously described [Chen et al, 2010; Guo et al, 2007]. Pulsatile shear flow (PS) or oscillatory shear flow (OS) was applied to ECs with a shear stress of 12 ± 5 dyn/cm^2 or 0.5 ± 5 dyn/cm^2, respectively. Samples for RNA sequencing analysis were collected at 1, 2, 3, 4, 6, 9, 12, 16, 20, and 24 hours after exposure to shear. ECs were also exposed to a "low flow" condition, i.e. 0.5 dyn/cm^2 laminar flow, and designated as a "static" flow condition (ST). Samples for RNA sequencing analysis were collected at hour 1 for the ST condition. Two biological replicates were collected for each time point and for each shear condition. All wet lab experiments were conducted by our collaborating labs.

RNA isolation and RNA-seq Library Preparation. The total RNA from HUVECs were extracted by using mirVana miRNA Isolation Kit (Ambion, Cat#AM1560). The RNA quality was

assessed by RNA Integrity Numbers (RIN), using an Agilent Bioanalyzer and all RNA samples used in this study had a RIN over 9.7 out of 10. The RNAs were polyA selected, fragmented, and random hexamer primed using Illumina TruSeq stranded mRNA sample preparation kit. Constructed cDNA library was subjected to single-ended 100-bp sequencing on Illumina Hi-seq 2000 instrument, generating an average of 30 million reads per sample. All wet lab experiments were conducted by our collaborating labs.

RNA-seq Data Analysis. RNA-seq fastq files were aligned to the Human Reference Genome (version hg19 / Human.B37.3) and converted to raw Refseq transcript count files (Refgene) using the Omicsoft Aligner (OSA) [Hu et al, 2012]. Raw refseq transcript counts were combined into raw gene counts by summation. Raw gene counts were normalized and analyzed for differential expression using DESeq2 [Love et al, 2014]. Pairwise comparisons of OS vs PS, OS vs ST and PS vs ST were performed for each time point. In comparisons against ST, the ST condition was always set to hour 1. In the case of OS vs PS, data were compared at each respective time point. After completing the differential expression analysis, only the genes with mean normalized read count greater than 5 were considered for further analysis. A raw p-value cutoff of 0.05 was used to determine differential expression of genes across conditions. Adjusted p-values were not used for differential expression analysis because of the prohibitively low number of differentially expressed genes in the early hours of shear exposure.

Temporal Map of Functional Regulation. Gene Set Enrichment Analysis (GSEA) was used for pathway enrichment analysis and was performed on the log fold change values of genes between OS vs PS [Subramanaian et al, 2005]. Pathway enrichment analysis was also conducted through the ConsensusPathDB platform spanning the KEGG, Reactome, and Wikipathways databases [Kanehisa et al, 2016; Kamburov et al, 2011] on all genes with p-

value < 0.05. Functional pathways describing temporal activation were constructed through a combination of curated pathways and manual literature search.

Cluster Analysis of TFs. In order to study the mechanistic regulation of TFs in ECs under shear, 1391 putative human TFs were considered based on a previous study [Vaquerizas et al, 2009]. These TFs were filtered by entrez ID. Ensembl IDs were translated to entrez IDs and gene names using Biomart (downloaded 7/31/2016) [Smedley et al, 2015]. Only TFs with a valid entrez ID were considered for further analysis. TFs with valid entrez IDs but with gene names that did not match with our data annotation were manually updated to reflect the annotation. TFs were then further filtered by (a) whether the gene existed within our annotation, (b) their base mean count in our time-series RNA-seq data being greater than 5, and (c) being differentially expressed in OS vs. PS in at least one time point. 500 unique TFs were isolated in this way.

The distance matrix for clustering of TFs was calculated using Euclidean distance. Figure 2.1A shows the resulting hierarchical clusters. An analysis of the cophenetic distances is provided in Figure 2.1B. Based on the size of the distances between clusters, it can be seen that the dendrogram is optimally split at three clusters, with diminishing returns for distinctiveness of clusters beyond k = 3. Figure 2.1A also displays the mean $\log_2$ fold change for each cluster, defined as the following:

$$Mean\ LFC_{cluster\ k} = \frac{\sum_{TF}\left(\frac{\sum_t LFC\ of\ TF\ in\ cluster\ k\ at\ time\ point\ t}{10}\right)}{\#\ of\ TFs\ in\ cluster\ k}$$

Through this analysis, it was determined that two of the three TF clusters were biased towards downregulation in OS vs PS, while the third cluster was approximately unbiased. These two downregulated clusters, when combined, comprise the 96 TFs that were designated as “distinctly downregulated in OS vs PS”.

The TFs of the remaining third cluster were further analyzed. Cluster analysis was repeated on these TFs exclusively (Figure 2.1C). Joint analysis of the cophenetic distance between clusters alongside mean LFC reveals that a selection of five clusters achieves adequate separation of clusters by expression bias while maintaining high distance between clusters (Figure 2.1D). Combination of clusters 4 and 5 (having mean LFC values of 0.230 and 0.323, respectively) comprises the 101 TFs that are designated as “distinctly upregulated in OS vs PS”.

Transcriptional Regulatory Network. Using the TRANSFAC database (Version 2015.4) [Matys et al, 2003], these TFs were mapped to their published transcriptional targets. The identified transcriptional networks, including TFs and its targets, were further filtered based on whether they were significantly differentially expressed in OS vs PS for at least one-time point. In cases where a TF was listed as a protein complex, we manually assessed all subunit genes of the complex and designated the complex to be upregulated or downregulated based on the direction of regulation of the majority of genes in the complex. TF complexes with no clear direction in expression among their subunits were omitted from the network analysis.

Results

ECs were exposed to OS, PS and static condition (ST). RNA-seq samples were collected for ten time points across 24 hours. Figure 2.2 summarizes the findings from the differential expression analysis. The number of differentially expressed genes increases over time across all pairwise conditions. As can be seen in Figure 2.2A, a majority of differentially expressed genes are common in both OS vs ST and PS vs ST. However, among this set of DE genes are hundreds of genes that are exclusively differentially expressed in OS vs ST but not PS vs ST, and vice versa. Figure 2.2B shows the number of differentially expressed genes in OS vs PS, further highlighting the importance of the type of shear stress in transcriptomic response. In order to study the specific impact of the type of shear on endothelial cell gene

expression, we focus on OS vs PS differential expression and its mechanistic and phenotypic consequences. We organize the results in terms of phenotypic responses studied as a function of time.

Cell Cycle. Prior work has shown that cell cycle activity in ECs is higher under OS than PS [Garcia-Cardeña et al, 2001]. Our GSEA results show consistent upregulation of cell cycle-related pathways under OS vs PS, with processes related to G1/S phase transition being among the top enriched Reactome pathways (Tables 2.1, 2.2, 2.3, 2.4). Figure 2.3 represents the G1/S transition pathway obtained from integration of our data, legacy pathways and literature [Kanehisa et al, 2016; Sur et al, 2016; Ertosun et al, 2016; Yamaguchi et al, 2010; Lolli et al, 2005; Siu et al, 2012].

Most of these genes are upregulated in OS beginning at hours 4 or 6. The major exceptions are CDKN2D and CCND2. CDKN2D is a repressor of G1/S progression, thus their downregulation in OS is consistent with literature [Henley et al, 2012]. The cyclin D genes do not show consistent differential expression, but the cyclin E genes, which present an alternative pathway to E2F1 activation, are upregulated over time. The expression profiles observed in our data is consistent with a previous study by Ohtani et al, which showed that E2F1 overexpression can upregulate cyclin E but not cyclin D [Ohtani et al, 1995]. E2F1-induced cyclin E further activates E2F1, perpetuating G1/S transition – which is reinforced in our data by the similar expression profiles between E2F1 and CCNE2. This is also supported by the dissimilarity between the expression profiles of cyclin D and cyclin E.

E2F1 is a TF activated by CDK-mediated phosphorylation of the retinoblastoma (Rb) protein. It is an essential contributor to the G1/S transition, and is responsible for the transcription of various cell cycle genes including E2F1 itself [Ertosun et al, 2016; Calzone et al, 2008]. E2F1 shows observable upregulation in OS vs PS as early as hour 6, achieving statistical significance in hour 24, which supports elevated G1/S transition activity in OS vs PS.

This finding is consistent with a previous finding that the level of phosphorylated Rb protein, which is necessary for the activation of E2F1, decreases in bovine arterial endothelial cells when exposed to laminar shear stress, particularly after 4 hours [Lin et al, 2000].

To further understand the timing of cell cycle transition in ECs under shear, global pathway analysis through the ConsensusPathDB platform was performed separately for OS vs ST and PS vs ST [Kanehisa et al, 2016; Kamburov et al, 2011]. We performed a cluster analysis on the ratio of p-values between OS vs ST and PS vs ST for all functional pathways. Importantly, starting at hour 6, many pathways specific to cell cycle progression begin to show differences in enrichment in OS vs PS (Figure 2.4). This finding, combined with the expression profile of E2F1 and the GSEA results in Table 2.1 – 2.4, suggests that cell cycle activities in OS and PS begins to differ between hours 4 and 6 after the initial exposure to shear.

The OS-upregulated cell cycle activity is concomitant to the enrichment of other prominent pathways such as ribosomal production and activity. The ribosome KEGG pathway is among the most strongly upregulated pathways, suggesting a global upregulation of ribosomal proteins (Table 2.3). The proteasome is also strongly enriched in hours 20 and 24 (Table 2.3), potentially indicating an increase in protein expression and degradation, which are in concert with increase of cell cycle activity and proliferation under OS.

Oxidative Stress. ECs under OS are known to undergo greater oxidative stress than those under PS [Hsieh et al, 2014; Hsiai et al, 2007; De Keulenaer et al, 1998]. Figure 2.5 represents a reconstruction of the oxidative stress pathway based on established genes involved in oxidative stress responses [Kim et al, 2014; Bhabak et al, 2010; Leopold et al, 2003; McDonagh et al, 2013; Reyes et al, 2016].

The genes that mediate superoxide and hydrogen peroxide production are upregulated in OS. For example, NOX4 is upregulated, achieving statistical significance as early as hour 2.

SOD1 is not differentially expressed, but mitochondrial SOD2 is distinctly upregulated in OS from hour 2 to hour 6, peaking at hour 4.

The shear-specific transcriptional response of oxidative stress genes varies in OS vs PS both in the identity of genes and the timing of their differential expression. Catalase is statistically significantly downregulated in OS vs PS starting at hour 9. Glucose-6-phosphate dehydrogenase (G6PD), a potent antioxidant protein involved in the pentose phosphate pathway, is downregulated in OS vs PS in some of the early hours and from hour 12 onward (Table 2.5). Genes for glutathione peroxidases (GPX1, GPX4, GPX8) and peroxiredoxins (PRDX2, PRDX4, PRDX6) were upregulated in OS vs PS beginning at hour 20.

NQO1, a gene that produces an antioxidant enzyme [Dinkova-Kostova et al, 2010], is distinctly downregulated in OS vs PS starting in hour 9 (Figure 2.6). JUNB, which transcribes NQO1 [Jaiswal et al, 2000], also has a distinct expression profile, being significantly downregulated in OS vs PS starting in hour 4. JUNB is the only member of the AP-1 family of TFs to be continuously downregulated in OS vs PS.

The metallothioneins are a family of proteins that are involved in protection from oxidative stress, and are also induced by hypoxia [Schulkens et al, 2014]. Several metallothioneins, particularly MT1X, were strongly upregulated in OS vs PS, with MT1X achieving statistical significance in hour 9. The OS vs PS expression of HIF1A, a hypoxia-inducible factor, is upregulated with a similar time course and achieves statistical significance at hour 4.

Inflammation. ECs exposed to OS exhibit pro-inflammatory phenotypes through the activation of NF-κB and the upregulation of several pro-inflammatory cytokines [Simmons et al, 2016; Szmitko et al, 2003]. The temporally-varying expression profiles of NF-κB targets and canonical pro-inflammatory genes show a sharp OS vs PS upregulation between hours 2 and 6 (Figure 2.7). Peak upregulation of NF-κB occurs at hour 4, coinciding with the expression

peaks for VCAM-1 and E-selectin. MCP-1 (CCL2) and IL8 have clear differences in the magnitude of OS vs PS differential expression, with both genes exhibiting statistical significance at hours 3, 4, and 6. NF-κB activation has been shown to be influenced by oxidative stress [Kim et al, 2014; Tornatore et al, 2012]. The temporal profile of the OS vs PS upregulation of NF-κB target genes is consistent with those of NOX4 and SOD2 beginning in hour 2 (Figure 2.5). SOD2 overexpression is abrogated after hour 6, consistent with the alteration of expression of NF-κB target genes within the same period.

Endothelial-Mesenchymal Transition. Endothelial to mesenchymal transition (Endo-MT) is a contributor to cardiovascular disease and has been observed in atherosclerotic lesions and severe vasculitis [Evrard et al, 2016; He et al, 2017]. We defined an "endothelial marker" gene set and a "mesenchymal marker" gene set from gene lists taken from literature [Evrard et al, 2016]. We used these gene sets in GSEA to examine Endo-MT enrichment in OS vs PS. Figure 2.8A shows that endothelial marker genes (e.g. NOS3, VWF, and CD34) are strongly downregulated as early as hour 6, whereas mesenchymal marker genes (e.g. CDH2, TPM1, and FBLN5) are upregulated at some time points beginning in hour 12. This suggests that Endo-MT initiation may occur in OS as early as hour 12.

Oxidative stress, hypoxia, and TGF-β signaling have been shown to drive Endo-MT [Evrard et al, 2016]. Genes pertaining to oxidative stress are observed to be upregulated in OS vs PS as early as hour 2, while HIF1A is upregulated as early as hour 4. An investigation of the TGF-β signaling receptors did not reveal a clear direction of regulation in ECs under shear, with a uniform upregulation of both inhibitors and activators being observed (Figure 2.8B).

Putative Endothelial Transcription Factor Network. TF-to-gene networks were constructed using the TRANSFAC database and sets of TFs determined to be distinctly downregulated (Figure 2.9) or upregulated (Figure 2.10) in OS vs PS. Several highly connected TFs were identified within these networks. Among these highly connected TFs

downregulated in OS vs PS are KLF4, considered to be one of the fundamental mechanosensitive TF genes [Hamik et al, 2007], and JUNB, described above as a transcriptional regulator of the antioxidant gene NQO1. Among the highly connected TFs upregulated in OS vs PS are E2F1, described above as crucial to cell cycle progression from G1 to S phase, and HIF1A, described above as an important TF in response to oxidative stress. Thus, we are able to identify key regulators of endothelial function as hubs that are involved in important endothelial pathways, while identifying additional TFs of interest such as CEBPB and EGR1.

Pathway analysis of these networks revealed several TF-TF interactions distinct in OS vs PS. The results show a KLF2-RARG-RARB regulatory pathway for preferential OS downregulation and a SRF-EGR1-HIF1A regulatory pathway for preferential OS upregulation. The expression profiles of KLF2 and RARG indicate that these are strongly downregulated in OS vs PS across all time points. The presence of both RXRA, RARG, and a RARG:RXRA node targeting RARB suggest that RARG and RXRA form a PS-specific complex for transcriptional regulation through interaction with retinoic acid response elements in the genome [Germain et al, 2006]. Similar to RARG and KLF2, RXRA is significantly downregulated in OS vs PS at most time points.

The early upregulation of EGR1 in the data in hour 1, followed by the upregulation of HIF1A starting in hour 2 and achieving statistical significance in hour 4, offers evidence to support the regulation of HIF1A by EGR1 in our system. EGR1 has previously been identified as a high-confidence target of retinoic acid receptors [Balmer et al, 2002]. The expression profile of EGR1 shows a trend of being inverse to that of RXRA. These suggest that the RARG-RXRA complex may repress EGR1 in PS. A schematic of these pathways and their expression profiles is shown in Figure 2.11.

Functional analysis of these targets reveals that the OS-specific transcription factors EGR1, HIF1A, and E2F1 regulate genes related to a variety of functions (Table 2.6). As expected, E2F1 is shown to regulate a variety of cell cycle genes, while EGR1 and HIF1A regulate several genes pertaining to inflammatory response and cellular adhesion.

Discussion

Our study offers the first detailed temporal map of EC regulation. This study, which uses multiple time-series RNA-seq data collected from human ECs subjected to physiological and pathophysiological flow conditions, examines systematically the functional contexts by which the shear-responsive pathways evolve over time. Our longitudinal analysis shows the dynamics of functional pathways, thus providing insights into the causal relationships between cellular response mechanisms as a function of time.

Secondary Effects of Differential Cell Cycle Activity. Several genes in the cell cycle pathway have functions external to cell cycle progression. The E2F transcription factors can modulate many genes relating to apoptosis, post-translational modifications, and metabolic functions [Ertosun et al, 2016; Denechaud et al, 2016; Muthusamy et al, 2015]. A recent study has implicated E2F1 in autophagy regulation, wherein E2F1 transcriptionally upregulates v-ATPase, thus promoting the trafficking of lysosomes to the cell periphery, activating mTOR, and inhibiting autophagy [Meo-Evoli et al, 2015]. This mechanism is consistent with previous findings about mTOR activation under OS [Guo et al, 2007] and our data on the strong downregulation of ATG9B in OS vs PS (Table 2.5). This is also consistent with the recent finding that laminar shear stress induces autophagy via a SIRT1-dependent mechanism [Liu et al, 2015]. Hence, disturbed flow regulation of E2F1 may have a role in the negative regulation of autophagy. Additionally, repression of autophagy is also associated with Endo-MT [Singh et al, 2015], which occurs in our study on ECs under OS towards the later hours. The activation of mTOR through E2F1-induced lysosomal trafficking may provide another mechanism by

which E2F1 participates in cell cycle progression in ECs under OS [Guo et al, 2007]. E2F1 upregulation begins in the mid-range hours of our time-series data on OS vs PS. This association is supported by the finding that several v-ATPase genes are also upregulated beginning in the mid-range hours (Table 2.5).

The cyclin-dependent kinases have been shown to have roles that are external to the cell cycle. Activated CDK4 has been shown to phosphorylate DNMT1 [Acevedo et al, 2016]. ECs under disturbed flow have been observed to undergo DNA hypermethylation via DNMT1 [Zhou et al, 2014]. The upregulation of CDK4 in OS vs PS in the middle hours and onward (Fig. 2A) suggests that DNA hypermethylation may also occur under OS within a similar time frame. It has also been shown that CDK2 can phosphorylate the histone methyltransferase EZH2 [Siu et al, 2012]. EZH2 expression is upregulated in OS vs PS in the later hours in our dataset (Table 2.5). CDK2 has been shown to inhibit FOXO1, a multi-functional TF that is coupled with cellular metabolic and survival pathways [Siu et al, 2012], and hence we examined the downstream targets of FOXO1 in order to infer the temporal profile of CDK2 phosphorylation activity in ECs. FOXO1 has been shown to transcribe CDK inhibitors such as CDKN2D and CDKN2B [Katayama et al, 2008]. As shown in Figure 2.3, we observe that cyclin E, a CDK2 activator, is upregulated in OS vs PS through most time points and achieves statistical significance at hour 24. We also observe CDKN2D to be strongly downregulated in OS vs PS at most time points. Therefore, we postulate that CDK2 may regulate FOXO1 at the protein level beginning in the early hours. These findings suggest that CDK2 is active and may be activating EZH2. The CDK2-mediated activation of EZH2, along with the CDK4-mediated activation of DNMT1, suggests that DNA hypermethylation in ECs under OS can be explained in part by increased cell cycle activity under OS. These findings also provide mechanistic links connecting shear stress-regulated cell cycle activity and epigenetic regulation.

Secondary Effects of Differential Oxidative Stress Activity. JUNB has been implicated in cell cycle processes in addition to transcribing antioxidant genes such as NQO1. JUNB, whose protein level is cell cycle-dependent [Piechaczyk et al, 2008], has been shown to inhibit G1-S phase transition via upregulation of CDK inhibitors such as CDKN2A, in addition to inhibiting the expression of cyclin D [Piechaczyk et al, 2008]. Thus, JUNB expression may have a role in suppressing cell cycle progression under PS, though these specific effects of JUNB may require additional data to explain. It has been suggested that peroxiredoxin 1 (PRDX1) is a mechanosensitive antioxidant that is upregulated in laminar shear stress in bovine aortic ECs [Mowbray et al, 2008]. Although PRDX1 was not significantly regulated in our dataset, PRDX2, PRDX4, and PRDX6 were all upregulated in OS vs PS in the later hours. Our observed results with these peroxiredoxin genes are partially corroborated with publicly available HUVEC data where transcripts are measured 72 hours after laminar shear stress, showing a significant downregulation of PRDX2 and PRDX4 [Maleszewska et al, 2016]. This suggests that the oxidative stress response under OS may involve multiple peroxiredoxin proteins.

TF Network - KLFs as Master Regulators. KLF4 has been shown to induce CEBPB expression and can bind directly to the CEBPB promoter [Birsoy et al, 2008]. Time-series ChIP-seq studies in mouse liver regeneration systems have shown that mouse Cebpb can bind within 2000 bp upstream of the mouse JunB transcription start site, with this interaction being strongly observed in hour 3 after a partial hepatectomy [Jakobsen et al, 2013]. In our study, CEBPB downregulation in OS vs PS occurs primarily in the early hours, achieving statistical significance in hour 4. This may suggest that CEBPB upregulation is dependent on the initial imposition of stress on the cell. JUNB is statistically significantly downregulated in OS vs PS in hour 4, concurrent with the downregulation of CEBPB. These findings suggest that a KLF4-CEBPB-JUNB regulatory pathway is activated in the early hours of shear

response. Based on our analysis, we postulate that this pathway has functional consequences in cell cycle suppression and oxidative stress response in the later hours. A previous study has demonstrated that KLF1 can inhibit E2F2 by binding to an intronic enhancer region [Tallack et al, 2009]. Hence, we investigated the ability of KLF2 to directly regulate E2F1. HUVEC chromatin state segmentation displayed on the UCSC genome browser [Rosenbloom et al, 2013; Ernst et al, 2010; Ernst et al, 2014] predicts an enhancer region within the first intron of E2F1 from the transcription start site. Scanning of this region using available binding motifs for these TFs from TRANSFAC [Matys et al, 2003] detected KLF4 and KLF2 motifs in both the forward and reverse strands. This raises the possibility that KLF2 and KLF4 regulate and repress E2F1. This repression can be observed at the mid hours in our data set, where E2F1 expression diverges between OS and PS.

Combined Gene-TF-Phenotypic Network. In this study, we present a dynamical view of the response of ECs to differential shear stress. Figure 2.12 shows the putative regulatory network derived from this system-wide time-series analysis between OS and PS. Through detailed pathway analysis, we elucidated several key molecular hubs (e.g. TFs) that instigate distinct transcriptomes in ECs subjected to atheroprone OS vs atheroprotective PS. Our study verifies and extends the previously studied mechanisms and pathways to provide an integrated perspective on the regulation of cellular functions that lead to defined endothelial phenotypes. Further, the transcriptional regulation mechanisms provide new insights into causality of mechanisms that lead to stress responses. The kinetic hypotheses derived in this work also serve to define experiments in animal models that can have implications for diseases such as atherosclerosis. The systems biology approach employed in this study may well serve as a template for future studies, where the input datasets would be derived via various high-throughput “-omics” techniques. These might provide a window on regulatory

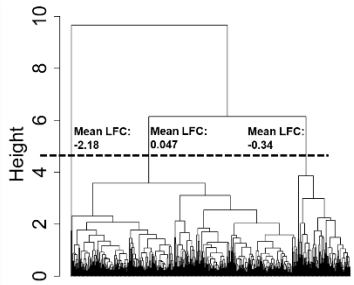
events at the epigenetic and post-translational levels that could have important pathophysiologic implications.

Acknowledgments

This work was supported in part by US NIH research grants R01HL106579 and HL108735 (SC, SS, and JS), U01CA198941 (SS), U19AI090023 (SS), K99/R00HL122368 (ZC) and NSF research grant CCF0939370 (SS). NA was supported by the NIH Training Grant T32GM8806 (PI: SS).

Chapter 2 is a modified presentation of “Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress" as it appears in PNAS 2017 by Nassim E Ajami, Shakti Gupta, Mano R Maurya, Phu Nguyen, Julie Yi-Shuan Li, John Y-J Shyy, Zhen Chen, Shu Chien, and Shankar Subramaniam. The dissertation author was the primary author of this material.

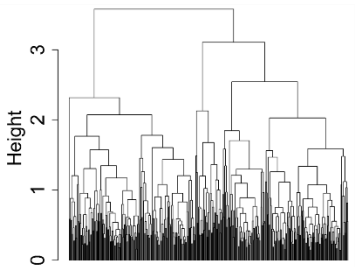
A



B

Number of Clusters (k)	1	2	3	4	5	6	7	8	9	10
Minimum Height Cutoff for k Clusters	9.66	6.14	3.86	3.58	3.11	2.99	2.54	2.32	2.24	2.15
Difference between cutoff for (k-1) clusters and cutoff for k clusters	-	3.52	2.28	0.28	0.47	0.12	0.44	0.23	0.07	0.09

C



D

Number of Clusters (k)	1	2	3	4	5	6
Minimum Height Cutoff for k Clusters	3.58	3.11	2.54	2.32	2.13	2.07
Difference between cutoff for (k-1) clusters and cutoff for k clusters	-	0.47	0.56	0.23	0.19	0.05
Mean LFC per cluster	1: 0.047	1: -0.051 2: 0.128	1: -0.051 2: 0.086 3: 0.323	1: -0.051 2: 0.012 3: 0.230 4: 0.323	1: -0.052 2: 0.137 3: 0.012 4: 0.230 5: 0.323	1: -0.052 2: 0.137 3: 0.012 4: 0.230 5: 0.294 6: 0.583

Figure 2.1: Details of Cluster Analysis of TFs in TF-to-Gene Network. (A) Dendrogram of 500 hierarchically clustered transcription factors that were selected based on their level of expression. The dendrogram was split into three clusters. Displayed are the mean log fold changes per cluster, which is calculated from the log fold change for each transcription factor from each time point. (B) Cluster analysis of dendrogram. The selection of three clusters was determined by the observed change in magnitude of height cutoff between three and four clusters. (C) Dendrogram of the largest of the three clusters with mean LFC 0.047, along with (D) further cluster analysis of this cluster. The selection of five clusters was determined by the observed stability of mean LFC from five clusters to six, in addition to the drop in difference between cutoffs.

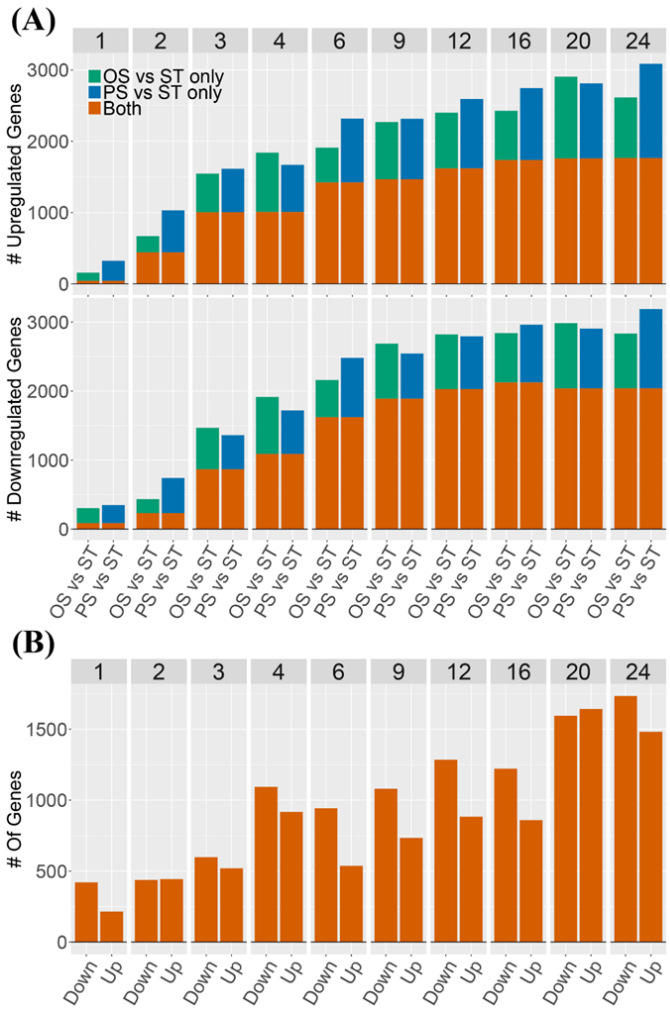


Figure 2.2: Number of differentially expressed genes. (A) Number of differentially expressed genes using the ST condition (hour 1) as control. Top panel: the number of upregulated genes; bottom panel: the number of downregulated genes. In each panel, at each time point, the left and right bars show the number of differentially expressed (DE) genes for OS vs ST and PS vs ST comparisons, respectively. (B) Number of differentially expressed genes in OS vs PS. In each panel, at each time point, the left and right bars show the number of differentially expressed (DE) genes that are downregulated and upregulated in OS vs PS, respectively.

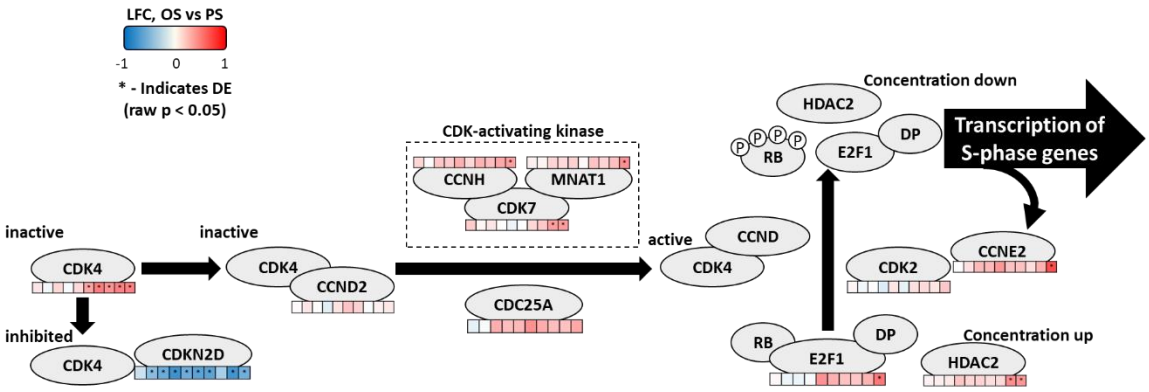


Figure 2.3: G1-to-S transition pathway. G1-to-S transition is dependent on the E2F1 activation through RB phosphorylation, which is facilitated in part by CDK2 (bound to cyclin E) and CDK4 (bound to cyclin D). The CDK proteins must be activated by interacting with the CDK-activating kinase. CDKs can also be inhibited by proteins such as CDKN2D. OS vs PS log fold-change (LFC) data projected onto the custom pathway consisting of genes and mechanisms exhibiting differential response between OS vs PS. The heat maps below the gene nodes show the time course of transcriptional changes, representing from left to right: hours 1, 2, 3, 4, 6, 9, 12, 16, 20, and 24.

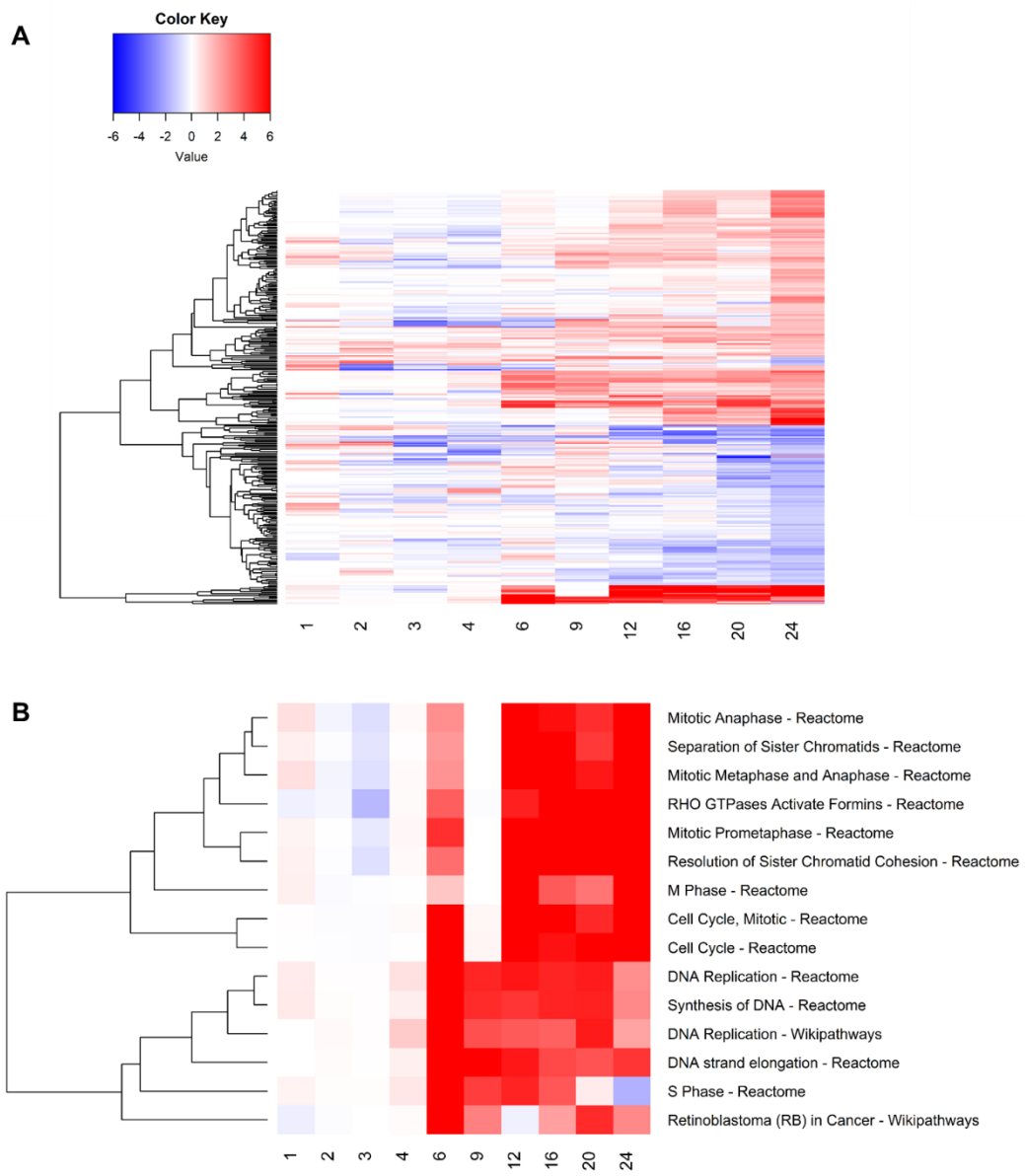


Figure 2.4: Cluster analysis of functional pathway enrichment. (A) Clustered heatmap of $\log_{10}$ratio of the p-values in OS vs ST and in PS vs ST. Positive values indicate that PS has the smaller p-value than OS, thus the pathway is more significantly enriched in PS. Negative values, vice versa. (B) When the heatmap is divided into two clusters (via hierarchical clustering), the smaller cluster shows a distinct difference in enrichment of several cell cycle pathways starting around hour 6. This is particularly pronounced with many pathways relevant to S phase.

