THE UNIVERSITY OF CHICAGO

ANALYSIS OF TIBETAN GENETIC ADAPTATION AND ITS FUNCTIONAL CONSEQUENCES IN MULTIPLE, HYPOXIA-RESPONSIVE CELL TYPES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY

OLIVIA ANNE GRAY

CHICAGO, ILLINOIS

JUNE 2021

Copyright © 2021 by Olivia Anne Gray
All rights reserved

TABLE OF CONTENTS

List of figures vii

List of tables viii

Acknowledgements ix

Abstract of the dissertation xii

Chapter 1: Introduction 1

1.1 The Genetics of Local Adaptation 1

1.2 Identification and interpretation of causal variants 2

1.3 High altitude hypoxia 4

1.4 Human adaptation to high altitude: the Tibetan Paradigm 7

1.5 New frontiers in functional genetics of human adaptation 10

1.6 Analysis of Tibetan genetic adaptation and its molecular consequences 12

Chapter 2: Positive selection in Tibetan populations has disrupted the activity of *EPAS1* enhancers in multiple, hypoxia-responsive cell types 13

2.1 Abstract 13

2.2 Introduction 14

2.3 Methods 17

2.3.1 ATAC-seq 17

2.3.2 Capture Hi-C 18

2.3.3 Luciferase enhancer assays 19

2.3.4 Generation of CRISPR-cas9 deletion endothelial lines and *mus*ENH5-KO mouse lines 20

2.3.5 Transcriptional profiling of WT and KO mice 23

2.4 Results 25

2.4.1 The regulatory landscape of the *EPAS1* locus in endothelial cells 26

2.4.2 *EPAS1* contains multiple enhancer regions with allele-specific effects in the Endothelium 29

2.4.3 Assessing enhancer and allele-specific activity in the kidney 29

2.4.4 Enhancer ENH5 displays hypoxia responsive, and allele-specific activity in cardiomyocytes 30

2.4.5 CRISPR-mediated deletion of ENH5 in endothelial cells results in downregulation of *EPAS1* and HIF-2α target gene *EGLN3* 32

2.4.6 *mus*ENH5 shows evidence of regulatory activity in multiple tissues *in vivo* 34

2.5 Discussion 40

2.5.1 Pleiotropy in hypoxic adaptation 41

2.6 Acknowledgements of work performed 43

Chapter 3: Development of a panel of Tibetan and Han Chinese iPSCs: a unique resource for the study of local adaptations 44

3.1 Abstract 44

3.2 Introduction 44

3.3 Methods 48

3.3.1 Sample Acquisition 48

3.3.2 LCL Generation and Genetic Ancestry analysis 48

3.3.3 iPSC Generation 50

3.3.4 RT-qPCR to characterize expression of endogenous and exogenous pluripotency factors 52

3.3.5 Embryoid body assays 54

3.3.6 RNA-seq of iPSCs 54

3.4 Results 55

3.4.1 Population Genetic Analyses of Tibetan and Chinese samples 55

3.4.2 Generating a 20 line Tibetan and Han Chinese iPSC panel 57

3.4.3 Embryoid body assays demonstrate functional pluripotency in panel 59

3.4.4 Analysis of gene expression in iPSC panel 60

3.5 Discussion 63

3.6 Acknowledgements of work performed 66

Chapter 4: Differentiation of a panel of Tibetan and Han Chinese iPSCs into vascular endothelium 67

4.1 Introduction 67

4.2 Methods 69

4.2.1 Sample acquisition and Generation of iPSC library 69

4.2.2 Differentiation of vascular endothelial cells 69

4.2.3 Immunofluorescent staining of iPSC-derived vascular endothelium 71

4.2.4 iPSC-derived EC harvest and bulk sequencing 71

4.2.5 Data Analysis: iPSC-derived endothelial cells 72

4.3 Results 72

4.3.1 All 20 lines successfully differentiate and are pulled down with high success rate 72

4.3.2 Differential expression (DE) analyses of endothelial cells recapitulate downregulation of *a* in Tibetans but are highly noisy 74

4.3.3 PLIER analysis of differential gene modules between populations 80

4.4 Discussion 81

4.5 Acknowledgements of work performed 83

Chapter 5: Conclusions 84

5.1 Tibetan genetic adaptations have downregulated *EPAS1* in multiple hypoxia-responsive cell and tissue types 84

5.2 Establishing an iPSC panel for the investigation of local adaptations 86

5.3 Future Directions 88

Bibliography 90

LIST OF FIGURES

Figure 2.1 A view of the UCSC genome browser 28

Figure 2.2 Chromatin HMM data 31

Figure 2.3 Representative ENH5 luciferase assay results 32

Figure 2.4 qPCR results for three genes in the endothelial ENH5 clones 33

Figure 2.5 Luciferase assay results for the *mus*ENH5 allele 35

Figure 3.1 Feeder iPSCs grown on CF-1 MEFs 51

Figure 3.2 Genetic ancestry analysis of TBC-CHB panel 56

Figure 3.3 Representative examples of feeder and feeder-free iPSCs 58

Figure 3.4 IF staining of embryoid bodies 60

Figure 3.5 Principal components analysis of iPSCs and differentiation products 61

Figure 3.6 Log-normalized expression counts by population for six pluripotency factors 62

Figure 4.1 Light microscopy image of P0 iPSC-derived vascular endothelial cells 73

Figure 4.2 Immunofluorescent staining of two iPSC-derived endothelial lines 74

Figure 4.3 PCA graphs of iPSC-derived vascular endothelium 76

Figure 4.4 Population differences in *EPAS1* log-CPM expression 79

LIST OF TABLES

Table 2.1 All primers used in CRISPR analyses 23

Table 2.2 Tibetan SNPs contained in each validated enhancer region 30

Table 2.3 Number and direction of differentially expressed (DE) genes 37

Table 2.4 Genes of interest identified in the mashr analysis 37

Table 2.5 Gene ontology and pathway analysis results 40

Table 3.1 Primers used for quality control of iPSC lines 53

ACKNOWLEDGEMENTS

First and foremost, I want to thank my PI, Anna Di Rienzo for her mentorship, her guidance, and her belief in me. She took a chance on me when she accepted me – a student with little previous education in genetics – into this program and then into her lab. She gave me the autonomy to explore, build, and direct my own project and research with her guidance, and this freedom gave me the push I needed to become a more critical thinker, an improved networker and collaborator, and a better scientist. Her constructive criticism in all aspects of my work has been truly invaluable. I've seen both my writing and presentations improve dramatically over the year thanks to her detailed feedback and insight. I am also incredibly thankful for the many members of both the Di Rienzo and Kupfer labs with whom I had the pleasure to work over the years. I gained a great deal from not only the opportunities I had to present my own work, but from those I had to hear about their work. I would like to especially acknowledge Jennifer Yoo and David Witonsky with whom I have worked closely. Jenny is the best technician I have ever had the opportunity to work with. Her expertise in mouse work and dedication to learning new skills in mammalian cell culture has made a huge difference to the amount and quality of results we were able to produce. David is an astounding computationalist and intrepid defender of the null. He's careful analyses and patience with my less computational background has been invaluable.

Next, I want to acknowledge my committee. I am both humbled and grateful to have had such an engaged, academically diverse, and supportive group of scientists guiding me. My chair, Marcelo Nobrega, has been an incredible source of knowledge, feedback, support, and inspiration. My project could not have happened without him. Yun Fang, likewise, has selflessly contributed his insights, training, and even equipment and lab space to this work. Luis Barreiro

agreed to join the committee just ahead of my penultimate committee meeting. Despite coming into the project so late, he was fully engaged and brought so much energy and so many new ideas to every meeting.

