

**Computational Investigation of Endothelial Heterogeneity
During Homeostasis and Inflammation**

BY

Ankit Jambusaria
B.S. - Bioengineering, University of California Santa Cruz, 2013

THESIS

Submitted as a partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Bioinformatics
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Jalees Rehman, Departments of Medicine, Pharmacology and
Bioengineering, Chair and Advisor
Asrar B. Malik, Department of Pharmacology, Advisor
Yang Dai, Department of Bioengineering
James Lee, Department of Bioengineering
Sang Ging Ong, Departments of Pharmacology and Medicine

© Copyright by

Ankit Jambusaria

2020

To my family for their love and support.

ACKNOWLEDGMENTS

I would like to convey my appreciation for my advisors – Drs. Jalees Rehman and Asrar B. Malik for their mentorship, genuine support, and valuable guidance. They have largely contributed to my development as a computational biologist, but also as a young professional. I am grateful for the time and effort they devoted towards developing my projects by providing their expertise and creativity. Dr. Jalees Rehman blossomed my interest in pursuing novel and clinically relevant directions in cardiovascular research. Thank you for not only broadening my vascular biology knowledge, but also motivating me to be informed about other fields of research. I am also grateful for Dr. Malik, who welcomed me into the Department of Pharmacology and provided me with the stellar environment and resources I needed to develop my computational and biological expertise. I will never forget the confidence both my advisors put in me through my Ph.D. studies.

I would also like to acknowledge the members of my thesis committee, Drs. Yang Dai, James Lee, and Sang Ging Ong for their time and expertise in shaping my dissertation. I would like to specially thank Dr. Yang Dai, who invited me to attend her group meetings, present my work, and provide computational feedback. I would also like to show appreciation to the fantastic collaborators who offered computational and cardiovascular expertise in my projects and publications.

To my fellow lab mates and colleagues, I would like to say thank you for cultivating an unforgettable environment. I am forever grateful that you all

ACKNOWLEDGMENTS (Continued)

willingly shared your knowledge and advice to help me succeed. From our thoughtful conversations about science to our philosophical conversations about life, you have been the most valuable resource when I have needed help.

I am forever indebted to my family for their love and support. I would have never pursued my Ph.D. research without the inspiration from them. My father's interest in engineering and his endless curiosity provided me with an upbringing that motivated me to pursue a career in computational biology. My mother's emotional support and encouragement proved to me that she is my number one fan through failure and success. Thanks to my loving brother, Sagar, who has always been someone I admire for his determination, hard work ethic, and humility. The three of them will always be in my heart.

Finally, I would like to thank my beautiful wife, Maithili, for her endless sacrifice, patience, and support. She has been beside me throughout my entire Ph.D. journey. Through all of the good and bad times, you stood behind me and provided me with the strength and motivation I needed to be the best that I can be.

CONTRIBUTION OF AUTHORS

Jambusaria, A., et al. (2018). "A computational approach to identify cellular heterogeneity and tissue-specific gene regulatory networks." *BMC Bioinformatics* 19(1): 217.

Ankit Jambusaria conceived and designed the computational analysis strategy, prepared the manuscript, developed the HeteroPath algorithm, analyzed the data, generated tissue-specific transcriptional networks, and revised the manuscript.

Jeff Klomp performed the PGSEA analysis and reviewed the computational analysis.

Zhigang Hong offered expertise regarding the biological context of the experimental data.

Yang Dai provided expertise to develop the computational analysis strategy, prepared the manuscript, and revised the manuscript.

Shahin Rafii designed the studies for the EC isolation and provided the mouse EC microarray data.

Asrar B. Malik conceived and designed experiments and revised the manuscript.

Jalees Rehman conceived and designed experiments, interpreted data, and revised the manuscript.

CONTRIBUTION OF AUTHORS (Continued)

Jambusaria, A., et al. (2020). "Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation." Elife 9.

Ankit Jambusaria conceptualized and designed computational methods, curated the data, performed formal computational analysis, validated the analysis, drafted the manuscript, and revised the manuscript.

Zhigang Hong conceptualized and implemented the RiboTag isolation method and performed proper validation studies.

Lianghui Zhang performed cell isolation and immunostaining experiments, and drafted parts of the manuscript.

Shubhi Srivastava performed cell isolation and immunostaining experiments, and drafted parts of the manuscript.

Arundhati Jana performed cell isolation and immunostaining experiments, and drafted parts of the manuscript.

Peter T. Toth designed imaging experiments and interpreted data.

Yang Dai provided expertise to develop computational analysis strategy and revised the manuscript.

Asrar B. Malik conceived and designed experiments and revised the manuscript.

Jalees Rehman conceived and designed experiments, interpreted data, and revised the manuscript.

SUMMARY

Endothelial cells (ECs) are a specialized cell type lining all vertebrate blood vessels and form an interface between circulating blood and the neighboring tissue. Understanding the tissue-specific characteristics of endothelial cells could markedly improve our understanding of the organ-specific roles of blood vessel function and the development of vascular disease. Computational approaches statistically modeling the gene expression data provided by high throughput sequencing technologies have been developed to analyze the cellular transcriptome. In our work, we present novel computational methods, namely HeteroPath and Subnetwork Signaling Entropy Analysis (SSEA), to analyze gene expression data and ascertain organ-specific endothelial heterogeneity during homeostasis and inflammation.

We hypothesized that characterizing cells from distinct tissues based on the heterogeneity of the molecular signaling would allow for the precise identification of clusters of genes which are uniquely upregulated or uniquely downregulated in each tissue. Using HeteroPath alongside traditional gene set enrichment analysis methods, we demonstrated endothelial transcriptomic heterogeneity. HeteroPath specifically identified organ-specific signaling pathways and provided a comprehensive characterization of EC heterogeneity in the healthy state. We next adopted the RiboTag mRNA isolation technique to directly isolate tissue-specific mRNAs undergoing translation without cell disassociation to understand the nature of the endothelial translome in vivo. By performing RNA-Sequencing and computationally analyzing the endothelial

SUMMARY (Continued)

translatome, we identified specific pathways, transporters, and cell-surface markers expressed in an organ-specific manner. In addition, we found that ECs adopt the characteristics of the tissue by expressing genes typically expressed in the surrounding tissue such as genes associated with synaptic function in the brain endothelium and cardiac contractile genes in the heart endothelium.

Once we established the organ-specific endothelial signature during homeostasis, we studied whether this heterogeneity persisted in response to a biological stimulus that induced systemic inflammation. Using differential expression approaches and our novel framework, SSEA, we quantified the organ-specific endothelial gene expression dynamics and found that the progression and resolution of endothelial injury during vascular inflammation in each organ is mediated by distinct endothelial signaling mechanisms. Using these methods and tools, we characterized organ-specific endothelial heterogeneity during homeostasis and inflammation and provided insights regarding the underlying endothelial biology and potential therapeutic targets.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
CONTRIBUTION OF AUTHORS.....	vi
SUMMARY	viii
LIST OF FIGURES	xiv
LIST OF EQUATIONS.....	xvi
CHAPTER 1: INTRODUCTION AND BACKGROUND	1
1.1 Origin and Development of the Vasculature	1
1.1.1 Vasculogenesis.....	1
1.1.2 Angiogenesis	2
1.2 Endothelial Functions	4
1.2.1 Endothelial Permeability	5
1.2.2 Endothelial Immune Cell Trafficking	6
1.2.3 Endothelial Nutrient Transport	7
1.2.4 Endothelial Metabolism.....	8
1.2.5 Endothelial Inflammation	10
1.3 Endothelial Heterogeneity	11
1.4 Computational Analysis of Endothelial Heterogeneity	13
1.4.1 Gene Set Analysis	14
CHAPTER 2: HETEROPATH: A COMPUTATIONAL APPROACH TO IDENTIFY CELLULAR HETEROGENEITY AND TISSUE-SPECIFIC GENE REGULATORY NETWORKS	17
2.1 Introduction.....	17
2.1.1 Problem Definition	19
2.2 Methods.....	19
2.2.1 HeteroPath Overview.....	19
2.2.2 HeteroPath Algorithm	22
2.2.3 Methods for Gene Set Analysis	24
2.2.4 HeteroPath Performance Evaluation	25
2.2.5 HeteroPath Performance Simulations.....	26
2.2.6 Annotated Transcriptional Regulators of Heterogeneous Genes.....	28
2.2.7 Putative Regulatory Transcription Factors	29
2.2.8 Constructing Gene Regulatory Networks.....	30
2.2.9 Pairwise Differential Gene Expression.....	33
2.3 Experiments & Results	34
2.3.1 Datasets	34

2.3.2	Identification of Heterogeneously Expressed Tissue-specific Pathways	34
2.3.3	HeteroPath Benchmarking.....	35
2.3.4	Comparison of HeteroPath, GSEA, and PGSEA using Simulated Data	40
2.3.5	Case Study 1: Endothelial Heterogeneity	45
2.3.6	Case Study 2: Neuronal Heterogeneity	52
2.4	Conclusions	59

CHAPTER 3: ENDOTHELIAL TRANSLATOME HETEROGENEITY ACROSS DISTINCT VASCULAR BEDS DURING HOMEOSTASIS..... 64

3.1	Introduction.....	64
3.1.1	Problem Definition	66
3.2	Methods.....	66
3.2.1	Experimental Animals	66
3.2.2	Isolation of Mouse Brain, Lung, and Heart.....	67
3.2.3	Ribosome Immunoprecipitation (IP).....	67
3.2.4	RNA-Sequencing	68
3.2.5	RNA-Sequencing Data Processing and Batch Correction	69
3.2.6	Baseline Tissue-specific Gene Signatures	69
3.2.7	Baseline Tissue-specific Pathway Analysis	70
3.2.8	Baseline Cell Surface Markers.....	70
3.2.9	Isolation of Brain ECs	70
3.2.10	Isolation of Lung ECs.....	71
3.2.11	Isolation of Heart ECs.....	71
3.2.12	Preparation of Cytospin Slides from Brain, Lung, and Heart Endothelial Cells	72
3.2.13	Immunofluorescence and Confocal Microscopy of Brain, Lung, and Heart Endothelial Cells	73
3.2.14	Assessing Baseline Endothelial Translatome Heterogeneity	73
3.2.15	Computational Assessment of mRNA Purity.....	74
3.2.16	Single-cell Endothelial Transcriptome Heterogeneity	75
3.2.17	Comparison of Organ-specific Endothelial Translatome and Endothelial Single-cell Transcriptomic Data	76
3.2.18	Online Endothelial Translatome Expression Database.....	77
3.3	Experiments & Results	77
3.3.1	Optimized Platform to Characterize Organotypic Endothelial Heterogeneity	77
3.3.2	RiboTag ^{EC} Endothelial mRNA Purity	85
3.3.3	Brain-specific Endothelial Molecular Signature.....	96
3.3.4	Lung-specific Endothelial Molecular Signature	100
3.3.5	Heart-specific Endothelial Molecular Signature	103
3.3.6	Single-cell Endothelial Heterogeneity	107

3.3.7	Organ-specific Parenchymal Gene Signature Exists in Endothelial scRNA-Seq	111
3.4	Conclusions	116
CHAPTER 4: TISSUE-SPECIFIC ENDOTHELIAL GENE EXPRESSION DYNAMICS DURING INFLAMMATION.....		122
4.1	Introduction	122
4.1.1	Problem Definition	127
4.2	Methods	127
4.2.1	Experimental Animals	127
4.2.2	Assessing Endothelial Heterogeneity using the RiboTag method	128
4.2.3	Tissue-specific Endothelial Kinetics following LPS-induced Injury	131
4.2.4	Early and Late Tissue-specific Inflammatory Markers	131
4.2.5	Online Endothelial Translatome Expression Database	131
4.2.6	Genome-scale Co-expression and Fold Change Networks (FCN)	132
4.2.7	Gene Community Detection in the Networks	135
4.2.8	Network Signaling Entropy for Subnetworks Identified by Community Detection Algorithms	136
4.3	Experiments & Results	138
4.3.1	In situ Organ-specific Endothelial Early- and Late-Inflammation Signature	138
4.3.2	Tissue-specific Dynamic Response following LPS-induced Inflammatory Activation	145
4.3.3	Louvain Community Detection in Co-expression Networks versus Fold Change Networks	152
4.3.4	Identifying Dysregulated Subnetworks associated with the Progression of LPS injury	154
4.3.5	Subnetwork Dynamics during the Progression of LPS Injury	158
4.4	Conclusions	160
CHAPTER 5: CONCLUSIONS		164
5.1	A Pathway-based Computational Modeling Approach to Identify Tissue-Specific Gene Expression Networks	166
5.1.1	Limitations and Future Research	169
5.2	Unraveling Tissue-Specific Endothelial Cell Heterogeneity Using Pathway-based Computational Modeling of Gene Expression	171
5.2.1	Limitations and Future Research	174
5.3	Capturing Tissue-specific Endothelial Translatome Dynamics during Systemic Inflammation	176

5.3.1	Limitations and Future Research	179
	CITED LITERATURE	181
	APPENDIX A: COPYRIGHT PERMISSIONS	197
	APPENDIX B: CURRICULUM VITAE	200

LIST OF FIGURES

Figure 1: Schematic Diagram of the HeteroPath framework designed for tissue-specific transcriptomic profiling.....	21
Figure 2: The HeteroPath algorithm for identifying heterogeneous pathways and gene sets.....	23
Figure 3: A) Hierarchical clustering of Endothelial cells from 7 mouse organs Intra- and inter-tissue heterogeneity.....	32
Figure 4: Comparison of pathways enriched using HeteroPath, GSEA, and PGSEA.....	39
Figure 5: Comparison of statistical power and type-I error rate between HeteroPath, GSEA, and PGSEA for DE Gene Set size of 50 genes.....	42
Figure 6: Comparison of statistical power and type-I error rate between HeteroPath, GSEA, and PGSEA for DE Gene Set size of 150 genes.....	44
Figure 7: Endothelial cell heterogeneity.....	49
Figure 8: Endothelial gene regulatory networks generated from HeteroPath tissue-specific pathways.....	50
Figure 9: Neuronal heterogeneity	55
Figure 10: Neuronal gene regulatory networks generated from HeteroPath tissue-specific pathways.	58
Figure 11: RiboTag Isolation of endothelial mRNA.....	82
Figure 12: Endothelial heterogeneity exists in classic endothelial functions.....	84
Figure 13: Characterization of Whole Brain RNA-Seq data.....	89
Figure 14: Characterization of Whole Lung RNA-Seq data	91
Figure 15: Characterization of Whole Heart RNA-Seq data	93
Figure 16: Kendall's Tau correlation supports endothelial mRNA isolation from RiboTag ^{EC} mice.....	95
Figure 17: Brain endothelial specific signature.	99
Figure 18: Lung endothelial specific signature	102
Figure 19: Heart endothelial specific signature.....	106
Figure 20: Single-cell endothelial heterogeneity	110
Figure 21: Spearman correlation scatter plots of average gene expression in RiboTag bulk RNA-Seq and Betsholtz scRNA-Seq.....	114
Figure 22: Expression Correlation Analysis between endothelial gene expression generated by RiboTag, Betsholtz, and Tabula Muris.....	115
Figure 23: RiboTag mRNA isolation and time-series transcriptomic analysis...	130
Figure 24: Subnetwork Signaling Entropy Analysis (SSEA).	134
Figure 25: Markers of early (6 hr) and late (24 hr) LPS-induced inflammation in brain, lung, and heart ECs.....	140
Figure 26: The early inflammation (6 hr) markers across organ-specific endothelial cells.....	141
Figure 27: The late inflammation (24 hr) markers across organ-specific endothelial cells.....	143
Figure 28: Organ-specific endothelial cells uniquely regulate endothelial genes during the progression and resolution of inflammation	149

Figure 29: Organ-specific endothelial cells uniquely regulate glycolysis genes during the progression and resolution of inflammation	150
Figure 30: The performance of Louvain on the co-expression network versus the fold change network	153
Figure 31: Top Pathways identified in the Brain EC Fold Change Subnetworks by SSEA.....	156
Figure 32: Top Pathways identified in the Brain EC Co-expression Subnetworks by SSEA.....	157
Figure 33: Brain Endothelium Wnt Signaling Subnetwork Dysregulation	159

LIST OF EQUATIONS

Equation 1	22
Equation 2	25
Equation 3	26
Equation 4	135
Equation 5	136
Equation 6	137
Equation 7	137

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Origin and Development of the Vasculature

Vascular endothelial cells (ECs) are an abundant cell type lining the inner wall of every blood vessel in the body. Blood vessels possess a thick wall of connective tissue composed of fibroblasts, adipocytes, mast cells, macrophages, lymphocytes, and smooth muscle cells. On the luminal side, ECs form a thin monolayer named the endothelium which lines the vessel wall and serves as an interface between the blood and the surrounding tissue¹.

The first functional organ system to arise in the mammalian embryo is the circulatory system which plays an essential role in the survival and growth of the developing embryo^{2,3}. Laboratory mice are a powerful model for elucidating the phenotypic and molecular mechanisms regulating vascular development. In mice, ECs and red blood cells first arise in the yolk sac by embryonic day 7 and begin to massively expand. ECs, like hematopoietic cells, arise from the mesoderm through differentiation of hemangioblasts^{4,5}. Hemangioblasts possess the multipotent characteristic enabling them to also give rise to the vascular smooth muscle cells⁶, skeletal muscle cells, and cardiac muscle cells⁷.

1.1.1 Vasculogenesis

Vasculogenesis is the formation of blood vessels during embryonic development⁸. Vasculogenesis primarily occurs during early embryogenesis during which *de novo* capillaries are formed from endothelial precursor cells. This observed biological process begins with the migration of blood and endothelial

precursor cells, referred to as angioblasts, from the mesoderm to the yolk sac^{9,10}. After angioblasts begin proliferating and differentiating into ECs, they migrate into the yolk sac and establish a simple capillary network. This primitive capillary plexus is composed of interconnected microvessels which are homogeneous in shape and size^{8,11,12}.

The de novo synthesis of the vascular network during embryonic vascular development is mediated by several signaling mechanisms. Vascular endothelial growth factor (VEGF) promotes EC migration and proliferation into a primitive vascular network¹³⁻¹⁶. Scl/T-cell acute lymphoblastic leukemia transcription factor (Scl) is a vital factor for the generation of blood and endothelial cells from progenitor cells¹⁷⁻¹⁹; and, vascular endothelial growth factor receptor 2 (Flk1) which is a regulator of the multipotent ability of progenitor cells during capillary plexus formation. The establishment of the network of blood vessels in the yolk sac is an essential step in the highly dynamic development of the vasculature²⁰.

1.1.2 Angiogenesis

This primitive network formed during embryonic vascular development is able to expand due to endothelial migration and proliferation from capillaries. This process, termed angiogenesis, consists of pre-existing vessel sprouting, bridging, and branching. Angiogenesis enables the existing vascular plexus to remodel into a hierarchically branched, highly organized vascular structure consisting of arteries and veins. For vessel stabilization, pericytes and smooth muscle cells are recruited to surround the endothelial layer. This vascular remodeling shapes vessels into low-resistance conduits optimized for tissue-specific functions

including transport of nutrients, oxygen, blood, and waste. The remodeling process regulated by both molecular and biochemical signaling is crucial for cardiovascular function and the viability of the growing embryo.

The well-known signaling pathways linked to angiogenesis include the Tie family of receptor tyrosine kinases, transforming growth factor β superfamily (TGF β), the Eph receptor tyrosine kinases and their membrane-bound ligands, Notch signaling, and platelet derived growth factor B (PDGFB). The expression of Tie1 and Tie2 appear early on during angioblast specification and play a critical role in vascular remodeling after the generation of the capillary plexus²¹⁻²³. Dysregulation of the TGF β signaling pathway or bone morphogenic proteins and activins resulted in stunted development of the yolk sac, leaky vessels, and defects in hematopoietic and endothelial differentiation²⁴⁻²⁷. After vascular remodeling occurs, the yolk sac and the embryo vasculature become compartmentalized with their own vascular network composed of arteries and veins responsible for providing the developing embryo with oxygen and nutrients. The ephrins are among the first identified factors which regulate distinct arterial and venous specifications of nascent vessels²⁸. Following the discovery of the ephrins, Notch signaling was found to be an upstream regulator of ephrin signaling in arterial venous identity. This arterio-venous specification and maintenance is essential for proper circulation. To further ensure vascular stability for circulation, ECs secrete PDGFB which binds to PDGF receptors on mural cells and activates signaling pathways that maintain the integrity of the vessel wall^{29,30}.

As development proceeds, in addition to arterio-venous cell fate identity the hierarchical vessel tree diversifies due to changes in hemodynamics, surrounding cell types, and the microenvironment. These microenvironmental cues guide the individual ECs into developing tissue-specific characteristics necessary for distinct organ function³¹⁻³⁴. The largely homogenous embryonic EC population must differentiate and mature in an organ-specific manner in order engage in endothelial functions such as regulating flow, transport, hormonal interactions, cellular trafficking, metabolism, and inflammation^{35 36} that are attuned to the requirements of a given organ.

1.2 Endothelial Functions

ECs engage in several different functions, some of which are unique to a specific organ and tissue and others which are universal EC functions in most tissues and organs. For example, ECs regulate hemostasis – the cessation of bleeding – in most organs but even for this near universal EC function, studies have identified varying distribution of EC factors involved in the formation of blood clots³⁷, thus suggesting that there is indeed a heterogeneity and organ-specific phenotype of ECs. Similar observations have been made for several pan-endothelial functions such as blood vessel permeability, nutrient transport, immune cell trafficking and inflammation, and the regulation of tissue metabolism.

Larger vessels such as arteries and veins branch into increasingly smaller vessels until they become microvascular capillaries which represent the connection between the arterial and venous circulation and which perfuse the

tissue, thus supporting organ-specific parenchymal cells such as cardiomyocytes in the heart or hepatocytes in the liver. The capillary endothelium in each organ consists of a thin sheet of cells functioning as a semi-permeable barrier between the blood and the tissue parenchymal cells. Adjacent endothelial cells dynamically modulate junctions between themselves to regulate permeability of the endothelium allowing for the passage of certain cell types and macromolecules. The multi-tasking nature of the endothelium is necessary for regulating physiological processes and pathophysiological responses including acute inflammation, antigen presentation and recognition, and immune surveillance. Persistent permeability due to prolonged disruption of endothelial cell-cell junctions can lead to excessive inflammation.

1.2.1 Endothelial Permeability

Activation of signaling mechanisms enables the endothelium increases and restore permeability by modulation of cell junctions³⁸⁻⁴⁰. The regulation of endothelial permeability is heterogeneous in response to an external stimulus within a vascular bed and also differs widely across distinct vascular beds as each organ-specific endothelium likely uses tissue-specific signaling mechanisms regulating permeability⁴¹. On one end of the spectrum there is the highly selective continuous endothelium forming the blood-brain barrier containing specialized tight junctions to protect neurons from toxins. On the other end of the spectrum are the discontinuous liver sinusoidal ECs which form a dynamic filtration barrier containing fenestrae with pore sizes ranging from 100-2000 nm

for endocytosis-based clearance of waste macromolecules and which therefore have very high permeability even during physiological homeostasis⁴².

1.2.2 Endothelial Immune Cell Trafficking

Aside from permeability, the vasculature is essential for the regulation of immune cell trafficking. Since ECs are among the first cells of the body (other than circulating blood leukocytes) to come into contact with circulating pathogens or toxins in the blood, ECs act as key sensors which can respond by activating the innate immune system. It has been shown that ECs express toll-like receptors (TLRs), nod-like receptors (NLRs), and chemokine receptors to detect foreign substances^{43,44}. In response to pathogen and toxin detection, ECs respond by producing and secreting pro-inflammatory cytokines^{45,46}. This recognition and response mechanism in endothelial cells initiates the recruitment of leukocytes^{47,48}.

Not only do ECs recruit immune cells to the inflammatory tissue, but ECs also physically interact with immune cells. Once the endothelium become activated during inflammatory conditions, the ECs begin to upregulate typical inflammatory adhesion molecules such as selectins, vascular cell adhesion molecule-1 (VCAM1), and intercellular adhesion molecule-1 (ICAM1) to promote immune cell trafficking across the endothelial barrier⁴⁹. The endothelial specific expression of VCAM1, ICAM1, E-Selectin, and P-Selectin is critical in mediating the leukocyte adhesion cascade of events. First, the ECs secrete chemokines to recruit and activate the leukocytes. This promotes leukocyte attraction to ECs at the site of injury. The leukocytes tether and roll along the endothelium until

binding of selectins and integrin results in leukocyte arrest⁵⁰. Once arrested, the leukocytes then transmigrate across the endothelium into the inflamed tissue while binding to adhesion molecules such as platelet endothelial adhesion molecule-1 (PECAM1) and ICAM1⁵¹. The transmigration can occur in two different fashions. One route is paracellular where the leukocytes pass between the ECs while the other is transcellular, directly through the ECs⁵². The paracellular migration requires a relaxation of the endothelial junctions to create large enough gaps for immune cells from the blood to pass through⁵¹. The transcellular migration, on the other hand, involves remodeling of the membrane where a conduit is repeatedly opened and closed⁵².

1.2.3 Endothelial Nutrient Transport

The endothelial barrier is further responsible for regulating delivery of nutrients from the blood into the tissue. In the energy-consuming heart tissue, the cardiac ECs adjust to the metabolic demands of the tissue by controlling vessel density and endothelial nutrient transfer. Contractions of the myocardium heavily relies on ATP generated from fatty acid catabolism and the cardiac ECs express high levels of fatty acid transporters such as CD36⁵³. In the brain tissue, neurons consume high levels of glucose for ATP production that is required for neuronal activity. Therefore, brain ECs have high levels of the glucose transporter, GLUT1 which transports glucose from the blood into the brain tissue. This transporter is essential for the brain because dysregulation of GLUT1 leads to reduced levels of glucose which has been linked to neuronal dysfunction and neurodegenerative disease⁵⁴⁻⁵⁶. Other solute transporters including solute carrier family 7 member 5

(SLC7A5) have been associated with autism spectrum disorders due to the modulated levels of essential amino acids in the brain⁵⁷.

1.2.4 Endothelial Metabolism

Emerging evidence has suggested that endothelial metabolism allows EC to adapt to the tissue specific functions are supply the tissue with the necessary nutrients that it imports from the circulating blood. Dysregulation of EC metabolism has been associated with many diseases including atherosclerosis, diabetes, neovascular eye disease, and cancer. EC metabolism is tightly regulated because it is imperative for ECs maximally preserve nutrients for the tissue the ECs are supplying and minimize consumption of these nutrients and metabolites by the ECs themselves. Under healthy homeostatic conditions, ECs are quiescent in nature but possess the ability to quickly adapt into a proliferative state for angiogenesis⁵⁸. Like cancer cells, ECs rely heavily on glucose to generate 85% of their ATP and therefore consume high levels of glucose via the glycolysis pathway⁵⁹. Glucose metabolism occurs via two distinct pathways: aerobic and anaerobic. The aerobic, or oxidative metabolism of glucose occurs in the mitochondria and results in the greatest production of ATP, while the anaerobic cycle occurs in the cytoplasm and is less efficient at producing ATP⁵⁹. During angiogenesis, ECs migrate into a hypoxic environment to sprout and generate new blood vessels. In this hypoxic environment, the ECs are unable to rely on oxidative metabolism and must thus activate anaerobic metabolism⁶⁰. To initiate the glycolytic pathway during angiogenesis, glucose must be taken up by the ECs via membrane-bound glucose transporters (GLUTs)⁶¹.

Another energy source for ECs is fatty acids. In the setting of low glucose, ECs adapt by increasing their fatty acid oxidation (FAO) through the activation of AMP-activated protein kinase (AMPK)⁶². It has been previously shown that during basal EC proliferation, VEGF-dependent signaling induces the expression of fatty acid binding protein (FABP4) for fatty acid uptake and trafficking⁶³. The role of FAO in the endothelium is not well understood. Due to the similarity in ATP generation by cancer cells and endothelial cells, it has been hypothesized that endothelial FAO may regulate EC redox homeostasis in addition to biosynthesis of macromolecules^{64,65}. It has been previously reported that the cardiac endothelium and endothelium in skeletal muscle express high levels of fatty acid binding proteins for fatty acid transport into the tissue⁶⁶. The cardiac and skeletal muscle tissues then use their β -oxidation machinery to metabolize the fatty acids and generate energy⁶⁷.

Besides glycolysis and fatty acid oxidation, in recent years it has been shown that ECs metabolize specific amino acids for survival, migration, and proliferation. Arginine, for example, has been implicated in EC homeostasis due to its role in converting citrulline to nitric oxide via endothelial nitric oxide synthase (eNOS)^{68,69}. In addition, a non-essential amino acid, glutamine, has been broadly studied as an easily accessible and abundant fuel source in the plasma⁷⁰. ECs consume more glutamine than any other amino acid for TCA cycle anaplerosis, biomass synthesis and redox homeostasis. Experiments depleting glutamine by pharmacological inhibition or knockdown of glutaminase 1 (GLS1), have shown impairment of EC proliferation⁷¹. Interestingly, it was found that loss

of EC function and protein synthesis by glutamine withdrawal in ECs can be partially rescued with asparagine⁷².

1.2.5 Endothelial Inflammation

At rest, the endothelium is responsible for inhibiting coagulation of the blood, controlling blood flow, regulating barrier permeability, and reducing leukocyte entry into the tissue. In the presence of infectious microbes or during tissue injury, the process of acute inflammation begins with ECs becoming activated and recruiting neutrophils. EC activation consists of two responses: type I and type II activation. Type I activation is the acute response independent of transcription that effect vascular tone, permeability, and leukocyte diapedesis. Type II activation is a delayed response because it is dependent on transcription of genes which modulate the function of ECs in response to bacterial products or cytokines⁷³. Type II endothelial activation result in a loss of barrier integrity, expression of adhesion molecules, a functional switch from anti-thrombotic to pro-thrombotic, and cytokine production⁷⁴. A proper inflammatory response requires ECs to become activated, although prolonged activation can lead to endothelial dysfunction and vascular disease. Type II EC activation can be characterized by activation of a pleiotropic transcription factor, nuclear factor- κ B (NF- κ B), expression of cell-surface adhesion molecules such as endothelial leukocyte adhesion molecule (E-selectin), VCAM1, and ICAM1. Furthermore, the presence of pro-inflammatory cytokines including TNF α and IL-6 induce type II EC activation and initiate leukocyte recruitment and attachment ⁷⁵.

The inflammatory process causes the production of reactive oxygen species which ultimately leads to endothelial dysfunction ⁷⁶. Previous work established that oxidative stress regulates cytokine production and secretion ⁷⁷⁻⁷⁹ thus implicating its role in endothelial dysfunction. To regulate vascular homeostasis the endothelium has built a system which regulates autocrine, paracrine, and endocrine signaling. These distinct forms of signaling within the endothelium and between the endothelium and neighboring cell types are driven by pathways which regulate vasodilation, vasoconstriction, and expression of cytokines and adhesion molecules. For example, reduced levels of NO due to less production and/or increased degradation is a hallmark characteristic for early onset of endothelial dysfunction ⁸⁰. This NO driven endothelial dysfunction and imbalance of vasodilators and vasoconstrictors results in a vasculature characterized by vasoconstriction, leukocyte adhesion, and vascular inflammation.

1.3 Endothelial Heterogeneity

The ECs which line the inner walls of blood vessels exhibit drastic functional and molecular characteristics tailored for supporting the tissue they reside in. Across different vascular beds, ECs show structural heterogeneity due to the varying levels of endocytic pathway components such as clathrin-coated pits, varying numbers of caveolae in the transcytosis pathway, and unique expression levels of intercellular junctions, tight junctions, adherens junctions, or gap junctions. This organ-specific phenotypic heterogeneity observed in ECs is linked to both intrinsic genetic factors and extrinsic factors such as location, soluble mediators,

cell to cell contact, or mechanical forces^{41,81,82,83}. ECs adopt organ-specific characteristics via mechanisms such as cytokine release, production of metabolites, biophysical forces, and cellular interaction between ECs and parenchymal cells. It has been previously discussed through transplantation studies that the endothelial plasticity is driven by ECs altering their tissue-specific gene expression patterns⁸⁴.

The tissue-specific interaction between ECs and surrounding cells occurs as early as during development, when, for example, brain ECs instruct neuronal differentiation^{85,86}. Such tissue-specific endothelial adaptations persist throughout adulthood when brain ECs form a highly selective barrier composed of specialized tight junctions to limit neurotoxicity⁸⁷. In the lung, ECs differentiate in parallel with epithelial cells to form gas exchange units which are in contact with the external environment and thus need to ensure a rapid immune response⁸⁸. Heart ECs, on the other hand, are specialized in a manner to ensure ready supply of fatty acids to voracious cardiomyocytes which rely on continuous supply of fatty acids as the primary fuel to generate ATP necessary for cardiac contraction⁴².

The tissue-specific gene expression patterns which regulate these processes remain poorly understood. By studying patterns of gene expression in different tissues, insights into the regulatory landscape of each endothelial population can be obtained. Identifying differences in the expression levels of selected genes in endothelial cells from different tissues provides some insights into the molecular underpinnings of endothelial heterogeneity, however unbiased

transcriptomic profiling is likely to yield a more comprehensive evaluation of the genes and regulatory pathways underlying endothelial heterogeneity.

1.4 Computational Analysis of Endothelial Heterogeneity

Computational analysis of microarray and RNA-Seq gene expression data of a given cell type obtained from distinct organs and tissues enables the unbiased identification of gene candidates. Endothelial cells, neurons, macrophages or fibroblasts that reside in different tissues or organs are thought to have distinct functions and are thus likely to have unique gene expression signatures that reflect their tissue-specific adaptation and function. Several recent publications addressing cellular heterogeneity have provided widely applicable biological insight in many areas including disease subtypes⁸⁹, candidate biomarkers⁹⁰, and molecular mechanisms of disease⁹¹. After generating gene expression data for different tissues, the goal is to identify the molecular heterogeneity characterizing the tissues. The heterogeneous populations may represent previously unidentified molecular profiles responsible for tissue-specific function. To characterize heterogeneous or differentially regulated genes, differential expression analysis⁸³ and gene co-expression network analysis⁹² are commonly used. A limitation of these methods however is that the single dimensional analysis of genes does not identify the causal molecular mechanisms that regulate them⁹³. These methods rely on ranking individual genes by differential expression and subsequently inferring the underlying pathways or transcription factors that maintain the heterogeneous gene expression profiles. Importantly, deriving the initial rank-list of the most differentially expressed genes using these

conventional computational approaches does not consider whether differential expression of tissue-specific genes is concentrated within functional groups. From a biological perspective, functionally related genes often have similar expression patterns which match cell-specific phenotypes ⁹⁴. In order to identify the molecular signature of distinct cell populations, new methods, in addition to existing gene set analysis methods need to be developed to interpret dynamic changes within a group of genes with common function.

1.4.1 Gene Set Analysis

Over-representation analysis (ORA), functional scoring (FCS) and pathway topology-based methods are the three predominant classes of gene set analysis methods ⁹⁵. ORA is the most widely used method because of its simplicity, robust statistical model, and ease of use. In a recent publication it was reported that across the 68 gene set analysis methods tools, 40 of them were ORA-based ⁹⁶. ORA is a first-generation method that evolved from single gene analysis. It requires a list of differentially expressed genes identified by a single gene analysis method as input. Given this input list of differentially expressed genes L , and a gene set, G_i , ORA-based methods statistically evaluate whether there is an over-representation or under-representation of L in G_i . The null hypothesis is there is no association between L and G_i . These methods are powered by a pre-defined list of differentially expressed genes. The significance of over-representation in a gene set is computed based on the extent to which the gene set is enriched in these differentially expressed genes. A gene set which contains significantly more genes from L than randomly expected is more likely to be truly

related to the biological condition. The next group of gene set analysis methods, functional class scoring (FCS) was developed in order to transition away from powering the analysis using differentially expressed genes defined by arbitrary thresholds. These methods hypothesize that in addition to genes that have large expression changes, minimal coordinated changes in gene expression are functionally relevant. There have been several implementations of this approach in recent years including gene set enrichment analysis (GSEA)⁹⁷, sigPathway⁹⁸, FunCluster⁹⁹, and many others.

The gene expression measurements generated by microarray or RNA-seq platforms are not fully exploited in gene set analysis. These values are only used in the context of identifying differentially expressed genes (ORA), or to rank the genes (FCS), but not to assess the modulation of a single gene set such as a biological signaling pathway. For example, ORA methods are unable to detect the difference between a subset of genes being differentially expressed at a small magnitude (e.g. 2-fold) versus the same subset of genes being dysregulated by many orders of magnitude (e.g. 100-fold). The FCS methods similarly provide a ranking for genes according to expression but rely on the correlation between the genes and observed phenotypes. These two gene set analysis methods miss the contextual information of the differentially expressed genes within gene sets.

Treating the gene sets as groups of genes ignores the contextual information provided by annotation databases including KEGG or Gene Ontology. To be more precise, all the interactions and dependencies among the

elements within a gene set which likely dictate the function of the gene set are disregarded. Pathway topology-based methods are a powerful class of methods which have integrated gene expression data with interaction databases. In addition to gene expression, these approaches leverage the known signaling and interactions of each gene to define whether the gene set is enriched in a phenotype. If a signaling pathway is regulated by a few driver genes and those particular genes are modulated during disease conditions, it is likely that this pathway is implicated in disease. On the other hand, if several genes in a pathway are modulated during disease, but they do not greatly impact the pathway, changes in their expression levels may not be relevant. Performing an ORA-based or FCS-based gene set analysis is limited to identifying differentially expressed genes but does not provide insight about whether these expression changes are critical for signaling.

Although current gene set analysis methods overcome some of the weaknesses of single gene analysis, there remain shortcomings when investigating cellular heterogeneity. In this thesis, we develop novel quantitative approaches for identification of the molecular signatures which explain organ-specific endothelial heterogeneity. In addition, we propose using powerful mRNA isolation approaches to assess the *in situ* endothelial molecular signature across several mice organs. Our studies provide a direction for further research on identification of molecular signatures driving cellular heterogeneity *in vivo*.

CHAPTER 2

HETEROPATH: A COMPUTATIONAL APPROACH TO IDENTIFY CELLULAR HETEROGENEITY AND TISSUE-SPECIFIC GENE REGULATORY NETWORKS

Previously published as:

Jambusaria, A., et al. (2018). "A computational approach to identify cellular heterogeneity and tissue-specific gene regulatory networks." BMC Bioinformatics **19**(1): 217.

2.1 Introduction

Addressing how to detect heterogeneity of cells in any given tissue requires studying whether a relationship between two random variables has changes across conditions. This question is often fundamental to a scientific inquiry. For example, a vascular biologist may ask which sets of genes or pathways are modified in brain endothelial cells versus lung endothelial cells. This example considers only two groups of cells. Imagine uncovering the differences between three or more groups of cells when evaluating 15,000 genes. To answer the heterogeneity question in this context, we developed a computational method, HeteroPath, which statistically infers functional changes from high-throughput mRNA data. HeteroPath is a powerful tool which allows for studying individual genes in the context of their function.

Previous publications focusing on cellular heterogeneity have made it clear that nearly all cellular systems are composed of a heterogeneous population of cells^{100,101}. This heterogeneity is thought to exist to maintain homeostasis, enhance survival rates, and support tissue-specific functionality. For example, population-level survival strategies in single-celled and multicellular

organisms such as bet-hedging promote risk spreading and provide these organisms with the resources necessary to be fit for survival when faced with exogenous stresses¹⁰².

Conducting large scale transcriptomics analyses is a common approach for detecting cellular heterogeneity. Although cells within an individual have nearly identical genomes, tissue-specific development generates tissue-specific cell types with distinct gene expression profiles. For example, during embryonic development the endothelium first forms clusters known as blood islands throughout the embryo to establish a vascular circulatory network¹⁰³⁻¹⁰⁵. During development, the immature network remodels in response to microenvironmental cues including neighboring cell types and hemodynamics. This maturing process of the network is critical for the establishment of tissue-specific functionality and the function of each organ³¹⁻³³. This developmental and maturation process ensures that the initial homogenous population of angioblasts differentiates into organ-specific endothelial cells fit for supporting the flow, transport, hormonal interactions, and cell trafficking in each organ^{35,36}. Since little is known about how each organ maintains the functional properties of its endothelium for instance, we developed HeteroPath to identify the molecular signature driving these functional properties.

The HeteroPath algorithm is designed to identify heterogeneous pathways and gene sets. HeteroPath aims to find the pathways or gene sets that are not only differentially expressed from the global median gene expression value but also appear to be responsible for the regulation of distinct cell types. Briefly, we

first integrate normalized gene expression data with pathway data for all pathways and gene sets. For each group of samples, we calculate a t-statistic based heterogeneity score for every gene in every pathway. By aggregating the heterogeneity score for all heterogeneous genes in each pathway, we can rank the pathways in order of heterogeneity score. This ranked list of pathways defines the functional and molecular signature of each group of samples.

2.1.1 Problem Definition

In this study, we design a novel computational method, HeteroPath, which first identifies heterogeneously expressed pathways in cell populations of unique organs or tissues. The algorithm aims to determine a pathway heterogeneity score which allows for individual elements of the pathway to be either upregulated or downregulated when compared to the median of all tissues. This computational model generates experimentally testable predictions for understanding the general architecture of the gene regulatory networks that define cellular heterogeneity.

2.2 Methods

2.2.1 HeteroPath Overview

We developed HeteroPath, a novel computational algorithm to identify significant heterogeneous pathways, in which the heterogeneity of pathways for a given set of cell types was evaluated (Figure 1). We applied a fixed gene set enrichment analysis which combined the gene expression and pathway data. We tested for bidirectionally perturbed pathways in the KEGG database. We were specifically interested in these bidirectionally perturbed pathways because they are individual

pathways which contain a significant number of elements that are upregulated and a significant number of elements which are downregulated.

As seen in Figure 2, the algorithm iterates through all KEGG pathways and assigns a heterogeneity score to each pathway. The heterogeneity score is generated by summing the absolute t-statistics of genes with sufficient magnitude to determine significant differences in gene expression across multiple tissue types. The t-statistic is used as a distance metric to quantify tissue-specific association between gene expression profiles on a per-gene basis. Therefore, the heterogeneity scores factor in both direction and magnitude of perturbation. Although a t-statistic based heterogeneity score of 0.975 or greater is equivalent to a p-value < 0.05 , it is not appropriate to calculate a p-value from this statistic because of the small number of genes associated with some pathways. Therefore, a permutation-based p-value is estimated in our algorithm. An adjusted p-value is then calculated to control for the false discovery rate. The stringent Benjamini-Hochberg correction method is applied to the raw p-values produced from the permutation-based calculation¹⁰⁶.