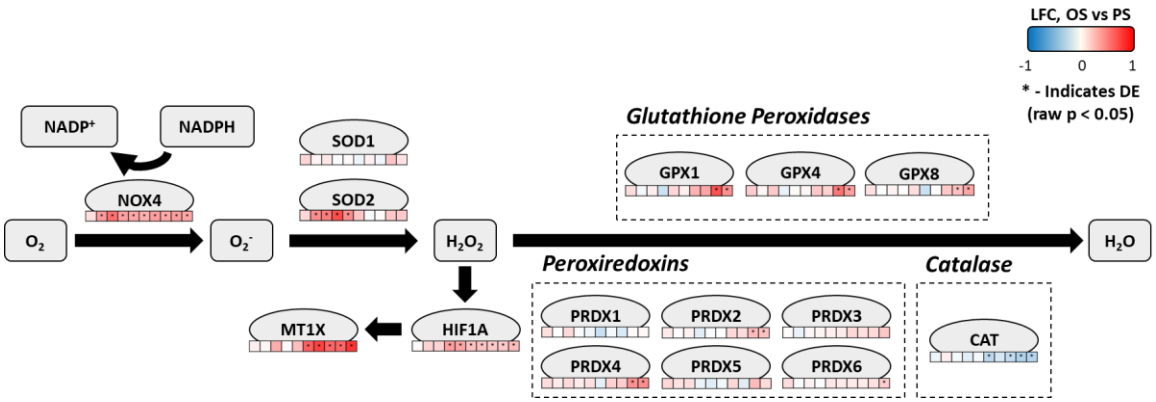


Figure 2.5: Reconstructed pathway of oxidative stress and superoxide metabolism. Oxygen in the cell is converted to superoxide, hydrogen peroxide, and finally water. Reactive oxygen species are also known to activate HIF1A, a marker of hypoxia. The metallothionein MT1X is also shown, which is thought to be hypoxia-responsive. OS vs PS log fold-change (LFC) data projected onto the custom pathway consisting of genes and mechanisms exhibiting differential response between OS vs PS. The heat maps below the gene nodes show the time course of transcriptional changes, representing from left to right: hours 1, 2, 3, 4, 6, 9, 12, 16, 20, and 24.

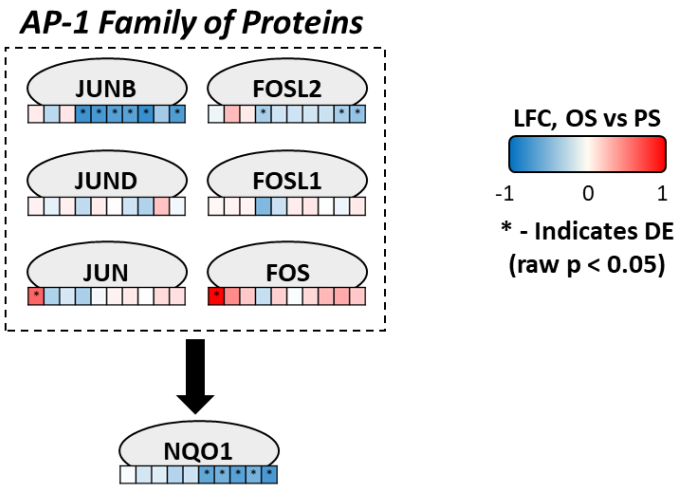


Figure 2.6: AP-1 family of genes and the antioxidant product NQO1. OS vs PS log fold-change (LFC) data projected onto the custom pathway consisting of genes and mechanisms exhibiting differential response between OS vs PS. The heat maps below the gene nodes show the time course of transcriptional changes, representing from left to right: hours 1, 2, 3, 4, 6, 9, 12, 16, 20, and 24.

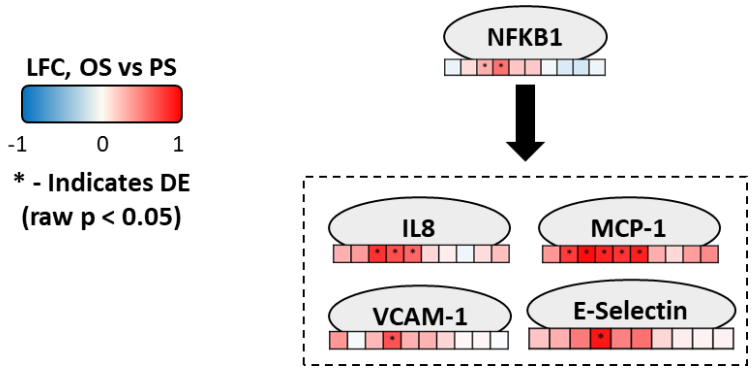


Figure 2.7: NF-κB and NF-κB target genes. OS vs PS log fold-change (LFC) data projected onto the custom pathway consisting of genes and mechanisms exhibiting differential response between OS vs PS. The heat maps below the gene nodes show the time course of transcriptional changes, representing from left to right: hours 1, 2, 3, 4, 6, 9, 12, 16, 20, and 24.

A

Downregulated Genes - Hour	1	2	3	4	6	9	12	16	20	24
ENDOTHELIAL_MARKERS	1	1	1	1	0.031	0	0	0	0	0
MESENCHYMAL_MARKERS	1	0.165	1	0.509	0.019	1	1	1	1	1
Upregulated Genes - Hour	**1**	**2**	**3**	**4**	**6**	**9**	**12**	**16**	**20**	**24**
ENDOTHELIAL_MARKERS	0.173	0.169	0.156	0.264	1	1	1	1	1	1
MESENCHYMAL_MARKERS	0.042	1	0.080	1	1	0.086	0.013	0.057	0.148	0.010

B

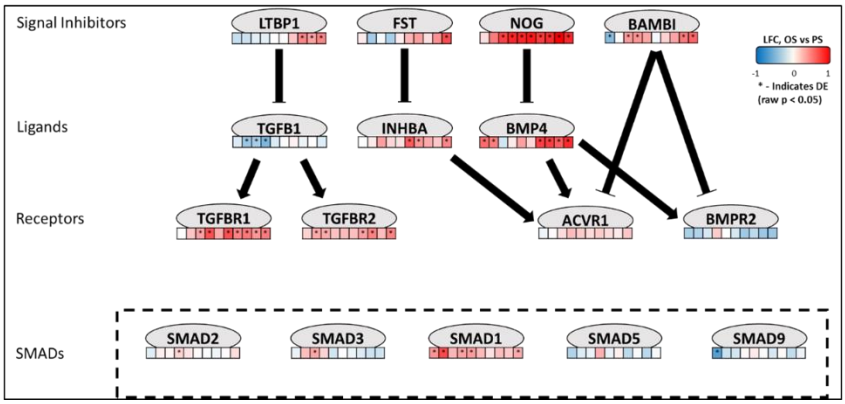


Figure 2.8: Endothelial-mesenchymal transition (EndoMT) and related pathways. (A) Raw p-values of custom-made EndoMT pathways through GSEA. P-values that are statistically significant are highlighted in red, with one borderline case highlighted in yellow. (B) TGF-beta signaling pathway, based on pathways in KEGG and Wikipathways [Gal et al, 2016].

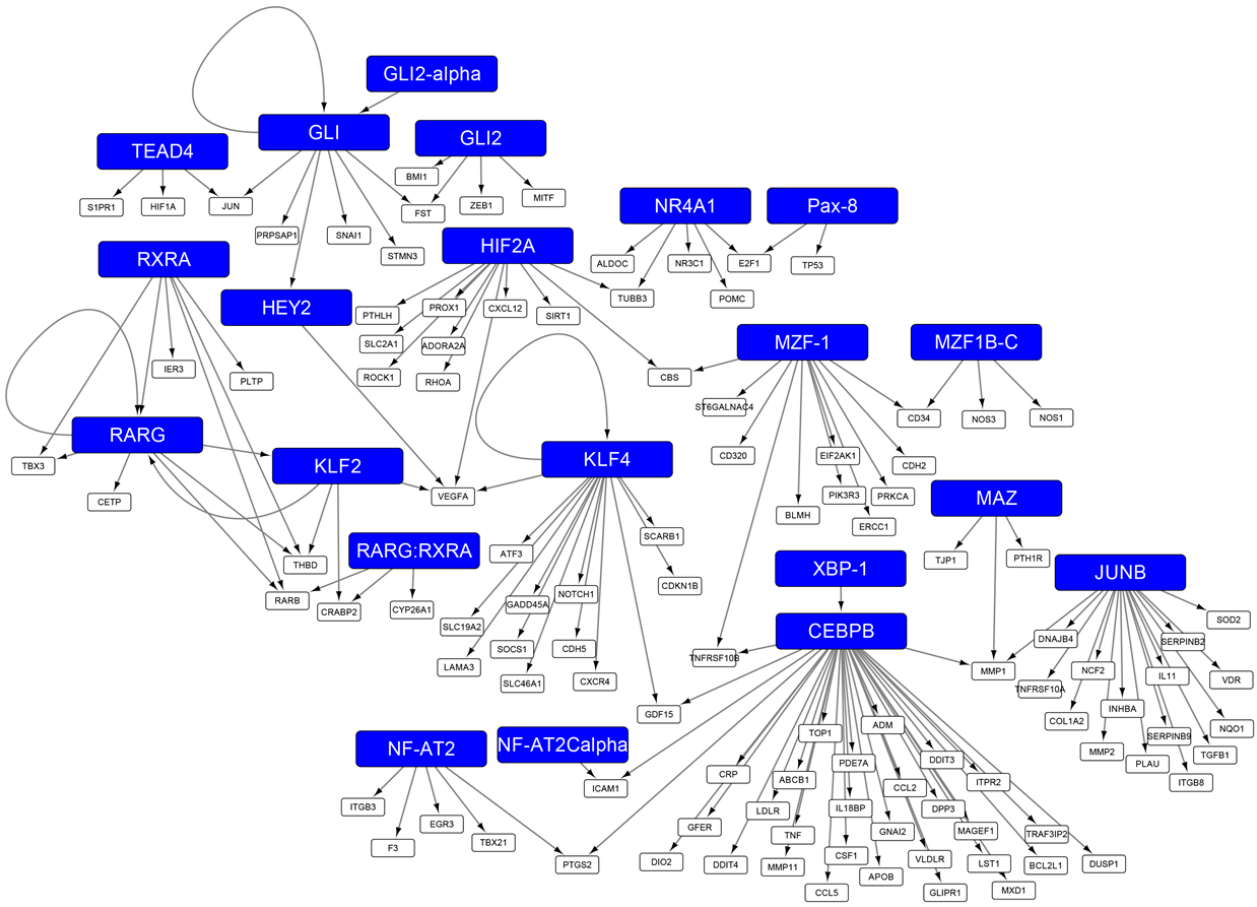


Figure 2.9: Largest contiguous portion of TF-to-target network for TFs that are downregulated in OS vs PS. TFs were chosen based on cluster analysis of expression data. Gene targets were chosen based on the presence of an entry in TRANSFAC, as well as whether the target was differentially expressed in at least one time point.

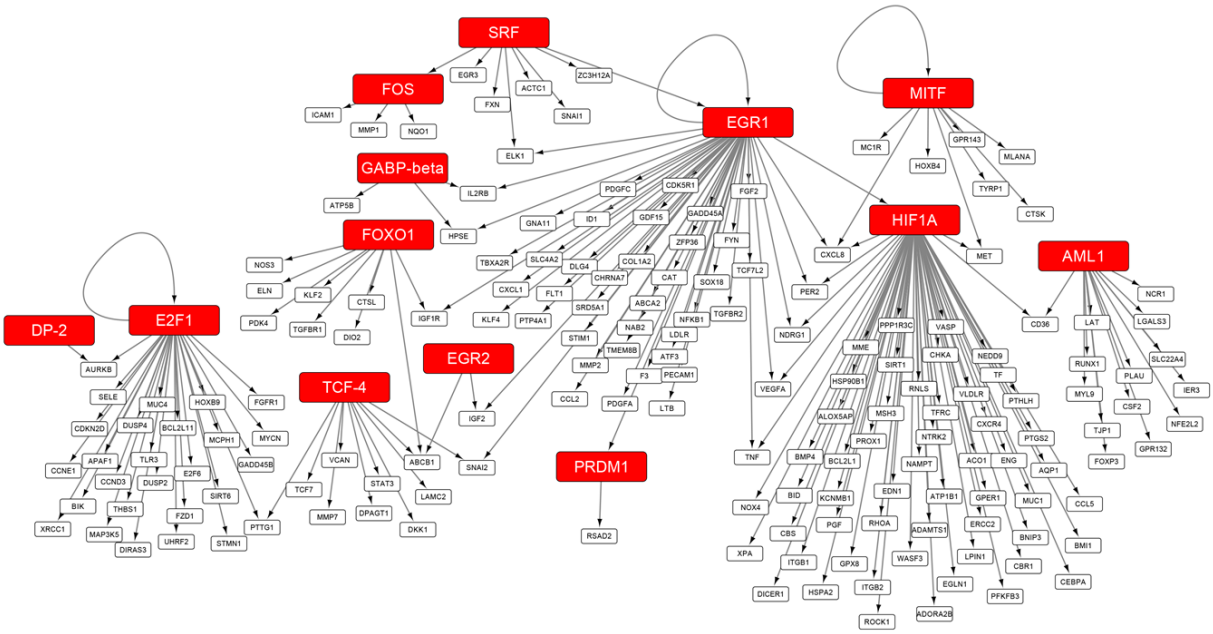


Figure 2.10: Largest contiguous portion of TF-to-target network for TFs that are upregulated in OS vs PS. TFs were chosen based on cluster analysis of expression data. Gene targets were chosen based on the presence of an entry in TRANSFAC, as well as whether the target was differentially expressed in at least one time point.

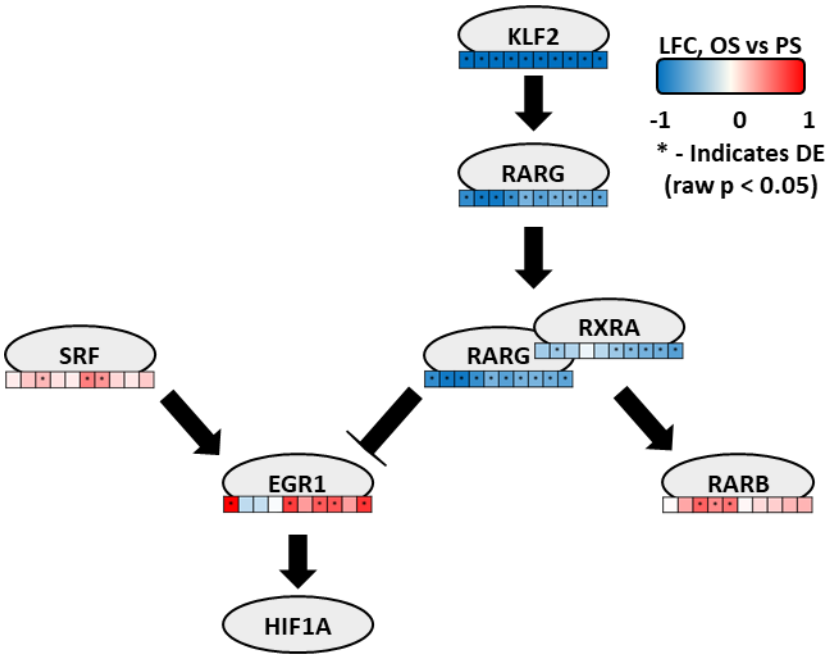


Figure 2.11: Longest paths in TF to gene networks. Pathways were found by observing the longest contiguous TF-to-TF pathways in the networks in figures 2.9 and 2.10. Because RXRA, RARG, and the RXRA:RARG complex all independently appear in the network in figure S4A and all target RARB, it was inferred that an RARG:RXRA complex is a part of this regulatory pathway. RARG:RXRA inhibition of EGR1 was inferred from manual literature curation [Balmer et al, 2002].

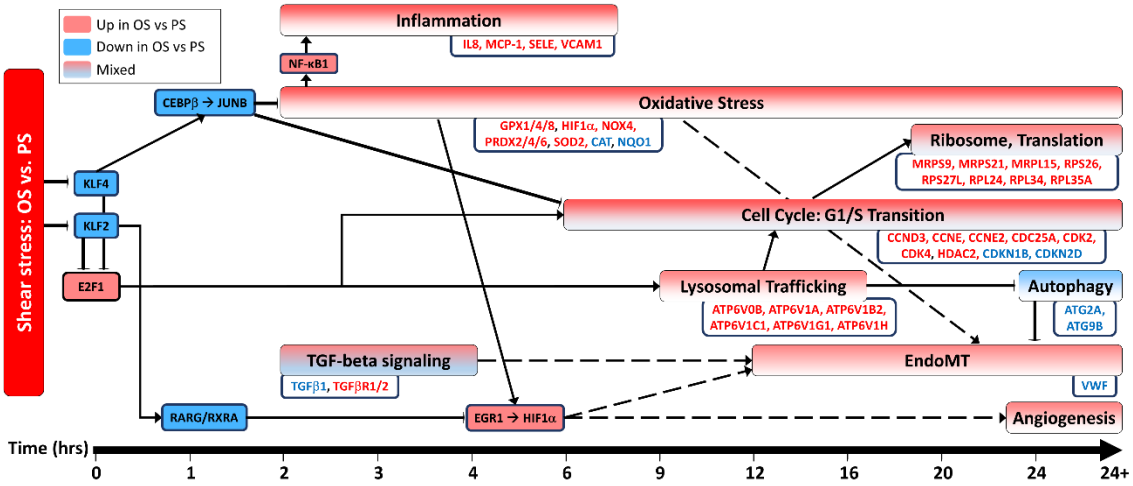


Figure 2.12: Relative differences in gene expression in ECs exposed to OS or PS over time. Pathways and genes in red are upregulated in OS vs PS, while those in blue are downregulated in OS vs PS. TGF-β signaling pathway is labeled “mixed” because its genes’ expression directions were mixed. Divergence in cell cycle activity between OS and PS begins by hour 6 with E2F1 upregulation in OS. This divergent cell cycle activity may occur through KLF2- and KLF4-mediated repression of E2F1 expression in PS. E2F1 may upregulate v-ATPase genes under OS, as observed in the intermediate hours post-shear. This promotes lysosomal trafficking to the cell periphery, thus activating mTOR and inhibiting autophagy by hour 20. The activation of mTOR by E2F1 may also contribute to divergence in cell cycle activity through S6K activation [Guo et al, 2007]. Through CEBPB, KLF4 activates JUNB at hour 4, which may contribute to KLF4-induced inhibition of cell cycle activity and to anti-oxidative stress activity in PS. Genes that contribute to ROS production are upregulated in OS vs PS by hour 2, and other oxidative stress-related genes exhibit changes up to hour 24. This increased ROS production activates NF-κB and promotes several early response genes pertaining to inflammation such as MCP-1 and VCAM-1. Inflammation-related genes show changes during hours 2-9. KLF2 transcriptionally activates RARG, which forms a heterodimer with RXRA and may repress EGR1 activity. EGR1, along with ROS production, transcriptionally activates HIF1A, which are all upregulated in OS vs PS. HIF1A is observed to be upregulated beginning in hour 4 and may contribute to angiogenesis and to Endo-MT. Endo-MT may occur in OS beginning in hour 12. Oxidative stress and autophagy repression, both of which occur in OS, also contribute to Endo-MT.

Table 2.1: Top enriched pathways through GSEA: The top 10 up-regulated enriched Reactome pathways. The analysis examined log fold changes of genes between OS and PS. Values represent P values that were corrected for familywise error rate with a Bonferroni correction. Values highlighted in red are <0.05. The pathways were sorted by increasing P values, with sorting on the first hour, then the second hour, until reaching hour 24.

Hour	1	2	3	4	6	9	12	16	20	24
REACTOME_PROCESSING_ OF_CAPPED_ INTRON_CONTAINING_PRE_MRNA	1	1	1	1	0.068	0	0	0	0	0
REACTOME_MRNA_PROCESSING	1	1	1	1	0.388	0	0	0	0	0
REACTOME_MRNA_SPLICING	1	1	1	1	0.584	0	0	0	0	0
REACTOME_MITOTIC_M_ M_G1_PHASES	1	1	1	1	0	0.026	0	0	0	0
REACTOME_DNA_REPLICATION	1	1	1	1	0	0.041	0	0	0	0
REACTOME_SYNTHESIS_OF_DNA	0.315	1	0.709	1	0	0.101	0	0	0	0
REACTOME_G1_S_TRANSITION	0.497	1	0.219	1	0	0.153	0	0	0	0
REACTOME_M_G1_TRANSITION	0.011	1	0.01	0.997	0	0.204	0	0	0	0
REACTOME_ASSEMBLY_OF_THE_ PRE_REPLICATIVE_COMPLEX	0.001	1	0.004	0.609	0.003	0.558	0	0	0	0
REACTOME_METABOLISM_OF_RNA	0	1	0	0.969	1	0.662	0	0	0	0

Table 2.2: Top enriched pathways through GSEA: The top 10 down-regulated enriched Reactome pathways. The analysis examined log fold changes of genes between OS and PS. Values represent P values that were corrected for familywise error rate with a Bonferroni correction. Values highlighted in red are <0.05. The pathways were sorted by increasing P values, with sorting on the first hour, then the second hour, until reaching the hour 24.

Hour	1	2	3	4	6	9	12	16	20	24
REACTOME_ABC_FAMILY_ PROTEINS_MEDIATED_TRANSPORT	1	1	1	1	1	0.92	0.945	0.855	0.07	0.021
REACTOME_GPCR_ DOWNSTREAM_SIGNALING	1	1	0.999	1	0.992	0.438	0.058	0.034	0.011	0.079
REACTOME_INTEGRIN_CELL_ SURFACE_INTERACTIONS	1	1	1	1	1	1	0.999	0.749	0.044	0.091
REACTOME_POTASSIUM_ CHANNELS	1	0.889	0.047	0.359	0.644	1	0.999	0.989	0.928	0.161
REACTOME_GPCR_ LIGAND_BINDING	1	1	1	1	0.947	0.581	0.639	0.082	0.202	0.225
REACTOME_SIGNALING_ BY_GPCR	1	1	1	1	0.999	0.775	0.333	0.115	0.098	0.426
REACTOME_ COMPLEMENT_CASCADE	1	1	1	1	0.124	0.996	0.781	1	0.999	0.431
REACTOME_G_ALPHA_I_ SIGNALLING_EVENTS	1	1	1	1	0.975	0.756	0.89	0.298	0.333	0.615
REACTOME_PLATELET_ HOMEOSTASIS	1	1	1	1	1	0.995	0.215	0.655	0.667	0.66
REACTOME_PLC_BETA_ MEDIATED_EVENTS	0.575	1	1	1	0.997	0.996	0.84	0.982	0.038	0.674

Table 2.3: Top enriched pathways through GSEA: The top 10 up-regulated enriched KEGG pathways. The analysis examined log fold changes of genes between OS and PS. Values represent P values that were corrected for familywise error rate with a Bonferroni correction. Values highlighted in red are <0.05. The pathways were sorted by increasing P values, with sorting on the first hour, then the second hour, until reaching the hour 24.

Hour	1	2	3	4	6	9	12	16	20	24
KEGG_RIBOSOME	0	0.001	0	0.187	0.268	0.998	0	0	0	0
KEGG_DNA_REPLICATION	0.952	1	1	1	0	0.006	0	0.005	0	0
KEGG_PROTEASOME	0	0.835	0.001	0.133	0.605	0.943	0	0.024	0	0
KEGG_SPLICEOSOME	0.784	1	0.981	1	1	0.007	0	0.027	0	0
KEGG_OXIDATIVE_ PHOSPHORYLATION	0	1	0	1	1	1	0	1	0	0
KEGG_PARKINSONS _DISEASE	0.002	1	0.001	1	1	1	0.032	1	0	0
KEGG_CELL_CYCLE	1	1	1	1	0.176	0.635	0.288	0.622	0.323	0.003
KEGG_NUCLEOTIDE_ EXCISION_REPAIR	1	1	1	1	0.522	0.983	0.049	0.707	0.051	0.005
KEGG_RNA_DEGRADATION	1	1	1	0.952	0.423	0.42	0.958	0.992	0.015	0.057
KEGG_HOMOLOGOUS_ RECOMBINATION	1	1	1	1	0.645	0.995	0.031	0.999	0.669	0.058

Table 2.4: Top enriched pathways through GSEA: The top 10 down-regulated enriched KEGG pathways. The analysis examined log fold changes of genes between OS and PS. Values represent P values that were corrected for familywise error rate with a Bonferroni correction. Values highlighted in red are <0.05. The pathways were sorted by increasing P values, with sorting on the first hour, then the second hour, until reaching the hour 24.

Hour	1	2	3	4	6	9	12	16	20	24
KEGG_ABC_TRANSPORTERS	1	1	0.997	0.997	1	0.565	0.113	0.042	0	0
KEGG_ECM_ RECEPTOR_INTERACTION	0.96	1	0.969	0.872	0.658	1	0.733	0.519	0	0.028
KEGG_GLYCEROPHOSPHOLIPID_ METABOLISM	1	0.634	0.026	0.153	0.181	0.573	0.058	0.214	0.919	0.171
KEGG_METABOLISM_OF_ XENOBIOTICS_BY_CYTOCHROME_P450	1	1	1	0.986	0.437	0.001	0.171	0.069	0.037	0.175
KEGG_NEUROACTIVE_ LIGAND_RECEPTOR_INTERACTION	1	1	0.444	1	0.861	0.319	0.265	0.195	0.175	0.203
KEGG_DILATED_ CARDIOMYOPATHY	1	0.998	0.048	0.052	0.128	0.7	0.265	0.264	0.089	0.257
KEGG_PHOSPHATIDYLINOSITOL _SIGNALING_SYSTEM	0.636	1	0.961	1	0.839	0.917	0.042	0.098	0.02	0.294
KEGG_VASCULAR_SMOOTH _MUSCLE_CONTRACTION	1	1	1	1	0.699	0.999	0.265	0.042	0.096	0.376
KEGG_STEROID_ HORMONE_BIOSYNTHESIS	1	1	1	1	1	0.09	0.149	0.227	0.587	0.428
KEGG_ARACHIDONIC_ ACID_METABOLISM	1	1	1	0.726	0.081	0.007	0.651	0.886	0.544	0.516

Table 2.5: Log2 fold change of select DE genes in OS vs PS across time. Values that achieve statistical significance (raw p < 0.05) are shown. The tables display G6PD, ATG9B, select v-ATPase genes, and EZH2.

		Hour									
Gene name	Base mean	1	2	3	4	6	9	12	16	20	24
G6PD	5,112.55	—	—	—	−0.33	—	—	−0.27	−0.37	−0.36	−0.48
ATG9B	75.19	−0.60	−0.88	−0.79	−1.18	−1.81	−1.89	−1.22	−1.26	−0.84	−1.49
ATP6V1A	3,142.13	—	—	—	0.49	0.51	0.46	—	0.42	0.45	0.58
ATP6V1B2	2,801.60	—	—	—	—	0.17	—	0.17	0.23	0.22	0.36
ATP6V1G1	2,461.78	—	—	—	—	—	—	—	—	—	0.31
ATP6V1G2	18.65	—	—	—	—	—	—	—	—	—	—
ATP6V0B	1,871.63	—	—	—	—	—	—	0.25	—	0.31	0.28
ATP6V1E1	3,375.57	—	—	—	—	—	—	—	—	0.28	—
ATP6V1C1	2,256.86	—	—	—	0.24	—	—	—	—	—	0.22
ATP6V1H	1512.97	—	—	—	—	—	—	—	—	0.21	0.22
EZH2	1,007.55	—	—	—	—	—	—	—	0.34	0.36	0.47

Table 2.6: Select functional pathways pertaining to relevant endothelial functions under shear, and listings of the number of their participating genes in the TRANSFAC-derived network displayed in Figure 2.10. These genes were then examined for their connection with the most central nodes to the original network – EGR1, HIF1A, and E2F1. The number of each TF's differentially expressed (in OS vs PS, $p < 0.05$) targets that are also among the listed pathways are listed in the three right-most columns.

Endothelial Function	Pathway Name - Database	# of DE gene targets In OS TRANSFAC Network	EGR1 Targets	HIF1A Targets	E2F1 Targets
Cell Cycle	Cell cycle - Homo sapiens (human) - KEGG	7	1	0	6
	G1/S Transition - Reactome	3			
	G2/M Checkpoints - Reactome	2			
	M/G1 Transition - Reactome	1			
Inflammation	Cytokines and Inflammatory Response - Wikipathways	7	12	5	1
	Cytokine-cytokine receptor interaction - Homo sapiens (human) - KEGG	23			
	Inflammatory Response Pathway - Wikipathways	4			
Cell Signaling and Stress Response	Oxidative Stress - Wikipathways	7	2	1	0
	Apoptosis - Wikipathways	9	4	3	2
	Focal adhesion - Homo sapiens (human) - KEGG	20	8	7	2
	VEGF signaling pathway - Homo sapiens (human) - KEGG	3	1	2	0
Lipid-related Processes	SREBP signalling - Wikipathways	4	2	2	0
	ABC transporters in lipid homeostasis - Reactome	1			
Organelle synthesis, cell uptake	Lysosome - Homo sapiens (human) - KEGG	3	6	5	0
	Endocytosis - Homo sapiens (human) - KEGG	12			
	Mitochondrial biogenesis - Wikipathways	1			
	Regulation of autophagy - Homo sapiens (human) - KEGG	0			
eNOS pathways	eNOS activation and regulation - Reactome	1	3	1	0
	Angiogenesis - Wikipathways	5			
Cell Metabolism, Transcription, Translation	Glycolysis / Gluconeogenesis - Homo sapiens (human) - KEGG	0	0	1	0
	Citrate cycle (TCA cycle) - Homo sapiens (human) - KEGG	1			
	Oxidative phosphorylation - Homo sapiens (human) - KEGG	1			
	Generic Transcription Pathway - Reactome	5			
	Translation - Reactome	0			

References

Acevedo M, Vernier M, Mignacca L, Lessard F, Huot G, Moiseeva O, Bourdeau V, Ferbeyre G. (2016) A CDK4/6-Dependent Epigenetic Mechanism Protects Cancer Cells from PML-induced Senescence. Cancer research 76(11):3252-3264

Balmer JE & Blomhoff R (2002) Gene expression regulation by retinoic acid. Journal of lipid research 43(11):1773-1808.

Bhabak KP & Mugesh G (2010) Functional mimics of glutathione peroxidase: bioinspired synthetic antioxidants. Accounts of chemical research 43(11):1408-1419.

Birsoy K, Chen Z, & Friedman J (2008) Transcriptional regulation of adipogenesis by KLF4. Cell metabolism 7(4):339-347.

Calzone L, Gelay A, Zinovyev A, Radvanyi F, & Barillot E (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Molecular systems biology 4:173.

Chen Z, Peng IC, Cui X, Li YS, Chien S, Shyy JY. (2010) Shear stress, SIRT1, and vascular homeostasis. Proceedings of the National Academy of Sciences of the United States of America 107(22):10268-10273.

De Keulenaer GW, Chappell DC, Ishizaka N, Nerem RM, Alexander RW, Griendling KK (1998) Oscillatory and steady laminar shear stress differentially affect human endothelial redox state: role of a superoxide-producing NADH oxidase. Circulation research 82(10):1094-1101.

Denechaud PD, Lopez-Mejia IC, Giralt A, Lai Q, Blanchet E, Delacuisine B, Nicolay BN, Dyson NJ, Bonner C, Pattou F, Annicotte JS, Fajas L. (2016) E2F1 mediates sustained lipogenesis and contributes to hepatic steatosis. The Journal of clinical investigation 126(1):137-150.

Dinkova-Kostova AT & Talalay P (2010) NAD(P)H:quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector. Archives of biochemistry and biophysics 501(1):116-123.

Ernst J & Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology 28(8):817-825.

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. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43-49.

Ertosun MG, Hapil FZ, & Osman Nidai O (2016) E2F1 transcription factor and its impact on growth factor and cytokine signaling. Cytokine & growth factor reviews 31:17-25.

Evrard SM, Lecce L, Michelis KC, Nomura-Kitabayashi A, Pandey G, Purushothaman KR, d'Escamard V, Li JR, Hadri L, Fujitani K, Moreno PR, Benard L, Rimmele P, Cohain A, Mecham B, Randolph GJ, Nabel EG, Hajjar R, Fuster V, Boehm M, Kovacic JC (2016) Endothelial to mesenchymal transition is common in atherosclerotic lesions and is associated with plaque instability. Nature communications 7:11853.

Gal N, Pico A, Hanspers K, Sklar S, Kelder T, van Iersel M. (2016) TGF-beta Receptor Signaling (Homo sapiens) [Accessed: 12/04/2016]. Available from: http://www.wikipathways.org/index.php/Pathway:WP560.

Garcia-Cardena G, Comander J, Anderson KR, Blackman BR, & Gimbrone MA, Jr. (2001) Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. Proceedings of the National Academy of Sciences of the United States of America 98(8):4478-4485.

Germain P, Chambon P, Eichele G, Evans RM, Lazar MA, Leid M, De Lera AR, Lotan R, Mangelsdorf DJ, Gronemeyer H. (2006) International Union of Pharmacology. LX. Retinoic acid receptors. Pharmacological reviews 58(4):712-725.

Guo D, Chien S, & Shyy JY (2007) Regulation of endothelial cell cycle by laminar versus oscillatory flow: distinct modes of interactions of AMP-activated protein kinase and Akt pathways. Circulation research 100(4):564-571.

Hamik A, Lin Z, Kumar A, Balcells M, Sinha S, Katz J, Feinberg MW, Gerzsten RE, Edelman ER, Jain MK. (2007) Kruppel-like factor 4 regulates endothelial inflammation. The Journal of biological chemistry 282(18):13769-13779.

He M, Chen Z, Martin M, Zhang J, Sangwung P, Woo B, Tremoulet AH, Shimizu C, Jain MK, Burns JC, Shyy JY. (2017) miR-483 Targeting of CTGF Suppresses Endothelial-to-Mesenchymal Transition: Therapeutic Implications in Kawasaki Disease. Circulation research 120(2):354-365.

Henley SA & Dick FA (2012) The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle. Cell division 7(1):10.

Hsiai TK, Hwang J, Barr ML, Correa A, Hamilton R, Alavi M, Rouhanizadeh M, Cadenas E, Hazen SL. (2007) Hemodynamics influences vascular peroxynitrite formation: Implication for low-density lipoprotein apo-B-100 nitration. Free radical biology & medicine 42(4):519-529.

Hsieh HJ, Liu CA, Huang B, Tseng AH, & Wang DL (2014) Shear-induced endothelial mechanotransduction: the interplay between reactive oxygen species (ROS) and nitric oxide (NO) and the pathophysiological implications. Journal of biomedical science 21:3.

Hu J, Ge H, Newman M, & Liu K (2012) OSA: a fast and accurate alignment tool for RNA-Seq. Bioinformatics 28(14):1933-1934.

Jaiswal AK (2000) Regulation of genes encoding NAD(P)H:quinone oxidoreductases. Free radical biology & medicine 29(3-4):254-262.

Jakobsen JS, Waage J, Rapin N, Bisgaard HC, Larsen FS, Porse BT. (2013) Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. Genome research 23(4):592-603.

Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. (2011) ConsensusPathDB: toward a more complete picture of cell biology. Nucleic acids research 39(Database issue):D712-717.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, & Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic acids research 44(D1):D457-462.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, & Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic acids research 44(D1):D457-462.

Katayama K, Nakamura A, Sugimoto Y, Tsuruo T, & Fujita N (2008) FOXO transcription factor-dependent p15(INK4b) and p19(INK4d) expression. Oncogene 27(12):1677-1686.

Kim YW & Byzova TV (2014) Oxidative stress in angiogenesis and vascular disease. Blood 123(5):625-631.

Leopold JA, Zhang YY, Scribner AW, Stanton RC, & Loscalzo J (2003) Glucose-6-phosphate dehydrogenase overexpression decreases endothelial cell oxidant stress and increases bioavailable nitric oxide. Arteriosclerosis, thrombosis, and vascular biology 23(3):411-417.

Lin K, Hsu PP, Chen BP, Yuan S, Usami S, Shyy JY, Li YS, Chien S. (2000) Molecular mechanism of endothelial growth arrest by laminar shear stress. Proceedings of the National Academy of Sciences of the United States of America 97(17):9385-9389.

Liu J, Bi X, Chen T, Zhang Q, Wang SX, Chiu JJ, Liu GS, Zhang Y, Bu P, Jiang F (2015) Shear stress regulates endothelial cell autophagy via redox regulation and Sirt1 expression. Cell death & disease 6:e1827.

Lolli G & Johnson LN (2005) CAK-Cyclin-dependent Activating Kinase: a key kinase in cell cycle control and a target for drugs? Cell cycle 4(4):572-577.

Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15(12):550.

Maleszewska M, Vanchin B, Harmsen MC, & Krenning G (2016) The decrease in histone methyltransferase EZH2 in response to fluid shear stress alters endothelial gene expression and promotes quiescence. Angiogenesis 19(1):9-24.

Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic acids research 31(1):374-378.

McDonagh EM, Bautista JM, Youngster I, Altman RB, & Klein TE (2013) PharmGKB summary: methylene blue pathway. Pharmacogenetics and genomics 23(9):498-508.

Meo-Evoli N, Almacellas E, Massucci FA, Gentilella A, Ambrosio S, Kozma SC, Thomas G, Tauler A. (2015) V-ATPase: a master effector of E2F1-mediated lysosomal trafficking, mTORC1 activation and autophagy. Oncotarget 6(29):28057-28070.

Mowbray AL, Kang DH, Rhee SG, Kang SW, & Jo H (2008) Laminar shear stress up-regulates peroxiredoxins (PRX) in endothelial cells: PRX 1 as a mechanosensitive antioxidant. The Journal of biological chemistry 283(3):1622-1627.

Muthusamy S, Hong KU, Dassanayaka S, Hamid T, & Jones SP (2015) E2F1 Transcription Factor Regulates O-linked N-acetylglucosamine (O-GlcNAc) Transferase and O-GlcNAcase Expression. The Journal of biological chemistry 290(52):31013-31024.

Ohtani K, DeGregori J, & Nevins JR (1995) Regulation of the cyclin E gene by transcription factor E2F1. (Translated from eng) Proceedings of the National Academy of Sciences of the United States of America 92(26):12146-12150 (in eng).

Piechaczyk M & Farras R (2008) Regulation and function of JunB in cell proliferation. Biochemical Society transactions 36(Pt 5):864-867.

Reyes I, Hanspers K, Pico A, & Rashid K (2016) Wikipathways: Oxidative Stress (Homo sapiens) [Accessed: 12/04/2016]. Available from: http://www.wikipathways.org/index.php/Pathway:WP408.

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. (2013) ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic acids research 41(Database issue):D56-63.

Schulkens IA, Castricum KC, Weijers EM, Koolwijk P, Griffioen AW, Thijssen VL. (2014) Expression, regulation and function of human metallothioneins in endothelial cells. Journal of vascular research 51(3):231-238.

Singh KK, Lovren F, Pan Y, Quan A, Ramadan A, Matkar PN, Ehsan M, Sandhu P, Mantella LE, Gupta N, Teoh H, Parotto M, Tabuchi A, Kuebler WM, Al-Omran M, Finkel T, Verma S. (2015) The essential autophagy gene ATG7 modulates organ fibrosis via regulation of endothelial-to-mesenchymal transition. The Journal of biological chemistry 290(5):2547-2559.

Siu KT, Rosner MR, & Minella AC (2012) An integrated view of cyclin E function and regulation. Cell cycle 11(1):57-64.

Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, Bonierbale M, Brookes AJ, Bucci G, Buetti I, Burge S, Cabau C, Carlson JW, Chelala C, Chrysostomou C, Cittaro D, Collin O, Cordova R, Cutts RJ, Dassi E, Di Genova A, Djari A, Esposito A, Estrella H, Eyras E, Fernandez-Banet J, Forbes S, Free RC, Fujisawa T, Gadaleta E, Garcia-Manteiga JM, Goodstein D, Gray K, Guerra-Assunção JA, Haggarty B, Han DJ, Han BW, Harris T, Harshbarger J, Hastings RK, Hayes RD, Hoede C, Hu S, Hu ZL, Hutchins L, Kan Z, Kawaji H, Keliet A, Kerhornou A, Kim S, Kinsella R, Klopp C, Kong L, Lawson D, Lazarevic D, Lee JH, Letellier T, Li CY, Lio P, Liu CJ, Luo J, Maass A, Mariette J, Maurel T, Merella S, Mohamed AM, Moreews F, Nabihoudine I, Ndegwa N, Noirot C, Perez-Llamas C, Primig M, Quattrone A, Quesneville H, Rambaldi D, Reecy J, Riba M, Rosanoff S, Saddiq AA, Salas E, Sallou O, Shepherd R, Simon R, Sperling L, Spooner W, Staines DM, Steinbach D, Stone K, Stupka E, Teague JW, Dayem Ullah AZ, Wang J, Ware D, Wong-Erasmus M, Youens-Clark K, Zadissa A, Zhang SJ, Kasprzyk A. (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic acids research 43(W1):W589-598.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. (2005) Gene set enrichment analysis: a

knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102(43):15545-15550.

Sur S & Agrawal DK (2016) Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. Molecular and cellular biochemistry 416(1-2):33-46.

Szmitko PE, Wang CH, Weisel RD, de Almeida JR, Anderson TJ, Verma S. (2003) New markers of inflammation and endothelial cell activation: Part I. Circulation 108(16):1917-1923.

Tallack MR, Keys JR, Humbert PO, & Perkins AC (2009) EKLF/KLF1 controls cell cycle entry via direct regulation of E2f2. The Journal of biological chemistry 284(31):20966-20974.

Tornatore L, Thotakura AK, Bennett J, Moretti M, & Franzoso G (2012) The nuclear factor kappa B signaling pathway: integrating metabolism with inflammation. Trends in cell biology 22(11):557-566.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, & Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nature reviews. Genetics 10(4):252-263.

Yamaguchi T, Cubizolles F, Zhang Y, Reichert N, Kohler H, Seiser C, Matthias P. (2010) Histone deacetylases 1 and 2 act in concert to promote the G1-to-S progression. Genes & development 24(5):455-469.

Zhou J, Li YS, Wang KC, & Chien S (2014) Epigenetic Mechanism in Regulation of Endothelial Function by Disturbed Flow: Induction of DNA Hypermethylation by DNMT1. Cellular and molecular bioengineering 7(2):218-224.

Chapter 3: Novel Mechanisms of Endothelial Response to Shear Derived From "Omics" Data Analysis

Introduction

In chapter 2, we emphasized a "systems biology" view of endothelial response to shear, creating a temporal map of functional endothelial response. In chapter 3, we sought the detection of novel patterns in gene expression in the time-series RNA-seq data in order to build on existing knowledge of endothelial response to shear. Where applicable, the networks and temporal maps developed in chapter 2 will serve as a framework for understanding the context of these discoveries. Herein we describe several putative mechanisms of endothelial response to shear, inferred from analysis of our data sets. I especially discuss LINC00520, which will also be referred to as LEENE (**l**ncRNA that **e**nhances **eN**OS **e**xpression). I computationally detected LINC00520 and proposed LINC00520's role as an enhancer RNA, providing putative regulatory targets for the LINC00520 enhancer region, particularly NOS3. Our collaborators validated LINC00520 and further developed LINC00520's proposed mechanism of action as a long non-coding RNA that enhances eNOS expression.

Methods – "Omics" Data Generation and Processing"

Cell culture, shear stress experiments, and RNA-seq data processing are identical to that in chapter 2. All wet lab experiments were conducted by our collaborating labs. The ten point RNA-seq time-series data set was analyzed for distinctly expressed genes, as defined by discrete categories of expression profiles. Given that there are three possible pair-wise comparisons (OS vs PS, OS vs ST, and PS vs ST), and given that there are three possible designations for gene regulation (upregulated, downregulated, or not differentially expressed), there are 27 possible descriptions for the expression profile of a given gene at a given time point (Figure 3.1). Genes were defined as "PS-distinct" if they were differentially upregulated in PS vs ST, and differentially downregulated in both OS vs ST and OS vs PS in at least one time point. Genes were defined as "OS-distinct" if they were differentially downregulated in PS vs ST, and differentially upregulated in both OS vs ST and OS vs PS in at least one time point.

This method will be referred to as “point-wise state analysis”. In order to minimize the occurrence of false positives, I would add an extra layer of filtration, wherein I would check to see if PS-distinct genes were also upregulated in public PS vs ST shear data, collected at 72 hours [Maleszewska et al, 2016]. I would also perform this check to see if OS-distinct genes were also downregulated in this public PS vs ST shear data.

A similar analysis, which will be referred to as “dynamical state analysis”, was conducted with comparisons across time points. Consider, for example, two time points for three possible pair-wise comparisons. Then one can consider the 27 possible descriptions for the expression profile of a given gene at time point 1 (as described in the previous paragraph), the 27 possible descriptions for the expression profile of a given gene at time point 2, and the 27 possible comparisons of the expression profile for a given condition between time point 2 and time point 1 (e.g. OS hour 2 vs OS hour 1). This allows for the identification of genes that not only exhibit dramatic differential expression across conditions, but also exhibit monotonic expression behavior across time. Because the RNA-seq data in this analysis has ten time points, the space of possible descriptions for the expression profile based on differential expression (up, down, or not differentially expressed) is very large. A simple indexing system was developed that allows the user to define the expression profile of interest, and searches for the specified expression profile among the analyzed set of genes using regular expressions.

4C and histone ChIP-seq data sets used identical cell culture and shear stress experimental design as RNA-seq data sets, described in chapter 2. The construction of 4C libraries was performed following previously published protocol [Splinter et al, 2012]. Briefly, ECs were crosslinked with 2% formaldehyde, which was quenched with 0.1 M glycine. The cross-linked DNA underwent two rounds of digestion respectively by DpnII and CviqI recognizing 4 bp restriction sites. Each digestion was followed by a reaction with T4 DNA

ligase for proximity ligation. The resulting 4C template was used for the subsequent PCR reactions, of which 16 were pooled and purified for next-generation sequencing. The 4C sequencing reads were tested for the quality and aligned to human reference genome version hg19 by Bowtie 2 [Langmead et al, 2012]. Read numbers in given genomic location were counted by BEDTools [Quinlan et al, 2010] and normalized by the total mapped reads per sample.

Pre-processed 5-kb resolution inter-chromosomal hi-C matrices for HUVEC were accessed and downloaded online (GEO accession: GSE63525) [Rao et al, 2014; Sanborn et al, 2015]. The LINC00520 genomic region was defined as chr14:56240000-56290000 in order to span the LINC00520 gene, the detected 4C signals, and the local H3K27ac marks. All genomic regions with non-zero signals in the inter-chromosomal matrices that were associated with the defined LINC00520 genomic region were located and annotated by gene (or closest gene if the region with detected signal was intergenic). These LINC00520-associated genes were further filtered depending on whether they were detected to be differentially expressed in OS vs PS at any time point.

KLF2 and KLF4 binding sites were predicted by using regular expression matches in R, based on the motifs from TRANSFAC database (Version 2015.4) [Matys et al, 2003]. Genomic regions were downloaded from the UCSC genome browser. KLF2 and KLF4 binding sites were predicted using DNA sequences from the LINC00520 genomic region spanning -20 kb to +5 kb of its TSS in hg19 and from the BY707159 mouse genomic region (chr14: 47,786,094-47,815,319, mm10).

Methods – LINC00520 Validation Experiments

All validation experiments described in this section, along with the accompanying analysis of the data from those experiments, were conducted by our collaborating labs, particularly Dr. Zhen Chen and Dr. Yifei Miao.

Absolute quantification of lncRNA copy number. The RNA copy number was performed following previously described protocol [Tripathi et al, 2010]. In brief, in vitro transcribed LEENE RNA fragments were used as standard and subjected to reverse transcription. The RT product were serially diluted to generate calibration curve. Total RNA from known quantity of ECs were subjected to the same RT procedure. The copy number of LEENE RNA from resulting RT products were then determined by qPCR.

Subcellular Fractionation and RNA Isolation. Subcellular fractionation was performed following published protocol [Bhatt et al, 2012] with minor modification. Briefly, HUVECs from three confluent 150 mm culture dishes were applied as independent triplicates. The cells were collected in 200 μL cold cytoplasmic lysis buffer (0.15% NP-40, 10 mM Tris pH 7.5, 150 mM NaCl) and incubated on ice for 5 min. The lysate was layered onto 500 μL cold sucrose buffer (10 mM Tris pH 7.5, 150 mM NaCl, 24% sucrose w/v) and centrifuged. The supernatant containing cytoplasmic component was quickly added to TRizol LS for RNA extraction. The nuclear pellet was gently suspended into 200 μL cold glycerol buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT). An addition of cold nuclei lysis buffer (20 mM HEPES pH 7.6, 7.5 mM MgCl2, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, 1% NP-40, 1 mM DTT) was added, followed by vortex and centrifuge. The supernatant containing nucleoplasmic fraction was mixed with TRizol LS for RNA extraction. 50 μL of cold PBS was added the remaining pellet and gently pipetted. After vigorous vortex to resuspend the chromatin, chromatin-associated RNA was extracted by adding 100 μL chloroform and TRizol reagent. RNA samples from three different fractions were dissolved with same amount

of RNase-free water and same volume of RNA was used for reverse-transcript and quantitative PCR.

CRIPSR-Cas9 gene editing. Our collaborators designed multiple single-guide RNAs (sgRNA) to target the genomic region of LEENE. The sequences of sgRNAs are listed in Table 3.1. The designed sgRNAs were sub-cloned into the CAS9-T2A-GFP-expression vector (Addgene: pX458) using designed BbsI cloning site. All sgRNAs were tested with its cutting efficiency in HEK293 cells using the Surveyor Mutation Detection Kit from IDT (Figure 3.2).

Cell transfection. Two Antisense LNA™ GapmeRs specifically targeting two different regions of LEENE (NR_026797) were designed and purchased from Exiqon (Table 3.2). siRNA with scrambled or KLF2 targeting sequence were designed and purchased from Qiagen (SI03650318 and SI04275110). LNAs or siRNA were separately transfected into ECs with Lipofectamin RNAiMAX following the protocol provided by the manufacturer. ECs were cultured for another 48 hr after transfection before further analysis. Transfection of ECs with GFP-Cas9 with or without sgRNAs were performed with Cytofect™ HUVEC transfection kit (Cell Applications). Respective vectors (2 μg) were transfected per well of 6-well plates as the cells reached 80% confluency. After one-hour incubation with transfection mixture, antibiotics-free growth medium was added for another 48 hr culture, before the cells were harvested.

Monocyte adhesion assay. Monocytes adhesion was performed as previously described [Huang et al, 2017]. THP1 monocytes (ATCC) cells were labeled with CellTracker™ Green CMFDA Dye (Thermo Fisher #C2925) and seeded on endothelial cell monolayers for 15 minute. After two washes with M199 complete medium, attached THP-1 cell numbers were calculated from five randomly selected views captured by fluorescence microscopy.

Chromatin isolation by RNA purification (ChIRP), RNA-immunoprecipitation (RNA-IP), and chromatin immunoprecipitation (ChIP). ChIRP was performed following the protocols as described in previous studies [Chu et al, 2011; Chu et al, 2012; Leveille et al,

2015]. Biotin-labeled anti-sense oligo probes were designed and purchased from Biosearch Technologies (Table 3.3) following several criteria: 1) number of probes = 1 probe/100 bp of RNA length; 2) target GC% = 45; 3) oligonucleotide length = 20 bp; 4) spacing length = 60-80 bp. The 'even' and 'odd' pools of probes were diluted into 100 μM concentration. After 24 hr treatment with 1 μM ATV or DMSO, 1×107 HUVECs were fixed with 1% glutaraldehyde for 10 min at room temperature. The pelleted cells were lysed and sonicated for 10 min using the '30 seconds ON, 30 seconds OFF' program. The sonicated samples were then centrifuged and 1/100 of supernatant was taken as RNA input and DNA input respectively. 100 pmol probes were hybridized with supernatant at 37°C for 4 hr. Afterwards, washed Streptavidin-conjugated magnetic beads were mixed with the reaction for another 30 min. Following several rounds of washing, beads were resuspended with 1 ml wash buffer and 100 μL mixture was taken for RNA isolation using TRIzol. The rest of the ChIRP precipitates underwent DNA isolation. qPCR analysis was performed to assess the RNA retrieval rate using β-actin as negative control and the LEENE-associated DNA sequences.

RIP was performed as previously described [Chen et al, 2013]. In general, after 24 hr treatment with 1 μM Statin or DMSO, 1×107 HUVECs were cross-linked by UV irritation and pelleted. Whole cells were lysed with 500 μL lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA, and 100 units/ml RNAse inhibitor) and incubated overnight at 4°C with 50 μL of Protein G dynabeads that were pre-washed and pre-mixed with antibodies or non-specific IgG control. Antibodies used for RIP assays include anti-RNA polymerase II (ab817, Abcam), anti-KLF4 (12173, Cell Signaling Technology) and anti-MED1 (A300-793A, Bethyl Laboratories). All of the antibodies have been previously authenticated for ChIP use [Eid et al, 2015; Riz et al, 2015; Pelish et al, 2015]. Following three times of wash to remove non-specific binding, RNA was extracted for qPCR analysis.

ChIP assays were performed as previously described [Chen et al, 2015] using the same antibodies as RIP. Briefly, HUVECs were treated with 0.75% formaldehyde for 20 minutes at room temperature. After sonication, the chromatin was immunoprecipitated by various antibodies conjugated to protein A or protein Dynabeads. Protein and RNA were degraded by proteinase K and RNase A, respectively. The purified chromatin DNA was then used as the template for quantitative polymerase chain reaction. As an isotype control, rabbit IgG was used in ChIP.

DNA fluorescence in situ hybridization (FISH). In-house probes detecting eNOS and LEENE genomic regions were generated from bacterial artificial chromosome (BAC) probes (Source BioScience LifeSciences). The clone IDs are eNOS, RP11-910F16 (length 183744 bp) and LEENE, RP11-105H21 (length 183093 bp). BAC probes were labeled by FISH tag DNA kit (Invitrogen). DNA FISH was performed following previously described protocols [Fanucchi et al, 2013; Bolland et al, 2013]. Briefly, HUVECs were seeded on the coverslides and fixed directly with 4% formaldehyde and permeabilized with 0.1% saponin/0.1% Triton X-100 in PBS for 10 min at room temperature. Cells were then equilibrated in 50% formamide/2x SSC for 10 min at room temperature and denatured for 3 min at 78^{o}C. Afterwards, cells were hybridized overnight in a humidified chamber at 37^{o}C in 10 μL Hyb buffer (40% dextran sulfate plus 8×SSC) combined with 30 ng DNA FISH probes that have been freshly denatured at 78^{o}C for 5 min and cooled on ice. On the second day, the slides were washed three times with wash buffer (0.1% Tween plus 4×SSC). Cells were counter stained with DAPI, mounted with prolong buffer and imaged with Zeiss Apotome. The two probes were considered as proximally associated when the signals were completely overlapped or the distance between the centers of the signals < 1 μm. Up to 20 pictures were randomly taken from each sample and three researchers were assigned to independently and blindly quantify the percentage of the cells showing proximity association.

NO bioavailability assay. The NO production from HUVECs was detected as the accumulation of nitrate/nitrite by using a Nitrate/Nitrite fluorometric assay Kit (Cayman Chemical) as previously described [Chen et al, 2015]. Briefly, the phenol-red free M199 medium used to culture ECs was collected and centrifuged. The fresh supernatant was used for NO assay. Nitrate was first converted to nitrite utilizing nitrate reductase, followed by DAN addition to form fluorescent product. The fluorescent signal was read using TECAN Infinite 200 pro (TECAN) under 360 nm excitation wavelength and 430 nm emission wavelength. The NO content was calculated based on the nitrate standard curve.

Nascent RNA capture. Newly synthesized mRNA species were isolated using Click-iT Nascent RNA Capture Kit (C10365, Invitrogen) according to manufacturer's protocol. Briefly, HUVECs were synchronized with 2% FBS in M199 medium for 8 hr, followed by incubation in 0.2mM of 5-ethymyluridine (EU, an alkyne-modified uridine analog, which is incorporated into the nascent RNA) for another 24 hr and total RNA was isolated using TRIzol reagent. A copper catalyzed click reaction was performed using 5 μg RNA with 0.5mM azide-modified biotin. The mixture was incubated at room temperature for 30 min following RNA precipitation. Biotin-label EU-RNA was then pulled down by mixing with Streptavidin T1 magnetic beads at room temperature for 30 min and the unbound RNA was washed away. The cDNA synthesis was performed directly on the beads using Superscript VILO cDNA synthesis kit (Invitrogen), followed by qPCR analysis.

Quantitative PCR (qPCR). Reverse-transcription of RNA into cDNA was performed with PrimeScript™ RT Master Mix containing both Oligo dT primer and random 6mers primer (Takara Bio Inc.). KAPA SYBR FAST ROX Low supermix was used for qPCR following manufacturer's suggested protocol. All the primer sequences used were listed in Table 3.4 and Table 3.5.

Western Blot analysis. Western blot was performed using antibodies against eNOS (9572S, Cell Signaling Technology) and β-actin (8457S, Cell Signaling Technology) following standard protocol.

LEENE homology analysis. The sequence similarity between the predominant transcript of LEENE (NR_026797.1) and BY707159.1 was calculated by using the EMBOSS Water tool, which was designed based on Smith-Waterman algorithm, with default parameters [Rice et al, 2000].

Animal studies. Animal study protocol was approved by Institutional Animal Care and Use Committee of City of Hope, Duarte. Five 8-12 weeks C57BL male mice were randomly chosen and euthanized and thoracic aorta (TA) and aortic arch (AA) were isolated in PBS to tease out perivascular adventitia. Cleaned vessels were immediately snap-frozen in liquid nitrogen following RNA extraction with TRIzol reagent. Based on our collaborators' previous experience, sample size was determined to have enough power to detect an estimated statistical difference between two groups. With a sample size of 5 in each group, this study can provide 80% powder to detect an effect size of 2 with 0.05 significant level using two-sided t-test between two groups. Our collaborators did not expect large variation between two groups since the chosen animals are identical or similar regarding their age, gender, and background and raised under same condition. In the given case, no blinding was needed.

Mouse lung endothelial cells isolation was performed following the protocol as previously described [Chen et al, 2015] with modifications. For each experiment, the lungs from six C57BL male mice at the age of ~6 weeks were pooled, collagen-digested, and CD31-Sorted.

Statistical analysis of low-throughput experimental results. First, the statistical distribution of each group was confirmed to be normal by using a χ^2 test. Statistical analysis was then performed using Student's t-test (two-sided) between two groups or ANOVA followed

by Bonferroni post-test for multiple group comparisons. If variances between two groups were significantly different (F-test), nonparametric Mann-Whitney test was applied. $p<0.05$ was considered as statistically significant. Power analyses for the animal study have been described in the previous section. At least three replicates were performed for all validation experiments unless specified.

Results – *in silico* Analysis

I used dynamical state analysis to interrogate the RNA-seq data for expression patterns of interest, including "consistent upregulation between conditions across time" and "nondecreasing fold changes in a condition across time". Among 22300 annotated genes, only 4 were identified as having expression profiles that were upregulated in PS vs ST across all time points, but NOT upregulated (i.e. downregulated or not differentially expressed) in OS vs ST in all time points. These four genes were KLF2, KLF4, CNR1, and DHH. Zero genes were identified as having the converse expression profile (i.e. upregulated in OS vs ST across all time points but NOT upregulated in PS vs ST in all time points).

Using point-wise state analysis, 97 genes were identified as "PS-distinct" and 14 genes were identified as "OS-distinct" (Table 3.6). Three of the four identified genes identified in the dynamical state analysis were also identified among the PS-distinct genes (KLF2, KLF4, DHH). A portion of these OS-distinct and PS-distinct genes were manually curated and assessed for how (a) how well-studied the gene is in endothelial cells, (b) how well-studied the gene is in shear conditions, (c) if a hypothesis can be proposed about the relevance of the gene to endothelial cells under shear, how clear is the hypothesis and how consistent the hypothesis is with existing literature on the gene.

Several genes known to be relevant to endothelial response to shear were identified using point-wise state analysis, including KLF2, KLF4, NOS3, THBD [Rochfort et al, 2015], and VEGFA [dela Paz et al, 2012]. These genes were considered to already be well-

characterized and therefore were not considered for further study. Another cohort of genes were found to be well-studied in endothelial cells, but not specifically under conditions of shear. These genes include ABLIM3, ADAMTS4, APOLD1, ATG9B, CASKIN2, GCH1, GMFG, GPER, GRK5, and HEG1. These genes were found to have some studies that already describe their functional role, and would be ideal candidates for a future study on novel mechanosensitive genes. ATG9B, for example, has known relevance to autophagy [Zavodszky et al, 2013] and neighbors NOS3 on the genome, making it a promising candidate for a small and focused study on its role in shear stress.

Novel Hypotheses of Mechanisms Revealed from Distinctly Expressed Genes. Some of these identified genes were not well-studied in endothelial cells or in shear, but were sufficiently well-studied in other contexts, allowing for the formulation of a shear-specific mechanistic hypothesis. One such PS-distinct gene is DHH, which is involved in the hedgehog signaling pathway and activates the GLI family of transcription factors [Falkenstein et al, 2014]. DHH has previously been shown to be mechanosensitive [Ni et al, 2010], and the TF-to-gene network described in chapter 2 (Figure 2.9) indicates that GLI1 has a role in the PS-specific network together with several differentially expressed target genes. Analysis of the expression profiles of these target genes reveals that STMN3, a stathmin protein involved in the function of microtubules, has an expression profile is strongly downregulated in OS vs PS similar to DHH's expression profile. Therefore, we propose that PS can modulate cell morphology directly via the hedgehog signaling pathway through the production of DHH (Figure 3.3).

Another such promising PS-distinct gene is C1orf21. A previous proteomics study [Rolland et al, 2014] implicated C1orf21 as interacting with DRAM1, a gene that regulates autophagy via the lysosome [Zhang et al, 2013]. Although DRAM1 is not differentially expressed, a future study of C1orf21 and DRAM1 on the protein level could reveal a mechanosensitive mechanism by which endothelial autophagy is modulated via C1orf21

expression. Other PS-distinct genes of potential interest that are not well-studied in shear conditions are GCKR (also known as GKRP), an inhibitor of the metabolic protein glucokinase [de la Iglesia et al, 1999], and DUSP4, a phosphatase responsible for activating p38 [Barajas-Espinosa et al, 2015].

Other genes offer a more complex picture of endothelial regulation, being understudied in endothelial cells and shear, but also having several studies in other contexts that present contradictory or unclear conclusions. CXCR4, which is an OS-distinct gene, is known to be directly inhibited by KLF4 [Li et al, 2015], and interacts with the ligand CXCL12. CXCR7, on the other hand, is a PS-distinct gene that shares the same ligand. The CXCR7-CXCR4-CXCL12 axis of regulation is well-studied in cancer and is involved in a variety of cellular functions such as proliferation and migration [Pozzobon et al, 2016; Duda et al, 2011]. The role of this regulatory axis has not been studied in ECs under shear, despite the distinct expression and mechanosensitivity of these chemokine receptors. A future study that explores the dynamics of these proteins in shear is likely to produce novel insights into mechanisms of atherosclerosis.

Another such set of genes are HYAL1 and HYAL2, lysosomal acid hydrolases that were both identified as PS-distinct. HYAL2 in particular was identified via dynamical state analysis as a gene whose expression profile is nondecreasing in PS vs ST, suggesting a long term ramping in upregulation of HYAL2 in PS over time. HYAL proteins are involved in the degradation of hyaluronan (HA), a glycosaminoglycan that is part of the extracellular matrix and can form high molecular weight polymers [Triggs-Raine et al, 2015; Chowdhury et al, 2016]. HYAL2 proteins are localized to the cellular surface in order to target high-molecular weight HA, and HYAL1 proteins are localized to the lysosome in order to target low-molecular weight HA [Triggs-Raine et al, 2015]. Previous studies reveal that HA levels may modulate inflammation, eNOS activity, proliferation, and angiogenesis [Genasetti et al, 2008; Kong et al,

2016], but the direction of regulation is not clear. The role of HA in the extracellular matrix proximal to the endothelium, along with the relationship between HA and shear, offer questions that would make for promising exploratory studies.

In silico Identification and Characterization of LINC00520. Another PS-distinct gene of interest was LINC00520 (aka C14orf34). This was the only gene detected through point-wise state analysis that encoded a long non-coding RNA (lncRNA). Long non-coding RNAs are a large class of non-coding RNAs that are >200 bp in length (as opposed to, for example, microRNA). Over 27,000 lncRNAs have been predicted/annotated in the human genome10, but relatively little is known about their biological function, and classification of an RNA molecule as a lincRNA can be ambiguous due to a generic lack of functional characterization [St Laurent et al, 2015].

In order to develop a more sophisticated hypothesis on the function of LINC00520, I then examined the LINC00520 genomic region in the 4C-NOS3 and H3K27me3, H3K27ac, and H3K4me1 ChIP-seq data sets. These data sets, generated under OS and under PS with two replicates apiece, show distinctive shear-specific patterns in the LINC00520 region. Figure 3.4 offers a visualization of this data taken together. The preponderance of H3K27ac and H3K4me peaks detected in the LINC00520 region suggest that this is an active enhancer region [Rivera et al, 2013], making LINC00520 a potential enhancer RNA. The detected 4C peaks suggest that this putative enhancer region may interact with the NOS3 gene (aka eNOS), an important factor in maintaining vascular homeostasis that also appeared among the PS-distinct genes [Tousoulis et al, 2012]. Strikingly, the 4C peaks are only present under PS, and the acetylation peaks are higher under PS as well compared to OS.

In order to further clarify the unique standing of LINC00520 expression as a lncRNA, I isolated all noncoding RNA (any transcript with a refseq ID that began with “NR”) and exclusively analyzed all exclusively ncRNA-producing genes using the pipeline illustrated in

Figure 3.5A. Among the 2054 lncRNAs identified in the RNA-seq, I first filtered for those differentially regulated by PS vs. OS at hour 24. These flow-regulated lncRNAs are listed and ranked based on their differential expression in the heatmap in Figure 3.5B. Because of the observation that NOS3 and LINC00520 may interact, I then filtered for lncRNAs that were positively correlated with NOS3 over the entire time course (correlation coefficient >0.8). The RNA level of LINC00520 emerged as the top-ranked candidate and was highly correlated with NOS3 expression levels (correlation coefficient 0.85, Figure 3.5C). The temporal course of flow-regulated NOS3 and LINC00520 RNA levels showed similar patterns, i.e. a sustained and robust induction by PS. Figure 3.5D compares the expression profiles of NOS3 and LINC00520, along with KLF2, KLF4, and the pro-inflammatory vascular cell adhesion molecule 1 (VCAM1). Notably, mRNAs encoding KLF2 and KLF4, which are key TFs of NOS3 [Lin et al, 2005; Jiang et al, 2014] and are also PS-distinct genes, were significantly induced by PS as early as hour 1, reached their highest levels at hours 4 and 6, and remained induced at hour 24.

When examining our KLF2 overexpression RNA-seq data, it was also found that LINC00520 is overexpressed when KLF2 is overexpressed. Because KLF2 is mechanosensitive and is upregulated in PS, this further reinforces the hypothesis that LINC00520 is mechanosensitive and PS-distinct, while also suggesting that its mechanism of expression may involve KLF2. The timing of differential expression is also consistent with this hypothesis – KLF2 is statistically significantly differentially expressed ($p < 0.05$) from hour 1 onward, while LINC00520 slowly increases in PS upregulation over time, achieving statistically significant differential expression in hour 6 and onward. Finally, KLF2 can transcriptionally regulate NOS3 [Lin et al, 2005], offering yet another potential regulatory link between LINC00520 and NOS3.

In order to further understand the role of the LINC00520 genomic region as an enhancer, Hi-C HUVEC data (GEO ID: GSE63525) was mined and searched for interactions with the LINC00520 genomic region [Rao et al, 2014; Sanborn et al, 2015]. A total of 2794 genes were found to overlap with regions interacting with the LINC00520 genomic region. Of these genes, 1177 were differentially expressed in OS vs PS. Figure 3.6 showcases the enriched pathways for these 1177 genes. Pathways pertaining to hemostasis (i.e. wound repair), focal adhesion, and VEGFA signaling feature prominently among the list of enriched pathways. The enrichment of VEGFA pathways is particularly interesting, as VEGFA is among the PS-distinct genes in table 2 alongside LINC00520.

While this computational approach was sufficient to propose hypotheses about the genomic region of LINC00520, further work was required for characterizing the role of the LINC00520 molecule itself. Depending on their subcellular localization (i.e. in the nucleus or cytoplasm), lncRNAs can regulate gene expression through diverse mechanisms. A group of lncRNAs has previously been identified as nucleus-retained and chromatin-associated [Wang et al, 2016; Viereck et al, 2016; Tripathi et al, 2010; Michalik et al, 2014]; they can serve as scaffolds or guides in cis or in trans to recruit TFs, transcriptional co-activators, or chromatin remodelers, and/or to promote long-range DNA (e.g. promoter-enhancer) interaction, thus resulting in transcriptional activation [Trimarchi et al, 2014; Wang et al, 2011; Hacisuleyman et al, 2014]. For example, the lncRNA Firre has been shown to be localized around its site of transcription in X-chromosome in the embryonic stem cells and mediate trans-chromosomal interaction [Hacisuleyman et al, 2014]. LncRNAs can also be classified depending on their encoded genomic locations (i.e. intragenic, intergenic, or enhancer regions) and the associated histone modifications [St Laurent et al, 2015]. A new class of lncRNAs have emerged as lnc-eRNA or elncRNA, which are encoded in enhancer regions marked by Histone

3 lysine 4 monomethylation (H3K4me1) and Histone 3 lysine 27 acetylation (H3K27ac) [Li et al, 2016; Devaux et al, 2015].

The regulatory role of lnc-eRNAs, especially those in the vascular ECs, had not been explored. Furthermore, literature on LINC00520 was scant, with fewer than 20 papers mentioning the gene (even in passing as part of a broader GWAS data analysis) as of December 2016. In that same period of time, no papers existed that connected LINC00520 and shear stress in any way. The putative connection with NOS3 in particular was very promising. Endothelial nitric oxide synthase (eNOS, also known as NOS3), which is central to endothelial homeostasis and vascular function, is regulated at multiple levels [Forstermann et al, 2012], including post-translational modifications (such as phosphorylation and acetylation) [Fulton et al, 1999; Chen et al, 2010] and transcriptional regulation by transcription factors (TFs) [Shaul et al, 2002]. It has been established that eNOS transcription is largely regulated by Krüppel-like factors 2 (KLF2) and 4 (KLF4), two key TFs in endothelial identity and vascular homeostasis [Sangwung et al, 2017]. The expression and activity of KLF2 and KLF4 can be altered by a number of mechanical (e.g. hemodynamic flow), biochemical (e.g. pro-inflammatory stress), and pharmacological stimuli (e.g. cardiovascular protective drugs), leading to differential transcriptional regulation of eNOS as well as other genes important in endothelial biology [SenBanerjee et al, 2004; Zhou et al, 2012]. There is also evidence that eNOS expression can be regulated through histone modifications [Fish et al, 2005; Gan et al, 2005]. However, whether and how long-range DNA interaction coordinates with TF binding and histone modification to modulate eNOS transcription in endothelial cells (ECs) remains essentially unknown. LINC00520 appeared to be a putative mechanosensitive lncRNA, associated with an enhancer region, relevant to eNOS expression. Therefore LINC00520, a putative "**l**ncRNA that **e**nhances **eN**OS **e**xpression", will be referred interchangeably as "LEENE" and as its official name LINC00520.

Results – Validation of LINC00520 (LEENE) as a mechanosensitive lnc-eRNA relevant to endothelial homeostasis.

All validation experiments were conducted by our collaborating labs, particularly Dr. Zhen Chen and Dr. Yifei Miao.

LEENE is known to have two transcripts. The predominantly expressed transcript in ECs contains exons 1, 3, and 4, while the less abundant transcript in ECs contains all four exons. We referred to FANTOM5 [Lizio et al, 2015] and confirmed that neither LEENE transcripts have any coding potential. Using absolute quantification assay with qPCR, we determined the copy number of LEENE in untreated HUVECs to be ~10 copies and the copy number of LEENE in HUVECs exposed to PS to be ~40 copies. Both LEENE transcripts were upregulated in PS compared to OS (Figure 3.5E). As shown in Figure 3.5F, qPCR with LEENE RNA-specific primers also revealed the significantly higher level of LEENE in ECs subjected to PS compared to OS. To confirm the flow regulation of LEENE in ECs and explore its relevance to endothelial function, we tested whether LEENE is differentially regulated by tumor necrosis factor alpha (TNFα, which exerts pro-inflammatory effects similar to OS, and atorvastatin (ATV), which confers endothelial protective effects similar to PS. Resembling the opposite effects of OS and PS, TNFα decreased, while ATV increased the level of LEENE. These findings are in line with the differential levels of KLF2/KLF4-eNOS signaling (Figure 3.5G).