In addition to my own lab and committee members, I have many other labs and lab members to acknowledge here. As the sole grad student in my lab taking on a project that became increasingly interdisciplinary, I would have been lost without the unbelievable amount of training, collaboration, advice, and support (both academic and personal) that was selflessly donated by so many different PIs, students, and postdocs in the program. First and foremost among these is Debora Sobreira. Debora is not only one of the best scientists I have ever had the good fortune to work with, she is also an incredible friend and mentor. I am forever grateful for the time she took training me, and for the hours spent side-by-side culturing stem cells and chatting on the 3rd floor. I also want to thank Ivy Aneas from Marcelo's lab, Matt Kraus from Yun's lab, and many members of the Gilad, He, and Stephens labs at the University of Chicago. The collaborative and generous nature of the faculty and students in Human Genetics is such an asset to the program.

Beyond academia, I also want to express my thanks for members of my cohort, friends at the university, and friends from my many homes across the country. It truly takes a small village of support to make it through a Ph.D., and I am incredibly fortunate to have had a collection of strange, wild, brilliant, wonderful friends with whom to share this journey. At the university, I've gotten to share the trials and tribulations of grad school with Kevin Furlong, Meghan O'Connell, Carlo Berti, Becca Satterwhite, Linsin Smith, Tina Roman and Ryan Dohn. Beyond academia, I have felt so much love and support from Alex Romanzow, Alesha Sedasey, Josh Cox, Katie

Kennedy, Jeremy Alisaukas, and John Fredrickson. Thank you all, so much. I could not have done without you, and I hope to return the favor.

Finally, I have to thank my family, who have supported me in my academic ambitions from a young age. My mother, Karen Gettman, gave me the strength to pursue my dreams. She taught me to be bold and unapologetic in a world that doesn't often encourage these traits in women. My father, Terence Gray, gave me my love of science and research. He taught me that curiosity and delight in discovering new things doesn't end with childhood. My brother, Ian Gray, kept me connected to the world beyond academics, sharing his love of nature and politics and reminding me of the wider world outside of graduate school stressors. I love you all. Thank you.

ABSTRACT OF THE DISSERTATION

Tibetan populations have lived at high-elevation since the late Pleistocene, adapting to an oxygen concentration that is just 60% what is available at sea-level. Understanding the consequences of such local adaptations is a major goal in human genetics research. However, although decades of research have provided a strong foundation of evidence concerning Tibetan phenotypes and genetic signatures of adaptation, disentangling the functional and molecular impact of these selection signals remains a challenge. In the dissertation work presented here, I comprehensively examine functional and molecular impacts of Tibetan genetic adaptation using an interdisciplinary approach. First, I dissect the regulatory architecture of *EPAS1* - the gene which carries the strongest signatures of selection in the Tibetan genome – and assess the contribution of variation at this locus to adaptive phenotypes. I uncover evidence of striking pleiotropic activity in regulatory elements impacted by Tibetan variation, implicating adaptive pleiotropy in Tibetan adaptation. Next, I describe the generation of a panel of induced pluripotent stem cells (iPSCs) from Tibetan and Han Chinese populations. These cells offer an unprecedented opportunity to explore gene networks shaped by local adaptation in cell types of interest in a controlled and well-balanced experiment. Finally, I discuss my initial differentiation of this iPSC panel into vascular endothelium, and detail the challenges and possible solutions to managing inter-individual variation in population panels such as ours. Taken together, the work described here provides both novel insights into the role of pleiotropy in Tibetan adaptation to hypoxia, and new tools to further investigate cell-type specific gene expression networks and phenotypes which result from Tibetan genetic variation.

CHAPTER 1

INTRODUCTION

1.1 The Genetics of Local Adaptation

Local adaptation, defined as adaptation to specific, local environmental features in a heterogeneous or changing environment has long been observed and studied within and between species [1]. It is a crucial component to the ability to survive and respond to a changing environment, allowing certain populations or individuals to persist and become genetically distinct from closely related groups. Human evolutionary history, is often thought to have been guided by a propensity to adapt and survive in a wide variety of environments in an ecological era of rapid climatic change[2]. The rapid 'Out of Africa' expansion, ~50-70kya, is a remarkable event in evolutionary history. In just tens of thousands of years, the human population expanded into widely variable environments, through bottlenecks, admixture, and rapid population growth. Unique among most mammals, humans colonized climates ranging from the arctic to the tropics, lowland plains, to arid plateaus. Along the way resource availability and novel modes of subsistence altered the nutrients available to local populations, creating new selective pressures[3]–[6]. In addition, humans were exposed to a range of new environmental pressures which drove local adaptation including changes in UV exposure[7]–[9], arctic temperatures[6], high-altitude hypoxia[10]–[13], and exposure to novel pathogens[14]–[16].

The precise mechanism of local adaption in human populations has long been studied, particularly with the increasing data availability and computational power in modern genetics. While historically, natural selection has been viewed as acting upon a single novel variant which is quickly swept to high frequency or even fixation [17], increasingly modern research has pointed to more complex mechanisms of adaptations and noted that single-allele hard sweeps

appear rare in the history of human evolution [18]–[20]. Instead, soft sweeps via standing variation[20], [22], [23], introduction of novel variants and haplotypes via introgression events [21], and perhaps most prominently, genome-wide polygenic adaptation [18] are thought to underlie much of human evolutionary adaptation.

1.2 Identification and interpretation of causal variants

Although population genetic approaches allow the identification of variants carrying signatures of natural selection, they provide no information on the mechanism(s) by which they exert a fitness effect. Moreover, due to linkage disequilibrium, multiple variants at the same locus often carry selection signals making it hard to unambiguously pinpoint the causal variant(s) under selection. Despite decades of study and recent innovation in genomics tools, answering these questions still remains a monumentally challenging task.

In essence, there are two routes one can take to interpreting these variants. The first is a GWAS-based approach which leverages the ever-increasing wealth of GWAS data to link variants carrying selection signals with specific phenotypes [22], [23]. This is an incredibly powerful tool, but one which suffers from several disadvantages. First, it requires some *a priori* information about phenotypes you expect or know to be under selection. For some selective pressures this may be obvious, or there may be phenotypes previously linked to fitness advantages, for others however clearly advantageous phenotypes may not be evident, or one may be interested in identifying phenotypes under selection in an unbiased manner. The second major issue with GWAS-based approaches is power. Dependent upon effect size and allele frequency, large sample sizes are needed to gain power to detect associated variants[24], [25]. Sample sizes in the thousands or even tens of thousands, as are now commonly seen in consortium GWA studies, may be impractical or impossible to obtain from remote, locally adapted populations in

some of the globe's most extreme environments[11], [26]. More recently, the advent of new methodologies for testing polygenic selection in a population of interest by leveraging allele frequency data from within a population of interest and GWAS data from larger global databases, obviate some of these issues, though they are still somewhat dependent on the availability of GWAS data for the adaptive phenotype [27]–[29]. Yet even these new methodologies cannot overcome the issues inherent in purely computational assessments of selected alleles, which is the disentangling the causal SNP(s), i.e. the true target(s) of selection, and correlated SNPs.

Therefore, a second route can be taken or used in addition to a GWAS-based approach, the use of humanized animal models to investigate the effect of putatively selected alleles *in vivo*. This technique has been used to great effect, for example, in characterizing the EDARV370A variant, which was predicted to affect hair thickness in humans[30], [31]. Through a humanized mouse model, this phenotype was not only confirmed, but additional phenotypes were identified, such as an increase in eccrine sweat glands which may had a fitness effect[32]. Although this is an undoubtedly powerful approach still requires some prediction of phenotypic outcome and an assumption that it will be visible and measurable in mice. This assumption becomes much more challenging when, unlike missense variant EDARV370, the SNPs under investigation are non-coding. The majority of selection signals identified in human populations are non-coding and presumed to be regulatory [30], [33]. Understanding how, when, through which gene(s), and in what tissue(s) these regulatory variants exert a fitness effect remains a crucial goal in understanding human evolution and adaptation.