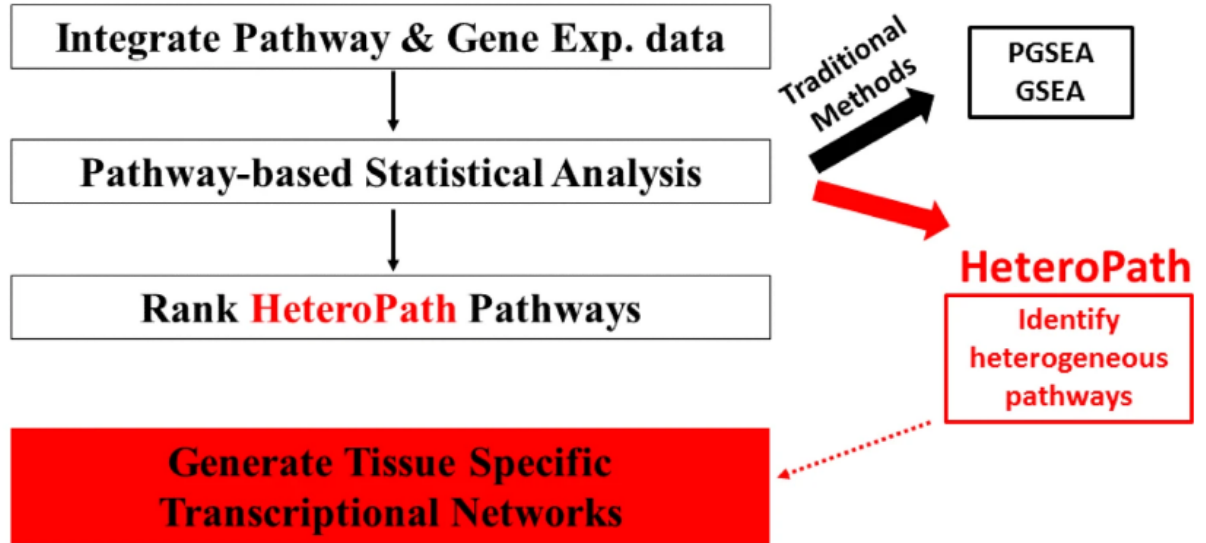


Figure 1: Schematic Diagram of the HeteroPath framework designed for tissue-specific transcriptomic profiling. First, the gene expression data is preprocessed and normalized. Then, the gene expression data and gene set data are integrated together. Each KEGG pathway is statistically evaluated using the traditional algorithms GSEA, PGSEA, and the novel HeteroPath algorithm to identify tissue-specific pathways. Next, the tissue-specific gene regulatory networks are constructed by identification of heterogeneous genes and their regulatory transcription factors as determined by motif enrichment analysis using the ENCODE database.

2.2.2 HeteroPath Algorithm

1. For each of the N genes in the gene expression matrix, calculate the t-statistic for each tissue by performing an individual-gene analysis:

Equation 1:

$$t_i = \frac{M_1(i) - M_2(i)}{s(i)}$$

Where $M_1(i)$ is the median expression level of gene i in an individual tissue, $M_2(i)$ is the median expression level of gene i across all tissues, and $s(i)$ is a pooled standard deviation over the two groups (individual tissue vs median of all tissues).

2. Filter out genes which have less than a threshold for fold-change (The value 2 is set as default).
3. Compute the Heterogeneity score (HS) corresponding to pathway/set S :

$$HS = \sum_{i \in S} |t_i|$$

4. Permute the labels of the phenotype P in the data matrix and repeat (1) and (2). Repeat until all permutations are considered.
5. Compute empirical p-value for the association of S and P as the fraction of the HSs from the permuted datasets from (3) that is larger than the observed HS statistic from (2).
6. Repeat the analysis for multiple gene sets and estimating false discovery rates (FDRs) from p-values of individual sets using the q-values¹⁰⁶.

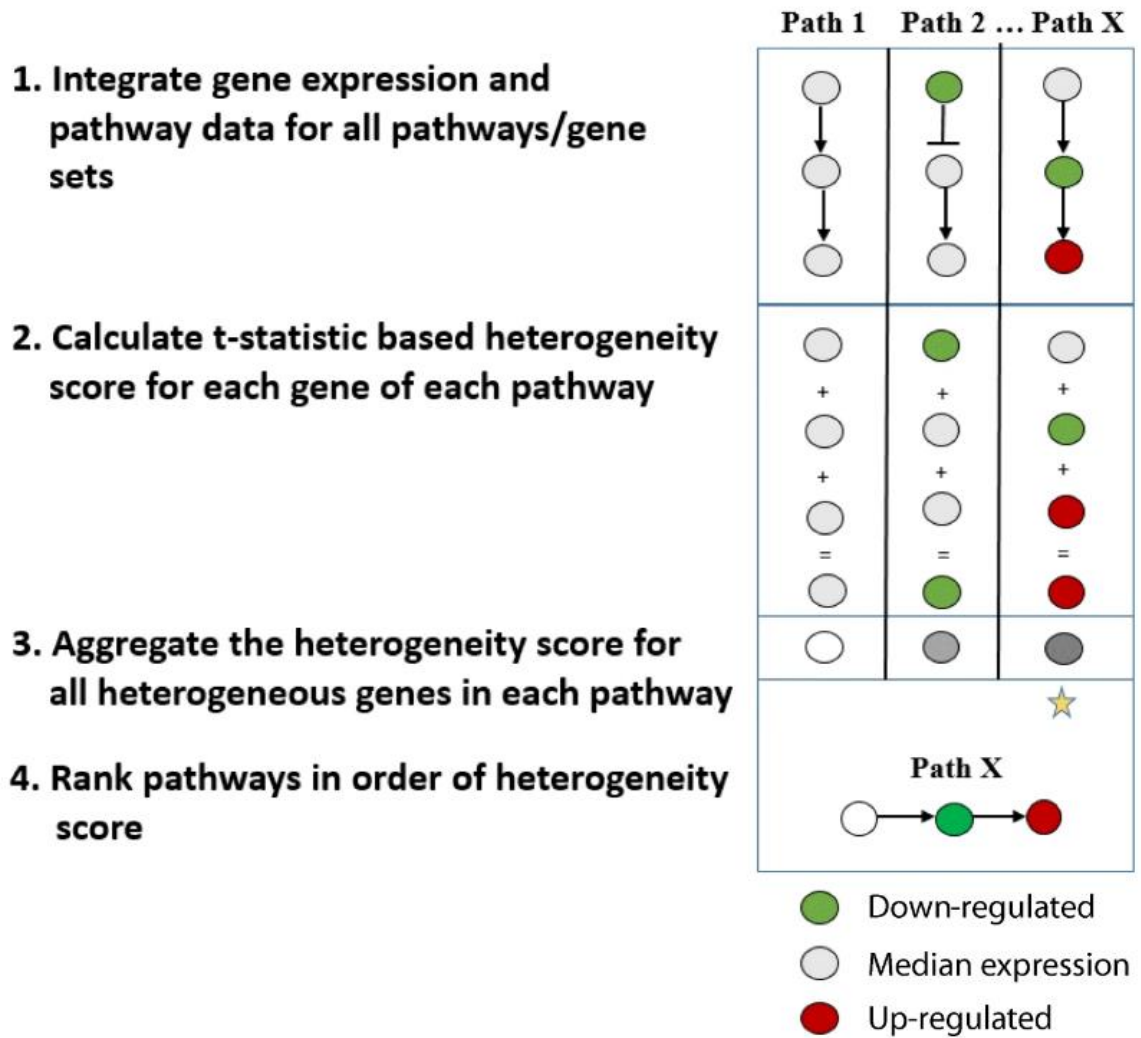


Figure 2: The HeteroPath algorithm for identifying heterogeneous pathways and gene sets. HeteroPath aims to find the pathways/gene sets that are not only differentially expressed from the global median gene expression value but also appear to be responsible for the regulation of distinct cell types.

2.2.3 Methods for Gene Set Analysis

To reveal the biological relevance of the gene expression profiles obtained from HeteroPath, comparison studies were performed using GSEA⁹⁷ and parametric analysis of gene set enrichment (PGSEA)¹⁰⁷. The GSEA algorithm tests whether the distribution of the ranks of genes in the gene set differs from a normal distribution using a weighted Kolmogorov-Smirnov test. PGSEA is an algorithm used to analyze a gene expression data set for enrichment in gene sets, often by testing whether the average fold-change of a gene set is different from zero. Gene enrichment scores for each of the KEGG pathways within each tissue sample were calculated using both GSEA and PGSEA. The GSEA procedure allows for selection of a main parameter. The final output of enriched pathways is affected by the ranking metric which measures the level of difference in gene expression between phenotypes. Therefore, we compared the GSEA results using different ranking parameters and observed a strong overlap among the enriched pathways when a t-test statistic (t-test) or the Pearson correlation coefficient was used for quantitative studies. Since HeteroPath is a t-statistic based algorithm, using the t-statistic quantitative measure for GSEA was more appropriate.

To visualize the degree of heterogeneity identified by different gene set enrichment analysis methods, GSEA, PGSEA, and HeteroPath, we calculated a pathway z-score for significantly differentially expressed pathways identified by the three independent algorithms and generated heatmaps. The Z-score for each pathway was calculated using the PGSEA method. In both microarray data sets,

we first calculated the fold change values for every gene by comparing each of the three tissues individually with the median expression across all tissues. Using those fold-change values we next calculated the mean of the total fold change values (μ) and the standard deviation of the total fold change values (σ). We denoted the mean of fold change values for a given pathway as x_p and the number of genes in a given pathway as p , and then calculated the pathway Z score as

Equation 2

$$Z = (x_p - \mu) * \frac{1}{\sigma \sqrt{p}}$$

By calculating the pathway Z score for all significant pathways, we were able to generate individual heatmaps for each of the algorithms and visualize the degree of heterogeneity identified by the significant pathways from each respective algorithm.

2.2.4 HeteroPath Performance Evaluation

The performance of HeteroPath was evaluated by calculating receiver operating characteristic (ROC) curves and area under the curve (AUC) values for each dataset at varying fold-change thresholds (1.5, 2, and 3) using the R package pROC¹⁰⁸. We first defined a list of bona fide true positive and true negative pathways in each dataset by using the gene expression values to identify pathways which either had a q-value < 0.01 (“positive”) or q-value > 0.2 (“negative”) in GSEA and PGSEA. We then drew a ROC curve for distinct fold change threshold values in the HeteroPath algorithm. Since HeteroPath identifies

significantly differentially expressed genes in each tissue sample by comparing each tissue with the median gene expression of all tissues, varying the fold change threshold influences the number of genes included to calculate the heterogeneity score for each pathway. Using the heterogeneity score, we identified the HeteroPath enriched pathways and compared them to the true positives and true negatives. By calculating the AUCs of the ROC curves based on binary classification of the pathways to the ground truth at three distinct fold change thresholds we were able to evaluate the performance of HeteroPath.

We further evaluated the significance of the AUC values by performing a permutation test. By randomly permuting the class labels and running HeteroPath, we recorded AUC values. We repeated the process 1000 times and recorded all the “random” AUCs. Finally, we compared the observed AUC with the empirical distribution of “random” AUCs from the permutation tests to obtain a p-value which is defined by the fraction of “random” AUCs greater than or equal to the observed AUC value.

2.2.5 HeteroPath Performance Simulations

To further test the validity and performance of HeteroPath we designed a simulation study. The simulation studies were designed using a linear additive model to generate normalized microarray data on m genes and n samples¹⁰⁹. The samples were divided in three groups representing a scenario involving gene set enrichment analysis for three tissues:

Equation 3

$$y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}$$

where $\alpha_i \sim N(\mu = 0, \sigma = 1)$ is a gene-specific effect, such as a probe-effect, with $i = 1, \dots, m$, $\beta_j \sim N(\mu_j, \sigma_j)$ is a sample-effect with $j = 1, 2, 3$ and $\epsilon_{ij} \sim N(\mu = 0, \sigma = 1)$ corresponds to random noise.

To assess statistical power and false positive rate (type-I error), we designed a microarray gene expression data set with $m = 5,000$. Next, we simulated two differently sized differentially expressed gene sets. The first containing 50 genes and the second containing 150 genes. We considered different numbers of samples, $n = 10, 20, 40, 60$, and varying conditions leading to different simulation scenarios for each gene set size. We performed the simulation study varying fractions of differentially expressed genes in the gene set (25%, 50% and 80%) and varying the signal-to-noise ratio (the magnitude of the mean sample effect in differentially expressed genes for one of the sample groups).

In the differentially expressed genes scenario, for β_j , we set $\mu_1 = 0.5$, $\mu_2 = \mu_3 = 0$ for the weak effect; $\mu_1 = 1$, $\mu_2 = \mu_3 = 0$ for the strong effect; and $\sigma_1 = 0.5$, $\sigma_2 = \sigma_3 = 1$ for both cases. For the non-differentially expressed genes case we set $\mu_1 = \mu_2 = \mu_3 = 0$ with $\sigma_1 = \sigma_2 = \sigma_3 = 1$.

We simulated 500 independent data sets using these parameters. For each of the gene set enrichment methods we generated an enrichment score matrix for both gene sets (differentially expressed and non-differentially expressed). We then performed an ANOVA on the score matrix for the two gene sets for a difference in mean between the three groups of samples at a significance level $\alpha = 0.05$. Across the 500 simulations, we estimated the

statistical power as 1 minus the ratio of non-rejections of the differentially expressed gene set and the empirical type-I error as the ratio of rejections of the non-differentially expressed gene set at a significance level $\alpha = 0.05$.

2.2.6 Annotated Transcriptional Regulators of Heterogeneous Genes

To identify sets of unique transcription factors associated with heterogeneous pathways, we searched for transcription factors that have been experimentally proven by ChIP-seq to bind annotated motifs from the ENCODE project¹¹⁰ in the promoter regions of the heterogeneous genes. Our goal was to identify transcription factors motifs that were statistically over-represented in the set of DNA promoter sequences of the heterogeneous genes from a single pathway.

The method requires three steps. First, we extracted 2.5 kb upstream of the transcription start site for each of the heterogeneous genes and examined enrichment for transcription factor binding sites (TFBSs) based on the TRANSFAC¹¹¹ and JASPAR¹¹² databases. We then derived a significance score which is a comparison of the enriched motif found in the set of upstream sequences as compared to a randomly selected set of sequences. In order to calculate significance, we first needed to identify the probability distributions of TFBS for a single transcription factor between the heterogeneous gene set and the randomly selected genes in the mouse genome. Then, we derived a p-value for the number of TFBS using the randomly selected background sequence set, which explains the probability of obtaining the number of TFBS observed merely by chance. Low p-values ($p < 0.05$) suggested that the motif was significantly over-represented. By identifying enriched motifs within each of the

heterogeneous pathways, we were able to infer candidate transcription factors that were associated with tissue-specific endothelial cell signaling pathways.

2.2.7 Putative Regulatory Transcription Factors

We then predicted which transcription factors regulate the heterogeneous gene expression between distinct cell types by identifying transcription factors predicted to bind the overrepresented motifs. This was performed by scanning the promoter regions of the heterogeneous genes and assessing the propensity of a transcription factor to bind a given sequence based on the PWM scores obtained from TRANSFAC¹¹¹ and JASPAR^{112,113}. We performed this analysis using the MATCH algorithm¹¹⁴. This search algorithm uses a matrix similarity score (MSS) and a core similarity score (CSS) to measure the quality of the match between the PWM score and the sequence. This score ranges between 0.0–1.0. If the score is above 0.7, we consider these as putative regulatory transcription factors.

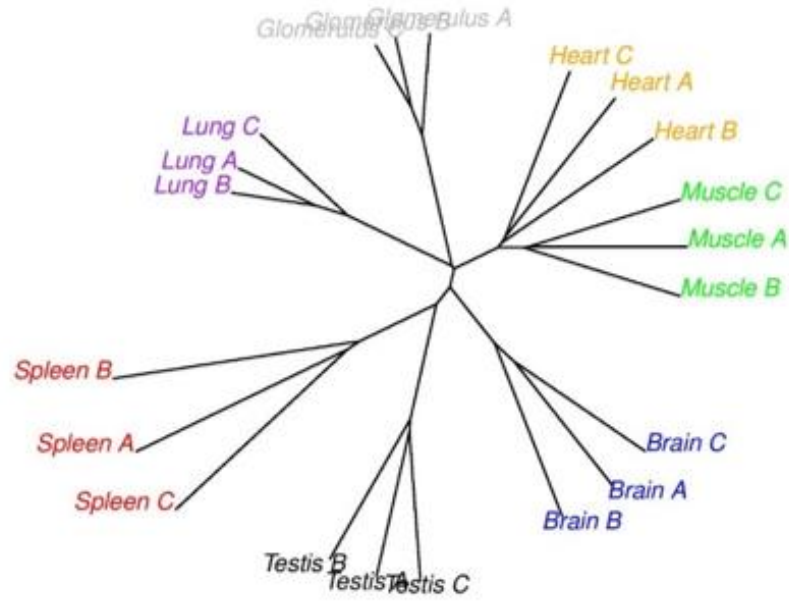
The MATCH algorithm has tunable cutoffs that allow for minimizing the false negative rate (minFN), minimizing the false positive rate (minFP), and minimizing the sum of both errors (minSum). We utilized the minSum cutoff which computes a sum of both false positive and false negative rates to find cutoffs that give an optimal number of false positives and false negatives. The number of matches found in the exon sequences for each matrix is computed using minFN cutoffs which define 100% of false positives. The sum of percentages for false positives and false negatives is then computed for every

cut-off ranging from minFN to minFP. The minimum sum cut-off is then defined as the minSum cut-off.

2.2.8 Constructing Gene Regulatory Networks

The known and putative gene regulatory networks were reconstructed in R using the RTN package¹¹⁵ for visualization. This computational framework establishes interactions and structure of the network by mapping the interactions between upregulated transcription factors identified through motif enrichment and their respective heterogeneous genes identified by HeteroPath.

A



B

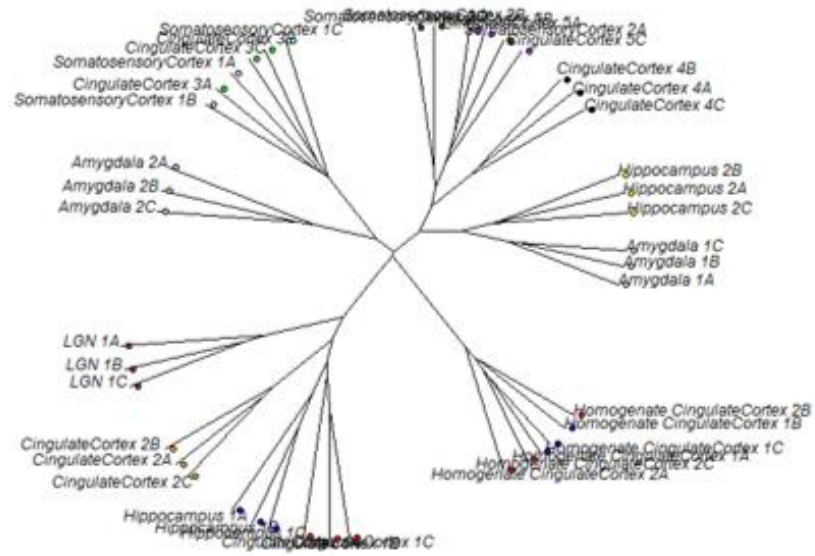


Figure 3: A) Hierarchical clustering of Endothelial cells from 7 mouse organs Intra- and inter-tissue heterogeneity. Tree plot generated via hierarchical clustering of 500 most variable genes across all distinct tissue endothelial cell samples B) Hierarchical clustering of Neuronal cells from 5 different regions of the mouse forebrain Intra- and inter-tissue heterogeneity. Tree plot generated via hierarchical clustering of 500 most variable genes across all distinct tissue neuronal cell samples.

2.2.9 Pairwise Differential Gene Expression

In a separate analysis, prior to applying our novel method of identifying heterogeneous pathways, we first performed hierarchical clustering and generated tree plots shown in Figure 3 to evaluate the differential expression across the studied datasets. Further, to assess the level of heterogeneity among the tissues being analyzed we identified individual differentially expressed genes using the Linear Models for Microarray and RNA-Seq Data (limma) module contained in the R/Bioconductor software package. We compared the gene expression among three of the endothelial cell tissues from Figure 3A and among three of the neuronal cell tissues from Figure 3B to identify the heterogeneity of endothelial cells and the heterogeneity of neuronal cells. To address the degree of differential expression, we assigned confidence intervals to the differential expression. The transcripts were ranked based on the degree of differential expression using the fold-change (FC) in expression level metric^{116,117}. These statistics were computed using the biological replicates and the variance between the replicates to assign a probability value that indicates an incorrect classification of a gene as being differentially regulated. These statistical techniques allowed for a robust analysis that iterates through the transcriptomic cohort to identify genes that are differentially expressed. Gene expression differences were assessed in limma with false discovery rate (FDR) correction for multiple testing¹¹⁸. Genes with an adjusted $p \leq 0.05$ and a $FC \geq 2$ were considered significantly differentially expressed. This analysis did not allow us to

sufficiently understand the underlying heterogeneity biology therefore we sought out to elucidate characteristic pathways explaining the heterogeneity.

2.3 Experiments & Results

2.3.1 Datasets

The statistical modeling framework we developed show in Figure 1 was applied to microarray data sets downloaded from the Gene Expression Omnibus (GEO) at <http://www.ncbi.nlm.nih.gov/geo/>. The mouse endothelial cells (GSE47067) were freshly isolated from mouse organs, labeled by antibodies specific to endothelial cells, isolated via flow sorting and immediately processed for RNA extraction, amplification and hybridization⁸³. The mouse forebrain neurons (GSE2882) were fluorescently labelled neurons isolated from five different regions of the forebrain. Datasets of three of the endothelial cell tissues (brain, lung, and heart) and three of the neuronal cell tissues (hippocampus, cingulate cortex, and amygdala) were used in this study.

2.3.2 Identification of Heterogeneously Expressed Tissue-specific Pathways

First, we used a parametric and a non-parametric gene set enrichment analysis, PGSEA¹⁰⁷ and GSEA⁹⁷ respectively, as gene set enrichment methodologies followed by our novel algorithm HeteroPath to analyze organ-specific endothelial and tissue-specific neuronal transcriptomics data. We evaluated three distinct tissues with a well-balanced coverage of three samples per tissue in both datasets. PGSEA identifies differentially expressed gene sets by testing whether the average expression of genes in a gene set deviates from the overall

expression of all genes in the sample. GSEA aims to test the up- or downregulation of gene sets by testing the expression levels of individual genes. In this type of analysis, no threshold is set to select for significantly differentially expressed genes, but rather all genes are used to determine the differential expression of the pathway. Furthermore, GSEA makes the assumption that the more differentially expressed a gene is, the more biological relevance it has. We implemented our novel algorithm HeteroPath which assigns a heterogeneity score to each pathway based on how distinct its elements are across all tissues versus a “virtual median cell type”, i.e. the virtual endothelial cell that represents the median of endothelial cells from all tissues. As seen in Figure 2, a higher heterogeneity score is attributed to the pathways containing the most heterogeneously expressed genes when comparing distinct tissues independent of cumulative upregulation or downregulation. For example, if three elements of a pathway are similarly upregulated and three elements of pathway are also similarly downregulated in a given tissue when compared to the gene expression of the virtual median cell, HeteroPath would rank this as a highly heterogeneous pathway while PGSEA and GSEA would not consider this as a significant pathway.

2.3.3 HeteroPath Benchmarking

Two comprehensive studies were conducted to assess the performance of our novel pathway-based algorithm to detect tissue-specific gene regulatory networks. In the first study, we analyzed endothelial cells from three out of nine mouse organs. Each sample consisted of 28,815 probes that were mapped to

the 186 KEGG pathways. All of the KEGG pathways were evaluated in the analysis to simulate the differential expression change for all annotated biological processes. More specifically, the fold change in differential expression was drawn from a normal distribution with the mean set at 1.5, 2, 3 and the standard deviation at 0.5.

In the second study with neuronal cell populations, all of the parameters were the same, with the only difference being that there were 22,690 probes representing the genome. The performance of the HeteroPath algorithm was evaluated by calculating the receiver operating characteristic (ROC) and area under the curve (AUC) values for each dataset using the R package pROC.

In the endothelial cell heterogeneity study, the results showed that all three algorithms identified significantly enriched gene sets to distinguish the three endothelial cell populations. Furthermore, HeteroPath identified the least number of significant sets while PGSEA identified the largest number of significant sets (Figure 4A). Of the significant sets identified, only 20% of the significant sets identified were unique to the GSEA algorithm while 25 and 29% were unique to HeteroPath and PGSEA, respectively. PGSEA demonstrated a less stringent functional class scoring technique with significantly higher enrichment scores and more significant p-values (Fig. 2a). In the study of neuronal heterogeneity (Figure 4B), HeteroPath obtained the highest enrichment score, most significant p-values, and highest percentage of unique significant sets (55%). These results suggest that HeteroPath performs more optimally when the heterogeneity of pathways is not unidirectional but includes upregulated and downregulated

genes when compared to the virtual median cell type and may thus reflect a tight regulation of pathways.

The AUC values ranging from 0.7856 to 0.8633 in the endothelial study (Figure 4C) and 0.7192 to 0.9549 in the neuronal study (Figure 4D) indicate that the power of the HeteroPath algorithm increased as the total proportion of genes increased and the fold change increased. Importantly, the HeteroPath algorithm increased in power significantly in the neuronal dataset because it contained a higher number of differentially expressed genes.

a

Pathway Analysis Methods	Top 10 pathways (enrichment score)	Top 10 p-values	Number of Significant Sets	Unique significant sets (%)
HeteroPath	2.19	4.3E-4	56	25%
GSEA	1.53	6.8E-2	81	20%
PGSEA	10.90	8.8E-28	96	29%

b

Pathway Analysis Methods	Top 10 pathways (enrichment score)	Top 10 p-values	Number of Significant Sets	Unique significant sets (%)
HeteroPath	6.72	3.3E-7	39	55%
GSEA	1.07	4.2E-2	48	46%
PGSEA	1.33	9.8E-3	51	45%

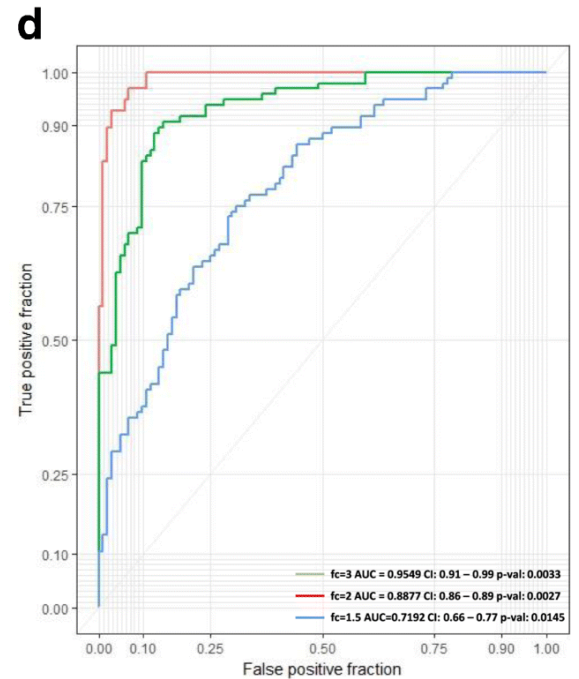
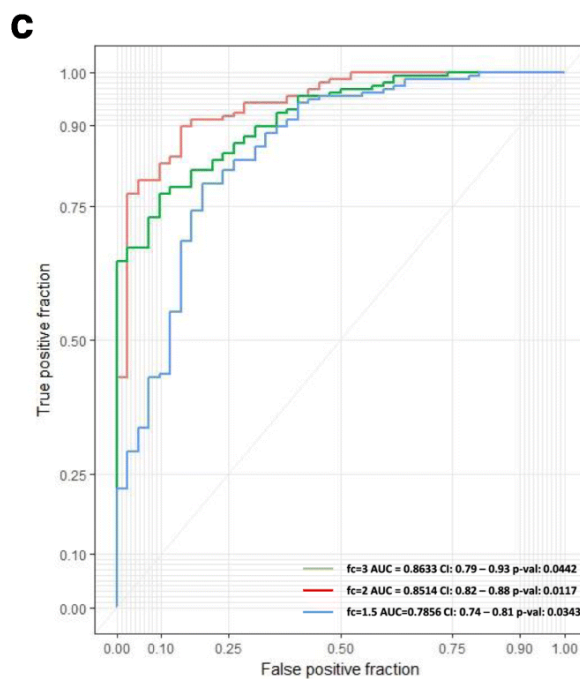


Figure 4: Comparison of pathways enriched using HeteroPath, GSEA, and PGSEA. A) The significantly enriched experimental sets and canonical pathways in mouse endothelial cells were inferred by HeteroPath, GSEA, and PGSEA. Top 10 enrichment scores, p-values, numbers of significant gene sets, and percentage of unique gene sets are shown. B) The significantly enriched experimental sets and canonical pathways in mouse neurons were inferred by HeteroPath, GSEA, and PGSEA. Top 10 enrichment scores, p-values, numbers of significant gene sets, and percentage of unique gene sets are shown. C) ROC curves for the HeteroPath algorithm using the endothelial cell dataset. fc = fold-change; AUC= area under curve. D) ROC curves for the HeteroPath algorithm using the neurons dataset fc = fold-change; AUC= area under curve.

2.3.4 Comparison of HeteroPath, GSEA, and PGSEA using Simulated data

HeteroPath is a pathway-based algorithm which yields tissue-specific enrichment scores. Therefore, we evaluated the statistical power and type I error of HeteroPath, PGSEA and GSEA using simulated data. We simulated microarray data using a linear additive model with sample and probe effects for 5000 genes and three groups of samples (see Methods for details). Using simulated data for each scenario, we calculated the pathway enrichment scores using HeteroPath, PGSEA, and GSEA. For the differentially expressed gene set, we estimated the statistical power for each method as a function of the sample size. At the same time, for the non-differentially expressed gene set, we estimated the empirical type-I error rate. The results of this simulation (Figure 5 and Figure 6) illustrate that HeteroPath performs with comparable statistical power while maintaining similar control of the type-I error rate when compared to GSEA and PGSEA.

Gene Set size: 50 genes, Weak and Strong signal-to-noise ratios, Fraction of DEGs: 25%, 50%, and 80%

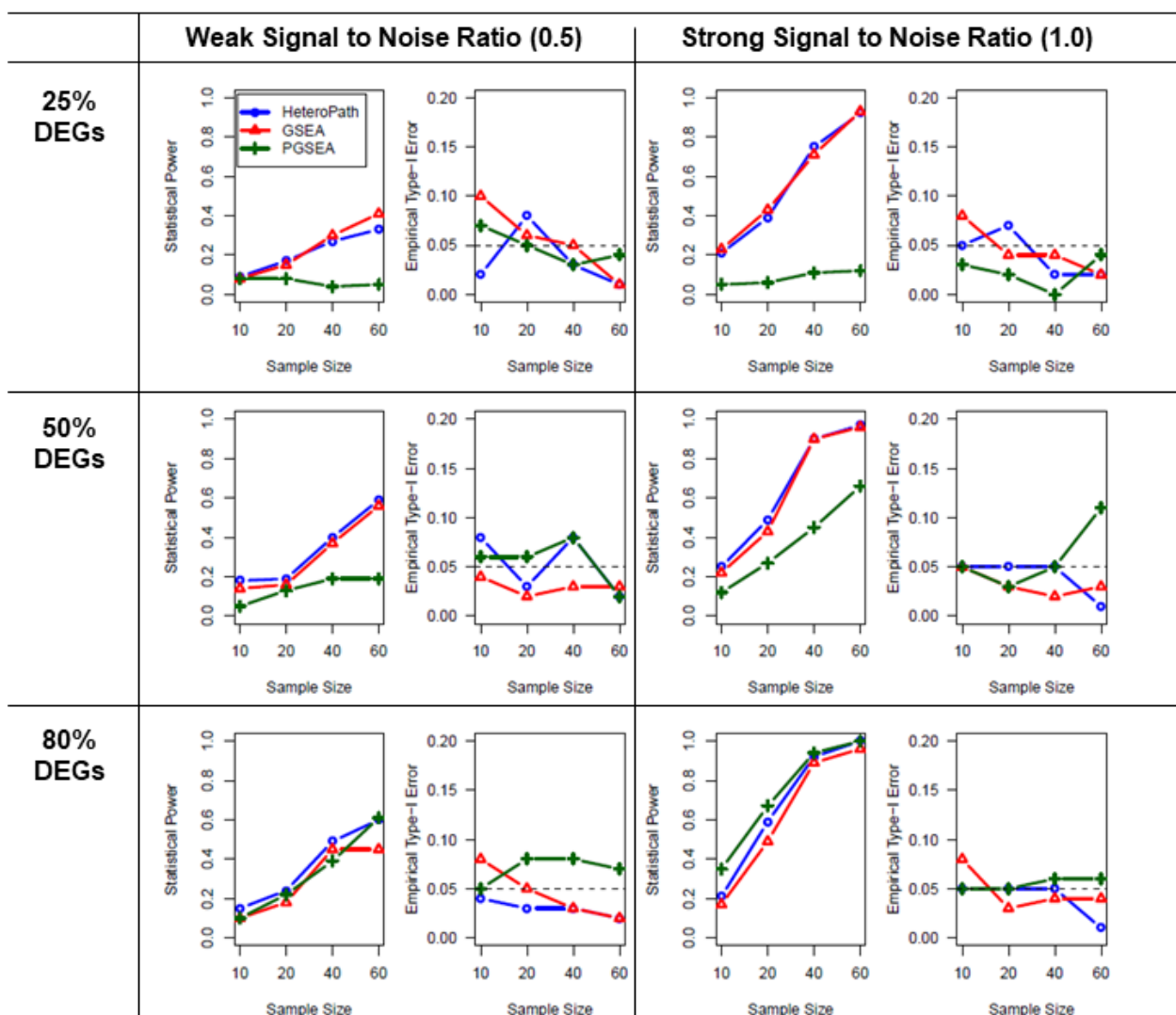


Figure 5: Comparison of statistical power and type-I error rate between HeteroPath, GSEA, and PGSEA for DE Gene Set size of 50 genes. The averaged results of 500 simulations are depicted as function of the sample size on the x-axis, for each of the methods. On the y-axis either the statistical power or the empirical type-I error rate is shown. GSE scores were calculated with each method with respect to two gene sets, one of them differentially expressed (DE) and the other one not. Statistical power and empirical type-I error rates were estimated by performing an ANOVA on the DE and non-DE gene sets, respectively, at a significance level of $\alpha = 0.05$.

Gene Set size: 150 genes, Weak and Strong signal-to-noise ratios, Fraction of DEGs: 25%, 50%, and 80%

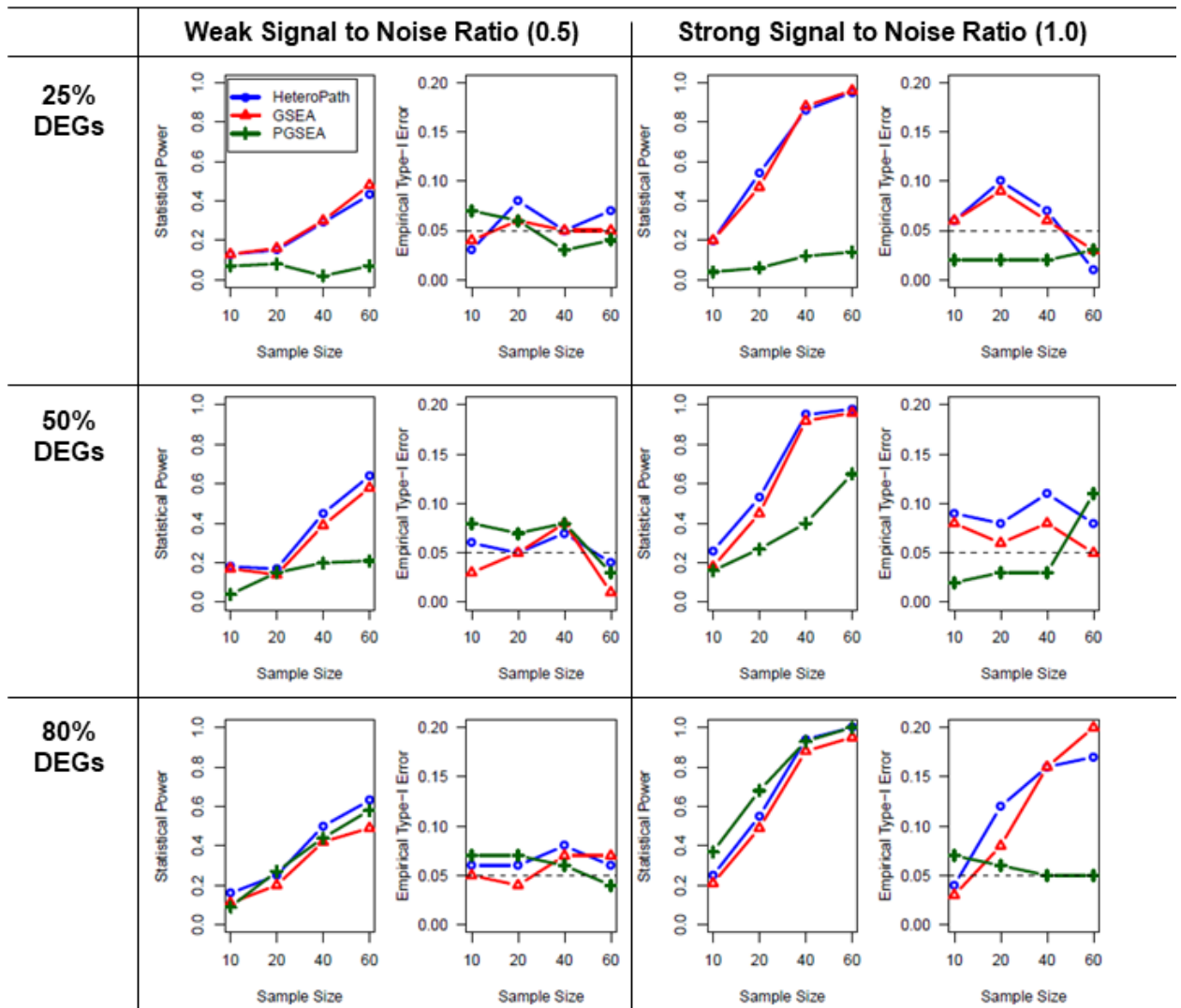


Figure 6: Comparison of statistical power and type-I error rate between HeteroPath, GSEA, and PGSEA for DE Gene Set size of 150 genes. The averaged results of 500 simulations are depicted as function of the sample size on the x-axis, for each of the methods. On the y-axis either the statistical power or the empirical type-I error rate is shown. GSE scores were calculated with each method with respect to two gene sets, one of them differentially expressed (DE) and the other one not. Statistical power and empirical type-I error rates were estimated by performing an ANOVA on the DE and non-DE gene sets, respectively, at a significance level of $\alpha = 0.05$.

2.3.5 Case Study 1: Endothelial Heterogeneity

In order to assess the degree of organ-specific endothelial heterogeneity we applied the three distinct functional class scoring algorithms to freshly isolated mouse endothelial cells from several organs. The results obtained from the HeteroPath algorithm display the most heterogeneous pathways in endothelial cells for three of the nine vascular beds studied. The heterogeneous pathways were ranked in the order of largest to smallest heterogeneity score where only top statistically significant pathways are shown (Figure 7A). The most prominent upregulated pathways identified using HeteroPath were the “Wnt signaling” and “adherens junction” pathways in brain endothelial cells; “focal adhesion”, “PPAR signaling”, “PI3K-Akt signaling” pathways in lung endothelial cells; and “cardiac muscle contraction” and “cytokine-cytokine receptor interactions” pathways in heart endothelial cells (Figure 7A). The HeteroPath algorithm assigned tissue specificity to the heterogeneous pathways when 60% of the heterogeneous elements of the pathway have unique expression in a specific organ (Figure 7A). GSEA analysis primarily identified molecular pathways involved in regulating global molecular function such as RNA and protein synthesis, processing, and degradation (Figure 7B) which were statistically significant but often represented minimal or moderate gene expression changes in terms of magnitude when compared to the virtual median endothelial cell. PGSEA analysis (Figure 7C) also identified signature pathways that were greater in magnitude than GSEA, but the pathways were distinct from those identified by the HeteroPath algorithm. For example, PGSEA revealed amino acid metabolism pathways as being

differentially expressed in brain endothelial cells. This finding likely reflects the importance of individual metabolic enzymes required in brain endothelial cells¹¹⁹.

Of the 186 KEGG gene sets assessed, PGSEA detected 96 gene sets as statistically significant at $p < 0.05$ in this data set whereas GSEA detected 81 gene set at this significance (Fig. 3d). The p-values obtained by PGSEA were generally smaller than p-values of corresponding gene sets obtained by GSEA. These methods specifically target pathways cumulatively regulated in a single direction, but do not consider that tissue-specific heterogeneity which may involve both heterogeneous upregulation and downregulation of elements within a single pathway unlike HeteroPath which ranks overall heterogeneity of a pathway by assessing the cumulative gene expression distance from that of the “virtual median endothelial cell” for each gene within a pathway.

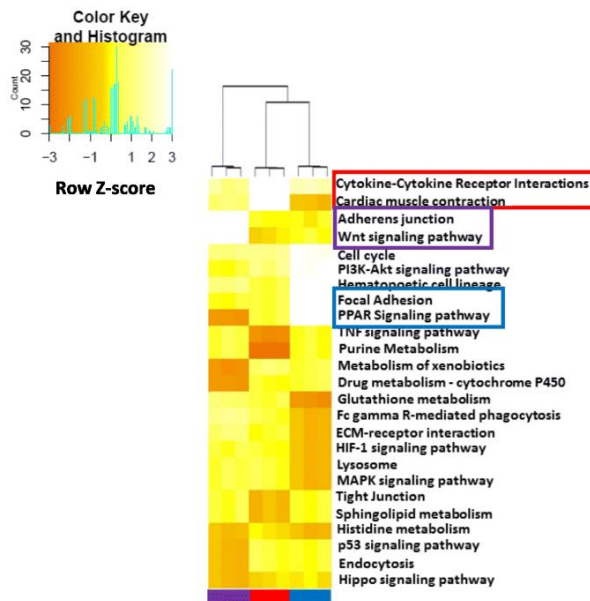
We performed comparative analysis to determine the number of significant sets which were exclusive to a particular algorithm (Figure 7D). For example, HeteroPath uncovered 14 unique pathways. Furthermore, with consistent thresholds applied to the different functional class scoring techniques, HeteroPath identified the least number of significant sets (56), while GSEA and PGSEA identified 81 and 96 significant sets, respectively (Figure 7D).

Based on the design of the HeteroPath algorithm, each heterogeneous pathway reflected the simultaneous upregulation and downregulation of several member genes within a pathway in each tissue. To visualize the role of each significant heterogeneous element within one of the brain endothelium-specific pathways and one of the hippocampal neuronal pathways, we generated

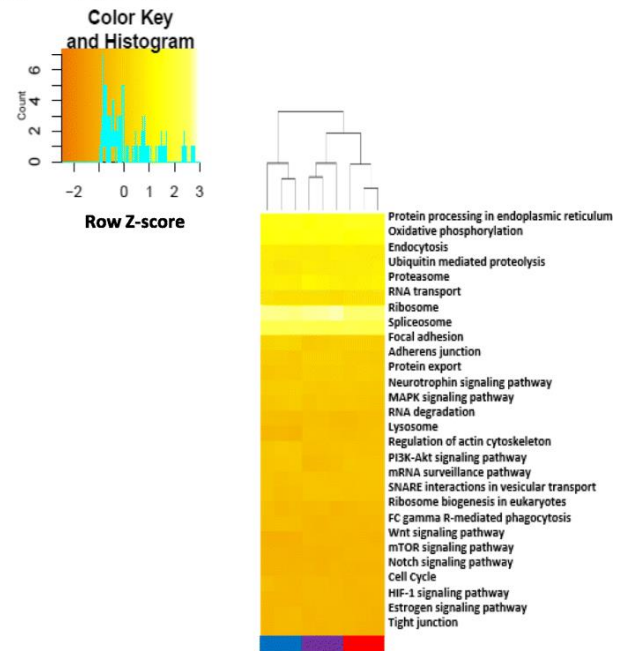
respective gene expression heat maps for Wnt signaling and oxidative phosphorylation.