Because KLF2/KLF4 can transactivate eNOS through TF binding sites (TFBS) in the eNOS promoter regions [Atkins et al, 2007], we next searched for TFBS in the DNA region in and near LEENE locus. As illustrated in Figure 3.7A, the region spanning -20 kb ~ +5 kb of LEENE TSS contains multiple TFBS for KLF2 and KLF4. We subsequently overexpressed KLF2 and KLF4 in ECs to experimentally verify whether these key TFs can upregulate LEENE. Indeed, we found increased levels of LEENE by the overexpression of KLF2 or KLF4 in ECs, with eNOS as a positive control (Figure 3.7B and C). To confirm the association of such TFs on the promoter of LEENE, we have performed ChIP-qPCR, which detected a robust binding

between KLF4 and multiple regions within the promoter region of LEENE (marked by H3K4me3 peaks, Figure 3.7A); these interactions were significantly increased by Ad-KLF4, which mimics the effect of PS and ATV (Figure 3.7D). In contrast, when we knocked down KLF2 in ECs, the PS-regulated LEENE was substantially decreased (Figure 3.7E). The combined computational analysis and experimental results in Figure 3.5 and Figure 3.7 suggest that LEENE is 1) co-regulated with eNOS downstream of KLF2 and KLF4, and 2) induced in conditions that promote endothelial homeostasis but suppressed by stimuli that impair endothelial function.

LEENE RNA is nucleus-localized and its genomic locus is in proximity association with eNOS promoter. To gain insights into the biological function of LEENE, we first determined its subcellular localization. As shown in Figure 3.8A, LEENE RNA transcripts were predominantly detected in the nucleus of ECs, i.e., as chromatin-associated and nucleoplasm-localized, with only a minor fraction in the cytoplasm, suggesting that its biological function is mainly in the nucleus. To this end, we used MALAT1 as a positive control, which has been previously identified as nucleus-enriched lncRNA [Tripathi et al, 2010]. We also used DANCR as the cytoplasm-enriched lncRNA control and CasC7 and TUG1 as the controls for lncRNAs localized in both nucleus and cytoplasm [van Heesch et al, 2014; Long et al, 2016] (Figure 3.9). We then assessed the genomic features and neighboring genes of LEENE. As recently described, LEENE is located 110 kb downstream of KTN1 and 321 kb upstream of PELI2 [Henry et al, 2016]. Unlike LEENE, the mRNA levels of neither KTN1 nor PELI2 were differentially regulated by PS or OS in ECs (Figure 3.10A). These results suggest that LEENE is transcribed independently from its neighboring genes. The strong enhancer marks (i.e. H3K27ac and H3K4me1 peaks) surrounding LEENE including the 5', gene body, and 3' regions (Figure 3.8B) suggest that LEENE genomic locus may act as a distal enhancer to mediate transcriptional activation in ECs. We did find that PS led to significant increase in

the H3K27ac in the LEENE region as measured by ChIP-qPCR, indicating the activation of LEENE as an enhancer in ECs subjected to PS vs. OS (Figure 3.8C).

Among the 1177 genes that were identified in the Hi-C data analysis and were differentially expressed in OS vs PS RNA-seq data, 81 of these genes were found to be highly correlated with LEENE (Figure 3.11), and eNOS was among the top hit with the highest correlation (Figure 3.11). The inter-chromosomal interaction between eNOS (chr7: 150,700,000-150,705,000) and LEENE (chr14: 56,280,000-56,285,000) in HUVEC is illustrated in Figure 3.8E. Of note, such interaction is absent in human epithelial and HeLa cells (Figure 3.12), which do not express detectable level of endogenous eNOS.

To confirm the proximal association between eNOS and LEENE, we performed DNA FISH, which has been commonly used to validate the chromosomal association revealed by chromatin conformation capture-based methods [Giorgetti et al, 2016]. Indeed, we observed the localization or juxtaposition of eNOS and LEENE probes in 8-10% of ECs under PS, the physiological flow condition (Figure 3.8F). To further confirm and quantitatively compare the LEENE-eNOS inter-chromosomal interaction in ECs under different flow conditions, we performed high-resolution 4C-seq in HUVECs subjected to PS and OS using the H3K27ac- and H3K4me1-enriched peak region in the eNOS promoter as the bait (Figure 3.13). This region was previously identified to be crucial for endothelial-specific eNOS expression [Fish et al, 2005]. Consistent with the Hi-C data, 4C-seq also revealed the chromosomal proximity between the LEENE enhancer and eNOS promoter and this interaction is substantially increased in ECs subjected to PS as compared with OS (Figure 3.8G).

LEENE enhancer region forms proximal association with eNOS to enhance eNOS transcription. To examine whether the LEENE-associated enhancer plays a role in positive regulation of eNOS transcription, we employed CRISPR-cas9 gene editing to remove the ~10 kb enhancer region of LEENE immediately upstream of its TSS, as illustrated in Figure 3.14A.

The sgRNA-guided cas9 cutting efficiency was first verified using the surveyor assay in human embryonic kidney (HEK) 293 cells (Figure 3.15A) and then in ECs using genomic PCR assay with primers probing the 5' and 3' ends of the targeted region (Figure 3.15B). As a result of the enhancer ablation, the transcription of LEENE and eNOS was significantly suppressed, in both DMSO (a control vehicle) and ATV-treated ECs (Figure 3.14B). These changes in gene expression were attendant with similar changes in the proximity association between eNOS and LEENE, as revealed by DNA FISH (Figure 3.14C-D). We also deleted the coding region of LEENE in ECs and examined the eNOS expression in ECs. As shown in Figure 3.14E, the coding region deletion of LEENE also significantly decreased eNOS. Taken together, the LEENE enhancer forms long-range DNA interaction with eNOS promoter, serves as a prerequisite for eNOS expression under both untreated and statin-induced conditions. On the other hand, LEENE RNA transcript may also mediate, at least in part, this positive regulation of eNOS.

LEENE RNA transcript regulates eNOS expression and endothelial function at the transcriptional level. To further address the role of LEENE RNA transcript, we inhibited LEENE using LNA Gapmers, which can effectively silence the target nuclear RNA via an RNase H-mediated degradation [Watts et al, 2012]. First, we tested two LNAs targeting two separate regions of LEENE in ECs under basal condition (Figure 3.16A). Compared with the scrambled control, both LEENE-inhibiting LNAs decreased the eNOS mRNA levels in HUVECs (Figure 3.16A). To confirm this result, we also silenced LEENE in human aortic endothelial cells (HAoECs), i.e., ECs with a different origin, and observed a similar effect in the suppression of eNOS mRNA expression (Figure 3.17). In contrast to the suppressive effect on eNOS, LEENE LNAs led to an increased transcription of pro-inflammatory molecules intercellular adhesion molecule 1 (ICAM1) and VCAM1 (Figure 3.18). Next, we further demonstrated that inhibition of LEENE RNA decreased eNOS expression at the protein level

in ECs under pharmacological or physiological stimuli, i.e. ATV or PS (Figure 3.14B and Figure 3.19). To examine the functional regulation of LEENE in ECs, we performed a monocyte adhesion assay, which has been commonly used to assess the eNOS-mediated endothelial function. As shown in Figure 3.16C and Figure 3.16D, inhibition of LEENE significantly increased the number of monocytes adhering to ECs subjected to PS.

To mimic the effect of LEENE induction by PS, ATV, and KLF2/KLF4, we overexpressed LEENE in its predominant form (encoded by Exons 1, 3, and 4) in ECs using a CMV-driven and GFP-tagged adenovirus. With comparable transfection efficiency as control GFP vector (Figure 3.20), LEENE overexpression increased the mRNA levels of eNOS in both HUVECs and HAoECs (Figure 3.16E and Figure 3.20). In line with the increased eNOS transcription, LEENE overexpression also led to increased eNOS protein level (Figure 3.16F and Figure 3.19) and eNOS-derived NO production (Figure 3.16G). Collectively, results in Figure 3.16 suggest that LEENE RNA positively regulates eNOS expression and its associated endothelial function.

LEENE RNA promotes RNA Pol II-mediated nascent mRNA transcription of eNOS. We next examined the molecular mechanism that explains how LEENE promotes the eNOS transcription. Because enhancer-promoted transcription typically requires TFs, Mediator (Med) complex [Lai et al, 2013] and RNA Polymerase II (Pol II) [Li et al, 2012], and lncRNAs have been suggested to bind these factors/complexes to promote transcription [Yang et al, 2013], we hypothesized that LEENE RNA transcript may promote eNOS transcription by facilitate the recruitment of one or more of these transcriptional activators in the LEENE-eNOS loci.

We first tested whether there is an increased binding between LEENE and these TFs in ECs treated with ATV. As shown in Figure 3.21A-C, RNA-IP revealed that the associations of LEENE RNA with Pol II, KLF4, MED1, and were substantially enhanced in ECs treated with

statin. As an isotype control, IgG did not pull down significant amount of LEENE RNA, and the association of LEENE RNA with IgG was not altered by statin treatment (Figure 3.22). In order to test whether LEENE RNA associate with LEENE-eNOS loci, we performed chromatin isolation by RNA purification (ChIRP) assay with two pools of biotin-labeled RNA probes (even-numbered and odd-numbered), each with 5 probes containing sequences complementary to the respective regions of LEENE (Figure 3.21D). We were able to recover/enrich LEENE RNA specifically and efficiently, with β-actin RNA as a negative control (Figure 3.23A). Of note, this enrichment of LEENE was not achieved with biotin-labeled LacZ probes (Figure 3.23B). In the LEENE-enriched chromatin precipitates, the DNA sequences in LEENE enhancer region and eNOS promoter were also detected, suggesting that LEENE RNA indeed interact with these chromosomal regions (Figure 3.21E). Furthermore, these interactions were increased by statin treatment, which induces eNOS and LEENE (Figure 3.21E). As an additional control, we performed ChIRP assay in HEK293 cells, which do not express detectable level of endogenous eNOS. Compared to ChIRP performed using ECs, LEENE ChIRP using HEK293 cells revealed virtually no binding between LEENE RNA and the genomic loci of LEENE and eNOS (Figure 3.23C). Importantly, the statin-induced interaction between LEENE RNA and eNOS locus appears to be region-specific because this was absent for the 150 kb up- or downstream of LEENE encoding PELI2 and KTN1 respectively (Figure 3.23D).

Next, to test whether LEENE is required for the recruitment of KLF2/KLF4, Med1, and RNA Pol II to enhancer eNOS transcription, we determined the association of these proteins with the eNOS promoter in LEENE-depleted cells. As shown in Figure 3.21F, compared with ECs transfected with scramble LNA, LEENE LNA resulted in the reduced association between RNA Pol II and multiple eNOS promoter regions, although that between KLF4 or Med1 and eNOS promoter regions did not change (Figure 3.24). In line with the inhibitory effect of

LEENE LNA in Pol II binding to eNOS promoter, LEENE LNA caused a significant decrease in nascent eNOS mRNA level, which is quantified by nascent RNA pulldown combined with qPCR (Figure 3.21G). Collectively, Figure 3.21 suggests that LEENE RNA regulates the transcription of eNOS gene by facilitating the recruitment of RNA Pol II and the resultant nascent RNA transcription.

Mouse homologue of human LEENE. Next, we explored the conservation of LEENE between human and mouse. First, we compared the genomic structure of chromosomal region between KTN1 and PELI2 in human vs. mouse and found an expressed sequence tag (EST) (identifier BY707159.1) located in the similar region in mouse chromosome 14 as LEENE. Similar to the human LEENE, BY707159.1 is also transcribed from the negative strand (Figure 3.25A), and the surrounding DNA region contains multiple KLF2/KLF4 binding sites (Figure 3.25B). Comparison of the sequences of BY707159.1 showed that 472 out of 680 bp were aligned to the Exons 1, 3, and 4 of the human LEENE (Figure 3.25C). To explore its functional and disease relevance, we examined the level of BY707159.1 in the mouse artery. It is well established that the mouse thoracic aorta (TA) and aortic arch (AA) are associated, respectively, with distinct flow patterns and opposite endothelial phenotypes, and that eNOS is expressed at a significantly higher level in TA than AA31. Hence, we determined the transcription level of BY707159.1 in TA and AA isolated from C57BL mice. As shown in Figure 3.25D, the level of BY707159.1 was ~8 fold higher in TA than AA; this recapitulates the PS-induction and OS suppression of LEENE levels in the human ECs (Figure 3.5). Further, we have isolated lung ECs from C57BL mice and examine the potential regulation of eNOS by LEENE homolog in mouse. As shown in Figure 3.25E, LNA inhibiting of LEENE homolog indeed decreased the mRNA level of eNOS. These findings suggest that LEENE regulation of eNOS may be a conserved mechanism in mouse and human.

Discussion Of LEENE Study

We have identified LEENE, a lncRNA encoded by a distal enhancer region that forms proximal association with the eNOS locus. LEENE RNA enhances RNA Pol II binding to the eNOS promoter, thereby enhancing eNOS transcription. Inhibition of LEENE at either genomic (i.e. DNA) or transcriptional (i.e. RNA) level suppresses eNOS transcription, whereas overexpression of LEENE increases levels of eNOS and its derived NO bioavailability. Elucidation of this mechanism provides novel insights into the epigenetic modulation of endothelial gene expression in health and disease. All of this was successfully pursued through the analysis of transcriptomic, epigenetic, and chromatin structure "omics" data, followed by robust experimental validation.

To identify the lncRNAs that potentially regulate eNOS transcription, we employed a systems biology approach to profile flow-regulated endothelial transcriptomes. Among all the lncRNAs that are differentially regulated by PS vs. OS, LEENE ranked at the top (log2FC=1.93, correlation coefficient=0.85) (Figure 3.5B-D). Indeed, LEENE RNA was found to be regulated in concert with eNOS in ECs under hemodynamic, biochemical, and pharmacological stimuli (Figure 3.5F-G). At the transcriptional level, LEENE and eNOS are co-regulated by KLF2 and KLF4 (Figure 3.7). The hierarchical regulation of KLF2 and KLF4 upstream of LEENE and eNOS is reflected by the early induction of KLF2/KLF4, preceding that of LEENE and eNOS (Figure 3.5D). The identification of KLF2/KLF4-induced LEENE expands the repertoire of these TF-regulated transcriptional targets. In line with this notion, a recent report identified globally enriched TF binding motifs for KLFs in ECs using ChIP-seq and ATAC-seq [Zhou et al, 2017]. Therefore, KLFs may regulate a broader spectrum of transcriptional targets, including not only the protein-coding genes [Atkins et al, 2007], and miRNAs [He et al, 2017; Hergenreider et al, 2012], but also lncRNAs.

While there have been extensive studies on the regulatory mechanisms of eNOS expression at multiple levels, there is a lack of information on the role of epigenetic modulation, particularly through lncRNAs and long-range DNA interaction. Summarizing the findings from our study, LEENE may enhance eNOS expression through two layers of regulations: 1) LEENE enhancer serves as a distal enhancer that forms proximal association with eNOS promoter (Figure 3.8 and Figure 3.14); 2) in such a chromosomal context, LEENE RNA transcript induced by KLF2/KLF4 facilitate the binding of RNA Pol II to promote nascent RNA transcription of eNOS (Figure 3.21). To tease out the reciprocal requirement of these two layers, we found that in ECs with LEENE enhancer ablated, overexpression of LEENE failed to induce eNOS expression (Figure 3.26A). Furthermore, in these LEENE enhancer-deleted cells, the association between LEENE transcript and eNOS locus is significantly decreased under both untreated and ATV-treated conditions (Figure 3.26B). These findings support the notion that the proximity association between LEENE and eNOS loci would be a prerequisite for the association of LEENE RNA to eNOS promoter; without the LEENE enhancer region, LEENE RNA is not sufficient to enhance eNOS transcription. This hypothesis is illustrated in Figure 3.27. It remains to be explored how do the LEENE and eNOS loci come in proximity and form the promoter-enhancer contact in ECs but not other cells types, and whether LEENE RNA transcript per se further stabilize such inter-chromosomal interaction.

Considering the genomic feature of LEENE, one may classify LEENE as an enhancer RNA (eRNA) as LEENE is encoded in a ~300 kb H3K27ac-enriched and H3K4me-enriched region. However, comparing to most eRNAs reported to promote nearby gene transcription in cis [Espinosa, 2016], LEENE does not seem to affect its neighboring genes, because 1) its RNA level is discordant with neighboring KTN1 and PELI2, which do not show differential expression in ECs subjected to different flow patterns or overexpression of KLF2/KLF4 (Figure 3.10A-B); and 2) neither LNA knockdown nor CRISPR deletion of LEENE locus affects its

adjacent KTN1 expression (Figure 3.10C-D). Intriguingly, LEENE transcription is highly concordant with eNOS, the key endothelial molecule encoded on chromosome 7. Despite the seemingly distinct chromosomal territories, LEENE and eNOS loci show proximal association in ECs, both under untreated or ATV/PS-treated conditions (Figure 3.8E-G and Figure 3.14C-D). This is in line with the emerging notion that lncRNAs may facilitate co-regulation of genes involved in similar biological processes [Engreitz et al, 2013]. Given the molecular mechanisms identified in this study, we reason that LEENE would be an example of lnc-eRNA or elncRNA, which is a transcript with initiation sites overlapping with enhancer regions and presence in current lncRNA databases [Li et al, 2016; Rothschild et al, 2017].

Other than eNOS, LEENE may interact with genomic loci encoding a set of genes that are involved in multiple pathways crucial for endothelial homeostasis, e.g., cell adhesion and VEGF signaling (Figure 3.6). In addition, LEENE may also regulate other genes important for endothelial function through indirect mechanism. For example, we found that LEENE LNA decreases, whereas LEENE overexpression increases thrombomodulin (Tm), another KLF2 transcriptional target in ECs (Figure 3.28) [Lin et al, 2005]. In reference to the HUVEC Hi-C data, there is a lack of direct interaction between LEENE and Tm loci (Figure 3.29). The complete repertoire of LEENE-regulated transcriptome remains to be characterized.

In addition to LEENE, we also identified a number of other lncRNA loci in the eNOS 4C libraries, such as MALAT1 and MIAT (Figure 3.30), all of which have been shown to be abundantly transcribed and play functional roles in ECs [Michalik et al, 2014; Yan et al, 2015]. The chromosomal contacts of other lncRNAs with eNOS may recruit additional chromatin remodelers to modulate eNOS transcription. These mechanisms may coordinate with TF-binding and histone modifications to organize the chromatin conformation of eNOS, contributing to its transcriptional control. Given the recent study demonstrating the poor CpG content and the lack of flow-altered DNA methylation status in eNOS promoter [Jiang et al,

2014], our findings suggest that the lncRNA-mediated chromatin remodeling may be an important factor other than DNA methylation in epigenetic regulation of eNOS, with both spatial and temporal control.

The cross-species conservation of lncRNAs is a challenging and key topic in the epigenetics field as the estimated sequence homology between human and mouse lncRNAs is only 20% [Hezroni et al, 2015]. We identified BY707159.1, which is similar to human LEENE in several aspects including sequences, genomic structure, TFBS enrichment, differential regulation by flow patterns, and its gene regulation of eNOS (Figure 3.25). Given the conservation of Firre in the repeating RNA domains [Hacisuleyman et al, 2014], it is possible that LEENE is conserved between human and mouse in regions/domains important for its molecular function. Our findings set the stage for future systematic exploration of the functional roles of LEENE as an epigenetic player in cardiovascular health and diseases.

Acknowledgments

The authors would like to thank Drs. Arthur Riggs, Rama Natarajan, Amy Leung, Dustin Schones, Kevin Morris, and Jing-Kuan Yee at Beckman Research Institute, City of Hope, and Drs. John Y.-J. Shyy, Michael (Geoff) Rosenfeld, Bing Ren, and David Gorkin at UCSD for the useful discussions and their valuable suggestions, and Drs. Wouter de Laat and Adrien Melquiond for their consultation on 4C-seq and chromatin conformation study. This work was supported in part by US NIH research grants K99/R00HL122368 (Z.C.), R01HL106579 and HL108735 (S.C. and S.S.), Beckman Research Institute Startup Fund (Z.C.), and American Heart Association Postdoctoral Fellowship 17POST33410101 (T.-S.H.). The Genome Editing and Analytical Cytometry Cores at City of Hope were supported by the National Cancer Institute of the National Institutes of Health under award number P30CA033572. Chapter 3 is a modified presentation of “Systems Biology Analysis of Longitudinal Functional Response of Endothelial Cells to Shear Stress" as it appears in PNAS 2017 by Nassim E Ajami, Shakti

Gupta, Mano R Maurya, Phu Nguyen, Julie Yi-Shuan Li, John Y-J Shyy, Zhen Chen, Shu Chien, and Shankar Subramaniam. The dissertation author was the primary author of this material.

Chapter 3 is a modified presentation of “Enhancer-Associated Long Non-Coding RNA LEENE Regulates Endothelial Nitric Oxide Synthase and Endothelial Function" as it appears in Nature Communications 2018 by Yifei Miao, Nassim E Ajami, Tse-Shun Huang, Feng-Mao Lin, Chih-Hong Lou, Yun-Ting Wang, Shuai Li, Jian Kang, Hannah Munkacsi, Mano R Maurya, Shakti Gupta, Shu Chien, Shankar Subramaniam, and Zhen Chen. The dissertation author was a co-primary author of this material.

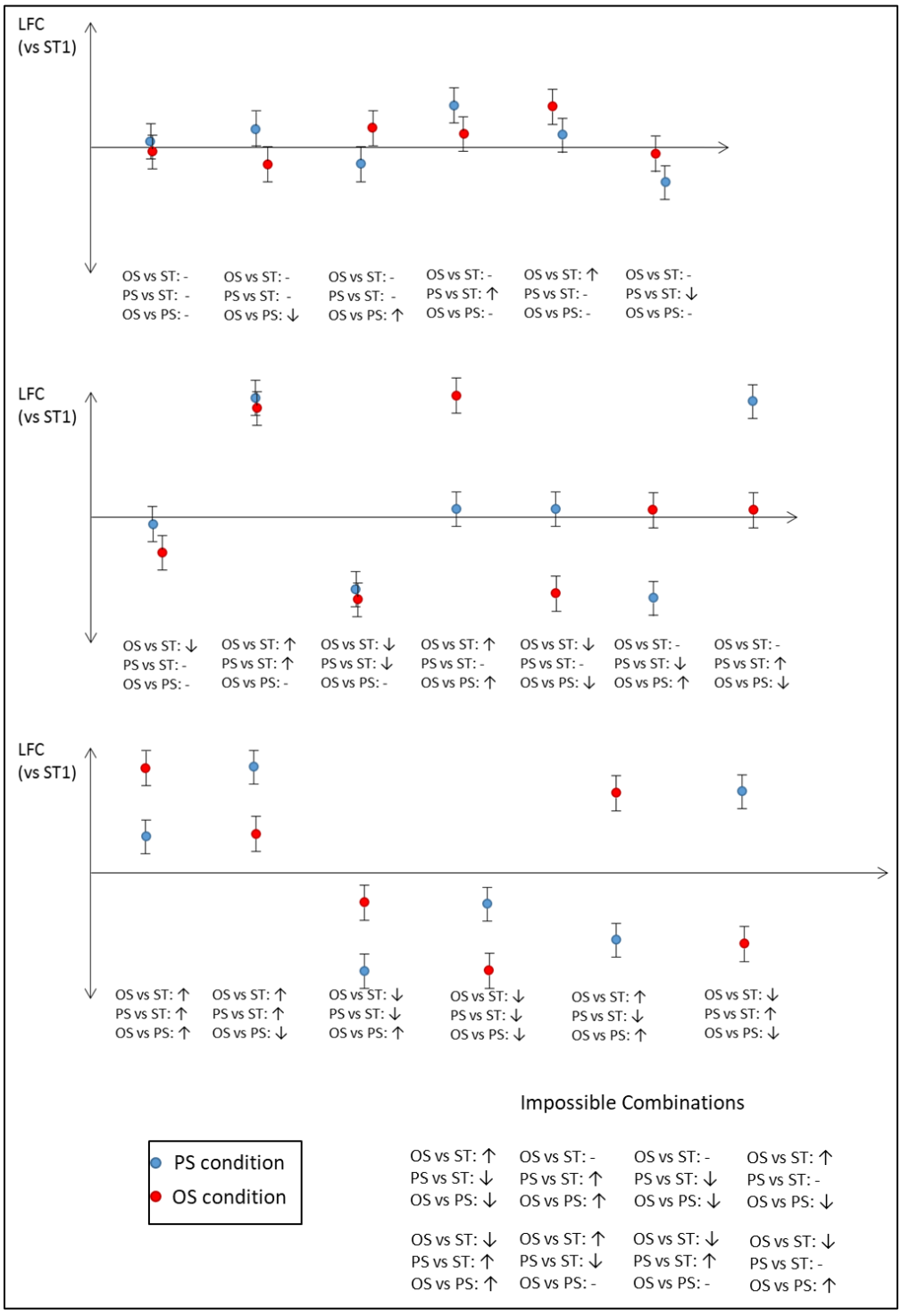


Figure 3.1: Visualization of the 27 possible point-wise states for a given gene's expression across three pairwise conditions and three possible states of regulations.

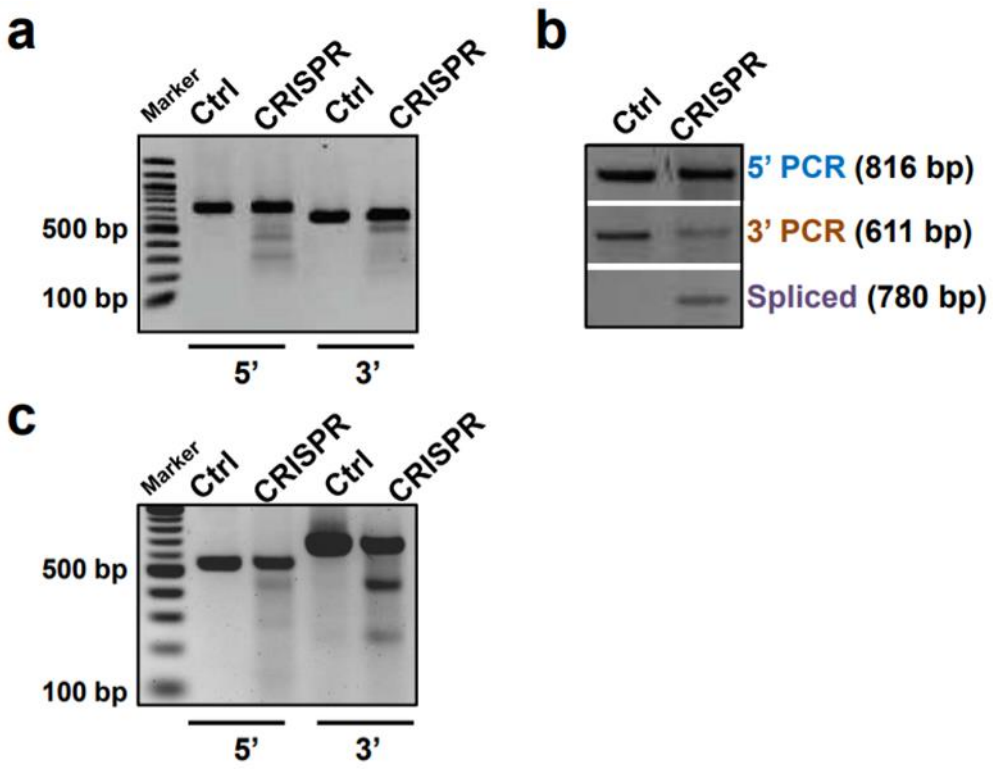


Figure 3.2: Quality control of CRISPR-Cas9 gene editing system targeting LEENE promoter/enhancer (ED) (a and b) and coding region (CD) (c). (a) Surveyor assay for ED sgRNA-guided cas9 cutting efficacy in HEK293 cells. Select sgRNAs targeting respectively 5' or 3' ends were used to test the cutting efficiency as compared to control vector. (b) PCR validation of ED region of Cas9 targeting position as illustrated in Fig. 4a. The spliced and ligated 10 kb region is amplified in CRISPR-targeted but not control ECs. (c) Surveyor assay for CD sgRNA-guided cas9 cutting efficacy in HEK293 cells. Select sgRNAs targeting respectively 5' or 3' ends were used to test the cutting efficiency as compared to control vector.

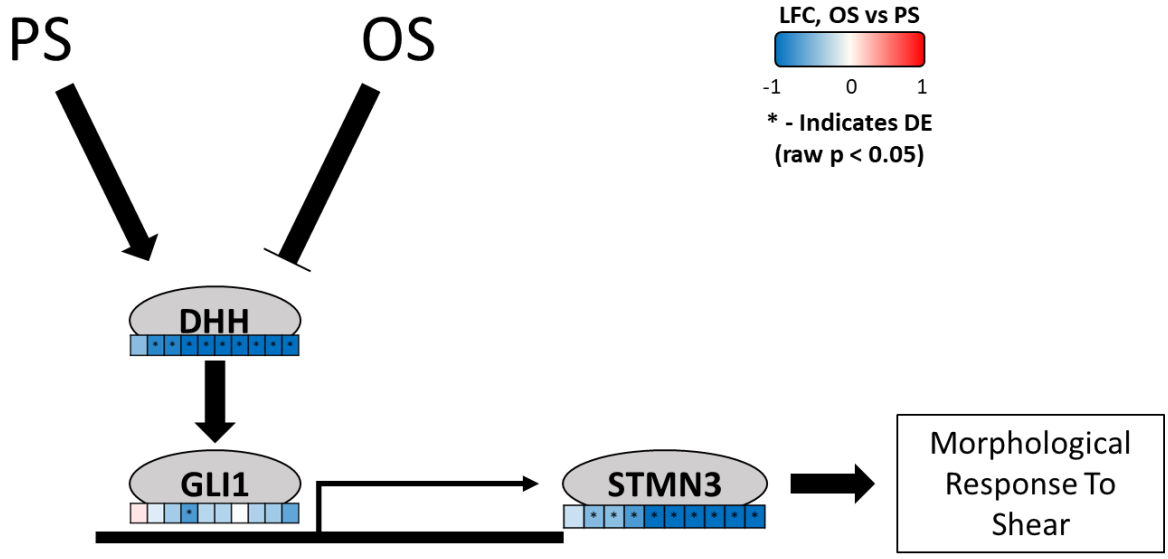


Figure 3.3: Proposed mechanosensitive mechanism of regulation of endothelial morphology. DHH, an integral component of the hedgehog signaling pathway, is upregulated in PS vs ST while being downregulated in both OS vs ST and OS vs PS. DHH signaling activates the Gli family of transcription factors. Among the Gli targets is STMN3, a stathmin protein that is involved in microtubule formation and function. The expression profile of STMN3 exhibits strong similarity with the expression profile of DHH, indicating a potential signal transduction from shear to transcriptional regulation.

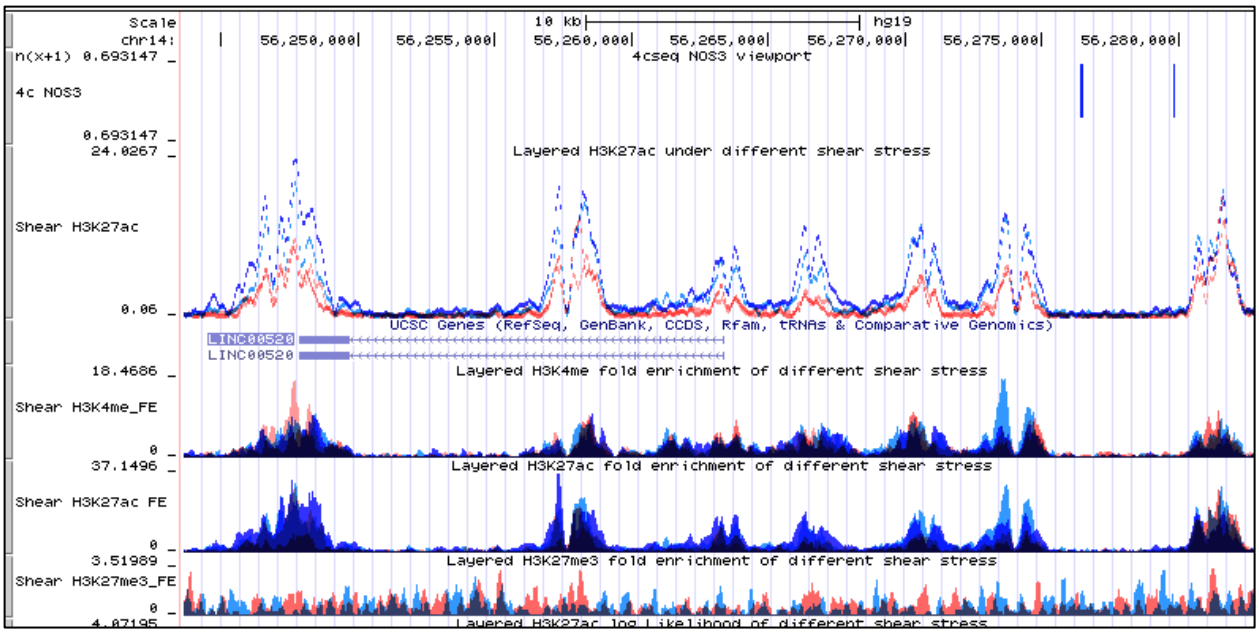


Figure 3.4: LINC00520 region with visualized histone ChIP-seq and 4C datasets. Red data represents the OS condition, while blue data represents the PS condition. Seven distinct acetylation and methylation peaks can be observed on and upstream of LINC00520's gene body. This is a screenshot of the UCSC Genome Browser [Rosenbloom et al, 2013; Ernst et al, 2010; Ernst et al, 2014].

