Defining endothelial cell functional heterogeneity and plasticity using single-cell RNA-sequencing

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013060">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013060</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Defining endothelial cell functional heterogeneity and plasticity using single-cell RNA-sequencing

A dissertation presented

by

Aditya Sreemadhav Kalluri

to

The Committee on Higher Degrees in Biophysics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biophysics

Harvard University

Cambridge, Massachusetts

June 2019
Defining endothelial cell functional heterogeneity and plasticity using single-cell RNA-sequencing

Abstract

The state of the endothelium defines vascular health – ensuring homeostasis when intact and driving pathology when dysfunctional. The endothelial cells (ECs) which comprise the endothelium are necessarily highly responsive and reflect their local mechanical, biological, and cellular context, making a concrete definition of endothelial state elusive. Heterogeneity is the hallmark of ECs and has long been studied in the context of structural and functional differences between organs and vascular beds. More recently, local heterogeneity within vessels has also been demonstrated. Understanding the impact of local heterogeneity on EC function in health and disease requires complete description of the heterogeneous states occupied by ECs under physiologic conditions and methods for defining the components of and variability in plasticity in EC state in response to disease-relevant stimuli. Recent technological advances in transcriptional characterization with single-cell resolution including single-cell RNA-sequencing (scRNA-seq) enable high-resolution study of the EC response to a range of stimuli.

The premise of this work is that by applying scRNA-seq we can more precisely define individual EC state in terms of functional correlates and differentiate heterogeneous EC responses to pathologically relevant stimuli singly and in combination. Thus, using droplet-based scRNA-seq profiling of the murine aorta we
defined three distinct EC states present under physiological conditions \textit{in vivo}: two major populations specialized for adhesion and lipid handling, and a third primarily lymphatic. These subpopulations persist and demonstrate a unified transcriptional response to Western diet, an environmental stressor.

To further define the components of and heterogeneity in stimulus-induced EC plasticity, isolated ECs were examined using a controllable bioreactor system for flow culture and picowell-based scRNA-seq for transcriptional characterization. Shear stress was applied as a paradigmatic mechanical stimulus singly and in combination with rapamycin as a paradigmatic chemical stimulus. The modular components of shear-induced EC state, defined based on previously identified shear-responsive transcription factors, showed heterogeneous expression among shear-cultured ECs. Computational definition of the components of rapamycin-induced plasticity demonstrates that physiological shear-cultured ECs have a more complex plastic response relative to altered shear-cultured ECs. EC subpopulations with transcriptomes dominated by characteristic functions including mitosis and stress states were identified in the shear-cultured populations. These subpopulations recapitulate the altered-shear response to rapamycin but not the physiologic-shear response to rapamycin, suggesting that ECs in a physiologic state are maximally plastic while shear alteration or other stress stimuli induce transcriptional states that limit plasticity to a second stimulus.

These results suggest that locally observed EC heterogeneity may be the result of variable and interdependent responses to multiple stimuli, and that EC dysfunction may limit pharmacologic manipulation. Further studies to investigate EC heterogeneity
and its impact on plasticity are proposed using both scRNA-seq and other molecular phenotyping methods with the aim of understanding the translational relevance of these results.
Table of Contents

Title Page ........................................................................................................ i
Copyright ........................................................................................................ ii
Abstract ........................................................................................................ iii
Table of Contents ........................................................................................... vi
Acknowledgements ......................................................................................... vii

1. Background and Significance ...................................................................... 1
   1.1 Introduction .......................................................................................... 1
   1.2 Endothelial plasticity in health and disease ........................................ 2
   1.3 Structural endothelial heterogeneity .................................................... 4
   1.4 Functional heterogeneity of endothelial cells ....................................... 6
   1.5 Application of transcriptomics to define EC state ............................. 9
   1.6 Single-cell resolution transcriptional analysis .................................... 10
   1.7 Single-cell RNA sequencing of vasculature ..................................... 14
   1.8 Summary: Combining EC heterogeneity and plasticity into a unified paradigm of endothelial response ................................................. 17

2. Functional endothelial heterogeneity in vivo in the physiologic mouse aorta ... 21
   2.1 Introduction .......................................................................................... 21
   2.2 Methods .............................................................................................. 23
   2.3 Results .................................................................................................. 29
   2.4 Discussion ............................................................................................ 57

3. Defining heterogeneous components of EC plasticity in vitro in response to mechanical and chemical stimulation ............................................. 63
   3.1 Introduction .......................................................................................... 63
   3.2 Methods .............................................................................................. 67
   3.3 Results .................................................................................................. 84
   3.4 Discussion ............................................................................................ 108

4. Conclusions and Future Research Directions ............................................ 118
   4.1 Overview and challenges ..................................................................... 118
   4.2 Functional endothelial heterogeneity and future steps ....................... 120
   4.3 Single-cell resolution EC plasticity and future steps ......................... 122
   4.4 Long-term research goals .................................................................... 123

Appendix A: Ibidi Protocols .......................................................................... 125
   Seeding cells in Ibidis .............................................................................. 125
   Connecting Ibidis to flow ........................................................................ 125
   Addition of rapamycin to flow loops ....................................................... 127
   Isolation of cells from Ibidis ................................................................. 127

Appendix B: Computational Resources ........................................................ 129
   R helper functions for analysis of scRNA-seq data .................................. 131

References ...................................................................................................... 138
Acknowledgements

This dissertation is the product of mentorship, support, and encouragement from numerous individuals to whom I am profoundly grateful. Academic research training is still very much in the apprenticeship model, and the PhD advisor serves as instructor, role model, and guide. I am very fortunate to have had Professor Elazer Edelman serve in these capacities. As an instructor, he is a dedicated teacher who delights in sharing and receiving knowledge. As a role model, he is a paragon of the physician-scientist whose career exemplifies the delicate but rewarding balance between these two arenas. As a guide, he provided me encouragement when I despaired, realism when I was grandiose and untethered, and perspective when I developed tunnel vision. Beyond his role as an academic mentor, he connects with his trainees on a personal level and treats them as his extended family. I will always value his continued mentorship and friendship.

As a dissertation grows more defined and more complex, it is supported not only by the advisor but by a dedicated and expert committee. Each member of my thesis committee has contributed their valuable time, knowledge, and guidance to this work. Professor Alex Shalek gave me generous and unfettered access to his lab, shared his expertise in single-cell RNA-sequencing and high-dimensional molecular phenotyping, and delivered a new profound insight each time I sat down on his office sectional. Professor Rajat Gupta showed me the nuts and bolts of the scientific process from the ground up, taught me the value of ‘just doing the experiment’, and provided an insider understanding of the clinician-scientist career path. Professor Peter Sorger showed me
the value in gleaning insights from related fields and in asking the simple questions which turn out to be big questions. I am grateful for their lessons and their guidance.

Mentors lift us to greater heights, but our friends and companions in the trenches support us through the daily struggles and triumphs. The Edelman Lab is an eclectic combination of diverse individuals with the common goal of using any technique necessary to define new paradigms in translational cardiovascular science. It has been my absolute privilege to train there alongside excellent scientists, teachers, and friends. Many thanks to my fellow graduate student, co-TA, and co-conspirator Brian Chang for teaching me that high standards and a fearless approach elevate any pursuit, whether scientific, educational, or recreational. Or Gadish and Natalia Drosu welcomed me into their circle of vascular biologists, and provided the kind of wisdom only senior students can. My gratitude also to my MD-PhD program classmates, particularly Mark Kalinich and Senan Ebrahim, who make the longest roads seem breezy and familiar with their companionship and good humor.

Nothing can be built without a foundation. My family is my foundation, and anything I have accomplished is based on their support. My grandparents and extended family have loved and nurtured me always while exemplifying personal values and professional excellence. Special thanks to Swetha Dravida and Teja Voruganti for accompanying me on the long journey in pursuit of an MD-PhD. My brother Aravind is my oldest confidant and most unwavering believer, while providing just enough sibling rivalry to motivate timely completion of a doctorate (before he does, of course). My parents Sreeramesh and Manjula have been an inexhaustible source of love and encouragement in all things; they are my first and best teachers. Finally, my eternal
gratitude to the real Dr. (Saxena) Kalluri, my wife Nikita – I admire every day how she utilizes her boundless energy and infinite optimism to navigate our personal and professional milestones together.
1. Background and Significance

1.1 Introduction

The understanding of endothelial cells (ECs) as contributors to health and disease has evolved over time from the "nucleated cellophane" of Lord Florey's time [1] to the modern view of a complex and highly diverse cell type involved in homeostasis, immune function, stromal support of parenchymal tissue, and many other diverse functions [2]. Depending on their location and the pathology under study, translational scientists are interested in ECs as permeability “gatekeepers”, as regulators of hemostasis and flow sensing, as lynchpins of the immune response, and as paracrine regulators with diverse phenotypes. As the known functions of the EC have expanded, the complexity of describing EC state has dramatically increased.

The nature of endothelial cells (ECs) as cells whose plastic phenotypes reflect their local context [2] limits classic means of cell identification and even the most sophisticated of correlative statistics. Endothelial cells can display heterogeneity in any of their characteristic functions, including but not limited to regulation of leukocyte trafficking, expression of coagulation mediators, and secretion of paracrine factors that control the biology of local basement membrane and adjacent cells like smooth muscle cells [3]. The very nature of the EC is that its phenotype is plastic, or mutable; therefore, knowing the state of an individual cell or an average population phenotype is not that enlightening. However, modern advances in single-cell characterization of transcriptional state in cell populations promise to enable new insight into both the diversity and plasticity of EC states. The premise of this work is that harnessing single-cell transcriptomics enables definition of discrete EC states in vivo under physiologic
and disease-simulating conditions as well as dissection of the myriad biological programs composing plastic EC state shifts in response to disease-relevant stimuli in vitro.

1.2 Endothelial plasticity in health and disease

The heterogeneous response of EC populations to physiologic and pathologic stimuli is the result of the inherent plasticity of EC phenotype. ECs respond to a wide range of mechanical, chemical, and cell-derived environmental stimuli. The “state” of an EC is a superposition of its participation in a number of physiologic functions including mechanical support, barrier function, coagulation, inflammation, metabolism, and paracrine signaling. Because ECs perform so many complementary functions in their role as the parenchymal cell of the vasculature, each potential stimulus can result in a wide array of overlapping phenotypic changes. This is the basis of EC plasticity.

In the absence of pathological stimulation, mature ECs following development exist in a non-migrating, non-proliferating state of quiescence [4]. EC dysfunction in response to pathology disturbs this quiescent state, manifesting as endothelial “activation”. Although the distinction between quiescent and activated endothelium is often presented as a binary, activated states vary widely depending on the nature of the inciting stimulus. For example, in response to the hyperglycemic environmental stimulus associated with diabetes mellitus, ECs alter their glucose metabolism and handling of reactive oxygen species (ROS). The downstream effects of EC interaction with advanced glycation end products also results in alteration of eNOS signaling and crosslinking of vascular extracellular matrix, resulting in an EC-mediated mechanical
alteration of vessel properties [5]. The hyperinsulinemia associated with diabetes may also result in enhanced expression of inflammatory adhesion molecules on ECs, which results in greater migration of immune cells across the vascular barrier [6]. Thus, a single pathological environmental stimulus results in multiple overlapping phenotypic changes to the state of the plastic EC, which in turn affect multiple biological outputs of the cell. The same is true of other well-known vascular disorders. Atherosclerosis is also characterized by alterations in eNOS signaling and increase in production of ROS [5]. The inflammatory component of atherosclerosis results in the activation of pleiotropic transcription factors including NF-kB, which results in activation of a characteristic pattern of effector proteins including class II MHCs, endothelial-leukocyte adhesion molecules, procoagulant molecules, and secreted chemokines. VCAM-1, an important EC adhesion molecule, is specifically linked to atherosclerosis; it is stimulated by oxidized lipoproteins and spatially expressed in those regions of the vasculature most prone to lesion formation in animal models of atherogenesis [7]. Thus, EC plasticity can manifest both as general transcriptional programs activated by multiple pathological stimuli and as more specific molecular cues linked directly to disease pathogenesis. EC responses to pathologies as varied as sepsis and hypertension can include plasticity in inflammatory, metabolic, mechanical, and coagulation functions of ECs [8]–[10]. The wide range of EC functions allows for a highly plastic EC state as ECs respond to various environmental stimuli with both general and specific departures from the physiologic state of homeostatic quiescence.

Plasticity in response to specific conditions can often be related to specific transcriptional programs as defined by transcription factor (TF) expression and activity;
these TFs and their targets are often used as biological linchpins which enable the organization of complex EC plasticity into discrete “building block” programs. For example, a recent genome-wide approach combining epigenetic and transcriptional data identified the ETS and AP1 TF families as primary modulators of human aortic endothelial cell (HAEC) biology, and further associated ERG, NRF2, NF-kB, CEBD, and IRF1 with the EC response to inflammatory stimuli [11]. This type of study illustrates the inherent complexity of EC plasticity even in response to a relatively straightforward stimulus. New methods for defining existing EC states, demonstrating plastic responses elicited by stimuli, and resolving the modular components of those responses are necessary to advance our understanding of the endothelium.

1.3 Structural endothelial heterogeneity

Achieving a complete biological definition of endothelial state is complicated by the nonuniformity of the endothelium, or endothelial cell heterogeneity. The study of endothelial cell heterogeneity originated primarily with the discussion of organotypic vasculature, categorizing the differences between organ-specific vascular beds like the sinusoidal endothelium of the liver, the fenestrated endothelium of the kidney, and the ubiquitous continuous endothelium of most microvasculature [12]. Heterogeneity among different levels of the vascular tree – ECs from arteries, capillaries, and veins – is also well-elucidated, and arises from diversity in vessel morphology [13], [14]. In both between-organ and between-vessel type differences, heterogeneity arises from the necessity for endothelial specialization – ECs in these differing vascular beds have unique molecular functions which drive their particular structure and molecular
phenotype. In organs, these differences are attributable to the diverse biological processes in which ECs participate. The sinusoids of the liver, spleen, and bone marrow allow for free conductance of plasma proteins and other large solutes. Fenestrated capillaries as found in the kidneys, intestines, and choroid plexus have intracellular windows with endothelium-spanning diaphragms that increase the passage of water and small solutes. High endothelial venules in lymph nodes serve to transport lymphocytes from the blood into lymph nodes [12].

Differences in ECs throughout the body are not limited to morphology or changes in the permeability of the basic EC barrier. The endothelium in various organs secretes growth factors, termed ‘angiocrine factors’, which serve to support organ patterning, regeneration, and homeostatic function. The secretion of these factors in an organ and tissue-specific manner by ECs is another important component of structural endothelial heterogeneity. Upregulation of angiogenic factors including VEGF-A, VEGF-C, FGF-2, or the angiopoietins in response to injury or as a component of homeostasis result in elaboration of diverse angiocrine factors which support regeneration via tissue-specific stem cells (bone marrow, testes, brain), parenchymal regeneration (liver), or metabolic alterations (pancreas, adipose tissue) [15]. These organ-specific EC secretory functions further underscore the important role that structural origin plays in driving the observed heterogeneity of ECs throughout the body.

The origin of these differences has been linked to three major classes of drivers: cell-intrinsic transcriptional programs, mechanical forces in the environment, and external metabolic signals including paracrine growth factors and cell-cell interactions [14]. Heterogeneity in ECs primarily arises during embryonic development as a result of
differences in angiogenic mechanisms, the balance between resident and circulating cellular precursors, and in some cases clonal expansion of vascular cell populations. The structural difference between arteries, veins, and capillaries has been confirmed as arising from characteristic networks of transcription factors [14]. This structural heterogeneity between vessels from different organs and different vascular beds has important physiologic and pathologic consequences; it allows for organ-specific homing of leukocytes, supports diverse end-organ functions, and leads to vascular diseases or tumors that express preferentially in specific vascular beds [14]. These factors are an important component of our understanding of endothelial heterogeneity as a field, and the techniques and ideas used to define endothelial heterogeneity were as a rule first developed to characterize this type of variation.

1.4 Functional heterogeneity of endothelial cells

In addition to structural heterogeneity in spatially diverse locations, EC heterogeneity has been demonstrated on a local level within specific vessels and vascular locations. We term this type of EC heterogeneity ‘functional heterogeneity’, as it arises not as a consequence of the differences in structure and biology between spatially distinct regions, but instead as a consequence of the physiologic or pathologic functioning of a single, localized endothelium.

Functional endothelial heterogeneity arises in myriad biological contexts. In a physiologic context, specific ACh-sensitive “detector cells” have been identified within vessels that serve as focal points for the overall vascular response to ACh [16]. These cells co-occur in spatially isolated clusters and demonstrate increased ACh receptor
expression relative to the majority of neighboring cells, demonstrating that within a local endothelial cell population ECs can specialize for complementary functions. Additionally, von Willebrand Factor (vWF), a molecular participant in coagulation and canonical marker of EC identity, is expressed in a “mosaic” pattern in the capillaries of vascular beds and the aorta. This diversity is dynamically regulated, and has been modeled as a bistable, stochastic, phenotypic “switch” whose function is to support homeostasis in the adult vasculature [17]. Although targeted at individual molecules rather than overall EC state, these findings support the paradigm that local endothelia are comprised of a heterogeneous population of individual ECs, which may be functionally specialized even in homeostasis.

Functional EC heterogeneity has also been demonstrated in pathological contexts. Endothelial cells within tumors in colorectal cancer have demonstrated differing phenotypes based on tumor microenvironment primarily defined by the matrix protein SPRC1 and correlating with known tumor phenotypes that impact clinical prognosis [18]. Local endothelial heterogeneity is also relevant in the pathobiology of solid organ transplantation, where the presence of chimeric vessels composed of a mixed endothelial population has been investigated as a potential biomarker of graft state [19]. Functional endothelial heterogeneity plays a role in the response to injury. Regeneration in large vessels of adults has been shown to involve proliferation of endothelial cells proximal to the injury; these cells fall into two distinct populations with differential proliferative capacity by single-cell imaging of labeled EC clonal groups [20]. These observations in healthy and pathologic endothelia suggest that EC heterogeneity is not only the result of developmental drivers or spatial location, but can also have a
dynamic functional significance within local EC populations. This local heterogeneity is fundamentally related to the plasticity of the overall endothelium as it responds to external environmental drivers; however, whether heterogeneous populations arise transiently in response to stress or persist through homeostasis and pathology remains unknown.

The idea that each EC in a local population can be modeled as a plastic entity whose state is the result of a superposition of intrinsic and extrinsic cues has been summarized in a “dynamical systems” model [21]. Under this model, endothelial heterogeneity arises from the presence of multiple potential “stable states” in a theoretical EC phenotypic landscape. ECs move within this landscape in response to external stimuli and may be biased towards particular states based on intrinsic cues like developmental origin. External stressors induce a phenotypic response in individual endothelial cells which may exhibit variable plasticity; this alters the overall population distribution, resulting in a shift in proportion between stable states or the development of new or intermediate states as a consequence of plasticity. Such a model highlights the importance of EC plasticity in driving local heterogeneity, and conceptualizes the response of EC populations to physiologic and pathologic stimuli as the interaction between a multistable landscape of potential attractor states and the environmental stimuli that may or may not be sufficient to displace cells from those attractor states.

While this model is an intuitive paradigm for linking endothelial plasticity with endothelial heterogeneity, its use for hypothesis generation and understanding vascular biology is limited by the difficulty in adequately resolving distinct EC states and in describing endothelial plasticity.
1.5 Application of transcriptomics to define EC state

Selection of an appropriate molecular phenotyping method to capture EC plasticity and heterogeneity requires an understanding of previous strategies used to study alterations in EC state. Transcriptional phenotyping is a proven method to capture a snapshot of cell state under different conditions and to compare alterations in cell state following exposure to environmental stimuli. Initially, RNA microarray was applied to identify pan-EC transcriptional markers that differentiate ECs from other cell types. This approach was utilized to differentiate ECs from differing vascular beds as part of the early studies on EC heterogeneity [22]. This study demonstrated that no individual transcriptional marker was sufficient to distinguish ECs from non-ECs or ECs in differing vascular beds from each other; instead, a transcriptional profile composed of multiple genes was necessary to capture EC state. The RNA microarray approach has also been applied to the study of EC heterogeneity arising from organotypic differences; in this case a transcriptional profile composed of multiple RNA species was sufficient to cluster ECs derived from similar organs and to differentiate those from different organs, demonstrating that transcriptional phenotyping can capture important similarities and differences among EC populations [23]. Additionally, in this study RNA microarray was utilized to demonstrate the alterations in EC state arising from interventions including irradiation and hepatectomy. This implies that transcriptomics are useful not only to capture EC heterogeneity, but also to demonstrate EC plasticity. Transcriptional profiling has also been used to capture EC plasticity in response to mechanical alteration including prolonged fluid shear stress [24]. Studies of transcriptional networks in ECs have been utilized to summarize EC state in both homeostasis and under stress.
conditions; by focusing on transcription factors and their targets, complex stimulus classes can be linked to specific molecular profiles which enables a more precise discussion of alteration in EC state [11]. Transcriptomics have also been used to define circulating ECs, cells removed from their normal tissue context and found circulating in the blood, usually defined by EC-expressed surface markers [25], implying that transcriptomics can also be useful when ECs are detached from their typical local microenvironments. Transcriptional analysis like scRNA-seq best captures the combination of genomic and environmental influences which drive the functional state of a cell [26]. This is especially relevant in vascular disease, which is often the result of both a susceptible genetic background and local mechanical influences like hypertension and flow disruption.