Using the heterogeneous elements in the Wnt signaling pathway (Figure 8A), we examined the role of putative transcription factors responsible for the brain endothelial cell specific gene expression signature by identifying transcription factors which have been experimentally proven as identified by the ENCODE database¹¹¹ to bind motifs in the promoter regions of the heterogeneously expressed genes (Figure 8B, Figure 8C). For the Wnt signaling pathway, lymphoid enhancer-binding factor 1 (LEF1) and friend leukemia integration 1 (FLI1) were the top candidate transcription factors (Figure 8C). The Wnt signaling gene regulatory network (Figure 8B) contains upregulated genes in brain endothelial cells such as LEF1, Wnt family member 5A (WNT5A), transforming growth factor beta receptor 2 (TGFB2), and Axin-related protein (AXIN2) as well as downregulated genes such as cyclin D1 (CCND1) and cyclin D2 (CCND2) (Figure 8A).

a HeteroPath

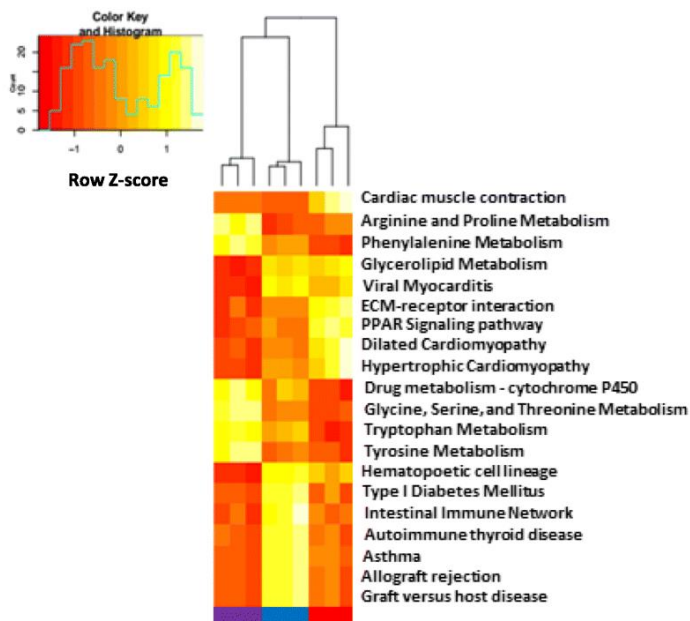


b GSEA



Brain ECs Heart ECs Lung ECs

c PGSEA



d

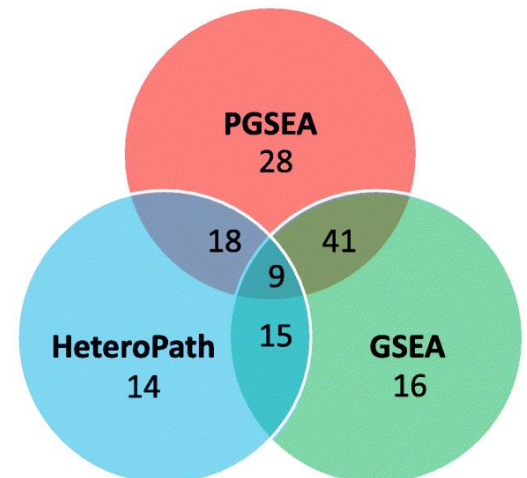
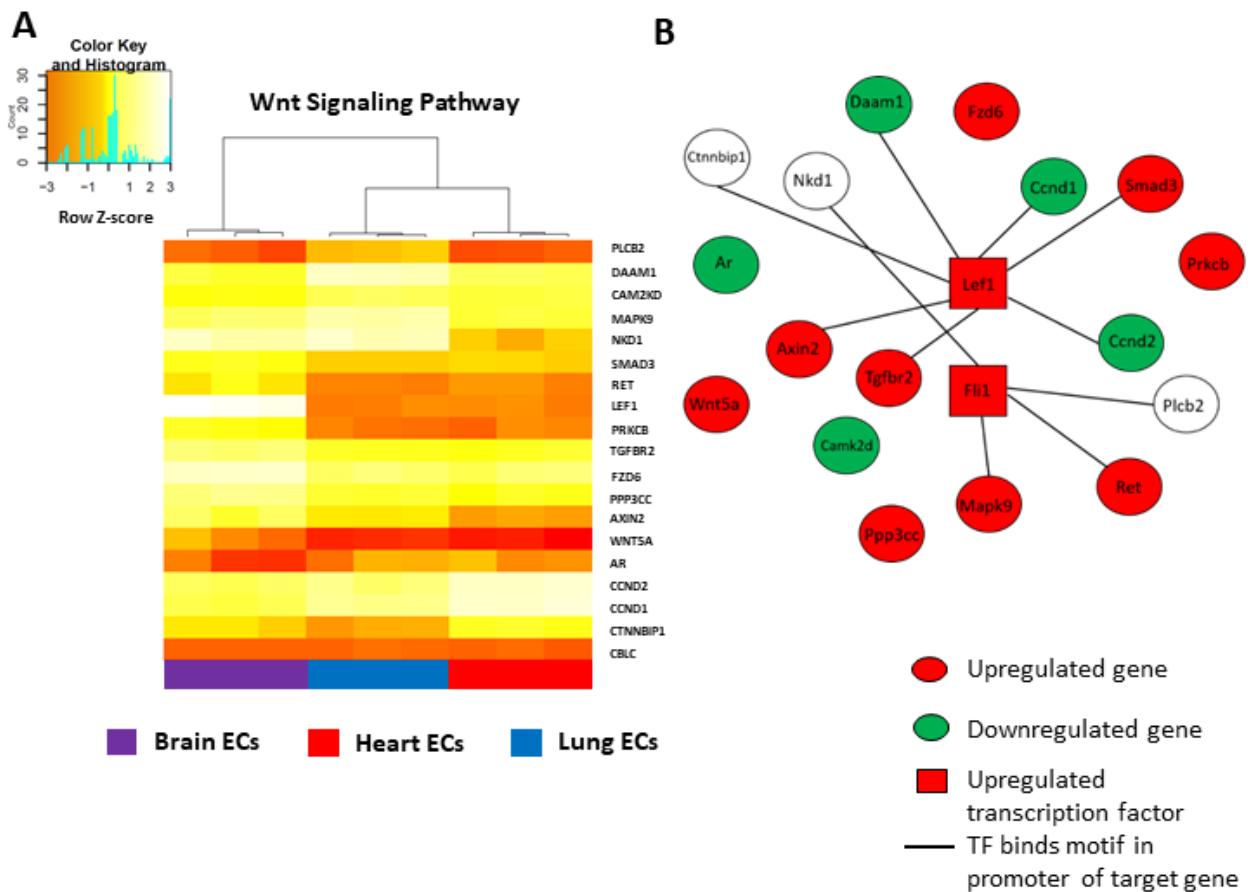


Figure 7: Endothelial cell heterogeneity. A) Heat map of heterogeneous pathways identified by HeteroPath from Brain, Lung, and Heart endothelial cells. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. Upregulated tissue-specific pathways are highlighted in colored boxes. B,C) The results of enriched PGSEA and GSEA pathways from Brain, Lung, and Heart endothelial cells. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. D) A Venn diagram displaying the number of overlapping and unique KEGG pathways identified by HeteroPath, PGSEA, and GSEA.



C

Enriched Motifs in Wnt Signaling Heterogeneous Genes				
Motif Consensus Sequence	Hits	Sig Value	Upregulated TFs which bind motif	P-Value
CTTTGA	193	69.9	LEF1	1.81e-03
AAGAAGARR	27	65.7		
NRYTTCCGGH	111	49.4	FLI1	3.87e-03
CTTTGT	54	48.6	LEF1	2.93e-03
CGCNNNCCC	68	44.5		

Figure 8: Endothelial gene regulatory networks generated from HeteroPath tissue-specific pathways. A) The heat map shows the normalized mRNA expression level in Brain, Lung, and Heart endothelial cells for the heterogeneous genes of the Wnt signaling pathway. B) Wnt signaling gene regulatory network including upregulated transcription factors which bind motifs

in the promoter region of brain-specific heterogeneous elements. C) Enriched Wnt Signaling Motifs from Brain endothelial cells. The table shows the five most enriched motifs in ChIP-seq peaks and the associated transcription factors. Significance values and significant p-values ($p \leq 0.05$) are shown.

2.3.6 Case Study 2: Neuronal Heterogeneity

To assess the relative fidelity of neuronal heterogeneity we applied HeteroPath, GSEA, and PGSEA to neurons isolated from 12 regions of the mouse forebrain¹²⁰. To perform a comparison between the three algorithms, we focused on three distinct regions namely the hippocampus, the cingulate cortex, and the amygdala. The results showed large statistical differences between the three independent algorithms which emphasizes the fundamental molecular difference between neurons at their basal state in various regions of the brain. The HeteroPath algorithm identified the most distinct tissue-specific pathways among the three neuronal populations. For instance, hippocampal neurons exhibited an upregulation of “oxidative phosphorylation” and “GABAergic synapse”; “Hedgehog signaling” and “regulation of autophagy” were upregulated in cingulate cortex neurons; while “taste transduction” and “ribosome” were upregulated in amygdala neurons (Figure 9A). In the case of GSEA analysis, the subsets of neurons in distinct regions of the brain exhibited similar molecular signatures (Figure 9B). Across entire gene sets, there were no tissue-specific pathways. In fact, the algorithm was unable to cluster the neuronal populations into distinct groups. PGSEA, on the other hand, was able to differentiate the three different populations of neurons and identify pathways which contained several upregulated genes in a single tissue (Figure 9C). PGSEA primarily identified pathways in which there was a significant upregulation in one of the tissues relative to the downregulation in the other tissues. For instance, the entire fatty acid metabolism pathway was downregulated in the cingulate cortex and

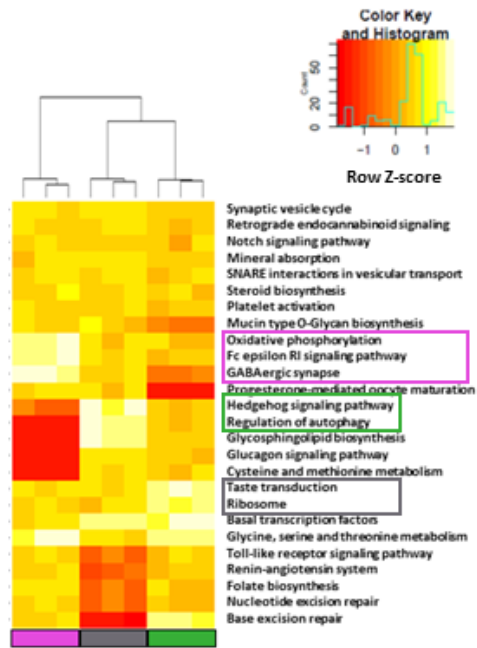
amygdala neuron populations and hence upregulated in the hippocampal neurons (Figure 9C).

In the analysis of neuronal cellular heterogeneity, the three independent algorithms showed that identifying tissue-specific pathways requires prioritizing the up- and downregulation of individual genes within a single pathway. PGSEA and GSEA detected similar numbers of significantly differentially regulated gene sets while HeteroPath detected the least number of differentially expressed pathways (Figure 9D), but these pathways segregated the neuronal populations most distinctively and thus elucidate pathways descriptive of each neuronal subpopulation. In addition, more than half of the significant sets identified by GSEA and PGSEA overlapped while HeteroPath detected 17 unique pathways which likely contributed to the distinctive clustering of the neuronal subpopulations (Figure 9D).

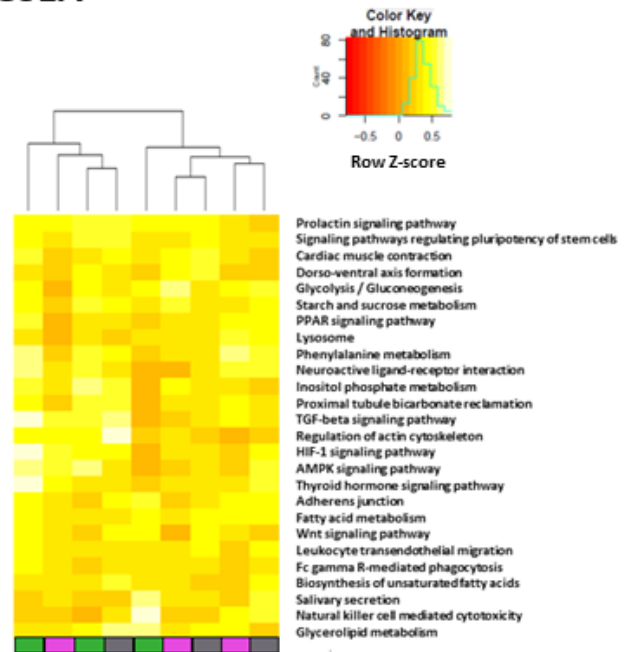
By examining the tissue-specific neuronal pathways, we identified oxidative phosphorylation as a key upregulated pathway in the hippocampal neurons. Analysis of the heterogeneous elements in the oxidative phosphorylation pathway demonstrated that the high heterogeneity score was driven by the significant upregulation of cytochrome c oxidase family genes as well as mitochondrial ATP synthase genes in hippocampal neurons (Figure 10A). Using analogous methods to uncover regulatory transcription factors, we identified three central transcription factors which may drive the upregulation of oxidative phosphorylation in hippocampal neurons: cAMP response element binding protein (CREB), serum response factor (SRF), and Dimethyladenosine

transferase 1, mitochondrial (TFB1M) (Figure 10B, Figure 10C). From these results, we generated a hippocampal neuron specific gene regulatory network which included the regulatory transcription factors and the oxidative phosphorylation heterogeneous genes (Figure 10B).

A HeteroPath

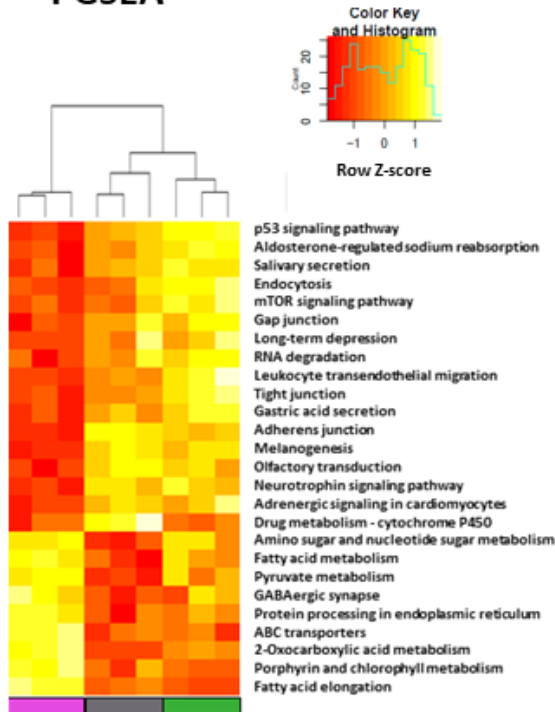


B GSEA



Amygdala Neurons Hippocampal Neurons Cingulate Cortex Neurons

C PGSEA



D

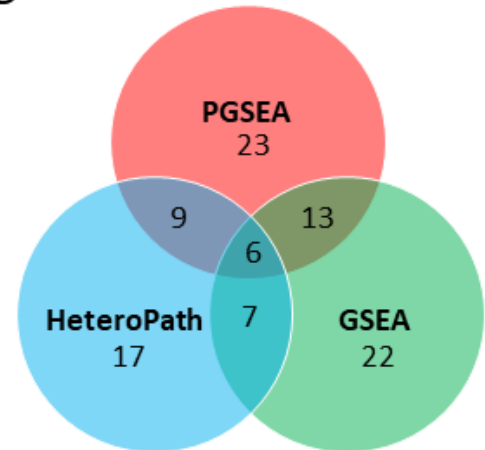
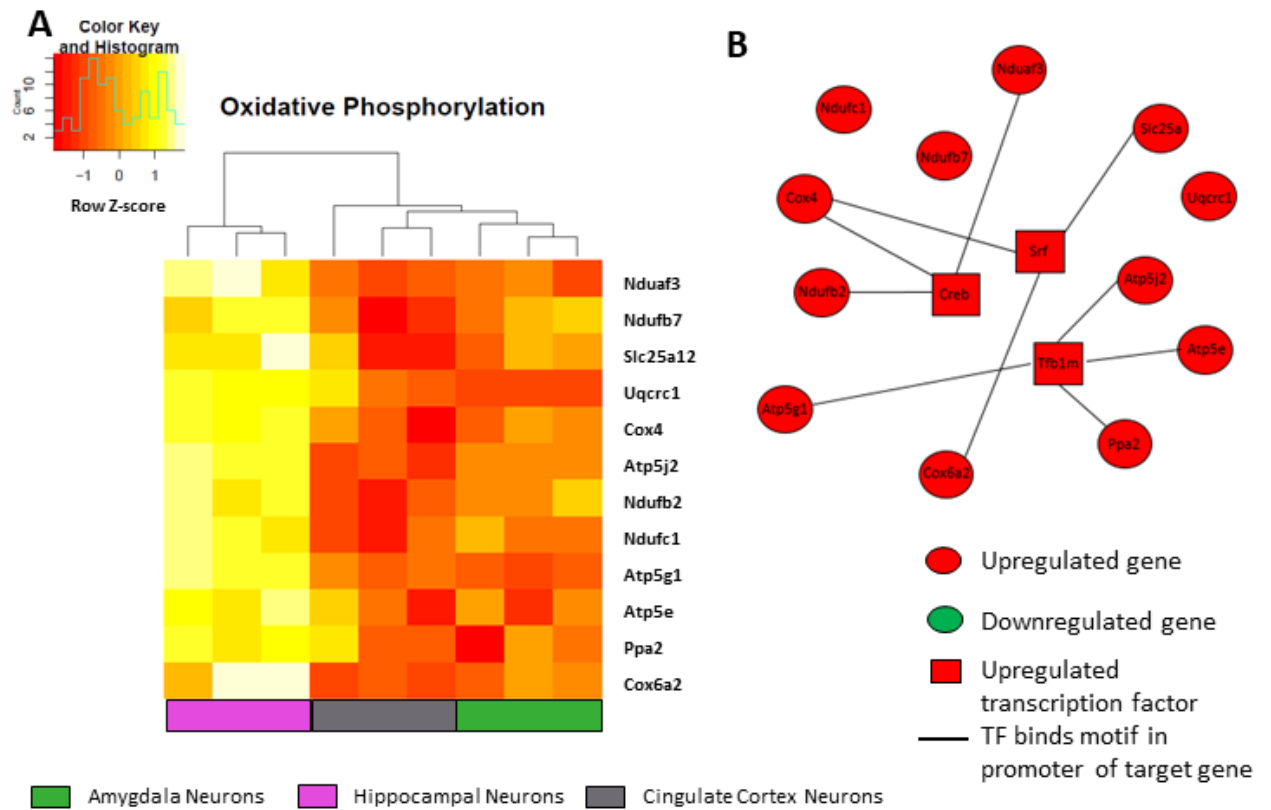


Figure 9: Neuronal heterogeneity A) Heat map representation of heterogeneous

pathways identified by HeteroPath from hippocampal, cingulate cortex, and amygdala neurons. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. Upregulated tissue-specific pathways are highlighted in colored boxes. B,C) The results of enriched PGSEA and GSEA pathways from hippocampal, cingulate cortex, and amygdala neurons. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. D) A Venn diagram displaying the number of overlapping and unique KEGG pathways identified by HeteroPath, PGSEA, and GSEA.



C

Enriched Motifs in Oxidative Phosphorylation Heterogeneous Genes				
Motif Consensus Sequence	Hits	Sig Value	Upregulated TFs which bind motif	P-Value
BCSGVB	1321	867.6		
TGACGTCA	246	243.8	CREB	2.23e-02
TCCTBYBYBYTCCT	48	194.0	SRF	5.36e-03
CGCCG	136	143.3	TFB1M	
CGGAG	124	100.8		

Figure 10: Neuronal gene regulatory networks generated from HeteroPath tissue-specific pathways. A) The heat map shows the normalized mRNA expression level in hippocampal, cingulate cortex, and amygdala neurons for the heterogeneous genes of the oxidative phosphorylation pathway. B) Oxidative phosphorylation gene regulatory network including upregulated transcription factors which bind motifs in the promoter region of hippocampal-specific heterogeneous elements. C) Enriched Oxidative Phosphorylation Motifs from Hippocampal Neurons The table shows the five most enriched motifs in ChIP-seq peaks and the associated transcription factors. Significance values and significant p-values ($p \leq 0.05$) are shown.

2.4 Conclusions

Organ-specific endothelial cells and tissue-specific neuronal cells display remarkable cellular heterogeneity in both their genotypic and phenotypic characteristics⁴¹. Although it is well established that phenotypes of cell populations in different regions are distinct, the unique transcriptomic signatures that define cellular heterogeneity are less clear. Evaluating tissue-specific gene expression is critical for identifying tissue-specific mechanisms of disease ¹²¹. Here we applied HeteroPath to transcriptomic data of mouse endothelial cells from distinct tissues⁸³ and neuronal cells from regions of the mouse forebrain to identify signature gene regulatory networks.

Since endothelial cells are extraordinarily plastic and are known to change their phenotype in culture¹²², the data from these freshly isolated endothelial cell populations was particularly relevant for identifying tissue-specific endothelial pathways⁸³. It is also known that regionalization in the brain is directed by molecular gradients during development¹²³. The degree to which this regionalization causes neurons to express genes heterogeneously was previously unknown. In our work, we observed that the gene expression distance between hippocampal, cingulate cortex, and amygdala neurons isolated from mouse forebrain significantly surpassed the range of the expression distances between replicate experiments, thus indicating gene expression profile specificity for each of the studied brain regions. One of the major challenges in neurobiology has been to create signature gene regulatory networks which are

associated with cell type specific phenotypes such as morphology, firing patterns, connectivity and synaptic transmission¹²⁰.

To address this challenge in both endothelial cells and neurons, we developed the computational algorithm HeteroPath which performs a contextual analysis by assigning a higher heterogeneity score if multiple elements are heterogeneous within a single pathway. Furthermore, this computational model suggests experimentally testable predictions for understanding the general architecture of the gene regulatory networks that establish how basal cellular identity is maintained¹²⁴.

In this study, our objective was to design an algorithm, which first identified heterogeneously expressed pathways in cell populations of unique organs or tissues. The key principle in our analysis was that we determined a pathway heterogeneity score which allowed for individual elements of the pathway to be either upregulated or downregulated when compared to the median of all tissues.

In order to show the application of identifying heterogeneous pathways, transcription factors and gene regulatory networks were generated for HeteroPath, GSEA, and PGSEA. When comparing the novel HeteroPath analysis with PGSEA¹⁰⁷, which ranks genes according to their relative expression levels without prior identification of heterogeneous pathways, we found that HeteroPath uniquely identified signature gene regulatory networks for defining tissue specificity. Thus, the HeteroPath approach is well-suited for identifying tissue-specific druggable signaling targets or regulatory signaling pathways

because it particularly identifies tissue-specific regulated pathways. Furthermore, the HeteroPath analysis differs from GSEA because GSEA ranks pathways by cumulative perturbation of genes in a pathway but does not consider the extent of differential expression for each element within the pathway in establishing or maintaining tissue-specific heterogeneity. GSEA primarily identified minimally differentially expressed pathways as tissue specific in some cases and was unable to identify any tissue-specific pathways in other cases. Therefore, GSEA analysis of tissue specificity may be more appropriate for assessment of global cellular quiescence or activity as a function of subtle gene expression changes in distinct tissues.

Traditional over-representation analysis (ORA) methods such as Fisher's exact test treat genes in a gene set or a pathway simply as gene labels with equal importance, and then test the significance of the over-representation of the gene set among a list of interesting genes. In this type of analysis, the magnitude and direction of change are not evaluated and used to identify tissue-specific gene sets. To complement this approach, we designed HeteroPath to calculate a pathway score that factors in the magnitude and direction of change to identify characteristic pathways segregating distinct populations of cells.

A fundamental question in cellular heterogeneity is defining the nature of interactions of cells from different organs or tissues with the underlying parenchymal cells. Recent studies have described an angiocrine mechanism by which the signals from surrounding cell types influence functions of tissue cells such as their growth and differentiation characteristics⁸⁸. It is also likely that

specialized signals from a heterogeneous population of cells influence interactions underlying cells such as in the case of endothelial cells, vascular smooth muscle cells and pericytes¹²⁵. Including gene upregulation and downregulation in the analysis along with the extent of differential expression to define tissue-specific gene expression generates comprehensive tissue-specific signatures as opposed to those obtained by existing gene set enrichment analyses based only on cumulative unidirectional gene regulation. Downregulation of specific genes and pathways is essential for the development of tissues such as during mesodermal differentiation when downregulation of Flk1 followed by a later induction of Flk1 expression is required for the formation of cardiac progenitors¹²⁶. In addition, downregulated genes can act as “valves” which maintain low levels of baseline gene expression and enable upregulation as a response to stressors or stimuli.

In the brain endothelium, we uncovered the Wnt signaling pathway as being significantly heterogeneous when compared to heart or lung endothelium. Analysis of regulatory transcription factors that could maintain the brain EC specific upregulation of the Wnt signaling pathways allowed us to identify the upregulation of Lef1, which is known to interact with β -catenin and regulate brain vascularization as well as differentiation of the BBB in vivo¹²⁷. Additionally, the Wnt-associated beta-catenin/TCF7 transcriptional complex has been shown to regulate vascular remodeling through the regulation of smooth muscle cell proliferation and EC growth¹²⁸⁻¹³⁰. Similarly, the Wnt pathway member and transcription factor FLI1 was also upregulated in brain endothelial cells and is

thought to be among the earliest transcription factors involved in endothelial cell development¹³¹.

In hippocampal neurons, we showed the oxidative phosphorylation biological process to be upregulated in a tissue-specific manner compared to cingulate cortex and amygdala neurons. We predicted that CREB, SRF, and TFB1M are crucial transcription factors driving the upregulation of the oxidative phosphorylation process in a tissue-specific manner. In neurons, CREB is known to be phosphorylated under conditions of hypoxia and oxidative stress which suggests that the CREB activation is a survival program during harmful stimuli and may play a role as a cellular form of defense¹³². In addition, the molecular mechanisms underlying SRF-dependent axon growth have been reported in mouse hippocampal neurons¹³³. Furthermore, the mitochondrial transcription factor TFB1M has been implicated in cellular systems in which its upregulation induces mitochondrial biogenesis¹³⁴.

CHAPTER 3

ENDOTHELIAL TRANSLATOME HETEROGENEITY ACROSS DISTINCT VASCULAR BEDS DURING HOMEOSTASIS

Previously published as:

Jambusaria, A. et al. (2020). "Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation." *eLife* 9. e51413

3.1 Introduction

Endothelial cells (ECs) line blood vessels in all tissues and organs, where they form a barrier which tightly regulates the trafficking of oxygen, metabolites, small molecules and immune cells into the respective tissue⁷⁵. Previous studies have suggested that the morphology of the microvascular endothelium and the expression of selected genes can vary when comparing the vasculature of multiple tissues, thus allowing ECs to take on tissue-specific EC functions^{82,84,135}. Environmental signals from the tissue microenvironment including mechanical forces, metabolism, cell-matrix, cell-cell interactions, organotypic growth factors likely play an important role in regulating this endothelial heterogeneity¹³⁵.

Identifying differences in the expression levels of selected genes in endothelial cells from different tissues or organs provides some insights into the molecular underpinnings of endothelial heterogeneity, however unbiased gene expression profiling is likely to yield a more comprehensive evaluation of the genes and regulatory pathways underlying endothelial heterogeneity. Microarray profiling has been used to identify paracrine factors and signaling pathways that characterize endothelial cells in different organs^{83,136}. Single-cell transcriptomic

analysis of endothelial cells has provided a molecular atlas of the brain and lung endothelial subpopulations at a single cell level¹³⁷. While single cell RNA-sequencing is ideally suited for identifying subpopulations within a single vascular bed, current single cell technologies are limited in their ability to detect the expression of individual genes in a given cell¹³⁸⁻¹⁴². The endothelial signatures defined using these transcriptomic approaches are potentially influenced by disassociation and isolation of endothelial cells, a process affecting cellular mRNA levels when cells are removed from their native niche^{120,143,144}. Furthermore, conventional global mRNA and single cell mRNA transcriptomic profiling does not discriminate between the total mRNA pool and those mRNAs preferentially translated due to translational regulation^{145,146}.

In this work, to understand the variegated nature of the endothelium, we used the RiboTag transgenic mouse model, in which LoxP mice express an HA-tag on the ribosomal Rpl22 protein¹⁴⁷. These mice enable direct isolation of tissue-specific mRNAs undergoing translation without cell disassociation¹⁴⁷. Using an endothelial-specific RiboTag model, we show that organ-specific ECs have distinct translome patterns of gene clusters during homeostasis. We found that ECs express tissue-specific genes involved in vascular barrier function, metabolism, and substrate-specific transport. In addition, we found that ECs expressed genes thought to be primarily expressed in the surrounding tissue parenchyma, suggesting a previously unrecognized organ-specific endothelial plasticity and adaptation. To allow other researchers to explore the organ-specific EC translome heterogeneity, we have generated a searchable

database (<http://www.rehmanlab.org/ribo>), in which users can visualize gene expression levels of individual genes.

3.1.1 Problem Definition

Blood vessels are lined by endothelial cells engaged in distinct organ-specific functions, but little is known about their characteristic gene expression profiles. In this study, we performed RNA-Sequencing of the brain, lung, and heart endothelial transcriptome and identified specific pathways, transporters and cell-surface markers expressed in the endothelium of each organ. We found that endothelial cells express genes typically found in the surrounding tissues such as synaptic vesicle genes in the brain endothelium and cardiac contractile genes in the heart endothelium. Complementary analysis of endothelial single cell RNA-Seq data identified the molecular signatures shared across the endothelial transcriptome and single cell transcriptomes. Our study defines endothelial heterogeneity and plasticity and provides a molecular framework to understand organ-specific vascular disease mechanisms and therapeutic targeting of individual vascular beds.

3.2 Methods

3.2.1 Experimental Animals

RiboTag (*Rpl22^{HA/+}*) mice were purchased from Jackson Labs. Endothelial-specific VE-cadherin-Cre mice were provided by Dr. Ralf Adams. We crossed the RiboTag mice (*Rpl22^{HA/+}*)¹⁴⁷ with the endothelial-specific VE-cadherin-Cre mice^{148,149} to generate RiboTag^{EC} (*Cdh5^{CreERT2/+}; Rpl22^{HA/+}*) mice. Following tamoxifen-induced recombination at week 4, HA-tagged *Rpl22* was specifically

expressed in endothelial cells. To investigate the mechanisms of organ-specific EC injury, repair, and regeneration we performed RNA-Seq analysis of gene expression in ECs isolated at 6 hr, 24 hr, 48 hr, 72 hr, and 1 week post-LPS challenge (10 mg/kg LPS i.p., Sigma-Aldrich Cat#: L2630) with PBS-injected mice serving as controls.

The C57BL/6J mice were purchased from the Jackson Laboratory. All animal experiments were conducted in accordance with NIH guidelines for the Care and Use of Laboratory Animals and were approved by the IACUC of the University of Illinois (IACUC Protocol #19–014, IACUC Protocol #13–175 and IACUC Protocol #16–064).

3.2.2 Isolation of Mouse Brain, Lung, and Heart

After surgically opening the mouse chest, the brain, lung and heart were harvested after a one-time perfusion of 20 mL PBS through the left and right ventricular chamber.

3.2.3 Ribosome Immunoprecipitation (IP)

The tissue samples were extracted from RiboTag^{EC} mice, flash-frozen in liquid nitrogen and then stored at –80°C. The samples were then homogenized on ice in ice-cold homogenization buffer (50 mM Tris, pH7.4, 100 mM KCl, 12 mM MgCl₂, 1% NP-40, 1 mM DTT, 1:100 protease inhibitor (Sigma), 200 units/mL RNasin (Promega) 1 mg/mL heparin and 0.1 mg/mL cycloheximide (Sigma) in RNase free DDW) 10% w/v with a Dounce homogenizer (Sigma) until the suspension was homogeneous. To remove cell debris, 1 mL of the homogenate was transferred to an Eppendorf tube and was centrifuged at 10,000xg and 4°C

for 15 min. Supernatants were subsequently transferred to a fresh Eppendorf tube on ice, then 100 μ L was removed for 'input' analysis and 3 μ L (=3 μ g) of anti-HA antibody (ab91110, Abcam) or 3 μ L (=1 μ g) of mouse monoclonal IgG1 antibody (Sigma, Cat# M5284) or 6 μ L anti-RPL22 (Invitrogen Cat# PA5-68320) was added to the supernatant, followed by 1 hr of incubation with slow rotation in a cold room at 4°C. Meanwhile, Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific), 100 μ L per sample, were equilibrated to homogenization buffer by washing three times. At the end of 1 hr of incubation with antibody, beads were added to each sample, followed by incubation 1 hr in cold room at 4°C. After that, samples were washed three times with high-salt buffer (50 mM Tris, 300 mM KCl, 12 mM MgCl₂, 1% NP-40, 1 mM DTT, 1:200 protease inhibitor, 100 units/mL RNasin and 0.1 mg/mL cycloheximide in RNase free DDW), 5 min per wash in a cold room on a rotator. At the end of the washes, beads were magnetized, and excess buffer was removed, 350 μ L Lysis Buffer was added to the beads and RNA was extracted with RNeasy plus Mini kit (Qiagen). RNA was eluted in 30 μ L H₂O and taken for RNA-Sequencing.

3.2.4 RNA-Sequencing

RNA quality and quantity were assessed using an Agilent Bio-analyzer. RNA-Seq libraries were prepared using Illumina mRNA TruSeq kits as protocolled by Illumina. Library quality and quantity were checked using an Agilent Bio-analyzer and the pool of libraries was sequenced using an Illumina HiSeq4000 and Illumina reagents.

3.2.5 RNA-Sequencing Data Processing and Batch Correction

The sequenced reads from all samples were aligned to the mouse reference genome (mm10) using STAR v. 2.4.2¹⁵⁰. The aligned reads were then processed to calculate relative mRNA expression levels by using HTSeq-count v. 0.6.1¹⁵¹. Gene symbols were mapped to the ENSEMBL features using the biomaRt package v. 2.26.1¹⁵². Preliminary unsupervised analysis of normalized and processed profiles by principal component analysis (PCA) revealed separation into three major clusters. These clusters largely corresponded to the distribution of samples by sequencing batch. Consistent with the PCA plots, the distribution of samples by sequencing batch differed significantly but not by tissue type. To better harmonize profiles prior to analyses reported here, we normalized expression data of all samples using ComBat¹⁵³. This correction ameliorated the separation by sequencing batch without substantially affecting distributions tissue type.

3.2.6 Baseline Tissue-specific Gene Signatures

We calculated the differential expression level of genes using a one versus others approach in order to identify signature genes which were upregulated for each tissue at baseline. For instance, to identify the genes significantly upregulated in brain ECs at baseline, we compared the 0 hr brain EC samples to 0 hr lung ECs and 0 hr heart ECs. We performed these analyses for all three tissues to identify baseline organ-specific EC signatures. We utilized the limma R package and applied the standard limma pipeline¹⁵⁴ to RNA-Seq data after voom transformation¹⁵⁵. For each gene, the log fold-change (logFC) in expression level

is derived from the limma analysis. Genes with $FDR < 0.05$ were identified as being differentially expressed. All upregulated genes for each tissue were plotted using the heatmap.2 function from the gplots v.3.0.1.1 R package. The top 10 significantly differentially expressed genes by logFC were listed.

3.2.7 Baseline Tissue-specific Pathway Analysis

To define the biological function associated with the molecular signature of the tissue-specific ECs, we specifically performed gene set enrichment analysis (GSEA)⁹⁷ on the genes which were significantly upregulated ($\log FC > 1$) in the tissue of interest. GSEA was performed on significantly upregulated genes ranked by their p-value using the clusterProfiler package¹⁵⁶ in R with gene ontology (GO) gene sets downloaded from the Molecular Signatures Database (MSigDB)¹⁵⁷. The top 20 most enriched GO terms were plotted.

3.2.8 Baseline Cell Surface Markers

Tissue-specific cell surface markers were identified by intersecting tissue-specific differentially expressed genes with predicted cell surface markers, as reported in the Cell Surface Protein Atlas (www.proteinatlas.org)¹⁵⁸. The top 10 significantly differentially expressed cell surface proteins by logFC were plotted.

3.2.9 Isolation of Brain ECs

The forebrains of C57BL/6J mice were micro dissected and minced in collagenase/dispase (Roche, Cat#: 11097113001) and DNase (Worthington Biochemical Cat#: LK003170) and incubated for 1 hr at 37°C. Myelin Removal Beads (Miltenyl Biotec, Cat#: 130-096-433) and LS columns (Miltenyl Biotec, Cat#: 130-042-401) were used. The resulting pellet after myelin removal

contained microglia, astrocytes and endothelial cells. The endothelial cells were further enriched by using CD31 microbeads (Miltenyl Biotec, Cat#: 130-097-418

3.2.10 Isolation of Lung ECs

The C57BL/6J mice mouse lungs were minced and digested with 3 mL collagenase A at 1 mg/mL in PBS (Roche, Cat#: 10103586001) at 37°C water bath for 1 hr. Mixtures were titrated with #18 needles and then pipetted through a 40 µm disposable cell strainer. After centrifuging 500xg for 5 min and washing with 1x PBS, the isolated cells were treated with red blood cell lysis buffer (Biolegend, Cat#: 420301) for 5 min. After washing with 1x PBS twice, cells were incubated in suspension buffer (Ca²⁺ and Mg²⁺ free PBS, 0.5% BSA, 4.5 mg/mL D-glucose, and 2 mM EDTA) with 5 µg anti-CD31 antibody (BD Pharmingen, Cat#: 553370) at 4°C for 60 min with gentle tilting and rotation. After washing, cells were then incubated in suspension buffer with pre-washed Dynabeads (20 µL beads in 1 mL buffer, Invitrogen Cat#: 11035) at 4°C for 60 min with gentle tilting and rotation. After washing with 1x PBS three times using magnetic separation, lung ECs were dissociated from magnetic beads with trypsin.

3.2.11 Isolation of Heart ECs

Isolated C57BL/6J mice hearts were minced and digested with prewarmed Collagenase/Dispase mix (1 mg/mL) (Roche) at 37°C for 30 min. 75 µL DNase I per 10 mL cell suspension (1 mg/mL) was added and the suspension was incubated at 37°C for 30 min. The digested tissue was filtered using 70 µm cell strainer followed by RBC lysis in RBC lysis buffer (Biolegend, Cat#: 420301) for 7

min at room temperature. The cell suspension was diluted with 10 mL of MACS buffer (Prepared in phosphate-buffered saline (PBS), pH 7.2, 0.5% bovine serum albumin (BSA), and 2 mM EDTA) by diluting MACS BSA Stock Solution (Cat#: 130-091-376) 1:20 with autoMACS Rinsing Solution (Cat#: 130-091-222)) and cells were passed through 40 μ m cell strainer followed by centrifugation at 500xg for 5 min to pellet the cardiomyocytes. The supernatant containing endothelial cells was centrifuged at 800xg for 5 min to pellet down the ECs. The endothelial cell enriched pellet was resuspended in 500 μ L of MACS buffer and the isolated cells were counted. Endothelial cells were further purified by using CD31 microbeads (Miltenyl Biotec, Cat#: 130-097-418) and Miltenyl Biotec MS columns (Miltenyl Biotec, Cat#: 130-042-201) through affinity chromatography according to the manufacturer's protocol.

3.2.12 Preparation of Cytospin Slides from Brain, Lung, and Heart Endothelial Cells

The Thermo Shandon Cytospin three was used to generate Cytospin slides. Briefly, the Cytoslide with filter card were inserted into a Cytoclip. The Cytoclip was fastened and placed in a recess of the Cytospin rotor after sliding a Cytofunnel into it. The required volume of the cell suspension was pipetted into the Cytofunnel after cell counting and calculation. The Cytospin was centrifuged for 500 rpm for 5 min. The slide was fixed with 4% paraformaldehyde for 10 min and stored in 1x PBS at 4°C.

3.2.13 Immunofluorescence and Confocal Microscopy of Brain, Lung, and Heart Endothelial Cells

The slides were permeabilized and blocked with 10% donkey serum, 2% BSA, 0.05% tween in PBS for 1 hr at room temperature. For lung cells, the slides were incubated with primary antibodies anti-CD31 (BD Pharmingen, Cat#: 550274, 1:25) and anti-RAGE (Abcam, Cat#: Ab3611, 1:3200) at 4°C overnight. The brain ECs were incubated with primary antibodies anti-CD31 (BD Pharmingen, Cat#: 550274, 1:25) and anti-PTN (Santa Cruz Biotechnology, Cat#: sc-74443, 1:3200) at 4°C overnight. For the heart samples, primary antibodies anti-AQP7 (Novus Biologicals, Cat#: NBP1-30862, 1:3200) and anti-CD31 (BD Pharmingen, Cat#: 550274, 1:25) were used and incubated at 4°C overnight. The next day, slides were washed and incubated with the fluorescence-conjugated secondary antibody (AF488 donkey anti-rat 1:300, Invitrogen Cat#: A-21208; AF594 donkey anti-rabbit 1:300, Invitrogen Cat#: A-21207; AF594 goat anti-mouse 1:300, Invitrogen Cat#: A11032), followed by washing with 1x PBS. Cells were stained with DAPI and mounted on ProLong Gold mounting medium (Invitrogen, Cat#: P36934). Images were taken with a confocal microscope LSM880 (Zeiss) and analyzed by Zen software (Zeiss).

3.2.14 Assessing baseline endothelial translome heterogeneity

Tissue-specific baseline gene expression heatmaps were generated for gene sets related to endothelial function including classical endothelial markers, glycolysis, fatty acid metabolism, and solute transport. The individual genes listed

in the heatmaps contain the tissue-specific differentially expressed genes which overlapped with each of the respective gene sets.

The classical endothelial markers gene set contains 152 mouse endothelial cell markers downloaded from PanglaoDB¹⁵⁹. The mouse glycolysis and fatty acid metabolism gene sets containing 67 and 52 genes respectively were downloaded from the Rat Genome Database (RGD) <https://rgd.mcw.edu/>¹⁶⁰. For the transport gene set, the solute carrier family including 423 membrane transport proteins located in the cell membrane were downloaded from the HUGO Gene Nomenclature Committee database (<https://www.genenames.org/>)¹⁶¹.

3.2.15 Computational Assessment of mRNA Purity

Due to the endothelial cells being surrounded by other tissue-resident cell types, it is likely that the mRNA isolated from endothelial-specific RiboTag^{EC} samples could contain non-endothelial mRNA. For this reason, we assessed the mRNA purity of RiboTag endothelial samples isolated from whole tissue by comparing the gene expression levels of the endothelial-specific RiboTag samples to the gene expression levels of mRNA from whole tissue. We compared endothelial-specific RiboTag^{EC} mRNA expression levels from brain, lung, and heart tissue to whole brain, lung, and heart mRNA expression levels.