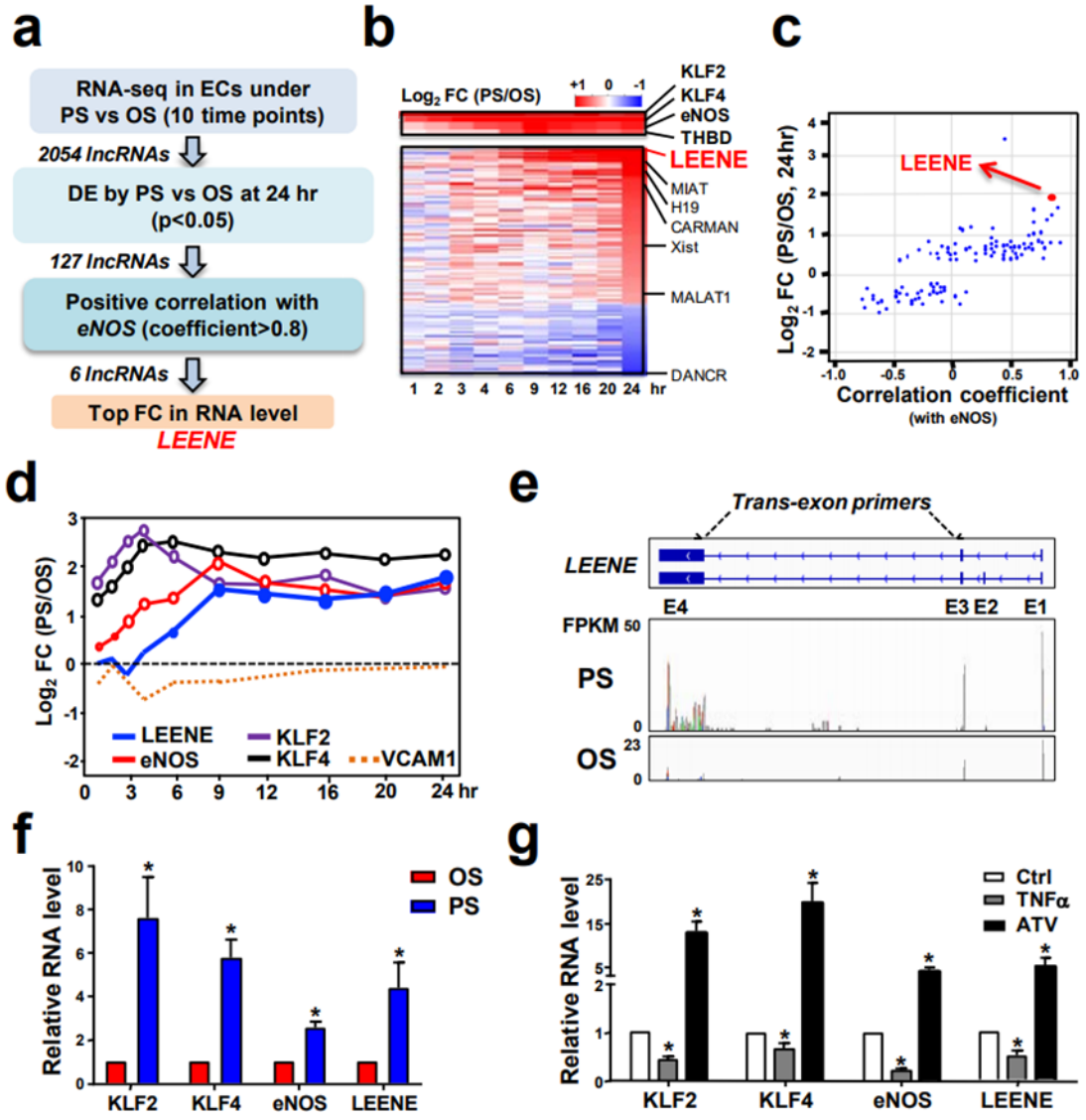


Figure 3.5: Co-regulation of LEENE and eNOS. (a) Pipeline for assessing LINC00520's unique regulatory standing as a lncRNA. (b) Heatmap of RNA levels of flow-regulated lncRNAs derived from RNA-seq. (c) Scatter plot of the flow-regulated lncRNAs ranked by differential expression (DE) fold change (FC) at 24 hr (PS/OS) and correlation with eNOS mRNA level. (d) Time course of log2FC of mRNAs encoding various genes. (e) Structure of LEENE gene encoding two RNA transcripts. Trans-exon primers used in qPCR were designed to amplify fragments flanking Exons 3 and 4. RNA-seq tracks depicting FPKM of LEENE in ECs under PS or OS for 24 hr. (f, g) qPCR detection of various RNA transcripts in ECs subjected to PS or OS (in f) or TNFα (100 ng/ml) or atorvastatin (ATV) (1 μM) (in g) for 24 hr. Data are presented as mean±SEM, n=5 in each group. *p<0.05.

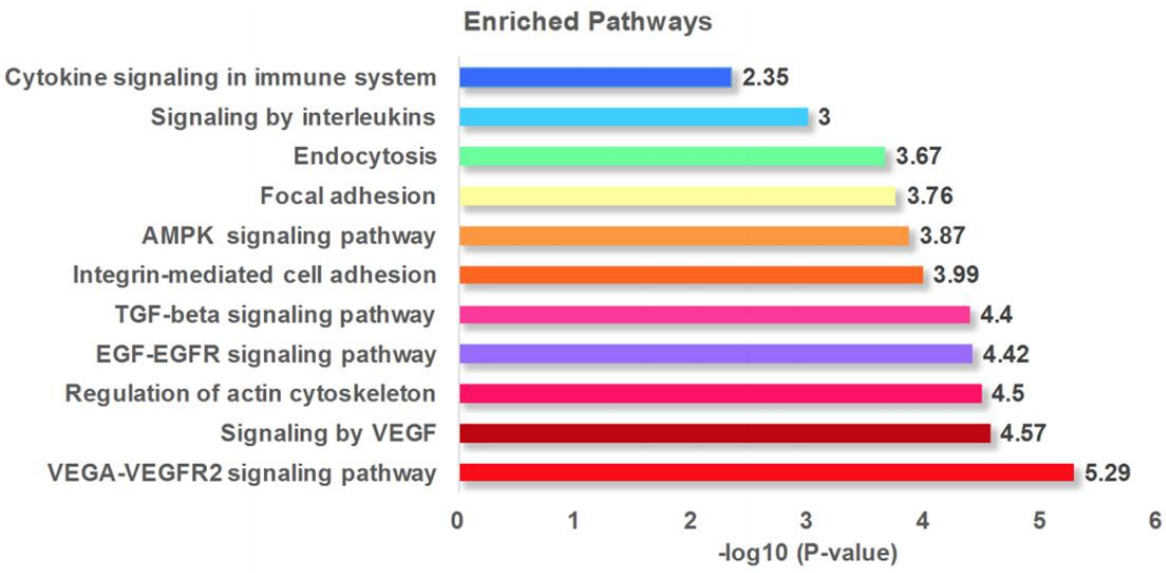


Figure 3.6: Pathway enrichment analysis from 1177 genes that are differentially regulated between OS and PS and interact with LEENE genomic loci. Pathway enrichment analysis was conducted using CPDB [Kamburov et al, 2011].

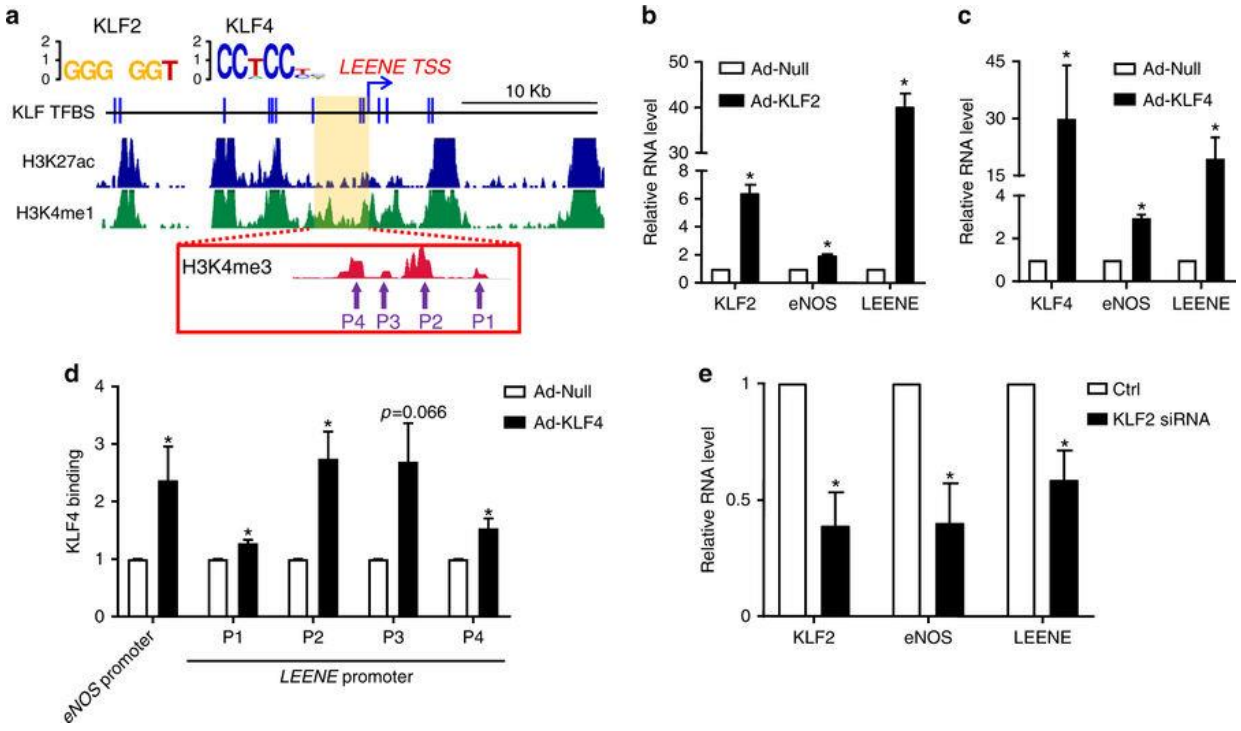


Figure 3.7: KLF2 and KLF4 transcriptionally regulate LEENE. (a) Putative KLF2 and KLF4 binding sites in LEENE enhancer/promoter based on the conserved KLF2 and KLF4 binding motifs (shown on the top). Middle tracks display H3K27ac and H3K4me1 ChIP-seq signals in LEENE locus, and the inset shows H3K4me3 ChIP-seq signals in the putative LEENE promoter region from ENCODE HUVEC data; arrows indicate regions detected in ChIP-qPCR (Fig. 2d). (b–d) HUVECs were infected with respective adenoviruses for 48 h. RNA levels of KLF2, KLF4, LEENE, and eNOS were detected by qPCR (in b and c) and KLF4 binding to promoters of eNOS and LEENE was quantified by ChIP-qPCR analysis (in d). (e) qPCR of respective RNA levels in ECs transfected with scramble control (Ctrl) or KLF2 siRNA. Data are presented as mean ± SEM, n = 5 in each group. * indicates $p < 0.05$ compared to respective controls using Student's t-test.

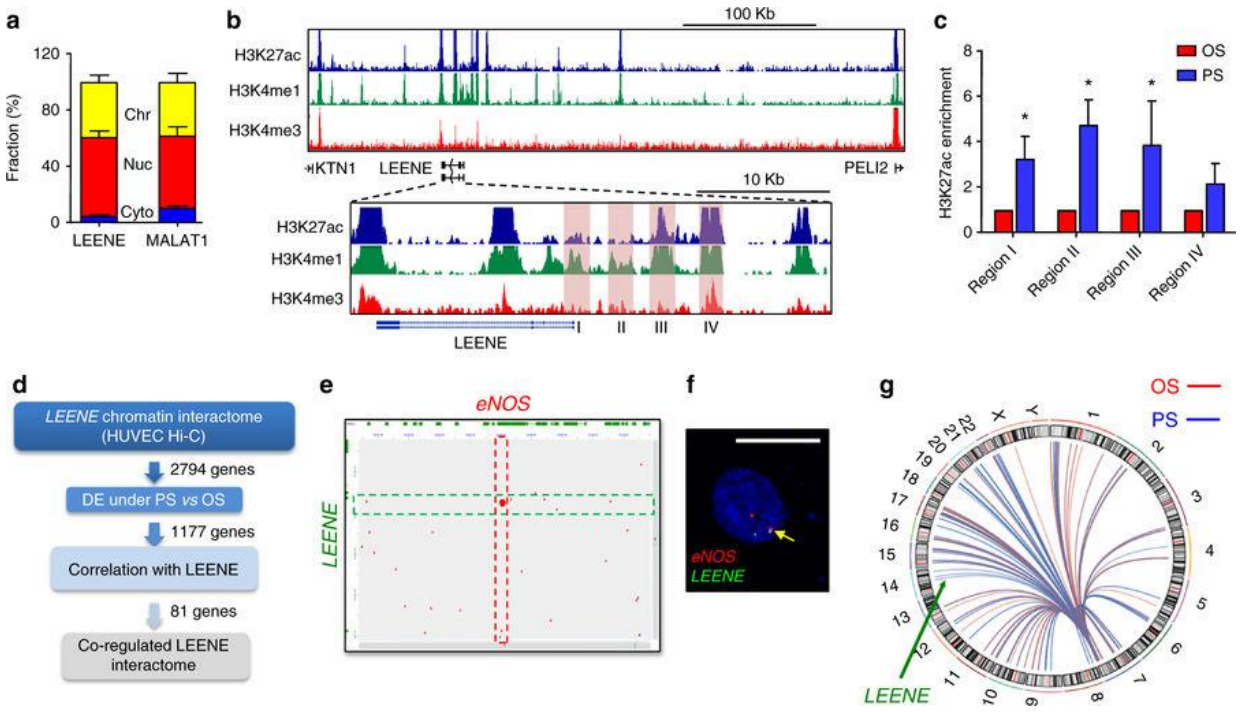


Figure 3.8: LEENE RNA is nucleus-localized and its DNA lies in enhancer region interacting with eNOS promoter. (a) qPCR quantitation of LEENE and MALAT1 in subcellular fractions from ECs, plotted as percentages in association with chromatin (Chr), nucleoplasm (Nuc), and cytoplasm (Cyt). (b) ENCODE HUVEC ChIP-seq signals in 400 kb (top tracks) and 50 kb (bottom tracks) regions surrounding LEENE. Regions in shades were selected for H3K27ac ChIP-qPCR in (c). (d) Flow chart of integrative Hi-C and RNA-seq analyses. (e) LEENE–eNOS interaction map generated from GEO HUVEC Hi-C analysis. Red pixels represent interactions between two regions, respectively, in chr7 (X-axis) and chr14 (Y-axis). The highlighted regions correspond to eNOS promoter and LEENE enhancer regions. (f) Representative image of DNA FISH with respective probes recognizing LEENE and eNOS genomic loci. Arrow indicates proximity association between two loci. Scale bar = 10 μm. (g) 4C-seq mapping of inter-chromosomal interactions between eNOS bait (249 bp) and lncRNAs listed in Figure 3.5B. Each line in the circoplot represents an interaction and the color intensity reflects the normalized reads of ligated DNA ends. Chromosomes are numbered around the circle. Data are presented as mean ± SEM, n = 5 in each group. * indicates $p < 0.05$ compared to OS based on Student's t test.

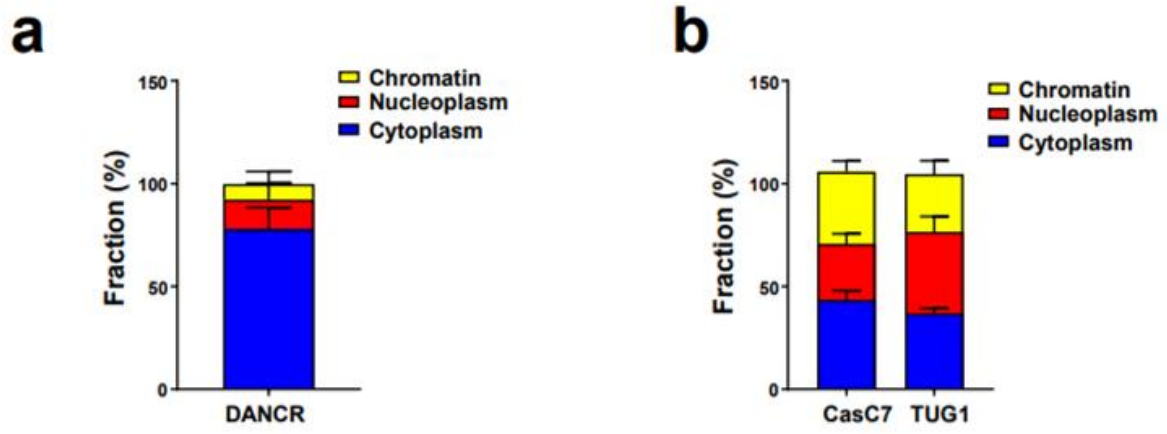


Figure 3.9: Subcellular quantitation for lncRNAs DANCR, Casc7, and TUG1. qPCR quantitation of DANCR (a), CasC7, and TUG1 (b) in subcellular fractions from ECs, plotted as percentages in association with chromatin, nucleoplasm, and cytoplasm. Data are presented as mean±SEM. ANOVA followed by Bonferroni post-test is applied. n=5/group.

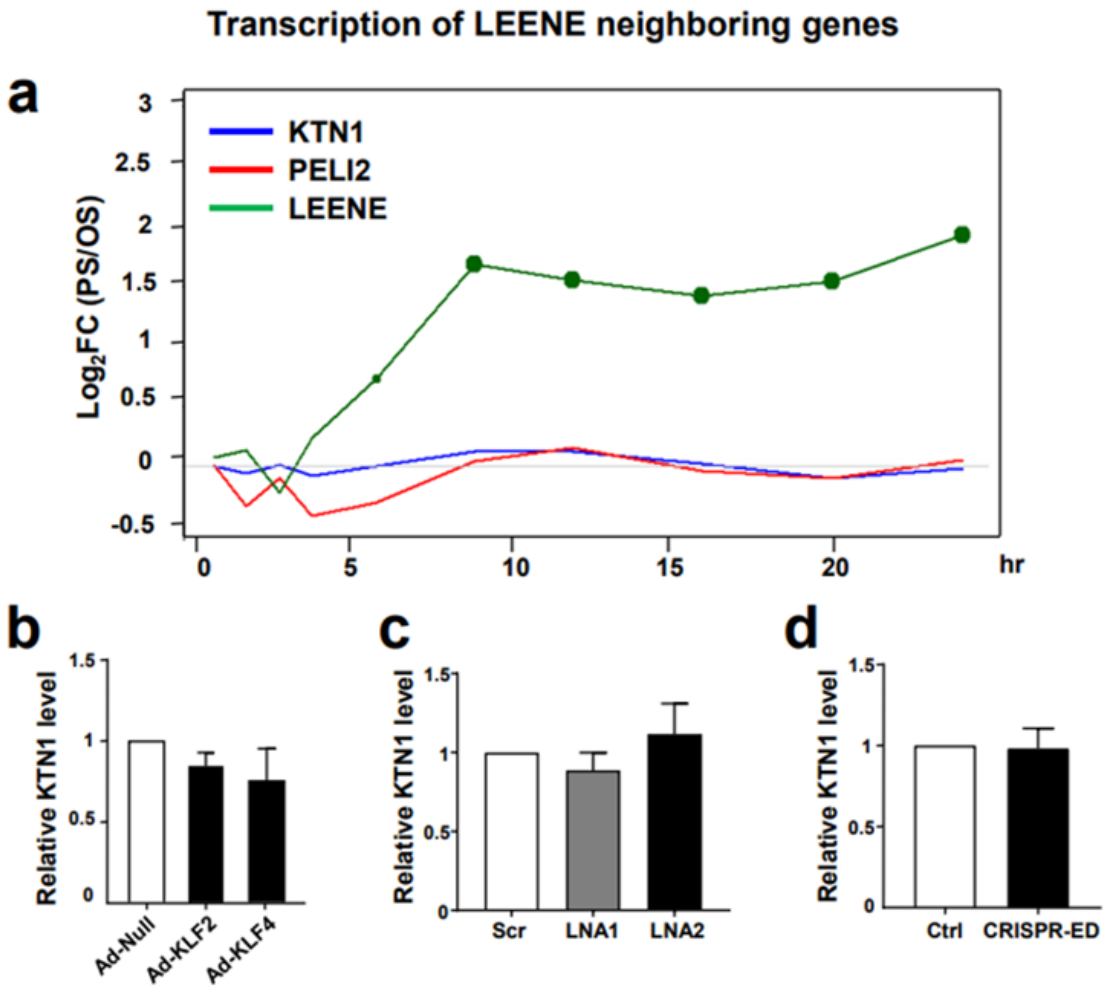


Figure 3.10: Transcription of LEENE neighboring genes. (a) Time course of log2FC (PS/OS) of LEENE and mRNAs encoding KTN1 and PELI2. (b-d) qPCR quantification of KTN1 in ECs with (b) KLF2 or KLF4 overexpression, (c) LEENE knock-down with two different gapmerLNA (50 nM), or (d) CRISPR-cas9 gene editing targeting LEENE promoter/enhancer regions. Data are presented as mean±SEM. ANOVA followed by Bonferroni post-test is applied, n=5/group.

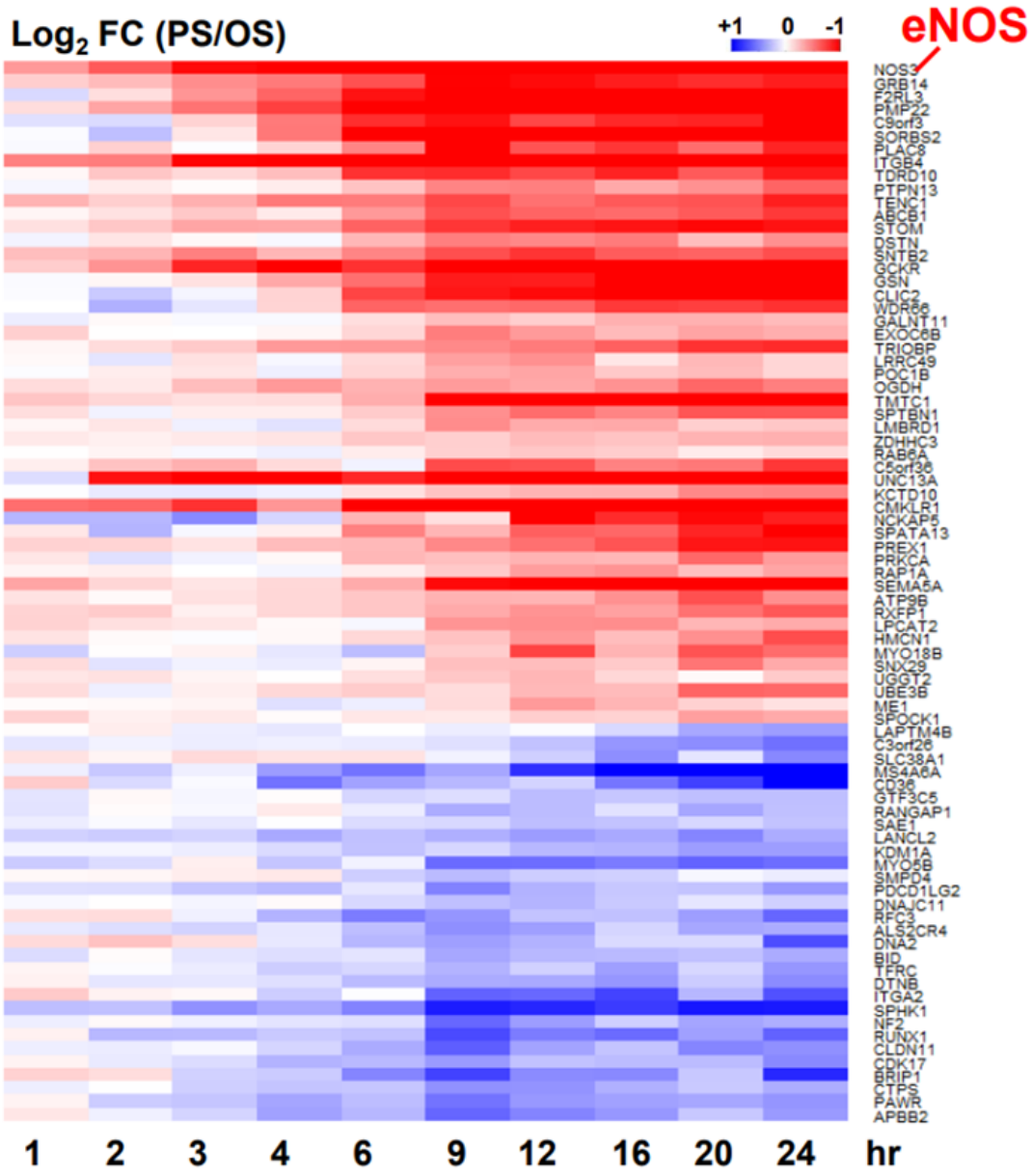


Figure 3.11: Heatmap of flow-regulated 81 genes in LEENE interactome. Heatmap demonstrates the flow-regulated 81 genes in the LEENE interactome in Figure 3.8D, with eNOS as one of the top correlated transcripts with LEENE.

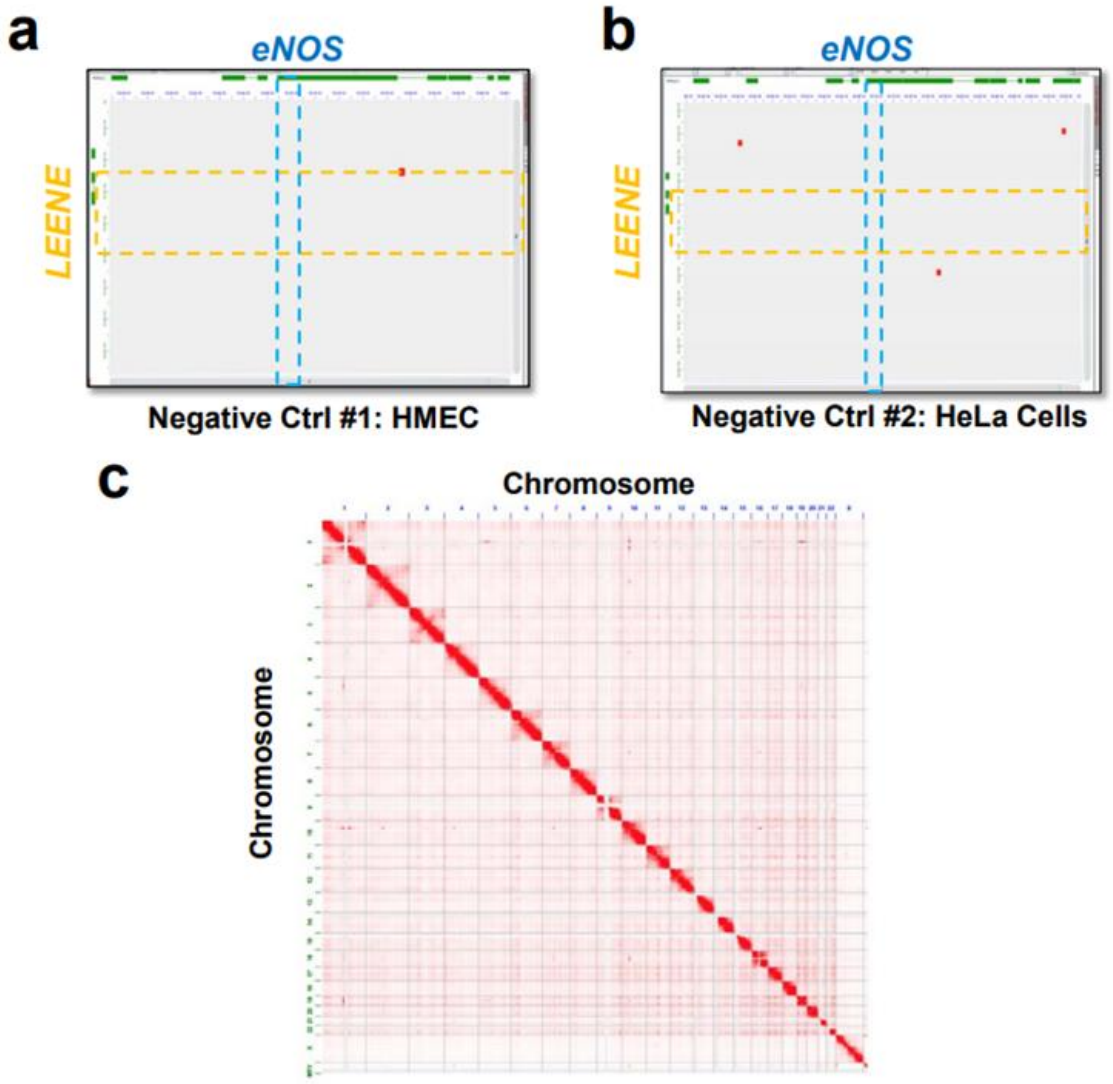


Figure 3.12: Hi-C analysis of LEENE-eNOS interaction map in non-EC cell types. LEENE-eNOS interaction map generated from Hi-C data from human mammary epithelial cells HMEC (a) or HeLa cells. Red pixels represent interactions between two regions respectively in chr 7 (X axis) and chr 14 (Y axis). The highlighted regions correspond to eNOS promoter and LEENE enhancer regions. Note that no red pixel was shown in the overlapped highlighted region. (c) A genome-wide contact matrix produced by Hi-C. Each pixel represents a chromosomal interaction using 5-kb windows; intensity corresponds to the total number of reads.

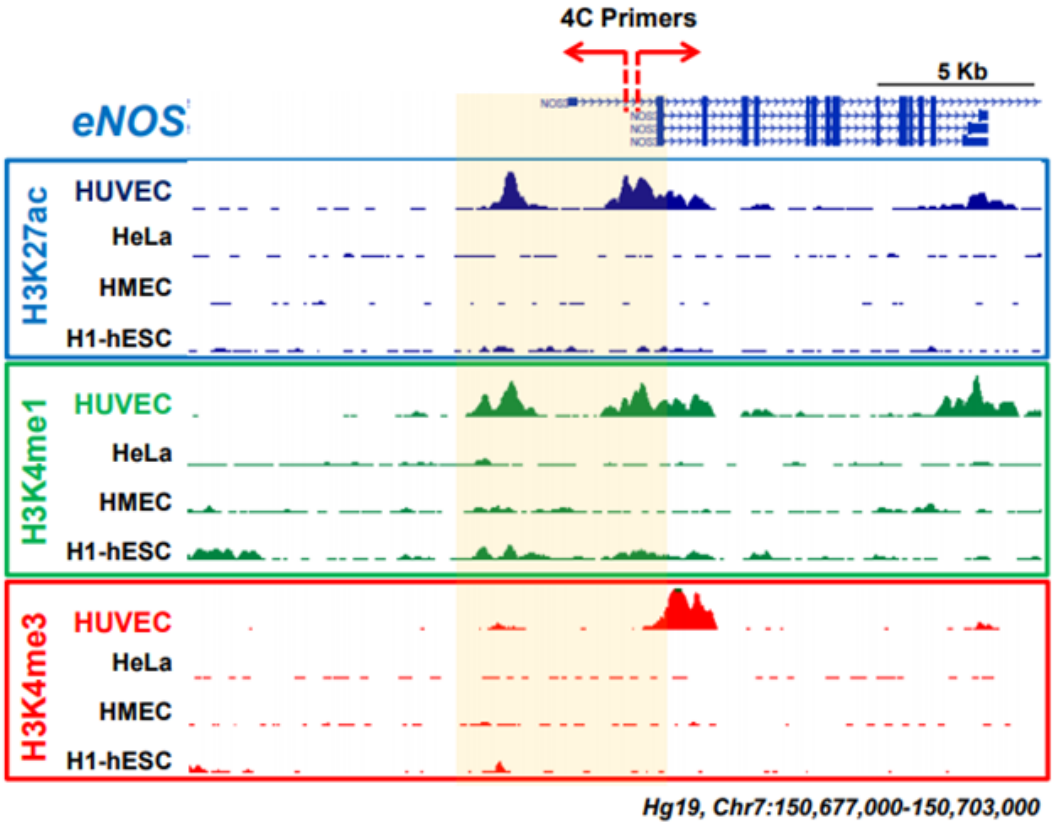


Figure 3.13: Histone ChIP-seq tracks of eNOS promoter/enhancer region from four different cell types. ENCODE histone ChIP-seq signal from four different cell types surrounding eNOS promoter/enhancer region. Shaded area indicates eNOS promoter/enhancer regions. The red arrows indicate 4C library primers directions and locations. HMEC, human mammary epithelial cells; H1-hESC, human embryonic stem cells.

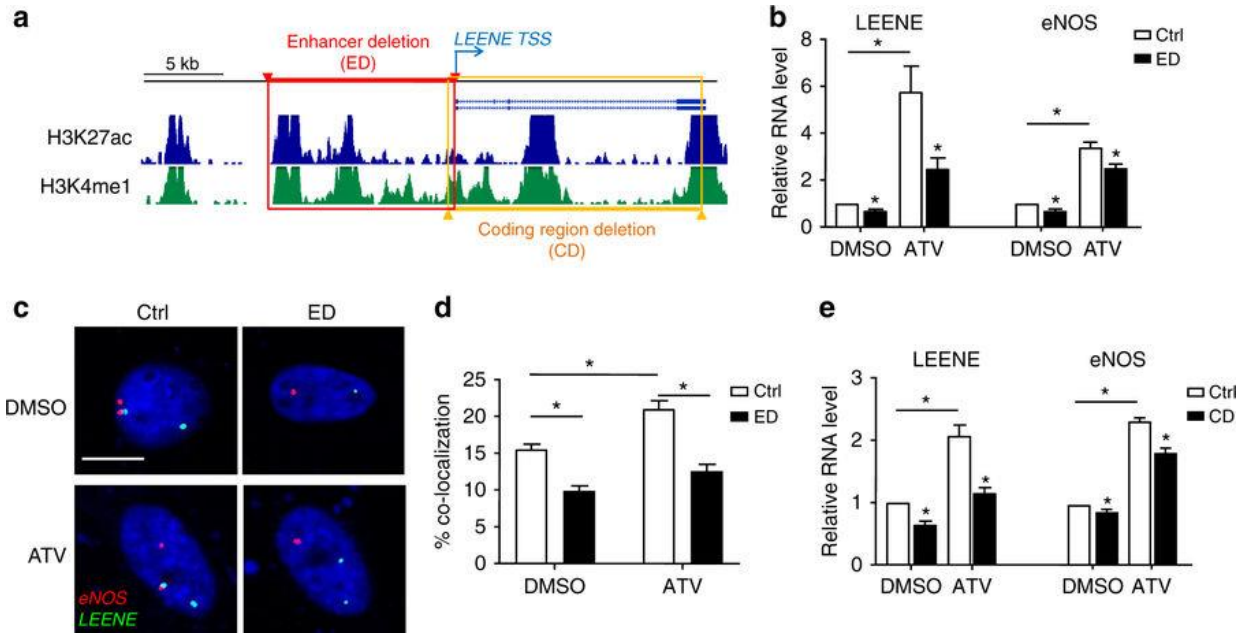


Figure 3.14: Gene editing of LEENE locus influences eNOS transcription. (a) Schematic illustration of CRISPR-Cas9 targeting strategy. Regions in red and orange indicate, respectively, the upstream enhancer/promoter region or the coding region deleted by sgRNA-guided Cas9, resulting in enhancer deletion (ED) and coding region deletion (CD) in LEENE locus. (b) LEENE and eNOS RNA levels in ECs transfected with control Cas9 plasmid (Ctrl) or "ED" Cas9-sgRNAs were quantified using qPCR. (c) DNA FISH for proximity association of LEENE and eNOS genomic loci. ECs transfected with control (Ctrl) or "ED" Cas9-sgRNAs were treated with DMSO or ATV (1 μM) for 24 h. Scale bar = 10 μm. (d) Percentage of cells with LEENE and eNOS proximity association (distance <1 μm). n = 678 in "DMSO-Ctrl" group; n = 425 in "DMSO-ED" group; n = 632 in "ATV-Ctrl" group; n = 581 in "ATV-ED" group. (e) qPCR quantification of LEENE and eNOS RNA levels in ECs transfected with control Cas9 (Ctrl) or "CD" Cas9-sgRNAs. All data are presented as mean ± SEM. n = 5 in each group unless specified. *$p < 0.05$ compared with "Ctrl" or between indicated groups based on Student's t-test.

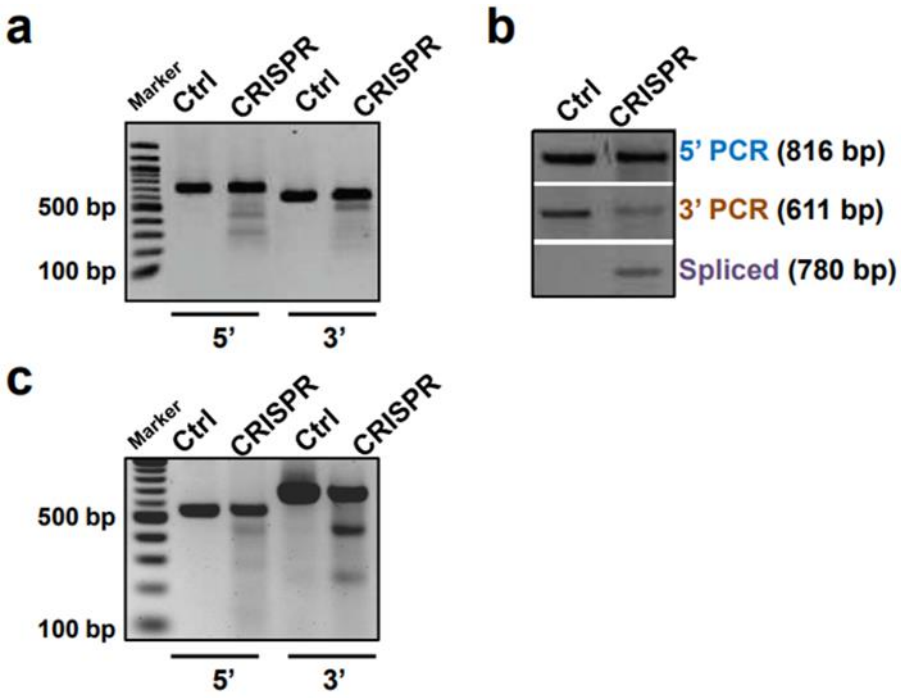


Figure 3.15: Quality control of CRISPR-Cas9 gene editing. Quality control of CRISPR-Cas9 gene editing system targeting LEENE promoter/enhancer (ED) (a and b) and coding region (CD) (c). (a) Surveyor assay for ED sgRNA-guided cas9 cutting efficacy in HEK293 cells. Select sgRNAs targeting respectively 5' or 3' ends were used to test the cutting efficiency as compared to control vector. (b) PCR validation of ED region of Cas9 targeting position as illustrated in Figure 3.14A. The spliced and ligated 10 kb region is amplified in CRISPR-targeted but not control ECs. (c) Surveyor assay for CD sgRNA-guided cas9 cutting efficacy in HEK293 cells. Select sgRNAs targeting respectively 5' or 3' ends were used to test the cutting efficiency as compared to control vector.