1.3 High altitude hypoxia

For the purposes of the research presented here, we are interested primarily in local adaptation to hypobaric hypoxia experienced at high altitude. High altitude environments, defined as more than 2,500m above sea level, have been independently settled by several human populations across the globe. These populations have had to cope with and adapt to a local oxygen concertation just 60% of that experienced at sea level. In addition to being a severe pressure on human physiology, hypoxia is distinct from other common environmental pressures in that it is universally and evenly experienced by all members of a population residing at high-altitude [12]. No amount of material cultural innovations or socioeconomic status can mitigate the experience of hypoxia, whereas in arctic climates, for instance, changes in material culture may yield better protection from the elements and socioeconomic status may afford more protective housing and equipment [34], [35]. In many ways, settling in such an environment is a shocking testament to the power of adaptation. Animal life depends, absolutely, upon oxygen delivery to every cell to support mitochondrial respiration and survival.

It is therefore unsurprising that the response pathway to conditions of low oxygen is highly conserved. This conserved network of genes known as the Hypoxia Inducible Factors (HIF) pathway, which is broadly involved in sensing hypoxia and activating response genes [36], [37]. The HIF pathway is driven by the HIF complex, a transcription factor complex consisting of an α and β subunit. In the presence of oxygen, HIF-α subunits are targeted for degradation through oxygen-dependent hydroxylation of a specific proline residue by a prolyl hydroxylase domain protein (PHD1-3). This in turn targets the HIF-α protein for degradation through the ubiquitin proteasome pathway[38]. However, in hypoxic conditions, HIF-α is stabilized. Once

stable, one of the three α-subunits (HIF1-3α) heterodimerizes with a HIF-β subunit, translocates to the nucleus and activates target genes by binding hypoxia response elements (HRE) [39]. These target genes differ both between tissues and amongst the specific HIF-α isoform involved, but they are broadly involved in directing the molecular response to hypoxia, including angiogenesis, glucose metabolism, inflammation, vasodilation/vasoconstriction, and erythropoiesis to name a few[39]–[42].

The HIF pathway may be activated by any source of tissue hypoxia or hypoxemia, both environmental hypoxia such as is found at high altitude, and biological hypoxia found in low oxygen niches within the body and developing embryos [42]–[45]. These niches develop in the embryo, activating the HIF pathway which plays a critical role in the development of the cardiovascular and hematopoietic systems. Within the adult body, local hypoxic microenvironments keep the HIF pathway active in specific regions where it plays an essential role in, for example, the maintenance of hematopoietic stem cells [46]–[48]. In cases of environmental hypoxia, the activation of the HIF pathway drives the acclimatization response to both acute and chronic hypoxia. The acclimatization response is a bevy of physiological changes including elevated ventilation, blood pressure, and hemoglobin concentration; all of which aim to increase inspired oxygen, blood oxygen concentration and oxygen delivery to the organs in order to counteract the effects of hypoxia [12], [41], [49].

However, overactivation or aberrant activation of the HIF pathway either due to hypoxia or improper oxygen homeostasis in the body can lead to or exacerbate a range of human pathologies[40], [50]–[53]. For example, mountaineers or lowlanders settling at high-altitude commonly experience a syndrome known as chronic mountain sickness (CMS) [49], [54]–[56] which can be characterized by excessive erythrocytosis, hypoxemia, and increased hypoxic

vasoconstriction, ultimately leading to high pulmonary arterial pressure, congestive right heart failure and even pulmonary edema and death[55], [57]. It is interesting to note that although excessive erythrocytosis is commonly considered a defining feature of CMS and one correlated with increased pulmonary arterial pressure (PAP) and pulmonary hypertension, this relationship is not observed in east Asian high-altitude residents such as the acclimatized Chinese and Tibetan populations[55]. Research in these populations suggests that excessive vasoconstriction rather than polycythemia-induced increases in blood viscosity underlies CMS pathogenesis[55]. In human pathology, aberrant activation of the HIF pathway can be driven by a number of diseases. In asthma, for instance, hypoxic tissue in the lungs drives activation of the HIF pathway resulting in hypervascularization and airway remodeling, the degree of which is directly correlated with asthma severity[58]–[60].

This capacity to respond to changes in environmental oxygen availability through changes in patterns gene expression is an example as phenotypic plasticity. Activation of the HIF pathway is one mechanism underlying plasticity in gene expression, which in turn underlies specific plastic phenotypes such as the erythropoietic response. The impact of directional selection on phenotypic plasticity remains contentious[61]–[64]. However, a growing body of evidence, including empirical data from animal populations, have posited that some plasticity is maladaptive, and that strong directional selection acts to reduce this maladaptive phenotypic plasticity [62], [65]. Or, put another way, that selection favors a dampening of the plastic response towards a reduced transcriptional plasticity. In this paradigm, it is easy to see how HIF activation could contribute to maladaptive plasticity in hypoxic response. For example, the plastic increase in Hb levels in response to hypoxia is advantageous in fluctuating environments, but in the long run it may lead to erythrocytosis and perpetuate chronic mountain sickness.

Though to our knowledge no empirical studies have been performed in humans to assess this paradigm, it has been suggested that Tibetan adaptation to high-altitude has favored a dampening of the phenotypic and molecular response to hypoxia[13], [62], [66]. Finding evidence in the gene expression data of a global downregulation of hypoxia response genes and gene modules would provide the first empirical evidence for maladaptive plasticity driving adaptive evolution in humans.

1.4 Human adaptation to high altitude: the Tibetan Paradigm

From an evolutionary perspective, given the severity and uniformity of selective pressure applied by hypoxia, high-altitude adapted populations represent a uniquely interesting study group. From a biomedical perspective, the hypoxia endured by high-altitude populations might be thought of as a proxy for the tissue hypoxia which causes or exacerbates so many cardiovascular diseases[52], [53], [58], [67]. Thus, high-altitude adapted populations may be seen as the output of a long forward genetics experiment to uncover the basis of coping with hypoxia. For all these reasons, high-altitude populations have long been the focus of intense interest in the field of evolutionary genetics and beyond[11], [13], [68]–[74]. A distinct feature of human adaptation to high altitude is that is has occurred multiple times in our history. Human populations have settled at and adapted to high-altitude multiple times in human history, at varying timepoints, spanning the globe: the Aymara and Quechua Native American populations on the Andean Altiplano, the Amhara people in the Ethiopian Highlands, and Tibetans and Sherpa on the Tibetan Plateau [12], [71].

While the Amhara are less well studied, decades of research have investigated the physiology and genetics of Tibetan and Andean natives. Intriguingly, the two populations seem to have adapted to the same hypoxic pressures by entirely distinct mechanisms[12]. One of the

most striking differences between the two populations is their blood oxygen content at high altitude. Andean natives have significantly higher arterial oxygen content than both native Tibetan highlanders and acclimatized lowlanders, whereas Tibetans have lower arterial oxygen content than both Andeans and closely related acclimatized lowlanders. This difference is driven by both a lower hemoglobin concertation and lower oxygen saturation among Tibetans [69], [72], [73], [75]. While Tibetans do experience the prototypical rise in hemoglobin as part of the acclimatization response when they ascend to high altitude, this increase is significantly less than that of either the Andeans or closely related lowlanders such as the Han Chinese[10]. Tibetans also display a unique ventilatory phenotype. Typically, in the process of acclimatization, ventilation increases initially, then decreases back towards sea level averages over time spent at high altitude. Tibetans, however, seem to maintain an elevated resting ventilation at high altitude relative to Andeans or Han Chinese[12], [13], [68], [76]. Despite this, Tibetans are still remarkably hypoxemic compared to all other high altitude inhabitants, and yet, have consistently better reproductive outcomes than acclimatized Han Chinese[70], [77].