We first acquired RNA-Seq data for whole brain, whole lung, and whole heart tissue from Array Express (Athar et al., 2019). The three whole brain samples and three whole lung samples were downloaded from accession number E-MTAB-6081, while the three whole heart samples were downloaded

from accession number E-MTAB-6798. Raw mRNA counts were processed, and batch corrected in a cohort including the 0 hr RiboTag brain, lung, and heart endothelial mRNA counts. The preprocessing and batch correction were performed in the same manner as described above.

To identify whether mRNA of tissue-resident cells was isolated during the RiboTag^{EC} mRNA isolation procedure, we calculated a Kendall's Tau rank coefficient between the most abundant genes in the RiboTag^{EC} mRNA and whole tissue mRNA. The Kendall's Tau rank coefficient, ranging between -1 and 1 , allowed us to test whether there was contamination of mRNA from the whole tissue in the RiboTag^{EC} samples. As the coefficient approaches -1 , the rank of most abundant genes differs in both sets of samples; while, as the coefficient approaches 1 , the rank of most abundant genes becomes identical. Using this test, we were able to infer that if the rank of the most abundant genes in the RiboTag^{EC} sample and the whole tissue is identical, there is contamination of non-endothelial mRNA in the RiboTag^{EC} mRNA samples. All samples were compared to each other and heatmaps with Kendall's Tau rank coefficients were generated to visualize the results.

3.2.16 Single-cell Endothelial Transcriptome Heterogeneity

To specifically analyze ECs at the single-cell level, we downloaded Tabula Muris data from <https://github.com/czbiohub/tabula-muris> and Betsholtz Lab data from NCBI Gene Expression Omnibus (GSE99235, GSE98816). We filtered out non-ECs from the Tabula Muris brain, lung, and heart data based on Cd31 and Cdh5 expression. We selected ECs from the Betsholtz Lab brain and lung data based

on Cd31 and Cldn5 expression. All genes that were not detected in at least 10% of all single cells were discarded. For all further analyses we used 2585 cells expressing 6802 genes from the Tabula Muris dataset and 873 cells expressing 8116 genes from the Betsholtz Lab dataset. Data were log transformed for all downstream analyses.

We analyzed the data utilizing the Seurat R package (<https://github.com/satijalab/seurat>; <http://satijalab.org/seurat/>)¹⁶². PCA analysis of organ-specific ECs was performed in each dataset separately using the 'RunPCA' function of the Seurat package¹⁶². Differential expression analysis for organ-specific endothelial cells was performed using a Wilcoxon rank-sum test as implemented in the 'FindAllMarkers' function of the Seurat package. GSEA was performed on significantly upregulated genes ranked by their p-value using the clusterProfiler package¹⁵⁶ in R with gene ontology (GO) gene sets downloaded from the Molecular Signatures Database (MSigDB)¹⁵⁷.

3.2.17 Comparison of Organ-specific Endothelial Transcriptome and Endothelial Single-cell Transcriptomic Data

Cross-platform comparisons between bulk RNA-Seq data and scRNA-Seq data required computing the fold change of each gene relative to a housekeeping gene. We calculated the relative fold change by dividing the expression value for every gene in every sample by an invariable housekeeping gene. We chose *Sap30l* as a housekeeping gene because it was invariable in all three datasets. By generating the fold change matrix in all three datasets, we were then able to use these values to compare relative abundances for genes of interest. We next

calculated Spearman's correlation coefficients for all genes shared between the organ-specific endothelial translome, Tabula Muris scRNA-Seq, and Betsholtz scRNA-Seq datasets, and then separately for all parenchymal (non-endothelial) genes.

3.2.18 Online Endothelial Translome Expression Database

The endothelial translome expression database (www.rehmanlan.org/ribo) is hosted on Amazon S3. The website was constructed using Angular 8.0, JavaScript, HTML5, and CSS. Bar plots and heatmaps were generated for genes of interest using Tableau Public. The visualizations were integrated into the web application using the Tableau JavaScript API. RiboTag log₂ normalized baseline translome expression data were uploaded to Tableau. The averages were computed using Tableau calculated fields. Tableau dashboards and workbooks were created to generate bar plots and heatmaps for online publishing.

3.3 Experiments & Results

3.3.1 Optimized Platform to Characterize Organotypic Endothelial Heterogeneity

To precisely investigate the in-situ organ-specific EC molecular signature in brain, lung, and heart tissue we crossed the RiboTag mice (*Rpl22^{HA/+}*)¹⁴⁷ with the endothelial-specific VE-cadherin-Cre mice^{148,149} to generate RiboTag^{EC} (*Cdh5^{CreERT2/+}*; *Rpl22^{HA/+}*) mice. At 4 weeks post tamoxifen administration, ribosomes in the endothelial cells of all tissues expressed the HA tag, thus allowing for the specific isolation of mRNA undergoing ribosomal translation from ECs in the brain, heart and lung during homeostatic conditions.

We also isolated brain, lung, and heart endothelial mRNA at several time points following systemic inflammatory injury, induced using a sublethal dose of the bacterial endotoxin lipopolysaccharide (LPS), ranging from the acute injury phase at 6 hr post-LPS to the recovery phase at 1 week post-LPS (Figure 11). Log fold change (logFC) values were calculated between endothelial mRNA (immunoprecipitated by an anti-HA antibody) versus whole tissue mRNA (immunoprecipitated with control antibody, anti-RPL22) using quantitative PCR (qPCR). The analysis of the qPCR data confirmed enrichment of endothelial-specific RNA similar to what has been reported in other studies using the RiboTag model¹⁴⁸ and also demonstrated minimal expression of RNA from other tissue-resident cell types (Figure 11B–11F).

After confirming the enrichment of endothelial RNA using qPCR, we performed global transcriptional profiling with RNA-Seq on the RiboTag^{EC} brain, lung, and heart samples. Principal component analysis (PCA) of the RNA-Seq data for endothelial mRNA from brain, lung, and heart tissue from all time points showed a clear separation between the replicate brain, lung, and heart transcriptomes, indicating that ECs from each tissue demonstrated a distinct transcriptional identity at baseline that is maintained even in the setting of profound systemic inflammatory injury (Figure 12A). In order to identify the genes responsible for these distinct tissue-specific EC profiles, we performed a differential expression analysis on the RNA-Seq data. The differential expression analysis was concordant with the PCA and identified 1692 genes which were differentially expressed in brain ECs (versus ECs from the other two tissues),

1052 genes which were differentially expressed in lung ECs, and 570 genes which were differentially expressed in heart ECs (Figure 12B).

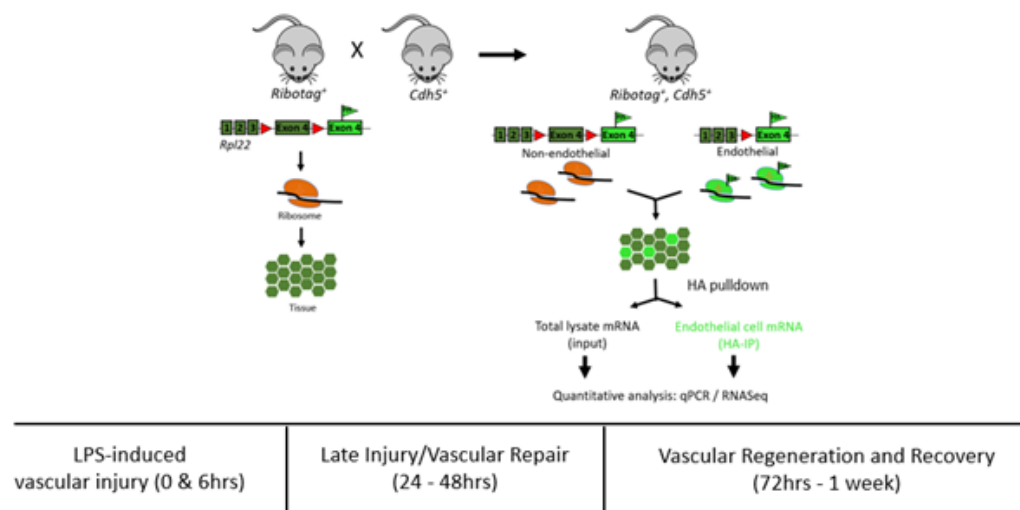
We next analyzed the baseline heterogeneity of ECs obtained from brain, lung and heart by assessing the gene expression levels of endothelial genes using established databases. We specifically focused our analysis on a pan-endothelial gene set¹⁵⁹, glycolysis and fatty acid metabolism gene sets¹⁶⁰, and a solute transport gene set¹⁶¹. Hierarchical clustering of the RNA-Seq profiles on merely 152 pan-endothelial genes from PanglaoDB¹⁵⁹ separated all replicate baseline samples, indicating that classical endothelial markers are sufficient to differentiate ECs from these three organs (Figure 1C). For example, genes upregulated in brain ECs included T-box transcription factor (*Tbx1*) and the glucose transporter 1 (*Slc2a1*), genes upregulated in the lung endothelium included claudin 5 (*Cldn5*) and the *Hes* related family BHLH transcription factor with YRPW Motif 1 (*Hey1*), whereas heart ECs demonstrated upregulation of vascular endothelial growth factor receptor 2 (*Kdr*) and the endothelial cell surface expressed chemotaxis and apoptosis regulator (*Ecsr*).

We next focused on the tissue-specific upregulation of metabolic genes. As seen in the glycolysis gene heatmap, we found that most tissue-specific EC genes involved in glycolysis were specifically upregulated in the brain endothelium (Figure 12D), but there were selected glycolytic genes specifically upregulated in other tissues such as 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (*Pfkfb3*) in lung ECs and alcohol dehydrogenase 1 (*Adh1*) in heart ECs. In contrast, fatty acid metabolism genes were most upregulated in

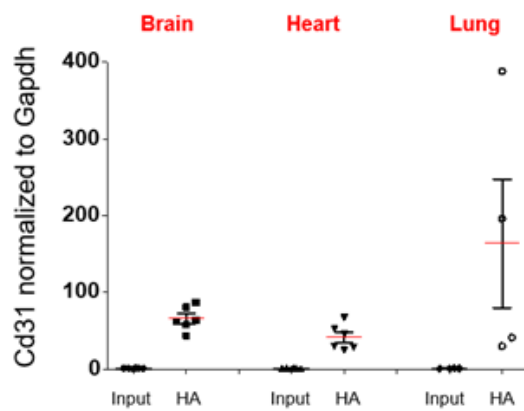
heart ECs consistent with the heavy reliance of the heart on fatty acids to generate ATP (Figure 12E). Heart ECs exhibited upregulation of 17 fatty acid metabolism genes whereas brain ECs and lung ECs only demonstrated upregulation of 9 and 4 metabolism genes, respectively.

Regarding solute transport genes, the brain endothelium showed the most specific upregulation of genes when compared to ECs of the other tissues, both in terms of number of transporters as well as the magnitude of upregulation. We found that 141 transporter genes were upregulated in brain ECs, whereas 43 and 44 genes were upregulated in lung and heart ECs, respectively. As seen in the heatmap (Figure 12F), the expression levels of brain EC-specific transporters were far greater than those of lung and heart ECs, indicative of the central role of solute transport regulation in brain EC function.

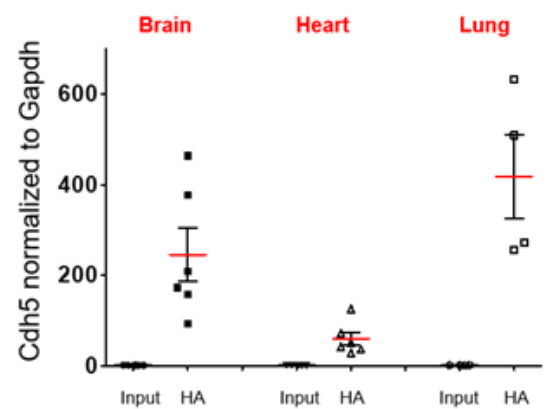
A



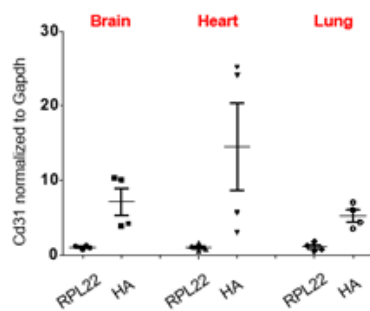
B



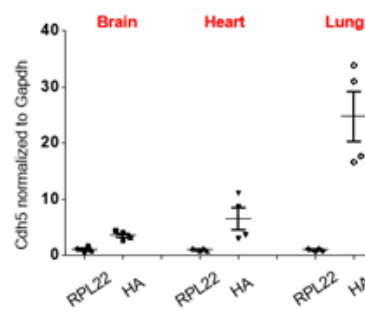
C



D



E



F

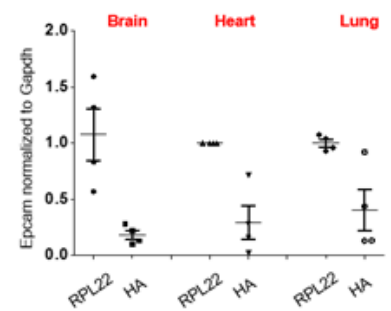
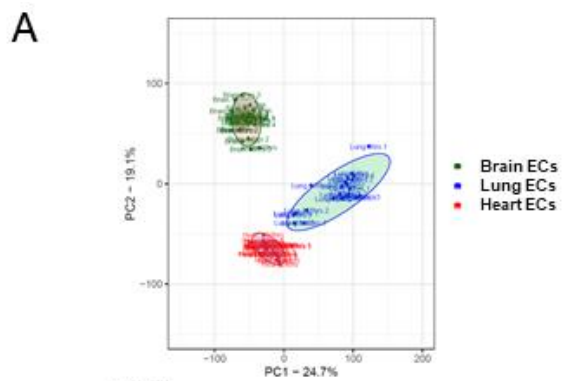


Figure 11: RiboTag Isolation of endothelial mRNA. A) Schematic of RiboTag method for isolating the endothelial-specific transcriptome at various time points during LPS induced inflammatory injury, repair, and regeneration. B–D) Mice brain, lung, and heart RiboTag^{EC} samples were homogenized with ice-cold homogenization buffer. mRNA from total lysate input and IP with HA antibody were used for RT-qPCR to compare (A) *CD31*, (B) *Cdh5* (VE-cadherin). D–F) RiboTag^{EC} mice brain, lung, and heart samples were homogenized with ice-cold homogenization buffer. Total ribosome associated mRNA (IP with anti-RPL22) and endothelial specific ribosome associated mRNA (IP with anti-HA) were used for RT-qPCR to compare (D) *Cd31* and (E) *Cdh5* (VE-cadherin) (F) *Epcam* mRNA levels.



B

Tissue	DEGs	% of total genes
Brain ECs	1692	9%
Lung ECs	1052	6%
Heart ECs	570	3%

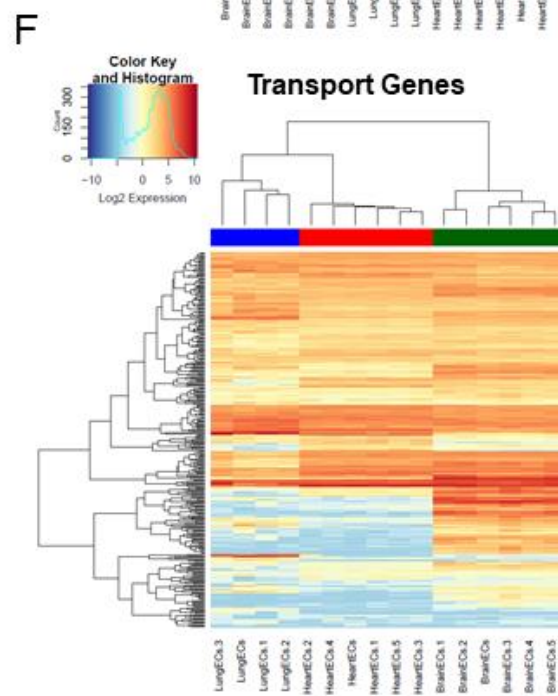
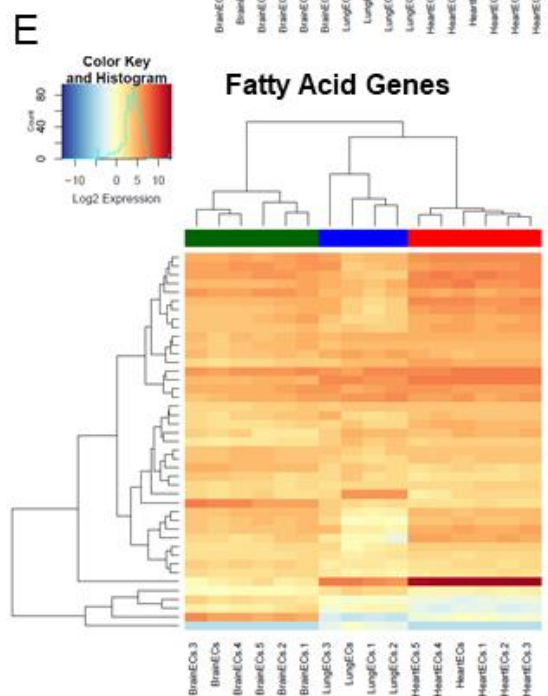
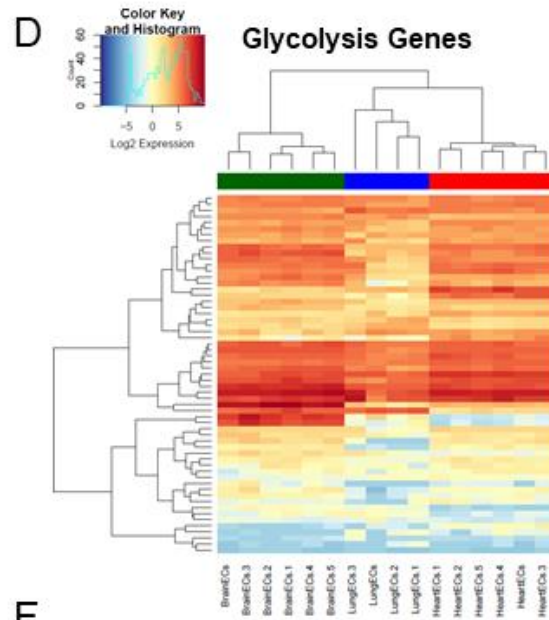
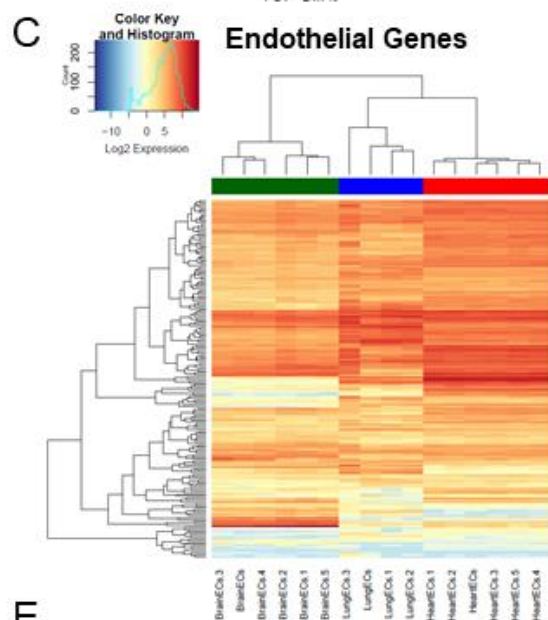


Figure 12: Endothelial heterogeneity exists in classic endothelial functions. A) Principal component analysis of RNA-Seq data generated from brain, lung, and heart endothelial samples isolated from RiboTag^{EC} mice displays the organ-specific in-situ endothelial clusters. B) Differential expression analysis of 18,910 genes which are expressed in brain, lung, and heart endothelium at baseline identified tissue-specific differentially expressed genes. (FDR < 5%) C–F) Hierarchical clustering of classical endothelial processes including (C) endothelial genes, D–E) metabolism, and (F) transporters results in distinct clustering of brain, lung, and heart endothelial baseline samples.

3.3.2 RiboTag^{EC} Endothelial mRNA Purity

After confirming the efficiency of the RiboTag immunoprecipitation protocol using qPCR, we next sought to perform an unbiased and systematic analysis of the utility of the RiboTag^{EC} model as a tool to study the organ-specific endothelial transcriptome heterogeneity. We therefore compared organ-specific RiboTag^{EC} RNA-Seq baseline profiles to healthy whole-tissue RNA-Seq profiles obtained from publicly available whole tissue RNA-Seq dataset¹⁶³. By applying normalization and batch correction techniques, we were able to directly compare the mRNA expression levels of RiboTag^{EC} endothelial samples with those of whole tissue samples.

To characterize the whole brain, lung, and heart samples, we identified the genes that were significantly upregulated in each of the tissues and generated a heatmap displaying the 1358 differentially upregulated whole brain-specific genes relative to whole lung and whole heart (Figure 13A). By performing a gene set enrichment analysis (GSEA) to ascertain the pathways associated with these genes, we confirmed the validity of the samples because the top pathways included 'neurotransmitter transport', 'synapse organization', 'synaptic vesicle cycle' (Figure 13B). The top 10 most abundant genes in the whole brain RNA-Seq data included myelin basic protein (*Mbp*), proteolipid protein 1 (*Plp1*), calmodulin 1 (*Calm1*), synaptosome associate protein 25 (*Snap25*), kinesin family member 5A (*Kif5a*), ATPase Na⁺/K⁺ transporting subunit alpha 3 (*Atp1a3*), sodium-dependent glutamate/aspartate transporter 2 (*Slc1a2*),

secreted protein acidic and cysteine rich (*Sparcl1*), carboxypeptidase e (*Cpe*), stearyl-coA desaturase 2 (*Scd2*) (Figure 13C).

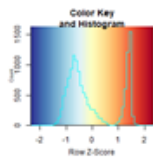
Whole lung samples were characterized by 1071 differentially expressed genes (Figure 14A) on which we performed GSEA (Figure 14B). The top 10 most abundant genes in the whole lung were desmoyokin (*Ahnak*), microtubule-actin crosslinking factor 1 (*Macf1*), actin beta (*Actb*), surfactant protein c (*Sftpc*), spectrin beta, non-erythrocytic 1 (*Sptbn1*), hypoxia inducible factor two alpha (*Hif2a*), stearyl-CoA desaturase (*Scd1*), filamin a (*Flna*), adhesion g protein-coupled receptor f5 (*Adgrf5*), and ldl receptor related protein 1 (*Lrp1*) (Figure 14C).

The signature of the whole heart derived from differential gene expression analysis was composed of 1351 genes (Figure 15A). GSEA indicated a preponderance of metabolic and muscle contraction pathways (Figure 15B). The top 10 most abundant cardiac genes were myosin heavy chain 6 (*Myh6*), ATPase sarcoplasmic/endoplasmic reticulum Ca^{2+} transporting 2 (*Atp2a2*), myoglobin (*Mb*), actin, alpha, cardiac muscle 1 (*Actc1*), phospholamban (*Pln*), myosin regulatory light chain 2 (*Myl2*), titin (*Ttn*), troponin t2, cardiac type (*Tnnt2*), tropomyosin 1 (*Tpm1*), and lipoprotein lipase (*Lpl*) (Figure 15C).

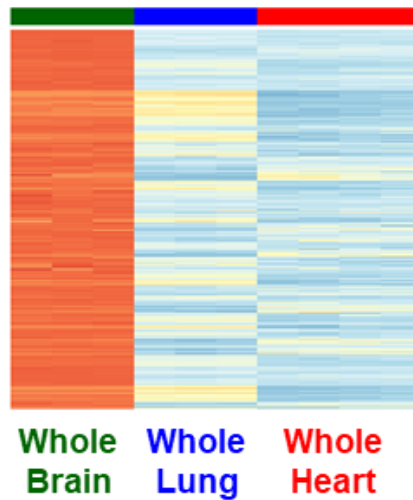
After establishing and confirming the molecular signatures of the whole brain, whole lung, and whole heart tissue, we next calculated a Kendall's Tau correlation coefficient to assess the rank correlation between the RiboTag^{EC} samples and the whole tissue samples. We surmised that if the rank of the most abundant whole tissue genes was the same as the rank of these genes in the

RiboTag^{EC} samples, then it would indicate possible contamination of the EC samples with whole tissue mRNA; however, if the abundance rank order of whole tissue genes was quite distinct from that in the RiboTag^{EC} samples, then it would indicate tissue specific programming of ECs in situ (Figure 16A). We assessed the Kendall's Tau rank correlation for all three tissues and plotted correlation heatmaps showing the results (Figure 16B–D). Our findings indicate that there was no significant correlation between the abundance rank of whole tissue genes and their rank order in the RiboTag^{EC} samples. The rank correlation in the brain samples ranged from -0.29 to 0.38 (Figure 16B). Since the cellular composition of the lung is 40–50% endothelial, we expectantly saw a higher rank correlation between whole lung samples and lung RiboTag^{EC} samples, ranging between 0.02 and 0.6 (Figure 16C). In the heart, we found a range of rank correlations between -0.29 to 0.24 (Figure 16D). These results provide mathematical evidence for the robustness and purity of the RiboTag^{EC} samples.

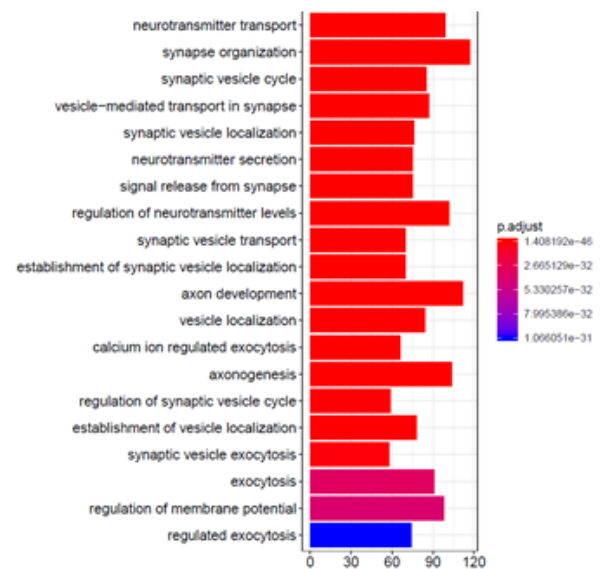
A



WholeBrain vs WholeLung/WholeHeart DEGs



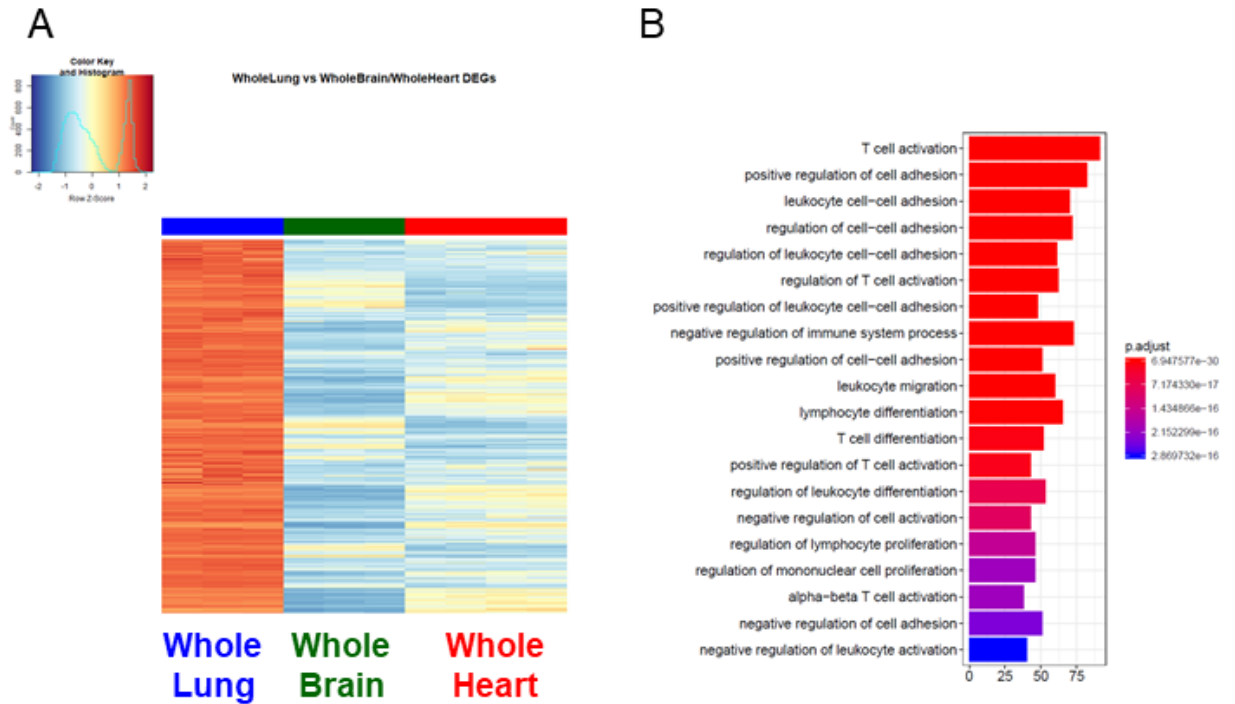
B



C

Most abundant whole brain genes	
Mbp	Myelin Basic Protein
Plp1	Proteolipid Protein 1
Calm1	Calmodulin 1
Snap25	Synaptosome Associated Protein 25
Kif5a	Kinesin Family Member 5A
Atp1a3	ATPase Na ⁺ /K ⁺ Transporting Subunit Alpha 3
Slc1a2	Sodium-Dependent Glutamate/Aspartate Transporter 2
Sparcl1	Secreted Protein Acidic And Cysteine Rich
Cpe	Carboxypeptidase E
Scd2	stearoyl-CoA desaturase 2

Figure 13: Characterization of Whole Brain RNA-Seq data. A) Heat map representation of differentially upregulated genes identified by comparing whole brain RNA-Seq samples to whole lung and whole heart tissues at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. B) The GSEA results of enriched GO terms from whole brain samples at baseline. C) Top whole brain signature genes ranked in order of abundance.



C

Most abundant whole lung genes	
Ahnak	Desmoyokin
Macf1	Microtubule-Actin Crosslinking Factor 1
Actb	Actin Beta
Sftpc	Surfactant Protein C
Sptbn1	Spectrin Beta, Non-Erythrocytic 1
Epas1	HIF2α
Scd1	Stearoyl-CoA Desaturase
Flna	Filamin A
Adgrf5	Adhesion G Protein-Coupled Receptor F5
Lrp1	LDL Receptor Related Protein 1

Figure 14: Characterization of Whole Lung RNA-Seq data. A) Heat map representation of differentially upregulated genes identified by comparing whole lung RNA-Seq samples to whole brain and whole heart tissues at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. B) The GSEA results of enriched GO terms from whole lung samples at baseline. C) Top whole lung signature genes ranked in order of abundance.

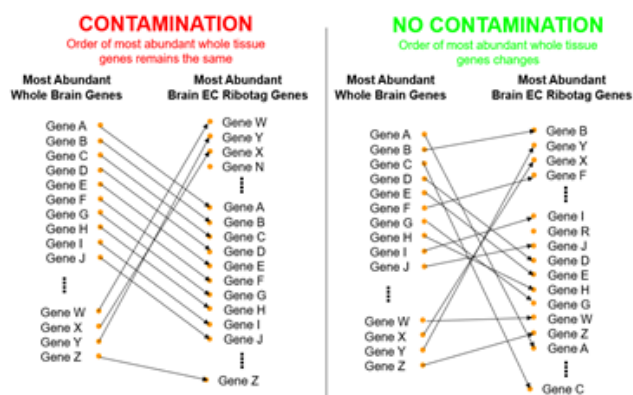


C

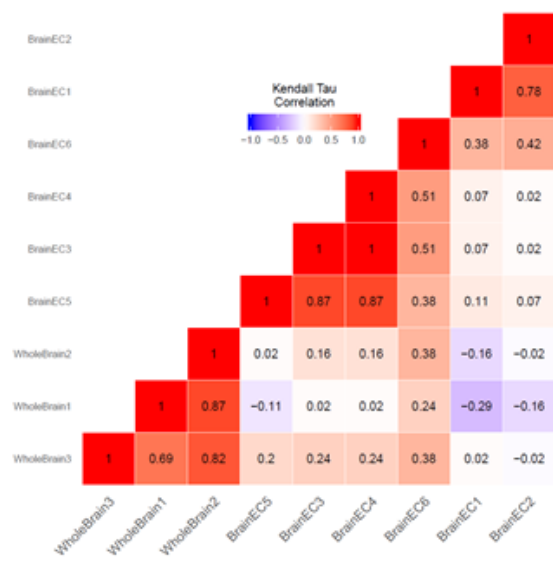
Most abundant whole heart genes	
Myh6	Myosin Heavy Chain 6
Atp2a2	ATPase sarcoplasmic/endoplasmic reticulum Ca ²⁺ transporting 2
Mb	Myoglobin
Actc1	Actin, Alpha, Cardiac Muscle 1
Pln	Phospholamban
MyI2	Myosin regulatory light chain 2
Ttn	Titin
Tnnt2	Troponin T2, Cardiac Type
Tpm1	Tropomyosin 1
Lpl	Lipoprotein Lipase

Figure 15: Characterization of Whole Heart RNA-Seq data. A) Heat map representation of differentially upregulated genes identified by comparing whole heart RNA-Seq samples to whole brain and whole lung tissues at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. B) The GSEA results of enriched GO terms from whole heart samples at baseline. C) Top whole heart signature genes ranked in order of abundance.

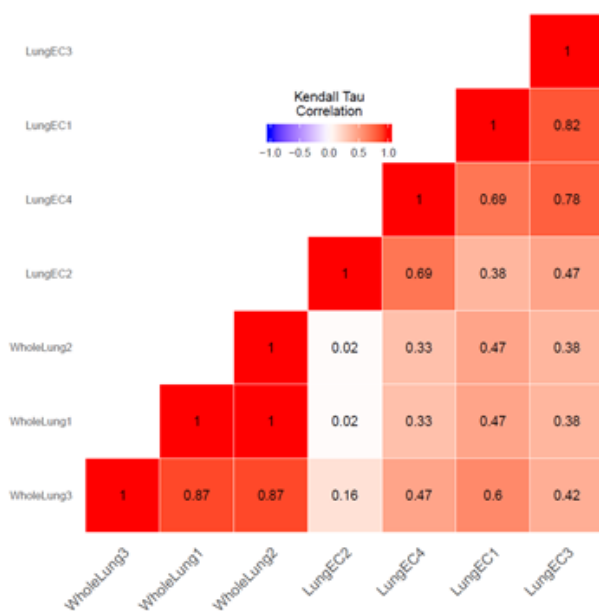
A



B



C



D

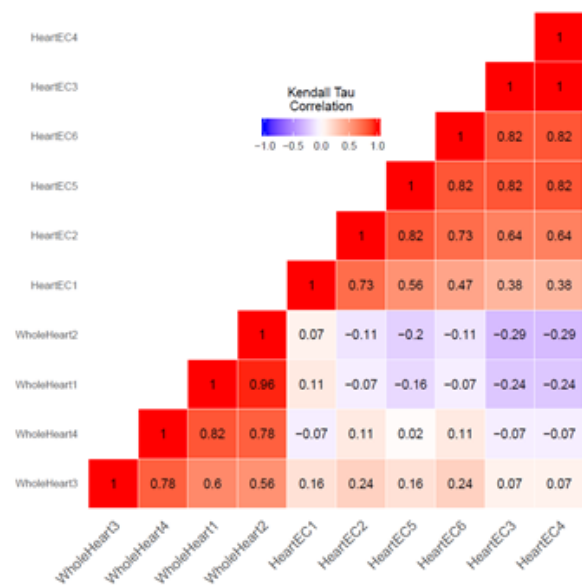


Figure 16: Kendall's Tau correlation supports endothelial mRNA isolation from RiboTag^{EC} mice. A) Schematic of assessing endothelial purity in RiboTag^{EC} using Kendall's Tau correlation shows rearrangement in the rank of most abundant genes between whole tissue and endothelial samples supports purity of the endothelial samples. B–D) Kendall's Tau non-parametric correlation plots for whole tissue and RiboTag^{EC} RNA-Seq samples for (B) brain, (C) lung, and (D) heart tissue.

3.3.3 Brain-specific Endothelial Molecular Signature

After confirming the robustness and purity of the RiboTagEC samples, we performed differential expression analysis to identify the significantly upregulated genes in the brain endothelial transcriptome (Figure 17, full list of genes available at: <https://cdn.elifesciences.org/articles/51413/elifesciences-sup1-v2.xlsx>). We used these upregulated genes as the input into GSEA to characterize the brain ECs (Figure 17B). Surprisingly, we found that genes involved in processes typically thought of being canonical neuronal functions such as synapse organization, neurotransmitter transport, axon development, and regulation of ion transmembrane transport were significantly enriched in brain ECs (Figure 17B). The top 10 most significantly upregulated genes in the brain ECs included: prostaglandin d synthase (*Ptgds*), ATPase, Na⁺/K⁺ transporting, alpha two polypeptide (*Atp1a2*), basigin (*Bsg*), apolipoprotein e (*ApoE*), glutamine synthase (*GluT*), apolipoprotein d (*ApoD*), pleiotrophin (*Ptn*), insulin like growth factor 2 (*Igf2*), osteonectin (*Spock2*), and glucose transporter 1 (*Slc2a1*) (Figure 17C). In order to identify brain EC-specific surface markers, which could be of great value for therapeutic targeting of brain ECs, we used the Cell Surface Protein Atlas database (Bausch-Fluck et al., 2015) and identified the top 10 surface markers for brain ECs (Figure 17D), which included the glutamate/aspartate transporter II (*Slc1a2*), thyroxine transporter (*Slc1c1*), glial fibrillary acidic protein (*Gfap*), ATPase Na⁺/K⁺ transporting subunit alpha 3 (*Atp1a3*), endothelin b receptor-like protein 2 (*Gpr37l1*), Delta/Notch like EGF repeat containing transmembrane (*Dner*), synaptic vesicle glycoprotein 2b (*Sv2b*), sodium voltage-gated channel

beta subunit 2 (*Scn2b*), glutamate ionotropic receptor NMDA type subunit 2a (*Grin2a*), and neurofascin (*Nfasc*). Individual boxplots for the log2 expression levels of each gene show that the expression levels of these cell surface markers are 6–8 log2 units higher in brain ECs than in the lung and heart endothelium. We freshly isolated individual ECs, performed a cytopspin and stained for the neurotrophic factor PTN and found that it was expressed on individual brain ECs but at much lower levels in heart or lung ECs (Figure 17E).

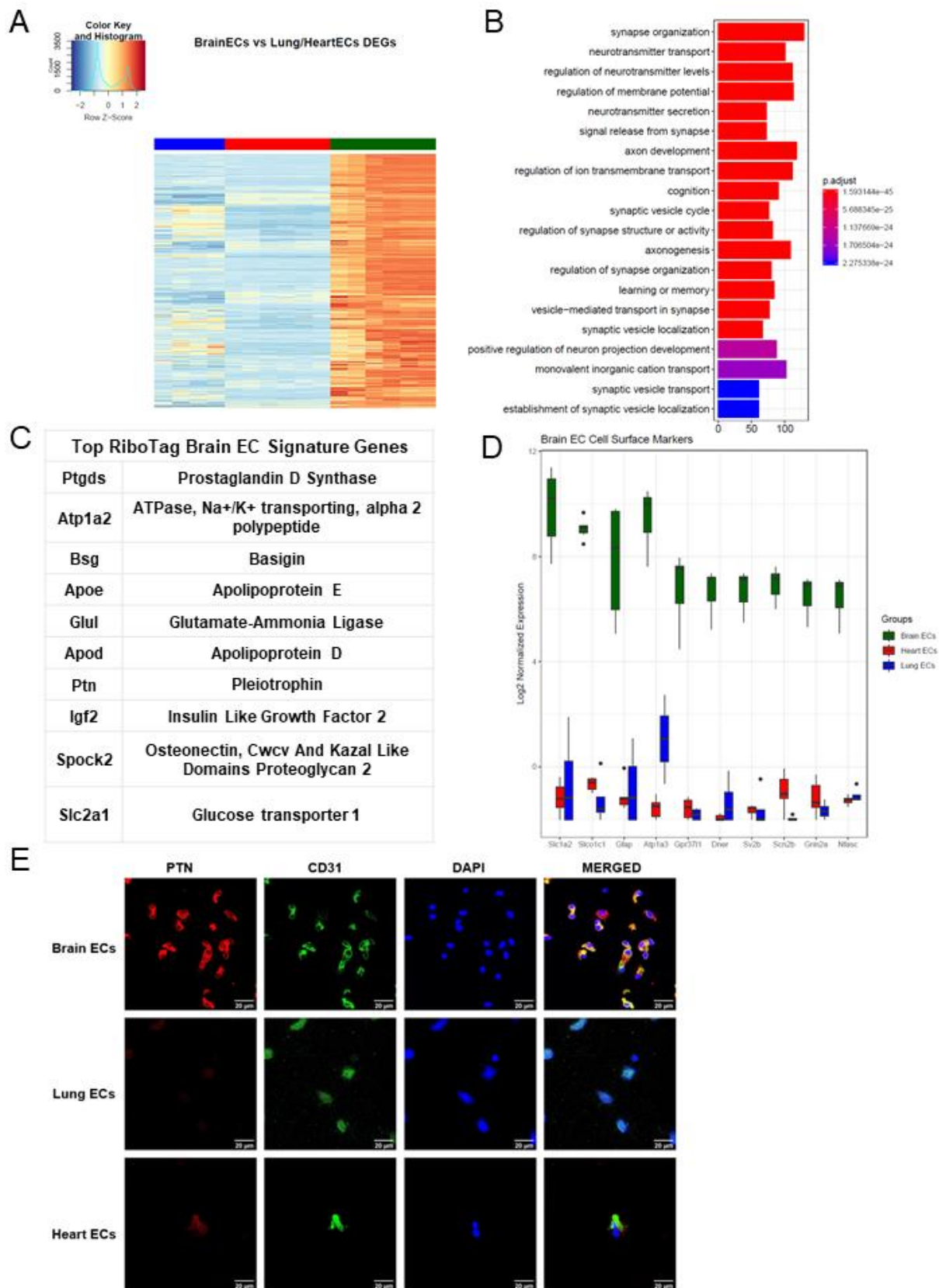


Figure 17: Brain endothelial specific signature. A) Heat map representation of differentially upregulated genes identified by comparing brain ECs to lung and heart ECs at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. Individual gene expression values can be visualized at www.rehmanlab.org/ribo. B) The GSEA results of enriched GO terms from RiboTag brain ECs at baseline. C) Top RiboTag brain EC signature markers ranked in order of logFC. D) Top RiboTag brain EC cell surface markers identified using the Cell Surface Protein Atlas. E) Confocal analysis was performed after brain, lung, and heart ECs were processed on a cytospin to assess brain EC PTN (Pleiotrophin) specificity. A scale bar of 20 μ m is included on all images.