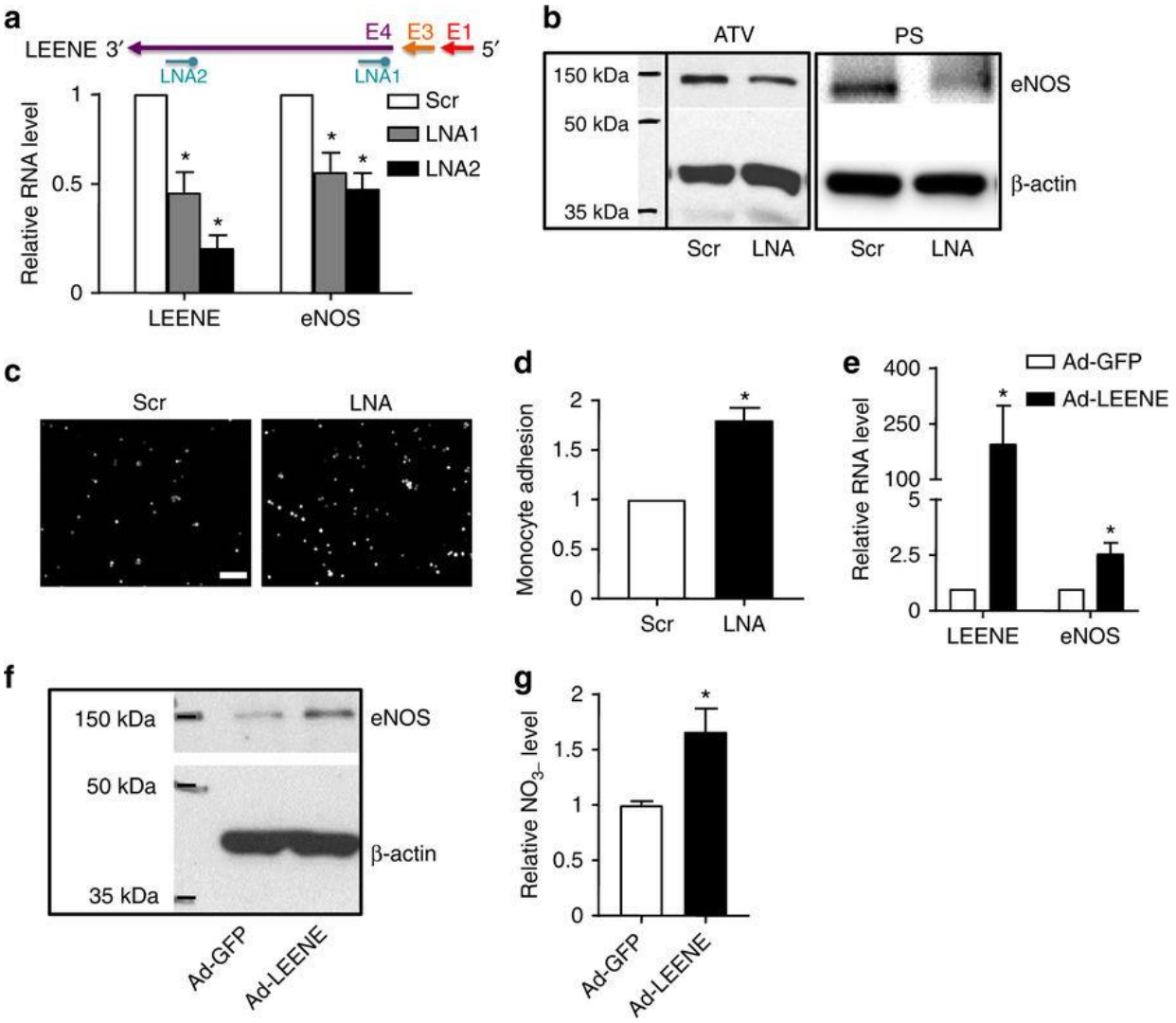


Figure 3.16: LEENE RNA regulates eNOS expression and EC function. (a–e) HUVECs were transfected with LNA (50 nM) targeting Exon 4 of LEENE. Basal RNA levels of LEENE and eNOS were detected by qPCR in (a). Protein levels of eNOS in HUVECs treated with ATV or PS were revealed by immunoblotting. (c, d) ECs were transfected with scramble or LEENE LNA before subjected to PS for 12 h. Fluorescence-labeled THP-1 cells were added to the EC monolayer, and the monocytes adhering to ECs were visualized by fluorescence microscopy (scale bar = 100 μm). The representative images are shown in (c) and the quantification based on five randomly selected fields per group per experiment are shown in (d). (e–g) HUVECs were infected with Ad-GFP or Ad-LEENE for 48 h. RNA levels of LEENE and eNOS were detected by qPCR (e), protein level of eNOS in HUVECs was revealed by immunoblotting (f), and NO production was measured by a fluorometric assay (g). Densitometry analysis of immunoblotting shown in (b) and (f) was performed (Figure 3.19). Data are presented as mean ± SEM. n = 3–5 in each group. Student's t test was used. * indicates $p < 0.05$ compared to scrambled control or Ad-GFP in respective experiments.

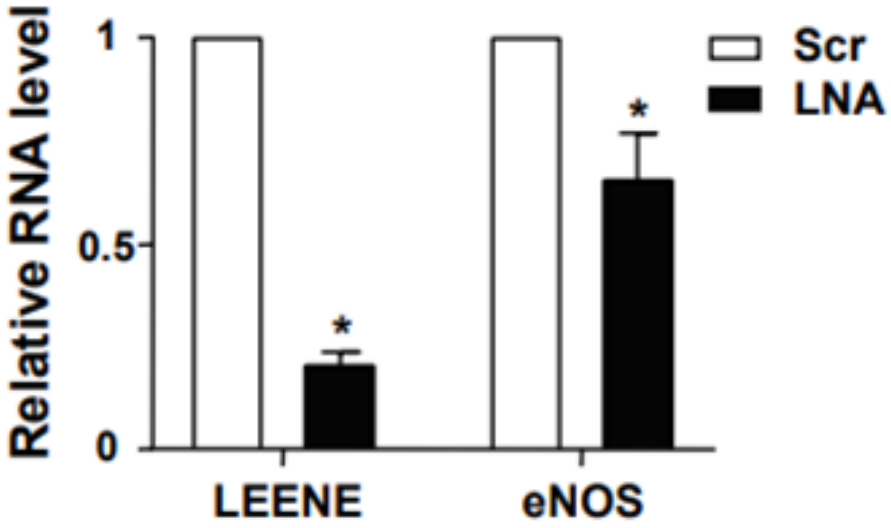


Figure 3.17: Effect of LEENE knock-down on eNOS mRNA level in HAoECs. Knock down of LEENE with LNA decreased mRNA expression of eNOS in HAoECs. Error bars represent mean±SEM, n=5/group. Student's t-test is applied, *p<0.05.

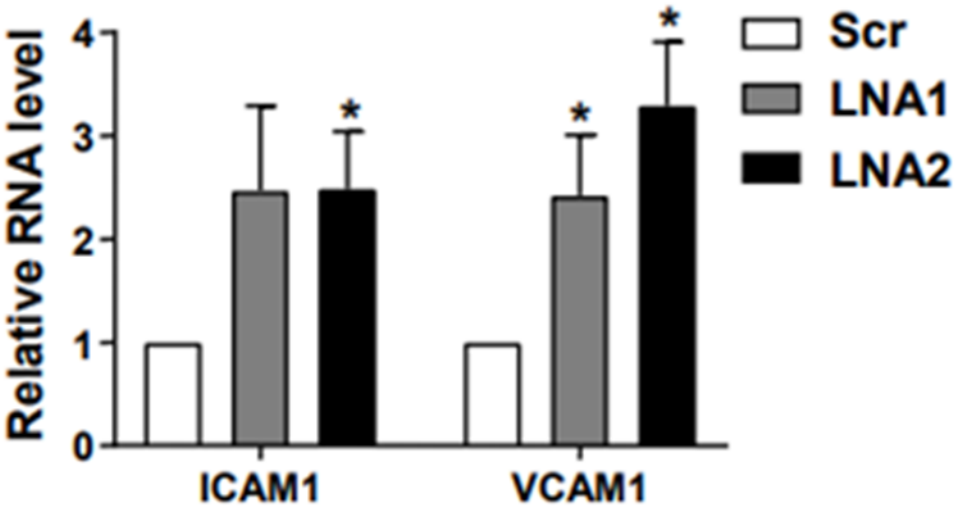


Figure 3.18: Effect of LEENE knock-down on pro-inflammatory gene expression. Knockdown of LEENE with LNAs induced ICAM1 and VCAM1 mRNA levels in HUVECs. Error bars represent mean±SEM, n=5/group. ANOVA followed by Bonferroni post-test is applied, *p<0.05.

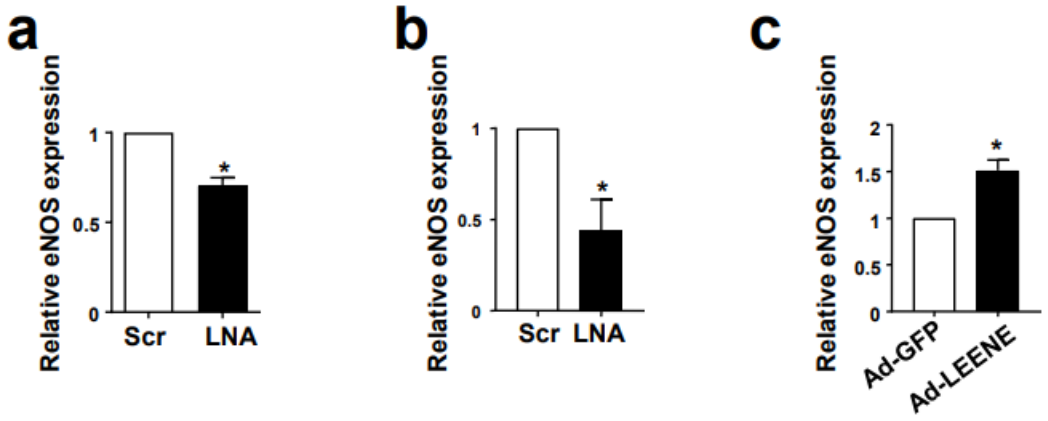


Figure 3.19: Densitometry quantification of Western Blot experiments. (a) Figure 3.16B, under statin treatment; (b) Figure 3.16B, under PS condition; (c): Figure 3.16F. Error bars represent mean±SEM, n=5/group. Student's t-test is applied, *p<0.05.

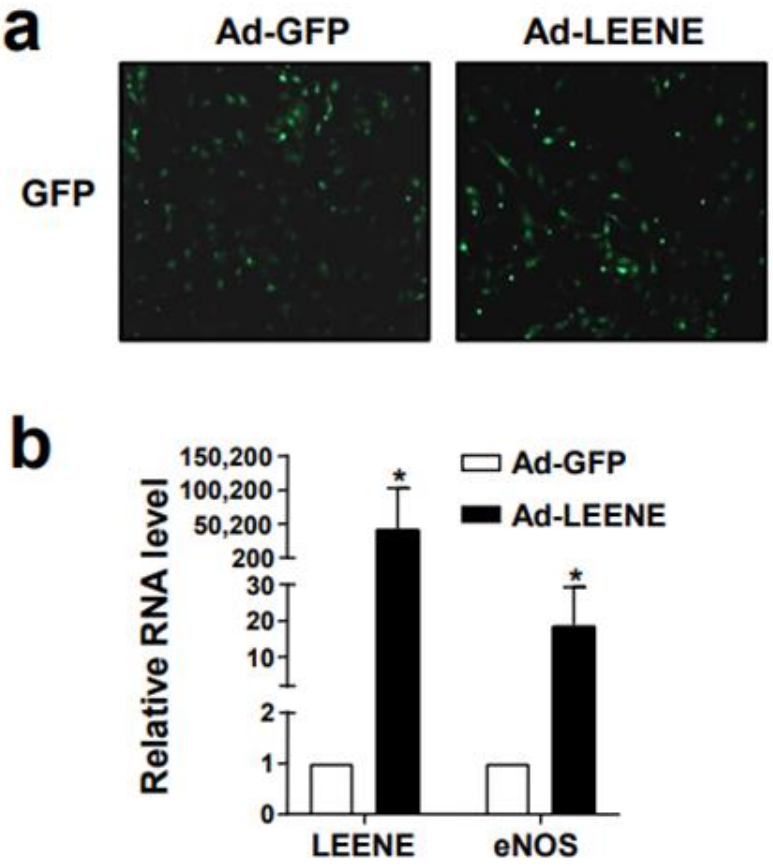


Figure 3.20: Overexpression of LEENE using adenovirus. (a) GFP images from HUVECs infected with Ad-GFP or Ad-LEENE adenovirus. (b) LEENE and eNOS RNA/mRNA in HAoEC infected with Ad-GFP or Ad-LEENE. Error bars represent mean±SEM, n=5/group. Student's t-test is applied, *$p<0.05$.

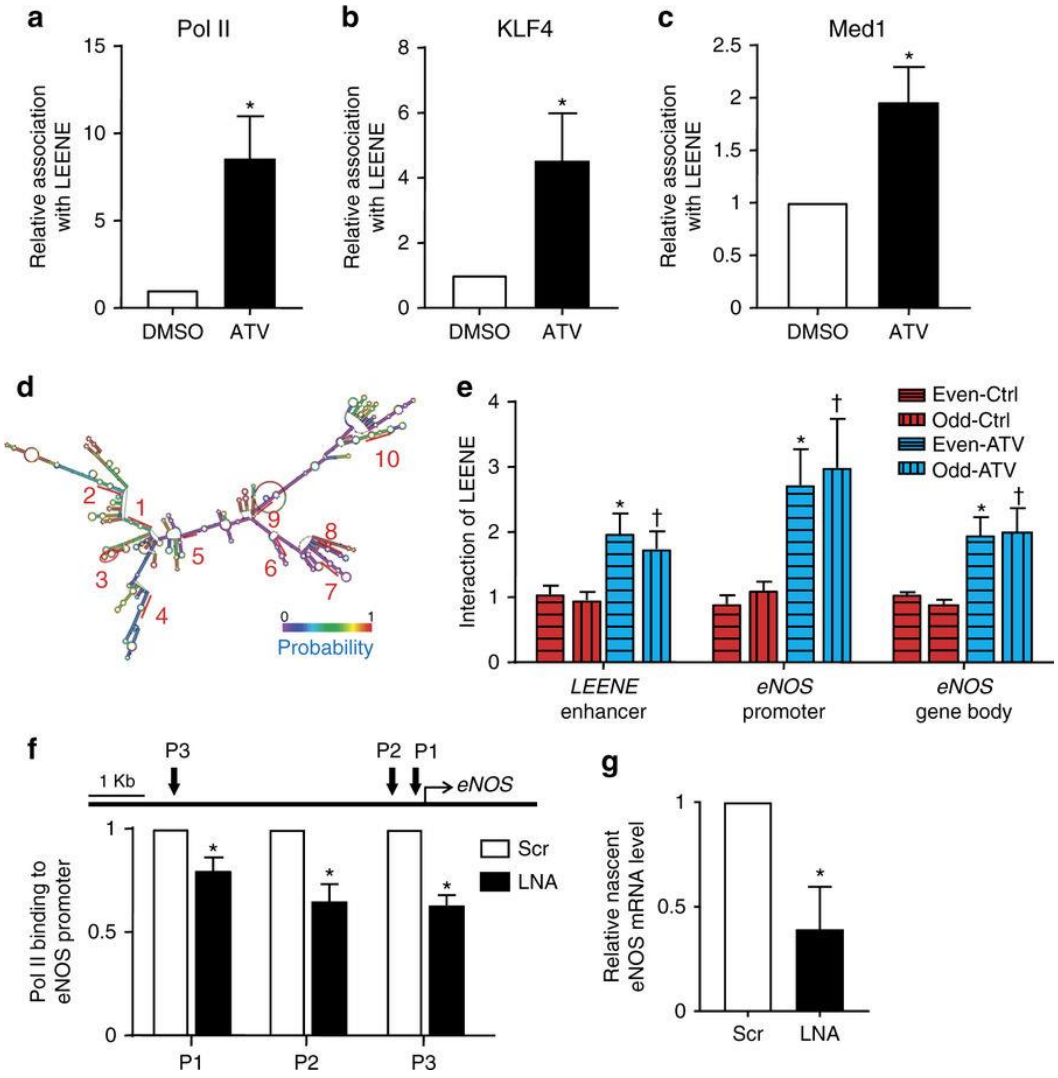


Figure 3.21: LEENE RNA promotes RNA Pol II binding and eNOS transcription. (a–c), (e) HUVECs were treated with ATV for 24 h. The binding of RNA Pol II, KLF4, and MED1 to LEENE RNA was determined by RIP followed by qPCR (a–c). (d) Predicted secondary structure of LEENE RNA based on minimum free energy (MFE) and fragments complementary to ChIRP probes are labeled with numbers 1–10. Color scale shows the probabilities for every nucleotide to hold the structural position. Following ChIRP, interactions between LEENE RNA and respective DNA regions of LEENE and eNOS were detected by qPCR (in e). (f, g) Static ECs were transfected with scramble or LEENE LNA. The binding of RNA Pol II to eNOS promoter was determined by ChIP-qPCR with three primer sets flanking three regions upstream of eNOS TSS (in f). (g) Nascent RNA was captured in static ECs transfected with scramble or LEENE LNA. eNOS mRNA level was detected by qPCR. Data are presented as mean ± SEM, n = 5 in each group. *$p < 0.05$ compared with respective control in each experiment. In (a–c), * denotes $p < 0.05$ compared with DMSO; in (e), * indicates $p < 0.05$ between Ctrl and ATV using even-numbered probes; † denotes $p < 0.05$ between Ctrl and ATV using odd-numbered probes; in (f) and (g), * means $p < 0.05$ between scramble vs. LNA groups. Student's t-test was applied.

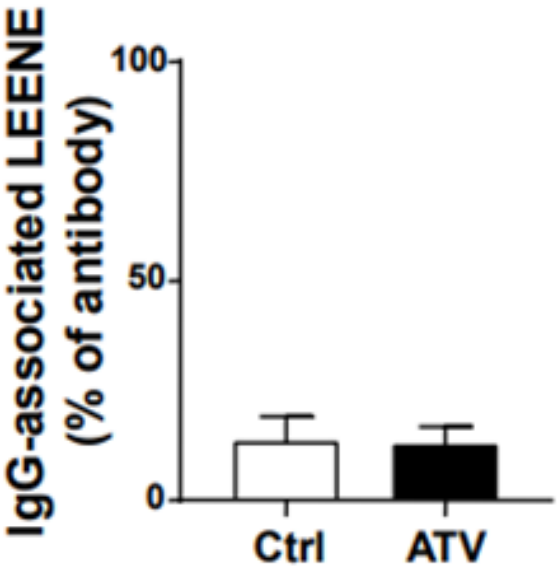


Figure 3.22: IgG isotype control for the RIP assay. Error bars represent mean±SEM, n=5/group. Student's t-test is applied.

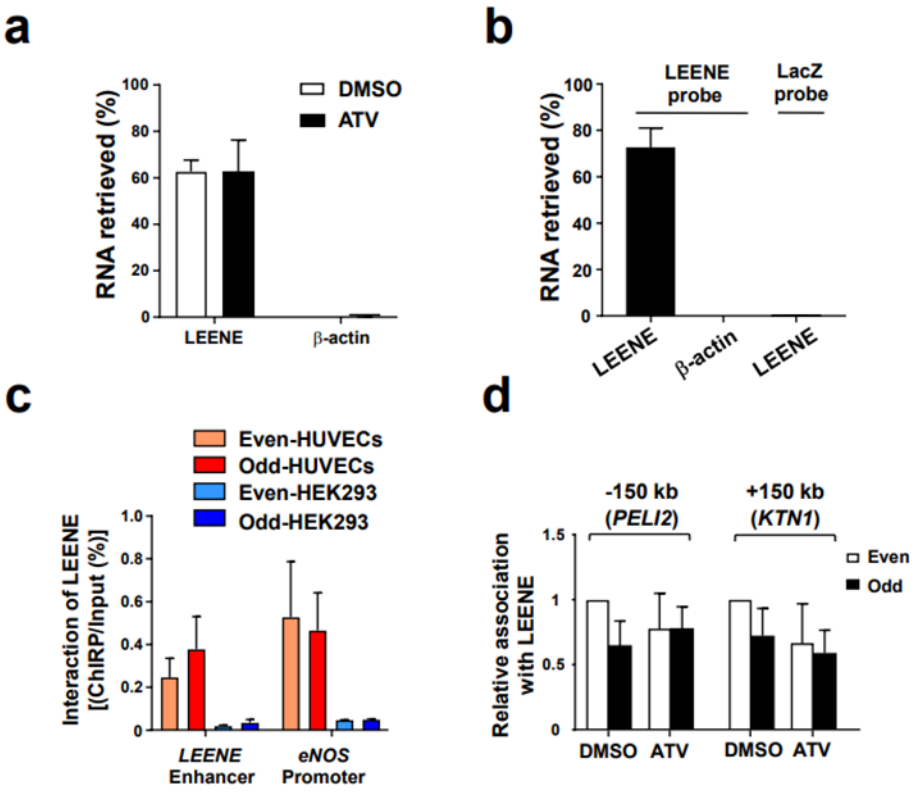


Figure 3.23: Quality and negative controls of LEENE ChIRP assay. (a) RNA retrieval rate using LEENE probe under DMSO or atorvastatin (ATV) treatment, quantified by qPCR. (b) RNA retrieval rate using LEENE or LacZ probe to pull-down LEENE RNA or β-actin under basal condition (DMSO). (c) LEENE ChIRP assay detecting LEENE RNA and LEENE enhancer/eNOS promoter region interactions using HUVECs or HEK293 cells. Error bars represent mean±SEM, n=5/group. (d) LEENE ChIRP qPCR detection of 150 kb up- an downstream regions respectively encoding PELI2 and KTN1. Error bars represent mean±SEM, n=5/group. ANOVA followed by Bonferroni post-test is applied.

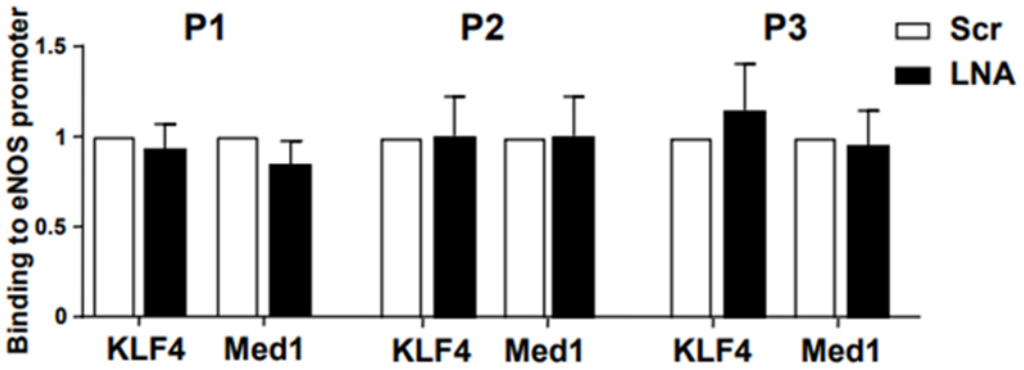


Figure 3.24: Effect of LEENE depletion on KLF4/Med1 binding to eNOS promoter. KLF4 and Med1 were immunoprecipitated from ECs transfected with scramble LNA or LEENE LNA and their bindings to eNOS promoter region were detected by qPCR. Error bars represent mean±SEM, n=5/group. Student's t-test is applied.

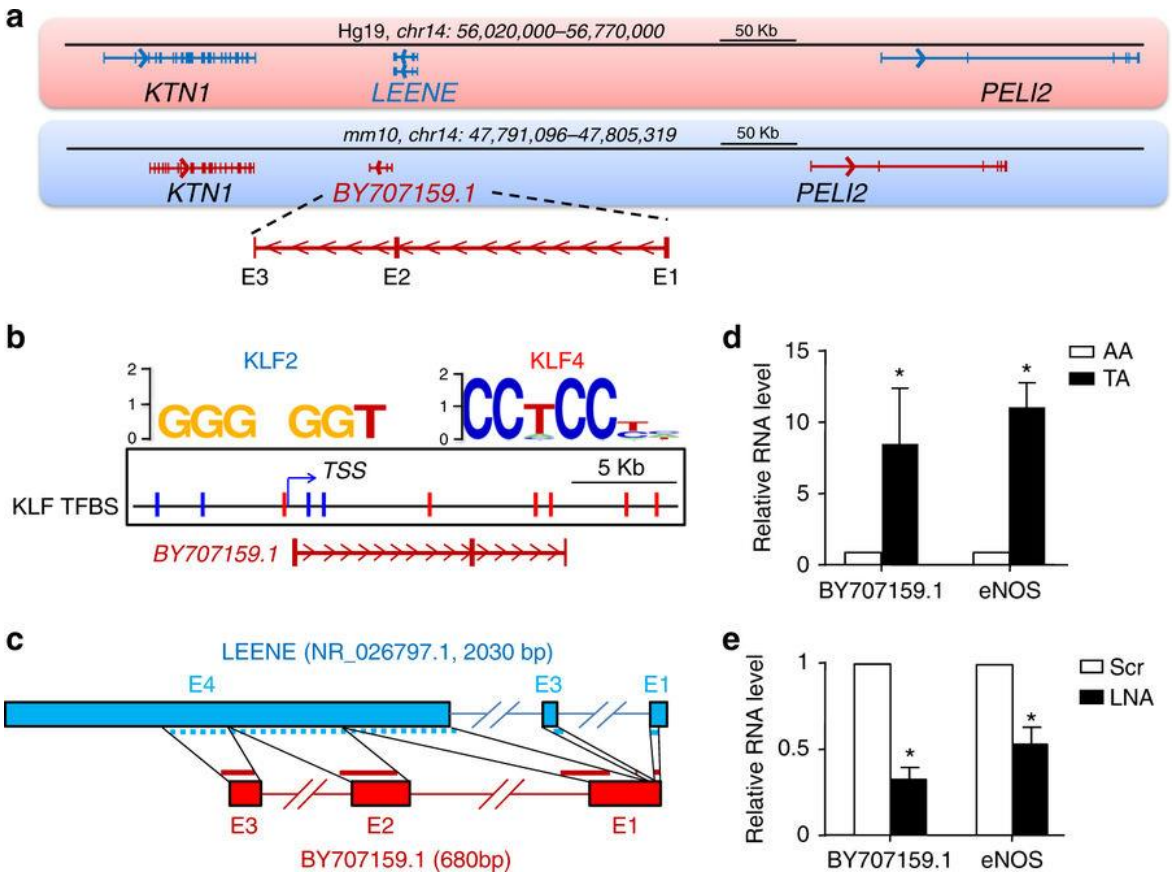


Figure 3.25: LEENE homolog in mouse. (a) Comparison of human LEENE and mouse BY707159.1 loci. (b) Putative KLF2/4 TFBS in DNA region encoding BY707159.1, indicated by blue (for KLF2) and red (for KLF4). (c) Sequence alignment between human LEENE and BY707159.1. (d) Measurement of BY707159.1 RNA levels in thoracic aorta (TA) and aortic arch (AA) using qPCR. (e) qPCR of BY707159.1 and eNOS RNA level in isolated mouse lung ECs transfected with scramble or LEENE LNA. In each experiment, lungs from four animals were pooled for isolation and transfection. Data are average from four independent experiments. Error bars present mean ± SEM. * indicates $p < 0.05$ between AA and TA in (d) and between scramble and LNA in (e) based on Student's t-test.

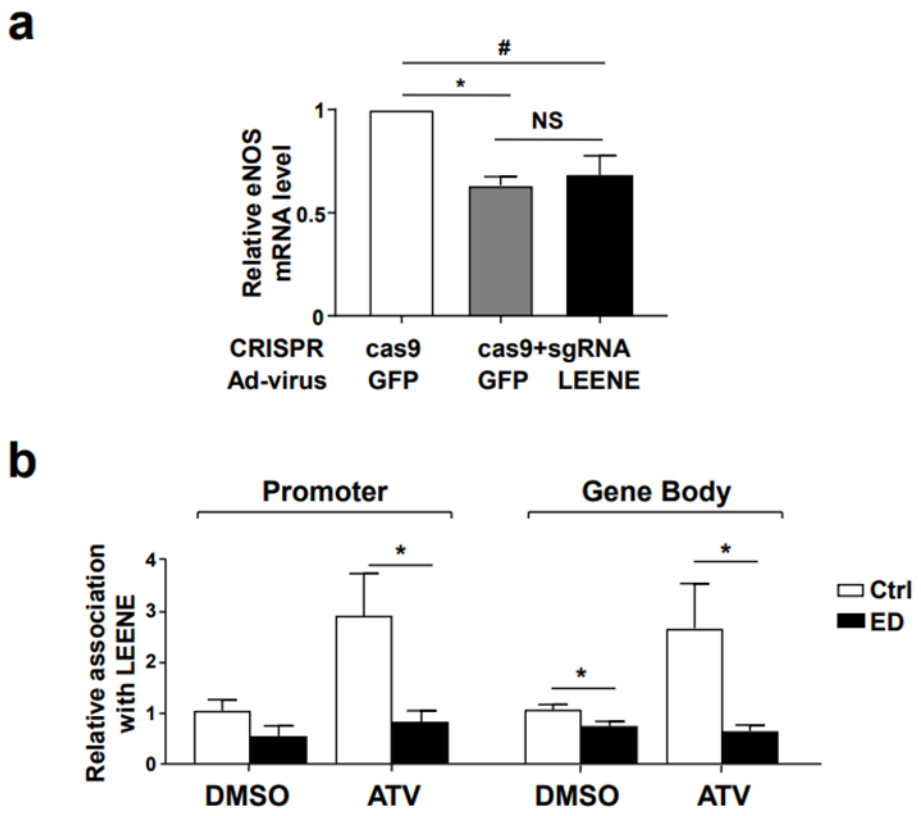


Figure 3.26: LEENE enhancer is prerequisite to LEENE RNA-eNOS locus association and LEENE RNA is not sufficient to enhance eNOS level. (a) HUVECs were transfected with control (Ctrl) cas9 plasmid or Cas9-sgRNA targeting LEENE enhancer region (CRISPR-ED) 24 hr before infection with Ad-GFP or Ad-LEENE for another 48 hr. eNOS mRNA level was quantified by qPCR. (b) ChIRP detection of eNOS genomic locus binding with LEENE RNA in ECs transfected with Cas9 and sgRNA targeting the enhancer of LEENE and treated with DMSO or ATV. Error bars represent mean±SEM, n=5/group. ANOVA followed by Bonferroni post-test is applied, *,#$p<0.05$.

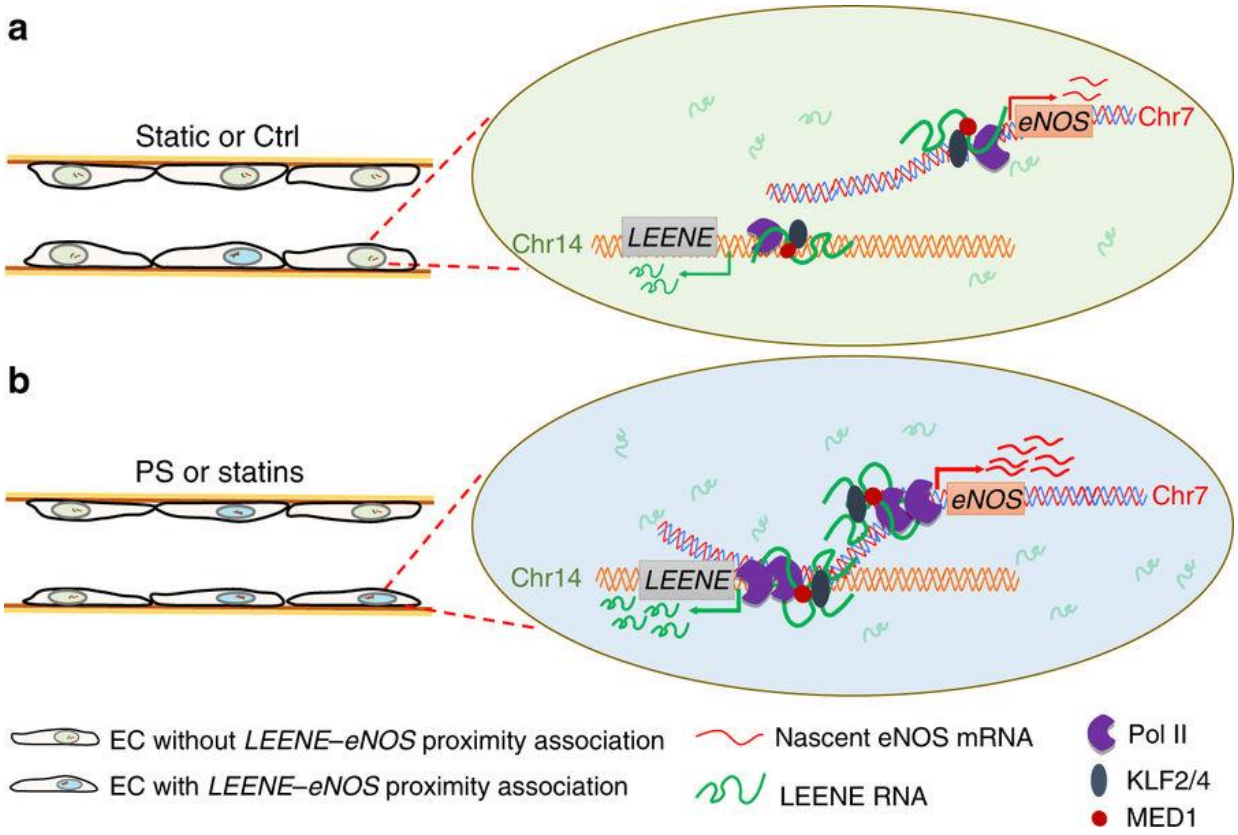


Figure 3.27: Schematic illustration of LEENE–eNOS regulatory mechanism. The LEENE-associated enhancer (located in chr14) forms proximity association with eNOS locus (in chr7) under both static/basal/control (ctrl) (a) and stimulated conditions (PS or statins) (b), but is at a higher probability in the latter condition. In both conditions, KLF2 and KLF4 transcriptionally regulate LEENE and eNOS through binding to TFBS in the promoters of both genes. The LEENE RNA transcripts serve as guides to facilitate RNA Pol II binding to the promoter of eNOS. This enhancer lncRNA-mediated transcriptional regulation positively modulates the nascent eNOS mRNA synthesis to promote endothelial function.