This seeming contradiction has only grown more remarkable with the advent of high throughput sequencing and population genetics studies. Multiple studies have examined the Tibetan genome for evidence of a selection signal both using phenotype-agnostic methods and using GWAS of distinct Tibetan phenotypes. These studies have found that not only is the unelevated hemoglobin phenotype observed in Tibetans highly heritable, but the alleles associated with reduced hemoglobin concentration are the same alleles which carry the strongest signal of selection found in the Tibetan genome[11], [74], [78]–[80]. These alleles fall into an extended haplotype in the introns of the Endothelial PAS Domain 1 (*EPAS1*) gene. This gene is readily linked to unelevated hemoglobin as its protein product, Hypoxia Inducible Factor 2 alpha

(HIF-2α), is known to activate erythropoietin (EPO) in the kidney, which drives the erythropoietic response to hypoxia [81]. Tibetans are known to produce less EPO at high altitude than acclimatized Han Chinese, moreover *EPAS1* is known to be downregulated relative to the Han in a number of cell types including umbilical vein endothelial cells and peripheral blood[13], [78]. What remains unclear is whether or not unelevated hemoglobin is, in itself, a phenotype under selection or if instead it is merely a pleiotropic effect of selection favoring a reduction of HIF-2α in one or more cell types. The evidence is contradictory. On the one hand, many have argued that reducing hemoglobin concentration may aid in blood flow or reduce risk during pregnancy loss[70], [76], [82]–[85]. On the other, our lab has repeatedly found that although low hemoglobin is associated with better reproductive outcomes among Tibetan women, there is no evidence of polygenic selection favoring reduced hemoglobin concentration, which is a highly polygenic trait[11], [86].

Thus, there remains an open question as to how the high-altitude haplotype at *EPAS1* influences fitness and through what tissues and phenotypes in might exercise this effect. This question strongly motivated the research described in Chapter 2 of this dissertation.

This haplotype is particularly interesting not only as the source of the strongest selections signal in the Tibetan genome, but also as a site of archaic introgression from the Denisovans, an archaic hominin group contemporaneous with humans and Neanderthals [87], [88]. The extended derived haplotype in *EPAS1*is shared only between Tibetans and Denisovans, and to a very small extent with Chinese populations, suggesting that admixture may have occurred between Denisovans and ancestral Tibetans, where it was driven to high frequency through natural selection. This introgression event is considered one of the most clear-cut examples of adaptive archaic introgression events in humans, yet it is unclear how the adaptive alleles functioned or

why they existed in the Denisovan genome. Until recently, all known fossils and information about Denisovans came from Denisova cave in the Altai region of Russia [89], which is not at high altitude (700m elevation). Although evidence of Denisovan admixture across Asia and Melanesia exists in the genomic data from local populations [90], [91] it was unclear if the adaptive alleles at *EPAS1* were mere standing variation that became advantageous to humans when they migrated to low oxygen environments, or whether it could be the result of a much older adaptation within the Denisovan lineage. However, in 2019, half of a hominin mandible that had been found in Baishiya Karst Cave on the Tibetan plateau in 1980 was subjected to palaeoproteomic analysis and determined to have come from a Denisovan, providing the first evidence for Denisovan habitation outside of the Altai mountains and specifically at high altitude (3,280 m elevation)[92]. The possibility that one or more alleles in the Denisovan *EPAS1* haplotype may have been specifically adapted to the conditions of the Tibetan plateau is of great interest and further prompts the need to characterize the effects of causal alleles at this locus. The first portion of this dissertation will focus on untangling the genetic architecture of the *EPAS1* locus in order to identify alleles with functional effects, identify the relevant tissue types for these effects, and begin to parse in implications of the archaic introgression and subsequent selective event on the high-altitude *EPAS1* haplotype in Tibetans.

1.5 New frontiers in functional genetics of human adaptation

In addition to uncovering novel information about known selection signals and phenotypes in Tibetans, we also now have the tools to look more broadly at the Tibetan genome and perform more agnostic analyses to identify unknown or unexplored phenotypes and gene networks involved in adaptation to high-altitude. Additionally, innovations in induced stem cell technologies now enable us to examine gene expression in cell types that would previously have

been impractical or impossible to gather from the populations of interest. For example, work in the Di Rienzo lab has uncovered evidence of polygenic selection favoring a lower pulse in Tibetan populations[11]. However, in practice, it would be exceedingly difficult to collect enough primary heart tissue from Tibetans and a control group such as the Han Chinese living in hypoxic conditions to comprehensively analyze gene expression differences in the heart and identify novel gene networks that may have been shaped by local adaptation. However, induced pluripotent stem cells (iPSCs) have been increasingly utilized in genetics and biomedical research to study the genetics of gene regulation, regulatory networks, and molecular phenotypes in cell types of interest[93]–[96]. These cell lines are easily generated from lymphoblastoid cell lines (LCLs), a particularly useful starting cell type that is both readily available for many global populations and also very straightforward to generate from as little as 3ml of blood from donors. LCLs can be rapidly expanded and can be reprogrammed into iPSCs with a greater that 95% success rate [95], [97], [98]. The generation of iPSCs has been well characterized, and they have been found to maintain individual patterns of gene expression as well as capturing individual-specific phenotypic variation upon differentiation into terminal cell types[94]–[96]. These properties have made iPSCs increasingly popular in the study of human disease [99]–[102]. Yet these same properties open up the possibility of using iPSCs to answer previously unanswerable questions about human variation and adaptation among populations in controlled environmental conditions. As described above, there remain open questions about the nature of phenotypic plasticity in human evolution, with tantalizing evidence in Tibetan populations for maladaptive plasticity favoring a dampening of the hypoxic response relative to the Han Chinese. iPSCs now provide a novel tool in which Tibetan and Chinese stem cells may be differentiated in parallel into relevant hypoxia-responsive cell types and analyzed under controlled hypoxic conditions.

1.6 Analysis of Tibetan genetic adaptation and its molecular consequences

The purpose of this dissertation is to comprehensively examine the functional impact of Tibetan adaptation to high-altitude hypoxia using a two-pronged approach: 1) dissecting the regulatory architecture of *EPAS1* and the contribution of variation at this locus to adaptive phenotypes, and 2) uncovering novel gene networks shaped by adaptive evolution in vascular cell types using an iPSC-based approach. For the purposes of this dissertation, I have focused primarily on endothelium, which is essential to both the pathological and biological responses to hypoxia [53], [54], [85] and has already been heavily implicated in Tibetan adaptation[11], [12], [78], [103]. While this is certainly not the only cell types relevant to the adaptation to high-altitude hypoxia, it represents an important first step towards elucidating the functional molecular consequences of Tibetan adaptive variants.

In Chapter 2, I will describe an interdisciplinary approach combining computational, *in vitro*, and *in vivo* techniques to examine the genetic architecture of the *EPAS1* locus. I will discuss the identification an allele-specific and hypoxia-responsive enhancer of *EPAS1* with pleiotropic effects in endothelial and cardiac cell types. In Chapter 3, I will discuss the development of an iPSC panel that enables us to ask nuanced questions about the effects of local adaptation in different cell types in a highly-controlled environment. Finally, In Chapter 4, I will detail our first investigations using this iPSC panel to differentiate vascular endothelial cells.