1.6 Single-cell resolution transcriptional analysis

Understanding the functional significance of EC heterogeneity and its relationship to vascular pathology requires molecular methods for individually characterizing these highly plastic cells. In recent years, the ability to determine a single cell’s functional state has rapidly advanced with the development of techniques for profiling DNA, RNA, protein, epigenetic modifications, and chromatin accessibility in individual cells [26]. Of these, the best established are methods for single-cell RNA sequencing (scRNA-seq) [27].

Single-cell transcriptional sequencing methods for tissue-derived cell suspensions and for cultured cell populations have been an area of intense development for the last decade. Use of next-generation sequencing to characterize the
entire transcriptome of a single cell was first demonstrated in 2009 [28], following the basic paradigm of reverse transcription for conversion of cellular mRNA into cDNA, followed by DNA amplification via PCR, library preparation, and sequencing. In the following five years, the number of cells that could be simultaneously profiled has increased by orders of magnitude with the introduction of multiplexing [29], FACS [30], and microfluidic circuits [31] for automated cell sorting and manipulation of lysates and amplification products. In the past five years, the options for highly parallel scRNA-seq of thousands to tens of thousands of cells using barcode-based multiplexing have exploded with technical variations including reactions in plates [32], droplets [33], and microwell arrays [34]. The introduction of the commercial Chromium System from 10X Genomics in 2016 has further supported the meteoric rise in the use of scRNA-seq. The latest methods seek to push the envelope further by enabling sequencing of preserved samples via nuclear isolation [35] and by enabling high-dimensional characterization of the transcriptome with preservation of spatial information [36]–[38].

Employing scRNA-seq to understand the functional heterogeneity of ECs and its impact on EC plasticity is a logical extension of the major currently demonstrated applications of single cell transcriptomics. Demonstrating and characterizing the heterogeneity present in a complex tissue is the first major application of scRNA-seq. This often takes the form of using computational clustering methods applied to the high-dimensional single-cell transcriptional dataset in order to identify the discrete cellular populations or subpopulations comprising a sample. Populations identified in this way can be differentiated by enumeration of marker genes expressed selectively relative to other populations. Cross-reference with canonical markers of known cell types can
allow for labeling of major cell populations, and new rare or previously unappreciated populations can also be identified in this manner. Additionally, resolution of multiple subpopulations all corresponding to a given canonical cell type allows for the demonstration of discrete states available to a given cell, or of cellular subtypes. This type of analysis is evident in studies describing the cellular composition of metastatic melanoma and the cellular states available to tumor cells [39], or in the “atlas” type work documenting the cell types (often including endothelial cells) present in complex organs like the intestine [40], heart [41], lungs [42], or cerebral cortex [43]. A recent major undertaking in this domain was the Tabula Muris, a truly titanic effort to sequence 100,000 cells from 20 organs and tissue types in the mouse to generate a full and comprehensive atlas of the whole-body single-cell transcriptome [44]. Similar analyses have been also applied to identify component subpopulations in tissue engineered constructs, such as liver organoids [45] or iPSC-derived cardiomyocytes [46]. Single cell RNA-seq also enables a parallel description of heterogeneity in the form of coexpressed gene networks; even within a particular cell type, the availability of thousands of cellular transcriptomes allows for the identification of gene sets whose expression is tightly correlated within a population, defining a gene module. These modules may be variably coupled to other modules, resulting in a basis vector-like description of cellular population biology [47], [48]. Transcription factors are often used as the basis of this type of characterization, as their targets should logically form this type of coexpressed module [39], [42], [46].

Appropriately, plasticity is the second broad theme in recent scRNA-seq studies. Correlated gene modules can be used to partition the response of a seemingly
homogeneous cell population to a given stimulus (e.g. dendritic cells and inflammatory mediators) [47], [49]. In a complex tissue composed of multiple cell types, these data sets can also demonstrate which aspects of a biological response – for example, the intestinal response to infection [40] or the planarian response to injury [50] -- are conserved across all cells and which are unique to specific cell types. Comparison between physiologic and pathologic states is an extension of this type of study. Tumor and non-tumor cells have been compared directly in the context of lung cancer [42], as have the differences in heterogeneity between homeostatic and post-injury hepatocytes and biliary epithelial cells [51].

The third major idea in single-cell transcriptomics is the identification of cellular trajectories. Many cellular applications and contexts do not result in fully discrete subpopulations, but instead in a dynamic continuum of states corresponding to an axis of biological activation or a developmental trajectory. Single-cell transcriptomics can elucidate the genes which form the axis of these trajectories to provide insight into the biological nature of the continuum. This strategy has been applied to good effect in the developmental context of ordering neural stem cells in mice [52]. Pseudotemporal trajectories are also useful for understanding continuum disease processes like the trajectory of nasal epithelial cells from non-polyp to polyp states with allergic inflammation [53]. A recent attempt to document the full series of developmental trajectories in mouse embryos was conducted in over 2 million cells [54] and illustrates the potential power of this type of analysis. Thus, both the recent explosion in single-cell transcriptomic methods in the last five years and the currently demonstrated
applications of scRNA-seq suggest its utility as the chief tool to illustrate the presence of functional EC heterogeneity and its impact on endothelial plasticity.

### 1.7 Single-cell RNA sequencing of vasculature

Of course, the field of vascular biology has not been asleep at the wheel during this period of tremendous growth in the promise and potential of scRNA-seq. Primarily in the last two years, there has been a steady and growing interest in applying scRNA-seq to vascular contexts to distinguish heterogeneous cellular populations, refine our understanding of gene modules comprising response to stimuli, and define continuous axes of cellular phenotype. Several studies have explored the cellular subpopulations present in genetic mouse models of atheroma development. Studies which used cellular presorting for CD45 expression to obtain immune cell single-cell transcriptional profiles in ApoE KO mice identified multiple subpopulations of immune cells within major cell groups [55], [56]. One found 11 leukocyte clusters and heterogeneity within established immune cell categories; intriguingly, differences in subpopulation frequency within plaque-derived samples correlated with clinical outcomes [55]. Another study identified a novel macrophage subpopulation with an osteoclast-like signature suggesting potential involvement in lesion calcification [56], demonstrating that scRNA-seq profiling of the vasculature can contribute to both translational and basic applications of vascular biology. scRNA-seq of the vascular adventitia in ApoE KO and WT mice identified multiple cell types including endothelial cells, with a primary focus on immune cells and identification of novel progenitor-like and proinflammatory populations of adventitial cells.
[57] to expand our understanding of the composition of the vascular niche in health and disease.

Single-cell transcriptional studies have also focused specifically on ECs in various physiologic organ-specific contexts, including barrier function in the lung [58]. Much of this work has focused on continuous zonation of EC phenotype. A study of arteriovenous zonation in brain ECs described a continuum of states for endothelial cells which tracked arteriovenous progression, contrasted with a punctuated gradation of phenotype in mural cells [59]. Another used a paired-cell sequencing approach to define the genes which define zonation of liver ECs in tandem with the spatially graded phenotype of hepatocytes. Transcripts describing this spatial gradient included Wnt9b, Rspo3, Pcdhgc5, Ecm1, Ltbp4, and Efnb2 [60]. From a developmental perspective, transcriptional characterization of arteriovenous determination of progenitor cells in coronary arteries demonstrated a key role for the transcription factor NR2F2 [61]. By delving deeper into the transcripts and gene modules that correspond to anatomic spatial axes (like the arteriovenous axis and the central-portal hepatic axis), this type of study provides pushes forward our understanding of structural EC heterogeneity in specific contexts. Further studies in the development and differentiation domains have clarified the continuum aspect of the endothelial-hematopoietic transition of hematogenous endothelium [62] and on the developmental trajectory of iPSC-derived ECs [63]. Interestingly, the latter study demonstrated 4 major subpopulations of iPSC-derived ECs associated with CLDN5, APLNR, GJA5, and ESM1, indicating that the potential presence of discrete endothelial states requires further investigation.
Additional vascular applications of scRNA-seq have focused on describing differences between EC populations in complex tissues including the lung [42] and kidney [64] in physiological states compared to pathology or disease models. Comparison of tumor and nontumor ECs in lung ascribed immune activation functions to non-tumor ECs and angiogenic function to tumor ECs, but did not analyze any discrete heterogeneity within the normal EC population [42]. Notably, the study of transcription factors as an important set of module-defining genes is common to many of these reports [42], [61]. A translationally-minded experiment to isolate and analyze single ECs isolated from implanted intracranial stents in pigs successfully obtained scRNA-seq profiles, although they did not demonstrate any discrete EC subtypes in these samples [65]. This offers the tantalizing vision of the future use of scRNA-seq to assay patient vasculature from minute samples to obtain clinically actionable information on the state of the endothelium.

Although the broad themes of heterogeneity and plasticity are present in the studies of single-cell vascular transcriptomes to date, the questions of what basal heterogeneity is present in homeostatic endothelium and how the elements of EC functional heterogeneity impact endothelial plasticity remain open. Early efforts have begun to probe these ideas in highly specialized vascular contexts. A recent study demonstrated that in high endothelial venules (HEVs), continuous heterogeneity in expression of HEV EC (HEC) marker genes exists during homeostasis and that some HECs may express inflammatory markers even in the absence of inflammation [66]. Moreover, following the induction of inflammation, not all HECs upregulate inflammatory markers to the same degree. This type of study describing the response of a specialized
endothelium to a single stimulus supports concrete next steps to be taken in this arena. The studies described here fit nicely into the current context of scRNA-seq of endothelium by elucidating discrete functional heterogeneity that exists in a stereotypic vascular bed with and without disease-relevant conditions and by demonstrating the use of a controlled in vitro environment to define heterogeneity in a plastic EC population.

1.8 Summary: Combining EC heterogeneity and plasticity into a unified paradigm of endothelial response

The incredible plasticity of the endothelium as a participant in homeostasis and pathology is one of the oldest and richest areas of investigation in vascular biology. The longstanding interest in the structural heterogeneity distinguishing endothelia from distinct organs and distinct vascular beds has recently been supplemented by an interest in the local, functional heterogeneity of ECs within a single vessel. We propose an overarching paradigm for understanding the state of the endothelium as a whole and of the individual ECs which comprise it as the superposition of a plastic response with basal functional heterogeneity (Figure 1.1). Under this model, the endothelial cell population demonstrates some basal phenotype that may involve the presence of multiple discrete subpopulations. In the presence of a stress of any type (e.g. environmental, metabolic, genetic), the plastic response of the endothelium is convolved with this basal phenotype. The result from an average perspective is the state of the endothelium, which corresponds to our current notions of quiescence vs. activation.
Figure 1.1 Unifying paradigm of endothelial and EC state as the superposition of plasticity and heterogeneity.
Individual cells within the population, however, may have heterogeneous states following their plastic shift as per the dynamical systems model. The nature of this heterogeneity may be dependent on the nature of the inducing stimulus (stimuli-dependent or extrinsic), or it may be a consequence of heterogeneity in the underlying basal phenotype (stimuli-independent or intrinsic). Although this model is intuitive and simple, its usefulness and validity depend on careful investigations into the heterogeneity of EC state in homeostasis and under stress, as well as into defining the distinct components of endothelial plasticity in response to diverse stressors singly and in combination. The following chapters of this dissertation describe studies that seek to examine this paradigm.

**Chapter 2** details an unbiased description of the normal mouse aorta under homeostatic conditions, with particular emphasis on discrete endothelial subpopulations that were identified and functionally characterized using scRNA-seq. These subpopulations persisted and exhibited a conserved response with a diet-induced stressor, suggesting that basal EC heterogeneity exists and plays a role in variation of response to pathology.

**Chapter 3** details the use of a controlled bioreactor system to study plasticity of an endothelial cell population under mechanical stress alone as compared with coupled mechanical and chemical stresses. The components of flow-induced plasticity are dissected and described using scRNA-seq and the potential impact of heterogeneity induced by one stimulus class on plasticity to another stimulus class is explored.
Chapter 4 summarizes the described contributions and details future experiments that may be useful in elucidating this core paradigm.

Appendices following the main body of the text include experimental protocols and other resources.
2. Functional endothelial heterogeneity *in vivo* in the physiologic mouse aorta

*Note:* This chapter has been adapted from work accepted for publication as of June 2019:


### 2.1 Introduction

#### 2.1.1 Context

In the first chapter, we described a paradigm for studying the heterogeneity and plasticity of populations of endothelial cells in physiologic and pathologic states. We proposed that the state of the endothelium in a given vascular bed can be interpreted as the convolution of a basal heterogeneity that exists at physiologic baseline with plastic responses to superimposed stressors. The result in the aggregate is the state of the endothelium, and within the endothelium may exist functional heterogeneity of individual ECs. Single-cell RNA-sequencing (scRNA-seq) offers the possibility of supporting this paradigm by defining for the first time both basal heterogeneity and functional heterogeneity following a plastic response to stress.

In this chapter, we report the use of scRNA-seq to characterize the transcriptional profile of the physiologic mouse aorta. In doing so, we uncover basal EC heterogeneity via clustering of aortic ECs into three subpopulations. In order to understand the significance of this heterogeneity, we apply functional interpretation including gene set enrichment and transcription factor module scoring. These results
demonstrate multiple stable states available to ECs in vivo under homeostatic conditions. Additionally, we demonstrate a unified plasticity of these populations to Western diet, an environmental stressor. These results both reveal basal functional heterogeneity and demonstrate the utility of scRNA-seq in elucidating our paradigm of EC state response.

2.1.2 Vascular single-cell transcriptional profiling

As covered in the previous chapter, this work stands in the context of the burgeoning applications of scRNA-seq to the vascular milieu. While previous EC-specific studies have focused on vascular zonation [59], [60], this study analyzes a single large vessel with multiple spatial axes to bring to light functional heterogeneity within a single vessel. Previous studies have also focused on particular cell types by presorting cells using FACS for canonical markers, primarily targeted at immune cell subpopulations [55], [56]. Our results instead first generate an unbiased cell profile to capture all cell populations, then analyze the discrete EC subpopulations identified in this context. This ensures that we capture all cellular subpopulations in an unbiased manner and enables the identification of true pan-cell type markers.
2.2 Methods

2.2.1 Mice

Whole mouse aortas were harvested from 12-week old female C57/BL6 mice on either chow diet or 8 weeks of Western Diet (Research Diets). Four mice were included in each group, with 2 dissociated aortas from each condition sequenced at low-depth (17,000 reads/cell) and 2 samples sequenced at high-depth (145,000 reads/cell). The high-depth samples from the chow diet (n=2) and Western diet (n=2) were used for subsequent analyses. All mouse protocols were approved by the Broad Institute IACUC and all protocols were in accordance with institutional guidelines. The aorta was dissected from the root (distal to the aortic valve) to the femoral artery bifurcation. The isolated aorta included aortic arch, ascending, descending, thoracic, and abdominal portions. Perivascular fat was dissected from the vascular tissue prior to dissociation and single cell analysis.

2.2.2 Aortic dissociation

Preparation of a single cell suspension of aortic cells was performed using a previously described enzymatic digestion protocol [67]. Briefly, the isolated whole aorta was finely cut and incubated in 1X Aortic Dissociation Enzyme Solution (125 U/mL collagenase type XI, 60 U/mL hyaluronidase type 1-s, 60 U/mL DNase I, and 450 U/mL collagenase type I) for 1 hour at 37°C. The cell suspension was strained through a 30 μm filter, treated with ACK lysis buffer for 5 minutes at room temperature, and washed twice with PBS. The cells were resuspended in 0.4% BSA-PBS at a final concentration of 8x10^5 cells/mL. To determine if the dissociation protocol resulted in under-representation of certain cell types, a second aortic dissociation protocol [68] using
elastase (0.5mg/mL) and collagenase A (2mg/mL) for 30 minutes at 37ºC was analyzed by flow cytometry and droplet-based single cell RNA-seq. Both dissociation protocols had a similar yield of endothelial cells and identified the six major vascular cell types.

2.2.3 Droplet-based scRNA-sequencing

Single cells were processed through the GemCode Single Cell Platform using GemCode Gel Bead, Chip and Library Kits (10X Genomics) as per the manufacturer’s protocol. In brief, single cells were sorted into 0.4% BSA–PBS solution. 9,000 cells were added to each channel. The cells were then partitioned into Gel Beads in emulsion in the GemCode instrument, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification, shearing and 5’ adaptor and sample index attachment. Libraries were sequenced on an Illumina NextSeq 500.

2.2.4 Pre-processing of droplet-based scRNA-seq data.

De-multiplexing, alignment to the mm10 transcriptome and unique molecular identifier (UMI)-collapsing were performed using the Cellranger toolkit (version 2.1.0) provided by 10X Genomics. The output was a corrected expression matrix, which was used as an input for further analysis.

2.2.5 Single-cell data analysis

Dimensional reduction, clustering, and analysis of single-cell RNA sequencing data were performed using the R package Seurat (Version 2.3.1) [69], [70]. Cells with expression of fewer than 200 or more than 4000 genes and cells with greater than 25% expression of mitochondrial genes were filtered out of the analysis. Normalized expression values $E_{i,j}$ for gene $i$ in cell $j$ were calculated by dividing unique molecular
identifier (UMI) counts for gene i by the sum of the UMI counts in cell j, to normalize for differences in coverage, multiplying by 10,000 to create transcript per million (TPM)-like values, and finally computing log2(TPM + 1). Variable genes were identified using the Seurat FindVariableGenes method with the LogVMR dispersion function parameter; genes with log-normalized expression values between 0.125 and 4 and with a dispersion of at least 0.5 were considered variable. The Seurat ScaleData function was used to scale and center expression values in the dataset for dimensional reduction.

Principal component analysis (PCA) for dimensional reduction was performed using Seurat functions based on the variable genes previously identified. PCs 1-12 were selected for further study based on manual examination of the contribution of each PC to overall variability and based on the genes contributing to each PC. t-SNE was performed using Seurat functions based on PCs 1-12, and clustering to define cell identity was performed using the Seurat FindClusters function with resolution=0.5. Marker genes for each cluster were determined using the Wilcoxon rank-sum test via the FindAllMarkers function in Seurat.

2.2.6 Endothelial cluster gene set signatures

Gene Ontology (GO) biological process-associated genes were identified using the R package biomaRt [71, 72] and gene set scores for each set were defined as described previously. The FindAllMarkers Seurat function was utilized to find positive markers of each EC subpopulation using the Wilcoxon rank-sum test. Pathway enrichment analysis was performed using ReactomePA [73] to identify gene sets from the Reactome [74] database with false discovery rate (FDR)<0.05 enrichment in marker sets. Functional gene signatures for each endothelial cell subpopulation were defined.
as the intersection between the enriched Reactome pathway and the subpopulation markers. The Seurat function AddModuleScore was used to define a score for each of the gene signatures defined this way. Briefly, a background gene set was created for each signature gene set by selecting the 10n nearest neighbors to the signature set genes as determined by mean expression and detection frequency in all cells. For each cell, the gene set score was defined as the mean expression of the genes in the signature set minus the mean expression of the background set genes.

2.2.7 Transcription factor target scores

Targets for transcription factors of interest were identified using Version 2 of the TRRUST database [75]. To capture the net effect on expression of targets for a given transcription factor, gene set scores were computed for all cells for gene sets comprising known targets of TF activation. Gene set scores were computed using the Seurat AddModuleScore function as previously described. Significant differences in gene set score distributions between subpopulations were determined using the Mann-Whitney U-test in R.

2.2.8 Integrated analysis of normal and high-fat diet scRNA-seq data

Aortic single cell RNA-seq datasets from normal and high-fat diet fed mice were combined using established methods for combined analysis of two single cell data sets in Seurat [76]. Briefly, this approach uses canonical correlation analysis (CCA) to identify similarities between the datasets and PCA to identify differences. The Seurat RunCCA function generates shared canonical 'basis' vectors which captured shared correlation. These basis vectors were aligned using Seurat functions to generate a
unified aligned low-dimensional space representing both data sets. Subpopulations shared between the two datasets were identified using the Seurat FindClusters function as previously described using the top 15 canonical correlates from CCA and a resolution of 0.5. Markers of each subpopulation were identified as previously described using Seurat. In order to identify condition-specific genes that varied between datasets, the FindMarkers function was applied to unaligned data in a low-dimensional space defined by PCA as previously described.