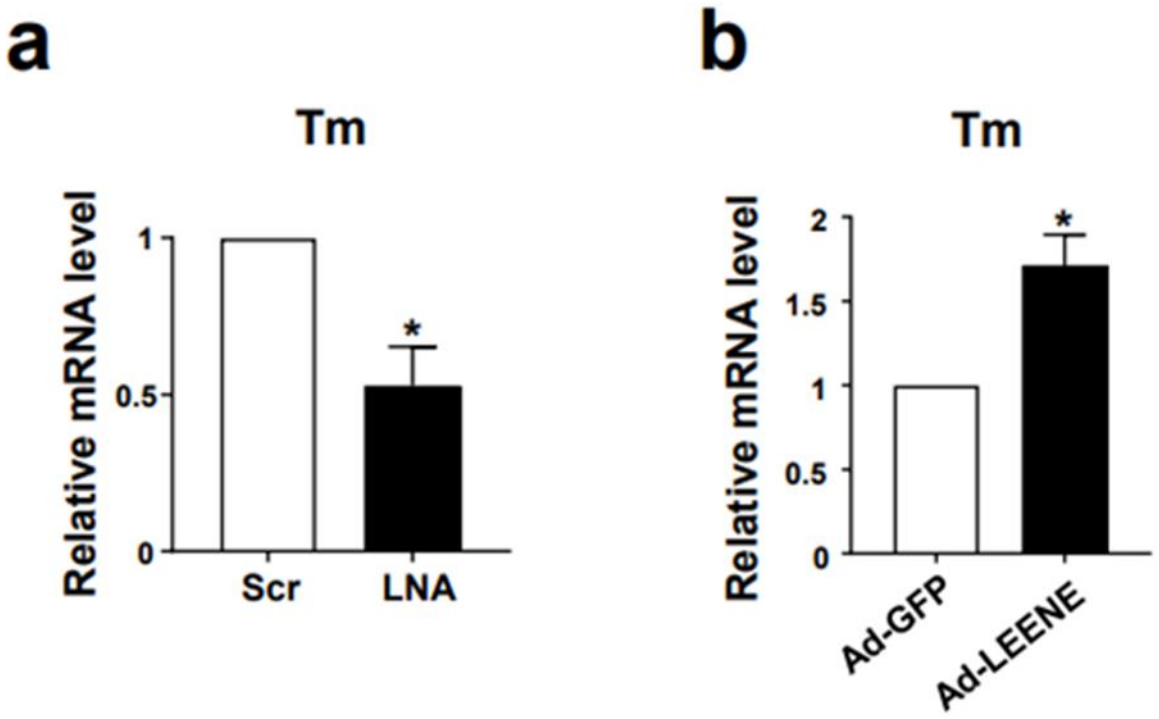


Figure 3.28: Effect of LEENE knock-down or overexpression on Tm mRNA level. Inhibition of LEENE by LNA decreased (a), while LEENE overexpression increased thrombomodulin mRNA level (b). Error bars represent mean±SEM, n=5/group. Student's t-test is applied, *p<0.05.

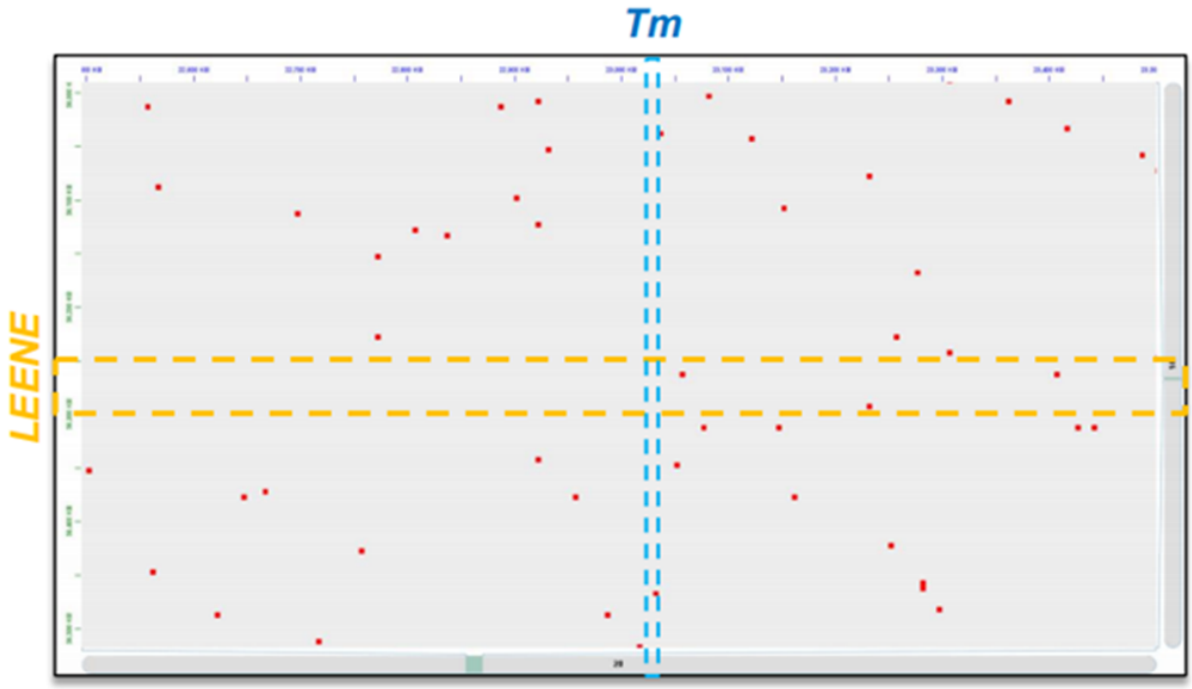


Figure 3.29: Hi-C analysis of LEENE-Tm interaction map in HUVECs. LEENE-Tm interaction map generated from Hi-C data from HUVECs. Red pixels represent interactions between two regions respectively in chr 20 (X axis) and chr 14 (Y axis). The highlighted regions correspond to Tm promoter and LEENE enhancer regions. Note that no red pixel was shown in the overlapped highlighted region.

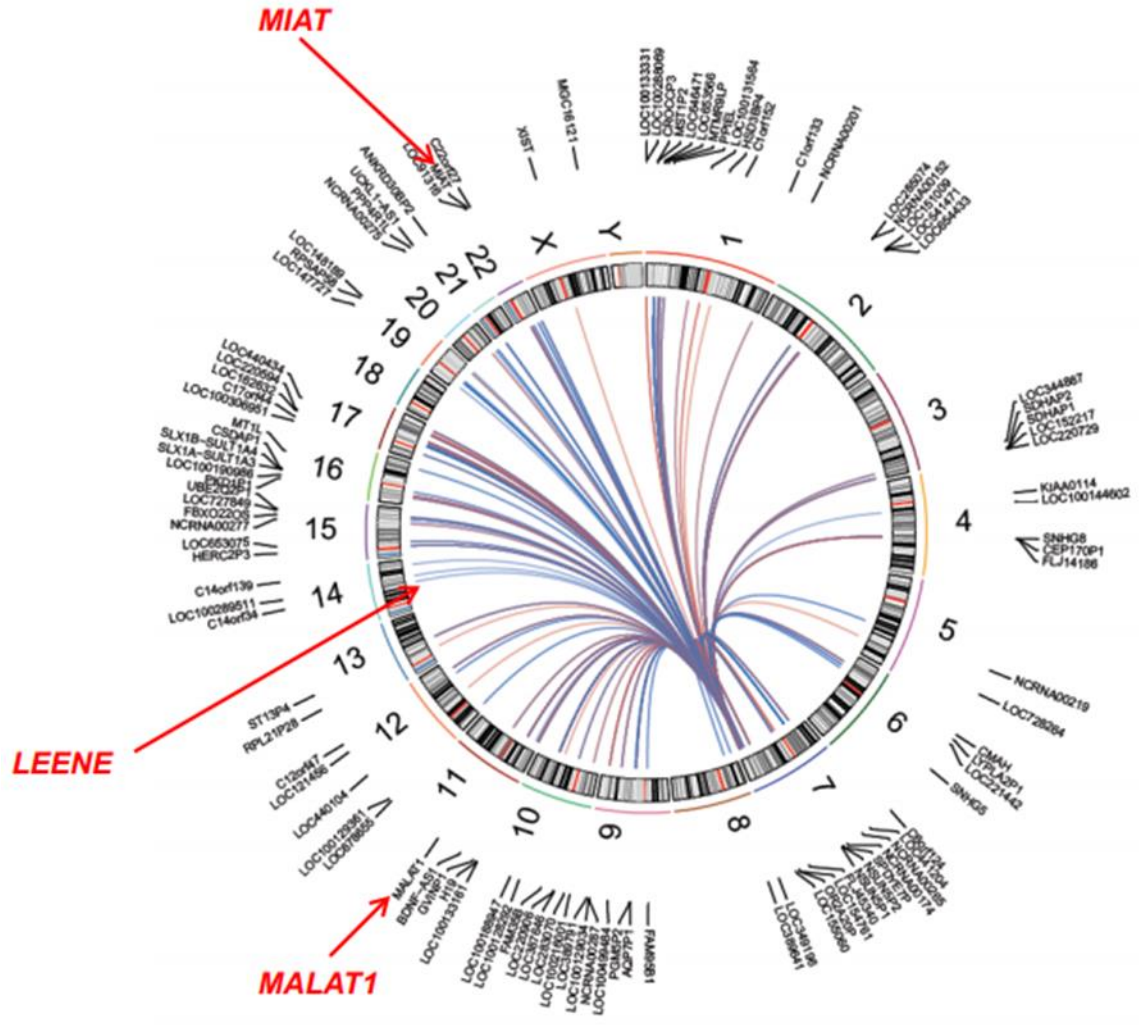


Figure 3.30: Detailed gene information of 4C-seq circoplot. Each line in circoplot depicts a chromosomal interaction between lncRNAs and eNOS bait as revealed by 4C-seq. Blue lines: PS; red lines: OS.

Table 3.1: sgRNA sequences.

sgRNA name	Sequence
ED_5'-1	ATGGGTGCAACACACCAACA
ED_5'-3	GGTGGGGAACATCACACACT
ED_5'-6	GAGGGGGGAGGGATAGCATT
ED_3'-7	GTCATATCTTACATGGCAGA
ED_3'-9	GAGCACAGTCATATCTTACA
CD_5'-1	TACCATAACACTAGACCAGC
CD_5'-2	GCTCTCAGTCATTATATGAT
CD_5'-3	AGTAAAATCACCCCGTGTAG
CD_3'-1	GAAACCCCGAGCCAAACGTA
CD_3'-2	TGTGTTACGGTCTTCAATGG
CD_3'-3	CAGCCTCCATTGAAGACCGT

Table 3.2: LEENE LNA-GapmeRs sequences.

	Sequence	Position (NR_026797)
LNA-1	TTTGATGAGTGAGTCG	218-233
LNA-2	GTTACGGTCTTCAATG	1628-1643
Scramble	AACACGTCTATACGC	

Table 3.3: LEENE ChIRP probe sequences.

Number	Sequence	Position
1	ATAGAATCTTGCTTGGGCAG	170-198
2	CTGAGTGGATTGTAGGTGTT	443-462
3	TCCTGTCTTCTTACTTGTAC	629-648
4	CCCCAAAATCCTTTAAGGTA	927-946
5	TGTCTCTGGGAAGAGGAGAG	1072-1091
6	TGGATGTAAAGACTGGTGCC	1173-1192
7	CTGAGCTGTAGAATCCACAG	1277-1296
8	CTGACATCTCATCAAGGGAG	1377-1396
9	GCTCATCAAGAAGCAGCTAG	1488-1507
10	AGTCACAAGAAGTCCAAGGC	1594-1613

Table 3.4: Human primer sequences.

Name (Human)	Forward	Reverse
qPCR primer		
LINC00520 (LEENE)	TTTCCCTCTTTGGGGTCTCA	GCCCTTTGATGAGTGAGTCG
eNOS	TGATGGCGAAGCGAGTGAAG	ACTCATCCATACACAGGACCC
KLF2	CATGTGCCGTTTCATGTGC	GCAAGACCTACACCAAGAGTTCG
KLF4	CCCACATGAAGCGACTTCCC	CAGGTCCAGGAGATCGTTGAA
VCAM1	GTCAATGTTGCCCCCAGAGA	TTTTCGGAGCAGGAAAGCCC
ICAM1	GTGTCCTGTATGGCCCCCGACT	ACCTTGCGGGTGACCTCCCC
ACTB	CATGTACGTTGCTATCCAGGC	CTCCTTAATGTCACGCACGAT
MALAT1	GTGCTACACAGAAGTGGATTC	CCTCAGTCCTAGCTTCATCA
DANCR	AGGAGTTCGTCTCTTACGTCT	TGAAATACCAGCAACAGGACA
CasC7	GAGGCTGACCTAAACTGCTATC	CCACTACAGGAACTGAACTGAC
TUG1	CTGTGACCCAGAAGAGTTAAG	CATATCCCAGGGACTCAAAC
KTN1	AAGGAAAGGCAGCAACAGGT	CTGACCCTGAAGTTCCAGCC
Tm	ACCTTCCTCAATGCCAGTCAG	CGTCGCCGTTCAGTAGCAA
4C primer		
eNOS 4C Primer	AGGACTCAAGGGTGGGGATC	GCAGGTCAGCAGAGAGACTA
H3K27ac ChIP-qPCR primer		
Region I	GGCCAATGCATTTGACCTCC	GCTTGTTGGCCACCTGATTT
Region II	CGGTCACTTGCCTGACATCT	GGGCACAGTGATCCACTTCA
Region III	ACTGACACTGAGCTCTCCCT	CCCAGCCGGGATTTACCTTT
Region IV	ACCTTTGCTACCGGAACTCG	GAGAGAGAACGTGACGTGGG
ChIP-qPCR primer		
KLF4 ChIP		
eNOS Promoter	GCCCAGCAAGGATGTAGTGA	GTGTGGGGGTTCCAGGAAAT
LEENE P1	CAGCCTCGTACACAGAAGCT	AACTCATTCCATCTTTAACCTGC
LEENE P2	CTGGTTTCAAACTCGTGGCT	AACGTAGATAGGAGGCCAGG
LEENE P3	GATAGGGCCTGGTGAATTGC	TTATGGAAGGAAGGGCAGGG
LEENE P4	CCATGGCTTTGAAACTGCC	GCAAAGACAACCACCCTGTA
Pol II, KLF4, and Med1 ChIP		
eNOS Promoter P1	ATCCCCTGCTACAGAAACGG	AGCCACCAGGGGGTCATAAA
eNOS Promoter P2	GCCCAGCAAGGATGTAGTGA	GTGTGGGGGTTCCAGGAAAT
eNOS Promoter P3	GCCGAACACCAAATCTCCAAC	AGCCCTGCCAAGAATGATGC
ChIRP-qPCR primer		
LEENE Enhancer	CAGCCCTCACATGTATCCCC	GTGCTATGGAAGGTGGCGTA
eNOS Promoter	GCCGAACACCAAATCTCCAAC	AGCCCTGCCAAGAATGATGC
eNOS Gene Body	AACACATAGGCCCTGATTGGG	TGATGATGTGGTCTCATCCCG
150 Upstream (*KTN1*)	AGAGGTGATAGTGAAGAATGACT	CAGAAGACAGGTGAAAGGGT
150 Downstream (*PEIL2*)	GAAAGAGGCTGTGAGGATGA	TTTCAACTCAGCTGCTATGC
CRISPR PCR primer		
5' site	CAATGATAGGCTGGATTAAGAAAAT GTGGC	GATGTTCTGATAAAATCATTGCAAT CTGGGC
3' site	TGTCATGGGAGGGACTGAATGG	GGCCACATATTGTATGATTCCATT TATATGAGATG
Spliced region	CAATGATAGGCTGGATTAAGAAAAT GTGGC	GGCCACATATTGTATGATTCCATT TATATGAGATG

Table 3.5: Mouse primer sequences.

Name (Mouse)	Forward	Reverse
qPCR primer		
BY707159.1 EST	GGACCTCTGGCTAAGGTGAG	TCCTTGCTTGTCCTGTCAGT
eNOS	CACCTACGACACCCTCAGTG	CTTGACCCAATAGCTGCTCAG

Table 3.6: List of PS-Distinct and OS-Distinct genes. PS-distinct genes are statistically significantly (a) upregulated in PS vs ST, (b) downregulated in OS vs ST, (c) downregulated in OS vs PS, and (d) upregulated in 72-hr PS vs ST data from Maleszewska et al, 2016. OS-distinct genes are statistically significantly (a) downregulated in PS vs ST, (b) upregulated in OS vs ST, (c) upregulated in OS vs PS, and (d) downregulated in 72-hr PS vs ST data from Maleszewska et al, 2016.

PS-Distinct Genes
ABLIM3, ACVRL1, ADAM15, ADAMTS4, ADARB1, ALS2CL, AMMECR1L, APOLD1, ARHGEF15, ARRDC2, ASAP2, ATG9B, C1orf21, C1QTNF6, CALML4, CASKIN2, CCDC69, CCM2L, CLCF1, CRTAC1, CXCR7, DENND4B, DGKA, DHH, DLL1, DNASE1L1, DOCK6, DUSP4, EHBP1L1, ENDOD1, ESAM, FAM65A, FOXP1, GCH1, GCKR, GPER, GRK5, HEG1, HR, HSPA12B, HSPA1A, HYAL1, HYAL2, IL3RA, INPP5K, ITGB5, ITPR3, ITSN2, JOSD1, KCNJ12, KIAA1161, KLF13, KLF2, KLF4, LINC00520, LPCAT4, MAP4K2, MN1, MOB2, MYO1C, NAGA, NDRG4, NDST2, NOS3, NPEPPS, PLCD3, PLXND1, PMP22, RAB11A, RECK, RNF213, RUSC1, S100A10, SEMA4B, SIGIRR, SLCO2A1, SPTBN1, SPTBN5, ST3GAL1, ST8SIA6, STARD8, STOM, STXBP1, SULF1, SYNE3, TBC1D1, TBC1D2, TECPR1, TENC1, THBD, TNFAIP1, TOB2, TRAK1, TSC22D3, VANGL1, VEGFA, ZNF467
OS-Distinct Genes
C9orf89, CCZ1, CXCR4, EIF2AK1, ENC1, ETV6, GLIPR1, GMFG, LYPLA2, MAGEF1, MEX3A, PCSK1, STX6, TNFRSF21

References

Atkins GB, Jain MK. Role of Kruppel-like transcription factors in endothelial biology. Circulation research 100, 1686-1695 (2007).

Barajas-Espinosa A, Basye A, Angelos MG, & Chen CA (2015) Modulation of p38 kinase by DUSP4 is important in regulating cardiovascular function under oxidative stress. Free radical biology & medicine 89:170-181.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell 150, 279-290 (2012).

Bolland DJ, King MR, Reik W, Corcoran AE, Krueger C. Robust 3D DNA FISH using directly labeled probes. Journal of visualized experiments : JoVE, (2013)

Chen Z, Lai TC, Jan YH, Lin FM, Wang WC, Xiao H, Wang YT, Sun W, Cui X, Li YS, Fang T, Zhao H, Padmanabhan C, Sun R, Wang DL, Jin H, Chau GY, Huang HD, Hsiao M, Shyy JY. Hypoxia-responsive miRNAs target argonaute 1 to promote angiogenesis. The Journal of clinical investigation 123, 1057-1067 (2013).

Chen Z, Peng IC, Cui X, Li YS, Chien S, Shyy JY. Shear stress, SIRT1, and vascular homeostasis. Proceedings of the National Academy of Sciences of the United States of America 107, 10268-10273 (2010).

Chen Z, Wen L, Martin M, Hsu CY, Fang L, Lin FM, Lin TY, Geary MJ, Geary GG, Zhao Y, Johnson DA, Chen JW, Lin SJ, Chien S, Huang HD, Miller YI, Huang PH, Shyy JY. Oxidative stress activates endothelial innate immunity via sterol regulatory element binding protein 2 (SREBP2) transactivation of microRNA-92a. Circulation. 2015 Mar 3;131(9):805-14.

Chowdhury B, Hemming R, Faiyaz S, Triggs-Raine B. Hyaluronidase 2 (HYAL2) is expressed in endothelial cells, as well as some specialized epithelial cells, and is required for normal hyaluronan catabolism. Histochem Cell Biol. 2016 Jan;145(1):53-66.

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Molecular cell 44, 667-678 (2011).

Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). Journal of visualized experiments : JoVE, (2012).

de la Iglesia N, Veiga-da-Cunha M, Van Schaftingen E, Guinovart JJ, & Ferrer JC (1999) Glucokinase regulatory protein is essential for the proper subcellular localisation of liver glucokinase. FEBS letters 456(2):332-338.

dela Paz NG, Walshe TE, Leach LL, Saint-Geniez M, D'Amore PA. Role of shear-stress-induced VEGF expression in endothelial cell survival. J Cell Sci. 2012 Feb 15;125(Pt 4):831-43.

Devaux Y, Zangrando J, Schroen B, Creemers EE, Pedrazzini T, Chang CP, Dorn GW 2nd, Thum T, Heymans S; Cardiolinc network. Long noncoding RNAs in cardiac development and ageing. Nat Rev Cardiol. 2015 Jul;12(7):415-25.

Duda DG, Kozin SV, Kirkpatrick ND, Xu L, Fukumura D, Jain RK. (2011) CXCL12 (SDF1alpha)-CXCR4/CXCR7 pathway inhibition: an emerging sensitizer for anticancer therapies? Clinical cancer research : an official journal of the American Association for Cancer Research 17(8):2074-2080.

Eid R, Demattei MV, Episkopou H, Augé-Gouillou C, Decottignies A, Grandin N, Charbonneau M. Genetic Inactivation of ATRX Leads to a Decrease in the Amount of Telomeric Cohesin and Level of Telomere Transcription in Human Glioma Cells. Mol Cell Biol. 2015 Aug;35(16):2818-30.

Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, 1237973 (2013).

Ernst J & Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology 28(8):817-825.

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. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43-49.

Espinosa JM. Revisiting lncRNAs: How Do You Know Yours Is Not an eRNA? Molecular cell 62, 1-2 (2016).

Falkenstein KN & Vokes SA (2014) Transcriptional regulation of graded Hedgehog signaling. Seminars in cell & developmental biology 33:73-80.

Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. Cell 155, 606-620 (2013).

Fish JE, Matouk CC, Rachlis A, Lin S, Tai SC, D'Abreo C, Marsden PA. The expression of endothelial nitric-oxide synthase is controlled by a cell-specific histone code. The Journal of biological chemistry 280, 24824-24838 (2005).

Forstermann U, Sessa WC. Nitric oxide synthases: regulation and function. European heart journal 33, 829-837, 837a-837d (2012).

Fulton D, Gratton JP, McCabe TJ, Fontana J, Fujio Y, Walsh K, Franke TF, Papapetropoulos A, Sessa WC. Regulation of endothelium-derived nitric oxide production by the protein kinase Akt. Nature 399, 597-601 (1999).

Gan Y, Shen YH, Wang J, Wang X, Utama B, Wang J, Wang XL. Role of histone deacetylation in cell-specific expression of endothelial nitric-oxide synthase. The Journal of biological chemistry 280, 16467-16475 (2005).

Genasetti A, Vigetti D, Viola M, Karousou E, Moretto P, Rizzi M, Bartolini B, Clerici M, Pallotti F, De Luca G, Passi A. Hyaluronan and human endothelial cell behavior. Connect Tissue Res. 2008;49(3):120-3.

Giorgetti L, Heard E. Closing the loop: 3C versus DNA FISH. Genome Biol. 2016 Oct 19;17(1):215.

Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, Morse M, Engreitz J, Lander ES, Guttman M, Lodish HF, Flavell R, Raj A, Rinn JL. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nature structural & molecular biology 21, 198-206 (2014).

He M, Chen Z, Martin M, Zhang J, Sangwung P, Woo B, Tremoulet AH, Shimizu C, Jain MK, Burns JC, Shyy JY. miR-483 Targeting of CTGF Suppresses Endothelial-to-Mesenchymal Transition: Therapeutic Implications in Kawasaki Disease. Circulation research 120, 354-365 (2017).

Henry WS, Hendrickson DG, Beca F, Glass B, Lindahl-Allen M, He L, Ji Z, Struhl K, Beck AH, Rinn JL, Toker A. LINC00520 is induced by Src, STAT3, and PI3K and plays a functional role in breast cancer. Oncotarget 7, 81981-81994 (2016).

Hergenreider E, Heydt S, Tréguer K, Boettger T, Horrevoets AJ, Zeiher AM, Scheffer MP, Frangakis AS, Yin X, Mayr M, Braun T, Urbich C, Boon RA, Dimmeler S. Atheroprotective communication between endothelial cells and smooth muscle cells through miRNAs. Nature cell biology 14, 249-256 (2012).

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell reports 11, 1110-1122 (2015).

Huang TS, Wang KC, Quon S, Nguyen P, Chang TY, Chen Z, Li YS, Subramaniam S, Shyy J, Chien S. LINC00341 exerts an anti-inflammatory effect on endothelial cells by repressing VCAM1. Physiol Genomics. 2017 Jul 1;49(7):339-345.

Jiang YZ, Jiménez JM, Ou K, McCormick ME, Zhang LD, Davies PF. Hemodynamic disturbed flow induces differential DNA methylation of endothelial Kruppel-Like Factor 4 promoter in vitro and in vivo. Circ Res. 2014 Jun 20;115(1):32-43.

Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. (2011) ConsensusPathDB: toward a more complete picture of cell biology. Nucleic acids research 39(Database issue):D712-717.

Kong X, Chen L, Ye P, Wang Z, Zhang J, Ye F, Chen S. The role of HYAL2 in LSS-induced glycocalyx impairment and the PKA-mediated decrease in eNOS-Ser-633 phosphorylation and nitric oxide production. Mol Biol Cell. 2016 Dec 15;27(25):3972-3979.

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494, 497-501 (2013).

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359 (2012).

Léveillé N, Melo CA, Rooijers K, Díaz-Lagares A, Melo SA, Korkmaz G, Lopes R, Akbari Moqadam F, Maia AR, Wijchers PJ, Geeven G, den Boer ML, Kalluri R, de Laat W, Esteller M, Agami R. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. Nature communications 6, 6520 (2015).

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

Li W, Jiang Z, Li T, Wei X, Zheng Y, Wu D, Yang L, Chen S, Xu B, Zhong M, Jiang J, Hu Y, Su H, Zhang M, Huang X, Geng S, Weng J, Du X, Liu P, Li Y, Liu H, Yao Y, Li P (2015) Genome-wide analyses identify KLF4 as an important negative regulator in T-cell acute lymphoblastic leukemia through directly inhibiting T-cell associated genes. Molecular cancer 14:26.

Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet. 2016 Apr;17(4):207-23.

Lin Z, Kumar A, SenBanerjee S, Staniszewski K, Parmar K, Vaughan DE, Gimbrone MA Jr, Balasubramanian V, García-Cardeña G, Jain MK. Kruppel-like factor 2 (KLF2) regulates endothelial thrombotic function. Circ Res. 2005 Mar 18;96(5):e48-57.

Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, 't Hoen PA, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest AR, Kawaji H; FANTOM consortium. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015 Jan 5;16:22.

Long J, Badal SS, Ye Z, Wang Y, Ayanga BA, Galvan DL, Green NH, Chang BH, Overbeek PA, Danesh FR. Long noncoding RNA Tug1 regulates mitochondrial bioenergetics in diabetic nephropathy. The Journal of clinical investigation 126, 4205-4218 (2016).

Maleszewska M, Vanchin B, Harmsen MC, & Krenning G (2016) The decrease in histone methyltransferase EZH2 in response to fluid shear stress alters endothelial gene expression and promotes quiescence. Angiogenesis 19(1):9-24.

Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic acids research 31(1):374-378.

Michalik KM, You X, Manavski Y, Doddaballapur A, Zörnig M, Braun T, John D, Ponomareva Y, Chen W, Uchida S, Boon RA, Dimmeler S. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. Circulation research 114, 1389-1397 (2014).

Ni CW, Qiu H, Rezvan A, Kwon K, Nam D, Son DJ, Visvader JE, Jo H. (2010) Discovery of novel mechanosensitive genes in vivo using mouse carotid artery endothelium exposed to disturbed flow. Blood 116(15):e66-73.

Pelish HE, Liau BB, Nitulescu II, Tangpeerachaikul A, Poss ZC, Da Silva DH, Caruso BT, Arefolov A, Fadeyi O, Christie AL, Du K, Banka D, Schneider EV, Jestel A, Zou G, Si C, Ebmeier CC, Bronson RT, Krivtsov AV, Myers AG, Kohl NE, Kung AL, Armstrong SA, Lemieux ME1, Taatjes DJ, Shair MD. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature. 2015 Oct 8;526(7572):273-276.
Pozzobon T, Goldoni G, Viola A, & Molon B (2016) CXCR4 signaling in health and disease. Immunology letters 177:6-15.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842 (2010).

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680 (2014).

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends in genetics : TIG 16, 276-277 (2000).

Rivera CM, Ren B. Mapping human epigenomes. Cell. 2013 Sep 26;155(1):39-55.
Riz I, Hawley TS, Hawley RG. KLF4-SQSTM1/p62-associated prosurvival autophagy contributes to carfilzomib resistance in multiple myeloma models. Oncotarget. 2015 Jun 20;6(17):14814-31.

Rochfort KD, Cummins PM. Thrombomodulin regulation in human brain microvascular endothelial cells in vitro: role of cytokines and shear stress. Microvasc Res. 2015 Jan;97:1-5.
Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabási AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M. (2014) A proteome-scale map of the human interactome network. Cell 159(5):1212-1226.

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. (2013) ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic acids research 41(Database issue):D56-63.

Rothschild G, Basu U. Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. Trends Genet. 2017 Feb;33(2):143-154.

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings of the National Academy of Sciences of the United States of America 112, E6456-6465 (2015).

Sangwung P, Zhou G, Nayak L, Chan ER, Kumar S, Kang DW, Zhang R, Liao X, Lu Y, Sugi K, Fujioka H, Shi H, Lapping SD, Ghosh CC, Higgins SJ, Parikh SM, Jo H, Jain MK. KLF2 and KLF4 control endothelial identity and vascular integrity. JCI insight 2, e91700 (2017).

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. The Journal of experimental medicine 199, 1305-1315 (2004).

Shaul PW. Regulation of endothelial nitric oxide synthase: location, location, location. Annual review of physiology 64, 749-774 (2002).

Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58, 221-230 (2012).

St Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. Trends Genet. 2015 May;31(5):239-51.

Tousoulis D, Kampoli AM, Tentolouris C, Papageorgiou N, Stefanadis C. The role of nitric oxide on endothelial function. Curr Vasc Pharmacol. 2012 Jan;10(1):4-18.

Triggs-Raine B, Natowicz MR. Biology of hyaluronan: Insights from genetic disorders of hyaluronan metabolism. World J Biol Chem. 2015 Aug 26;6(3):110-20.

Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A, Aifantis I. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell 158, 593-606 (2014).

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010 Sep 24;39(6):925-38.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome biology 15, R6 (2014).

Viereck J, Kumarswamy R, Foinquinos A, Xiao K, Avramopoulos P, Kunz M, Dittrich M, Maetzig T, Zimmer K, Remke J, Just A, Fendrich J, Scherf K, Bolesani E, Schambach A, Weidemann F, Zweigerdt R, de Windt LJ, Engelhardt S, Dandekar T, Batkai S, Thum T. Long

noncoding RNA Chast promotes cardiac remodeling. Science translational medicine 8, 326ra322 (2016).

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120-124 (2011).

Wang Z, Zhang XJ, Ji YX, Zhang P, Deng KQ, Gong J, Ren S, Wang X, Chen I, Wang H, Gao C, Yokota T, Ang YS, Li S, Cass A, Vondriska TM, Li G, Deb A, Srivastava D, Yang HT, Xiao X, Li H, Wang Y. The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. Nature medicine 22, 1131-1139 (2016).

Watts JK, Corey DR. Silencing disease genes in the laboratory and the clinic. The Journal of pathology 226, 365-379 (2012).

Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, Tao ZF, Song YC, Chen Q, Jiang Q. lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. Circulation research 116, 1143-1156 (2015).

Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature. 2013 Aug 29;500(7464):598-602.

Zavodszky E, Vicinanza M, Rubinsztein DC. Biology and trafficking of ATG9 and ATG16L1, two proteins that regulate autophagosome formation. FEBS Lett. 2013 Jun 27;587(13):1988-96.

Zhang XD, Qi L, Wu JC, & Qin ZH (2013) DRAM1 regulates autophagy flux through lysosomes. PloS one 8(5):e63245.

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. The Journal of clinical investigation 122, 4727-4731 (2012).

Zhou P, Gu F, Zhang L, Akerberg BN, Ma Q, Li K, He A, Lin Z, Stevens SM, Zhou B, Pu WT. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. eLife 6, (2017).

Chapter 4: Development and Application of a Web-Based “omics” Data Analysis and Integration Platform

Introduction

Many tools already exist to statistically analyze “omics” data. All sequencing data comes in the form of discrete count data, which conforms to Poisson and negative binomial statistical models. Several statistical frameworks exist for comparative analysis of count data, including edgeR, DESeq2, and FourCSeq [Schurch et al, 2016; Klein et al, 2015]. Analysis of ChIP-seq data typically involves a “peak calling” step, wherein genomic areas with enriched number of reads are identified as protein-bound regions. Several tools exist for comparative analysis of ChIP-seq data [Steinhauser et al, 2016]. However, the consistency of results between these packages is reported to be low. Furthermore, although peak-calling can identify locations on the genome that constitute “signal” in ChIP-seq data, peaks in similar regions of the genome across multiple data sets may not align very well with each other, making it difficult to conduct a proper 1:1 comparison.

Furthermore, it can be difficult for traditional “wet lab” biologists to analyze their own NGS data. Different NGS analysis packages (of which there are numerous) can produce different results for the same data set, while also demanding high technical programming skills. It remains an open challenge to create useful analysis pipelines that do not impose demands of computational skill on their users. Various web suites and web-based workflows have emerged in order to meet this demand [Li et al, 2016; Sun et al, 2017; Afgan et al, 2016; Dhariwal et al, 2017].

We present D-ChIP, an analysis platform for count data that is designed with data management, statistical analysis, and data integration in mind. Users can securely upload data sets and select which data files correspond to “experimental” and “control” data sets for on-site comparative analysis using DESeq2’s statistical framework. Count data are binned into genome-wide non-overlapping windows, allowing for efficient 1:1 comparative analysis of genomic regions across omics data sets. Users can visualize the features of their data sets

through an on-site genome browser. Users can also download summary files that describe the results of the comparative analysis in binned windows proximal to annotated genes. Overall, these tools provide users with an efficient and streamlined means of comparative analysis of multiple omics data sets.

Implementation - *Methods*

The D-ChIP analysis process takes variableStep WIG files as input and approximates the area under the curve within fixed intervals of chromosomal coordinate space. Figure 4.1 illustrates the concept – Figure 4.1A shows the projection of a graphical file of count data on the UCSC genome browser. This projection generates a curve whose area positively scales with the amount of read coverage. Figure 4.1B illustrates the means of area estimation – WIG files define the number of counts detected per genomic location at a user-defined "span" parameter of genomic locations (default WIG span: 10). The span can be thought of as the width of an individual bar in a bar graph, while the number of counts can be thought of as the height. The area under the curve of a given window in the graph in figure 4.1A can therefore be estimated by summing the areas of these bars within the interval, a basic calculus technique. The result is a series of areas per genomic window, where the size of the window can be user-defined. Figure 4.1C is a post-analysis visualization of 1000-nucleotide window areas along the same region of the genome represented in Figure 4.1A. The shape of the plot is retained between the UCSC genome browser projection and the analyzed data visualization, with the added information of statistical significance. Because these genomic windows are standardized in size, there is no need for a peak-calling or peak-overlapping analysis step.

The D-ChIP algorithm takes advantage of the sequential properties of a WIG file, which report chromosomes in ascending order and chromosomal positions within each chromosome in ascending order. An active window is defined independently of the data, consisting of a

chromosome, a beginning position, and an ending position that is determined by the window size. The data from each wig file is parsed line-by-line in Python, with each line of data occupying its own element in a list. This allows for the simultaneous mapping of all wig files to the active window, while retaining individual information from each data set. The area within each window is the sum of the value of each WIG file line (the “height” of the bar graph) multiplied by the WIG span (the “width” of the bar graph). The final output is a tab-delimited spreadsheet where each row is a genomic window and each column is the value of the area within the genomic window in each provided data set. In order to reduce the file size of the final output, genomic windows that uniformly have area values of 0 across all conditions are not saved.