CHAPTER 2

POSITIVE SELECTION IN TIBETAN POPULATIONS HAS DISRUPTED THE ACTIVITY OF *EPAS1* ENHANCERS IN MULTIPLE, HYPOXIA-RESPONSIVE CELL TYPES

2.1 Abstract

Tibetans have lived in conditions of high-altitude hypoxia for over 10,000 years and have evolved several striking phenotypes including a dampening of the erythropoietic response resulting in unelevated hemoglobin levels. Population genetic studies in Tibetans have revealed that strong signals of positive selection and association with lower hemoglobin levels co-localize to a 17kb segment in the endothelial PAS domain 1 (*EPAS1*) gene, whose protein product, Hif-2α, is a central regulator of the hypoxia response. Here, we report the discovery of three regions within this segment, which confer allele-specific enhancer activity and loop to the *EPAS1* promoter. The activity of these enhancers is significantly disrupted by Tibetan alleles in several key cell types (endothelial, kidney, and cardiac). We further characterize one of these enhancers (ENH5) whose activity was found to be not only allele-specific, but also hypoxia responsive in all three cell types. Deletion of this enhancer resulted in downregulation of *EPAS1* and HIF-2α targets in hypoxia. This hypothesis is confirmed by *in vivo* deletion of the orthologous region in mice, resulting in downregulation of *EPAS1* and differential gene expression across multiple interrogated tissues. These results have important implications for understanding the impact of positive selection on regulatory variation.

2.2 Introduction

Improper oxygen homeostasis, low blood oxygen concentration (hypoxemia), and hypoxic stress in other tissues is central to a wide range of human pathologies, including chronic obstructive pulmonary disease, pulmonary hypertension, asthma, ischemic heart disease, and many others. The pathological hypoxia that causes or exacerbates these diseases leads to overactivation of the Hypoxia Inducible Factors (HIF) pathway, which drives the molecular response to hypoxia [40, p. 2], [104]. This pathway is driven by the HIF complex, which is made up of an α and β subunit. HIF-α subunits are targeted for degradation in the presence of oxygen, and are stabilized under conditions of hypoxia. Once stable, one of the three α-subunits (HIF1-3α) heterodimerizes with a HIF-β subunit, and activates target genes by binding hypoxia response elements (HRE) [39]. Under conditions of biological hypoxia such as hypoxic microenvironments in the adult body or produced during development, the HIF pathway drives essential biological processes such as development of the cardiovascular and hematopoietic systems and maintenance and propagation of hematopoietic stem cells [5]–[8],[46]–[48]. Additionally, the HIF pathway drives the acclimatization response to chronic or acute exposure to atmospheric hypoxia, resulting in an elevation of ventilation, blood pressure, and hemoglobin concentration, which conspire to mitigate the drop in inspired oxygen and blood oxygen concentration[12], [41], [49]. However, chronic exposure to both atmospheric or pathological hypoxia can lead to aberrant HIF-pathway activation resulting in or worsening many diseases[40], [50]–[53]. For example, in cancers, solid tumors become hypoxic at their center stabilizing HIF-α isoforms and activating the HIF pathway. This activation is a crucial tumorigenic event, driving tumor angiogenic, cell proliferation, and ultimately metastasis[105].

Similarly, natives and visitors exposed to long-term high-altitude hypoxia risk developing chronic mountain sickness (CMS), a disease characterized by serious complications such as pulmonary hypertension and edema[49], [54], [56].

Populations adapted to high altitude have emerged as a novel system in which to study the response to hypoxia[11], [13], [70], [78]–[80], [106]–[108]. Tibetan populations, for example, have inhabited elevations above 3500m since the late Pleistocene, living at an elevation with just 60% of the oxygen pressure available at sea level[12], [41], [109], [110]. Exposure to such low oxygen applies a serious environmental pressure to populations inhabiting these elevations, yet the Tibetans have adapted a resistance to their environmental hypoxia and have thrived on the plateau for millennia. This unique resistance to the pathological pressure of hypoxic exposure has made Tibetan populations a central focus of scientific inquiries, which have revealed that the improved health and reproductive outcomes Tibetans display at altitude result from a suite of unique adaptations which distinguish Tibetans from acclimatized lowlanders. For instance, the typical response to the increasing hypoxemia experienced at high-altitude would be pulmonary vasoconstriction often leading to pulmonary hypertension, and yet Tibetans display no such vasoconstriction resulting in significantly lower rates of both pulmonary hypertension and CMS[49], [56], [103], [111]. In addition, Tibetan women display reduced high-altitude intrauterine growth restriction and resultant reproductive losses than do Han Chinese women at the same altitude [70], [112]. Another prominent phenotype which emerged in the comparison of Tibetans to acclimatized Chinese is the lower hemoglobin (Hb) concentrations at high altitude. However, Tibetans maintain normal oxygen dissociation from Hb resulting in low arterial oxygen content and severe hypoxemia[12], [108].

Unelevated Hb concentration is highly heritable, and a number of studies have identified alleles associated with low Hb concentration in the endothelial PAS domain protein 1 gene (*EPAS1*), which codes for HIF-2α, that have high frequency in Tibetans but are absent or very rare elsewhere[11], [74], [78], [79]. These same alleles are inferred by numerous studies and selection metrics to carry the strongest signal of selection found in the Tibetan genome, and indeed one of the strongest signatures of selection identified in any human population[11], [41, p. 2], [77], [78, p. 1], [79], [80]. Intriguingly, many of these alleles are part of a haplotype introgressed from Denisovans – an archaic hominin – suggesting that adaptive introgression may be acting at this locus[88].

The link between *EPAS1* and Hb levels appears mechanistically straightforward. HIF-2α is a known activator of erythropoietin (EPO) in epo-producing interstitial fibroblasts in the kidney, which in turn stimulates erythropoiesis in a hypoxia-dependent manner [81]. This mechanism is consistent with studies showing that EPO levels among Tibetans are reduced compared to other high-altitude inhabitants [12], [113]. The coincidence of selection signal at alleles associated with reduced hemoglobin in Tibetans has been interpreted to be due to a selective advantage being conferred by a dampened induction of EPO and erythropoietic response to hypoxia. Moore et al. (2001), for example, hypothesized that decreased Hb during pregnancy in Tibetans permits an increased uterine artery blood flow that is strongly associated with improved pregnancy outcomes[70]. Other researchers have hypothesized that a reduction in erythropoiesis would result in lower blood viscosity and risk for pulmonary hypertension and edema[71], [114]. However, evidence from lowlanders with and without CMS living at high altitude has suggested that pulmonary vasoconstriction rather than increased blood viscosity due to polycythemia is the leading cause of hypoxic pulmonary hypertension and eventual pulmonary

edema in CMS patients[55]. Moreover, hemoglobin concentration is known to be a highly polygenic phenotype both amongst Tibetans[11], [79] and in larger European cohorts[86], [115], yet work by the Di Rienzo lab has found no statistically significant evidence of polygenic selection on decreased hemoglobin levels beyond the *EPAS1* SNPs in Tibetans, raising the possibility that reduced hemoglobin may not be the true trait under selection[11]. It is thus unclear if reduced Hb concentration alone is the phenotypic target of adaptation in Tibetans. And, if it were a target of selection, there remains an open question of how Tibetans cope with the resultant hypoxemia.

Work in the Di Rienzo lab has narrowed the region of colocalized Hb association and selection signal, to a 17kb fragment in intron 2 spanning X SNPs with high frequency alleles in Tibetans. Because of their near-perfect linkage disequilibrium, the phenotypic effects of these SNPs cannot be distinguished through mapping approaches. In this study, we focus first on the endothelium, as the main site of *EPAS1* expression and a key tissue in the etiology of PH, to characterize the regulatory landscape of *EPAS1* and to assess the allelic effects on enhancer activity in normoxia and hypoxia. We hone in on an enhancer region, ENH5, which has hypoxia-dependent allele-specific effects and demonstrate these effects in multiple cell types. We further characterize this enhancer by *in vivo* deletion in mice and show that it has marked transcriptional consequences in multiple organs involved in the physiological response to hypoxia.