2.2.9 **Partitioning cell type contribution to aortopathy-related gene expression**

A list of all genes linked to Mendelian forms of inherited aortic dissection syndromes was compiled from the Online Mendelian Inheritance in Man (OMIM) database (https://omim.org). The average expression of each aortopathy gene in each stromal cell type (EC, Fibroblast, and VSMC) was computed using the AverageExpression function in Seurat. Average expression values for each gene were normalized to determine proportion of stromal expression in each cell type for ternary plotting using the R package ggttern [77].

2.2.10 **Statistical analysis**

Marker genes for transcriptional subpopulations in scRNA-seq profiles were identified using the FindAllMarkers Seurat function with a minimum log-fold change threshold of 0.25 and with p-values computed using a Wilcoxon rank-sum test. Pathway enrichment analysis was implemented using the enrichPathway ReactomePA function; p-values were computed using a hypergeometric test and adjusted for multiple hypothesis correction with a Benjamini–Hochberg procedure. Gene set scores and
imaging characteristics were compared between cell populations using the Mann-Whitney \( U \) test.

### 2.2.11 Data Sharing

2.3 Results

2.3.1 Single cell profile of the normal aorta

We transcriptionally profiled four wild-type C57/BL6 mouse aortas using droplet-based massively parallel scRNA-sequencing. Two aortas were sequenced at low-sequencing depth (17,000 reads/cell), and two aortas were sequenced at high-sequencing depth (145,000 reads/cell). There was no difference in the number of clustered cell populations based on sequencing depth. The high-sequencing depth data was used for all subsequent analyses, and yielded approximately 6,200 cells and 1,900 genes/cell (Figure 2.1). Cells were enzymatically dissociated over 1 hour using either collagenase/hyaluronidase or collagenase/elastase protocols, with no difference in the number of isolated cell types (Figure 2.2). Single cells were then individually bar-coded and sequenced using massively parallel droplet-based sequencing (Figure 2.1A). The individual samples were independently analyzed to confirm correlation between replicates, and then normalized and aggregated for joint analysis. The samples yielded a mean of 145,000 post-normalization reads per cell, which corresponds to a median of 1,900 genes per cell. We eliminated cells that expressed gene counts greater than 4,000 or less than 200 prior to analysis. The reproducibility between samples was validated by ensuring that all biological replicates were represented in all cell clusters.

Unsupervised graph clustering partitioned the cells into groups, which we visualized using t-distributed stochastic neighbor embedding (t-SNE, Figure 2.1B). Individual clusters were labelled for cell type using known marker genes (Table 2.1). The full set of transcriptional data was used to generate the clusters. Here we
Figure 2.1 Single cell RNA-sequencing atlas of aortic cells types. a) Schematic overview of experimental approach. Four mouse aortas were dissected from aortic root to femoral take-off, enzymatically dissociated for 1 hour at 37°C, and then the single cell suspension was sequenced at a depth of 1,900 median genes per cell using droplet-based RNA-sequencing methods. b) t-SNE representation of single cell gene expression shows the 6 identified major aortic cell types. c) Dotplot demonstrates the top markers of each aortic cell type. Dot size corresponds to proportion of cells within the group expressing each transcript and dot color corresponds to expression level.
Figure 2.2 Comparison of collagenase/hyaluronidase/DNAse dissociation protocol (original, left) with elastase/collagenase dissociation protocol (right). t-SNE plots summarizing single-cell RNA-seq profiles for each protocol with vascular cell populations labeled. EC indicates endothelial cells; Fibro, fibroblasts; Macro, macrophages; Mono, monocytes; VSMC, vascular smooth muscle cells.
Table 2.1 Markers used to identify aortic cell types

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Marker Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibroblasts</td>
<td>Pdgfra</td>
</tr>
<tr>
<td>VSMC</td>
<td>Cnn1, Myh11</td>
</tr>
<tr>
<td>EC</td>
<td>Cdh5</td>
</tr>
<tr>
<td>Monocytes</td>
<td>Cd68, Cd115, Lyz2</td>
</tr>
<tr>
<td>T/B cells</td>
<td>Cd3d, Cd19</td>
</tr>
<tr>
<td>Macrophage/DC</td>
<td>H2-Ab1 (MHC-II), Cd68</td>
</tr>
<tr>
<td>RBC</td>
<td>Hbb</td>
</tr>
</tbody>
</table>
present clustering analysis using 1,100 variable genes per cell to identify cells with similar profiles. Vascular cell types represented in the distinct clusters included ECs, VSMCs, arterial fibroblasts, and immune cells, as well as a small neuronal cluster likely arising from adjacent tissue. A small population of red blood cells (RBCs) was also present following whole-aorta preparation; these cells were excluded from further analysis.

2.3.2 scRNA-sequencing identifies novel cell type-specific gene expression markers

The full set of differentially expressed markers for each cell population were identified from the RNA-sequencing data. Cell type specific markers for a cluster were defined as the five genes with the highest differential expression relative to all other cells (Figure 2.1C). For example, the top 5 VSMC specific markers are Myh11, Tpm2, Myl9, Tagln and Acta2. These genes include canonical VSMC genes, and therefore confirm the assignment of the cell clusters as VSMCs. We also identified all genes with log fold enrichment greater than 2 for each cell type relative to all other cells (Figure 2.3). The heatmap of the gene expression shows the number of genes with significantly different expression (p-value<0.01 by Wilcoxon rank-sum test and log fold enrichment > 2) for each cell type. Monocytes have the largest number of unique transcripts as defined by this specificity standard compared with other cell clusters.

With expression data for ~1,900 genes per cell, we were able to identify different categories of cell type-specific markers. The first set of markers are widely expressed in all cells of that type, and not expressed in other clusters. In Figure 2.1C the size of each dot corresponds to the percentage of cells within a cluster expressing the gene, and the
Figure 2.3 Heatmap identifying all genes with log fold change $>2$ for each aortic cell type relative to all other cells.
color represents the level of expression. Several markers are expressed in over 75% of the cells and with an average normalized expression of >0.75. Examples of these markers are Myh11 (VSMCs), C1qa (Monocytes), and Pecam1 (ECs). These markers represent the most specific and sensitive markers for the entire population of the given cell type. This is in contrast to markers that are highly expressed in a subset of a cell type. These genes remain specific markers, as they are not expressed in other cell types, but are not sensitive to capture all the cells of a pre-specified group. Examples include Gpihbp1 (ECs) and Cd52/Rac2 (monocytes). A third marker type we identified were genes with high expression in one cell type, but also significant expression in other cell clusters. While these genes are not classically considered cell type specific markers, they did contribute to identification of cell identity in our clustering algorithm. Using all genes to cluster the cells allows for higher resolution detection of cell subpopulations and cellular heterogeneity compared with using only a small number of canonical cell-type specific markers.

2.3.3 Whole transcriptomic data identifies cellular subpopulations within aortic cell types

Following the analysis of cell types, we examined the 10 individual groups defined by clustering to define subpopulations within each aortic cell type (Figure 2.4A). Fibroblasts, monocytes/macrophages, and ECs all clustered into multiple subpopulations which were more similar to each other than to other aortic cell types as described by a cluster dendrogram (Figure 2.4B).
Figure 2.4 Subpopulations within cell types identified from single cell RNA-sequencing. a) t-SNE demonstrating 10 clusters identified as comprising the aorta. b) Dendrogram summarizing similarity between aortic cell subpopulations. EC indicates endothelial cells; Fibro, fibroblasts; Macro, macrophages; Mono, monocytes; VSMC, vascular smooth muscle cells.
VSMCs comprise the largest population of cells in our analysis, accounting for 39% of all cells. VSMCs clustered into one subpopulation which expresses the canonical VSMC markers Myh11 and Cnn1.

The second largest population of cells were fibroblasts, which account for 33% of all cells. These cells are defined by higher expression of Pdgfra and collagens/collagen-binding proteins (Col1a1, Col1a2, Dcn, Lum) along with reduced expression of VSMC-associated contractile proteins (Myh11, Cnn1) and were split into 2 subgroups by clustering. One canonical marker of VSMC identity, alpha smooth muscle actin (Acta2), was highly expressed in VSMCs, but also present in a large proportion of fibroblasts, though at lower levels (Figure 2.5).

Endothelial cells cluster into three distinct groups, defined by their common expression of canonical marker Cdh5. Certain canonical EC-specific genes such as Von Willebrand Factor (Vwf) and VEGF-receptors (Flt1, Kdr) showed heterogeneous expression by cluster, and therefore do not serve as effective markers for identifying all ECs in the aorta.

Immune cells clustered into 3 broadly defined groups – macrophages (H2-Ab1) and two subpopulations of monocytes (both expressing Lyz2). There is overlap in gene expression in the monocyte and macrophage/DC clusters, and multiple markers were necessary to discriminate the two clusters. Finally, a small population of cells expressed markers of neuronal or nerve-associated identity including Mbp and Cnp. These likely represent a small amount of contamination from neighboring neuronal tissue during dissection of the aorta.
Figure 2.5 Acta2 expression in aortic fibroblasts vs. VSMCs. Violin plot demonstrates increased expression of Acta2 in aortic VSMCs relative to aortic fibroblasts, but nonzero expression in both populations. Violin plot y-axis demonstrates normalized transcript expression values. Fibro indicates fibroblasts; VSMC, vascular smooth muscle cells.
In order to determine whether separate clusters identified within each cell type represent discrete subpopulations or a continuous phenotypic gradient, the markers for each of the 10 cellular subpopulations were plotted via heatmap (Figure 2.6). Each of the three EC clusters express distinct and non-overlapping markers, suggesting the presence of discrete subpopulations. Conversely, the two fibroblast clusters expressed overlapping markers, suggesting the presence of a continuous phenotypic gradient rather than true subpopulations.

Previous work on small vessels in the brain [59] has suggested that immediate early genes (IEG) including Fos, Fosb, Jun, and Junb may be enriched in cells isolated from solid organs or vascular tissue for single-cell profiling. This IEG signature likely represents an artifact of single cell dissociation, as there is no evidence these genes are expressed in situ. The IEG score for each of the cellular subpopulations shows similar levels of expression in each cluster (Figure 2.7), which suggests that IEG activation is not an explanation for the cellular heterogeneity in ECs and fibroblasts.

In order to explore the minimum number of sequencing reads per cell to achieve effective partitioning of aortic cell types using single-cell RNA-sequencing, we also performed clustering and identification of cell types in a single cell RNA-seq library of much lower depth (~17,000 reads/cell). The clustering of cells was essentially identical to data obtained from higher depth RNA sequencing. The same major vascular wall cell types and immune populations were identified with lower depth RNA sequencing (Figure 2.8). The clustering of ECs into three discrete subpopulations is present at low sequencing depth (~17,000 reads/cell) and high sequencing depth (~145,000 reads/cell).
**Figure 2.6** Heatmap identifying markers of each cellular subpopulation. EC indicates endothelial cells; Fibro, fibroblasts; Macro, macrophages; Mono, monocytes; VSMC, vascular smooth muscle cells.
Figure 2.7 Immediate-early gene (IEG; Fos, Fosb, Jun, Junb) activation score for aortic cellular subpopulations. Immediate-early gene activation score does not differentiate the 3 EC subpopulations identified in this study, and these genes are minimally expressed (scores < 0) in aortic ECs, VSMCs, and fibroblasts. EC indicates endothelial cells; Fibro, fibroblasts; Macro, macrophages; Mono, monocytes; VSMC, vascular smooth muscle cells.
Figure 2.8 Resolution of major aortic cell types using lower-depth sc-RNAseq. Seurat clustering analysis pipeline applied to low-depth (17k reads/cell) scRNA-seq library demonstrates resolution of the same major aortic cell populations, including 3 distinct EC populations. EC indicates endothelial cells; Fibro, fibroblasts; Macro, macrophages; Mono, monocytes; VSMC, vascular smooth muscle cells.
2.3.4 Three endothelial cell subpopulations have distinct gene expression profiles

Clustering analysis of all cells in the wild-type mouse aorta identified three distinct subpopulations of ECs as determined using the canonical marker Cdh5, which has previously been identified as one of the few markers constitutively expressed throughout the vascular tree (Figure 2.9A-B) [3]. Analysis of these clusters relative to other vascular cell populations using a Wilcoxon rank sum test reveals novel transcriptional markers of all aortic ECs (Figure 2.9C). These include known endothelial-specific participants in angiogenesis and blood vessel formation including Sdpr [78], Egfl7 [79], Ptprb [80], and Ecscr [81] as well as several adhesion and transport molecules that have previously been reported in ECs and other cell types including Cldn5 [82], Lcam2 [83], and Slc9a3r2 [84].

In order to characterize these subpopulations, we identified the transcriptional markers that differentiate each cluster. Unique markers were defined by highest differential expression using the Wilcoxon rank sum test with Bonferroni correction (Figure 2.10A). The largest population (EC 1) is defined by higher expression of many “canonical” EC markers (Vcam1) as well as other genes with known functions in ECs such as Clu, Gkn3, and Eln. The second EC population (EC 2) expresses genes involved in lipid transport (Cd36, Fabp4, Lpl, and Gpihbp1) and angiogenesis markers (Flt1). The EC 3 population expresses markers characteristic of lymphatic endothelium, including Lyve1 [85]. In order to elucidate the distinction between the two distinct subpopulations of blood endothelial cells, the lymphatic ECs in EC 3 were filtered out from downstream analysis.
Figure 2.9 Endothelial cells cluster into 3 distinct populations. a) t-SNE plot of EC subpopulations highlighted relative to other aortic cell types. The three EC subpopulations have distinct transcriptional profiles compared with each other and with all other aortic cells. b) t-SNE plot of all VE-Cadherin positive cells in mouse aorta. EC subpopulations were extracted from whole aorta data, and separately analyzed. In this t-SNE plot the 3 subpopulations are re-identified along with a small (5%) population of doublet cells which are an artifact of droplet-based single cell RNA-sequencing. c) Multiple genes show EC-specific expression and are expressed in all 3 EC subpopulations. Genetic markers expressed in all EC clusters compared with all other cells are divided into angiogenic EC marker genes and adhesion/transport EC marker genes. Violin plot y-axis demonstrates normalized transcript expression values.
Figure 2.10 Gene expression signatures of EC subpopulations. a) Top 20 genes with specific expression for each EC subpopulation. Dot plot shows the percentage of cells expressing each gene (dot size) and the expression level (dot color). b) Differential expression of transcription factors by EC subpopulation. c) Expression of PPAR-G targets in EC 1 and EC 2 subpopulations. *** indicates Mann-Whitney U-test p-value < 0.001.
Each of the subpopulations is characterized by a unique transcription factor profile (Figure 2.10B). In order to characterize the contribution of these transcription factor-based signaling networks to subpopulation identity, the known targets activated by each transcription factor were identified using the TRRUST database [75]. These targets were used to define a transcription factor network signature that was computed for all endothelial cells and compared between the two major subpopulations EC 1 and EC 2. EC 2 showed significantly greater expression of Pparg-activated genes (Figure 2.10C, p-value < 2.2 x 10^{-16} by Mann-Whitney U test), consistent with the greater expression of Pparg and other lipid-handling genes in that subpopulation. The differential expression of multiple transcription factors and their targets in EC subpopulations suggests that they have broadly different functional properties.

2.3.5 Gene set enrichment profiles identify distinct functions for the EC subpopulations

To systematically identify cellular functions that differ between EC subpopulations, we investigated the expression profiles enriched for each cell subpopulation. In order to confirm the functional identity of identified EC populations, two endothelial cell-relevant biological process gene sets for EC development and EC differentiation were located in the Gene Ontology (GO) database. These annotated gene sets were used to create EC gene set scores (Table 2.2). Generating expression scores for all cells in the dataset demonstrated that these well-characterized EC signatures did differentiate all the endothelial subpopulations from other aortic cells, but were not sufficient to resolve the differences between the subpopulations (Figure 2.11A).
<table>
<thead>
<tr>
<th>Development</th>
<th>Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add1</td>
<td>Rap1a</td>
</tr>
<tr>
<td>Arhgef26</td>
<td>Acvr1</td>
</tr>
<tr>
<td>Ccm2</td>
<td>Acvr2b</td>
</tr>
<tr>
<td>Cdh5</td>
<td>Hoxb5</td>
</tr>
<tr>
<td>Cldn1</td>
<td>Add1</td>
</tr>
<tr>
<td>Cldn5</td>
<td>Rapgef2</td>
</tr>
<tr>
<td>Clic4</td>
<td>Id1</td>
</tr>
<tr>
<td>Col18a1</td>
<td>Rdx</td>
</tr>
<tr>
<td>Ctnnb1</td>
<td>Arhgef26</td>
</tr>
<tr>
<td>Dmd</td>
<td>S1pr2</td>
</tr>
<tr>
<td>Eng</td>
<td>S1pr3</td>
</tr>
<tr>
<td>Ezr</td>
<td>Sox18</td>
</tr>
<tr>
<td>F11r</td>
<td>Rock1</td>
</tr>
<tr>
<td>F2rl1</td>
<td>S1pr2</td>
</tr>
<tr>
<td>Gpx1</td>
<td>Tjp2</td>
</tr>
<tr>
<td>Gstm3</td>
<td>Tnf</td>
</tr>
<tr>
<td>Heg1</td>
<td>Tnfrsf1a</td>
</tr>
<tr>
<td>Icam1</td>
<td>Tnmd</td>
</tr>
<tr>
<td>Id1</td>
<td>Vezf1</td>
</tr>
<tr>
<td>Ikbkb</td>
<td>Cldn1</td>
</tr>
<tr>
<td>Il1b</td>
<td>Ctnnb1</td>
</tr>
<tr>
<td>Marveld2</td>
<td>Dll1</td>
</tr>
<tr>
<td>Met</td>
<td>Ppp1r16b</td>
</tr>
<tr>
<td>Msn</td>
<td>Eng</td>
</tr>
<tr>
<td>Myadm</td>
<td>Ppp1r16b</td>
</tr>
<tr>
<td>Myd88</td>
<td>Ezr</td>
</tr>
<tr>
<td>Notch4</td>
<td>F11r</td>
</tr>
<tr>
<td>Pde2a</td>
<td>F2rl1</td>
</tr>
<tr>
<td>Pde4d</td>
<td>Gdf2</td>
</tr>
<tr>
<td>Ppp1r16b</td>
<td>Heg1</td>
</tr>
<tr>
<td>Ptprs</td>
<td>Hey1</td>
</tr>
<tr>
<td></td>
<td>Rap2c</td>
</tr>
<tr>
<td></td>
<td>Rapgef1</td>
</tr>
</tbody>
</table>

Table 2.2 GO set-based gene lists
Figure 2.11 Pathway analysis of EC subpopulations. a) GO-annotated EC development and differentiation gene set scores in endothelial subpopulations and non-endothelial aortic cells (Other). b) Functional gene sets identified from Reactome pathway enrichment and subpopulation markers differentiate the two major EC populations EC 1 and EC 2. *** indicates Mann-Whitney U-test p-value < 0.001. c) Angiogenic tip cell gene set score differentiates EC 1 and EC 2. *** indicates Mann-Whitney U-test p-value < 0.001.
In order to identify the EC functions that showed the greatest heterogeneity between the major subpopulations EC 1 and EC 2, pathway enrichment was performed on the top 50 markers of each subpopulation against the Reactome pathway database [74] using the R package ReactomePA [73] (Figure 2.12). Extracellular matrix organization and integrin cell surface interaction pathways showed selective enrichment in the markers of EC 1, while the plasma lipoprotein assembly, remodeling, and clearance pathway demonstrated selective enrichment in EC 2. Functional pathway scores based on these signatures (Table 2.3) were significantly different (p-value < 2.2 x 10^{-16} by Mann-Whitney U test) between the two EC subpopulations and provided a functional biological hypothesis for the observed separation in transcriptional space (Figure 2.11B).

Because the subpopulation markers specific to EC 2 included multiple angiogenic factors, we also investigated the ability of angiogenic gene sets to differentiate the two subpopulations. Angiogenic heterogeneity in ECs is often summarized in the distinction between “tip” and “stalk” cells [86]. We used a previously identified gene signature for tip cells [87] (Table 2.4) to generate a tip cell score for all observed endothelial cells in our data set. This score significantly differentiated EC 1 and EC 2, with increased expression in EC 2 (Figure 2.11C, p-value < 2.2 x 10^{-16} by Mann-Whitney U test).