Once these area estimates are complete, the raw area counts are normalized using the between-sample normalization technique described in DESeq2. Statistical analysis between selected experimental and control data sets is done using DESeq2, where the features of the count data are the genomic windows. All outputs other than the raw area counts are comma-delimited.

Implementation - Structure and Interface

D-ChIP was implemented using Django 1.8 and uses a SQLite database. All scripts are written in Python 2.7, with the exception of the DESeq2 script, which is written in R and executed through Python. Figure 4.2 summarizes the workflow of D-ChIP and the prerequisite navigation. Figure 4.3 provides an outline of the system architecture, while Figure 4.4 offers screenshots of the web interface itself. There are three primary analysis steps when using D-ChIP: Initial data submission, differential expression analysis, and basic data integration.

Initial Data Submission. Users can submit data for analysis by navigating to the “Analyze Data” page linked on the top header. The provided form (as seen in Figure 4.4) prompts the user to define which genome annotation applies to their data (GRCh37, GRCh38,

GRCm38), their desired window size (default: 1000), a name for their submission, and which files they would like to upload for analysis. Only bam and wig files are valid for submission. Users are expected to upload multiple files per submission. At minimum, a user must upload enough data per submission in order to perform a differential expression analysis (a control set and an experimental set). For users with large amounts of data (e.g. time-series data), it is to the user's discretion whether they would prefer to upload portions of their data sets across multiple submissions, or whether they would prefer all of the data to be uploaded in one submission.

Once the files are successfully uploaded, D-ChIP converts any submitted bam files into wig files using samtools [Li et al, 2009]. All wig files are then run through the D-ChIP algorithm, which calculates the "area under the curve" for each window (of user-specified size) across the genome. All data sets in a given submission are normalized together using DESeq2's built-in normalization method. Users are notified by email when the window-binning calculations are complete. They are also notified by email if these calculations fail to complete.

Differential Expression Analysis. In order to perform comparative analysis on their successfully uploaded and windowed data sets, users can navigate to the individual submission page listed on the "Analyze Data" page. All uploaded data sets will be listed within the submission page, with the option to toggle the classification of each data set as "Experiment", "Control", or "Neither". This is designed for the case where a user may have many uploaded data sets in a given submission, but is interested in comparing two subgroups at a time. The selected experimental and control conditions are compared using DESeq2's standard analysis. When conducting DESeq2 analysis, only the selected experimental and control data sets are included in the design matrix and statistical analysis.

In order to reduce computational time and eliminate low-confidence results, some genomic windows are filtered out prior to DESeq2 analysis. Windows are eliminated if their

average normalized count across all conditions in the submission are lower than the window size divided by the wig file step size. This is designed such that the cutoff parameter becomes more stringent when the resolution of the data is low (i.e. broader step sizes). Larger genomic windows are expected, on average, to provide larger area values than smaller genomic windows, and the cutoff adjusts its stringency accordingly. Eliminated windows remain within the analyzed file, but are temporarily removed during the DESeq2 analysis and therefore are not assigned p-values. Users are notified by email when the differential expression analysis is complete. They are also notified by email if these calculations fail to complete.

There are two output files from this analysis. The first output is a larger file that lists a genomic window per row, along with each window's normalized area per compared condition and the statistical output given by DESeq2 analysis. The second output is a condensed file aimed towards emphasizing genomic regions that contain genes – which we call the "Gene Summary" file. Each row corresponds to a different Ensembl Gene ID (annotations downloaded from from BioMart), along with the gene name and chromosome. The spreadsheet also lists "Most significant log2FC" and "Adj. P Value". D-ChIP defines a gene's region as the gene body, the 5000-base upstream region of a gene, and 5000-base downstream region of a gene. "Adj. P Value" indicates the most significant p-value among the genomic windows in a given gene's region, while "Most significant log2FC" indicates the $\log_2$ fold change of the genomic window with the most significant p-value. Lastly, the second output contains a "% significant windows in gene region" column, which indicates the percentage of windows within a gene region that have statistically significant p-values (FDR-adjusted $p < 0.05$).

Furthermore, users have the option to visualize completed analyses on the D-ChIP website. The submission page post-analysis will indicate summary information on the number of windows found to be differentially modified, along with statistics by chromosome. A plotting

tool will allow users to specify an Ensembl Gene ID, a gene name, or chromosomal coordinates (e.g. chr1:100000-102000). D-ChIP will display the normalized area values for each genomic window for each compared condition. If a gene name or Ensembl Gene ID is selected, the plot will display the gene body, 5000 nucleotides upstream, and 5000 nucleotides downstream. The plot will also indicate which genomic windows are differentially modified, showing which windows have raw p-values < 0.05 and which windows have FDR-adjusted p-values < 0.05.

Users are also permitted to upload raw count data in the form of a comma-separated value (.csv) file if they wish to conduct standard differential expression analysis on RNA-seq data using DESeq2. This allows for the use of RNA-seq data sets in downstream integration analysis using D-ChIP, as well as offering a simple means to conduct differential expression analysis on data. Visualization for RNA-seq differential expression is not offered, and the only output file is the standard DESeq2 output file.

Basic Data Integration. Users can access D-ChIP's data integration features by navigating to the "Integrate Data" page linked on the top header. There are two available features – integration by genomic windows and integration by genes.

The "integration by windows" feature allows users to compare different DESeq2 results for commonly significant genomic windows. Only comparisons that have successfully completed differential expression analysis will appear listed on the "Integrate Data" page. The output is a spreadsheet that combines summary statistical data for each comparison, as well as two columns that indicate the frequency across data sets that a genomic window is statistically significant – one column for counting raw p-values, the other for FDR-adjusted p-values. Comparisons can only be selected together if they have identical window sizes, and RNA-seq data analyses cannot be selected as they do not incorporate windows in their analysis.

The “integration by genes” feature allows users to compare different results for commonly significant genes. This uses the “Gene Summary” files for all data sets regardless of window size, as well as the standard DESeq2 results from RNA-seq analysis. The output is similar to the “integration by windows” output, only using genes instead of windows for the comparison. Because of the smaller size of the output file and because of the relative ease of identifying biologically relevant genomic features using genes as compared to windows, some summarizing information on the “integration by genes” analysis is displayed on the data integration page upon analysis completion, including a total number of genes differentially expressed across all selected conditions and a preview of the list of genes.

User-Data Management and Access. D-ChIP is designed to facilitate user-friendly organization of uploaded data while maintaining high security of data. Users must create personal accounts and authenticate their account with a valid email address. Data sets uploaded by a user can only be accessed by that user. Submissions are listed on the “Analyze Data” page as soon as the user’s files complete uploading. All uploaded files are stored on the database in a secure folder specified for the user. All analyzed files on the submission page, such as differential expression analysis results, are also stored on the database for quick user access. If users have already analyzed multiple combinations of data sets within one submission, then they will be able to rapidly toggle between different combinations without redoing the analysis. All results can be downloaded as comma-delimited files. All submissions have the option to be deleted by the user – deleting a submission deletes all associated files with the submission, including both uploaded and analyzed files. Data files generated on the “Integrate Data” page are always generated upon the user’s request, and are not stored in the database long-term.

Example Analysis

As part of the testing process for D-ChIP, the H3K27ac ChIP-seq data set generated by our collaborating labs was analyzed using the D-ChIP pipeline. Gene summary results were sorted by minimum adjusted p-value within the gene region, followed by the percentage of significant windows within the gene region. 221 genes were detected to have 25% or more 1000-nucleotide windows within their genomic region to be differentially acetylated between OS and PS. Among these genes include KLF2 (28.57% differentially acetylated, most significant p-value 1.75×10^{-9}), KLF4 (37.5% differentially acetylated, most significant p-value 3.13×10^{-9}), and LINC00520 (48.15% differentially acetylated, most significant p-value 1.89×10^{-6}). All three of these genes have been discussed extensively in the previous chapters. The recapitulation of these established shear-relevant genes in our H3K27ac ChIP-seq data analysis demonstrates the utility of D-ChIP's methodology, as well as the utility of % differentially modified regions per gene region as a ranking metric.

Furthermore, 6 other genes that were detected as PS-distinct (as described in chapter 3) were detected among these top ranking genes. The detected genes were ADAMTS4, CASKIN2, DHH, DLL1, HSPA12B, and INPP5K. DHH (37.5% differentially acetylated, most significant p-value 4.28×10^{-7}) has been extensively described in chapter 3 as a PS-distinct gene with a proposed mechanism influencing endothelial morphology. ADAMTS4 (32% differentially acetylated, most significant p-value 3.95×10^{-4}) and CASKIN2 (34.62% differentially acetylated, most significant p-value 2.24×10^{-10}) were described in chapter 3 as PS-distinct genes that were found to be previously studied in endothelial cells, but not specifically under conditions of shear. A previous study by Hsu et al connected ADAMTS4 expression to anti-angiogenesis [Hsu et al, 2012]. A 2016 dissertation written by Sarah Beth Mueller of Duke University discusses Caskin2 protein in extensive detail as a suppressor of EC proliferation and migration and an eNOS interactor [Mueller, 2016]. By the filtration

analysis of RNA-seq data described in chapter 3, along with the unbiased analysis of differential acetylation using D-ChIP, ADAMTS4 and CASKIN2 have been identified as excellent candidates for a future shear stress experiment with a clear connection to endothelial phenotype (angiogenesis, proliferation, and eNOS regulation).

Compared to the “percentage of significant windows per gene region” (to be hereafter refered to as “percentage”) metric, the “minimum adjusted p-value” metric offers significantly less specificity. 6267 genes were detected to have a minimum adjusted p-value < 0.05. This can be interpreted to mean that 6267 genes have at least one 1000-nucleotide genomic window within its genomic region that is differentially acetylated. It is unclear if either metric has a clear advantage in pinpointing biologically meaningful results. Although a gene scoring highly with the percentage metric would intuitively have a higher likelihood of being a true positive result due to the higher frequency of statistically significant windows, the percentage metric is also biased against longer genes, whose genomic regions feature a larger number of windows. Furthermore, there are some mechanosensitive genes that are known to be epigenetically regulated are not detected with the percentage metric but are detected by the p-value metric. For example, HDAC3, induced by OS, has been reported to repress eNOS expression through deacetylation [Dunn et al, 2015]. This is accurately reflected in the acetylation data analysis, with NOS3’s minimum adjusted p-value being 0.014 with an OS vs PS log fold change of -1.15, indicating that NOS3 is differentially acetylated and down-acetylated in OS vs PS. However, NOS3 only has a percentage metric of 8.82%, being the 1381st-highest percentage in the list. This implies that NOS3, a mechanosensitive gene of established physiological importance, would have likely been overlooked, had the percentage metric been used exclusively.

One effective strategy is to find the intersection of statistically significant features across data sets, thereby incorporating data integration while also producing a result whose

size is a compromise between the two metrics of feature selection available via D-ChIP. There are 3216 differentially expressed genes in the OS vs PS comparison in RNA-seq (hour 24). There are approximately 1148 genes that compose the intersection between these DE genes and the 6267 differentially acetylated genes from the D-ChIP output. This comparison can easily be done thanks to D-ChIP's gene summary output, which allows for a 1:1 comparison for other analyses based around genes, such as RNA-seq data. Pathway enrichment analysis of this gene set via CPDB reveals enrichment of VEGF signaling pathways, extracellular matrix organization, and focal adhesion.

Discussion

The D-ChIP web suite offers an approach to the analysis of count data that is standardized to the genome. The web suite's user-friendly interface allows for accessibility across many academic disciplines, particularly those that do not involve heavy computational or programming training. D-ChIP differentiates itself from other online analysis tools by introducing its window-based analysis and being designed with simplicity in mind. Implementing the genomic window-based analysis on multiple types of count data for the purpose of downstream data integration is a simple and novel approach, and D-ChIP excellently handles this analysis. This suite therefore offers a substantially useful resource to biologists of all disciplines.

Acknowledgements

Chapter 4 is a modified presentation of material currently being prepared for submission for publication as "D-ChIP: An accessible online suite for the comparison of broad-peaked ChIP-seq data sets" by Nassim E Ajami, Mano R Maurya, Shakti Gupta, and Shankar Subramaniam. The dissertation author is the primary author of this material.

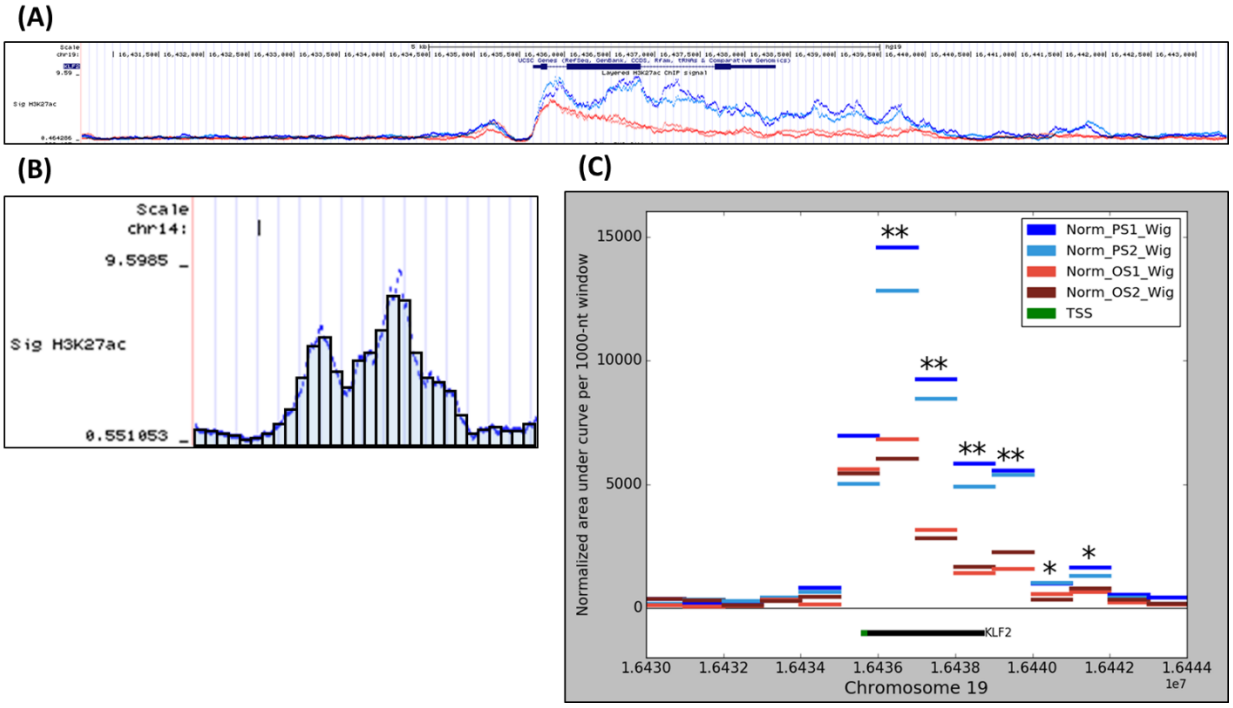


Figure 4.1: Visual representation of the D-ChIP analysis method. (A) Screenshot of standard visualization of histone ChIP-seq data over the KLF2 gene in UCSC genome browser, with different ChIP-seq conditions outlined in shades of blue and red. (B) Visual mockup of the D-ChIP windowing process, which will associate height values for steps along the histone same histone ChIP-seq data set and sum the area values for windows of a user-specified length. (C) Screenshot of the D-ChIP suite's visualization of the same histone ChIP-seq data over the KLF2 gene, showcasing lengths of windows and associated statistical significance within windows across different conditions. One asterisk (*) indicates a raw p-value < 0.05, while two asterisks (**) indicates an FDR-corrected p-value < 0.05.

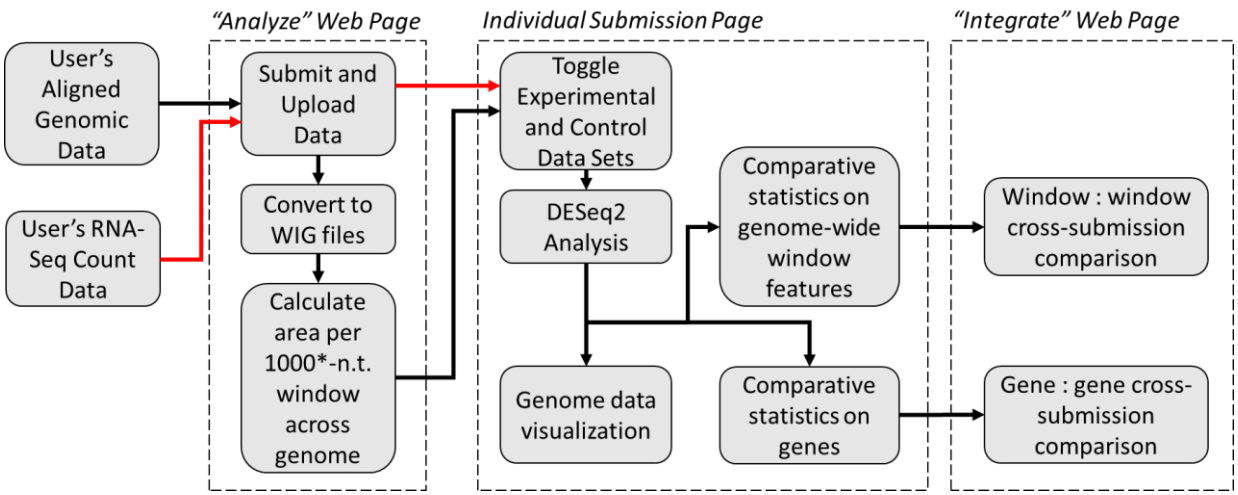


Figure 4.2: Summary of analysis pipeline and location of the website where each analysis step occurs. Standard D-ChIP pipeline is delineated with black arrows. D-ChIP pipeline when using RNA-seq data (which uses DESeq2 analysis using the standard gene features) bypasses the window-generating step, which is indicated with the red arrows.

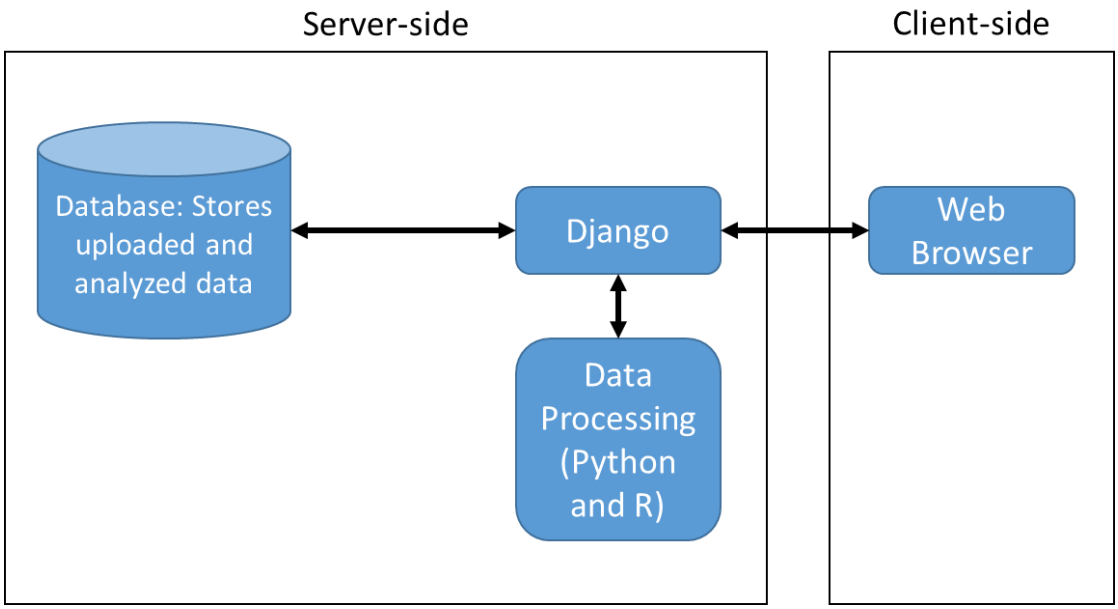


Figure 4.3: D-ChIP systems architecture.

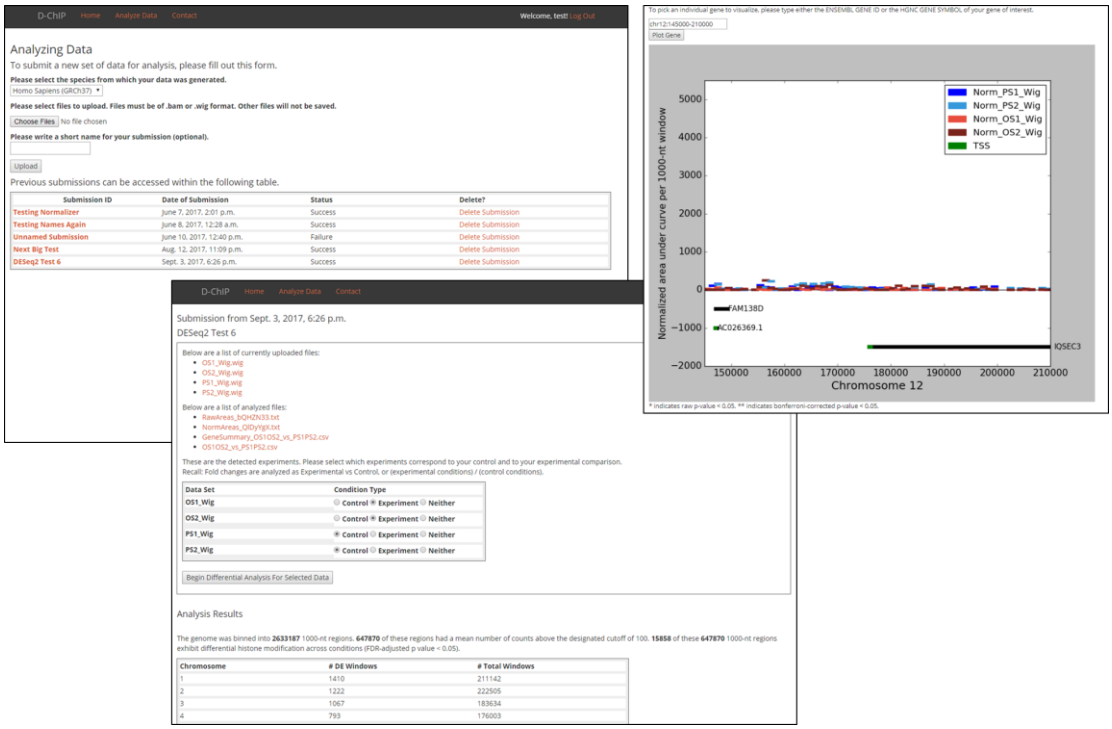


Figure 4.4: Screenshots of D-ChIP's analysis pages. These screenshots highlight the organization of submitted data sets, the ability to toggle between different experimental and control conditions, the visualization capabilities, and the ability to download any and all analyzed files.

References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W3-W10.

Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 2017 Jul 3;45(W1):W180-W188.

Dunn J, Simmons R, Thabet S, Jo H. The role of epigenetics in the endothelial cell shear stress response and atherosclerosis. Int J Biochem Cell Biol. 2015 Oct;67:167-76.

Hsu YP, Staton CA, Cross N, Buttle DJ. Int J Exp Pathol. Anti-angiogenic properties of ADAMTS-4 in vitro. 2012 Feb;93(1):70-7.

Klein FA, Pakozdi T, Anders S, Ghavi-Helm Y, Furlong EE, Huber W. FourCSeq: analysis of 4C sequencing data. Bioinformatics. 2015 Oct 1;31(19):3085-91.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

Li W, Richter RA, Jung Y, Zhu Q, Li RW. Web-based bioinformatics workflows for end-to-end RNA-seq data computation and analysis in agricultural animal species. BMC Genomics. 2016 Sep 27;17(1):761.

Mueller, S. (2016). Characterizing the Role of the Previously Undescribed Protein Caskin2 in Vascular Biology. PhD. Duke University.

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016 Jun;22(6):839-51.

Steinhauser S, Kurzawa N, Eils R, Herrmann C. A comprehensive comparison of tools for differential ChIP-seq analysis. Brief Bioinform. 2016 Nov;17(6):953-966.

Sun X, Pittard WS, Xu T, Chen L, Zwick ME, Jiang X, Wang F, Qin ZS. Omicseq: a web-based search engine for exploring omics datasets. Nucleic Acids Res. 2017 Jul 3;45(W1):W445-W452.

Chapter 5: Challenges, Conclusions, and Future Directions

I have presented a systems biology approach to modeling the regulation of endothelial cells under shear stress using multiple types of “omics” data. We have demonstrated how the integration of “omics” data sets and literature curation can lead to the development of novel and complex hypotheses relevant to endothelial response to shear. As atherosclerosis research continues to advance, it is highly likely that it will significantly utilize multi-“omics” data in order to better understand the pathogenesis of the disease. As we were able to show with LINC00520, it is similarly likely that future studies will be able to identify novel mechanosensitive genes and biomolecules, offering novel potential targets for drugs that can improve vascular health. However, this approach is still relatively nascent compared to the more traditional approaches to biology, and the field will have to address several challenges over the course of its development.

Challenges And Future Directions – Biological

All experimental designs have their limitations, and the design used in this thesis is no exception. Although we were able to provide substantial insights to the onset of atherosclerosis through the work presented in this thesis, future studies could incorporate modified experimental designs that could reveal further novel insights.

It is well known that biological phenomena manifest at different time scales between *in vitro* and *in vivo* studies. The formation of atherosclerosis takes decades to manifest, typically beginning in patients at a young age and causing vascular complications in middle age. Our *in vitro* studies examine the effects of shear stress over the course of 24 hours. It remains unclear if there are more complex molecular mechanisms occurring in endothelial cells when exposed to shear at a time period that is beyond the scope of our data sets. A future study may incorporate longer time scales in order to examine the continued development of endothelial phenotypes under OS and PS.

There is substantial evidence that OS and PS can be used to represent disturbed flow and laminar flow *in vitro*. However, in human physiology, disturbed flow can be more complex than what is simulated in the flow chamber. Disturbed flow does not simply include the reciprocating flow that causes oscillatory shear stress, but also encompasses the phenomena of flow separation, flow reattachment, recirculation eddies, and other complex flow patterns. In the future, endothelial phenotypes under more complex flow conditions need to be explored in a controlled manner. With the advent of 3D printing technology, it will likely be possible in the future to devise custom flow chambers that more closely emulate physiological geometries and flow dynamics.

The experimental design in this thesis explored the response to shear stress by endothelial cells and its pertinence to atherosclerosis. Understanding the endothelium's importance in atherosclerosis as a cellular monolayer exposed to hemodynamic forces is essential to understanding the disease. However, atherosclerosis is a complex disease that also involves leukocytes, smooth muscle cells, macrophages, platelets, and other cellular factors. Their mechanisms of pathophysiological responses can depend on endothelial response in some cases – for example, endothelial cells can secrete factors such as nitric oxide and prostacyclin, which are then received by smooth muscle cells and promote a quiescent phenotype within smooth muscle cells [Zhou et al, 2014]. It is possible that some of the devised mechanistic hypotheses about endothelial response to shear is simply a page in an incomplete story – the hypothesized mechanism, upon activation, may have consequences that span beyond endothelial cells alone. Future studies may glean novel insights into the pathogenesis of atherosclerosis through using an experimental design that used multiple relevant cell types. Such a future "omics" study could simultaneously interrogate the transcriptomic profiles of multiple cell types during shear exposure.

There are several types of endothelial cells. Our study uses human umbilical vascular endothelial cells, which are derived from umbilical cords and are a well-established model system for studying the general endothelium [Jiménez et al, 2013]. However, endothelial cells from different anatomical locations are known to vary considerably in structure and in function. Venous endothelial cells, for example, show different patterns in gene expression compared to arterial endothelial cells. Human aortic endothelial cells, human dermal microvascular endothelial cells, and lymphatic endothelial cells are examples of the diversity of endothelial cells, and it cannot be expected that each of these endothelial cells respond identically to the same stimuli. Our study, while using a cell type with substantial precedence of use in the laboratory to study the endothelium, cannot perfectly describe all endothelial behavior under shear stress. A future study could examine the effects of shear stress of an endothelial cell type other than HUVEC.

Challenges – Technical and Systemic

Although “omics” data are very potent sources of biological information, proper analysis of this data can require finesse. Public data sets from disparate experiments may have unintended variation due to differences in experimental protocol and handling of samples (so-called “batch effects”), making it difficult to directly compare data sets from different sources [Conesa et al, 2016]. Before any meaningful information can be extracted from an “omics” data set, the data set also must be checked for quality control, including quality of sequencing reads and reproducibility among replicates.

Two crucial factors in any “omics” data set are the sequencing depth of the data set and the number of replicates performed for each condition. Sequencing depth in RNA-seq is especially important for detecting genes with low expression (which includes lncRNAs), and replicates are crucial for increasing the confidence of any observation of differential expression, particularly when the difference in gene expression is not dramatic. A truly

comprehensive RNA-seq analysis – one that generates data capable of providing useful information on low-expressing genes, ensuring a thorough interrogation of the transcriptome – requires approximately 80 million reads as its sequencing depth [Sims et al, 2014]. For ChIP-seq and Hi-C datasets, the estimation is even higher – saturation for ChIP-seq data sets are estimated to begin at approximately 110 million reads, and Hi-C data easily requires hundreds of millions of reads [Sims et al, 2014]. Furthermore, a 2016 study by Schurch et al on 48-replicate RNA-seq data found that at least six to twelve biological replicates are necessary for a comprehensive identification of significantly differentially expressed genes [Schurch et al, 2016]. With only three biological replicates, Schurch et al found that a majority of their tested RNA-seq analysis tools found only 20-40% of known DE genes in the data set.

In our own RNA-seq study, each condition per time point averaged approximately 30 million reads and only had two replicates. There are practical reasons for the chosen experimental design and its deviation from the previously described “ideal” number of replicates and read counts. First and most significantly, the cost of producing “omics” data sets is still prohibitive. Our experimental design involved the generation of RNA-seq data for two conditions across ten time points, plus one “static” condition at one time point. With two replicates, this amounts to 42 individual RNA-seq samples. It would cost approximately $5000 in order to run 40 RNA-seq samples with current technology and available infrastructure. A third replicate for all conditions and for all time points would have led to a 1.5-fold increase in this cost.

Secondly, the questions being explicitly posed in the experiment arguably do not necessitate a greater sequencing depth or number of replicates. Indeed, a study that explicitly kept its scope focused on mechanisms, particularly in a time series experiment where neighboring time points can offer information on the quality of the signal in the data, will not need as much resolution as a study that explicitly aims to detect low-expressing genes.

Furthermore, even if the data set with more sequencing depth/replicates would offer greater potential to answer other biological questions in the long term, it may not be necessary to answer a particular question of interest in the short term. Although generating fewer replicates lowers the statistical power in data analysis, it also allows for the reservation of funds for potential validation experiments. Such experiments typically begin with established protocol such as quantitative polymerase chain reaction (qPCR) experiments, which can very cheaply validate the expression for a particular gene in RNA-seq data. In such a scenario, 2-replicate RNA-seq data may only be able to identify genes of interest with middling statistical power, but such genes can be further screened with literature review and then cheaply screened and validated with qPCR before moving on to validating more complex mechanisms pertaining to those genes. This 2-replicate scenario would still be cheaper and more potentially impactful than merely identifying genes of interest from triplicate (or more) RNA-seq data alone with very high statistical confidence. This poses a crucial cost-benefit question that weighs statistical power of "omics" data – and the potential novelty that its conclusions can offer – against the ability to efficiently answer present biological questions for which there is already some prior knowledge.

These decisions are representative of systems biology's growing pains, where "omics" data sets are getting cheaper to generate, but are not yet cheap enough to be trivially generated at ideal quality. There are many significant recent publications in high profile journals that utilized 2-replicate (and in some cases single replicate) "omics" data sets [Velasco et al, 2017; Pillai et al, 2017; Dierickx et al, 2017; Madak-Erdogan et al, 2016; Liu et al, 2017]. Costs are still prohibitive enough where researchers must carefully consider whether their differential expression studies should emphasize greater sequencing depth or greater replication [Liu et al, 2014]. All of this amounts to limiting the potential of "omics" data sets for the sake of answering specific short-term questions, with little to no explicit consideration for

how the data could be useful many years in the future to hitherto unimagined questions in biology. This may ultimately be a systemic issue, where the current incentives for scientists do not exclusively encourage maximizing the scientific value of their research [Higginson et al, 2016].

Although these incentive structures are unlikely to change, it is highly likely that sequencing costs will continue to decrease as sequencing technologies improve. These lower costs may encourage future researchers to generate higher quality data sets – not because they are necessary to answer their immediate scientific question, but because they can do so without significant consequences for their budgets. For a field so dependent on finding patterns in large sets of data, the accumulation of high-quality public “omics” data sets in accessible, scalable, well-maintained databases is essential to the growth of systems biology as a discipline.

Conclusions

High-throughput technologies can provide novel insights on the onset of atherosclerosis. Due to the diversity of omics data sets, these insights span the transcriptional, epigenetic, and proteomic modulations in cellular systems relevant to shear stress response and to atherosclerosis. Bioinformaticians, experimental biologists, and medical doctors must work in unison in order to effectively characterize new and effective interventions for atherosclerosis.

In this thesis, I utilized multiple types of data sets from high-throughput technologies in order to construct a systems level model of endothelial response to shear stress as it pertains to the onset of atherosclerosis, as well as to define novel mechanisms by which identified mechanosensitive genes may contribute to endothelial response to shear stress. The data set most central to my work was a ten time point RNA-seq data set, which was used to provide information on the development of physiological and pathophysiological phenomena over time,

as well as the interdependence of these phenomena. I have also devised a platform with which future studies utilizing multiple “omics” data sets will be facilitated, permitting a more rapid generation of systems analyses. Taken together, this systems approach has the potential to greatly contribute to our understanding of the onset of atherosclerosis and the underlying molecular and cellular mechanisms.

References

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016 Jan 26;17:13.

Dierickx P, Vermunt MW, Muraro MJ, Creyghton MP, Doevendans PA, van Oudenaarden A, Geijsen N, Van Laake LW. Circadian networks in human embryonic stem cell-derived cardiomyocytes. EMBO Rep. 2017 Jul;18(7):1199-1212.

Higginson AD, Munafò MR. Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. PLoS Biol. 2016 Nov 10;14(11):e2000995.

Jiménez N, Krouwer VJ, Post JA. A new, rapid and reproducible method to obtain high quality endothelium in vitro. Cytotechnology. 2013 Jan;65(1):1-14.

Liu Y, Chen S, Wang S, Soares F, Fischer M, Meng F, Du Z, Lin C, Meyer C, DeCaprio JA, Brown M, Liu XS, He HH. Transcriptional landscape of the human cell cycle. Proc Natl Acad Sci U S A. 2017 Mar 28;114(13):3473-3478.

Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics. 2014 Feb 1;30(3):301-4.

Madak-Erdogan Z, Kim SH, Gong P, Zhao YC, Zhang H, Chambliss KL, Carlson KE, Mayne CG, Shaul PW, Korach KS, Katzenellenbogen JA, Katzenellenbogen BS. Design of pathway preferential estrogens that provide beneficial metabolic and vascular effects without stimulating reproductive tissues. Sci Signal. 2016 May 24;9(429):ra53.

Pillai ICL, Li S, Romay M, Lam L, Lu Y, Huang J, Dillard N, Zemanova M, Rubbi L, Wang Y, Lee J, Xia M, Liang O, Xie YH, Pellegrini M, Lusis AJ, Deb A. Cardiac Fibroblasts Adopt Osteogenic Fates and Can Be Targeted to Attenuate Pathological Heart Calcification. Cell Stem Cell. 2017 Feb 2;20(2):218-232.e5.

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 2016 Jun;22(6):839-51.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014 Feb;15(2):121-32.

Velasco S, Ibrahim MM, Kakumanu A, Garipler G, Aydin B, Al-Sayegh MA, Hirsekorn A, Abdul-Rahman F, Satija R, Ohler U, Mahony S, Mazzoni EO. A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. Cell Stem Cell. 2017 Feb 2;20(2):205-217.e8.

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.