3.3.4 Lung-specific Endothelial Molecular Signature

We next analyzed the lung EC signature using differential expression analysis (Figure 18A, full list of genes available at <https://cdn.elifesciences.org/articles/51413/elifesciences-sup2-v2.xlsx>). We found that the lung endothelium exhibits significant upregulation of genes involved in biological processes related to immune function such as leukocyte cell-cell adhesion, T cell activation, leukocyte migration, and regulation of immune system processes (Figure 18B). The 10 most significantly upregulated genes in lung ECs included surfactant protein c (*Sftpc*), advanced glycosylation end-product specific receptor (*Ager*), norepinephrine transporter (*Slc6a2*), chitinase-like protein 3 (*Chil3*), WAP four-disulfide cco domain 2 (*Wfdc2*), c-type lectin domain containing 7a (*Clec7a*), mucin 1 (*Muc1*), resistin like alpha (*Retnla*), lysozyme (*Lyz1*), homeobox a5 (*Hoxa5*) (Figure 18C). The top lung endothelial cell surface markers included norepinephrine transporter (*Slc6a1*), mucin 1 (*Muc1*), tumor necrosis factor c (*Ltb*), prostaglandin transporter (*Slco2a1*), epithelial membrane protein 2 (*Emp2*), ATPase sarcoplasmic/endoplasmic reticulum Ca²⁺ transporting 3 (*Atp2a3*), epithelial cell adhesion molecule (*Epcam*), leukocyte function-associated molecule one alpha chain (*Itga1*), interleukin three receptor subunit alpha (*Il3ra*), matriptase (*St14*) (Figure 18D). We validated our computational analysis by staining freshly isolated ECs for RAGE and found that RAGE was only expressed at significant levels in lung ECs but not heart or brain ECs (Figure 18E).

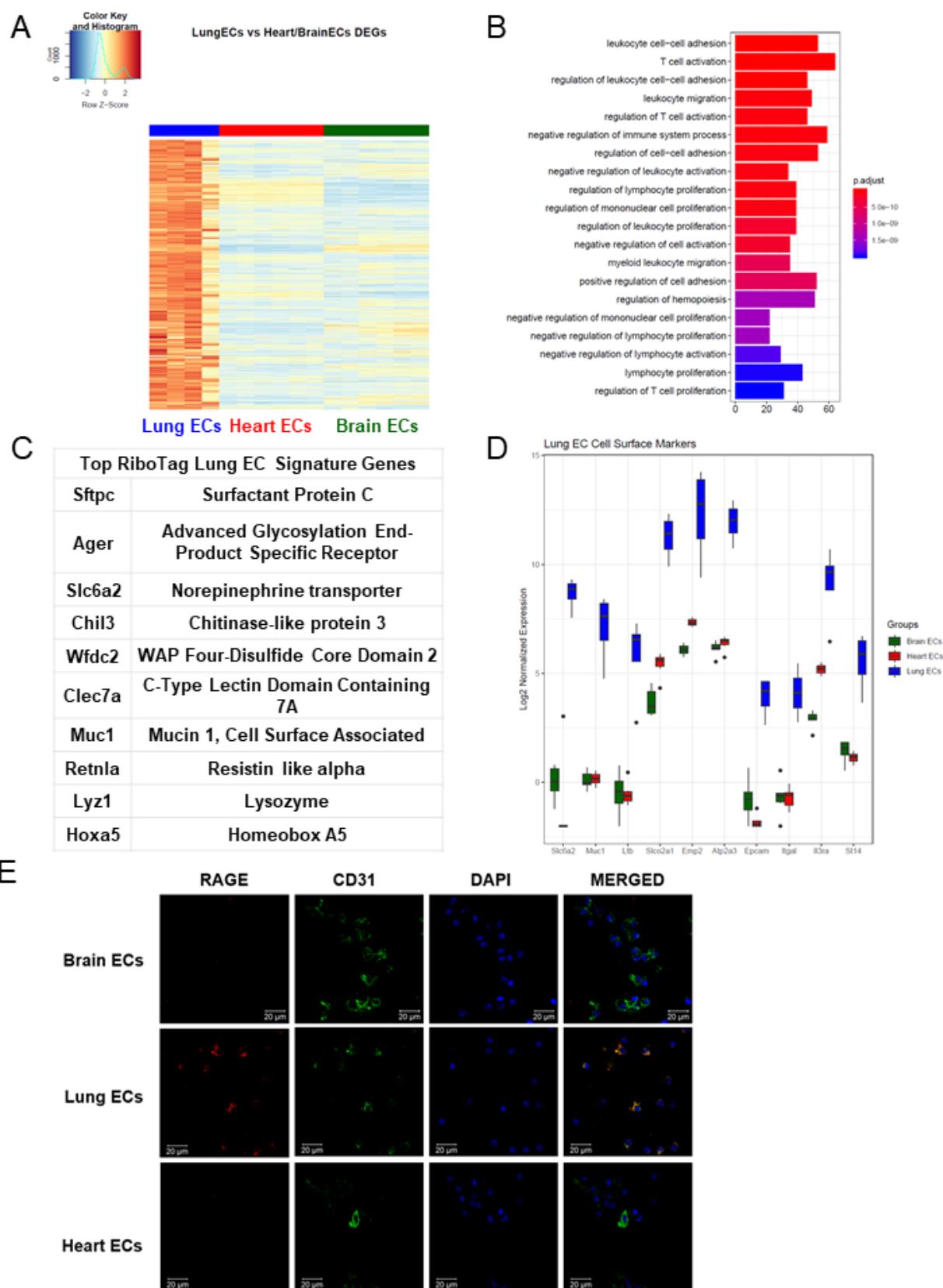


Figure 18: Lung endothelial specific signature. A) Heat map representation of differentially upregulated genes identified by comparing lung ECs to brain and heart ECs at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. Individual gene expression values can be visualized at www.rehmanlab.org/ribo B) The GSEA results of enriched GO terms from RiboTag lung ECs at baseline. C) Top RiboTag lung EC signature markers ranked in order of logFC. D) Top RiboTag lung EC cell surface markers identified using the Cell Surface Protein Atlas. E) Confocal analysis was performed after brain, lung, and heart ECs were processed on a cytospin to assess lung EC RAGE (Receptor for Advanced Glycation Endproducts) specificity. A scale bar of 20 μm is included on all images.

3.3.5 Heart-specific Endothelial Molecular Signature

We then studied the differentially expressed genes in the heart endothelium (Figure 19A, Full list of genes available at: <https://cdn.elifesciences.org/articles/51413/elifesciences-sup3-v2.xlsx>). GSEA identified pathways specifically upregulated in heart ECs, as compared to brain and lung ECs (Figure 19B). Strikingly, we found that the genes specifically upregulated in heart ECs were involved in processes such as cardiac muscle tissue development, myofibril assembly and cardiac contraction (Figure 19B). This suggested that the cardiac endothelium expresses genes canonically thought to be cardiomyocyte genes, analogous to the expression of canonical neuronal genes in the brain endothelium. The top expressing heart EC genes included myosin regulatory light chain 2 (*Myf2*), myosin regulatory light chain 3 (*Myf3*), aquaporin 7 (*Aqp7*), ADP-ribosylhydrolase like 1 (*Adprhl1*), alpha 2-HS glycoprotein (*Ahsg*), sodium-coupled nucleoside transporter (*Slc28a2*), xin actin binding repeat containing 2 (*Xirp2*), myoglobin (*Mb*), Butyrophilin like 9 (*Btnl9*), creatine kinase, mitochondrial 2 (*Ckmt2*), leucine rich repeats and transmembrane domains 1 (*Lrtm1*), and fatty acid binding protein 4 (*Fabp4*) (Figure 19C). The top 10 heart EC surface marker genes included alpha 2-HS glycoprotein (*Ahsg*), sodium-coupled nucleoside transporter (*Slc28a2*), titin (*Ttn*), tumor necrosis factor receptor superfamily member 27 (*Eda2r*), platelet glycoprotein 4 (*Cd36*), laminin subunit alpha 4 (*Lama4*), fibulin 2 (*Fbln2*), ectonucleotide pyrophosphatase/phosphodiesterase 3 (*Enpp3*), t-cadherin (*Cdh13*), steroid sensitive gene 1 (*Ccdc80*) (Figure 19D). We tested the heart EC

specificity of AQP7 using confocal analysis on freshly isolated brain, lung, and heart ECs and found that AQP7 was robustly expressed in heart ECs but minimally expressed in brain and lung ECs (Figure 19E).

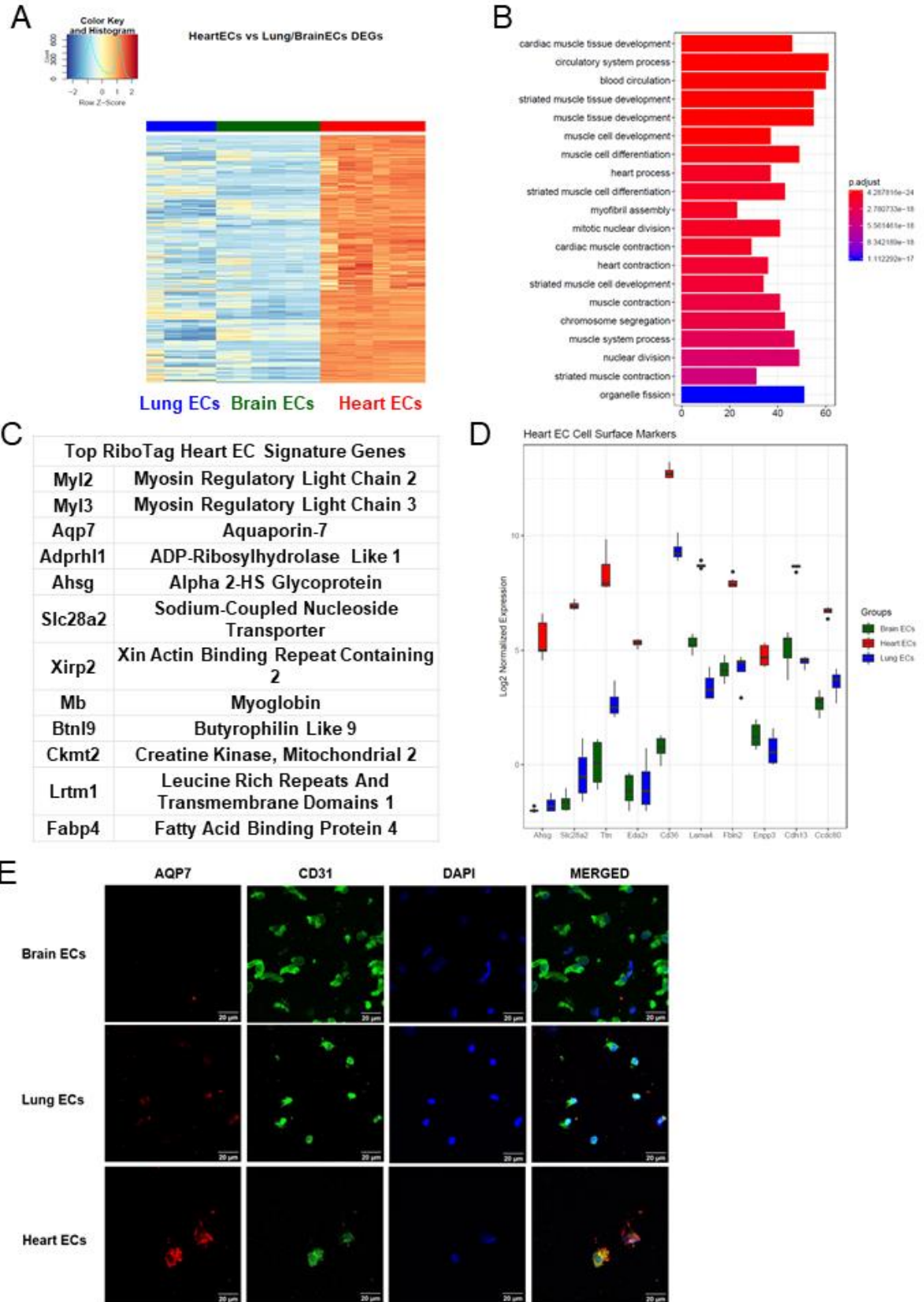


Figure 19: Heart endothelial specific signature. A) Heat map representation of differentially upregulated genes identified by comparing heart ECs to brain and lung ECs at baseline. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway. Individual gene expression values can be visualized at www.rehmanlab.org/ribo. B) The GSEA results of enriched GO terms from RiboTag heart ECs at baseline. C) Top RiboTag heart EC signature markers ranked in order of logFC. D) Top RiboTag heart EC cell surface markers identified using the Cell Surface Protein Atlas. E) Confocal analysis was performed after brain, lung, and heart ECs were processed on a cytospin to assess heart EC AQP7 (Aquaporin 7) specificity. A scale bar of 20 μm is included on all images.

3.3.6 Single-cell Endothelial Heterogeneity

In light of the surprising findings that endothelial cells express genes typically associated with surrounding parenchymal cells such as cardiomyocytes or neuronal cells, we next used single cell RNA-Seq analysis to assess whether the RiboTag^{EC} endothelial signatures are also found in individual endothelial cells by analyzing endothelial single-cell data from the Tabula Muris compendium¹⁶⁴ and the single cell RNA-Seq atlas of the brain and lung endothelium¹⁶⁵. Using expression of the endothelial genes *Cd31* and *Cdh5* as markers of ECs, we analyzed double positive cells for both markers in Tabula Muris brain, lung, and heart tissues and performed PCA to assess the extent of endothelial heterogeneity (Figure 20A). The PCA plot partitioned the cells into groups defined by their tissue of origin, indicating a tissue-specific EC signature even at the single cell level. Similarly, we performed PCA on ECs in the Betsholtz dataset (which relied on *Cd31* and *Cldn5* as EC markers) and also found that ECs similarly clustered according to their tissue of origin (Figure 20B).

We then used these two scRNA-Seq endothelial datasets for the three organs we had analyzed in our RiboTag experiments and intersected the differentially expressed genes for each organ-specific endothelial population. The intent of this was to ascertain which tissue-specific EC signature genes were present in the single cell datasets as well as our RiboTag^{EC} dataset. We found that the shared brain EC signature across all three datasets (Tabula Muris^{EC}, Betsholtz^{EC} and RiboTag^{EC}) for brain ECs was enriched for genes involved in ion transport, acid transport, synapse organization and neurotransmitter transport

(Figure 20C). This finding is consistent with the brain EC-specific enrichment of neuronal signaling pathways that had been identified by the RiboTag^{EC} analysis (Figure 17). We also found that the genes specifically upregulated in the Tabula Muris and Betsholtz lung ECs were involved in T cell activation, TGF β signaling, and antigen processing and presentation (Figure 20D), again consistent with the 'immune activation' signature identified by the RiboTag^{EC} analysis alone (Figure 18). Similarly, the shared upregulated genes in Tabula Muris single cell heart ECs were involved in processes such as cardiac muscle contraction, myofibril assembly and proliferation (Figure 20E, Figure 19).

We next quantified the intersection of brain, lung and heart endothelial marker genes across the Tabula Muris, brain and lung EC atlas, and RiboTag datasets. For the brain endothelium, 40 of the Tabula Muris top 50 brain EC specific genes were also brain EC specific genes in the RiboTag dataset. In the Betsholtz dataset, 27 of the top 50 brain EC specific genes were present in the RiboTag brain EC specific genes (Figure 20F). We found that 17 of the top lung endothelial specific genes in the Betsholtz data set were also found in the list of lung endothelial-specific genes in the RiboTag model (Figure 20G). Of the 24 top lung endothelial specific genes found in the Tabula Muris data set, the same genes were also found in the list of lung endothelial-specific genes in the RiboTag model (Figure 20G).

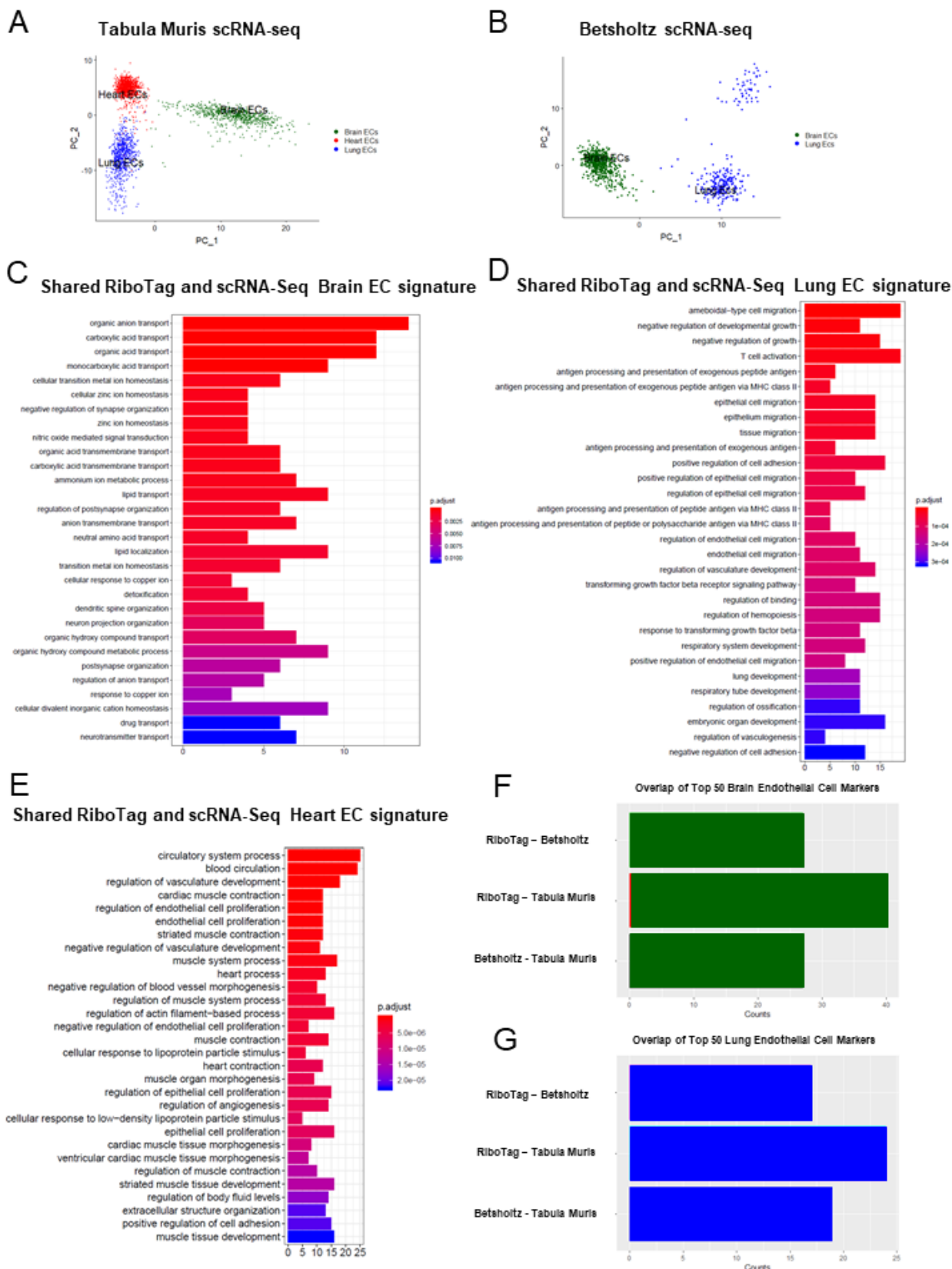


Figure 20: Single-cell endothelial heterogeneity. A) PCA of endothelial scRNA-Seq data from the Tabula Muris collection of mouse tissues colored by tissue. B) PCA of endothelial scRNA-Seq data from the Betsholtz Lab of mouse tissues colored by tissue. The GSEA results of enriched GO terms from overlapping differentially expressed genes between RiboTag and Betsholtz or Tabula Muris for C) brain ECs, D) lung ECs, and E) heart ECs. F) Overlap of top 50 scRNA-Seq brain EC marker genes with RiboTag brain EC marker genes. G) Overlap of top 50 scRNA-Seq lung EC marker genes with RiboTag lung EC marker genes.

3.3.7 Organ-specific Parenchymal Gene Signature Exists in Endothelial scRNA-Seq

To address further that the parenchymal signatures (Brain endothelial parenchymal signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl4-v2.xlsx>, Lung endothelial parenchymal signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl4-v2.xlsx>, Heart endothelial parenchymal signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl6-v2.xlsx>) identified in the endothelial transcriptome were simply not driven by low abundance of transcripts, we performed a Spearman correlation analysis to compare organ-matched RiboTag bulk RNA-Seq data with scRNA-Seq data generated by the Betsholtz and the Tabula Muris Compendium (Figure 21, Figure 22). In each dataset, we first determined the fold change for all genes using a housekeeping gene, Sap30l which we identified as being stably expressed across all datasets, and thus ideally suited to perform relative abundance comparisons (Brain endothelial shared signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl7-v2.xlsx>, Lung endothelial shared signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl8-v2.xlsx>, Heart endothelial shared signature: <https://cdn.elifesciences.org/articles/51413/elifesciences-suppl8-v2.xlsx>). Using the fold change values, we calculated the correlation coefficients between the brain endothelial transcriptome and single cell brain ECs from the Betsholtz and Tabula Muris datasets. We found that the correlation between RiboTag and Betsholtz was 0.53 for all genes detected in the

brain endothelium (Figure 21A) while the correlation between RiboTag and Tabula Muris was 0.47 (Figure 22A). We then specifically tested whether the parenchymal signature genes in the brain endothelium were correlated with the Betsholtz and Tabula Muris individual brain ECs. The correlation of the parenchymal gene expression between RiboTag brain EC samples and Betsholtz brain ECs was 0.31 (Figure 21B) while with Tabula Muris brain ECs the correlation was 0.28 (Figure 22B). Importantly, the brain EC parenchymal genes including synaptosome associated protein 47 (Snap47) and synaptotagmin 11 (Syt11) were expressed at similar or higher levels in the single cell brain ECs from the Betsholtz and Tabula Muris datasets than in the RiboTag brain EC samples (Figure 21C). We performed identical analysis for the lung and heart endothelium (Figure 21D–I, Figure 22), and found that similar correlation values ranging between 0.37 to 0.68. Of note, the heart endothelial gene expression was the most correlated organ across the distinct platforms (Figure 21 G–H). In the lung and heart endothelium, we also found that individual genes representing the parenchymal signature were expressed at similar or higher levels in the single cell samples (Figure 21F, Figure 21G–I), such as the cardiac contractile gene Tropomyosin (Tpm1), which was expressed at higher levels in individual heart ECs from the Tabula Muris dataset.

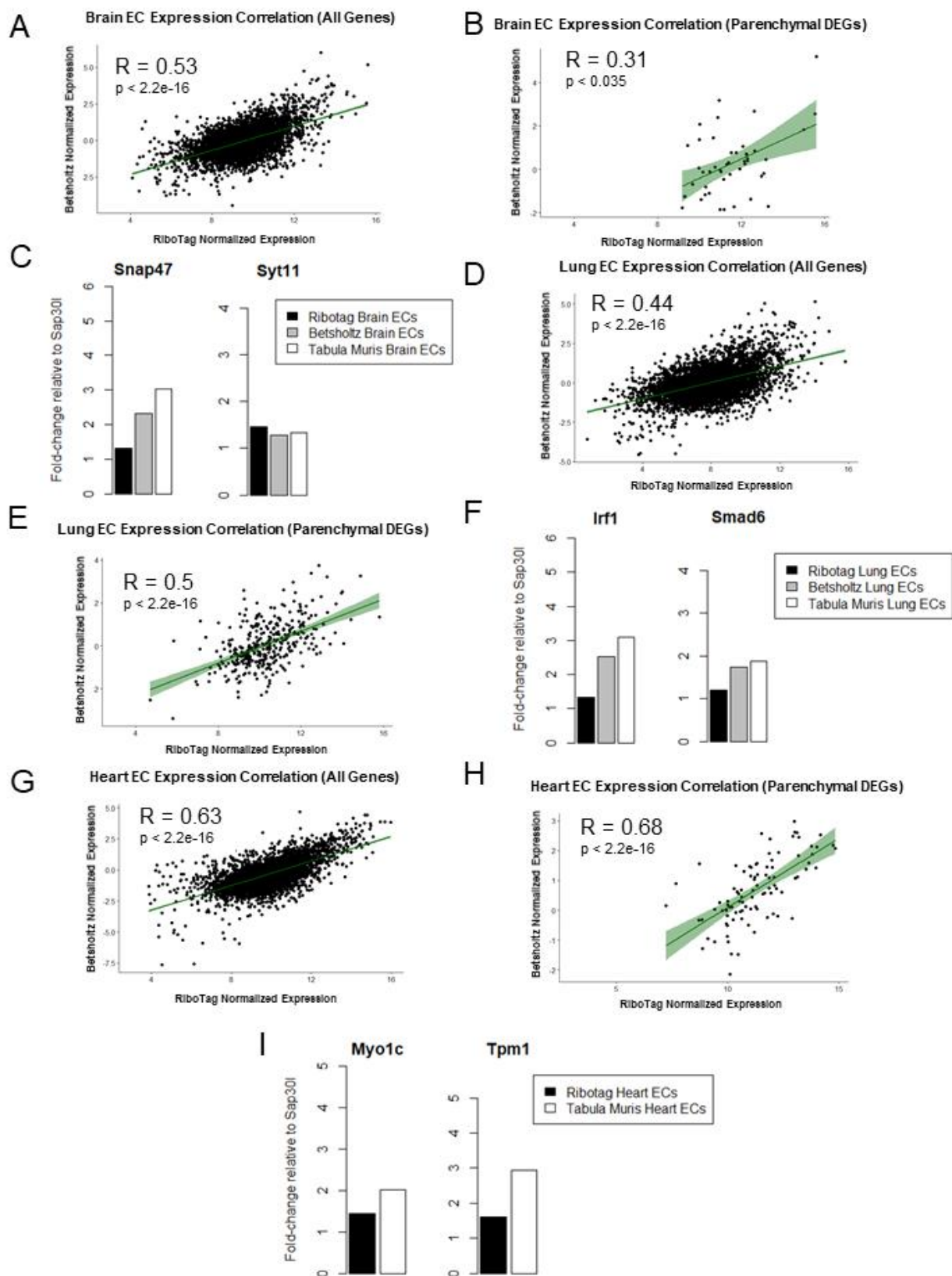


Figure 21: Spearman correlation scatter plots of average gene expression in RiboTag bulk RNA-Seq and Betsholtz scRNA-Seq were generated for A) all genes detected in brain ECs and B) parenchymal (non-endothelial) genes detected in brain ECs. C) Average expression level of representative brain EC parenchymal genes. Spearman correlation scatter plots of average gene expression in RiboTag bulk RNA-Seq and Betsholtz scRNA-Seq were generated for D) all genes detected in lung ECs and E) parenchymal (non-endothelial) genes detected in lung ECs F) Average expression level of representative lung EC parenchymal genes. Spearman correlation scatter plots of average gene expression in RiboTag bulk RNA-Seq and Tabula Muris scRNA-Seq were generated for (G) all genes detected in heart ECs and (H) parenchymal (non-endothelial) genes detected in heart ECs (I) Average expression level of representative heart EC parenchymal genes.

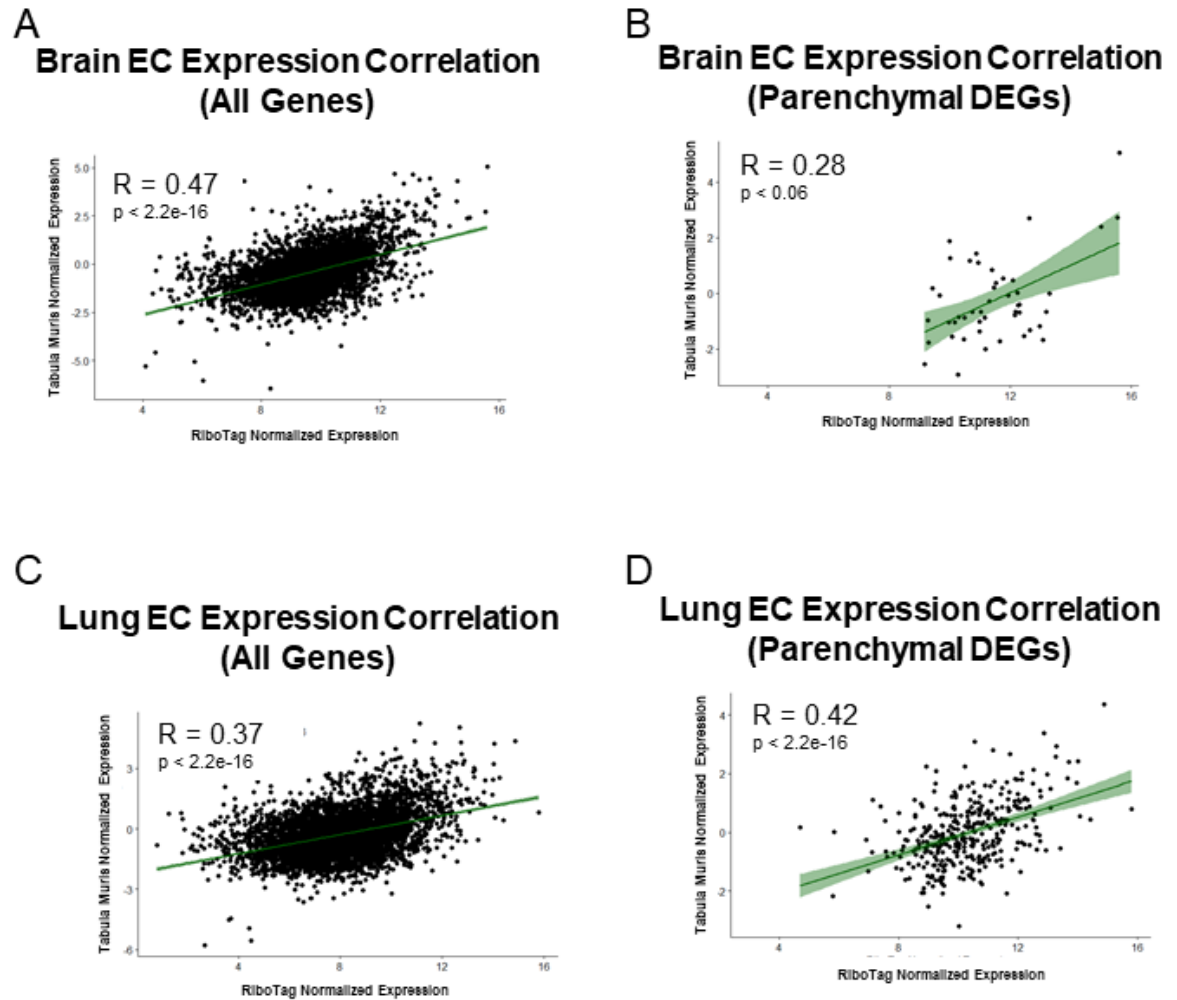


Figure 22: Expression Correlation Analysis between endothelial gene expression generated by RiboTag, Betsholtz, and Tabula Muris Spearman correlation scatter plots of average gene expression in RiboTag bulk RNA-Seq, Betsholtz scRNA-Seq (Smart-Seq2), and Tabula Muris scRNA-Seq (10x Genomics). A) All genes detected in brain ECs. B) Parenchymal (non-endothelial) genes detected in brain ECs. C) All genes detected in lung ECs. D) Parenchymal (non-endothelial) genes detected in lung ECs.

3.4 Conclusions

The endothelium which lines the entire vasculature matures in a tissue-dependent manner during embryonic development to control organ development, homeostasis, and tissue regeneration¹⁶⁶. Under normal physiological conditions, the endothelium maintains a quiescent interface between the blood and tissue. During stimulation, the endothelium becomes actively responsible for controlling blood flow, vascular permeability, leukocyte infiltration, and tissue edema¹⁶⁷. Understanding the organotypic endothelial heterogeneity that exists at baseline is essential for understanding endothelial plasticity and its implications in tissue-specific function¹⁶⁸⁻¹⁷¹.

The RiboTag strategy was originally applied to expression profiling of neurons and Sertoli cells¹⁴⁷. Cell type specificity of the approach depends on the accuracy of the Cre driver that is combined with the *Rpl22^{HA}* allele. This aspect is highlighted in our study and we revealed the precision of the inducible system for achieving endothelial specificity. Our results demonstrate that the RiboTag approach provides a useful method to identify distinct molecular gene expression signatures of tissue-specific endothelium. Performing high-throughput gene expression analysis on the transcriptome using the RiboTag approach enabled us to establish tissue-specific molecular signatures underlying in situ endothelial heterogeneity. During homeostasis, we found that the endothelial transcriptome in each organ is uniquely characterized by a signature adapted to the surrounding parenchymal tissue. The metabolic adaptation of the endothelium is less surprising as the endothelium plays a critical role in supplying nutrients to the

host tissue^{171,172}. The upregulation of the glucose transporter 1 (*Slc2a1*) in brain ECs is consistent with the massive glucose consumption of the brain¹⁷³, whereas the upregulation of the fatty acid metabolism genes *Cd36* and *Fabp4* in the heart likely reflects the importance of fatty acids to meet the bioenergetic demands of cardiomyocytes^{63,174}. Similarly, the upregulation of immune and stress response genes in the lung endothelium is expected due to the lung's continuous exposure to environmental stressors and pathogens contained in the inhaled air^{175,176}.

However, the adaptation of the endothelium appears to extend far beyond the supply of metabolites and nutrients to the parenchyma. We surprisingly found that there exists a multidirectional molecular crosstalk of vessel wall cells with the cells of their microenvironment. In the brain endothelium, synapse organization and neurotransmitter transport genes such as glutamine synthase (*Glul*) were highly enriched, which discloses the molecular mechanisms underlying how excitatory neurotransmitters such as glutamate can be transported among brain endothelial cells, neurons, and astrocytes^{177,178}. We also found that lung ECs expressed genes typically found in the lung epithelium such as Surfactant Protein C (*Spc*) and Mucin1 (*Muc1*), again indicative of a key interaction of the lung endothelium with the lung parenchymal epithelium. The upregulation of genes involved in cardiomyocyte contraction such as *Myl2* and *Ckmt2* again points to an unexpected adaptation of the cardiac endothelium to the surrounding cardiomyocytes, possibly suggesting a key role for the endothelium in modulating cardiac contractility^{179,180}.

One of the requisites for targeted therapies is the need to deliver such agents to specific organs, thus underscoring the importance of leveraging organ-specific endothelial heterogeneity for such approaches. It has been suggested that vascular endothelial cells in different organs or disease states express specific markers, or 'zip codes'¹⁸¹, so that ligands directed against organ-specific vascular endothelial cell surface markers could be used to deliver effector molecules to specific vascular beds. To address this concept, we expanded our analysis by analyzing 1296 cell surface glycoproteins, including 136 G-protein coupled receptors and 75 membrane receptor tyrosine-protein kinases. This allowed us to establish EC surface markers that were specifically upregulated in each vascular bed. Not only was this integrative analysis valuable for the establishment of EC 'zip codes' based on the organs they are derived from, but it may also provide insights about tissue-specific cell-cell contacts of ECs that allow them to interact with niche or parenchymal cells in each tissue^{182,183}.

Among the most intriguing findings of our study was the prominent 'parenchymal' signature of endothelial cells in each organ such as contractile genes in the cardiac endothelium and neurotransmitter transport or synaptic vesicle genes in the brain endothelium. A rank-based statistical analysis demonstrated that only selected genes of surrounding parenchymal cells were expressed in the endothelium of each organ. In the setting of a possible contamination, the most abundant genes expressed in the surrounding cells would also be the most abundant genes found in the cell of interest. That the rank order of parenchymal genes abundance in the endothelium differed from

that found in the parenchyma suggests tissue-specific programming and adaptation of the endothelium.

To further address the concern of possible mRNA contamination by neighboring cells in the RiboTag^{EC} data, we systematically analyzed two independent endothelial single cell RNA-Seq datasets^{164,165}, which can exclude contaminating tissue cells by examining the identity of each sequenced cell. We found that EC signature genes identified by our RiboTag^{EC} approach such as the synaptic vesicle gene Snap47 and cardiac contractile gene Tropomyosin were also expressed in individual brain and heart ECs as identified by scRNA-Seq. Importantly, we found a substantial overlap of individual signature genes across our data and both scRNA-Seq datasets. Even though the approaches to obtain the data were so different, this is a remarkable degree of consilience. We used a genetic VE-cadherin-Cre to label endothelial ribosomes, whereas the Tabula Muris scRNA-Seq dataset relied on mRNA markers of endothelial cells, and the Betsholtz dataset used Claudin5 lineage tracing combined with endothelial gene expression markers to identify individual ECs.

Although the bulk of scRNA-Seq tissue-specific genes were found in the Ribotag dataset, the converse was not true. Not all RiboTag^{EC} signature genes were present in the single cell RNA-Seq datasets. We think this likely reflects the greater depth and sensitivity of Ribotag RNA-Seq because current single cell technologies are limited in their ability to detect the expression of individual genes in a given cell¹³⁸⁻¹⁴². Not all single ECs expressed parenchymal genes such as Tropomyosin or Snap47 but those expressing them did so at an even

higher levels than what we found in the RiboTag^{EC} data. The reason for this might be that RiboTag^{EC} data represent an aggregate of all ECs in a tissue. It is therefore possible that the tissue adaptation of individual ECs may be most prominent in anatomically distinct ECs, for example those in close proximity to parenchymal cells such as neurons and astrocytes. Furthermore, if the expression of parenchymal gene signatures such as synaptic vesicle genes or cardiac contractile genes in the endothelium is dependent on environmental cues from neighboring cells or the extracellular matrix, the disassociation of the cells required for single cell RNA-seq may have further reduced mRNA levels of these genes^{120,143,144}. Sequencing a larger number of individual ECs in these tissues may enable identification of additional EC subsets with the most prominent parenchymal signatures, and a single cell sequencing approach that preserves the anatomy of the tissue such as Slide-Seq¹⁸⁴ may also be useful to address the in situ transcriptomic signature.

Using the RiboTag model, we were able to characterize the endothelial translome profile from distinct tissues. Our analysis uncovered a previously unrecognized degree of endothelial plasticity and adaptation to the parenchymal tissues, raising intriguing questions about the role that the endothelium plays in modulating parenchymal tissue function that likely go far beyond the classically ascribed roles of supplying oxygen, metabolites and solutes. Further studies such as endothelial-specific deletion of neurotransmitter transport or cardiac contractile genes will be required to establish the functional roles of these tissue-specific genes expressed in the endothelium of each organ. Understanding the

biological significance of endothelial plasticity and adaptation to the parenchyma will be important in providing a fuller picture of endothelial function during homeostasis and stress in each tissue.

CHAPTER 4

TISSUE-SPECIFIC ENDOTHELIAL GENE EXPRESSION DYNAMICS DURING INFLAMMATION

Parts have been previously published as:

Jambusaria, A., et al. (2020). "Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation." *Elife* 9.

4.1 Introduction

The heterogeneity of vascular endothelial cells across tissue types represents a major challenge for studying biological mechanisms of tissue-specific vascular disease as well as for therapeutic targeting of distinct tissues because the vascular endothelium represents an important gateway for delivering therapeutic drugs to the various tissues and organs⁸⁹. Endothelial cells that reside in different tissues or organs have distinct functions that are likely driven by unique gene expression signatures which reflect their tissue-specific adaptation and function. Computational methods have been developed and applied to identify tissue-specific gene regulatory networks from transcriptomic data may provide important insights into the mechanisms underlying the cellular heterogeneity of cells in distinct organs and tissues⁹³. In order to identify the molecular signature of distinct cell populations, new methods, in addition to existing methods such as GSEA⁹⁷ and PGSEA¹⁰⁷, have been developed to interpret dynamic changes within a group of genes with common function. In Chapter 2, we proposed a novel approach, HeteroPath, which assesses heterogeneously upregulated and downregulated genes within the context of pathways to provide a more

comprehensive understanding of the molecular underpinnings of tissue-specific phenotypes.

Recent advances in RNA-Sequencing make time-series (longitudinal) analyses more affordable and provide a window in the dynamic changes in vascular injury or disease. In our work, we used the RiboTag^{EC} (*Cdh5*^{CreERT2/+}; *Rpl22*^{HA/+}) mouse model to directly isolate endothelial mRNA without requiring endothelial cell isolation to minimize any changes in mRNA levels that might occur during the cell disassociation process¹⁴⁷. We injected RiboTag^{EC} mice with the bacterial endotoxin lipopolysaccharide (LPS) to induce systemic vascular endothelial inflammation and injury. We measured gene expression levels using RNA-Seq at several time points namely, 0 hr, 6 hr, 24 hr, 48hr, 72hr, and 1-week post injury. Computational analysis of the RNA-Seq data from the *in vivo* time series experiments led to a selection of candidate signaling mechanisms that are activated during the LPS-induced endothelial injury which can be validated with biochemical studies.

One of the major questions in the context of cellular heterogeneity is whether cellular responses to biological stimuli and stressors routinely involve widespread changes in tissue-specific gene expression¹⁸⁵. It is thought that biological gene expression networks are highly modular and contain dense communities of genes which work in concert to regulate cellular processes¹⁸⁶, but these modular communities of genes change dynamically as cells respond to stimuli over time. While some methods have been developed to infer networks and analyze their topological variation, there is a need to develop methods which

evaluate biological network reconfiguration over time. Therefore, we investigated genome-scale transcriptomic networks derived from distinct time points to identify tissue-specific subnetworks mediating response to external stimuli.

To be precise, we specifically studied whether the dynamics of tissue-specific gene expression in response to a stimulus can be characterized by shifts in network signaling entropy. When cells are challenged with an external stimulus, cells typically respond with shifts in gene expression patterns which return to a steady-state level over time. This endothelial injury-response pattern during systemic inflammation is different from other classes of temporal responses where gene expression changes in an oscillatory fashion, for example as classically observed during cell cycle. We propose that by integrating the gene expression fold-change data we produce with database-derived protein-protein interaction (PPI) networks, we can determine the relationship between the changes in network signaling entropy and endothelial phenotype. The use of fold-change networks (FCNs) is a novel approach to study signaling pathways in response to stressors. In our novel framework, Subnetwork Signaling Entropy Analysis (SSEA), we quantitatively characterize fold change subnetwork dynamics using network signaling entropy to quantitatively monitor gene regulatory network dynamics in biological systems.

In Chapters 2 and 3, we have established the molecular signatures at baseline of the distinct endothelial tissues. Therefore, being able to track the tissue-specific endothelial response to systemic inflammatory stimulation would help to understand the molecular mechanisms mediating the endothelial

dysfunction. To do this, we designed a network-theoretical framework to calculate network signaling entropy for time series endothelial transcriptomic data and interpret basic endothelial biology underlying vascular disease. To precisely study the endothelium, we used the RiboTag system and isolated *in situ* endothelial-specific mRNA actively being translated from brain, heart, and lung tissue from a time series ALI model induced by a bacterial endotoxin, lipopolysaccharide. By pairing this injury model with RNA-Seq, we were able to explore the tissue-specific molecular dynamics in mice during endothelial injury, repair, and recovery. Identifying the individual gene and network-based biomarkers provides a better understanding of the tissue-specific signaling mechanisms mediating the disruption and repair of the endothelial barrier.

We propose a generalized approach termed Subnetwork Signaling Entropy Analysis (SSEA) to evaluate cellular heterogeneity from transcriptomics data and produced comprehensive tissue-specific signatures. We hypothesize that the network signaling entropy dynamics of subnetworks derived from genome-scale fold-change networks represent the observed tissue-specific phenomena. We believe that during stimulation, genes that experience similar magnitude of gene expression change are more likely to be functionally related (fold-change networks) than genes which have similar expression levels at distinct time points (co-expression networks). We believe that using fold-change networks serves to provide insights about the manifestation of tissue-specific biological mechanisms in response to an external stimulus. In this work, we evaluate network signaling entropy to understand tissue-specific inflammatory

injury as well as determine the distinct states of endothelial injury and repair within a single tissue. We specifically demonstrate the power of assessing network signaling entropy to characterize cellular states of the brain endothelium during lipopolysaccharide (LPS) induced vascular injury and repair.