2.3.6 Single-cell RNA-seq in aortas from Western diet-fed mice reveal conserved and diet-dependent markers of endothelial subpopulations

In order to determine the effect of diet on the endothelial subpopulation markers, we analyzed aortic scRNA-seq profiles from mice fed normal and Western diet.
Figure 2.12 Reactome pathway enrichment for the top 50 markers of EC 1 (left) and EC 2 (right) using R package ReactomePA. Dot size indicates number of marker genes included in pathway and dot color indicates Benjamini-Hochberg adjusted p-value. Reactome pathways with adjusted p-values < 0.05 are displayed.
**Table 2.3** EC subpopulation pathway signature gene lists

<table>
<thead>
<tr>
<th>ECM Organization</th>
<th>Integrin Signaling</th>
<th>Lipoprotein Handling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bgn</td>
<td>Vcam1</td>
<td>Gpihbp1</td>
</tr>
<tr>
<td>Vcam1</td>
<td>Fn1</td>
<td>Lpl</td>
</tr>
<tr>
<td>Mfap5</td>
<td>Thbs1</td>
<td>Scarb1</td>
</tr>
<tr>
<td>Bmp4</td>
<td>Col4a3</td>
<td>Mylip</td>
</tr>
<tr>
<td>Fn1</td>
<td>Icam1</td>
<td>Angptl4</td>
</tr>
<tr>
<td>Lox</td>
<td></td>
<td>Apoc1</td>
</tr>
<tr>
<td>Pcolce2</td>
<td></td>
<td>Pltp</td>
</tr>
<tr>
<td>Fbln5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efemp1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thbs1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col4a3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col5a2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Icam1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col8a1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dcn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2.4 Tip gene signature gene list

<table>
<thead>
<tr>
<th>Tip Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nrp2</td>
</tr>
<tr>
<td>Cxcr4</td>
</tr>
<tr>
<td>Cd34</td>
</tr>
<tr>
<td>Dll4</td>
</tr>
<tr>
<td>Apln</td>
</tr>
<tr>
<td>Kdr</td>
</tr>
<tr>
<td>Plxnd1</td>
</tr>
<tr>
<td>Unc5b</td>
</tr>
<tr>
<td>Efnb2</td>
</tr>
<tr>
<td>Angpt2</td>
</tr>
<tr>
<td>Nrp1</td>
</tr>
<tr>
<td>Robo4</td>
</tr>
<tr>
<td>Flt4</td>
</tr>
<tr>
<td>Pdgfb</td>
</tr>
</tbody>
</table>
Differentiating conserved subpopulation markers from markers altered by diet requires two distinct methods for dataset alignment, as previously described [76]. PCA-based alignment (Figure 2.13A) highlights variation between the datasets. CCA-based alignment (Figure 2.13B) accentuates the similarities between the two datasets. A clustering algorithm applied to the datasets aligned for similarity (Figure 2.13C) showed three major subpopulations as previously identified in normal aortic data. These subpopulations are distinguished by the same marker profiles as previously determined. EC 1 specifically expresses Cyt11, Clu, and Gkn3; EC 2 specifically expresses lipid-handling genes including Fabp4, Gpihbp1, and Lpl; and EC 3 expresses markers of lymphatic identity including Lyve1. All EC subpopulations expressed the EC-specific markers differentiating them from other aortic cell types under both dietary conditions. Two minor populations identified via clustering consisted of far fewer cells than the major populations and expressed markers from multiple major clusters, possibly signifying the presence of cell doublets.

This analysis also identified markers of the response to Western diet which are common to all EC subpopulations (Figure 2.13D). Markers of contractile function including myosin light-chain 9 (Myl9), transgelin (Tagln), and alpha smooth muscle actin (Acta2) were increased in all subgroups with exposure to high-fat diet. Comparison of the two scRNA-seq datasets using CCA-based alignment demonstrates that the presence of distinct EC subpopulations persists after exposure to Western Diet, and they share a common response feature of upregulation in contractile gene expression.
Figure 2.13 Conserved and diet-dependent markers of EC subpopulations. a) Unaligned normal- and Western diet EC scRNAseq profiles represented in a PCA-based space with t-SNE. b) Aligned normal- and Western diet scRNAseq profiles represented with CCA and t-SNE. c) Conserved subpopulations of ECs identified in both datasets. d) Conserved diet-independent markers of EC subpopulation and pan-subpopulation markers of diet-induced change. Diet-independent subpopulation markers are the same as those identified previously while Western diet induces contractile proteins Myl9, Tagln, and Acta2 in all EC subpopulations.
2.3.7 Expression of Mendelian aortopathy genes identifies relevant cells and biologic pathways

We next applied this aortic single cell gene expression data to determine the role different cell types play in the pathogenesis of aortic diseases. We determined the cell-specific expression of 30 genes with known coding mutations associated with disease. The plurality of the genes show highest expression in VSMCs, though several display significant expression in other cell types (Figure 2.14A). As expected, genes known to affect muscle contraction (e.g. Myhlk, Myh11) are predominantly expressed in VSMCs. Genes with extracellular matrix functions are expressed most highly in fibroblasts (e.g. Col5a2, Col5a1, Col3a1). Several genes associated with thoracic aortic aneurysm and dissection (TAAD) syndromes are expressed in multiple cell types, though the major cell types are VSMCs, fibroblasts, and ECs. When cell type specific expression is limited to these 3 relevant cell types, the TAAD-causing mutations separate into distinct patterns of cellular expression (Figure 2.14B). First, four VSMC-specific genes—Myhlk, Myh11, Acta2, and Flna. Several ECM-related genes lie along the VSMC-Fibroblast axis, and have minimal EC expression. These genes include Col4a5, Col5a1, Col5a2, Fbn1, Mfap5, and Col3a1. Only Notch1 has exclusive expression in ECs. There are, however, several genes with significant expression in all three cell types. These include the TGFβ pathway genes (Tgfb2, Smad2, Smad3, Smad4, Tgfbr1) and two ECM-related genes (Lox, Eln).
Figure 2.14 Expression of aortopathy genes by cell identity. a) Expression of thirty genes associated with Mendelian aortopathy by cell type. The majority of these genes have highest expression in VSMCs. Several show highest expression in fibroblasts. Two genes (Mat2a and Notch1) have highest expression in ECs. b) Ternary plot to show expression of aortopathy genes in ECs, VSMCs, and fibroblasts. Several genes are exclusively expressed in VSMCs, while the majority show expression predominantly in VSMCs and fibroblasts. Genes with contractile function are in green, TFGb response genes are in red, ECM-related genes are in blue, and unassigned genes are in black. EC indicates endothelial cells; Fibro, fibroblasts; VSMC, vascular smooth muscle cells.
2.4 Discussion

2.4.1 Summary

In this chapter, we present the use of scRNA-seq to generate an unbiased transcriptional profile of the murine aorta, which enables the definition of functional heterogeneity in three EC subpopulations \textit{in vivo}. Droplet-based scRNA-seq of the normal mouse aorta identified all major vascular cell types as determined by canonical markers. Three EC subpopulations identified by clustering analysis and expression of Cdh5 variably expressed canonical EC markers such as Vwf, but did all express a set of “pan-EC” markers linked to angiogenesis and adhesion/transport. Functional enrichment of subpopulation-specific markers suggested the two major EC subpopulations had adhesion and lipid-based profiles. The three EC populations were differentiated by distinct transcription factor profiles, including a separation between the two major populations on the basis of expression of targets activated by PPAR-gamma. Functional gene scores based on subpopulation markers and enriched gene sets also differentiated these populations. These subpopulations persisted in response to an alteration in diet and exhibited a unified contractile response to Western diet. These findings establish the presence of basal endothelial heterogeneity as described in the introduction. Additionally, these results demonstrate that \textit{in vivo} this heterogeneity is not eliminated by diet-induced endothelial plasticity.

2.4.2 EC heterogeneity \textit{in vivo}

The data presented in this chapter demonstrate heterogeneous subpopulations of ECs within the aortic endothelium whose markers suggest distinct functional characteristics. These subpopulations were identified via computational clustering of the
full EC transcriptome, and therefore meaningfully capture distinct EC ‘states’ in a way that builds upon and exceeds previous single-molecule studies of functional EC heterogeneity [16], [17]. In the dynamical systems model [21], these states would represent local minima in the phenotypic energy landscape; in the paradigm for EC heterogeneity and plasticity presented in the introduction, these specialized groups present under physiologic conditions represent basal heterogeneity. The origin of these subpopulations is an open question – two major potential hypotheses are that these cells have a distinct developmental origin or that they are the result of differences in microenvironmental stimuli which shift differently localized ECs into distinct stable states. Results obtained by our collaborators and included in the paper on which this chapter is based [88] demonstrate that Vcam1, a marker of EC 1, and Cd36, a marker of EC 2, have nonoverlapping distributions in the mouse aorta. Vcam1 demonstrated increased expression in the lesser curvature of the aorta while Cd36 demonstrated increased expression in the greater curvature of the aorta. With the caveat of using single molecules as proxies for the identified EC states, this finding suggests that shear stress, which is different between these two zones, contributes to the basal functional heterogeneity between EC 1 and EC 2. Beyond shear stress alone, proximity to branch vessels is another possible microenvironmental determinant of state. Elucidating the origins of the observed heterogeneity will require a deeper understanding of how the effect of any given stimulus – whether mechanical, chemical, or metabolic – may vary across a population of ECs. Possible future studies would focus on superimposing multiple stimuli on ECs in vitro to demonstrate the ability to drive cells towards EC 1 or EC 2-like states, or on demonstrating a shift in EC subpopulation membership in the in
vivo aorta in response to local alterations (e.g. injury, aortic banding) or systemic alterations (e.g. hyperlipidemia, hyperglycemia, pharmacology). These studies would establish the possibility of state transitions in ECs, which would in turn suggest an environmental origin for the subpopulations observed in this study.

### 2.4.3 Functional characterization of endothelial subpopulations

Moving from identification of distinct EC subpopulations towards a meaningful definition of EC state requires a method to assign subpopulations a biological identity or significance. Statistical identification of subpopulation markers results in a list of enriched transcripts, which must be interpreted via our existing understanding to achieve systems-level state descriptions. Enrichment analysis against known and previously annotated gene sets provides a basis for future functional understanding. Our results demonstrate that previously compiled lists used to discern EC identity (GO EC Development and GO EC Differentiation) are insufficient to characterize the subpopulations observed here. Instead, we propose the use of functional gene signatures defined as the intersection of subpopulation markers and enriched gene sets, which capture a distinct functional theme describing a subpopulation. In our data, such gene set signatures for multiple EC functions – ECM organization, integrin signaling, and lipoprotein handling – serve to differentiate the two major EC populations. Differentiation using a set of tip identity marker genes also suggests that the difference between EC 1 and EC 2 exists along multiple functional axes. Coupling of EC functions to each other (for example, integrin handling and lipoprotein handling in different subpopulations) has intriguing implications for EC behavior and merits further investigation. Transcription factor-based modules are a second method to organize
subpopulation-specific transcripts and interpret observed heterogeneity. Here we demonstrate that by scoring the targets of PPAR-gamma we can differentiate EC 1 and EC 2, consistent with the identified theme of lipid handling. This transcription factor-focused analysis suggests a modular approach to functional characterization based on TF targets that will be applied in future studies.

2.4.4 Understanding the impact of EC heterogeneity on plastic response to disease-relevant stimuli

CCA-based alignment of ECs under normal diet and Western diet conditions demonstrated that the three EC subpopulations identified in this scRNA-seq aortic profile are not limited to the physiologic milieu, but persist with the application of a pathology-simulating stress. This supports the claim that the heterogeneity we observe is a “basal heterogeneity” in our paradigmatic model, which exists stably in the physiologic state and interacts with pathologic stressors. The upregulation of contractile gene expression in ECs after exposure to Western diet is particularly interesting given that endothelial-mesenchymal transition (EndMT) is a known “final common pathway” in EC dysfunction. EndMT in atherosclerosis has previously been linked to an increase in smooth muscle markers [89], including Acta2. Our data suggest that in the setting of a Western diet, this alteration is part of a set of increased contractile transcripts that can be detected in all three EC subpopulations. Demonstration of a unified response by all EC subpopulations suggests that endothelial “activation” may be the result of coordinated changes among endothelial subgroups. Further investigation of this and other scRNA-seq datasets will hopefully also establish which elements of the in vivo
pathologic response differ among these subpopulations, providing valuable insight into their function and origins.

This work is limited in its consideration of only one stimulus type, and future extensions should seek to investigate the effect of multiple stimuli singly and in combination on the subpopulations described here. We would predict that given the putative lipid handling function of EC 2, its response to a lipidogenic alteration like a Western diet differs from that of EC 1. The functional descriptions of each subpopulation provided here should suggest stimuli that might interact differentially with each population (for example, injury might disproportionately affect the ECM/integrin interaction characteristic of EC 1) that can be the basis of future study.

2.4.5 Single-cell profiling of vascular tissue

Stepping back from ECs alone, these results are also novel for being an unbiased description of the complete aortic single-cell transcriptional milieu incorporating all vascular cell types. Previous studies in this domain typically utilize presorting on cell-specific markers to select for cell types of interest such as macrophages [56] or endothelial cells [59]. However, our results on heterogeneous expression of ‘canonical’ markers suggest that pre-sorting in this fashion might select against smaller subpopulations within a given cell type. This type of profiling also suggests future work that may be of interest using similar methods – for example, fibroblasts in this dataset segregated into two populations that may be functionally characterized in a similar manner. Different dissociation conditions had minimal effect on our results in the EC domain, but might optimize the generation of a single-cell suspension for the analysis of other cell types (for example, including elastase in the
Finally, our results suggest that lower-depth sequencing is sufficient to discriminate major cell types and cell type subpopulations in the aorta, which may be applicable in future vascular scRNA-seq studies.

2.4.6 Relevance of findings to aortopathy

Generation of single-cell profiles for all aortic cell types has the added benefit of enabling an investigation into the cell type origins of vascular disease. Here we apply this idea to Mendelian aortopathies by partitioning the expression of known genes associated with aortopathy among three cell types. This analysis suggests that EC-specific genes implicated in aortopathy are rare, and that the direct impact of ECs on the development of aortopathy is likely secondary to the indirect effect on other vascular cells. These conclusions are limited, however, because our partitioning strategy is only suggestive, rather than causal. The knowledge that certain aortopathy genes are most expressed in certain cell types must be confirmed by directly perturbing those genes in VSMCs, ECs, or fibroblasts and directly examining the resulting effects on phenotype. Hopefully this analysis will serve as a starting point to suggest which cell type or combination of cell types is most interesting for any given gene to begin investigation of the pathophysiological link to aortopathy. This strategy can be applied to other vascular diseases in the future as the knowledge of disease-gene links expands, and hopefully provide a useful touchstone as we seek to understand the biological significance of disease-linked genes identified via population-level screens.
3. Defining heterogeneous components of EC plasticity in vitro in response to mechanical and chemical stimulation

3.1 Introduction

3.1.1 Context

In the previous chapter, we established the presence of functional basal heterogeneity in the endothelium of large arteries, and demonstrated that these populations are preserved during, and exhibit unified plasticity in response to, an environmental stimulus. The next step in understanding the potential nature and significance of this heterogeneity is to determine how a population of endothelial cells (ECs) change their state in response to a specific stimulus and how the presence of multiple simultaneous stimuli can result in variation in EC plasticity. These types of highly controlled, stimulus-specific analyses require an isolated system for EC culture with well-defined environmental conditions. Demonstration of functional heterogeneity in response to in vitro stimuli singly or in combination would support the prevailing model that discrete EC states observed in vivo arise from the superposition of numerous microenvironmental drivers. Moreover, it is likely that the responses to stimuli are not linear and additive independent effects but instead complexly interlinked and co-dependent.

In this chapter, we report the use of single cell RNA sequencing (scRNA-seq) to characterize the plasticity of an EC population in response to a unimodal mechanical stimulus – shear stress – and to a bimodal superposition of mechanical and chemical stimuli – shear stress and rapamycin. We investigate plasticity in terms of both “supervised” transcriptional modules based on prior knowledge and “unsupervised”
transcriptional modules based on computational identification of correlated genes. These results demonstrate that the response to even a single stimulus in an isolated EC population can be heterogeneous, and that this heterogeneity has functional consequences for plasticity to subsequent stimuli.

3.1.2 Shear stress as a mechanical stimulus in ECs

Understanding the response of ECs to shear stress is a cornerstone of vascular biology, and has been the subject of intense investigation for decades [10], [90], [91]. The interaction of ECs and shear stress generated by blood flow is not only a specific force-response relationship but in many respects the sine qua non that defines the cell that comprises the endothelium. Suffice to say here, there is almost no function of ECs which is unaffected by alterations in shear stress. In general – and grossly simplified – terms, a distinction is drawn between laminar shear stress, which is vasoprotective, and nonuniform or disturbed shear stress, which is deleterious to the vasculature [91]. The notion of a physiologic or vasoprotective shear stress is typically conflated with high or laminar shear stress, while pathologic or disturbed shear stress is modeled using low or oscillating shear stresses. Steady, laminar shear stress of high magnitude upregulates cytoprotective and antioxidant genes, while low or disturbed shear alters morphology, cytoskeletal dynamics, proliferation, migration, permeability, and gene expression [91]. The biology of the endothelial response to shear is well-explored and expansive, and therefore is often best understood via the identification of key transcription factors (TFs) that drive larger transcriptional programs in response to shear stimuli [90]. This idea will form the basis for the characterization of the shear-induced EC state presented in this chapter.
3.1.3 Rapamycin as a chemical stimulus in ECs

The choice of rapamycin as a paradigmatic chemical stimulus in this work was motivated by its clinical relevance to vascular pathology. Currently, rapamycin and its derived compounds are a major class of agents utilized in drug-eluting stents to prevent the neointimal hyperplastic response to angioplasty and endovascular manipulation that leads to clinical restenosis. Their pharmacology and biology are complex [92].

Rapamycin’s mechanism of action in ECs involves interaction with a variety of tissue and plasma proteins, interaction with vasoactive factors even in absence of classic ligands [93] and profound protein binding to enhance local tissue retention. Rapamycin induces complex biologic responses including not only inhibition of smooth muscle cell (SMC) proliferation but also acceleration of neoatherosclerosis [94]. Rapamycin’s inhibition target, the molecular or mammalian Target of Rapamycin (mTOR), interacts with a number of ancillary proteins that comprise different complexes. The two main complexes currently identified as targets have differential effects on ECs – activation of mTORC1 leads to endothelial migration and proliferation which is implicated in intimal hyperplasia, while activation of mTORC2 leads to endothelial survival processes including retention of endothelial integrity and barrier formation. Rapamycin or sirolimus is typically a selective mTORC1 inhibitor, but can act on mTORC2 with more sustained administration; alternate non-mTOR pathways for rapamycin impairment of EC barrier formation act via RyR2-mediated calcium dynamics [94].

Thus, given the complexity of its action and the opposing pathways at play, the nature and magnitude of rapamycin’s effect on ECs are still areas of active investigation. On the one hand, sirolimus has been shown to have a minimal effect on
EC migration *in vitro* with no dose-dependence [95], but on the other, administration of rapamycin in ECs has demonstrated induction of gene expression of multiple pivotal vascular genes including KLF-2, TF, and PAI-1 [96]. Our work thus has the potential not only to examine the effect of superimposed stimuli on EC plasticity but to do so with a compound that is of great clinical significance [97] and partially defined biology. These rationales motivate the choice of rapamycin as our paradigmatic chemical stimulus whose plastic effect on ECs can be illuminated by scRNA-seq.
3.2 Methods

3.2.1 Endothelial cell culture

Primary human aortic endothelial cells (HAECs, PromoCell) at passages P5-P7 were used for all experiments. The media used for cell culture (Lonza EGM-2 or PromoCell Growth Medium 2) and in the flow-controlled bioreactor was supplemented to a total serum concentration of 7% using fetal bovine serum (FBS). Media was additionally supplemented with 100 U penicillin/mL, 100 µg streptomycin/mL, and 1 µg amphotericin/mL.

3.2.2 Multimodal culture of ECs in a flow-controlled bioreactor

A previously described controllable flow loop bioreactor system was utilized for culture of HAECs [98]. Briefly, the flow loop was driven by a peristaltic pump (ISIATEC MCP, Product No. 78002-00) with an attached 8 roller, 12 channel pump head (ISIATEC Type IBM733A, Product No. 78002-36). ISIATEC PharMed BPT pump tubing with an internal diameter of 2.79 mm (Cole Parmer) was utilized in the flow loop. Silastic laboratory tubing with an internal diameter of 0.188 in/4.78 mm and an outer diameter of 0.312 in/7.92 mm (Dow Corning) connected the pump tubing and the cell culture chamber to a custom glass media reservoir, and 3/16" x 3/32” tubing connectors (Cole Parmer) connected the pump tubing and the Silastic tubing. The non-cell chamber components of the flow loop were preassembled and autoclaved prior to flow loop assembly.