2.3 Methods

2.3.1 ATAC-seq

ATAC-seq was performed on primary human aortic endothelial cells (HAECs) as described in [116]. Triplicate samples of cells cultured for 24 hours in either normoxia (21% O_2) and hypoxia

(1% O_2), were harvested in parallel and processed into ATAC libraries using the Nextera DNA library prep kit (FC-121-1030). Indexed samples were pooled and sequenced on one lane of the Illumina Hi-Seq using Single-end 50bp at the University of Chicago Genomics Core. Data was analyzed as described in [117]. Briefly, adaptor sequences were trimmed and reads aligned to hg19 using bowtie. Low quality reads, duplicate reads, and mitochondrial DNA reads were removed using Samtools. Peaks were called independently with macs2 callpeak (parameters: callpeak –
nomodel –shift -100 –extsize 200) and Homer findPeaks (parameters: -gsize 2.5e9 -F 4 -L 4) . Differentially accessible (DA) peaks were assessed with EdgeR looking either at the union of peaks identified by both HOMER and macs2 in each condition and merging peaks overlapping within 300bp as described in [117], or, more agnostically, by dividing up the genome into 300bp windows and assessing differential read counts in each window between conditions.

2.3.2 Capture Hi-C

In situ Hi-C was performed as described previously [118]. Briefly 5 million HAECs were grown in normoxia for 24 hours, then harvested and counted. Cells were resuspended in 1X DPBS and 37% formaldehyde was added to a final concertation of 1%. Formaldehyde treatment was carried out for 10 minutes to cross-link interacting DNA loci. Cross-linked chromatin was digested with MboI endonuclease (New England Biolabs, R0147). Subsequently, the restriction fragment overhangs were filled in and the DNA ends were marked with biotin-14-dATP (Life Technologies, 19524-016). The biotin-labeled DNA was sheared and pulled down using Dynabeads MyOne Stretavidin T1 beads (Life Technologies, 65602). The in situ Hi-C library was amplified directly off of the T1 beads with 6 cycles of PCR, using Illumina primers and protocol (Illumina, 2007). Promoter capture was performed as described previously (Lindsey

paper). The in situ Hi-C library was hybridyzed to 81,735 biotinylated 120-bp custom RNA oligomers (Custom Array) targeting promoter regions (4 probes/ RefSeq transcription start sites). After hybridization, post-capture PCR (8 amplification cycles) was performed on the DNA bound to the beads via biotinylated RNA. Each library was sequenced on a full lane of an Illumina HiSeq 4000 machine to achieve sufficient read depth for interaction calling. Standard capture Hi-C analyses were performed as described in [119]. Analyses were performed using both CHiCago[120] and HOMER[121]. The quality and biological significance of identified interactions was assessed by integrating publicly available RNA-seq and H3K27Ac Chip-seq[122] data from the same cell type and examining the correlation between interacting genes, gene activity, and bound transcription factors.

2.3.3 Luciferase enhancer assays

Luciferase reporter assays were performed using the pGL4.23[luc2/minP] (Promega, cat: E8411) vector backbone. Thirteen ~1000 bp windows were selected surrounding highly significant oxygenated hemoglobin GWAS SNPs previously identified by the Di Rienzo lab in [11]. For each region, portions of the Tibetan genome containing the high altitude or low altitude alleles were amplified from Tibetan DNA samples stored in the Di Rienzo lab and described in [11]. The low and high-altitude alleles were amplified and added to the PGL4 vector using Gibson Assembly (New England Biolabs, E5510S). The vectors were transfected together with the SV40 - pRL Renilla Luciferase control vector (GenBank® Accession Number AF025844) into telomerase immortalized human aortic endothelial cells (teloHAEC) using the Neon Electroporation system or into human embryonic kidney cells (HEK293T) using jetPRIME (PolyPlus, 114-07). Follow up transfections were performed in mouse immortalized atrial-derived cardiomyocytes (HL-1) using jetPRIME. HL-1s were chosen as a model for heart cells

as human cardiomyocytes are difficult to transfect and maintain in culture over multiple passage [123]. Following transfection, cells were placed in either normoxia or hypoxia for 48 hour to recover, then lysed and assessed for firefly luciferase expression which was normalized to the *Renilla* luciferase expression in accordance with the Dual Luciferase Reporter Assay system from Promega (cat: E1910). The 48 hour time point was chosen based on tests with GFP transfections to optimize expression of transfected vectors in this cell type. Assays were run on the GloMax-Multi Detection System with Dual Injector (E7031). In addition to the *EPAS1* constructs, additional cells were transfected with a negative control (the PGL4 vector containing a gene desert approximately the same size as the *EPAS1* regions) or a positive control (the PGL4 vector containing an SV40 early enhancer/promoter). Normalized expression from the *EPAS1* constructs were analyzed relative to the negative control in a given condition (normoxia or hypoxia) and paired t tests were used to test for differences between alleles. To account for experimental variability, triplicate construct DNA preps were made for each allele and each was transfected in triplicate; each genomic segment was tested in a minimum of 2 independent batches of cells (for a total of minimum 6 transfections per vector). This level of replication was necessary to detect subtle differences in allelic enhancer effect with statistical confidence. Tests with a coefficient of variation > 0.2 among triplicate transfection replicates were discarded. P-values from multiple tests were combined using Fisher's combined probability test.

2.3.4 Generation of CRISPR-cas9 deletion endothelial lines and musENH5-KO mouse lines

ENH5 deletion lines were generated in teloHAECs (ATCC) using a modified IDT ALT-R® RNP Protocol. Four crRNA's were designed to flank the region containing ENH5 using IDT's lt-R CRISPR HDR Design Tool. Guides were selected based on optimizing location and predicted efficacy, while minimizing the possibility of off target effects. The four crRNA's were

annealed separately to tracrRNA to create 4 complete guide RNAs. RNPs were complexed by incubating the guides with purified Alt-R® S.p. Cas9 Nuclease V3 (IDT, 1081058). The pool of four RNPs were transfected into teloHAECs using jetCRISPR (PolyPlus, 502-07). Due to the difficulty of transfecting endothelial cells, this transfection was serially performed 2-4x as described in [124] as cells were expanded and periodically checked by PCR for the presence of a deletion band. Once a strong deletion band was evident, the pool of edited cells were single-cell sorted into a 96 well plate on the FACSARIAIII at the University of Chicago Flow Core. Clones were genotyped using primers designed to flank the deletion site (table 2.1). Of all clones isolated, three knockout clones and three wildtype controls from the same pool of cells were selected for further analysis.

The three independent clonal isolates of ENH-5 KO and WT control endothelial lines were expanded and split into four 6-well cell culture plates and allowed to adhere overnight in normoxic incubators (37°C, 5% CO_2, 21%O_2). The next morning, the first plate of cells were lysed in plate using RLT Plus buffer from Qiagen in normoxic conditions. The remaining three plates were put into a normobaric hypoxia chamber in which neutral nitrogen gas is used to displace oxygen to desired oxygen concentration (Coy Lab Products, O2 Control InVitro Glove Box). Cells were placed in a humidified incubator contained within the box, and the entire box was maintained at 37°C, 5% CO2, 1%O2. At 24, 48, and 72 hours, one plate of cells was lysed inside of the chamber in hypoxic conditions, and the lysate immediately frozen at -80°C. Once all samples had been harvested, lysates were thawed together and RNA was extracted in parallel using RNeasy Plus Mini kit (Qiagen, 74134) in a Qiagen Qiacube. cDNA was generated using the iScript cDNA synthesis kit (biorad, 1708891) and RT-qPCR performed on a QuantStudio 6 Flex machine using qPCRBIO SyGreen Blue Mix Lo-ROX (PCR Biosystems, PB20.11-05).

Primers used may be found in table 2.1. Because hypoxia is an extreme stress on cells and causes a global change in gene expression, special considerations needed to be taken for qPCR analysis. Many typically used qPCR control genes, such as *GAPDH*, are known to be downregulated by hypoxic stress [125]. Therefore, *GAPDH* would be a poor internal control for detecting gene expression differences at different hypoxic time points. Instead, we chose *RPLP0* as an internal control as this gene has been shown to be stably expressed in hypoxia in the endothelium[125].