We first identified differentially expressed genes between each time point and baseline. We then integrated gene expression fold-change data with a database-derived protein-protein interaction (PPI) networks, to demonstrate the relationship between the changes in network signaling entropy and cellular phenotype. The use of fold-change networks (FCNs) is a novel approach we applied to study signaling pathways in response to stressors. In our novel framework, Subnetwork Signaling Entropy Analysis (SSEA), we quantitatively characterized fold change subnetwork dynamics using network signaling entropy to quantitatively monitor gene regulatory network dynamics in biological systems. We used the RiboTag system to isolate in situ endothelial-specific mRNA actively being translated from brain, lung, and heart tissue from a time series ALI model induced by a bacterial endotoxin, LPS. By pairing this model with RNA-Seq, we explored tissue-specific dynamics in mice which underwent endothelial injury, repair, and recovery. The individual genes and gene networks we uncovered provided us with a better understanding of the biomarkers and tissue-specific signaling mechanisms mediating the disruption and repair of the endothelial barrier during systemic inflammation.

4.1.1 Problem Definition

The circulating bacterial endotoxin lipopolysaccharide (LPS) is a key mediator of tissue inflammation and injury in patients with bacteremia and sepsis^{187,188}, therefore we exposed RiboTag^{EC} mice to LPS to induce systemic inflammatory injury and study the organ-specific endothelial transcriptome response. The tissue-specific heterogeneity of the endothelium during homeostasis is maintained during systemic in vivo inflammatory injury as evidenced by the distinct responses to inflammatory stimulation. Our study defines endothelial heterogeneity and plasticity during systemic inflammation and provides a molecular framework to understand organ-specific vascular disease mechanisms and therapeutic targeting of individual vascular beds.

4.2 Methods

4.2.1 Experimental Animals

RiboTag (*Rpl22^{HA/+}*) mice were purchased from Jackson Labs. Endothelial-specific VE-cadherin-Cre mice were provided by Dr. Ralf Adams. We crossed the RiboTag mice (*Rpl22^{HA/+}*)¹⁴⁷ with the endothelial-specific VE-cadherin-Cre mice^{148,149} to generate RiboTag^{EC} (*Cdh5^{CreERT2/+}; Rpl22^{HA/+}*) mice. Following tamoxifen-induced recombination at week 4, HA-tagged *Rpl22* was specifically expressed in endothelial cells. To investigate the mechanisms of organ-specific EC injury, repair, and regeneration we performed RNA-Seq analysis of gene expression in ECs isolated at 6 hr, 24 hr, 48 hr, 72 hr, and 1 week post-LPS challenge (10 mg/kg LPS i.p., Sigma-Aldrich Cat#: L2630) with PBS-injected mice serving as controls.

All animal experiments were conducted in accordance with NIH guidelines for the Care and Use of Laboratory Animals and were approved by the IACUC of the University of Illinois (IACUC Protocol #19–014, IACUC Protocol #13–175 and IACUC Protocol #16–064).

4.2.2 Assessing Endothelial Heterogeneity using the RiboTag method

In order to define cell populations, distinct cell populations must be isolated from physiological tissue, suspended, and then sorted by flow cytometry or magnetic bead sorting. To isolate the cells from an intact tissue, different levels of mechanical processing and enzymatic tissue digestion are required. By nature, these processing steps and use of enzymes manipulate the sample and introduce artifacts.

To bypass cell retrieval and obtain an in vivo snapshot of the tissue, new approaches have been developed to directly isolate cell-type-specific translomes by immunoprecipitation (IP) of tagged ribosomes from crude tissue extracts¹⁴⁷. The RiboTag strategy established by Sanz et al.¹⁴⁷ recommends the use of cell-type-specific Cre recombinase to activate expression of a epitope tagged to a ribosomal subunit (RPL22). By immunoprecipitating the epitope-tagged ribosomes out from the whole tissue, we can specifically isolate the cell-type-specific translome.

Here we report the application of the RiboTag approach to the study of endothelial-specific mRNA. We chose to use the Cdh5-Cre animals because their endothelial specific expression¹⁴⁸ would be ideal for profiling ECs using the RiboTag strategy (Figure 23). After the isolation of endothelial specific mRNA, we

performed a transcriptomic analysis on the transcriptome of brain, heart, and lung endothelium from animals challenged with the bacterial toxin LPS.

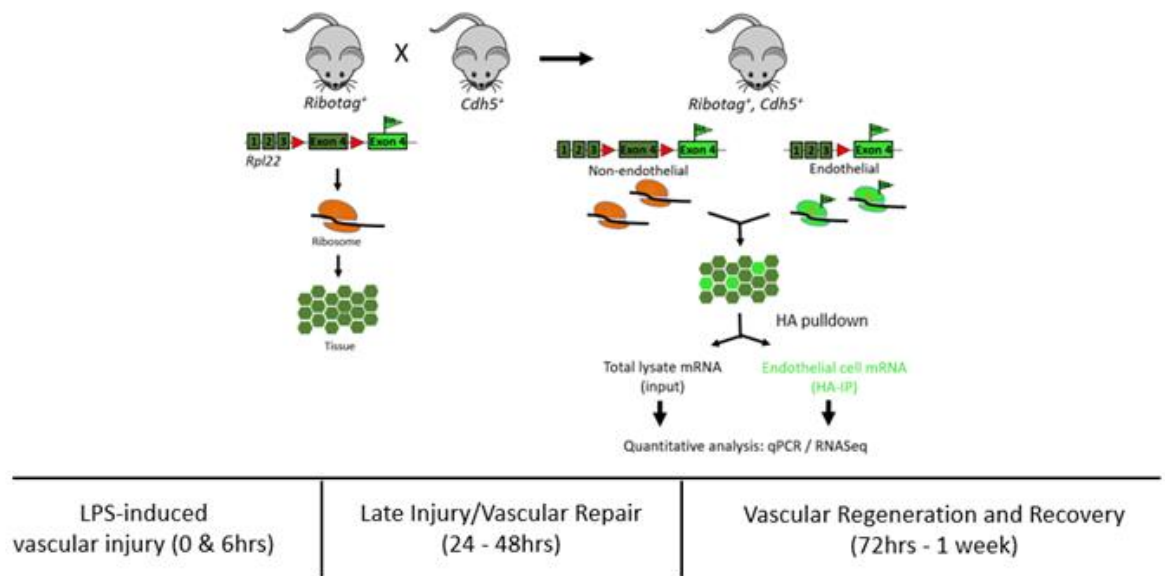


Figure 23: RiboTag mRNA isolation and time-series transcriptomic analysis. The RiboTag procedure allows simple and efficient isolation of ribosome-associated mRNAs from endothelial cells in complex tissues, including brain, lung, and heart.

4.2.3 Tissue-specific endothelial Kinetics following LPS-induced injury

To ascertain the kinetics of the tissue-specific endothelial signatures during inflammation we analyzed the time-series RNA-Seq data with the gene sets referenced in Chapter 2 and Chapter 3: classical endothelial markers, glycolysis, fatty acid metabolism, and transport. To visualize the tissue-specific dynamics for predominant endothelial functions, we plotted a heatmap which includes the tissue-specific differentially expressed genes for each gene set.

4.2.4 Early and late tissue-specific inflammatory markers

To identify the inflammatory genes that were upregulated in the LPS 6 hr samples as compared to the baseline samples, we applied the standard *limma* pipeline¹⁵⁴ for genes in the ‘inflammatory response’ gene ontology term (GO:0006954). The analysis was carried out on the tissue specific LPS treated samples against the baseline tissue-specific sample. *Limma* statistically evaluates each inflammatory gene and returns the genes which show statistically significant change between the inflammatory time point and baseline. We applied this approach to the early inflammation time point, 6 hr, and the late inflammatory time point, 24 hr. Heatmaps were generated to visualize the tissue-specific inflammatory genes and their kinetics.

4.2.5 Online endothelial translome expression database

The endothelial translome expression database (<http://www.rehmanlab.org/ribo>) is hosted on Amazon S3. The website was constructed using Angular 8.0, JavaScript, HTML5, and CSS. Bar plots and heatmaps were generated for genes of interest using Tableau Public. The

visualizations were integrated into the web application using the Tableau JavaScript API. RiboTag log₂ normalized baseline and inflammation time-course translome expression data were uploaded to Tableau. The averages were computed using Tableau calculated fields. Tableau dashboards and workbooks were created to generate bar plots and heatmaps for online publishing.

4.2.6 Genome-scale Co-expression and Fold Change Networks (FCN)

Integrating gene expression data with network data such as a protein interaction network may lead to novel discoveries about cellular states and dynamics. Let us say, for example, if two genes over a time series possess similar expression levels or similar changes in expression level, then these two genes may be functionally related and interact with each other. Furthermore, a gene set derived by clustering genes based on their fold change may identify novel signaling mechanisms that are modulated during the experiment¹⁸⁹.

We accumulated and preprocessed the RNA-Seq raw data from the brain, heart, and lung tissue from RiboTag+/Cdh5+ mice to construct co-expression and fold-change networks. The construction and analysis of the co-expression network were based on WGCNA¹⁹⁰, which is a typical algorithm (Figure 24). In the co-expression network, nodes are genes, and edges indicate the magnitude of their co-expression at 6hr post-LPS challenge.

To observe the effect of fold-change in expression of genes during the experimental condition, fold change networks were constructed. STRING¹⁹¹ interactions (confidence score>500) were used to construct the networks. In the fold change network of genes, nodes are genes, and edges indicate the

magnitude of the Pearson correlation of the fold changes calculated for each gene between 0hr and 6hr post treatment.

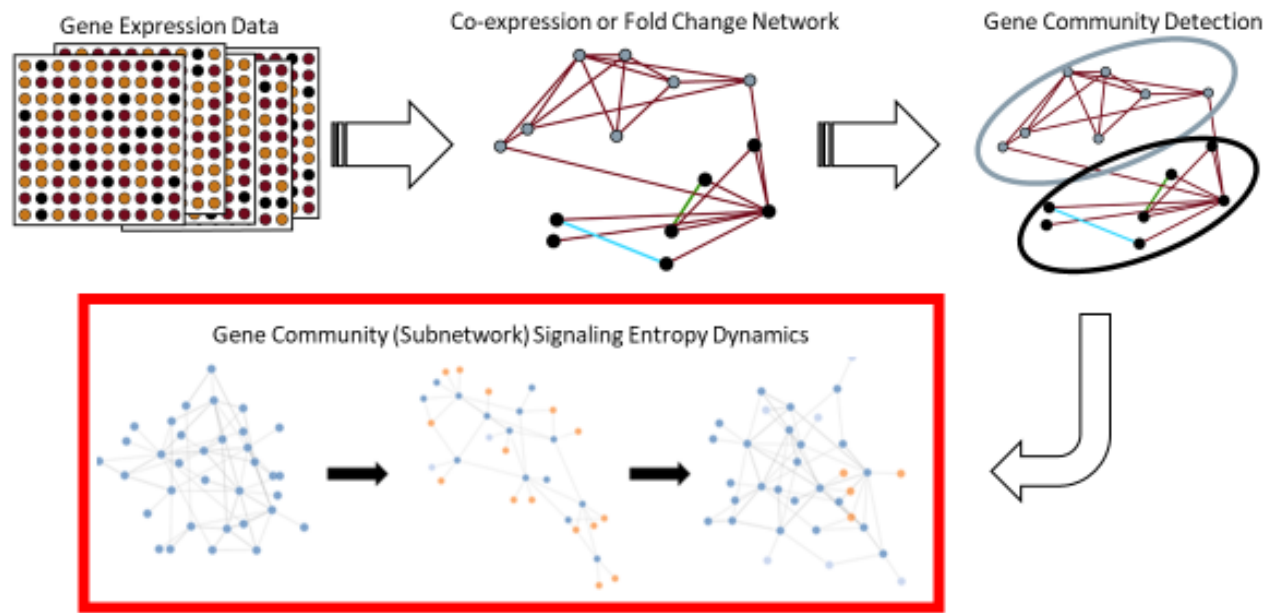


Figure 24: Subnetwork Signaling Entropy Analysis (SSEA). A schematic of the overall framework to identify tissue-specific subnetworks that are dysregulated in response to an external stimulus.

4.2.7 Gene community detection in the networks

We used the Louvain method^{192 193} to find the community structure of the co-expression and fold change networks. This approach is a greedy optimization method which optimizes the local modularity of a subnetwork within the network¹⁹³. Essentially, the modularity quantifies network structure by evaluating the number of edges within a cluster of genes and the number of edges that would be found in that cluster if it were derived from a random network^{194 195}. It is given by:

Equation 4

$$Q = \sum_{s=1}^{NC} \frac{E_s}{E} - \frac{\sum_s k_s^2}{2E}$$

Where Q varies between 0 and 1. Maximizing Q results in optimal clustering. In this equation, NC = number of clusters in the network, E = number of edges, E_s = the number of edges in the gene community s , and k_s = total degrees across all nodes in gene community s .

The Louvain approach is composed of two parts. In part 1, the algorithm strives to improve the network modularity, Q , in a local way to search for smaller gene communities. It then generates a new network by grouping genes from the same community, s , into a single node. This process is iterated through until the maximum value of modularity is achieved.

4.2.8 Network signaling entropy for subnetworks identified by community detection algorithms

The derivation of network signaling entropy heavily relies on the robustness of the PPI network which the gene expression data is superimposed on. After overlaying the gene expression data on the PPI, the edge weights are calculated by approximating the gene-gene interaction probabilities between the corresponding nodes in the graph. Signaling entropy is calculated by first estimating the stochastic matrix and then embedding the node interaction weights over the entire network.

Edge weights w_{ij} between nodes i and j are derived from the co-expression or fold change values of the genes i and j . The assumption here is that genes with similar expression level or changes in expression level over a time course are more likely to be functionally related rather than two genes which are not co-expressed or do not undergo similar changes in expression level over time. The underlying PPI network is required to separate interactions which are biologically observed rather than discovered by random chance. The edge weights between two nodes in the graph can be defined using several different quantification methods. Here, in our case, we used simple Pearson correlations. We define the weights as:

Equation 5

$$w_{ij} = \frac{1}{2}(1 + c_{ij})$$

With c_{ij} being the correlation between genes i and j . This definition of the correlation assumes that the positive correlation corresponds to an activating

interaction and a negative correlation corresponds to an inhibiting interaction. This assumption becomes very important because positive and negative interactions manifest in distinct downstream cellular signaling. This method to calculate edge weights does not assign edges to two nodes with an insignificant correlation. Once the gene-gene edge weights were calculated and normalized, they were then assigned to the stochastic matrix p_{ij} ,

Equation 6

$$p_{ij} = \frac{w_{ij}}{\sum_{K \in N_i} w_{ik}}$$

With N_i representing the PPI neighboring nodes of node i . Thus, p_{ij} can be defined as the signaling probability between nodes i and j . Notably, P is not a symmetric matrix because p_{ij} is not necessarily doubly stochastic. The bi-directional edges are assigned unique correlation-based weights to assign a quantitative metric for the transduction of information from node i to j , and vice versa ($p_{ij} \neq p_{ji}$).

Given a stochastic matrix, p_{ij} , the Shannon entropy for each individual gene i can be calculated using the following equation:

Equation 7

$$S_i = \frac{-1}{\log k_i} \sum_{K \in N_i} p_{ik} \log p_{ik} ,$$

Where k_i denotes the degree of gene i in the PPI network. The normalization is not required but scales the Shannon entropy for all genes to be between 0 and 1. If there exists only one edge weight in the network which is non-zero, then the entropy is equal to the lowest uncertainty value of zero. On the other hand, if all edge weights from a single node i are equal, the Shannon entropy is equal to the

greatest uncertainty and highest promiscuity of 1. Therefore, we can use the Shannon entropy metric to analyze distinct phenotypes and identify dysregulated genes or interactions in the signaling network. When Shannon entropy is estimated on the phenotypic scale, resampling approaches can be applied to generate differential entropy statistics. Deriving the entropy statistics is a quantitative approach for studying the association between factors such as node degree and entropy variance. When sample-specific network signaling entropies are calculated statistical tests such as rank-sum tests can be applied to generate robust *P*-values.

4.3 Experiments & Results

4.3.1 In situ organ-specific endothelial early- and late-inflammation signature

We first analyzed the dynamics of the EC inflammatory response in each tissue, focusing on the early response (6 hr post systemic LPS) and late response (24 hr post systemic LPS). At these time points, we identified the genes most upregulated by inflammatory injury in each tissue (Figure 25). In the brain endothelium, we identified several differentially expressed acute inflammatory factors including selectins, chemokine receptors, and interleukins which were strongly activated 6 hr post LPS treatment (Figure 26A–C). We analyzed the kinetics during the entire time course for the early inflammatory brain endothelial specific genes such as eosinophil chemotactic protein (*Ccl11*) (Figure 26C) and found that *Ccl11* is markedly upregulated at the 6 hr time point and remains significantly higher in the brain endothelium, but by one week post LPS injection

the expression level returns to the same level as that seen in lung and heart endothelium. In the lung endothelium, we discovered that the most upregulated inflammatory pathways included chemokines, response to cellular stress, hematopoiesis genes and early immune response mediators (Figure 26D–F). Lymphocyte antigen 96 (*Ly96*) was strongly upregulated (Figure 26D) whereas the apoptosis gene caspase 6 (*Casp6*) was markedly downregulated 6 hr post LPS treatment and remained lower in lung ECs than in brain or heart ECs throughout the injury period (Figure 26F). In heart ECs, leukocyte migration and neutrophil activation pathways were most upregulated by inflammatory injury (Figure 26G–I). At 24 hr post injury, we found the peak upregulation of inflammatory genes (Figure 27) with a substantial overlap of the inflammatory response pathways, predominantly associated with neutrophil and leukocyte chemotaxis and migration, in the brain (Figure 27A–C), lung (Figure 27D–F), and heart ECs (Figure 27G–I).

A

Top RiboTag Brain EC Early Inflammatory Markers	
Timp1	TIMP metalloproteinase inhibitor 1
Ccl11	C-C motif chemokine 11
Serpina3n	Alpha 1-antichymotrypsin
Acod1	Aconitate Decarboxylase
C3	Complement C3
Tnf	Tumor Necrosis Factor
Il6	Interleukin 6
Il1rn	Interleukin 1 Receptor Accessory Protein
Il1b	Interleukin 1 beta
Il1a	Interleukin 1 alpha

B

Top RiboTag Lung EC Early Inflammatory Markers	
Gata3	GATA Binding Protein
Ptgis	Prostaglandin I2 Synthase
Wfdc1	WAP four-disulfide core domain protein 1
Ly96	Lymphocyte antigen 96
Il1r2	Interleukin 1 receptor, type II
Casp6	Caspase-6
Ctnnbip1	Beta-catenin-interacting protein 1
Cxcl1	chemokine (C-X-C motif) ligand 1
Cxcl9	Chemokine (C-X-C motif) ligand 9
Saa3	Serum amyloid A

C

Top RiboTag Heart EC Early Inflammatory Markers	
Ccl3	Chemokine (C-C motif) ligand 3
Timp1	TIMP metalloproteinase inhibitor 1
Acod1	Aconitate Decarboxylase
Il6	Interleukin 6
Selp	P-selectin
Sele	E-selectin
Il1rn	Interleukin 1 Receptor Antagonist
Tspan2	Tetraspanin-2
Cxcl5	C-X-C motif chemokine 5
Cxcl1	C-X-C motif chemokine 1

D

Top RiboTag Brain EC Late Inflammatory Markers	
Ccl3	Chemokine (C-C motif) ligand 3
Timp1	TIMP metalloproteinase inhibitor 1
Ccl11	Chemokine (C-C motif) ligand 11
Serpina3n	Alpha 1-antichymotrypsin
Acod1	Aconitate Decarboxylase
C3	Complement C3
Il6	Interleukin 6
Il18rap	Interleukin 18 receptor accessory protein
Selp	P-selectin
Sele	E-selectin

E

Top RiboTag Lung EC Late Inflammatory Markers	
Mmp8	Neutrophil collagenase
Il10	human cytokine synthesis inhibitory factor
Acod1	Aconitate Decarboxylase
Mefv	pyrin innate immunity regulator
Wfdc1	WAP four-disulfide core domain protein 1
Pla2g7	platelet-activating factor acetylhydrolase
Il1r2	Interleukin 1 receptor
Il1rn	interleukin-1 receptor antagonist
Cxcl9	Chemokine (C-X-C motif) ligand 9
Siglece	sialic acid binding Ig-like lectin E

F

Top RiboTag Heart EC Late Inflammatory Markers	
Serpina3n	Alpha 1-antichymotrypsin
Itih4	Inter-alpha-trypsin inhibitor heavy chain H4
Acod1	Aconitate Decarboxylase
Apod	Apolipoprotein D
Il6	Interleukin 6
Selp	P-selectin
Sele	E-selectin
Il1rn	Interleukin 1 Receptor Antagonist
Cxcl5	C-X-C motif chemokine 5
Cxcl3	C-X-C motif chemokine 3

Figure 25: Markers of early (6 hr) and late (24 hr) LPS-induced inflammation in brain, lung, and heart ECs. A) Top RiboTag brain EC early (6 hr) inflammatory markers B) Top RiboTag lung EC early (6 hr) inflammatory markers C) Top RiboTag heart EC early (6 hr) inflammatory markers D) Top RiboTag brain EC late (24 hr) inflammatory markers E) Top RiboTag lung EC late (24 hr) inflammatory markers F) Top RiboTag heart EC late (24 hr) inflammatory markers.

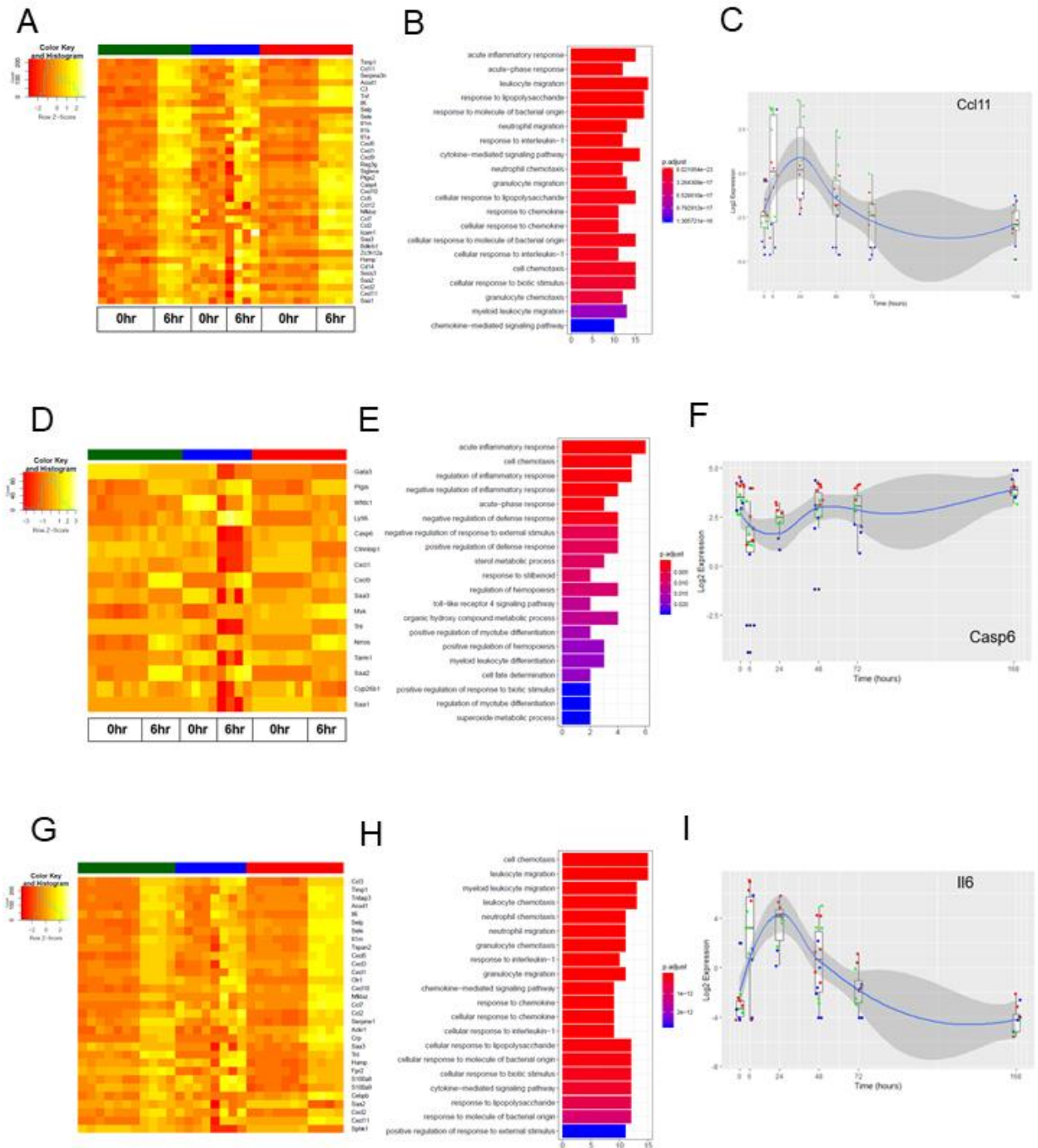


Figure 26: The early inflammation (6 hr) markers across organ-specific endothelial cells. A) Heat map representation of differentially expressed genes identified by comparing brain ECs to lung and heart ECs at the 6 hr time point. The orange to yellow to white gradient represents increasing expression of the

pathway with orange representing minimal expression while the white represents high expression of the pathway. B) The GSEA results of enriched GO terms from RiboTag brain ECs at the 6 hr time point. C) Tissue-specific kinetics of a specific RiboTag brain EC early inflammatory marker during the progression and resolution of inflammation. D) Heat map representation of differentially expressed genes identified by comparing lung ECs to brain and heart ECs at the 6 hr time point. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. E) The GSEA results of enriched GO terms from RiboTag lung ECs at the 6 hr time point. F) Tissue-specific kinetics of a specific RiboTag lung EC early inflammatory marker during the progression and resolution of inflammation. G) Heat map representation of differentially expressed genes identified by comparing heart ECs to brain and lung ECs at the 6 hr time point. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. H) The GSEA results of enriched GO terms from RiboTag heart ECs at the 6 hr time point. I) Tissue-specific kinetics of a specific RiboTag heart EC early inflammatory marker during the progression and resolution of inflammation.

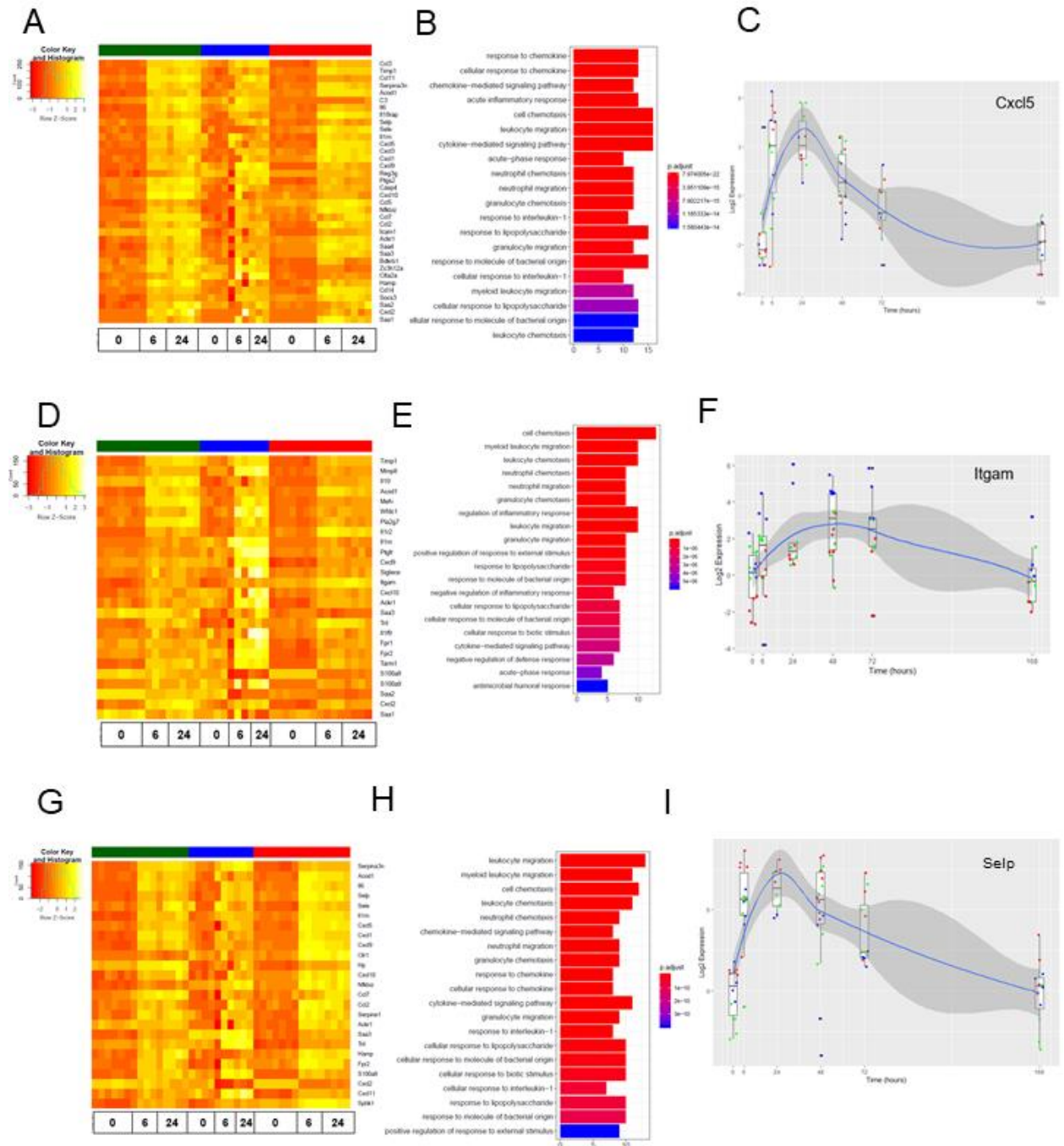


Figure 27: The late inflammation (24 hr) markers across organ-specific endothelial cells. Heat map representation of differentially expressed genes identified by comparing brain ECs to lung and heart ECs at the 24 hr time point. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents

high expression of the pathway. B) The GSEA results of enriched GO terms from RiboTag brain ECs at the 24 hr time point. C) Tissue-specific kinetics of a specific RiboTag brain EC late inflammatory marker during the progression and resolution of inflammation. D) Heat map representation of differentially expressed genes identified by comparing lung ECs to brain and heart ECs at the 24 hr time point. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. E) The GSEA results of enriched GO terms from RiboTag lung ECs at the 24 hr time point. F) Tissue-specific kinetics of a specific RiboTag lung EC late inflammatory marker during the progression and resolution of inflammation. G) Heat map representation of differentially expressed genes identified by comparing heart ECs to brain and lung ECs at the 24 hr time point. The orange to yellow to white gradient represents increasing expression of the pathway with orange representing minimal expression while the white represents high expression of the pathway. H) The GSEA results of enriched GO terms from RiboTag heart ECs at the 24 hr time point. I) Tissue-specific kinetics of a specific RiboTag heart EC late inflammatory marker during the progression and resolution of inflammation.

4.3.2 Tissue-specific dynamic response following LPS-induced inflammatory activation

After establishing the baseline heterogeneity of brain, lung and heart ECs, we next studied the dynamics of the organ-specific baseline endothelial signature during systemic inflammation, we collected transcriptome data of the brain, lung, and heart endothelium at several time points following LPS treatment. By computationally analyzing RiboTagEC mRNA from brain, lung, and heart at 0 hr, 6 hr, 24 hr, 48 hr, 72 hr, and 168 hr post-LPS administration, we were able to identify tissue-specific molecular mechanisms modulated in endothelial injury, repair, and regeneration.

We first investigated the tissue-specific baseline signatures over time in order to address the question of whether the baseline core endothelial functions were disrupted during inflammatory activation. The time-course of the brain endothelium specific endothelial genes were plotted to compare their kinetics to the lung and heart endothelium (Figure 28A). We found that selected genes which constitute the tissue-specific EC signature during homeostasis are modulated during inflammatory injury. For instance, the expression level of von Willebrand factor A domain containing protein 1 (Vwa1) which we found to be a brain endothelial gene during homeostasis decreases during early and late inflammation and then returns to baseline levels one-week post LPS injury, whereas its levels in lung and heart endothelium remain relatively low during the entire time course. On the other hand, there are signature genes such as glucose

transporter protein 1 (*Slc2a1*) which is consistently upregulated in brain ECs throughout the post-injury period.

From the analysis of the lung endothelium specific endothelial genes heatmap (Figure 28B), it is apparent that expression of nearly all the canonical endothelial genes drastically decrease during the early and late inflammatory time points. This is an important finding because it suggests that the lung endothelium experiences the most profound dysregulation of core endothelial genes following LPS injury. We also identified lung endothelial specific genes which are solely modulated in the lung endothelium during the inflammatory time course. For instance, the expression levels of forkhead-related transcription factor 1 (*Foxf1*) and tetraspanin8 (*Tspan8*) significantly decrease in the lung endothelium at 6 hr and 24 hr post LPS treatment and then gradually recover back to baseline levels, but both genes remain minimally expressed in the brain and heart endothelium.

The endothelial genes which were specifically upregulated in the heart endothelium at baseline do not appear to be affected to the extent that the brain and lung endothelium were during LPS stimulation. In the heatmap (Figure 28C), a few genes such as Rho family GTPase 1 (*Rnd1*) and platelet glycoprotein (*Cd36*) undergo a robust change in expression during the time course. From our analysis, we found that the endothelial genes specific to the heart endothelium are much more abundant in the heart versus the other tissues. For example, caveolin 1 (*Cav1*) and vascular endothelial growth factor receptor 2 (*Kdr*) maintained a high expression level in the heart endothelial samples during the

entire LPS time course whereas in the brain and lung endothelial samples, we see significantly lower expression.

Since, endothelial cells rely heavily on glycolysis for ATP generation, we next focused on the organ-specific endothelial glycolysis signature to investigate the tissue-specific dynamics of glycolytic genes. The brain endothelial basal transcriptome upregulated the greatest number of glycolytic genes compared to the lung and heart endothelium. Interestingly, when we analyzed the time course of these brain endothelial specific glycolysis genes, we found that they maintain similar levels during the progression and resolution of inflammation (Figure 29). There were only three glycolysis-related genes which were upregulated in the lung endothelium. When we analyzed these three genes over time, we found that two of them remained stable whereas 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (*Pfkfb3*) was dynamic in all three tissues. Even though this glycolysis regulatory enzyme was specifically upregulated in the lung endothelium at baseline, we found that it was activated in all tissues during late inflammation/early repair and then returned to baseline levels (Figure 29B). In the heart endothelium, we found that the upregulated glycolytic genes were not modulated during the LPS injury and recovery (Figure 29C).

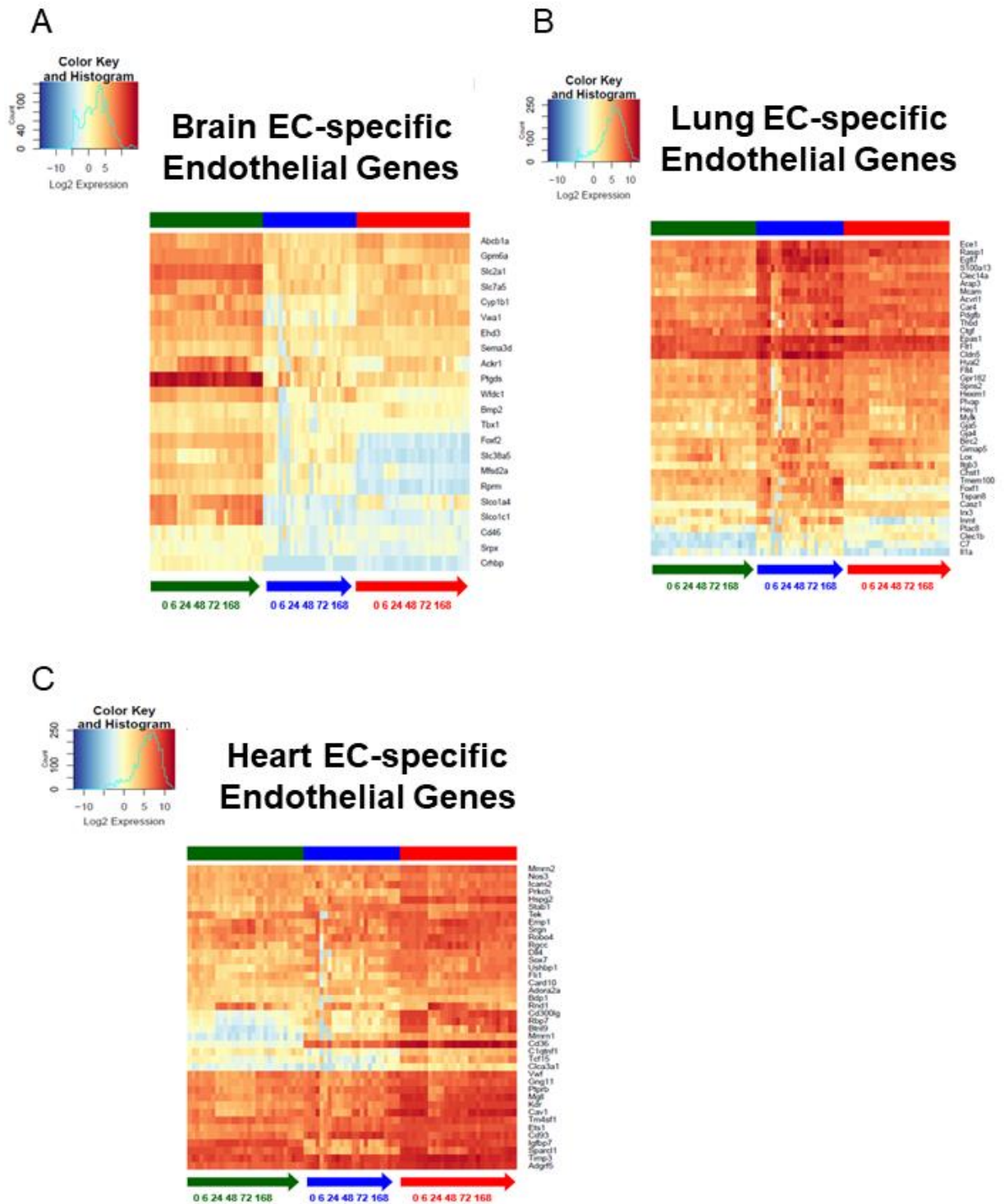


Figure 28: Organ-specific endothelial cells uniquely regulate endothelial genes during the progression and resolution of inflammation. A–C) Time-series heat map of significantly upregulated endothelial genes at baseline in A) brain ECs B) lung ECs and C) heart ECs. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway.

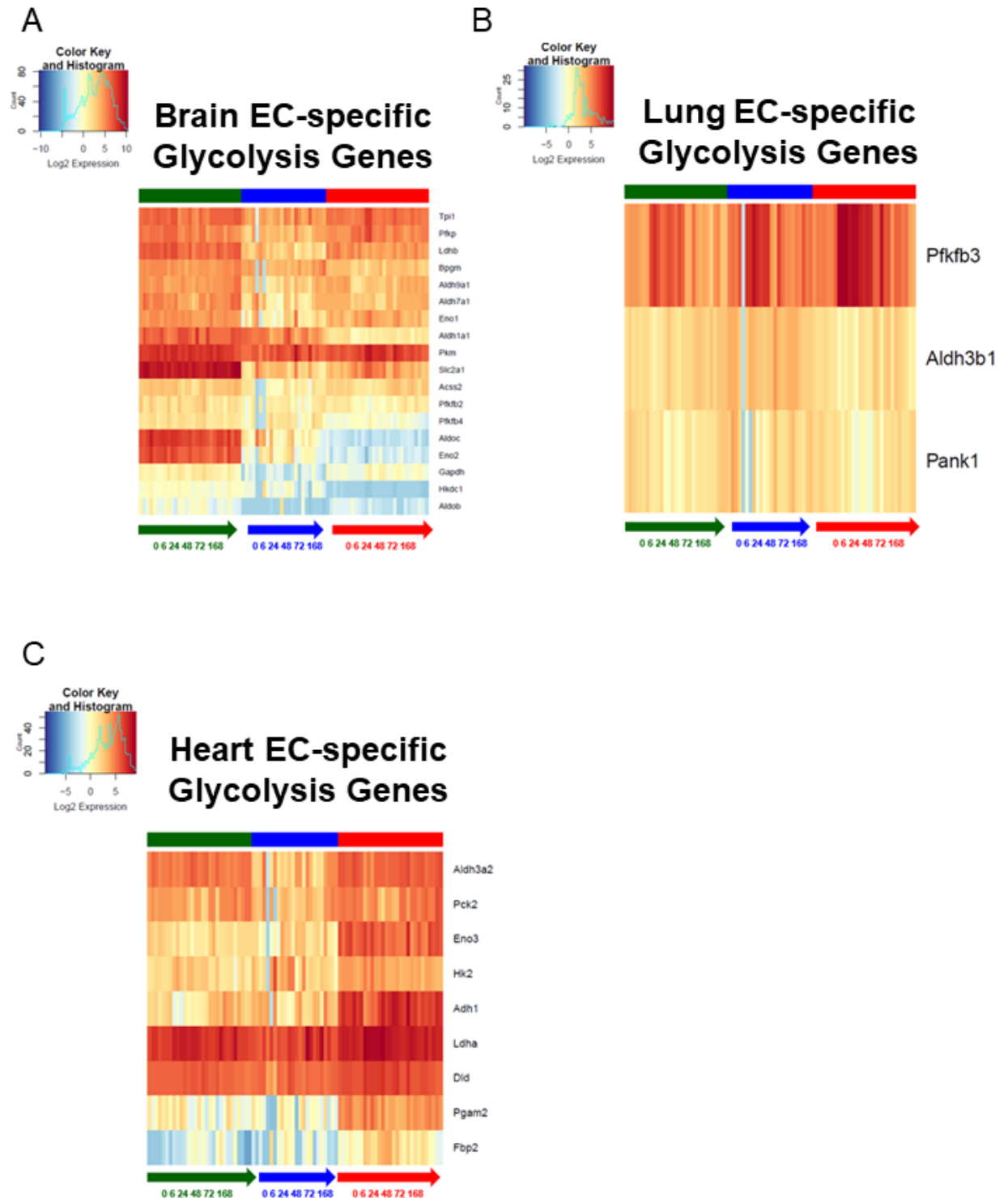


Figure 29: Organ-specific endothelial cells uniquely regulate glycolysis genes during the progression and resolution of inflammation. A–C) Time-series heat map of upregulated glycolysis genes at baseline in A) brain ECs B) lung ECs and

C) heart ECs. The blue to white to red gradient represents increasing expression of the pathway with blue representing minimal expression while the red represents high expression of the pathway.