HAECs were seeded in Ibidi µ-Slide I Luer0.8 flow chambers with ibiTreat proprietary tissue culture coating for all experiments. These chambers have a height of 800 microns, a channel volume of 200 µL, a channel length of 50 mm, a channel width
of 5 mm, a reservoir volume of 60 µL, and a growth area of 2.5 cm² per channel. Ibidis were seeded with 200 µL of HAECs at a 0.6-1.2 million cells/mL concentration at least 12 hours before connecting to flow; cell attachment and confluence were confirmed by microscopy before initiating shear conditioning.

The following equations were used to determine the pump speed necessary to achieve a desired shear stress on ECs cultured in the Ibidi:

\[
Pump\ speed\ [RPM] = \left(1.33\ \frac{\text{rotations}}{\text{mL}}\right) \left(\text{Flow\ rate}\ \frac{\text{mL}}{\text{min}}\right) + 0.0133\ RPM \quad (1)
\]

\[
Shear\ stress\ \frac{\text{dyn}}{\text{cm}^2} = \text{Dynamical\ viscosity}\ \frac{\text{dyn}\cdot\text{s}}{\text{cm}^2} \times 34.7\ \frac{\text{min}}{\text{mL}} \times \text{Flow\ rate}\ \frac{\text{mL}}{\text{min}} \quad (2)
\]

Equation 1 is derived from the ISMATEC MCP pump manual, which specifies that for a CA12 roller with 2.79 mm pump tubing the relationship between pump speed and flow rate should be linear with 1 rotation per minute (RPM) = 0.74 mL/min and 240 RPM = 180 mL/min, and equation 2 from Ibidi Application Note 11, “Shear Stress and Shear Rates for ibidi µ-Slides”. The viscosity of cell culture medium at 37 °C is estimated as 0.007 dyn-s/cm². Using these equations, the pump flow rates of 164.0 and 13.7 RPM used in these experiments achieved shear stresses of ~ 30 and 2.5 dynes/cm² respectively.

3.2.3 Unimodal shear stimulus experiment

HAECs were cultured using the flow loop bioreactor system under constant high/physiologic (30 dynes/cm²) or low/altered (2.5 dynes/cm²) shear for 5 days.
Following shear culturing, cells were isolated from Ibidis and used as input to Seq-Well for scRNA-seq.

3.2.4 Bimodal shear and rapamycin stimulus experiment

HAECs were cultured using the flow loop bioreactor system under constant high/physiologic (30 dynes/cm\(^2\)) or low/altered (2.5 dynes/cm\(^2\)) shear for a total of 3 days. 24 hours after initiation of flow, rapamycin (InSolution, Calbiochem/Sigma-Aldrich, Prod No. 553211-1MG) dissolved in media was added to half the flow loops to achieve a final concentration of 1 µM, which has been previously demonstrated to result in alterations in RNA expression in ECs [96]. An equal volume of normal media was added to the other flow loops. Following shear culturing, cells were isolated from Ibidis and used as input to Seq-Well for scRNA-seq.

3.2.5 Single-cell transcriptional profiling of cultured ECs via Seq-Well

Following flow conditioning, cells were harvested from Ibidi flow chambers and single-cell RNA sequencing was performed via Seq-Well, a picowell-based approach to scRNA-seq developed to be appropriate for low-input (~10\(^4\) cells) samples [34]. Seq-Well v2 using second strand synthesis was utilized for all samples. 10,000 cells per condition were loaded as the input to Seq-Well; two replicates per shear condition were included in the unimodal experiment and one array per condition was included in the bimodal experiment. Following Seq-Well, the Nextera XT DNA tagmentation method was used for library preparation with SPRI ratios of 0.6x and 0.8x applied to the amplified and tagmented product. Average fragment size distribution was between 400 and 600 bp for all sequencing libraries. Libraries were loaded at a final loading
concentration of 2.2 pM with a read structure as follows: Read 1 – 20 bp with 12-bp cell barcode and 8-bp unique molecular identifiers (UMIs), Read 2-50 bp. The bcl2fastq library was utilized to generate FASTQ files which were pre-processed for barcode and UMI information using the pipeline designed for Drop-seq [33]. Preprocessed reads were aligned to the hg19 genome using STAR aligner. Mapped reads filtered for quality were used to generate digital gene expression matrices (DGEs) for each sample. UMI-collapsed data were used to generate the final counts table used for analysis; DGEs were combined from separate array samples by retaining the genes detected in all samples to generate counts tables for further analysis.

3.2.6 Computational analysis and cell quality control

Raw counts tables were used as the input for analysis in the Seurat package for R (Version 2.3.1) [69], [70]. Genes detected in at least 10 cells and cells with at least 500 genes were retained for analysis; cells with >30% mitochondrial reads were excluded from analysis. Following quality control ~8,300 cells were analyzed from the unimodal experiment with an average of ~9,000 reads/cell, ~2,000 genes/cell, and ~5,500 UMI/cell. From the bimodal experiment, ~14,900 cells were analyzed with an average of ~10,000 reads/cell, ~1,900 genes/cell, and ~6,000 UMI/cell. Normalized expression values were computed for expression in each cell by scaling each cell to 10,000 total UMIs and log-normalizing these values with an added pseudocount of 1. Variable genes were identified using the Seurat FindVariableGenes function with the LogVMR dispersion function parameter; genes with average expression between 0.0125 and 3 and dispersion of at least 0.5 were considered variable.
3.2.7 Dimensional reduction of shear conditioned EC data

Dimensional reduction of the shear-conditioned EC scRNA-seq profiles was performed on scaled normalized count data using principal components (PCs) 2-10 based on examination of the PC significance plot (Figure 3.1A); PC 1 was excluded due to its correlation with QC metrics (Figure 3.1B). The Seurat implementation of the t-distributed stochastic neighbor embedding (t-SNE) algorithm was used to project PCs 2-10 into a 2D space for demonstration of shear state separation.

3.2.8 Differential expression analysis

Identification of differentially expressed transcripts between conditions (e.g. physiological vs. altered shear states or control vs. rapamycin) was performed using the FindMarkers function in Seurat with a minimum log-fold change threshold of 0.25 and with p-values computed using a Wilcoxon rank-sum test. Subpopulation-specific markers following cluster analysis were identified using the FindAllMarkers function in Seurat with a minimum log-fold change threshold of 0.25 and with p-values computed using a Wilcoxon rank-sum test. P-values were adjusted for multiple hypothesis correction in both cases using a Bonferroni correction based on the total number of genes in the analyzed dataset; adjusted p-value < 0.05 defined a significant marker.

3.2.9 Gene set enrichment for defined marker lists

Gene set enrichment on gene lists describing differentially expressed transcripts between two conditions or enriched markers of subpopulations was performed using the
Figure 3.1 Principal component selection for shear-cultured EC analysis. A) PC significance plot for unimodal shear-cultured EC scRNA-seq profiles. B) Correlation of PC 1 with QC metrics; nGene is the total number of genes/cell while nUMI is the total number of UMIs/cell.
R library clusterProfiler [99]. The clusterProfiler enricher function was applied to perform a hypergeometric test setting the universe parameter of background genes to all genes expressed in at least 1% of the cells in the analyzed dataset. P-values were adjusted using a Benjamini-Hochberg procedure with a q-value cutoff of 0.05. Annotated gene sets were drawn from a custom GMT file derived from the “canonical pathways” (CP) and Gene Ontology (GO) gene sets in v6.2 of the Molecular Signatures Database (MSigDB) [100]. These include gene sets from the Biocarta, Kyoto Encyclopedia of Genes and Genomes (KEGG) [101], Reactome [74], and GO [102], [103] databases.

3.2.10 Generation of supervised shear-altered transcription factor-based modules

Candidate transcription factors to be used as the basis for supervised modules were identified from the literature and from differentially expressed markers of high/low shear in the dataset (Table 3.1). The known targets of each transcription factor in humans were identified using the v2 of the TRRUST database [75]. For each transcription factor, targets that were known to be activated by that factor were included in one candidate gene set (annotated with the suffix ‘a’) and targets known to be repressed by that factor were included in a second set (annotated with suffix ‘r’). Targets with unknown relationships were excluded from gene set generation. In order to avoid gene set correlations driven by gene overlaps, the overlap between each gene set was assessed as a fraction of set size (Figure 3.2) and sets with overlap proportion > 0.33 with another set were merged into the larger set. Sets with fewer than 4 genes detected in the shear conditioned EC dataset were excluded, resulting in the final list of 11 supervised shear-responsive TF modules (Table 3.2). These modules had reduced overlap relative to the original candidate modules (Figure 3.3).
<table>
<thead>
<tr>
<th>TF</th>
<th>Source</th>
<th>Known EC function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPAS1</td>
<td>Lit + High markers</td>
<td>Expression reduced in HCAEC w/ LSS; angiogenesis</td>
<td>[104]</td>
</tr>
<tr>
<td>ETS2</td>
<td>High markers</td>
<td>Promotes inflammation, Promotes tube formation, leakiness</td>
<td>[105]</td>
</tr>
<tr>
<td>NCOA4</td>
<td>High markers</td>
<td>Hematopoietic - no EC connection found</td>
<td></td>
</tr>
<tr>
<td>KLF2</td>
<td>Lit + High markers</td>
<td>Up in high LSS, eNOS expression Maintenance of quiescent phenotype</td>
<td>[90], [106]</td>
</tr>
<tr>
<td>KLF4</td>
<td>Lit</td>
<td>Controls NOS; OSS decreases transcription vs. pulsatile SS Regionally variable in swine aorta</td>
<td></td>
</tr>
<tr>
<td>NFE2L2</td>
<td>Lit + High markers</td>
<td>Up in high LSS ROS response Antioxidant production</td>
<td>[90]</td>
</tr>
<tr>
<td>TCF4</td>
<td>Low markers</td>
<td>Interacts with beta-catenin Mechanism disrupted by NO</td>
<td>[107]</td>
</tr>
<tr>
<td>FUS</td>
<td>Low markers</td>
<td>Decreases in quiescent/confluent ECs</td>
<td>[108]</td>
</tr>
<tr>
<td>FHL2</td>
<td>Low markers</td>
<td>Inflammatory, increases NF-kB Deletion attenuates atheromas</td>
<td>[109], [110]</td>
</tr>
<tr>
<td>KLF9</td>
<td>High markers</td>
<td>Associated miRs increase with DM-like environment</td>
<td>[111]</td>
</tr>
<tr>
<td>NR4A1</td>
<td>High markers</td>
<td>Stimulated by ox-LDL Higher levels cause mitophagy, apoptosis in aortic ECs</td>
<td>[112]</td>
</tr>
<tr>
<td>MEF2A</td>
<td>High markers</td>
<td>Pro-proliferation</td>
<td>[113]</td>
</tr>
<tr>
<td>SNAI1</td>
<td>Lit</td>
<td>Up in HUVECs, PAECs with low SS Induces EndMT</td>
<td>[114]</td>
</tr>
<tr>
<td>TWIST1</td>
<td>Lit</td>
<td>Up in low shear stress Induces EndMT Enhances EC proliferation + leak Developmental angiogenesis</td>
<td>[115]</td>
</tr>
<tr>
<td>SMAD3</td>
<td>Lit</td>
<td>Up in OSS vs. LSS in HUVECs EndMT</td>
<td>[116], [89]</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Lit</td>
<td>Activated by OSS but not LSS Nuclear in atheroprone regions Promotes active, migratory phenotype</td>
<td>[116], [117]</td>
</tr>
<tr>
<td>FOXP1</td>
<td>High markers</td>
<td>Pro-angiogenic</td>
<td>[118]</td>
</tr>
<tr>
<td>ELK3</td>
<td>High markers</td>
<td>Anti-angiogenic, inhibits ETS1</td>
<td>[119]</td>
</tr>
</tbody>
</table>
Figure 3.2 Overlap percentages between candidate supervised TF sets. Percentages (pct1overlap) and color indicate the proportion of Set 1 genes (TF target sets on the y-axis) also included in Set 2 (TF target sets on the x-axis).
Table 3.2 Supervised shear-responsive TF modules used in analysis.

<table>
<thead>
<tr>
<th>Set ID</th>
<th>Annotation</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>epas1a</td>
<td>EPAS1-activated</td>
<td>CCR7, FLT1, MMP14, MSC, SERPINE1, VEGFA</td>
</tr>
<tr>
<td>ets2a</td>
<td>ETS2-activated</td>
<td>ANGPT2, ANPEP, ERG, FLI1, ICAM1, MDM2</td>
</tr>
<tr>
<td>klf2r</td>
<td>KLF2-repressed</td>
<td>CXCR4, FABP5, PPARG, WEE1</td>
</tr>
<tr>
<td>klf4r</td>
<td>KLF4-repressed</td>
<td>BIRC5, CCNB1, CCND1, CDKN1A, GDF15, IFITM3, LNX, ODC1, RARA, SOD1, TP53, VDR</td>
</tr>
<tr>
<td>nfe2l2a</td>
<td>NFE2L2-activated</td>
<td>BRCA1, CAT, HMOX1, MTHFR, NQO1, SOD1, SOD2</td>
</tr>
<tr>
<td>mef2aa</td>
<td>MEF2A-activated</td>
<td>JUN, MAPK14, MAPK7, MMP10, NRF1</td>
</tr>
<tr>
<td>twist1a</td>
<td>TWIST1-activated</td>
<td>AKR1C2, AKT2, ATM, BMI1, CD44, CTPS1, DNM3, FN1, FOS, GDF15, ICAM1, ITGB1, MAPK1, MAPK3, MMP2, NR2F1, YBX1</td>
</tr>
<tr>
<td>twist1r</td>
<td>TWIST1-repressed</td>
<td>ERBB3, ESR1, F2R, ILK, NF1, PFDN4, RAP1A, RBL2, SRPX, TIMP1</td>
</tr>
<tr>
<td>smad3a</td>
<td>SMAD3-activated</td>
<td>ANGPTL4, BAMBI, CCND2, CDKN1A, FST, PTHLH, SERPINE1, TNC, VDR</td>
</tr>
<tr>
<td>klf2-4a</td>
<td>KLF2 and KLF4-activated</td>
<td>CDKN1A, NOS3, THBD, ATF3, BDKRB2, CCND1, CD14, CDH5, CDKN1B, CDKN1C, HSPA8, IL6, KRT19, LAMA3, MMP2, PFKP, TP53</td>
</tr>
<tr>
<td>tcf4-ctnnb1a</td>
<td>TCF4 and CTNNB1-activated</td>
<td>ABCB1, CCND1, MYC, PLD1, VEGFA, VIM, JUN, MMP14, STARD7</td>
</tr>
</tbody>
</table>
Figure 3.3 Overlap percentages between supervised shear-responsive TF modules used in analysis. Percentages and color (pct1overlap) indicate the proportion of Set 1 target genes (y-axis) also included in Set 2 (x-axis).
3.2.11 Supervised module-based characterization of shear stress-induced EC state

A score was computed for each supervised TF module for each cell as the total number of post-normalization transcripts (out of the 10,000 normalized total) from genes in the module. Correlations between supervised TF modules across all shear-conditioned cells were determined using Pearson correlation and a dendrogram was generated using the R function hclust with the Euclidean distance and complete linkage clustering parameters. Correlations were also determined to technical variables (Figure 3.4) to ensure quality metrics were not driving the observations. Estimates of the significance of correlation strength between any two modules were determined by Monte Carlo approximation of the Pearson correlation probability distribution for two randomly generated gene sets with module size and overlap size identical to the tested modules. Linear discriminant analysis on scaled and centered supervised TF module scores was used to determine the linear combination of scaled scores that best discriminated high and low shear condition. To demonstrate heterogeneity in supervised TF module expression across shear-conditioned cells, cells in the dataset were clustered based on module scores using a Euclidean distance metric and complete linkage clustering.

3.2.12 Whole-genome correlation network analysis (WGCNA)-inspired unsupervised gene module identification

Differentially expressed genes between rapamycin and control conditions were identified in both physiologic (high) and altered (low) shear contexts as described above
Figure 3.4 Correlations between supervised shear-responsive TF modules and technical variables. Color and number indicate the Pearson correlation coefficient between technical variables (number of genes, nGene; number of UMIs, nUMI; or percentage of mitochondrial genes, percent.mito) and TF module scores.
using a logFC threshold of 0.25. Gene modules within these rapamycin response gene sets which were co-expressed were identified using functions from the R WGCNA library [120]. WGCNA was applied with a soft power of 6 used to create the adjacency matrix, transformation of adjacency matrix into a Topological Overlap Matrix (TOM), and clustering of the TOM using the cutreeDynamic function from WGCNA with method ‘tree’ using a minimum module size of 10 as previously described [121]. Module significance was estimated using a permutation test as previously described [121]. Briefly, genes from the total list of genes used for module discovery were binned by expression and 10,000 random modules with gene expression distributions like that of the tested module were generated. For each random module, a one-sided Mann-Whitney U test was performed to compare the distribution of dissimilarity values between the tested module and the random module. The p-values of these tests are corrected for multiple hypothesis correction with the Benjamini-Hochberg procedure and the number of tests with FDR>0.05 (indicating that the random module and tested module have similar dissimilarity distributions) was determined. If fewer than 5% of tests (p<0.05) fail to reject the null hypothesis of the Mann-Whitney U-test, the tested module was deemed to have significant correlation structure relative to background.

3.2.13 Correlation of unsupervised gene modules defining rapamycin response across shear conditions

To determine whether significant correspondence existed between unsupervised gene modules identified from rapamycin responses of physiologic shear and altered shear cells, we implemented a Fisher’s exact test for significance of gene overlap. The significance of overlap size between each pair of modules was computed using the set
of all differentially altered genes across both shear conditions as the gene universe. P-values were adjusted using the Benjamini-Hochberg procedure.

### 3.2.14 Canonical correlation analysis (CCA) for integrated identification of subpopulations within shear-conditioned EC populations

To identify subpopulations with potentially different responses to rapamycin within the populations of shear-conditioned cells, we used a CCA procedure as previously described [76] to analyze ~7,300 cells from all 4 conditions. Variable genes used for defining canonical variables were drawn from the union of the top 2000 variable genes across rapamycin-exposed cells and the top 2000 variable genes across control cells. CCA was performed using the Seurat function RunCCA and CCA subspaces were aligned across drug condition using the AlignSubspace function with CCs 1:13. Dimensional reduction for representation was performed via t-SNE using CCs 2:13; CC 1 was excluded due to correlation with QC variables (Figure 3.5). Subpopulations were identified using a shared nearest neighbors-based approach as implemented in the Seurat FindClusters function using a resolution parameter of 0.4. Markers were identified for each subpopulation using the Seurat FindMarkers function as described above.

### 3.2.15 Characterization of rapamycin-induced plasticity across subpopulations

The mean log fold change for a given gene module in a particular subpopulation was calculated by first determining the average normalized (to 10,000) expression of each gene in that set across all subpopulation-condition pairs. The average log fold change between the rapamycin and control conditions for a given gene in a
Figure 3.5 Correlation of CCs 1 and 2 with technical variables in CCA for integrated identification of shear-conditioned subpopulations. Cell color in nGene plot scales with total number of genes/cell; cell color in nUMI plot scales with number of UMIs detected/cell.
subpopulation was calculated as the natural log of the ratio of average expression in the subpopulation with rapamycin to the average expression in the subpopulation without rapamycin, with both expression values adjusted using a pseudocount of 1. The mean of this value across all genes in the module was taken as the mean log fold change for a given gene module in a given subpopulation.
3.3 Results

3.3.1 Single-cell RNA-seq resolves the effect of shear stress alteration on aortic endothelial cells

We used picowell-based massively parallel scRNA-seq to characterize HAEC populations subjected to either high (physiologic, 30 dyn/cm$^2$) or low (altered, 2.5 dyn/cm$^2$) laminar shear conditions in a controlled bioreactor system. Dimensional reduction using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) demonstrated that scRNA-seq profiles clearly resolved high and low shear-conditioned EC populations from each other (Figure 3.6A). Two shear replicates from each condition clustered with each other, demonstrating reproducibility of shear conditioning and transcriptional profiling (Figure 3.6B).