Similar to our approach used in the CRISPR-modified teloHAECs, mouse crRNA was designed to flank the orthologous ENH5 region in the mouse genome (~800bp). RNPs were generated as described above, but were resuspended in nuclease-free water rather than resuspension buffer prior to use. The mixture of two RNPs was given to the University of Chicago Mouse Core for microinjection and implantation. Founder mice were genotyped using both internal and external primers relative to the deletion site. The internal primers were necessary as external primers would preferentially amplify the deletion band and heterozygous mice were challenging to identify. A total of 7 heterozygous mice were identified to form the founder population and were bred for at least two generations prior to downstream analyses. Sanger sequencing around the deletion site was performed in the homozygous founders to ensure proper deletion without small insertion events at the cut site.

Table 2.1: All primers used in CRISPR analyses in human cell lines and mice. RT-qPCR primers were used to assess expression of genes in our endothelial ENH5 KO compared to WT controls. Genotyping primers were used to genotype and identify presence of the desired deletion in both human and mouse deletion lines.

Gene	Left Primer	Right Primer	purpose
h*EPAS1*	CGGAGGTGTTC TATGAGCTGG	AGCTTGTGTG TTCGCAGGAA	RT-qPCR
h*EGLN3*	ATC CGG GAAAT GGAACAGGT	GCAGGATCC CACCATGTAGC	RT-qPCR
h*PRKCE*	CGAGGCCGTGA GCTTGAAG	GCAATGTAGGG GTCGAGAAGG	RT-qPCR
h*RPLP0*	TCAGCAAGTGGGA AGGTGTAATC	TCAGCAAGTGGG AAGGTGTAATC	RT-qPCR
h*EPAS1*	ACC GGG GCT ATTTTTAGCAT	CAG GGT CTCT TCCCTCTGTG	External deletion genotyping primer
m*EPAS1*	ACTGACTGTA GCGCCCCATA	GTCCAGCACG CT TACTTCCA	External deletion genotyping primer
m*EPAS1*	CTGCCACCTG CTAGTTCACA	CTACATCCTG GCCATTTGCT	Internal deletion genotyping primer

2.3.5 Transcriptional profiling of WT and KO mice

The mouse work described here was approved by the Institutional Animal Care and Use Committee of the University of Chicago. All WT and KO mice were age, cage, and sex matched. The average weight of these animals was 22.67 ± 0.06 g for the KO's and 22.17 ± 0.06 g for the WT's. At age 9 weeks the mice were placed in pairs into a custom hypobaric hypoxia chamber described in ref [126] . The pressure was set to 470-480mmhg which is equivalent to ~8% O_2 and the mice were left overnight (~16.5 hours). The following morning, both mice were removed together and euthanized one at a time in a randomly selected order via urethane injection as described in [127]. Mice were rapidly dissected and left kidney, right adrenal gland, left lung, left and right atria, and left and right ventricle were all independently flash frozen in either RLT Plus buffer (kidney, adrenal gland, lung) or Trizol (all heart tissues) and stored at -80 degrees. Once all tissues had been collected from all three pairs of mice, tissues were removed from -80

and homogenized. RNA was extracted one tissue at a time using RNeasy Plus Mini kit (Qiagen, 74134) in a Qiagen Qiacube for all samples stored in RLT buffer and by hand using RNeasy plus mini kit with Qiagen Dnase treatment (RNase-Free DNase Set, cat#79254). RNA quality was assessed on an Agilent BioAnalyzer and RIN was recorded for each sample. RNA-seq library preparation was performed using the TruSeq RNA Library Preparation Kit v2, Set B (Illumina, RS-122-2002). Libraries were submitted to the University of Chicago Genomics Core where they were sequenced across 7 lanes of the Illumina HiSeq4000 (one organ per lane, balanced by genotype) with an average of 69.6M 50bp single end reads per sample.

Analysis of RNA expression data was performed independently in each tissue using a standard RNA-seq pipeline. Briefly, reads were aligned using STAR and read counts assessed with RSEM. Data were filtered to remove unannotated and lowly expressed genes (only genes with cpm >1 for at least 1/3 of the libraries in a given tissue were included) and differential expression analyses performed using limma-voom with standard settings[128], [129] Data were modeled as $\sim 0 + \beta_{\text{genotype}} Genotype_i + Batch + RIN\ (or\ Time)$. Where *Batch* is the pairs of mice that were treated together, and *Time* is the approximate time which the mice spent alive in normoxia prior to euthanasia. P-values determined by limma from each tissue were converted into z-scores and merged into a single matrix containing genes which appeared in all 7 datasets leaving 11,307 genes. This matrix of z-scores was put into the mashr program, which leverages data from all 7 tissues to infer patterns of differential gene expression[130]. Correlation due to sample overlap among tissues was accounted for by using estimate_null_correlation_simple()as described in the mashR vignettes (https://github.com/stephenslab/mashr/blob/master/vignettes/). Both canonical and data-driven covariance matrices were tested to generate a final dataset of differentially expressed (DE) genes. We ultimately chose to analyze the results using data-drive

covariance matrices as it demonstrated the best fit as measured by log-likelihood comparisons among models.

Gene ontology (GO) enrichment analysis was performed in single tissues using the goana function in the limma package, which take the output from limma differential expression analysis and look for overrepresentation of GO terms in the up and downregulated genes at an FDR of 0.1 [131]. GO analyses on the significant gene list from mashr was performed using the g:Profiler online platform, which uses a cumulative hypergeometric test for functional enrichment [132]. The platform offers multiple data sources including gene ontology [133], pathway enrichment [134], [135], and Human phenotype ontology [136], all of which were considered. In the case of both goana and g:Profiler, all annotated mouse genes are used as the background universe.

2.4 Results

All *EPAS1* variants with association and selection signals are non-coding and hence are likely to affect gene regulation. Previous work in the DiRienzo lab had noted that the Tibetan alleles carrying the top signatures of selection and Hb association in *EPAS1* fell into a genomic region decorated with H3K27Ac marks in Human Umbilical Vein Endothelial Cells (HUVECs) implicating that they may play a role in regulating *EPAS1* in the endothelium [11], [137]. The endothelium is the predominant site of *EPAS1* expression. As the first point of contact for oxygenated blood, the endothelium is a crucial tissue in the physiological and pathological responses to hypoxia [53], [138]–[140]. Several Tibetan phenotypes also implicate the endothelium as a potentially significant tissue in adaptation, in particular the low rates of hypoxia-induced pulmonary hypertension [103]. Indeed, HIF-2α activity in the endothelium is essential for the pathogenesis of hypoxia-induced pulmonary hypertension [40], [141], [142].

Reduction or deletion of HIF-2α in the endothelium results in protection from pulmonary hypertension (PH), pointing to this tissue as an important site of adaptations in Tibetans. For these reasons, we decided to start by focusing on endothelial cells to provide a thorough characterization of the regulatory landscape of *EPAS1* and of the allelic effects of the Tibetan SNPs at this locus.