4.3.3 Louvain Community Detection in Co-expression Networks versus Fold Change Networks

To assess co-expression and fold change networks, we compared the results of applying the Louvain clustering algorithm on the co-expression network versus the fold change network. We first performed the analysis without setting a limit on the cluster size and found that the Louvain algorithm is highly sensitive to the input order of the genes analyzed¹⁹¹. To account for this, we performed 10 independent runs and report the median results in Figure 30. Use of this algorithm in a fold-change network provides more robust results than in a co-expression network because the clusters are more defined and smaller in size. For example, Louvain clustering in a fold-change network resulted in clusters with the maximal size of approximately 1000 nodes, a number which is in the reasonable range for biological responses of defined transcription factor targets. When we directly clustered the co-expression network using the Louvain method, the largest cluster contained nearly one-third of the network (3000+ genes) making it difficult to interpret distinct gene response clusters. We next performed Louvain clustering by placing varying thresholds on cluster sizes. To be consistent with the “disease module identification” DREAM challenge, we forced the clusters to be between 3 and 100. Again, Louvain clustering on the fold change networks performs better than on the co-expression network, with bounded cluster sizes.

Method	Enriched Clusters	#NEC	%NEC	#NEC S	%NEC S
Co-expression Network	153/303 (50.50%)	2476.0	18.18%	871.0	7.61%
Fold Change Network	128/189 (67.72%)	1108.0	40.62%	645.0	25.11%

Figure 30: The performance of Louvain on the co-expression network versus the fold change network. NEC = “Nodes in Enriched Clusters”, %NEC = enrichment in the KEGG Pathways, and NEC S uses the same filtered KEGG Pathways. NEC S is the most meaningful because it reports results derived when we did not place a limit on the cluster size.

4.3.4 Identifying dysregulated subnetworks associated with the progression of LPS injury

The goal of network signaling entropy was to quantify the extent to which genes were influenced by the entire network, the gene community that the gene was assigned, and the other gene communities. We first constructed both a co-expression and fold change network including 9031 genes. For both networks, every gene in the network was initially assigned an entropy value. We first applied SSEA to our transcriptomic analysis of the endothelial transcriptome from animals exposed to bacterial toxin (LPS) challenge to demonstrate the effectiveness of the method in identifying dysregulated gene communities. We analyzed this dataset using fold change networks and co-expression networks.

Our novel method using fold change networks identified 21 gene communities with corrected p-values <0.01 (Figure 31). We then applied SSEA to the gene communities and found 10 significant pathways. When we compared SSEA on the fold change networks to the co-expression networks, we found that the two networks had five overlapping pathways. Interestingly, the pathways identified exclusively by SSEA in the fold change networks have previously been associated with the brain endothelium and LPS injury. For example, alterations of Wnt Signaling in brain endothelium plays an important role in the development of the blood brain barrier¹⁹⁶ while the focal adhesion pathway is linked to maintenance of the endothelial barrier¹⁹⁷. Furthermore, the pathways which were identified in both networks were more easily detected in the fold change communities than in the co-expression gene communities. These results suggest

that when SSEA was applied on the fold change networks, rather than the co-expression networks, it extracted meaningful biological subnetworks associated with brain endothelial injury.

To assess the performance of SSEA on co-expression subnetworks, we conducted analysis of the 6hr time point co-expression subnetworks in the RiboTag LPS dataset. The data matrix contained 9031 genes. We applied the new method SSEA to the data to identify subnetworks as well as the signaling entropy dynamics over the LPS time course for gene subnetworks that were co-expressed at 6hrs post LPS stimulation. The results of the subnetworks which exhibited significant changes in entropy during the time course, as well as their primary biological functions as defined by KEGG pathways are listed in Figure 32. In the co-expression subnetworks, signaling entropy was found to be primarily associated with cell growth and cell death pathways, signal transduction pathways, and metabolic pathways.

It is interesting to note that the co-expression subnetworks at the 6hr post LPS stimulation time point, which is canonically known to be the most severe inflammatory state, signaling entropy captures activated signaling transduction pathways and metabolic processes rather than pathways activated in response to LPS treatment which were identified by the fold-change networks. It appears that the signature of gene subnetworks responsible for responding to the inflammatory stimulation, exhibit a similar extent of fold-change in response to the stimulus rather than exhibiting similar gene expression levels.

Pathway	SSEA Fold Change Network FDR	SSEA Co- expression Network FDR	SSEA Fold Change Rank	SSEA Co- expression Rank
Wnt signaling pathway	<0.005	0.70	1	54
MAPK Signaling Pathway	<0.005	0.81	2	77
Focal Adhesion	<0.005	0.44	3	33
ECM-receptor interaction	<0.005	0.15	4	12
Cytokine-cytokine interaction pathway	<0.005	0.33	5	26
NF-kappa B signaling pathway	<0.005	0.27	6	22
Pathways in Cancer	<0.005	0.87	7	85
Cell cycle	0.009	0.21	8	14
JAK-STAT signaling pathway	0.011	0.77	9	58
Aldosterone-regulated sodium reabsorption	0.013	0.66	10	48

Figure 31: Top Pathways identified in the Brain EC Fold Change Subnetworks by SSEA with FDR < 0.01

Pathway	SSEA Co-expression Network FDR	SSEA Fold Change Network FDR	SSEA Co-expression Rank	SSEA Fold Change Rank
Apoptosis	<0.001	0.12	2	13
Pyruvate metabolism	<0.001	0.15	4	15
ECM-receptor interaction	<0.001	0.27	6	28
Pathways in Cancer	<0.001	0.28	10	29
Cell cycle	<0.001	0.35	8	31
Hematopoietic cell lineage	<0.001	0.44	3	36
P53 signaling pathway	<0.001	0.70	1	91
Fatty acid metabolism	<0.001	0.73	9	92
JAK-STAT signaling pathway	<0.001	0.77	5	94
Tyrosine metabolism	<0.001	0.87	7	103
Glutathione Metabolism	0.002	0.33	11	30
mTOR signaling pathway	0.003	0.66	12	82
RNA transport	0.003	0.81	13	96
MAPK signaling pathway	0.004	0.56	14	63
DNA replication	0.006	0.21	15	20
Ribosome	0.006	0.80	16	95
Amino sugar and nucleotide sugar metabolism	0.009	0.08	17	11

Figure 32: Top Pathways identified in the Brain EC Co-expression Subnetworks by SSEA with FDR < 0.01

4.3.5 Subnetwork Dynamics during the Progression of LPS Injury

The most significant fold change subnetwork identified in the brain endothelium by SSEA was the Wnt signaling pathway. The Wnt signaling pathway during homeostatic conditions was composed of several genes with large entropy values. More than 60% of genes were assigned larger entropy levels than the mean entropy value of all nodes in the network (0.09). We noticed that the pathway included multiple genes with large entropy values which indicated to us that the Wnt signaling pathway may be regulated by both expression level and number of interactions. As shown in Figure 33, the Wnt signaling subnetwork at baseline is characterized by high entropy (8.67), but then experiences a significant decrease in signaling entropy in response to LPS treatment 24hrs post LPS (6.03). Interestingly, this decrease in signaling entropy was reversed during the endothelial recovery and regeneration process and the signaling entropy level elevated at 1-week post LPS (7.84). SSEA clustered the genes in the Wnt signaling pathway to the top of the ranked subnetwork list. These observations indicate a strong connection between LPS induced inflammation in the brain and the Wnt signaling pathway. Indeed, it has been demonstrated that dysregulation of this subnetwork and change in Wnt signaling gene expression levels induce a disruption of the blood brain barrier¹⁹⁸.

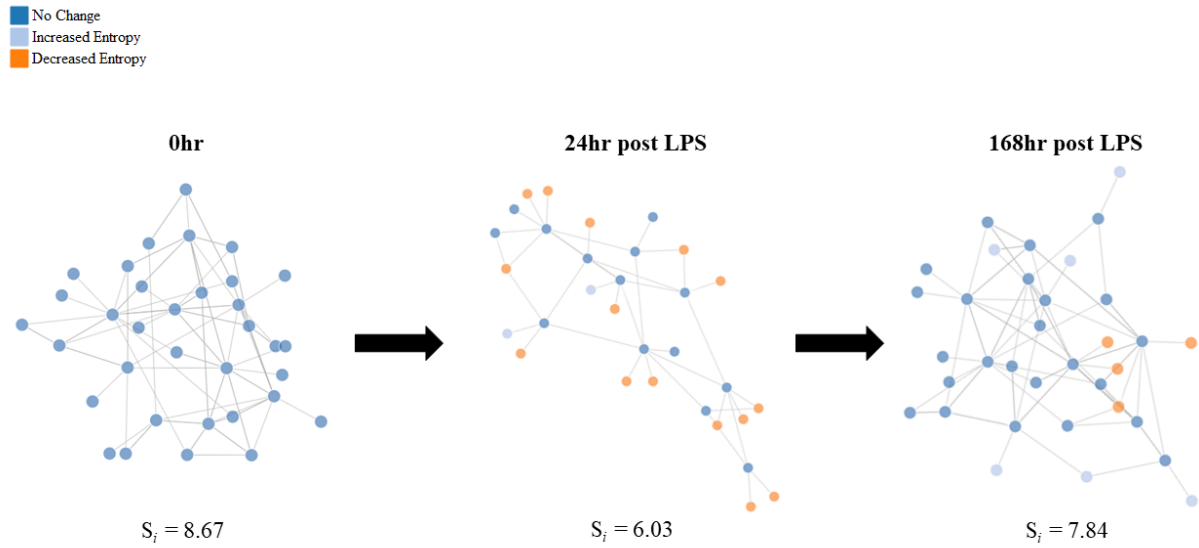


Figure 33: Brain Endothelium Wnt Signaling Subnetwork Dysregulation. In the Brain endothelium, a 34 node Wnt signaling subnetwork exhibited a network signaling entropy decrease 24hrs post inflammatory stimulation ($S_i = 8.67$ vs $S_i = 6.03$) followed by recovery at 168hrs post-injury ($S_i = 6.03$ to $S_i = 7.84$).

4.4 Conclusions

By computationally analyzing the in vivo endothelial transcriptome during systemic inflammation, we identified the heterogeneous organ-specific endothelial signaling dynamics in the brain, lung, and heart tissues. We elucidated the distinct mechanisms by which the endothelium modulates vascular injury, repair, and regeneration in an organ-specific manner. The existence of organ-specific endothelial cell processes has become a strong interest for vascular biologists but is severely understudied. For the first time, we have uncovered the molecular dynamics of brain, lung, and heart endothelial cells during LPS-induced systemic inflammation.

We developed a powerful model to study the endothelial transcriptome during endothelial injury, repair, and regeneration by administering a sub-lethal dose of LPS to RiboTag^{EC} mice. This model enabled us to study the translational relevance of endothelial destruction and regeneration across distinct organs during endotoxemia. It has been previously found that circulating LPS serves as a key mediator of disease in patients with bacteremia and sepsis¹⁸⁸. Our goal was to compare inflammatory transcriptomic responses in multiple vascular beds, therefore we needed to induce systemic inflammation in a controlled manner. LPS administration induced the release of inflammatory mediators and immune cell activation, thus allowing us to establish the heterogeneity of endothelial responses resulting from systemic inflammatory stimulation.

Studying endothelial heterogeneity in response to the systemic inflammatory stress induced by LPS, we found that the endothelium in each

tissue maintains a distinct organ-specific molecular identity. Even in the setting of a profound systemic inflammatory disease, the baseline heterogeneity of each vascular bed was maintained. However, each endothelial bed exhibited a highly characteristic inflammatory response. Brain and heart ECs express classical inflammatory adhesion molecules such as E-Selectin and P-Selectin, whereas lung ECs upregulate chemokines such as *Cxcl1* and *Cxcl9*. The gene expression shifts in the lung reflect the severe loss of lung endothelium recently observed during endotoxemia¹⁹⁹. The marked upregulation of P-Selectin in the heart and brain is especially interesting because P-Selectin is a key mediator of thrombosis and platelet aggregation¹⁹⁹, and both the brain and heart are especially vulnerable to thrombotic events. During the later stage of inflammation at 24 hr, the inflammatory gene expression pathways across all tissues demonstrated significant upregulation of leukocyte migration and chemotaxis genes, suggesting that despite the persistent heterogeneous signatures of the ECs in the respective organs, there is a broad shared program of inflammatory signaling pathways in response to systemic endotoxemia.

We found that the progression and resolution of vascular inflammation in each organ is mediated by the signaling mechanisms activated in the respective tissue-specific endothelium. Classically, when activated by inflammatory signals the endothelium changes in terms of resistance, vasomotor tone, and barrier function²⁰⁰. In our studies, we found that the dynamics and magnitude of change for gold-standard endothelial marker genes varied from organ to organ. The lung endothelium undergoes the most prominent loss of endothelial markers, while the

brain and the heart are more resistant to loss of endothelial markers. This finding is consistent with the severe loss of the lung endothelium in the endotoxemia model of EC injury and recovery²⁰¹. Furthermore, the brain and heart tissue which require consume large amounts of energy express numerous glycolytic factors such as Glut1 in the brain endothelium²⁰² and Eno3²⁰³ in the heart endothelium throughout the inflammatory time course. On the other hand, the lung endothelium heavily relies on a limited number of glycolytic factors such as Pfkfb3 which is known to be a master regulator of glycolysis in endothelial cells²⁰⁴.

When we applied our novel time-series based tool, SSEA, we found that signaling entropy, a measure of subnetwork promiscuity, is highly correlated with the dynamics of the brain endothelium and therefore identified signature pathways. In the fold change network, the most significant dynamic networks identified using SSEA were processes responsible for directly mediating the response to an external stimulus. For instance, Wnt signaling underwent a dramatic shift in the brain fold change network structure during vascular injury, thus suggesting a possible regulatory role of the brain endothelial Wnt signaling pathway during LPS-induced systemic inflammation. On the other hand, when we apply SSEA to co-expression subnetworks, pathways related to cell growth and cell death, metabolism, and genetic information processing. Interpreting signaling entropy as a measure of pathway promiscuity uncovered processes activated during the inflammatory injury, repair, and regeneration. Here we

demonstrated that the prognostic power of signaling entropy in LPS treated mice is indeed informative about endothelial inflammatory injury and resolution.

CHAPTER 5

CONCLUSIONS

Endothelial cells (ECs) require considerable molecular and phenotypic configuration to support organ-specific function. For example, to maintain a permeable endothelial barrier, liver and kidney ECs proliferate at a moderate pace, while lung and brain ECs double at a lower rate with slow turnover²⁰⁵. This heterogeneous endothelial proliferation rate is attributed to the blood brain barrier's (BBB) responsibility to tightly regulate the balance of ions, nutrients, and transport of molecules into the brain tissue²⁰⁶. ECs are an essential component of the BBB forming tight junctions between cells to minimize fenestrations. On the other hand, lung ECs primarily regulate the absorption of oxygen into the blood stream but also assist in immunity by removing exogenous pathogens from inhaled air²⁰⁷. A hallmark of heart ECs is experiencing higher glycolysis and oxidation rates as well as engaging in vascular growth or angiogenesis as to match oxygen demands of the heart²⁰⁸. These examples of organ-specific endothelial phenotypic heterogeneity are driven by the underlying molecular signaling. In this work, we designed and applied novel computational and experimental approaches to ascertain the organ-specific endothelial signatures required for organ-specific function during homeostatic and inflammatory conditions.

In Chapter 2 we first developed a novel computational tool, HeteroPath, to identify tissue-specific pathways from gene expression data by quantitatively evaluating the bidirectional expression of genes in the context of their molecular

signaling pathway. We applied this method to define organ-specific endothelial heterogeneity and tissue-specific neuronal heterogeneity. In both datasets, we found that HeteroPath segregated the distinct cellular populations by identifying signature pathways that were not identified by other pathway analysis methods. Using simulated datasets, HeteroPath demonstrated robustness that was comparable to what was seen using existing gene set enrichment methods. Furthermore, we generated tissue-specific gene regulatory networks involved in vascular heterogeneity and neuronal heterogeneity by performing motif enrichment of the heterogeneous genes identified by HeteroPath and linking the enriched motifs to regulatory transcription factors in the ENCODE database.

We enhanced the resolution of our endothelial heterogeneity studies in Chapter 3 by adapting the RiboTag isolation technique¹⁴⁷ to capture a snapshot of the endothelial translome. Performing RNA-Seq on the endothelial translome provided information about the mRNA transcripts undergoing translation. We generated endothelial specific RiboTag mice (RiboTag^{EC}: *Cdh5*^{CreERT2/+}; *Rpl22*^{HA/+}) and performed RNA-Sequencing on brain, lung, and heart endothelial translomes and identified specific pathways, transporters and cell-surface markers expressed in the endothelium of each organ. We found that endothelial cells express genes typically found in the surrounding tissues such as synaptic vesicle genes in the brain endothelium and cardiac contractile genes in the heart endothelium. In addition, we performed complementary analysis of endothelial single cell RNA-Seq data and found the molecular signatures shared across the endothelial translome and single cell transcriptomes.

After comprehensively establishing the molecular configuration underlying endothelial heterogeneity during homeostatic conditions, we expanded our work in Chapter 4 to study endothelial heterogeneity dynamics in response to inflammatory stimulation. We found that the baseline tissue-specific heterogeneity of the endothelium is maintained during systemic in vivo inflammatory injury as evidenced by the distinct translome responses to inflammatory stimulation.

In the following sections, we summarize endothelial heterogeneity and plasticity and discuss our novel computational frameworks implemented to identify organ-specific vascular disease mechanisms and therapeutic targets for individual vascular beds.

5.1 A Pathway-based Computational Modeling Approach to Identify Tissue-Specific Gene Expression Networks

We developed HeteroPath, a novel unbiased computational approach centered on analyzing gene expression within the context of pathways establishes tissue-specific gene expression signatures. As HeteroPath assesses bi-directional gene expression within pathways, it provides a more comprehensive description of cellular heterogeneity and identifies previously unrecognized tissue-specific therapeutic targets. The objective of our study was primarily to identify heterogeneously dysregulated pathways in populations of cells, organs, or tissues. Our key contribution was the establishment of a method to quantify the heterogeneity of a tissue-specific pathway by analyzing the directionality and magnitude of gene expression dysregulation. We applied HeteroPath to

characterize EC and neuronal heterogeneity from an inherently unbiased perspective. Since ECs and neurons exhibit significant heterogeneity in structure and function as well as propensity for disease, we sought to understand the molecular basis of tissue-specific gene expression signatures and networks to gain important insights into vascular and neuronal disease mechanisms.

In a cohort of freshly isolated mouse ECs from three distinct tissues (heart, brain, and lung), we analyzed the transcriptomic heterogeneity of ECs by comparing gene set enrichment analysis (GSEA) and parametric gene set enrichment analysis (PGSEA) to our novel gene interaction model that assesses the bi-directional gene expression state (HeteroPath). Ranking gene expression levels in the context of established signaling pathways allowed us to discover transcriptomic signatures specific to a vascular bed. We further characterized the tissue-specific signatures by constructing transcriptional networks consisting of the identified heterogeneous genes and their regulatory transcription factors as determined by motif enrichment analysis. GSEA ranks pathways by cumulatively summing the perturbations of all individual genes within the context of each pathway. However, one limitation of this method is that it factors in the magnitude of perturbation for every individual gene when establishing a pathway as tissue-specific. GSEA primarily labeled minimally differentially expressed pathways as tissue-specific and in some cases was unable to detect tissue-specific pathways. Therefore, applying GSEA to identify tissue-specificity would be powerful in settings when examining global cellular quiescence as a function of minute gene expression variation across several tissues. PGSEA predominantly identified

upregulation of amino acid metabolism pathways in the brain endothelium such as phenylalanine, tryptophan, and tyrosine metabolism. HeteroPath analysis, on the other hand, identified transcriptomic upregulation of signaling pathways in the brain such as the Wnt signaling and adherens junction pathways when compared to lung or heart endothelium.

The HeteroPath results observed from the endothelial, neuronal, and simulated datasets suggest that HeteroPath performs most optimally when the dysregulation of genes within pathways is both upregulation and down regulation rather than unidirectional dysregulation. This is evidenced by HeteroPath identifying the fewest significant tissue-specific pathways in the endothelial dataset. In the neuronal dataset, the output of HeteroPath produced the highest enrichment score, most significant p-values, and highest percentage of tissue-specific pathways. To test the robustness of HeteroPath, we manipulated the proportion and magnitude of dysregulated genes within a pathway and observed that as the proportion and magnitude of heterogeneous genes increased the specificity and sensitivity of HeteroPath detecting significant pathways increased. We further evaluated the statistical power and type-I error using a linear additive model and found that HeteroPath performs with comparable statistical power and control of type-I error to GSEA and PGSEA.

The upregulated and downregulated genes within pathways identified by HeteroPath are important factors in activating specialized signals which influence the interactions amongst a heterogeneous population of cells. We mapped these genes to gene expression networks to understand the underlying regulatory

interactions which drive the observed phenotype. From the brain endothelial Wnt signaling gene expression network, we identified the upregulation of the centroid Lef1, which is known to regulate brain vascularization and BBB differentiation¹²⁷. This finding, for example, generates the hypothesis that Lef1 is a key transcription factor mediating the development and maturation of brain ECs through the activation of Wnt signaling.

5.1.1 Limitations and Future Research

The identified tissue-specific gene regulatory networks include regulated pathways that would otherwise be overlooked by conventional analysis methods. Even though HeteroPath shows promise in generating tissue-specific gene regulatory networks, there are limitations of our approach that need to be considered. To assess the performance of HeteroPath, we randomly permuted the class labels and executed HeteroPath to calculate confidence intervals and p-values for each of the AUC values at distinct fold-change cut-offs. However, the sensitivity and specificity of these regulated pathways identified by HeteroPath could not be compared with GSEA and PGSEA using biological datasets since the method for generating the ground truth relied on GSEA and PGSEA. Therefore, we generated simulated microarray datasets and applied HeteroPath, GSEA, and PGSEA to detect differentially expressed gene sets. From this simulation study, we found that the statistical power and type I error rates for HeteroPath were comparable to existing gene set enrichment analysis methods.

Similar to PGSEA and GSEA, the HeteroPath algorithm relies on existing pathway databases. It is known that the curated pathways share multiple genes or describe similar phenomena²⁰⁹. For instance, we found occludin to be highly upregulated in the BBB, but it is not officially curated as a member of the adherens junction pathway even though it is known that occludins regulate adherens junction pathways in the brain endothelium²¹⁰. In the commonly used KEGG pathway database, which we also employed as a pathway reference database for our analysis, pathways often significantly overlap on a molecular and functional level. Therefore, one limitation of our analyses is the accuracy and comprehensiveness of the underlying pathway database curation. To resolve this issue, a dimension reduction machine learning technique could be implemented to curate and filter the pathways from existing databases. Furthermore, expanding existing annotations to include condition-, tissue-, and cell-specific functions for genes and pathways would allow for the prediction of system variation due to factors such as stimuli, mutations, or environmental change⁹⁵. The goal of our analysis and algorithm development was to develop tools that will allow researchers to predict signature gene regulatory networks for cell types, determine putative transcription factors that could drive the heterogeneity and generate novel biological hypotheses. It should be noted that the biological significance of each signature network would need to be confirmed using in vivo gene deletion or gene depletion studies.

In conclusion, we describe herein a computational algorithm which ranks pathways by assigning heterogeneity scores. This technique allowed us to

uncover additional endothelial cell and neuronal signature gene regulatory networks for each tissue that would not have been identified by traditional analyses such as GSEA or PGSEA. Even though our analysis focused on comparing two cohorts of cellular heterogeneity: three endothelial cell populations and three neuronal populations, the algorithm can be readily expanded to assessing pathway heterogeneity between several tissues and implemented in any cellular heterogeneity context. Thus, the described computational approach identifying distinct regulatory pathways and druggable therapeutic targets in endothelial and neuronal populations may be of value in understanding the complex heterogeneity of other tissues.

5.2 Unraveling Tissue-Specific Endothelial Cell Heterogeneity Using Pathway-based Computational Modeling of Gene Expression

In recent years, understanding endothelial heterogeneity and the tissue-specific endothelial niche has become important for defining vascular health during homeostasis²¹¹⁻²¹³. The extent of organ-specific endothelial heterogeneity and the plasticity of ECs in each respective organ to assist in organ-specific function is an important aspect of vascular biology²⁰⁸. In our work, we leveraged the RiboTag mouse model¹⁴⁷ and computational analysis of RNA-Seq data to identify the molecular signature of brain, lung, and heart endothelial transcriptome underlying organ-specific endothelial function. Our findings suggest that in each organ, the ECs adopt the characteristics of the neighboring cells within the tissue. In the brain, the endothelial cells actively translate mRNA transcripts associated with synaptic signaling and amino acid metabolism. In the lung

endothelial translome, we identified the upregulation of immune activation, cell migration, and tissue development. In the heart endothelium, we found a strong presence of cardiac muscle processes as well as proliferation and tissue maturation. While gene set enrichment analysis provided information about the biological processes that are characteristic of each organ-specific endothelium, it lacked the resolution necessary to detect individual gene enrichment. To understand which clusters of genes were enriched in a tissue and which genes potentially interact with each other in a tissue-specific manner, we performed differential expression analysis and established tissue-specific gene expression signatures. These tissue-specific endothelial translome signatures provide insight into the molecular mechanisms driving organ-specific endothelial heterogeneity.

The overarching goal of our systems biology approach was to develop an approach which blends genomic and pathway information to better understand endothelial biology. One of the current difficulties faced is defining the tissue-specific function of individual tissue-specific genes. This remains a challenge because the precise function of genes in multiple tissues are controlled by the microenvironment in which the cells are expressing the gene. Additionally, the tissue-specific gene-gene interactions are not well characterized due to the inability to perform high-throughput interaction measurements. To directly study the molecular configuration in brain, lung, and heart ECs, we integrated differentially expressed genes identified using RNA-Seq with publicly available curated signaling pathway databases to infer biological processes, functional

gene-gene interactions, and provide an enhanced understanding of organ-specific endothelial biology.

The tissue-specific endothelial translome we performed RNA-Seq on is an in vivo complement to the incessantly growing transcriptomic work on endothelial heterogeneity. It revealed the in vivo organ-specific endothelial signatures and generated hypotheses about the role of ECs in organ-specific function. In depth analysis of the endothelial translome data presented an opportunity to understand the tight regulation of endothelial signaling pathways within each organ. Since the organ-specific endothelial signatures we generated accurately weigh and integrate diverse molecular data, these signatures are the most comprehensive map of the relationship between genes, phenotypes, and tissues. These tissue-specific endothelial signatures become especially powerful for understanding vascular disease because they contain the information required to determine the dysregulation and reconfiguration of gene interactions manifested during different stages of disease. Transitions from healthy to diseased states as well as the progression of disease are likely driven by the molecular dynamics that occur in a tissue-specific pattern.

A key challenge in current drug development focuses on the question of tissue-specific targeting. One of the leading goals of drug delivery and therapeutic agents in vascular biology has followed the Folkman analogy of drugs acting only where they are needed¹⁸¹. In our work, we provide the comprehensive list of tissue-specific endothelial specific markers, or zip codes that can be targeted. Ligands which directly target the genes present in the

tissue-specific endothelial signatures would be effective delivery targets for toxic agents, diagnostic molecules, genes, and other effector molecules to localize in brain, lung, or heart endothelial beds²¹⁴. Modern drug delivery techniques for vascular injury should focus on reconstituting the tissue-specific homeostatic organ-specific endothelial signature.

5.2.1 Limitations and Future Research

Computationally analyzing the translome of organ-specific ECs has proven to be invaluable for understanding tissue-specific EC function in the healthy state. Relative to other translome profiling systems, the RiboTag system was easily paired with the *Cre* recombinase system to attain high specificity. This less technically challenging and time-consuming method was robust for targeting the endothelial cell population in each organ. Like all systems, the RiboTag system has inherent limitations which should be considered. For instance, the RiboTag system is not equipped to differentiate between mRNA primed to translate versus mRNA undergoing active translation. Due to this, the translational efficiency for individual mRNAs could not be quantified. If translation efficiency is an essential read out, then the RiboTag method could be performed alongside techniques including polysome profiling²¹⁵.

Additionally, since the RiboTag system directly isolates ribosome associated mRNAs, it cannot be used for single cell RNA-Seq. Although the RiboTag system loses individual cellular gene expression information by reducing the sample population into an aggregate profile, it more accurately detects minimally expressed genes. ScRNA-Seq studies are better suited for

understanding the inter-tissue composition, while tissue-specific transcriptome analyses are powerful for generating a more comprehensive signature since the data contains the depth necessary to make inferences about the tissue structure and function. When the behavior and properties of all the endothelial cells within a tissue are merged using a *Cre* driver, such as *Cdh5* in our case, we were able to comprehensively define the organ-specific endothelial differences.

In this work the RiboTag method allowed us to characterize organ-specific ECs in the healthy state. Now that the baseline organ-specific endothelial signature has been defined it would be quite exciting to compare it to each organ-specific endothelial transcriptome across different disease states. For example, which tissue-specific signaling pathways are activated in the brain endothelium during different stages of neurodegenerative diseases? Using the established baseline endothelial transcriptome signature, we can compare the activated brain EC transcriptome during early stages of Alzheimer's disease to the homeostatic brain EC transcriptome signature to identify endothelial mechanisms of Alzheimer's disease.

As the data we have generated continues to be analyzed we hope to learn more about the underlying mechanisms by which ECs adopt a tissue-specific molecular signature and contribute to tissue-specific function. An increasingly active area of research has been to generate tissue-specific EC populations for regenerative therapy²¹⁶. For example, there have been recent breakthroughs differentiating human pluripotent stem cells (hPSCs) to derive blood-brain barrier endothelial cells^{217,218} and endothelial cells with heart specificity²¹⁹. These studies

have been performed without considering the comprehensive tissue-specific endothelial signature. Now that we have established the pathways and factors that are distinctively expressed in each organ-specific endothelial population, it would be worthwhile to utilize this information to improve differentiating stem cells into organ-specific endothelial cells.

5.3 Capturing Tissue-specific Endothelial Translatome Dynamics during Systemic Inflammation

Even though ECs are situated on the inner lining of blood vessels and primarily act as regulators of transport between the blood and neighboring tissue, they are active mediators of tissue function and modulate tissue development and regeneration²²⁰⁻²²². In response to tissue injury, ECs signal tissue-resident stem cells to activate self-renewal and differentiation pathways to assist in the repair process⁸⁸. It has also previously been reported that ECs secrete chemokines and other paracrine factors to induce repair processes such as regenerative alveolarization in the lungs or liver regeneration²²³. The tissue-specific response ECs exhibit varies considerably in response to injury and is manifested in tissue-specific activation of transcription factors, chemokines, and growth factors²²⁴.

Delineating the molecular signature responsible for the tissue-specific ECs during tissue injury, repair, and regeneration during systemic inflammation required a novel strategy to genetically target the endothelial translatome in an organ-specific manner. We crossed RiboTag mice (*Rpl22^{HA/+}*) with endothelial-specific VE-cadherin-Cre mice to generate RiboTag^{EC} (*Cdh5^{CreERT2/+}; Rpl22^{HA/+}*) mice. Following tamoxifen-induced recombination at week 4, HA-

tagged *Rpl22* was specifically expressed in endothelial cells in all organs. To investigate the mechanisms of organ-specific EC injury, repair, and regeneration we performed RNA-Seq analysis of ECs isolated at 0 hr, 6 hr, 24 hr, 48 hr, 72 hr, and 1-week post-LPS challenge. We implemented this strategy and characterized the transcriptome for brain, lung, and heart ECs during systemic inflammation. One of our most important findings was that the organ-specific endothelial transcriptome signatures were not solely angiocrine factors, adhesion molecules, chemokines, transcription factors, cell surface receptors, or metabolic regulators, but rather a combination of several unique molecular processes. Thus, during systemic inflammation and more precisely endothelial injury, new strategies targeting different tissue-specific molecular processes are required.

The computational analysis of the brain, lung, and heart ECs in response to systemic inflammation induced by LPS identified that each organ-specific endothelial transcriptome maintained a distinct molecular signature. We began unraveling the tissue-specific endothelial heterogeneous responses by testing whether the different responses to LPS-induced injury could be attributed to the dysregulation of inflammatory genes. By focusing on inflammatory genes, we concluded that the inflammatory response genes in the brain, lung, and heart ECs time course significantly. Interestingly, in the brain endothelium there was delay in the progression of inflammation but an accelerated resolution of inflammation. Additionally, the brain endothelial inflammatory response was not as dramatic as the lung and heart endothelium due to the minimal permeability of the blood-brain barrier²²⁵. During early inflammation, the gene expression

dysregulation in the lung was likely a reflection of the significant EC loss. During late inflammation, the hallmark inflammatory gene expression profile in brain, lung, and heart ECs was upregulation of leukocyte migration and chemotaxis factors. This finding suggests that even though the endothelial transcriptome is heterogeneous across distinct organs at baseline, there is a shared inflammatory program in response to systemic endotoxemia.

To perform a temporal analysis as a function of previous time points rather than differential expression analysis of individual timepoints we developed an entropy-based model named Subnetwork Signaling Entropy Analysis (SSEA). When we applied SSEA to infer dysregulated tissue-specific gene regulatory subnetworks from the organ-specific endothelial transcriptome inflammation time series RNA-Seq data, we successfully identified signaling pathways which were misconfigured during inflammation and reconfigured during regeneration. For example, we found a subnetwork of the Wnt signaling pathway, with Lef1 as a centroid node, change its preferred targets during inflammation and then return to baseline interactions after the resolution of inflammation. We found that when we detected differentially regulated subnetworks, we were able to detect genes in the context of their interacting partners which may be genes with weaker differential expression or lower expression level. One of the major advantages of implementing SSEA was that SSEA provided insights into the time series responses specific for each of the distinct endothelial transcriptomes. This advantage did not only prove to be helpful in the case of analyzing a time-series in response to a stimulus but would also be important for analyzing time-series

datasets evaluating cellular differentiation or disease progression. By considering previous time points and the context of differentially expressed genes, we were able to identify a more comprehensive set of differentially expressed genes. In this fashion, we developed a novel method, SSEA, that detects subtle changes in the time series and broadens the scope of drug targets by providing the regulatory network for each differentially expressed gene.

5.3.1 Limitations and Future Research

In this work, we have identified differentially expressed genes by performing individual time point comparisons as well as identified dysregulated subnetworks by calculating subnetwork signaling entropy. Since the network signaling entropy is a function of the gene expression of each node in the network as well as the correlation between genes, it is important to identify why stimulation causes an increase or decrease in tissue-specific subnetworks. While network entropy was able to discriminate dysregulated tissue-specific subnetwork from unaffected subnetworks, it does not better classify individual genes as differentially expressed when compared to standard differential expression approaches. Since we did not perform rigorous analysis on other quantitative network measures, it remains an outstanding question whether other network measures may provide equally good discrimination between dysregulated subnetworks and unaffected subnetworks. Nevertheless, the loss of an edge between two nodes in the network indicated an increase in subnetwork entropy which provides an interpretable framework in which systemic changes in gene expression can be explored.

When studying dynamic biological phenomenon such as the tissue-specific endothelial response to LPS-induced systemic inflammation, datasets must be generated in a time series experiment. In our study, the brain, lung, and heart endothelial transcriptome were sampled and sequenced at several time points. By capturing the gene expression levels of the transcriptome we intended to identify factors responsible for mediating the tissue-specific endothelial response to LPS-induced inflammation. The analysis of this multi-time series involved analyzing several different samples simultaneously. As we increased the number of time points, the resolution of the data we generated better reflected the biological phenomenon, but the number of time points also heightened the complexity of the data and the cost of the experiment. To specifically handle the complexity of the obtained time series data, more powerful algorithms and methods need to be developed. So far, these algorithms have not been implemented due to technical infeasibilities. As tools such as cloud computing and parallelization continue to develop, pipelines and algorithms will have the resources necessary to be implemented to improve accuracy. As researchers continue to focus on high-throughput Omics data integration, the models will become fine-tuned and will further the understanding of time series biological processes including organ-specific endothelial response to systemic inflammation.

CITED LITERATURE

- 1 Aird, W. C. Spatial and temporal dynamics of the endothelium. *J Thromb Haemost* **3**, 1392-1406, doi:DOI 10.1111/j.1538-7836.2005.01328.x (2005).
- 2 Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725-732 (1998).
- 3 Asahara, T. *et al.* Isolation of putative progenitor endothelial cells for angiogenesis. *Science* **275**, 964-967, doi:10.1126/science.275.5302.964 (1997).
- 4 Kennedy, M. *et al.* A common precursor for primitive erythropoiesis and definitive haematopoiesis. *Nature* **386**, 488-493, doi:10.1038/386488a0 (1997).
- 5 Hirschi, K. K. Hemogenic endothelium during development and beyond. *Blood* **119**, 4823-4827, doi:10.1182/blood-2011-12-353466 (2012).
- 6 Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625-630, doi:10.1038/nature03122 (2004).
- 7 Motoike, T., Markham, D. W., Rossant, J. & Sato, T. N. Evidence for novel fate of Flk1(+) progenitor: Contribution to muscle lineage. *Genesis* **35**, 153-159, doi:10.1002/gene.10175 (2003).
- 8 Udan, R. S., Culver, J. C. & Dickinson, M. E. Understanding vascular development. *Wiley interdisciplinary reviews. Developmental biology* **2**, 327-346, doi:10.1002/wdev.91 (2013).
- 9 Risau, W. & Flamme, I. Vasculogenesis. *Annual review of cell and developmental biology* **11**, 73-91, doi:10.1146/annurev.cb.11.110195.000445 (1995).
- 10 Shalaby, F. *et al.* A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis. *Cell* **89**, 981-990, doi:10.1016/s0092-8674(00)80283-4 (1997).
- 11 Drake, C. J. & Fleming, P. A. Vasculogenesis in the day 6.5 to 9.5 mouse embryo. *Blood* **95**, 1671-1679 (2000).
- 12 Coultas, L., Chawengsaksophak, K. & Rossant, J. Endothelial cells and VEGF in vascular development. *Nature* **438**, 937-945, doi:10.1038/nature04479 (2005).
- 13 Leung, D. W., Cachianes, G., Kuang, W. J., Goeddel, D. V. & Ferrara, N. Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science* **246**, 1306-1309, doi:10.1126/science.2479986 (1989).
- 14 Klagsbrun, M. & Soker, S. VEGF/VPF: the angiogenesis factor found? *Current biology : CB* **3**, 699-702, doi:10.1016/0960-9822(93)90073-w (1993).
- 15 Dvorak, H. F., Brown, L. F., Detmar, M. & Dvorak, A. M. Vascular permeability factor/vascular endothelial growth factor, microvascular hyperpermeability, and angiogenesis. *The American journal of pathology* **146**, 1029-1039 (1995).

- 16 Petrova, T. V., Makinen, T. & Alitalo, K. Signaling via vascular endothelial growth factor receptors. *Experimental cell research* **253**, 117-130, doi:10.1006/excr.1999.4707 (1999).
- 17 Robb, L. *et al.* Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7075-7079, doi:10.1073/pnas.92.15.7075 (1995).
- 18 Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434, doi:10.1038/373432a0 (1995).
- 19 Visvader, J. E., Fujiwara, Y. & Orkin, S. H. Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. *Genes & development* **12**, 473-479, doi:10.1101/gad.12.4.473 (1998).
- 20 Kappel, A. *et al.* Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development. *Blood* **96**, 3078-3085 (2000).
- 21 Sato, T. N., Qin, Y., Kozak, C. A. & Audus, K. L. Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 9355-9358, doi:10.1073/pnas.90.20.9355 (1993).
- 22 Schnurch, H. & Risau, W. Expression of tie-2, a member of a novel family of receptor tyrosine kinases, in the endothelial cell lineage. *Development* **119**, 957-968 (1993).
- 23 Korhonen, J., Polvi, A., Partanen, J. & Alitalo, K. The mouse tie receptor tyrosine kinase gene: expression during embryonic angiogenesis. *Oncogene* **9**, 395-403 (1994).
- 24 Dickson, M. C. *et al.* Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice. *Development* **121**, 1845-1854 (1995).
- 25 Oshima, M., Oshima, H. & Taketo, M. M. TGF-beta receptor type II deficiency results in defects of yolk sac hematopoiesis and vasculogenesis. *Developmental biology* **179**, 297-302, doi:10.1006/dbio.1996.0259 (1996).
- 26 Oh, S. P. *et al.* Activin receptor-like kinase 1 modulates transforming growth factor-beta 1 signaling in the regulation of angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 2626-2631, doi:10.1073/pnas.97.6.2626 (2000).
- 27 Urness, L. D., Sorensen, L. K. & Li, D. Y. Arteriovenous malformations in mice lacking activin receptor-like kinase-1. *Nature genetics* **26**, 328-331, doi:10.1038/81634 (2000).
- 28 Klein, R. Eph/ephrin signalling during development. *Development* **139**, 4105-4109, doi:10.1242/dev.074997 (2012).
- 29 Armulik, A., Abramsson, A. & Betsholtz, C. Endothelial/pericyte interactions. *Circulation research* **97**, 512-523, doi:10.1161/01.RES.0000182903.16652.d7 (2005).

- 30 Gaengel, K., Genove, G., Armulik, A. & Betsholtz, C. Endothelial-mural cell signaling in vascular development and angiogenesis. *Arteriosclerosis, thrombosis, and vascular biology* **29**, 630-638, doi:10.1161/ATVBAHA.107.161521 (2009).
- 31 Adams, R. H. & Alitalo, K. Molecular regulation of angiogenesis and lymphangiogenesis. *Nature reviews. Molecular cell biology* **8**, 464-478, doi:10.1038/nrm2183 (2007).
- 32 Carmeliet, P. Angiogenesis in health and disease. *Nature medicine* **9**, 653-660, doi:10.1038/nm0603-653 (2003).
- 33 Eichmann, A. *et al.* Vascular development: from precursor cells to branched arterial and venous networks. *The International journal of developmental biology* **49**, 259-267, doi:10.1387/ijdb.041941ae (2005).
- 34 le Noble, F. *et al.* Flow regulates arterial-venous differentiation in the chick embryo yolk sac. *Development* **131**, 361-375, doi:10.1242/dev.00929 (2004).
- 35 Atkins, G. B., Jain, M. K. & Hamik, A. Endothelial differentiation: molecular mechanisms of specification and heterogeneity. *Arteriosclerosis, thrombosis, and vascular biology* **31**, 1476-1484, doi:10.1161/ATVBAHA.111.228999 (2011).
- 36 Red-Horse, K., Crawford, Y., Shojaei, F. & Ferrara, N. Endothelium-microenvironment interactions in the developing embryo and in the adult. *Developmental cell* **12**, 181-194, doi:10.1016/j.devcel.2007.01.013 (2007).
- 37 Yau, J. W., Teoh, H. & Verma, S. Endothelial cell control of thrombosis. *BMC cardiovascular disorders* **15**, 130, doi:10.1186/s12872-015-0124-z (2015).
- 38 McVerry, B. J. & Garcia, J. G. In vitro and in vivo modulation of vascular barrier integrity by sphingosine 1-phosphate: mechanistic insights. *Cellular signalling* **17**, 131-139, doi:10.1016/j.cellsig.2004.08.006 (2005).
- 39 Mehta, D. & Malik, A. B. Signaling mechanisms regulating endothelial permeability. *Physiological reviews* **86**, 279-367, doi:10.1152/physrev.00012.2005 (2006).
- 40 Komarova, Y. A., Kruse, K., Mehta, D. & Malik, A. B. Protein Interactions at Endothelial Junctions and Signaling Mechanisms Regulating Endothelial Permeability. *Circulation research* **120**, 179-206, doi:10.1161/CIRCRESAHA.116.306534 (2017).
- 41 Aird, W. C. Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms. *Circulation research* **100**, 158-173, doi:10.1161/01.RES.0000255691.76142.4a (2007).
- 42 Potente, M. & Makinen, T. Vascular heterogeneity and specialization in development and disease. *Nature reviews. Molecular cell biology* **18**, 477-494, doi:10.1038/nrm.2017.36 (2017).
- 43 Mitchell, J. A., Ryffel, B., Quesniaux, V. F., Cartwright, N. & Paul-Clark, M. Role of pattern-recognition receptors in cardiovascular health and disease. *Biochemical Society transactions* **35**, 1449-1452, doi:10.1042/BST0351449 (2007).