Markers of each shear condition were identified using the Wilcoxon rank-sum test as implemented in Seurat (Figure 3.6C). Top genes induced by high shear include a number of genes which have previously been associated with high shear stress in HAECs and other human endothelial cell populations including PI16, a peptidase inhibitor [122]; CD59, a glycoprotein [123]; ADAMTS1 [124]; and IGFBP5, a mechanosensor in ECs [125]. LYVE1, typically considered a marker of lymphatic ECs, was also a marker of high shear in this HAEC population. Top markers of low shear included CRIM1, CDH2, PTPRF, and TNFAIP8L1. Low shear was also marked by broad upregulation in mRNA expression of multiple ribosomal proteins and multiple mitochondrial proteins as broad signatures.
Figure 3.6 Shear-induced scRNA-seq profiles. A) t-SNE representation of HAEC scRNA-seq profiles under high and low shear. B) t-SNE representation of HAECs grouped by replicate. C) Heatmap of scaled expression of top 10 markers by shear condition and ribosomal and mitochondrial transcripts.
3.3.2 Shear stress-induced endothelial state can be partitioned into discrete supervised modules based on known shear-responsive transcription factors

Shear stress-associated transcriptional factors were identified from existing literature and the set of shear state-associated markers identified from the scRNA-seq dataset. The targets of these transcription factors were identified using the TRRUST database and used to define candidate supervised transcriptional modules to be used in partitioning shear-induced EC state. After combining modules with highly overlapping gene membership and imposing a minimum size constraint, 11 modules were included in the final analysis (Table 3.2). These modules each consisted of the targets of a transcription factor that are known to be activated (module designations ending in ‘a’) or repressed (module designations ending in ‘r’) for a given shear-responsive transcription factor.

Cells in the unimodal shear-only dataset were scored for total post-normalization expression of the genes in each shear-responsive module (Figure 3.7A). Average expression of each module for each shear condition was also computed (Figure 3.7B). The majority of supervised shear-responsive modules demonstrated a difference in average expression between low and high shear conditions, with the weakest shear-responsive effects observed in the ET2-activated module (ets2a), the MEF2A-activated module (mef2aa), and TWIST1-repressed module (twist1r). The majority of modules demonstrated average alterations with shear consistent with their described roles in the literature. The KLF4-repressed module (klf4r), which would be expected to decrease with high shear stress but instead increases, is a notable exception. The combined TCF4 activated/CTNNB1 activated module (tc4-ctnnb1a), which would be expected to increase with low shear stress, also behaves contrary to expectations.
Figure 3.7 TF module expression scores as a function of shear state. A) Violin plots of expression scores (transcripts from TF module-included genes as a fraction of the overall transcriptome) for 11 supervised shear-responsive TF modules. Scores are separated by shear condition. B) Average expression scores as a fraction of total transcriptome by shear condition.
Examination of the correlation of supervised module scores across all cells in the shear conditioning experiment (Figure 3.8) reveals how these shear-response programs are coupled. A Monte Carlo-based estimation method using randomly generated modules with equivalent size and overlap characteristics enabled quantification of the strength of any given relationship. For two modules of median size (9 genes) with no overlap, the expected 1st-percentile of correlation strength is ~0.05 and the expected 99th-percentile is 0.06. The correlation between the SMAD3-activated module (smad3a) and the EPAS1-activated module (epas1a), r=0.93, is particularly notable; Monte Carlo estimation suggests that this relationship is stronger than 99.9% of correlations using randomly generated gene sets with identical characteristics. The negative correlation between the KLF2/KLF4 activated set (klf2-klf4a) and the KLF2 repressed set (klf2r), r=-0.05, is also notable for being below the 1st percentile of the expected correlated strength distribution. The unexpected positive correlation between the KLF2/KLF4 activated set (klf2-klf4a) and the KLF4 repressed set (KLF4a), r=0.75, exceeds the 98th percentile of the expected correlation strength distribution.

Hierarchical clustering of TF modules based on their correlation across shear-conditioned cells demonstrates the modular organization of the shear response (Figure 3.9). The two major branches of the shear-responsive module dendrogram separate modules associated with high shear-induced state from modules associated with low shear-induced state. Notably, EPAS1-activated genes (epas1a) cluster together with SMAD3-activated genes (smad3a) and TWIST1-activated genes (twist1a) in the low shear-associated branch; all three transcription factors have previously been associated with endothelial-mesenchymal transition (EndMT) [89], [115], [126]. This ordering also
Figure 3.8 Correlation scores between supervised shear-responsive TF modules across shear-conditioned HAECs. Color and values indicate Pearson correlation coefficient between each pair of TF module scores across all cells in the unimodal shear-induced state experiment.
suggests the role of TFs whose functions have previously not been linked conclusively to shear stress; for example, the MEF2A-activated set (mef2aa) clusters with the high shear state-defining branch.

3.3.3 Supervised transcriptional modules discriminate shear-induced endothelial states and demonstrate functional heterogeneity in the shear stress response

Next, the relative utility of these supervised shear-responsive transcriptional modules in distinguishing shear-induced endothelial states was determined. Using linear discriminant analysis, a linear combination of shear-responsive module scores was identified which maximally distinguishes the high-shear EC population from the low-shear population (Figure 3.10A). The linear discriminant resolves the low and high-shear induced states with a partial overlap, indicating that these transcriptional modules are useful for differentiating shear-induced state but do not capture the distinction between shear-induced EC states as completely as the unbiased transcriptional profile obtained with scRNA-seq. The relative contribution of each supervised module to shear-induced EC state is determined by its coefficient in the shear-state distinguishing linear discriminant (Figure 3.10B). The modules most contributory to defining the low shear state are EPAS1-activated (epas1a), KLF2-repressed (klf2r), and TWIST1-activated (twist1a). The modules most contributory to defining the high shear state are NFE2L2-activated (nfe2l2a) and KLF2/KLF4-activated (klf2-4a). Ordering all individual cells by the shear-distinguishing linear discriminant demonstrates the correlation of each of the two major TF branches with shear state (Figure 3.10C).
Figure 3.9 Dendrogram summarizing relationships among shear-responsive TF modules. TF modules are grouped according to relative similarity as defined by Pearson correlation across all shear-cultured cells. Blue line indicates major division between high shear-associated and low shear-associated TF sets.
Figure 3.10 TF module-based discrimination of shear state. A) Linear discriminant (LD)-based separation of shear conditions using supervised TF modules. B) Contribution of supervised TF modules to LD as defined by module coefficient in the linear discriminant. C) Heatmap demonstrating supervised TF module contribution to definition of shear-induced EC state. LD1 indicates the linear discriminant-defined coordinate, flow indicates flow state, and scaled TF scores are included in the heatmap. Cells are organized by LD coordinate and TF modules are separated according to the major division in the TF score dendrogram to resolve low shear- and high shear-associated modules.
Clustering of individual cells by supervised module expression enables resolution of heterogeneity in shear-induced state. Although supervised shear-responsive modules broadly cluster into high shear-associated and low shear-associated modules, within each shear condition expression of shear-responsive modules is not monolithic (Figure 3.11). Instead, sets of TF modules are expressed by distinct groups of shear-responsive cells in each condition. This analysis demonstrates that the sets of shear-responsive modules which cluster separately in the TF set dendrogram are meaningfully expressed in different groups of cells. In the low shear-associated group, EPAS1-activated and SMAD3-activated genes are co-expressed, while TWIST1-activated genes, ETS2-activated genes, and KLF2-repressed genes define overlapping but distinct groups of cells. In the high shear-associated group, the KLF2/4-activated gene score, TCF4/CTNNB1-activated gene score, and KLF4-repressed score are co-expressed, while the MEF2A-activated score and TWIST1-repressed score define distinct groups of cells. This clustering is not a consequence of technical metrics of cell quality (Figure 3.11).

3.3.4 Rapamycin has multiple functional effects on endothelial state

To build on our results in EC populations responding to a single stimulus, we used scRNA-seq to profile EC populations cultured in a bimodal stress experiment under low or high shear (defined as previously) with or without administration of rapamycin.

First, the plastic response of a physiologically conditioned population of cells to rapamycin was characterized by identifying markers which discriminated the high shear
Figure 3.11 Heatmap demonstrating heterogeneity of expression in shear-responsive TF modules across all shear-conditioned cells. Cells are clustered by similarity. Flow annotation describes flow state. Technical variables (nUMI and nGene) are plotted in quartiles. Scaled TF module scores are plotted in heatmap. TF scores are separated using the previously defined dendrogram to demonstrate distinct expression patterns among TF modules across all cells.
population with rapamycin from the high shear population without rapamycin using a Wilcoxon rank-sum test as implemented in Seurat. Gene set enrichment was performed using clusterProfiler on significantly differentially expressed markers with a minimum 30% change in expression between the two conditions to identify the major functional components of the physiologic response to rapamycin (Figure 3.12A-B).

Multiple EC-relevant functional categories of genes exhibited reduced expression with administration of rapamycin (Figure 3.13). A set of endoplasmic reticulum-associated genes included heat shock proteins HSP90B1 and HSPA5, protein processing-associated genes PDIA3, PDIA4, and PDIA6, and collagens COL8A1 and COL1A2. A set of angiogenesis-associated genes included GJA5, CD34, EPHB4, and SOX17. A set of matrisome-associated genes included IGFBP4, HYAL2, ADAMTS18, and FBN1. Translational genes (including ribosomal transcripts) were the major functional category identified by gene set enrichment on transcripts upregulated with rapamycin expression in this model.

3.3.5 Unsupervised identification of co-expressed gene modules demonstrates reduced complexity of rapamycin response in altered shear-conditioned ECs

To leverage the ability of scRNA-seq to further dissect the components of EC plasticity in response to rapamycin, an unsupervised WGCNA-based method coupled with a significance test was employed to identify co-expressed gene modules consisting of at least 10 genes in the response of flow-conditioned ECs to rapamycin. The physiologic shear rapamycin response was defined as the set of differentially expressed genes between the high-shear cells with rapamycin and the high-shear cells without rapamycin, which consisted of 360 total genes. The altered shear rapamycin response
Figure 3.12 Gene set enrichments of rapamycin-altered genes. A) Enriched gene sets upregulated with rapamycin. B) Enriched gene sets downregulated with rapamycin. Bar width indicates negative log10 of false discovery rate (FDR) for enrichment.
Figure 3.13 Heatmap of genes altered by rapamycin in physiologically shear-cultured cells. Cells are separated by rapamycin condition (CTRL = control, RAPA = rapamycin). Genes associated with endoplasmic reticulum (ER), angiogenesis, matrisome, and translation/protein synthesis are plotted separately to demonstrate bulk functional effects of rapamycin.
was defined as the set of differentially expressed genes between the low-shear cells with and without rapamycin, consisting of 139 genes. The WGNCA-based module identification and significance testing procedure was applied separately to both sets. This process identified 8 distinct co-expressed modules of minimum size in the physiological (high) shear condition and 3 modules in the altered (low) shear condition (Table 3.3, Table 3.4, Figure 3.14A). Gene set enrichment was applied to obtain functional annotations for the modules when possible, although not all modules were enriched for known functional gene sets.

The co-expressed transcriptional modules comprising the response of physiologically shear-cultured HAECs to rapamycin included an ECM/surface adhesion molecule set including GJA4, GJA5, and HLA molecules; an angiogenic/vascular development set including TMEM100 [127] and BMPER [128]; an ER processing/antigen presentation set including B2M, CALR, and PDIA3; a peptide synthesis gene set; and a mixed set enriched for both heparin-binding genes including LTBP2 [129]. Three unannotated gene sets without known functional gene enrichments included a UBXN1/IGBP1/ATF4 module, a TUBB/TUBB6/MYO5A/FLT1 module, and a FBN1/F2R/EPHB4 module.

Co-expressed transcriptional modules comprising the simpler response of ECs cultured under altered (low) shear to rapamycin included a peptide synthesis gene set, an ER/protein folding gene set including CALR, PDIA3, and MMP14, and a larger set with no clear functional enrichment which included multiple genes with known relevance to vascular biology including GJA4, GJA5, VWF, FABP4, and PDIA4.

To leverage this method of decomposing the plastic response to rapamycin to
## Table 3.3 Unsupervised gene modules in the physiological shear rapamycin response.

<table>
<thead>
<tr>
<th>Module</th>
<th>Genes</th>
<th>Functional enrichments/ annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>phys.M1</td>
<td>VIM, HMGA1, HLA-E, EEF2, SNHG7, IGBP1, CAMLG, ATF4, CSNK2A2, UBXN1, EIF3D, SIGIRR, ANP32B, MPG</td>
<td>UBXN1/IGBP1/ATF4</td>
</tr>
<tr>
<td>phys.M2</td>
<td>SPARC, GJA5, GJA4, VIPR1, CYP1B1, PALLD, EFCA8B14, SLCO2A1, OSBPL8, TGM2, CD34, TSPAN13, AIF1L, KPNA2, TNFSF10, EMCN, PAPSS1, FLRT2, COL1A2, TSPAN18, SLC40A1, FAM198B, AQP3, STMN1, STOM, MMP15, GPR116, HLA-B, RGS3, UBE2E3, PDIA4, KIAA1161, CFH, PON2, NEDD9, PPAP2A, FSCN1, HLA-C, RAMP2, HLA-A, MMRN1, NCSTN, MARCKSL1, TFRC, LTA4H, VGLL4, LPCAT2, EFNB1, C19orf10, DAG1, ASPH, CD276, SHE, SSN1, NID1, FURIN, B4GALT5, CMIP, NGFRAP1, ZMAT3</td>
<td>ECM/Cell surface</td>
</tr>
<tr>
<td>phys.M3</td>
<td>HSP90B1, POSTN, RNASE1, HSPA5, IGBP4, DKK2, PI16, TMEM100, LAP3, FABP4, COL5A2, LGR4, LIM51, CALCRL, DDIT4, FAM213A, PRCP, KLHL13, EFNB2, CROT, PPAP2B, BMPER, PTPrB, HEG1, SLC16A1, ATP13A3, TMEM47, SELT, ELK3</td>
<td>Angiogenesis</td>
</tr>
<tr>
<td>phys.M4</td>
<td>B2M, CD9, CALR, ATP1B3, MMP14, PDIA3, CAT, CYB5B, IFITM3, PLTP, TANC2, LIPA, SNX3, AK3, SOX17</td>
<td>ER/Antigen processing</td>
</tr>
<tr>
<td>phys.M5</td>
<td>RPS21, GNB2L1, S100A6, RPS14, RPL37, RPL37A, RPS11, ZNF581, EIF1, ZFAS1, RPS18, RPS8, RPL30, FAU, RPS24, RPL35, S100A13, RPL23, RPS29, RPS15A, BR1, RPL34, RPL36, ATP5I, TOMM7, RPS25, PABPC4, EIF2A, C6orf48, NOB1, QARS, FAM211A</td>
<td>Peptide synthesis/ribosomal</td>
</tr>
<tr>
<td>phys.M6</td>
<td>GNG11, TUBB, TSC22D1, HSPA8, JAK1, ROBO4, PMP22, NAP1L1, TM4SF1, PTTG1IP, ACVRL1, CAST, EEF1A1, SQSTM1, FLT1, EMP1, TUBB6, ICAM2, MOY5A</td>
<td>TUBB/TUBB6/MYO5A</td>
</tr>
<tr>
<td>phys.M7</td>
<td>IL33, LIMCH1, ITGA9, GPC1, PTGFRN, KCTD12, DNAJB4, MGP, LAPT4M, CDH11, FBN1, IL1R1, CLEC14A, WSB1, HSD17B12, EPHB4, F2R, JAG2, PREX1, ITPR2</td>
<td>FBN1/F2R/EPHB4</td>
</tr>
<tr>
<td>phys.M8</td>
<td>PLXNA2, SERPINB1, PIEZO2, LTBP2, CLDN5, GABBR2, PGF, AHNAK2, TUBA1C, TFPI, NDRG1, TCF4, SASH1, ST3GAL1, MAP4K4, DPYSL2, BACE2, UBALD2, VEPH1, ITGA2, RRAS, PLEC, CCNI, NAV2, GNG12, KLHL5, PEAR1, RHBD1, SLC38A1, ARHGAP21</td>
<td>Mechanotransduction/heparin binding</td>
</tr>
</tbody>
</table>
### Table 3.4 Unsupervised gene modules in the altered shear rapamycin response.

<table>
<thead>
<tr>
<th>Module</th>
<th>Genes</th>
<th>Functional enrichments/annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt.M1</td>
<td>HSP90B1, POSTN, SPARC, HSPA5, CALR, COL5A2, PDIA6, P4HB, RPN2, CD93, MMP14, ATP1B3, UCHL1, ITGA5, PDIA3, PPIB, FSTL1, COLGALT1, CYTL1, TFPI2</td>
<td>ER/protein folding</td>
</tr>
<tr>
<td>alt.M2</td>
<td>GJA4, GJA5, FABP4, MYH10, EFEMP1, RGS3, TGM2, SERPINH1, CLDN11, VIPR1, WSB1, COL8A1, PROCR, PDIA4, CFH, STAT1, RNASE1, C10orf10, Ddit4, HYOU1, LIMCH1, VWF, FLRT2, ABLIM1, LAP3, PTGFRN, CDH11, AKR1C3, MAGED1, SAT1, PRDX4, KLHDC8B</td>
<td>GJAs/VWF/FABP4</td>
</tr>
<tr>
<td>alt.M3</td>
<td>S100A6, RPL31, RPL37, RPL23, RPS14, RPL35A, RPL32, RPL37A, RPL12, RPS21, RPL14, TMSB10, RPS9, RPL34, RPS24, RPS29, ATP5E, FAU, RPL38, RPS15A, RPL26, RPS27A, COX7C, RPL36, USMG5, TOMM7, RPS25, ZFAS1, ATP5I</td>
<td>Peptide synthesis/ribosomal</td>
</tr>
</tbody>
</table>
Figure 3.14 Gene modules in the physiological shear- and altered shear-cultured response to rapamycin. A) Dendrograms and significant gene modules for physiological shear and altered shear rapamycin response. Significant gene modules are labeled. B) MHC adjusted p-values for significance of overlap between modules.
understand the effect of shear on rapamycin response, we identified the modules from each shear condition which had statistically significant overlap (Figure 3.14B). The ribosomal gene sets present in each rapamycin response had strong correspondence. A significant overlap was also present between the ER/protein processing gene sets of the two responses; both modules included the common genes CALR, ATP1B3, MMP14, and PDIA3. Notably, the large unannotated module present in the altered shear set had nonsignificant overlaps with multiple modules in the physiological shear rapamycin response; it shared FABP4, RNASE1, DDIT4, and LAP3 with the angiogenesis set; GJA4, GJA5, RGS3, TGM2, VIPR1, PDIA4, CFH, and FLRT2 with the ECM/cell surface set; and WSB1, LIMCH1, PTGFRN, and CDH11 with the unannotated FBN1/F2R/EPHB4 set. A weaker overlap was also present between the ER/protein folding module in the altered shear plasticity and the angiogenesis module in the physiologic shear plasticity, consisting of the heat shock proteins HSP90B1 and HSPA5 as well as COLA2 and POSTN.

3.3.6 Distinct subgroups within shear-conditioned EC populations demonstrate heterogeneity in rapamycin-induced plasticity

Next, we used the scRNA-seq profiles of EC populations under bimodal stimulus conditions to identify heterogeneous subgroups of shear-conditioned ECs with distinct responses to rapamycin. Canonical correlation analysis of ~7,300 cells from four conditions enables the alignment of scRNA-seq profiles to identify corresponding subpopulations with and without stimulation [76]. By aligning scRNA-seq profiles with and without rapamycin, we project these transcriptional profiles into a space in which structure in the data is rapamycin-independent (Figure 3.15A). Computational clustering
in this space identified multiple subpopulations (Figure 3.15B) which were annotated on the basis of gene set enrichment on subpopulation-specific markers (Figure 3.16).

Based on the experimental condition of cells in each cluster (Figure 3.15C), the two largest clusters were identified as the primary high-shear and primary low-shear populations. The remaining smaller populations had functional signatures identified by direct examination and gene set enrichment of subpopulation-specific markers. One subpopulation was enriched in junctional/matrisomal genes including ESM1, APLN, and MT2A, and also expressed cytoskeletal/mesenchymal markers including TAGLN, MYL9, and MYL12B. A mitotic population expressed cell cycling markers including TOP2A, CENPF, and CENPE. Two populations had increased expression of either mitochondrial genes or ribosomal genes. The mitochondrial population also expressed genes associated with RNA splicing, including PNISR.