- 44 Opitz, B., Eitel, J., Meixenberger, K. & Suttorp, N. Role of Toll-like receptors, NOD-like receptors and RIG-I-like receptors in endothelial cells and systemic infections. *Thrombosis and haemostasis* **102**, 1103-1109, doi:10.1160/TH09-05-0323 (2009).
- 45 Gupta, S. K., Lysko, P. G., Pillarisetti, K., Ohlstein, E. & Stadel, J. M. Chemokine receptors in human endothelial cells. Functional expression of CXCR4 and its transcriptional regulation by inflammatory cytokines. *The Journal of biological chemistry* **273**, 4282-4287, doi:10.1074/jbc.273.7.4282 (1998).
- 46 Murdoch, C., Monk, P. N. & Finn, A. Cxc chemokine receptor expression on human endothelial cells. *Cytokine* **11**, 704-712, doi:10.1006/cyto.1998.0465 (1999).
- 47 Lee, M. *et al.* Transcriptional programs of lymphoid tissue capillary and high endothelium reveal control mechanisms for lymphocyte homing. *Nature immunology* **15**, 982-995, doi:10.1038/ni.2983 (2014).
- 48 Girard, J. P., Moussion, C. & Forster, R. HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes. *Nature reviews. Immunology* **12**, 762-773, doi:10.1038/nri3298 (2012).
- 49 Ley, K. & Reutershan, J. Leucocyte-endothelial interactions in health and disease. *Handbook of experimental pharmacology*, 97-133 (2006).
- 50 Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nature reviews. Immunology* **7**, 678-689, doi:10.1038/nri2156 (2007).
- 51 Muller, W. A. Mechanisms of transendothelial migration of leukocytes. *Circulation research* **105**, 223-230, doi:10.1161/CIRCRESAHA.109.200717 (2009).
- 52 Engelhardt, B. & Wolburg, H. Mini-review: Transendothelial migration of leukocytes: through the front door or around the side of the house? *European journal of immunology* **34**, 2955-2963, doi:10.1002/eji.200425327 (2004).
- 53 Coppiello, G. *et al.* Meox2/Tcf15 heterodimers program the heart capillary endothelium for cardiac fatty acid uptake. *Circulation* **131**, 815-826, doi:10.1161/CIRCULATIONAHA.114.013721 (2015).
- 54 Zhao, Z., Nelson, A. R., Betsholtz, C. & Zlokovic, B. V. Establishment and Dysfunction of the Blood-Brain Barrier. *Cell* **163**, 1064-1078, doi:10.1016/j.cell.2015.10.067 (2015).
- 55 Winkler, E. A. *et al.* GLUT1 reductions exacerbate Alzheimer's disease vasculo-neuronal dysfunction and degeneration. *Nature neuroscience* **18**, 521-530, doi:10.1038/nn.3966 (2015).
- 56 Jais, A. *et al.* Myeloid-Cell-Derived VEGF Maintains Brain Glucose Uptake and Limits Cognitive Impairment in Obesity. *Cell* **166**, 1338-1340, doi:10.1016/j.cell.2016.08.010 (2016).
- 57 Tarlungeanu, D. C. *et al.* Impaired Amino Acid Transport at the Blood Brain Barrier Is a Cause of Autism Spectrum Disorder. *Cell* **167**, 1481-1494 e1418, doi:10.1016/j.cell.2016.11.013 (2016).

- 58 Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146**, 873-887, doi:10.1016/j.cell.2011.08.039 (2011).
- 59 De Bock, K. *et al.* Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154**, 651-663, doi:10.1016/j.cell.2013.06.037 (2013).
- 60 Helmlinger, G., Endo, M., Ferrara, N., Hlatky, L. & Jain, R. K. Formation of endothelial cell networks. *Nature* **405**, 139-141, doi:10.1038/35012132 (2000).
- 61 Uldry, M. & Thorens, B. The SLC2 family of facilitated hexose and polyol transporters. *Pflügers Archiv : European journal of physiology* **447**, 480-489, doi:10.1007/s00424-003-1085-0 (2004).
- 62 Dagher, Z., Ruderman, N., Tornheim, K. & Ido, Y. Acute regulation of fatty acid oxidation and amp-activated protein kinase in human umbilical vein endothelial cells. *Circulation research* **88**, 1276-1282, doi:10.1161/hh1201.092998 (2001).
- 63 Elmasri, H. *et al.* Fatty acid binding protein 4 is a target of VEGF and a regulator of cell proliferation in endothelial cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **23**, 3865-3873, doi:10.1096/fj.09-134882 (2009).
- 64 Carracedo, A., Cantley, L. C. & Pandolfi, P. P. Cancer metabolism: fatty acid oxidation in the limelight. *Nature reviews. Cancer* **13**, 227-232, doi:10.1038/nrc3483 (2013).
- 65 Jeon, S. M., Chandel, N. S. & Hay, N. AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. *Nature* **485**, 661-665, doi:10.1038/nature11066 (2012).
- 66 Iso, T. *et al.* Capillary endothelial fatty acid binding proteins 4 and 5 play a critical role in fatty acid uptake in heart and skeletal muscle. *Arteriosclerosis, thrombosis, and vascular biology* **33**, 2549-2557, doi:10.1161/ATVBAHA.113.301588 (2013).
- 67 Grynberg, A. & Demaison, L. Fatty acid oxidation in the heart. *Journal of cardiovascular pharmacology* **28 Suppl 1**, S11-17, doi:10.1097/00005344-199600003-00003 (1996).
- 68 Tousoulis, D., Kampoli, A. M., Tentolouris, C., Papageorgiou, N. & Stefanadis, C. The role of nitric oxide on endothelial function. *Current vascular pharmacology* **10**, 4-18 (2012).
- 69 Bakris, G. L., Basile, J. N., Giles, T. D. & Taylor, A. A. The role of nitric oxide in improving endothelial function and cardiovascular health: focus on nebivolol. *The American journal of medicine* **123**, S2-8, doi:10.1016/j.amjmed.2010.04.012 (2010).
- 70 Eelen, G., de Zeeuw, P., Simons, M. & Carmeliet, P. Endothelial cell metabolism in normal and diseased vasculature. *Circulation research* **116**, 1231-1244, doi:10.1161/CIRCRESAHA.116.302855 (2015).
- 71 Kim, B., Li, J., Jang, C. & Arany, Z. Glutamine fuels proliferation but not migration of endothelial cells. *The EMBO journal* **36**, 2321-2333, doi:10.15252/embj.201796436 (2017).

- 72 Pavlova, N. N. *et al.* As Extracellular Glutamine Levels Decline, Asparagine Becomes an Essential Amino Acid. *Cell metabolism* **27**, 428-438 e425, doi:10.1016/j.cmet.2017.12.006 (2018).
- 73 Pober, J. S. & Cotran, R. S. The role of endothelial cells in inflammation. *Transplantation* **50**, 537-544, doi:10.1097/00007890-199010000-00001 (1990).
- 74 Hunt, B. J. & Jurd, K. M. Endothelial cell activation. A central pathophysiological process. *Bmj* **316**, 1328-1329, doi:10.1136/bmj.316.7141.1328 (1998).
- 75 Liao, J. K. Linking endothelial dysfunction with endothelial cell activation. *The Journal of clinical investigation* **123**, 540-541, doi:10.1172/JCI66843 (2013).
- 76 Mittal, M., Siddiqui, M. R., Tran, K., Reddy, S. P. & Malik, A. B. Reactive oxygen species in inflammation and tissue injury. *Antioxidants & redox signaling* **20**, 1126-1167, doi:10.1089/ars.2012.5149 (2014).
- 77 Bulua, A. C. *et al.* Mitochondrial reactive oxygen species promote production of proinflammatory cytokines and are elevated in TNFR1-associated periodic syndrome (TRAPS). *The Journal of experimental medicine* **208**, 519-533, doi:10.1084/jem.20102049 (2011).
- 78 Nakahira, K. *et al.* Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nature immunology* **12**, 222-230, doi:10.1038/ni.1980 (2011).
- 79 Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221-225, doi:10.1038/nature09663 (2011).
- 80 Wolin, M. S., Gupte, S. A., Neo, B. H., Gao, Q. & Ahmad, M. Oxidant-redox regulation of pulmonary vascular responses to hypoxia and nitric oxide-cGMP signaling. *Cardiology in review* **18**, 89-93, doi:10.1097/CRD.0b013e3181c9f088 (2010).
- 81 Aird, W. C. Phenotypic heterogeneity of the endothelium: II. Representative vascular beds. *Circulation research* **100**, 174-190, doi:10.1161/01.RES.0000255690.03436.ae (2007).
- 82 Chi, J. T. *et al.* Endothelial cell diversity revealed by global expression profiling. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10623-10628, doi:10.1073/pnas.1434429100 (2003).
- 83 Nolan, D. J. *et al.* Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Developmental cell* **26**, 204-219, doi:10.1016/j.devcel.2013.06.017 (2013).
- 84 Aird, W. C. *et al.* Vascular bed-specific expression of an endothelial cell gene is programmed by the tissue microenvironment. *The Journal of cell biology* **138**, 1117-1124, doi:10.1083/jcb.138.5.1117 (1997).
- 85 Bussmann, J., Wolfe, S. A. & Siekmann, A. F. Arterial-venous network formation during brain vascularization involves hemodynamic regulation of

- chemokine signaling. *Development* **138**, 1717-1726, doi:10.1242/dev.059881 (2011).
- 86 Matsuoka, H., Shima, A., Uda, A., Ezaki, H. & Michihara, A. The retinoic acid receptor-related orphan receptor alpha positively regulates tight junction protein claudin domain-containing 1 mRNA expression in human brain endothelial cells. *Journal of biochemistry* **161**, 441-450, doi:10.1093/jb/mvw092 (2017).
 - 87 Pozhilenkova, E. A., Lopatina, O. L., Komleva, Y. K., Salmin, V. V. & Salmina, A. B. Blood-brain barrier-supported neurogenesis in healthy and diseased brain. *Reviews in the neurosciences* **28**, 397-415, doi:10.1515/revneuro-2016-0071 (2017).
 - 88 Rafii, S., Butler, J. M. & Ding, B. S. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316-325, doi:10.1038/nature17040 (2016).
 - 89 Chen, F. *et al.* Multiplatform-based molecular subtypes of non-small-cell lung cancer. *Oncogene* **36**, 1384-1393, doi:10.1038/onc.2016.303 (2017).
 - 90 Su, Z. *et al.* An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. *Genome biology* **15**, 523, doi:10.1186/s13059-014-0523-y (2014).
 - 91 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).
 - 92 Dewey, F. E. *et al.* Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circulation. Cardiovascular genetics* **4**, 26-35, doi:10.1161/CIRCGENETICS.110.941757 (2011).
 - 93 Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nature genetics* **37 Suppl**, S38-45, doi:10.1038/ng1561 (2005).
 - 94 Glez-Pena, D., Gomez-Lopez, G., Pisano, D. G. & Fdez-Riverola, F. WhichGenes: a web-based tool for gathering, building, storing and exporting gene sets with application in gene set enrichment analysis. *Nucleic acids research* **37**, W329-334, doi:10.1093/nar/gkp263 (2009).
 - 95 Khatri, P., Sirota, M. & Butte, A. J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS computational biology* **8**, e1002375, doi:10.1371/journal.pcbi.1002375 (2012).
 - 96 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
 - 97 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
 - 98 Tian, L. *et al.* Discovering statistically significant pathways in expression profiling studies. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13544-13549, doi:10.1073/pnas.0506577102 (2005).

- 99 Henegar, C. *et al.* Clustering biological annotations and gene expression data to identify putatively co-regulated biological processes. *Journal of bioinformatics and computational biology* **4**, 833-852 (2006).
- 100 Elsasser, W. M. Outline of a theory of cellular heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5126-5129, doi:10.1073/pnas.81.16.5126 (1984).
- 101 Altschuler, S. J. & Wu, L. F. Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559-563, doi:10.1016/j.cell.2010.04.033 (2010).
- 102 Grimbergen, A. J., Siebring, J., Solopova, A. & Kuipers, O. P. Microbial bet-hedging: the power of being different. *Current opinion in microbiology* **25**, 67-72, doi:10.1016/j.mib.2015.04.008 (2015).
- 103 Coffin, J. D., Harrison, J., Schwartz, S. & Heimark, R. Angioblast differentiation and morphogenesis of the vascular endothelium in the mouse embryo. *Developmental biology* **148**, 51-62, doi:10.1016/0012-1606(91)90316-u (1991).
- 104 Hatzopoulos, A. K., Folkman, J., Vasile, E., Eiselen, G. K. & Rosenberg, R. D. Isolation and characterization of endothelial progenitor cells from mouse embryos. *Development* **125**, 1457-1468 (1998).
- 105 Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671-674, doi:10.1038/386671a0 (1997).
- 106 Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Statistics in medicine* **9**, 811-818, doi:10.1002/sim.4780090710 (1990).
- 107 Kim, S. Y. & Volsky, D. J. PAGE: parametric analysis of gene set enrichment. *BMC bioinformatics* **6**, 144, doi:10.1186/1471-2105-6-144 (2005).
- 108 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).
- 109 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).
- 110 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 111 Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic acids research* **31**, 374-378, doi:10.1093/nar/gkg108 (2003).
- 112 Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research* **32**, D91-94, doi:10.1093/nar/gkh012 (2004).
- 113 Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16-23, doi:10.1093/bioinformatics/16.1.16 (2000).
- 114 Kel, A. E. *et al.* MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic acids research* **31**, 3576-3579, doi:10.1093/nar/gkg585 (2003).

- 115 Castro, M. A. *et al.* Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nature genetics* **48**, 12-21, doi:10.1038/ng.3458 (2016).
- 116 Martinet, W. *et al.* Gene expression profiling of apoptosis-related genes in human atherosclerosis: upregulation of death-associated protein kinase. *Arteriosclerosis, thrombosis, and vascular biology* **22**, 2023-2029, doi:10.1161/01.atv.0000041843.44312.12 (2002).
- 117 Wuttge, D. M., Sirsjo, A., Eriksson, P. & Stemme, S. Gene expression in atherosclerotic lesion of ApoE deficient mice. *Molecular medicine* **7**, 383-392 (2001).
- 118 Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375, doi:10.1093/bioinformatics/btf877 (2003).
- 119 De Bock, K., Georgiadou, M. & Carmeliet, P. Role of endothelial cell metabolism in vessel sprouting. *Cell metabolism* **18**, 634-647, doi:10.1016/j.cmet.2013.08.001 (2013).
- 120 Sugino, K. *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature neuroscience* **9**, 99-107, doi:10.1038/nn1618 (2006).
- 121 Stranger, B. E. *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS genetics* **8**, e1002639, doi:10.1371/journal.pgen.1002639 (2012).
- 122 Aranguren, X. L. *et al.* Unraveling a novel transcription factor code determining the human arterial-specific endothelial cell signature. *Blood* **122**, 3982-3992, doi:10.1182/blood-2013-02-483255 (2013).
- 123 Fukuchi-Shimogori, T. & Grove, E. A. Neocortex patterning by the secreted signaling molecule FGF8. *Science* **294**, 1071-1074, doi:10.1126/science.1064252 (2001).
- 124 Artyomov, M. N., Meissner, A. & Chakraborty, A. K. A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS computational biology* **6**, e1000785, doi:10.1371/journal.pcbi.1000785 (2010).
- 125 Tallquist, M. D., Soriano, P. & Klinghoffer, R. A. Growth factor signaling pathways in vascular development. *Oncogene* **18**, 7917-7932, doi:10.1038/sj.onc.1203216 (1999).
- 126 Gessert, S. & Kuhl, M. The multiple phases and faces of wnt signaling during cardiac differentiation and development. *Circulation research* **107**, 186-199, doi:10.1161/CIRCRESAHA.110.221531 (2010).
- 127 Stenman, J. M. *et al.* Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* **322**, 1247-1250, doi:10.1126/science.1164594 (2008).
- 128 Quasnicka, H. *et al.* Regulation of smooth muscle cell proliferation by beta-catenin/T-cell factor signaling involves modulation of cyclin D1 and p21 expression. *Circulation research* **99**, 1329-1337, doi:10.1161/01.RES.0000253533.65446.33 (2006).

- 129 Venkiteswaran, K. *et al.* Regulation of endothelial barrier function and growth by VE-cadherin, plakoglobin, and beta-catenin. *American journal of physiology. Cell physiology* **283**, C811-821, doi:10.1152/ajpcell.00417.2001 (2002).
- 130 Wang, H. *et al.* Cyclin A transcriptional suppression is the major mechanism mediating homocysteine-induced endothelial cell growth inhibition. *Blood* **99**, 939-945 (2002).
- 131 De Val, S. & Black, B. L. Transcriptional control of endothelial cell development. *Developmental cell* **16**, 180-195, doi:10.1016/j.devcel.2009.01.014 (2009).
- 132 Lonze, B. E. & Ginty, D. D. Function and regulation of CREB family transcription factors in the nervous system. *Neuron* **35**, 605-623, doi:10.1016/s0896-6273(02)00828-0 (2002).
- 133 Li, C. L., Sathyamurthy, A., Oldenborg, A., Tank, D. & Ramanan, N. SRF phosphorylation by glycogen synthase kinase-3 promotes axon growth in hippocampal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 4027-4042, doi:10.1523/JNEUROSCI.4677-12.2014 (2014).
- 134 Chen, J. Q., Cammarata, P. R., Baines, C. P. & Yager, J. D. Regulation of mitochondrial respiratory chain biogenesis by estrogens/estrogen receptors and physiological, pathological and pharmacological implications. *Biochimica et biophysica acta* **1793**, 1540-1570, doi:10.1016/j.bbamcr.2009.06.001 (2009).
- 135 Potente, M. & Carmeliet, P. The Link Between Angiogenesis and Endothelial Metabolism. *Annual review of physiology* **79**, 43-66, doi:10.1146/annurev-physiol-021115-105134 (2017).
- 136 Jambusaria, A. *et al.* A computational approach to identify cellular heterogeneity and tissue-specific gene regulatory networks. *BMC bioinformatics* **19**, 217, doi:10.1186/s12859-018-2190-6 (2018).
- 137 Vanlandewijck, M. & Betsholtz, C. Single-Cell mRNA Sequencing of the Mouse Brain Vasculature. *Methods in molecular biology* **1846**, 309-324, doi:10.1007/978-1-4939-8712-2_21 (2018).
- 138 Bacher, R. & Kendzierski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome biology* **17**, 63, doi:10.1186/s13059-016-0927-y (2016).
- 139 Zhu, L., Lei, J., Devlin, B. & Roeder, K. A Unified Statistical Framework for Single Cell and Bulk Rna Sequencing Data. *The annals of applied statistics* **12**, 609-632, doi:10.1214/17-AOAS1110 (2018).
- 140 Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nature methods* **11**, 740-742, doi:10.1038/nmeth.2967 (2014).
- 141 Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome biology* **17**, 75, doi:10.1186/s13059-016-0947-7 (2016).
- 142 Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. & Marioni, J. C. Normalizing single-cell RNA sequencing data: challenges and

- opportunities. *Nature methods* **14**, 565-571, doi:10.1038/nmeth.4292 (2017).
- 143 Haimon, Z. *et al.* Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies. *Nature immunology* **19**, 636-644, doi:10.1038/s41590-018-0110-6 (2018).
- 144 Rossner, M. J. *et al.* Global transcriptome analysis of genetically identified neurons in the adult cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 9956-9966, doi:10.1523/JNEUROSCI.0468-06.2006 (2006).
- 145 Zhou, P. *et al.* Contraction of the type I IFN locus and unusual constitutive expression of IFN- α in bats. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2696-2701, doi:10.1073/pnas.1518240113 (2016).
- 146 Piccirillo, C. A., Bjur, E., Topisirovic, I., Sonenberg, N. & Larsson, O. Translational control of immune responses: from transcripts to translomes. *Nature immunology* **15**, 503-511, doi:10.1038/ni.2891 (2014).
- 147 Sanz, E. *et al.* Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13939-13944, doi:10.1073/pnas.0907143106 (2009).
- 148 Jeong, H. W. *et al.* Transcriptional regulation of endothelial cell behavior during sprouting angiogenesis. *Nature communications* **8**, 726, doi:10.1038/s41467-017-00738-7 (2017).
- 149 Sorensen, I., Adams, R. H. & Gossler, A. DLL1-mediated Notch activation regulates endothelial identity in mouse fetal arteries. *Blood* **113**, 5680-5688, doi:10.1182/blood-2008-08-174508 (2009).
- 150 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 151 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
- 152 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* **4**, 1184-1191, doi:10.1038/nprot.2009.97 (2009).
- 153 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).
- 154 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 155 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

- 156 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 157 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
- 158 Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PloS one* **10**, e0121314, doi:10.1371/journal.pone.0121314 (2015).
- 159 Franzen, O., Gan, L. M. & Bjorkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database : the journal of biological databases and curation* **2019**, doi:10.1093/database/baz046 (2019).
- 160 Shimoyama, M. *et al.* The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic acids research* **43**, D743-750, doi:10.1093/nar/gku1026 (2015).
- 161 Povey, S. *et al.* The HUGO Gene Nomenclature Committee (HGNC). *Human genetics* **109**, 678-680, doi:10.1007/s00439-001-0615-0 (2001).
- 162 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
- 163 Athar, A. *et al.* ArrayExpress update - from bulk to single-cell expression data. *Nucleic acids research* **47**, D711-D715, doi:10.1093/nar/gky964 (2019).
- 164 Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372, doi:10.1038/s41586-018-0590-4 (2018).
- 165 Vanlandewijck, M. *et al.* A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475-480, doi:10.1038/nature25739 (2018).
- 166 Augustin, H. G. & Koh, G. Y. Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology. *Science* **357**, doi:10.1126/science.aal2379 (2017).
- 167 Pober, J. S. & Sessa, W. C. Inflammation and the blood microvascular system. *Cold Spring Harbor perspectives in biology* **7**, a016345, doi:10.1101/cshperspect.a016345 (2014).
- 168 Chaqour, J., Lee, S., Ravichandra, A. & Chaqour, B. Absciscic acid - an anti-angiogenic phytohormone that modulates the phenotypical plasticity of endothelial cells and macrophages. *Journal of cell science* **131**, doi:10.1242/jcs.210492 (2018).
- 169 Dejana, E., Hirschi, K. K. & Simons, M. The molecular basis of endothelial cell plasticity. *Nature communications* **8**, 14361, doi:10.1038/ncomms14361 (2017).
- 170 Krenning, G., Barauna, V. G., Krieger, J. E., Harmsen, M. C. & Moonen, J. R. Endothelial Plasticity: Shifting Phenotypes through Force Feedback. *Stem cells international* **2016**, 9762959, doi:10.1155/2016/9762959 (2016).


- 171 Malinovskaya, N. A. *et al.* Endothelial Progenitor Cells Physiology and Metabolic Plasticity in Brain Angiogenesis and Blood-Brain Barrier Modeling. *Frontiers in physiology* **7**, 599, doi:10.3389/fphys.2016.00599 (2016).
- 172 Hamuro, J. *et al.* Metabolic Plasticity in Cell State Homeostasis and Differentiation of Cultured Human Corneal Endothelial Cells. *Investigative ophthalmology & visual science* **57**, 4452-4463, doi:10.1167/iovs.16-19807 (2016).
- 173 Tang, M. *et al.* Brain microvasculature defects and Glut1 deficiency syndrome averted by early repletion of the glucose transporter-1 protein. *Nature communications* **8**, 14152, doi:10.1038/ncomms14152 (2017).
- 174 Silverstein, R. L. & Febbraio, M. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Science signaling* **2**, re3, doi:10.1126/scisignal.272re3 (2009).
- 175 Al-Soudi, A., Kaaij, M. H. & Tas, S. W. Endothelial cells: From innocent bystanders to active participants in immune responses. *Autoimmunity reviews* **16**, 951-962, doi:10.1016/j.autrev.2017.07.008 (2017).
- 176 Kaparakis-Liaskos, M. & Ferrero, R. L. Immune modulation by bacterial outer membrane vesicles. *Nature reviews. Immunology* **15**, 375-387, doi:10.1038/nri3837 (2015).
- 177 O'Kane, R. L., Martinez-Lopez, I., DeJoseph, M. R., Vina, J. R. & Hawkins, R. A. Na(+)-dependent glutamate transporters (EAAT1, EAAT2, and EAAT3) of the blood-brain barrier. A mechanism for glutamate removal. *The Journal of biological chemistry* **274**, 31891-31895, doi:10.1074/jbc.274.45.31891 (1999).
- 178 Hawkins, R. A. The blood-brain barrier and glutamate. *The American journal of clinical nutrition* **90**, 867S-874S, doi:10.3945/ajcn.2009.27462BB (2009).
- 179 Cai, S. *et al.* Regulation of cytoskeletal mechanics and cell growth by myosin light chain phosphorylation. *The American journal of physiology* **275**, C1349-1356, doi:10.1152/ajpcell.1998.275.5.C1349 (1998).
- 180 Schnittler, H. J., Wilke, A., Gress, T., Suttorp, N. & Drenckhahn, D. Role of actin and myosin in the control of paracellular permeability in pig, rat and human vascular endothelium. *The Journal of physiology* **431**, 379-401, doi:10.1113/jphysiol.1990.sp018335 (1990).
- 181 Folkman, J. Angiogenic zip code. *Nature biotechnology* **17**, 749, doi:10.1038/11676 (1999).
- 182 Maoz, B. M. *et al.* A linked organ-on-chip model of the human neurovascular unit reveals the metabolic coupling of endothelial and neuronal cells. *Nature biotechnology* **36**, 865-874, doi:10.1038/nbt.4226 (2018).
- 183 Zamani, M., Karaca, E. & Huang, N. F. Multicellular Interactions in 3D Engineered Myocardial Tissue. *Frontiers in cardiovascular medicine* **5**, 147, doi:10.3389/fcvm.2018.00147 (2018).

- 184 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).
- 185 Proost, S. & Mutwil, M. CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic acids research* **46**, W133-W140, doi:10.1093/nar/gky336 (2018).
- 186 Pita-Juarez, Y. *et al.* The Pathway Coexpression Network: Revealing pathway relationships. *PLoS computational biology* **14**, e1006042, doi:10.1371/journal.pcbi.1006042 (2018).
- 187 Cross, A. Endotoxin: Back to the Future. *Critical care medicine* **44**, 450-451, doi:10.1097/CCM.0000000000001440 (2016).
- 188 Charbonney, E. *et al.* Endotoxemia Following Multiple Trauma: Risk Factors and Prognostic Implications. *Critical care medicine* **44**, 335-341, doi:10.1097/CCM.0000000000001404 (2016).
- 189 Ghaemmaghani, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737-741, doi:10.1038/nature02046 (2003).
- 190 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
- 191 Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* **45**, D362-D368, doi:10.1093/nar/gkw937 (2017).
- 192 Platig, J., Castaldi, P. J., DeMeo, D. & Quackenbush, J. Bipartite Community Structure of eQTLs. *PLoS computational biology* **12**, e1005033, doi:10.1371/journal.pcbi.1005033 (2016).
- 193 Blondel, V. D., Guillaume, J. L., Hendrickx, J. M., de Kerchove, C. & Lambiotte, R. Local leaders in random networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* **77**, 036114, doi:10.1103/PhysRevE.77.036114 (2008).
- 194 Newman, J. R. & Weissman, J. S. Systems biology: many things from one. *Nature* **444**, 561-562, doi:10.1038/nature05407 (2006).
- 195 Xu, G., Bennett, L., Papageorgiou, L. G. & Tsoka, S. Module detection in complex networks using integer optimisation. *Algorithms for molecular biology : AMB* **5**, 36, doi:10.1186/1748-7188-5-36 (2010).
- 196 Lengfeld, J. E. *et al.* Endothelial Wnt/beta-catenin signaling reduces immune cell infiltration in multiple sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E1168-E1177, doi:10.1073/pnas.1609905114 (2017).
- 197 Mehta, D. *et al.* Modulatory role of focal adhesion kinase in regulating human pulmonary arterial endothelial barrier function. *The Journal of physiology* **539**, 779-789, doi:10.1113/jphysiol.2001.013289 (2002).
- 198 Sabbagh, M. F. *et al.* Transcriptional and epigenomic landscapes of CNS and non-CNS vascular endothelial cells. *eLife* **7**, doi:10.7554/eLife.36187 (2018).
- 199 Merle, N. S. *et al.* P-selectin drives complement attack on endothelium during intravascular hemolysis in TLR-4/heme-dependent manner.

- Proceedings of the National Academy of Sciences of the United States of America* **116**, 6280-6285, doi:10.1073/pnas.1814797116 (2019).
- 200 Charo, I. F. & Ransohoff, R. M. The many roles of chemokines and chemokine receptors in inflammation. *The New England journal of medicine* **354**, 610-621, doi:10.1056/NEJMra052723 (2006).
- 201 Liu, M. *et al.* Sox17 is required for endothelial regeneration following inflammation-induced vascular injury. *Nature communications* **10**, 2126, doi:10.1038/s41467-019-10134-y (2019).
- 202 Patching, S. G. Glucose Transporters at the Blood-Brain Barrier: Function, Regulation and Gateways for Drug Delivery. *Molecular neurobiology* **54**, 1046-1077, doi:10.1007/s12035-015-9672-6 (2017).
- 203 Cascone, T. *et al.* Increased Tumor Glycolysis Characterizes Immune Resistance to Adoptive T Cell Therapy. *Cell metabolism* **27**, 977-987 e974, doi:10.1016/j.cmet.2018.02.024 (2018).
- 204 Cantelmo, A. R. *et al.* Inhibition of the Glycolytic Activator PFKFB3 in Endothelium Induces Tumor Vessel Normalization, Impairs Metastasis, and Improves Chemotherapy. *Cancer cell* **30**, 968-985, doi:10.1016/j.ccell.2016.10.006 (2016).
- 205 Craig, L. E., Spelman, J. P., Strandberg, J. D. & Zink, M. C. Endothelial cells from diverse tissues exhibit differences in growth and morphology. *Microvascular research* **55**, 65-76, doi:10.1006/mvre.1997.2045 (1998).
- 206 Gastfriend, B. D., Palecek, S. P. & Shusta, E. V. Modeling the blood-brain barrier: Beyond the endothelial cells. *Current opinion in biomedical engineering* **5**, 6-12, doi:10.1016/j.cobme.2017.11.002 (2018).
- 207 Mai, J., Virtue, A., Shen, J., Wang, H. & Yang, X. F. An evolving new paradigm: endothelial cells--conditional innate immune cells. *Journal of hematology & oncology* **6**, 61, doi:10.1186/1756-8722-6-61 (2013).
- 208 Marcu, R. *et al.* Human Organ-Specific Endothelial Cell Heterogeneity. *iScience* **4**, 20-35, doi:10.1016/j.isci.2018.05.003 (2018).
- 209 Jiang, Z. & Gentleman, R. Extensions to gene set enrichment. *Bioinformatics* **23**, 306-313, doi:10.1093/bioinformatics/btl599 (2007).
- 210 Tietz, S. & Engelhardt, B. Brain barriers: Crosstalk between complex tight junctions and adherens junctions. *The Journal of cell biology* **209**, 493-506, doi:10.1083/jcb.201412147 (2015).
- 211 Yuan, L. *et al.* A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity. *Nature communications* **7**, 10160, doi:10.1038/ncomms10160 (2016).
- 212 Cleuren, A. C. A. *et al.* The in vivo endothelial cell transcriptome is highly heterogeneous across vascular beds. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 23618-23624, doi:10.1073/pnas.1912409116 (2019).
- 213 Goveia, J. *et al.* An Integrated Gene Expression Landscape Profiling Approach to Identify Lung Tumor Endothelial Cell Heterogeneity and Angiogenic Candidates. *Cancer cell* **37**, 421, doi:10.1016/j.ccell.2020.03.002 (2020).

- 214 Burrows, F. J. & Thorpe, P. E. Vascular targeting--a new approach to the therapy of solid tumors. *Pharmacology & therapeutics* **64**, 155-174, doi:10.1016/0163-7258(94)90037-x (1994).
- 215 Liang, S. *et al.* Polysome-profiling in small tissue samples. *Nucleic acids research* **46**, e3, doi:10.1093/nar/gkx940 (2018).
- 216 Wilson, H. K., Canfield, S. G., Shusta, E. V. & Palecek, S. P. Concise review: tissue-specific microvascular endothelial cells derived from human pluripotent stem cells. *Stem cells* **32**, 3037-3045, doi:10.1002/stem.1797 (2014).
- 217 Lippmann, E. S., Al-Ahmad, A., Azarin, S. M., Palecek, S. P. & Shusta, E. V. A retinoic acid-enhanced, multicellular human blood-brain barrier model derived from stem cell sources. *Scientific reports* **4**, 4160, doi:10.1038/srep04160 (2014).
- 218 Lippmann, E. S. *et al.* Derivation of blood-brain barrier endothelial cells from human pluripotent stem cells. *Nature biotechnology* **30**, 783-791, doi:10.1038/nbt.2247 (2012).
- 219 Lui, K. O. *et al.* Driving vascular endothelial cell fate of human multipotent Isl1+ heart progenitors with VEGF modified mRNA. *Cell research* **23**, 1172-1186, doi:10.1038/cr.2013.112 (2013).
- 220 Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564-567, doi:10.1126/science.1064344 (2001).
- 221 Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559-563, doi:10.1126/science.1063889 (2001).
- 222 Ding, B. S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310-315, doi:10.1038/nature09493 (2010).
- 223 Ding, B. S. *et al.* Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. *Cell* **147**, 539-553, doi:10.1016/j.cell.2011.10.003 (2011).
- 224 Obermeier, B., Daneman, R. & Ransohoff, R. M. Development, maintenance and disruption of the blood-brain barrier. *Nature medicine* **19**, 1584-1596, doi:10.1038/nm.3407 (2013).
- 225 Banks, W. A. & Robinson, S. M. Minimal penetration of lipopolysaccharide across the murine blood-brain barrier. *Brain, behavior, and immunity* **24**, 102-109, doi:10.1016/j.bbi.2009.09.001 (2010).

APPENDIX A
COPYRIGHT PERMISSIONS

[Aims and scope](#)[Fees and funding](#)[Language editing services](#)**Copyright** [Preparing your manuscript](#)[Prepare supporting information](#)[Conditions of publication](#)[Editorial policies](#)[Peer-review policy](#)[Manuscript transfers](#)[Promoting your publication](#)

Copyright

- Copyright on any open access article in a journal published by BioMed Central is retained by the author(s).
- Authors grant BioMed Central a [license](#) to publish the article and identify itself as the original publisher.
- Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified.
- The [Creative Commons Attribution License 4.0](#) formalizes these and other terms and conditions of publishing articles.

In addition to BioMed Central's copyright policy, some journals also follow an Open Data policy and the [Creative Commons CC0 1.0 Public Domain Dedication waiver](#) applies to all published data in these journals. Further information can be found on the individual journals pages.

Where an author is prevented from being the copyright holder (for instance in the case of US government employees or those of Commonwealth governments), minor variations may be required. In such cases the copyright line and license statement in individual articles will be adjusted, for example to state '© 2016 Crown copyright'. Authors requiring a variation of this type should [inform BioMed Central](#) during or immediately after submission of their article. Changes to the copyright line cannot be made after publication of an article.

Exceptions to copyright policy

Our [policy pages](#) provide details concerning copyright and licensing for articles which were previously published under policies that are different from the above. For instance, occasionally BioMed Central may co-publish articles jointly with other publishers, and different licensing conditions may then apply. In all such cases, however, access to these articles is free from fees or any other access restrictions.

Information specifically regarding permissions and reprints can be found [here](#). Please [contact us](#) if there are questions.

BMC Series[Submit manuscript](#)[Editorial Board](#)[Editor Profiles](#)[Sign up for article alerts and news from this journal](#)**Follow****2018 Journal Metrics****Citation Impact**

2,511 - 2-year Impact Factor
2,970 - 5-year Impact Factor
0.855 - Source Normalized Impact per Paper (SNIP)
1,374 - SCImago Journal Rank (SJR)

Usage

4,129,368 downloads

Social Media Impact

4446 mentions

[More about our metrics](#)

Terms and conditions

Welcome and Consent to Terms

eLife Sciences Publications, Ltd. ("eLife") is a 501(c)(3) nonprofit initiative for the very best in science and science communication. In furtherance of our mission, eLife operates websites, including the websites at elifesciences.org, submit.elifesciences.org, payments.elifesciences.org and other eLife websites (the "eLife Sites"). We offer use of the eLife journal content free of charge to the public through the eLife Sites. Access to and use of the eLife Sites is provided by eLife subject to the following Terms and Conditions, which are a contract between eLife and you. Use of the eLife Sites constitutes your acceptance of these Terms and Conditions. If you do not accept these Terms and Conditions in full, you do not have permission to access and use the eLife Sites and should cease doing so immediately.

Ownership

The eLife Sites, the software ("eLife Software"), application programming interfaces ("eLife APIs"), content and trademarks used on or in connection with the eLife Sites are owned by eLife or its licensors, and are subject to US and international intellectual property rights and laws. Nothing contained herein shall be construed as conferring by implication, or otherwise any license or right under any trademark, copyright or patent of eLife or any other third party and all rights are reserved, except as explicitly stated in these Terms and Conditions.

License to Use Journal Articles and Related Content

Unless otherwise indicated, the articles and journal content published by eLife on the eLife Sites are licensed under a [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) (also known as a CC-BY license). This means that you are free to use, reproduce and distribute the articles and related content (unless otherwise noted), for commercial and noncommercial purposes, subject to citation of the original source in accordance with the CC-BY license.

APPENDIX B
CURRICULUM VITAE

Vita

Ankit Jambusaria

Bioinformatics, University of Illinois at Chicago
Chicago, IL

EMPLOYMENT

- | | |
|----------------------|---|
| Aug 2014 – May 2020 | Graduate Research Assistant
University of Illinois at Chicago
Advisor: Asrar Malik, Jalees Rehman, Yang Dai |
| April 2019 – Present | Bioinformatics Intern
Tempus Labs
Manager: Robert Tell |
| Sep 2017 – June 2018 | Bioinformatics Associate I at Genentech
PRO Unlimited, Inc.
Manager: Akshata Udyavara & Radia Johnson |
| June 2017 – Sep 2017 | Bioinformatics & Computational Biology Intern
Genentech, gRED Bioinformatics & Comp. Bio.
Manager: Akshata Udyavara |
| Dec 2012 – Jul 2013 | Genomics Researcher
Five3 Genomics
Manager: Steve Benz |
| Sep 2011 – Jun 2013 | Undergraduate Researcher - Bioengineering
University of California – Santa Cruz
Advisor: Mark Akeson |

EDUCATION

- | | |
|---------------------|---|
| Aug 2014 – Present | Doctor of Philosophy in Bioinformatics
University of Illinois at Chicago
Advisors: Asrar Malik, Jalees Rehman, Yang Dai |
| Aug 2008 – Jun 2013 | Bachelor of Science in Bioengineering
Minor in Bioinformatics
University of California – Santa Cruz |

PUBLICATIONS

1. **Jambusaria, A.**, et al. (2020). "Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation." *Elife* **9**.
2. Liu, M., Zhang, L., Marsboom, G., **Jambusaria, A.**, Xiong, S., Toth, P. T., Benevolenskaya, E. V., Rehman, J., Malik, A. B. (2019). "Sox17 is required

for endothelial regeneration following inflammation-induced vascular injury." Nat Commun 10(1): 2126.

3. **Jambusaria A.**, Klomp J., Hong Z., Rafii S., Dai Y., Malik AB., Rehman J.: A computational approach to identify cellular heterogeneity and tissue-specific gene regulatory networks. BMC bioinformatics 2018, 19(1):217.
4. Zhang L., **Jambusaria A.**, Hong Z., Marsboom, G., Toth, P. T., Herbert, B.-S., Malik A. B., Rehman, J. (2017). SOX17 Regulates Conversion of Human Fibroblasts into Endothelial Cells and Erythroblasts via De-Differentiation into CD34+ Progenitor Cells. Circulation. doi: 10.1161/circulationaha.116.025722

AWARDS AND HONORS

Jan 2019	Gordon Research Conference Travel Award
Oct 2017	UIC Department of Pharmacology Rising Star Award
Jan 2017	UIC Graduate Student Council Travel Award
Apr 2016	UIC Graduate Student Council Travel Award
Jun 2013	B.S., Bioengineering (Cum Laude), University of California – Santa Cruz

TEACHING

University of California – Santa Cruz, Santa Cruz, CA
Teaching Assistant: Math 19A, Calculus for Science, Engineering & Mathematics, Viktor Ginzburg instructor

PRESENTATIONS AT NATIONAL AND INTERNATIONAL CONFERENCES

"Assessing the dynamics of network signaling entropy in tissue injury". Ankit Jambusaria, Zhigang Hong, Yang Dai, Asrar B. Malik and Jalees Rehman. Gordon Research Conference: Quantitative Genetics and Genomics Lucca (Barga), IT February 13, 2019. (Poster Presentation)

"Assessing the dynamics of network signaling entropy in tissue injury". Ankit Jambusaria, Zhigang Hong, Yang Dai, Asrar B. Malik and Jalees Rehman. ISMB 2018. Chicago, IL July 7, 2018. (Poster Presentation)

"Unraveling Cellular Heterogeneity by Pathway-based Prediction of Tissue-Specific Gene Regulatory Networks". Ankit Jambusaria, Jeff Klomp, Zhigang Hong, Shahin Rafii, Asrar B. Malik and Jalees Rehman. UIC Bioengineering Symposium 2018. Chicago, IL April 20, 2018. (Oral Presentation)

"HeteroPath: A Pathway-based Computational Modeling Approach to Identify Tissue-Specific Gene Expression Networks". Ankit Jambusaria, Jeff Klomp,

Zhigang Hong, Shahin Rafii, Asrar B. Malik and Jalees Rehman. Experimental Biology 2017. Chicago, IL April 24, 2017. (Oral Presentation)

“Computationally generated transcriptional networks for heterogeneous organ-specific endothelial cells signaling pathways”. Ankit Jambusaria, Shahin Rafii, Zhigang Hong, Jeffrey Klomp, Asrar B. Malik and Jalees Rehman. International Vascular Biology Meeting 2016. Boston, MA. November 1, 2016. (Poster Presentation)

“Computational characterization of tissue-specific endothelial cell populations via transcriptional networks derived from heterogeneous signaling pathways”. Ankit Jambusaria, Jeffrey Klomp, Asrar B. Malik and Jalees Rehman. UIC Center for Cardiovascular Research Symposium 2016. Chicago, IL. September 14, 2016. (Oral Presentation)

“An unbiased K-means based single-cell algorithm uncovers heterogeneous gene expression signatures of endothelial cells post- inflammatory stimulus”. Ankit Jambusaria, Sarah Krantz, Nicole Pohl-Avila, Asrar B. Malik and Jalees Rehman. NAVBO Vascular Biology 2015. Boston, MA. October 20, 2015. (Oral Presentation)