To determine the effect that the heterogeneous presence of two large primary shear-conditioned populations and multiple smaller transcriptionally focused subpopulations had on rapamycin response, the alterations with rapamycin in each of the previously identified modules comprising rapamycin plasticity were quantified in each subpopulation (Figure 3.17). The mean log-fold change with rapamycin across the genes in each module was computed for each subpopulation. A mean log-fold change corresponding to a 30% alteration was used as a threshold. Of all the subpopulations present in shear-conditioned ECs, only the primary high (physiologic) shear population of cells demonstrated all of the components of rapamycin plasticity identified under physiologic shear; both the primary low shear populations and the other transcriptionally focused subpopulations showed a diminished response along these axes. Conversely,
Figure 3.15 Transcriptionally defined subpopulations in drug-aligned shear-cultured populations. A) t-SNE representation of bimodal scRNA-seq profiles CCA-aligned by drug status. B) Subpopulations identified by clustering and functional annotations. C) Cluster composition by original experimental condition.
Figure 3.16 Heatmap of subpopulation markers for CCA-identified subpopulations in aligned shear-cultured populations. Top markers for each subpopulation are plotted with scaled expression score and cluster functional annotations.
Figure 3.17 Quantification of mean changes in rapamycin response modules by subpopulation. The mean log fold-change in each module between rapamycin and control conditions is plotted by subpopulation for both the physiological shear (top) and altered shear (bottom) response to rapamycin. An arbitrary fold-change threshold corresponding to 30% alteration is plotted using the dashed lines. Functional annotations/representative genes are included for each module. Solid black lines indicate modules between physiologic shear and altered shear responses that demonstrated significant overlap in component genes.
all EC subpopulations except the mitotic population demonstrated plasticity in the modules identified in the altered shear rapamycin response, including both the primary high and low shear populations. The plastic response of physiologically shear-conditioned ECs to rapamycin is therefore more complex and unique to that population, while the less complex response of altered shear-conditioned ECs can be recapitulated by ECs in multiple transcriptionally defined states.
3.4 Discussion

3.4.1 Summary

In this chapter, we present the use of scRNA-seq to define the plasticity of ECs in a controllable flow-cultured bioreactor system in response to both a unimodal mechanical stimulus and a bimodal combination of mechanical and chemical stimulation. Application of single-cell transcriptomics enabled precise definition of the modular components of each plastic response, both in a supervised fashion using existing knowledge on shear-responsive transcription factors and in an unsupervised fashion using WGCNA. Our analysis demonstrated that while known shear-responsive transcriptional programs do differentiate high and low shear-conditioned EC states (Figure 3.10), the response to shear is not monolithic. The plastic response of an EC population to shear is heterogeneous, with different groups of ECs expressing distinct shear-responsive modules (Figure 3.11).

With the addition of a second chemical stimulus, we sought to determine the extent to which endothelial plasticity in multiple axes induced heterogeneity in the stimulus response. Our results elucidate the effect of rapamycin on the endothelial transcriptome at the population level (Figure 3.12) and partition this effect into discrete genetic modules in both physiologic shear and altered shear conditions, showing that the presence of a mechanical stressor (low shear) diminishes and reduces the complexity of rapamycin-induced plasticity (Figure 3.14). Single-cell profiling also enabled resolution of the heterogeneity in EC plasticity; when considering the response of shear-conditioned populations to rapamycin, multiple smaller populations with distinct transcriptomes signaling mitosis or cellular stress emerged (Figure 3.15). These smaller
populations recapitulated the simplified rapamycin plasticity of the main group of low
shear-conditioned cells, while only the main group of high shear-conditioned cells
demonstrated the full and complex response to rapamycin (Figure 3.17). These findings
elucidate the contribution of multiple stimuli to functional endothelial heterogeneity as
observed in the previous chapter in vivo and to the convolution of endothelial
heterogeneity with stimulus-induced plasticity as described in the introduction.

3.4.2 Definition of flow-induced endothelial state

The effect of different shear stresses on endothelial state has been a subject of
extensive investigation in vascular biology. This dataset enables an integrated view of
multiple discrete gene expression programs based on known shear-responsive
transcription factors in order to more fully understand the components of shear-induced
EC state and their heterogeneity within a population of ECs exposed to differential
shear stresses. From a bulk or population perspective, many of the markers of the high-
shear population in this dataset have been previously associated with high shear stress.
The low shear population in this dataset is differentiated on average by metabolic
dysfunction in the form of upregulation of both ribosomal and mitochondrial genes.

Low shear stress has previously been linked to endoplasmic reticulum stress and
induction of the unfolded protein response (UPR) [131]. The UPR is a final common
pathway of cellular dysfunction which arises when the balance between protein
synthesis and ER handling of protein folding is disrupted, resulting in the accumulation
of unfolded proteins within the cell. The UPR involves activation of multiple cellular
mechanisms to restore protein homeostasis. The PERK pathway involves activation of
eIF-2α, which results in global reduction in translation of proteins, while the
ATF6/XBP1/IRE1α pathway activates expression of ER protein folding chaperones and the ER-associated protein degradation system to eliminate misfolded proteins through ubiquitination and proteasome activity [132], [133]. The UPR pathway has been previously implicated as a primary driver of EC dysfunction, interacting with inflammatory pathways and handling of reactive oxygen species. UPR activation has been linked to endothelial dysfunction in response to multiple pathogenic stimuli including altered shear stress, hyperlipidemia, and hyperglycemia, as well as in the development of hypertension and pulmonary arterial hypertension [132]. Studies of athero-susceptible vs. atheroprotected vascular beds in vivo demonstrated an increase in expression of UPR components including ATF4, HSPA5 and XBP1 in atheroprone vascular beds, especially in the kidney, suggesting that this pathway is an important contributor to pathologically relevant EC heterogeneity [133]. UPR activation is generally associated with translational repression to restore protein balance, so our observation of increased ribosomal mRNA transcripts in low shear cells is at first surprising. However, a distinct feedback arm exists within the UPR mediated by ATF4 and GADD34 which has the effect of restoring protein synthesis; this pathway is linked to ribosomal biogenesis via NMP4 [134] and is thought to be particularly relevant in the setting of long-term ER stress [135], which may be the case in our multi-day shear-conditioning system. Intriguingly, this ATF4-mediated upregulation of protein synthesis interacts with mTORC1 [135], which may explain the relationship between rapamycin plasticity and shear-induced EC state that we observe here. The observed increase in mitochondrial biogenesis with low shear stress may be partially explained by a previous suggestion that mitochondrial biogenesis is protective against ER stress [136].
Using our scRNA-seq transcriptional profiles, we were also able to move beyond bulk responses to query the coupling of known shear-responsive transcription factor programs in defining shear-induced EC state. Many shear-induced TF modules demonstrated strong correlation in these shear-cultured EC populations and were arranged into a hierarchy of co-expression which suggests potential functional links. The extremely tight relationship between EPAS1 (also known as HIF-2α) and SMAD3-activated modules highlights the recently established role of both these factors in endothelial-to-mesenchymal transition (EndMT), another “final common pathway” of EC dysfunction [89], [126]. This is also supported by their link to TWIST1-activation [115]. This coupling suggests that EndMT is indeed a distinct state transition which can be identified in a larger population using scRNA-seq.

On the side of high shear-related programs, the link between activation of KLF2 and activation KLF4 as vasoprotective mechanisms is well-known [90], and their targets overlap such that they merited being combined into one module in this analysis. However, the activity of their known repression targets suggests that they may have distinct roles in defining shear-induced state. KLF2 repressed targets clustered as expected with the low shear-defining gene sets, while the putative targets of KLF4 repression actually clustered with the high shear-defining sets. Whether this is due to incomplete knowledge of KLF4’s targets or due to more complexity in KLF4’s function in the shear stress response remains an open question. Such coupling does provide an intriguing potential balance in TF biology wherein activation and repression of related factors provides a complex combination space of responses using only two TFs.
In this same vein, the role of TCF4 and CTNNB1 remain to be elucidated, as they cluster with the primarily high shear-defining TF programs; of particular note is the relationship to the KLF2-KLF4 program. This method also allowed for determination of the most important transcriptional programs for differentiating shear-induced states via contribution to linear discriminant analysis. This analysis suggests that NFE2L2 and KLF2/4 are most important for defining a high shear state while EPAS1 and TWIST1 are most important for defining a low shear state, potentially highlighting the functional importance of redox management and EndMT in this phenotypic axis.

In general, application of scRNA-seq to plastic cell populations including ECs in this manner is useful for reducing the states that arise in response to a known external condition into correlated transcriptional programs, and then determining the strength and structure of relationships between those functional basis vectors. Future work could extend this type of analysis to other endothelium-modulating environmental conditions like hyperglycemia or hyperlipidemia to determine if the network of TF set coupling shifts meaningfully with differential stimulus conditions.

### 3.4.3 Heterogeneity in the components of shear-induced endothelial state

A key finding from these results was that the components of shear state as defined by supervised transcriptional modules are not all expressed in the same cells, but instead in distinct subgroups within the broader endothelial population. This suggests that what we think of as the “shear response” is not monolithic, but instead consists of a number of correlated transcriptional programs which can be engaged singly or together. A progressive model of EC state development arises then, in which a group of ECs exposed to shear alteration shift through a number of successive shear-
induced states each defined by core transcriptional modules. Extending this idea lends credence to the model of an overriding endothelial state (for example, ‘quiescent’ vs. ‘activated’) which is made up of superimposed heterogeneous individual ECs, which occupy distinct transcriptional module-defined states in a spectrum of response. These distinct states are conceptually what we referred to as “stress-dependent heterogeneity” in the paradigm presented in the introduction. Further analysis using pseudotime or diffusion-based techniques [137] might elucidate this type of spectrum of shear responsiveness. In addition, the hierarchical nature of these TF sets might be illuminated by the use of probabilistic modeling that incorporates known information about gene regulatory networks to suggest which transcription factor sets are the furthest “upstream”, especially in the case of multiple TF sets linked to a single function (e.g., SMAD3, EPAS1, and TWIST1).

3.4.4 Rapamycin-induced endothelial plasticity

Our results from a bimodal experiment involving both shear stress conditioning and administration of rapamycin demonstrate multiple functional effects of rapamycin on physiologically conditioned ECs, including decrease in expression of transcripts implicated in ER handling, matrisome activity, and angiogenesis. The increase in ribosomal transcripts observed with rapamycin administration coupled with the observed decrease in ER-relevant transcripts (including HSPA5) is interesting given that in non-ECs rapamycin and its derivative temsirolimus have been recently found to both activate the UPR by inducing splicing of XBP1 and interact directly with ribosomes [138].
To partition the plastic response to rapamycin, which is less well understood than the response to shear stress, we employed an unsupervised method of identifying co-expressed gene modules based on WGCNA. Many of the modules identified in the response of physiologic (high) shear conditioned cells corresponded to the major functional sets identified in the bulk response, while other, smaller modules had no clearly enriched function but nonetheless included genes of known importance in EC biology. The increased complexity of the response of physiologic shear-cultured ECs to rapamycin relative to that of altered shear-cultured ECs demonstrates that stress-induced plasticity in one (e.g. a mechanical) domain can limit potential plasticity in another (e.g. chemical) domain. Such thinking has profound implications for the use of pharmacologic modulation in the vasculature, especially in the setting of drug-eluting stents, which not only injure the endothelium and underlying media but also impose profound alterations in shear stress on local endothelial cells [139]. The specific nature of the identified modules, particularly those without clear functional enrichments, is a valuable area of future investigation into the effect of rapamycin on ECs. The specific overlapping modules between the two rapamycin responses – ER processing and ribosomal genes – again suggest the potential importance of the UPR and ER stress in driving this phenomenon. The ER module in physiologic shear cells notably includes homeostatic functions such as antigen processing which are lost in the altered shear module, possibly due to the engagement of ER machinery by shear-induced UPR. This strategy for identification of transcriptional components of plasticity in an unsupervised manner will be most useful in future investigations of EC response to stimuli whose effects are insufficiently well-known for the generation of supervised modules. Such
stimuli may include other chemical/pharmacologic interventions as well as complex stimulation in pathologic environments where \textit{a priori} identification of modules of interest is difficult.

### 3.4.5 Heterogeneity in shear-induced EC populations and effect on plasticity to rapamycin

The CCA-based method presented here enables identification of corresponding subpopulations between the control and rapamycin-exposed conditions to identify groups with potentially distinct responses to rapamycin stimulation. While two major ‘pure’ populations of shear-induced cells are identified, the smaller subgroups with mixed memberships display distinct transcriptional phenotypes that suggest stress-induced EC states. The mitotic population is discrete owing to the profound extent to which mitosis commandeers the transcriptional output. The high ribosomal and high mitochondrial populations both have greater membership from the low shear-exposed cells, and may represent the effect of the “forcing” final common pathway of the UPR dominating the transcriptome in distinct ways. The high mitochondrial population may also be a cell quality-derived clustering, although these cells do have roughly equivalent library complexity to the other populations. Of greatest interest is the “junctional” population, enriched for expression of adhesion molecules, junctional proteins, and surface molecules. Gene set enrichment also identifies increased expression of cytoskeletal genes and TAGLN, a known mesenchymal marker that in ECs has been cited as a marker of EndMT [140]. This set of markers offers the tantalizing possibility that this population represents an early-phase EndMT population, as an increase in mesenchymal genes is the hallmark of EndMT, which has previously been associated
with shear disturbance [89]. If EndMT is observed as a separate final common pathway stress state accessed by cells in this experiment, it is thus reasonable that it be observed with a greater proportion in the low shear condition.

The finding that of all subpopulations, only the ‘pure’ high shear-responsive population demonstrates full plasticity in all the modules of the physiologic shear rapamycin response suggests that occupation of transcriptional stress states or ‘forcing’ states like mitosis prevents full plasticity in response to a chemical stimulus. This conclusion is augmented by the finding that almost all subpopulations, regardless of stress type, show a near-uniform response to rapamycin along the axes defined in the altered shear condition. This finding is true regardless of the origin of cells within the subpopulation, i.e. the ‘pure’ high-shear cell group recapitulates the altered shear response but not vice-versa. The fact that physiological shear-cultured ECs recapitulate the altered shear-cultured response but not vice versa suggests that ECs under physiologic conditions (our equivalent to ‘quiescence’) are capable of mounting the most complex and finely-tuned response to a chemical stimulus, while cells which are otherwise stressed and already committed to a specific transcriptional response demonstrate reduced plasticity to an additional stimulus. Additionally, low shear alone is sufficient to push ECs into a stress-related transcriptionally committed state, regardless of the presence or absence of a functional stress response like ribosomal or mitochondrial upregulation. This finding has strong implications for the pharmacologic treatment of the vasculature – if prevailing disease conditions including hypertension, hyperlipidemia, or altered flow are sufficient to force EC state and limit EC plasticity in response to chemical stimulus, then the efficacy of drugging ECs may diminish rapidly.
as disease develops and vasculature is forced into a stress state. The dominance of shear-stress induced state over chemically-induced plasticity is a demonstration of the primary nature of ECs as mechanosensitive cells.

Elucidation of this spectrum of EC responsiveness may be possible using pseudotime/trajectory-based analysis [137] to highlight the trajectory from stress-limited “underdeveloped” cells in the low shear population or stress subpopulations to the fully plastic high shear cells. Additionally, extension of this method to additional DES or vasoactive drugs like paclitaxel will demonstrate whether this phenomenon is common to all methods of chemical stimulation or whether it is a unique consequence of rapamycin and its interaction with the UPR stress response.
4. Conclusions and Future Research Directions

4.1 Overview and challenges

Recent studies on endothelial heterogeneity have focused not only on the differences between endothelial cells (ECs) in diverse vascular beds, but also on local heterogeneity within a given EC population [16], [17]. The extent to which discrete EC states exist in a local population, the functional significance of local heterogeneity, and the impact of this variability on endothelial plasticity are the chief questions underlying this important avenue in vascular biology. In this dissertation, we leverage the recent boom in the development of single-cell RNA-sequencing (scRNA-seq) methods [27] to address these questions in EC biology.

In Chapter 2, we used scRNA-seq to profile the mouse aorta under physiologic conditions to demonstrate the presence of a baseline heterogeneity in the aortic EC population. We identified discrete EC subpopulations in these context and used a functional lens to characterize these populations as distinctly underlying adhesion and lipid handling. We also concluded that the same populations persist in the presence of a pathologic stressor, and that they exhibit conserved responses to a pathological stressor that underlie the emergent plasticity of the endothelium.

In Chapter 3, we used scRNA-seq to profile EC populations cultured in an in vitro flow-controlled bioreactor system to study the heterogeneity inherent in the plastic response of ECs to a unimodal mechanical stimulus and to the bimodal combination of mechanical and chemical stimuli. We demonstrated that using scRNA-seq to characterize EC plasticity in this fashion reveals heterogeneity in the expression of known shear-responsive transcriptional modules in an EC population and allows for
organization and comparison of these modules. Furthermore, we demonstrated that the complexity of the EC response to a molecular modifier, here rapamycin, is dependent on previous stress exposure, here a shear-induced state. Our data suggest that stress stimuli commit EC into ‘forced’ transcriptional states, altering the response to subsequent stimuli. Shear stress takes a dominant role in governing EC state, consistent with the primary nature of ECs as flow-sensitive cells.

Viewing EC state as the result of superimposed responses to microenvironmental stimuli highlights the natural difference between programmed stochastic and random heterogeneity. A possible analogy is to snowflakes, which form via a set of physical rules that govern crystallization under specific conditions. If conditions are identical, then the results should be also. However, because the forces to which any given snowflake is exposed subtly differ in form, time, space, and sequence, no two snowflakes are identical. Similarly, ECs behave in well-defined ways in response to environmental stimuli, but in the complex vascular milieu no two ECs experience the same conditions, resulting in population heterogeneity. Our results suggest, however, that certain stimuli can dominate and result in the formation of stable EC states with specific functions occupied by local subpopulations.

The range of forces that EC can respond to is profound, which increases the complexity space for EC state. Unstressed cells under physiologic conditions are the most pluripotent in terms of stimulus response. Once a state-altering force is imposed, ECs are committed to specific transcriptional programs (e.g. mitosis, high ribosomal gene expression, and high mitochondrial gene expression) and demonstrate reduced plasticity. This is also what we observed in the response to rapamycin by altered shear-
cultured ECs. Plasticity to one abnormal stimulus reduces the potential plasticity of ECs in response to an additional stimulus.

The work presented here is an important step in understanding the nature of local EC heterogeneity and its consequences for EC function and plasticity, but challenges remain to be addressed in this space. Although we have demonstrated the presence of heterogeneous EC subpopulations in vivo using computational analysis of transcriptomes, confirmation that these represent stable states in situ will require simultaneous quantification of the spatial distribution of multiple identified subpopulation markers in parallel and demonstration that profiles corresponding to our populations show differential distributions. We demonstrate that initial conditioning with a mechanical stimulus diminishes plasticity to a chemical stimulus, but cementing this conceptual model will require evaluation of multiple different stimulus types and of whether the order of stimulus presentation affects this phenomenon. This will require a more complex experimental setup incorporating multiple stimulus types and allowing for presentation of stimuli in different orders. These advances will require harnessing new modalities for in situ transcriptomics and more complex bioreactor systems to further advance our understanding of these phenomena in vascular biology.

4.2 Functional endothelial heterogeneity and future steps

Future work to build on the functional EC heterogeneity established here under homeostatic conditions in vivo and in the stimulus response in vitro should focus on confirming the presence of distinct EC states in various contexts and understanding the
drivers of state transitions. In vivo, the canonical correlation analysis (CCA)-based method used here can be utilized in any disease-simulating condition to identify corresponding subpopulations between homeostatic and pathological contexts. Applying this method across a range of vasoactive stimuli including hyperglycemia, vascular injury, or inflammatory mediators will demonstrate how membership in these subpopulations can change, whether new pathogenesis-specific subpopulations arise, and which elements of endothelial plasticity are subpopulation-specific. Changes in subpopulation membership would be a particularly interesting finding suggesting the possibility of state transitions, providing further evidence that these states are due to different microenvironments applied to a common cell type rather than distinct developmental origins. The EC subpopulation findings reported here can also be translated to ECs in other contexts; for example, circulating ECs (CECs) detached from the vascular bed, which have previously been used as a biomarker for vascular damage [141]–[143]. Major questions in these space include how the transcriptome of CECs corresponds to that of attached ECs and whether CECs correspond to specific subpopulations of vascular ECs as identified by our analysis. Building on our results in this way would provide insight into the potential use of CECs as biomarkers for surveilling the vascular state.

Our results on the heterogeneity of shear-induced EC state in vitro also provide the basis for potential future investigations. Multiplexed fluorescence techniques like MERFISH [36], [37] can define the spatial correlates of the distinct but correlated transcriptional programs identified by our analysis. The methods used here to understand the structure of the EC response to flow can be applied to other stimulus
classes including metabolic and pharmacologic alterations to investigate how the network of transcription factors underlying EC state shifts in response to differential stimuli. Applying this approach across a range of stimulus classes will reveal previously unappreciated links between different transcription factors that are key drivers of EC biology. New studies integrating information from both halves of this work can investigate superpositions of stimuli \textit{in vitro} which may induce a phenotype similar to one of the subpopulations we identified \textit{in vivo}; this would illuminate the potential origins of basal EC heterogeneity from a first-principles approach and provide insight into what may drive state transitions.

\textbf{4.3 Single-cell resolution EC plasticity and future steps}

This dissertation also provides a blueprint for the future use of single-cell resolution techniques to further investigate EC plasticity. Our work with rapamycin and shear stress alteration is most directly applicable in the context of vascular interventions and manipulations like drug-eluting stents (DES), an area with immediate translational impact. Stents have previously been studied \textit{in situ} in rabbits [144], and similar animal models could be used and coupled with spatial transcriptomics to show that in this context the ECs adjacent to flow-disturbed regions of the vessel wall show altered plasticity relative to normal flow-exposed ECs. The idea that stimulus conditioned ECs are actually less amenable to pharmacologic manipulation supports a series of potential investigations on the magnitude of this phenomenon for different vasoactive drugs used clinically. Testing new vascular therapeutics with this idea in mind suggests that future
trials should study drug effects not only in pure EC populations, but also in ECs cultured with shear alteration and other in vivo stimulus-mimicking methods to reproduce the state-defining forces on ECs in the body.

Our results also suggest that ECs in various ‘final common pathways’ of EC dysfunction – e.g., the unfolded protein response and endothelial-to-mesenchymal transition – have limited plasticity to additional stimuli. These pathways are of great interest currently as contributors to a wide range of vascular pathologies. Future work may focus on restoration of plasticity as a key indicator of EC rescue from stress responses. If loss of plasticity is a potential hallmark of endothelial dysfunction, then assaying EC plasticity may be a useful method for early detection of vascular disease onset.

4.4 Long-term research goals

Ultimately my chief goal as a translational vascular biologist is to apply the insights we glean here from single-cell transcriptional characterization of EC heterogeneity and plasticity to develop new methods for diagnosing, characterizing, and treating human vascular disease. Understanding the heterogeneity of the native vasculature and how ECs in a population respond variably to stimuli enables the development of single cell-resolution biomarkers for vascular disease, whether drawn from vascular tissue samples or from circulating ECs in the blood. In both cases, identifying the EC phenotype of a limited clinical sample is only useful when paired with a knowledge of the potential EC states from which the sample is drawn. Similarly,
determining how EC heterogeneity impacts our attempts to pharmacologically manipulate the vasculature will inform the process we use to develop and test new vascular therapies. By leveraging these insights, I will continue to pursue research avenues that move from the basic principles of vascular biology illuminated in this dissertation using scRNA-seq to clinically actionable applications in vascular medicine.
Appendix A: Ibidi Protocols

**Seeding cells in Ibidis**

1. Trypsinize cultured cells and resuspend to final concentration in media of 0.8-1.2 x 10^6 cells/mL.

2. Transfer 200 µL cell suspension into Ibidi channel.

3. Incubate at 37° C in sterile incubator for 30 minutes – 1 hour.

4. Fill Ibidi reservoirs with 50 µL media each.

5. Incubate at 37° C in sterile incubator for 12 – 24 hours.

6. Confirm cell attachment via microscopy.

**Connecting Ibidis to flow**

1. Assemble flow loop supplies consisting of 1 glass media reservoir with cap, 3 25-cm segments of Silastic tubing, and 2 tubing connectors per flow loop.

2. Wash flow loop supplies in 0.2% SDS in distilled water in a tub on plate shaker for a minimum of 10 minutes.

3. Repeat Step 2 for washes in milli-Q water and 70% ethanol in distilled water.

4. Allow flow loop supplies to dry as completely as possible.
5. Assemble pump tubing segments consisting of 2 segments of 25-cm Silastic tubing attached to pump tubing using tubing connectors.

6. Individually package 1 media reservoir with cap, 1 pump tubing segment and 1 additional segment of Silastic tubing into an autoclave pouch. Autoclave pouches on Dry setting for 20 minutes sterilize time/20 minutes dry time.

7. Move freestanding peristaltic pump to biological hood and assemble stands for incubator rack.

8. Clean incubator rack with ethanol and place in biological hood on stands.

9. Arrange reservoir and flow loop tubing on incubator rack. Connect inferior port of media reservoir to intake of pump tubing and free Silastic tubing segment to lateral port of media reservoir. Connect pump holder to pump tubing and attach to peristaltic pump.

10. Remove seeded Ibidi from incubator and aspirate media from channel. Connect Ibidi to free ends of Silastic tubing.

11. Add media to media reservoir and run peristaltic pump at intermediate speed (e.g. 60 RPM) to circulate media throughout flow loop. If experiment involves later addition of stimuli to media, pre-measure media added to loop.

12. Ensure no bubbles persist in Ibidi channel or reservoir inlet/outlet before securing flow loop to incubator rack. Complete assembly and filling of one entire flow loop before assembling additional loops on incubator rack.

13. Return incubator rack to incubator and initiate flow.
Addition of rapamycin to flow loops

1. Assemble incubator rack stands in biological hood.

2. Pause flow and disconnect pump holders from peristaltic pump in incubator.

3. Move incubator rack to stands in biological hood.

4. Add rapamycin directly to media reservoirs in experimental condition loops to achieve desired concentration. Add equivalent amount of fresh media to control loops.

5. Return incubator rack to incubator, re-attach pump holder tubing and briefly run pump at intermediate flow rate (e.g. 60 RPM) to circulate media.

6. Restart flow at experimentally appropriate rate.

Isolation of cells from Ibidis

1. Disconnect Ibidis from flow tubing and allow media from flow loop to drain into a 50 mL tube.

2. Aspirate media from Ibidi reservoirs.

3. Using two 1000 µL pipets, flush 200 µL PBS into channel while simultaneously removing channel volume at distal reservoir to wash channel. Repeat this step for 2-3 total PBS washes.

4. Aspirate Ibidi channel minimizing residual fluid in channel.
5. Add 200 µL trypsin to channel and place in incubator for 3 minutes (low shear condition) to 8 minutes (high shear condition), or until cells are rounded and mobile when viewed under microscope.

6. Agitate Ibidi chips and shake parallel to flow channel to dislodge attached cells.

7. Add 100 uL media to channel to inactivate trypsin.

8. Using two 1000 µL pipets, perform two 200 µL washes with serum-containing media as in Step 3. Collect output in a 1.7 mL Eppendorf tube. Retain samples on ice during collection.

9. Visualize cells remaining in channel using microscope.

10. Using two 1000 µL pipets, perform 2 washes of channel using 400 µL media to collect remaining cells. Pipet media back and forth in channel between 1000 µL pipets to maximize number of collected cells.

11. Spin down cells at 300 g for 5 minutes at 4º C.

12. Resuspend cells in desired volume of media and count cells using a manual hemocytometer (e.g. inCyto C-Chip DHC-N01) for downstream analysis.
Appendix B: Computational Resources

All analyses of scRNA-seq data were performed in R (Versions 3.3.0-3.4.1). The following additional libraries were utilized:

**Chapter 2:**
- plyr (Version 1.8.4)
- dplyr (Version 0.7.4)
- Seurat (Version 2.3.1)
- ggthemr (Version 1.1.0)
- ggtern (Version 2.2.1)
- biomaRt (Version 2.34.2)
- clusterProfiler (Version 3.6.0)
- org.Mm.eg.db (Version 3.4.1)
- ReactomePA (Version 1.22.0)

**Chapter 3:**
- plyr (Version 1.8.4)
- dplyr (Version 0.7.4)
- Seurat (Version 2.3.1)
- Matrix (Version 1.2-14)
- Matrix.utils (Version 0.9.6)
- gridExtra (Version 2.3)
ggrepel (Version 0.7.0)
reshape2 (Version 1.4.2)
gplots (Version 3.0.1)
GGally (Version 1.4.0)
writexl (Version 1.1)
readxl (Version 1.0.0)
pheatmap (Version 1.0.12)
dendsort (Version 0.3.3)
MASS (Version 7.3-51.1)
RColorBrewer (Version 1.1-2)
ramify (Version 0.3.3)
WGCNA (Version 1.67)
flashClust (Version 1.01-2)
clusterProfiler (Version 3.6.0)
bionaRt (Version 2.34.2)
org.Hs.eg.db (Version 3.5.0)
GSEABase (Version 1.40.1)

Functions for implementation of WGCNA in scRNA-seq data were obtained from collaborators in the Shalek Lab.
R helper functions for analysis of scRNA-seq data

Helper functions for clustering and pathway enrichment analysis of Seurat objects:

#Functions to aid in clustering and enrichment of Seurat objects
#Aditya S. Kalluri

#Required libraries
library(clusterProfiler)
library(Seurat)
library(biomaRt)
library(writexl)
library(readxl)
library(org.Hs.eg.db)
library(GSEABase)
library(pheatmap)
library(RColorBrewer)

#Function to get ReactomePA enrichPathway object from subsetted marker matrix
markerEnrichment = function(markerSubset, num, org="org.Hs.eg.db", organism="human", pvalueCutoff=0.05){
  genes = markerSubset$gene
  entrez = bitr(genes,fromType="SYMBOL",toType="ENTREZID",OrgDb=org)
  entrez = entrez$ENTREZID
  return(enrichPathway(entrez[1:num], organism=organism, pvalueCutoff=pvalueCutoff, readable=T))
}

#Function to convert mouse genes to human genes using biomaRt
cconvertMouseGeneList <- function(x){
  human = useMart("ensembl", dataset = "hsapiens_gene_ensembl")
  mouse = useMart("ensembl", dataset = "mmusculus_gene_ensembl")

  genesV2 = getLDS(attributes = c("mgi_symbol"), filters = "mgi_symbol", values = x , mart = mouse, attributesL = c("hgnc_symbol"), martL = human, uniqueRows=T)
  humanx <- unique(genesV2[, 2])

  # Print the first 6 genes found to the screen
  print(head(humanx))
  return(humanx)
}

#Helper reorder function from tutorial:
reorder_cormat = function(cormat){
  # Use correlation between variables as distance
  dd <- as.dist((1 - cormat)/2)
  hc <- hclust(dd)
  cormat <- cormat[rev(hc$order), rev(hc$order)]  # Reverse order enables
  makeGeneCorrHeatmap to read top--->bottom
}

# Function to create a correlation heatmap for a given set of genes in
# a Seurat object
makeGeneCorrHeatmap = function(seurat, genes, lowlim=-0.5,
  highlim=0.5, scaled=T){
  if(scaled){
    vardata = as.data.frame(as.matrix(seurat@scale.data[genes,]))
  } else{
    vardata = as.data.frame(as.matrix(seurat@data[genes,]))
  }
  vardata_corr = cor(t(vardata))
  vardata_corr = reorder_cormat(vardata_corr)
  vardata_corr = melt(vardata_corr, na.rm=TRUE)
  heatmap = ggplot(vardata_corr, aes(Var2, Var1, fill = value)) +
    geom_tile(color = "white") +
    scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                        midpoint = 0, limit = c(lowlim,highlim),
                        space = "Lab",
                        name="Pearson\nCorrelation") +
    theme_minimal() +  # minimal theme
    theme(axis.text.x = element_text(angle=90,hjust=1,size=4),
          axis.text.y = element_text(size=4)) +
    coord_fixed()
  return(heatmap)
}

# Functions to perform gene clustering and enrichment analysis from
# Seurat object and list of genes
# Function to write enrichment results to file
writeEnrichResult = function(cluster_assigns, enrichResult_list,
  filename){
  # cluster_assigns is a named vector of cluster assignments
  (names=genes, values=clusters)
  # enrichResult_list is a list of enrichResults from clusterProfiler,
  but some entries may be NULL if no genes can map
  # filename should end in .xlsx
  output = lapply(enrichResult_list, "slot", "result")
  clustersDF = data.frame("gene"=names(cluster_assigns),
                          "cluster"=cluster_assigns)
  output[ ["Assignments"] ] = clustersDF
  write_xlsx(output, filename)
# Gene correlation --> clusters function
geneCorrClust = function(seurat, genes, cutFactor, fileseed){
    # cutFactor -- height factor at which to cut tree, max height divided
    # by cutFactor = cut level
    # fileseed -- start of filename for saving files generated in this
    # function call
    hr = hclust(as.dist(1-cor(t(seurat@scale.data[genes,]),
                         method="pearson")), method="complete") # Clustering
    genes_reorder = genes[hr$order] # Reorder list of genes according to
                                    # clustering
    mycl = cutree(hr, h= max(hr$height/cutFactor))[hr$order] # Cut and
                                                            # reorder in same order as heatmap

    # Put clusters into a list
    clust_c = vector("list", length=max(mycl)) # Declare an empty list of
    # the right size
    for(i in 1:max(mycl)){
        clust_c[[i]] = data.frame("genes"=names(mycl)[mycl==i])
        colnames(clust_c[[i]]) = paste0("clust", i)
    }
    clustersDF = data.frame("gene"=names(mycl), "cluster"=mycl)
    output = clust_c
    output[["Assignments"]]["clusters.xlsx"] = clustersDF # Write output file
    filename = paste0(fileseed,
                       ", clusters.xlsx")
    write_xlsx(output, filename)
    return(list("Assignments" = mycl, "Clusters" = clust_c))
}

# Make an object that can be used if no genes are mapped by enricher
setClass("nullResult", slots="result")

# Clusters --> enrichment function
clustEnrich = function(clust_list, fileseed, universe=NULL, qvalueCutoff=0.05, gmtfile = "data\msigdb.v6.2.symbols.gmt"){ # clust_list -- list that should contain a list of dataframes
    (clusters) with one column (genes) named "Clusters"
    # as well as gene-cluster assignments in "Assignments"
    gmtfile -- MSigDB gmt file, default was downloaded 03-25-2019
    fileseed -- start of filename for saving files generated in this
               # function call
    # universe -- gene background universe, NULL to use clusterProfiler
    default (all annotated)
    clust_c = clust_list[["Clusters"]]
    # Take the list of 1-D cluster
    dataframes from the input nested list
    mycl = clust_list[["Assignments"]]
    # Gene-cluster assignments
#Perform enrichment
msigdb = read.gmt(gmtfile)
clust_c_msigdb = vector('list')  #Declare an empty list
for (k in 1:length(clust_c)){
    clust = paste0("clust",k)
    e = enricher('gene'=clust_c[[k]][,1], universe = universe,
    TERM2GENE = msigdb, qvalueCutoff = qvalueCutoff)
    if(is.null(e)){
        clust_c_msigdb[[clust]] = new("nullResult",
        result=data.frame("Result"="No mapped genes found"))
    }else{
        clust_c_msigdb[[clust]] = e
    }
}
#Write output file
filename = paste0(fileseed, 
    "_clustEnrich.xlsx")
writeEnrichResult(mycl, clust_c_msigdb, filename)
return(clust_c_msigdb)

#Multi-sheet excel reader helper function
#Based on nhttps://stackoverflow.com/questions/12945687/read-all-
worksheets-in-an-excel-workbook-into-an-r-list-with-data-frames
read_excel_allsheets = function(filename, tibble = FALSE) {
    sheets = readxl::excel_sheets(filename)
    x = lapply(sheets, function(X) readxl::read_excel(filename, sheet = X))
    if(!tibble) x = lapply(x, as.data.frame)
    names(x) = sheets
    return(x)
}

#Make a ggplot2 neglog FDR barplot for top enrichments from a given
cluster in a list from a writeEnrichResult xlsx
enrichBarplot = function(enrichList, clustNum, title,
    numEnrichments=5){
    enrichResult = enrichList[[paste0("clust",clustNum)]]
    enrichPlot = head(enrichResult, numEnrichments)
    enrichPlot$neglog10FDR = -log10(enrichPlot$qvalue)
    b = ggplot(enrichPlot, aes(x=reorder(ID, neglog10FDR),
    y=neglog10FDR, fill=neglog10FDR)) +
        geom_bar(stat='identity') +
        coord_flip() +
        xlab(element_blank()) +
        ylab(expression(paste(-log[10], "(FDR)", sep=''))) +
        labs(fill=expression(paste(-log[10], "(FDR)", sep=''))) +
        ggtitle(title)
    return(b)
}
#Function to read a marker-cluster dataframe from Seurat to form required by clustEnrich function
clustReshape = function(markers){
  #markers -- a data frame like that produced by Seurat FindAllMarkers with marker-gene assignments
  clustnames = names(table(markers$cluster))
  clusters = vector('list', length=length(clustnames))
  for(i in 1:length(clustnames)){
    genes = filter(markers, cluster==clustnames[i])$gene
    clusters[[i]] = data.frame('genes'=genes)
  }
  assignments = setNames(markers$cluster, markers$gene)
  output = list("Assignments"=assignments, "Clusters"=clusters)
}

#Extract gene list from a clustEnrich xlsx for a desired cluster and enrichment set ID
#Returns a vector of gene names
enrichGenes = function(clustEnrichResult, clustName, setName){
  #clustEnrichResult -- result of clustEnrich read by read_excel_allsheets
  #clustName -- cluster ID of desired result
  #setName -- gene set ID of desired result
  ce = clustEnrichResult[[clustName]]
 rownames(ce) = ce$ID
  genes = ce[setName, 'geneID']
  return(strsplit(genes, '/'))[[1]]
}

Helper function for finding TF targets from TRRUST database file

#Read the list of TF targets from TRRUST database
#http://www.grnpedia.org/trrust/
trrust = read.table("data\trrust_rawdata.mouse.tsv", stringsAsFactors=FALSE)
colnames(trrust) = c("TF","Target","Effect","ID")

#Function to return targets from trrust
trrustTargets = function(tf, trtrustMat, activated=TRUE, includeUnknown=TRUE){
  type = ifelse(activated,"Activation","Repression")
  alltargets = filter(trtrustMat, TF==tf)
  if(includeUnknown){
    targets=unique(filter(alltargets, Effect==type | Effect=="Unknown"))
  }else{
    targets=unique(filter(alltargets, Effect==type))
  }
  return(list(targets$Target))
}
Helper function for performing Mann-Whitney-Wilcoxon test on Seurat object module score

```r
moduleWilcox = function(seurat, module, identName, ident1, ident2){
  s = SetAllIdent(seurat, identName)
  s = SubsetData(s, ident.use=c(ident1, ident2))
  df = data.frame("ident"=factor(s@ident),
                   "score"=s@meta.data[module])
  colnames(df)[2] = "score"
  mww = wilcox.test(score~ident, data=df)
  return(mww$p.value)
}
```

Helper function for computing mean avg_logFC across genes in a set for two identity classes in a Seurat object

```r
#Function that given a seurat object and a list of genes returns the mean avg_logFC between two broad identity classes of interest
gsMeanFC = function(seurat, genes, splitIdent, split1, split2){
  #seurat -- Seurat object, should already be set to ident to group results by
  #genes -- character vector of genes to calculate
  #splitIdent -- ident to split clusters by to calculate FC
  #split1 -- numerator group for FC
  #split2 -- denominator group for FC
  avgExp = AverageExpression(seurat, genes.use=genes, add.ident=splitIdent, show.progress=F)
  avg_logFC = data.frame(matrix(nrow=length(genes), ncol=0))
  for(i in levels(seurat@ident)){
    #Calculate avg_logFC the way Seurat does w/pseudocount per https://github.com/satijalab/seurat/issues/741
    avg_logFC[,i] = log(avgExp[,paste(i,split1,sep="_")]+1) - log(avgExp[,paste(i,split2,sep="_")]+1)
  }
  return(colMeans(avg_logFC)) #Mean avg_logFC across genes in set
}
```

Helper functions for computing expression score of a given gene set in a Seurat object (e.g., TF targets) and computing significance of correlations in such scores

```r
#Function for computing new scores
exprScore = function(seurat, geneList){
  return(Matrix::colSums(expm1(seurat@data[unlist(geneList),])))
  #expm1 to return to #transcripts/10000
}
```

```r
#Find p-values for each of these correlations
gsCorrDist = function(seurat, set1, set2, numTests=10000){
  #seurat -- seurat object to analyze
  #set1 -- gene set 1 list
```
#set2 -- gene set 2 list
#returns an ecdf function that can be used to determine percentile value of a corr
set1 = unlist(set1)
set2 = unlist(set2)
ns1 = length(set1); ns2 = length(set2)
totgenes = length(union(set1,set2))
corrvals = vector(length=numTests)
for(i in 1:numTests){
  if(i%%100==0){
    print(i)
  }
  genes = sample(rownames(seurat@data), totgenes, replace=F)
  rs1 = list(head(genes, ns1))
  rs2 = list(tail(genes, ns2))
  rs1scores = exprScore(seurat, rs1)
  rs2scores = exprScore(seurat, rs2)
  corrsvals[i] = cor(rs1scores, rs2scores)
}
return(ecdf(corrvals))
}
References


[79] G. Hong et al., “EGFL7: Master regulator of cancer pathogenesis, angiogenesis