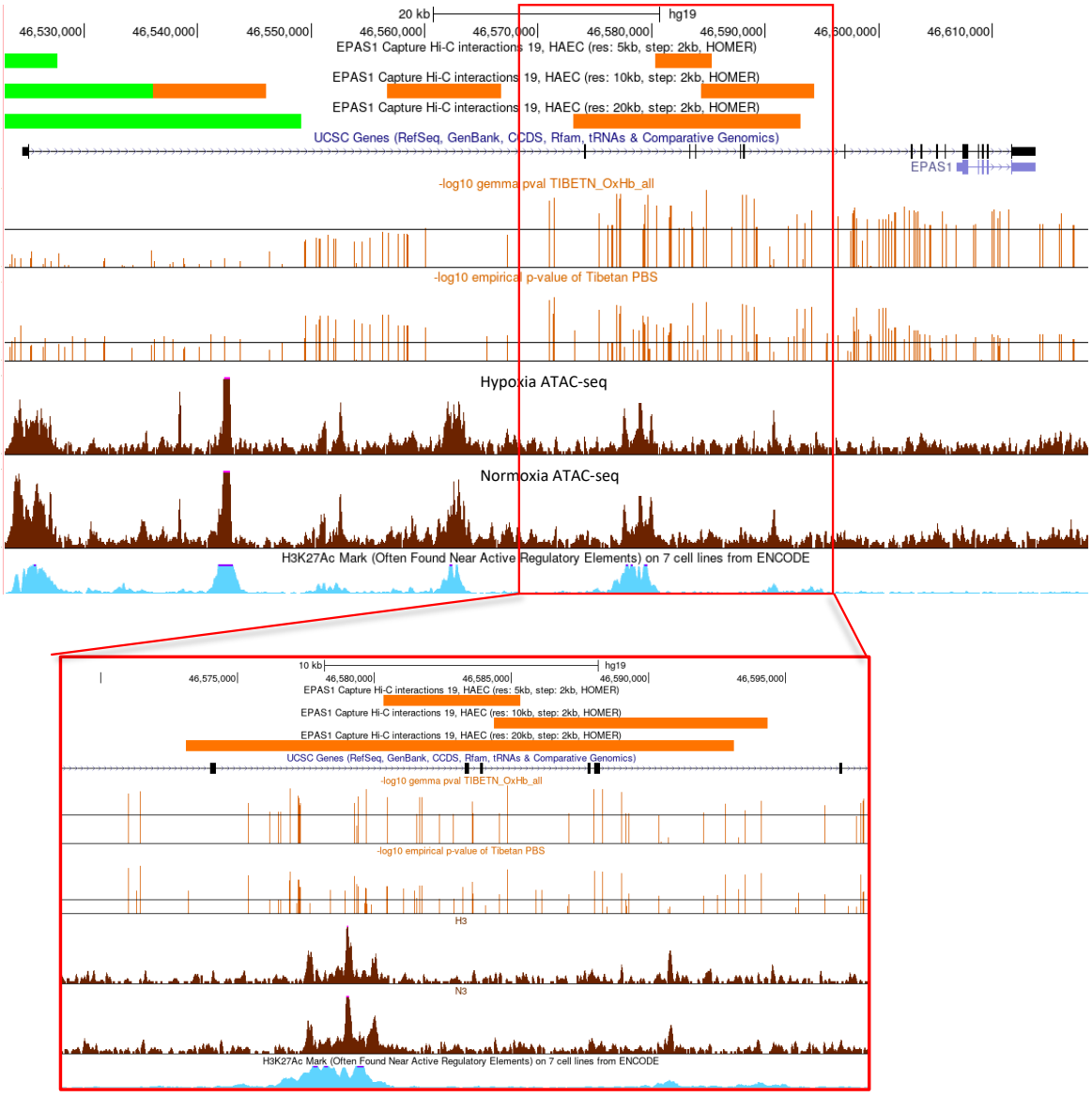
2.4.1 The regulatory landscape of the *EPAS1* locus in endothelial cells

To this end, we performed ATAC-seq in primary human aortic endothelial cells (HAECs) cultured in normoxia and hypoxia. This not only allowed us to visualize the chromatin accessibility at the *EPAS1* locus, but also allowed hypoxic changes in chromatin conformation in a primary endothelial context to be quantified genome wide. Because hypoxia is a strong cellular stress which causes broad changes in gene expression and chromatin accessibility [143], [144], it was crucial to know if regions predicted to be open and have regulatory activity in publicly available normoxic data remained so in hypoxia. At the *EPAS1* locus, no differentially open peaks were identified in the area.

Next, to determine if the alleles carrying the selection and association signals were looping to the *EPAS1* promoter or some other more distal gene, we examined chromatin looping using promoter capture Hi-C in HAECs.. We found evidence of extensive looping between the *EPAS1* promoter and upstream regions of the genome, containing the gene Protein Kinase C Epsilon (*PRKCE*). This serves as a positive control as these upstream regions contain dozens of GWAS signals for hematological traits observed in European cohorts as well as eQTLs for *EPAS1*[115], [145]. Within *EPAS1* itself, however, we only identified five regions, three of which were overlapping (ascertained in different window sizes), that interacted with the *EPAS1* promoter (fig 2.1). Critically, the large area of interaction defined by the three overlapping

regions encompasses nearly all of the strong Hb association signals and selection signals indicating that the Tibetan alleles in this region exert a regulatory effect on *EPAS1* in endothelial cells.

Figure 2.1: A view of the UCSC genome browser [146] containing, from top to bottom: Capture Hi-C results in HAECs (green represents promoter region, orange is the interacting regions inferred by homer on variable window sizes); GWAS results for oxygenated hemoglobin concentration and selection signal as measured by the population branch statistic, both from [11], each vertical bar represents one SNP, and the height of the bar is the -log10 P-value; two tracks showing open chromatin detected by ATAC-seq in HAECs grown in either hypoxia (1% O_2, 24 hours) or normoxia (20% O_2 , 24 hours); publicly available H3K27Ac data in human umbilical vein endothelial cells (HUVECs) from ENCODE (the peaks mirror open chromatin closely, as would be expected for active enhancer regions). The zoom focuses on the inferred regions of chromatin looping that falls into the region containing the strongest GWAS signals.



2.4.2 *EPAS1* contains multiple enhancer regions with allele-specific effects in the Endothelium

In order to identify the Tibetan allele(s) with regulatory effects among those in strong LD with each other, we generated a library of luciferase enhancer constructs to tile across the entire region of significant Hb GWAS and selection signals. These constructs contained, respectively, the high-altitude and low-altitude alleles. In total thirteen genomic segments, each with two alleles, were tested in parallel in HAECs cultured in both normoxia and hypoxia (1% O_2, 48 hours). This analysis identified three enhancer regions - ENH4, ENH5 and ENH8 – which showed both enhancer activity relative to the negative controls and significant allelic effects in with the high altitude construct showing weaker activity than the corresponding low altitude one (table 2.2). All three regions are located in or near open chromatin peaks identified in our ATAC-seq data and within the chromatin found to loop the *EPAS1* promoter in our capture Hi-C data, consistent with being active enhancers in this cell type.

2.4.3 Assessing enhancer and allele-specific activity in the kidney

Because the alleles found within the endothelial enhancers are highly associated with hemoglobin concentration in Tibetans, we reasoned that these enhancers may exert pleiotropic effects in the kidney, where HIF-2α activate *EPO* to stimulate the erythropoietic response [11], [79], [81], [147]. Alternatively, other enhancers and alleles within the extended haplotype at the *EPAS1* may be active in the kidney and may be responsible for the association with Hb levels. In order to ascertain the pleiotropic effect of the Tibetan haplotype at *EPAS1* in the kidney, we transfected the same 13 luciferase constructs into embryonic kidney cells (HEK293T) and measured their activity in normoxia and hypoxia. We discovered that the same three enhancers we had previously identified maintained consistent effects in the kidney. One additional region,

ENH6, demonstrated enhancer activity and allelic effects in the same direction that was unique to HEK293T cells.

Table 2.2: A table of Tibetan SNPs contained in each validated enhancer region. Boldface SNPs have genome-wide significant p-value in oxygenated Hb GWAS, and those with * have a derived alleles shared between Tibetans and Denisovans [11]

Region	Significant SNPs	Cell Type	p-value normoxia	p-value hypoxia
ENH4	**rs76242811***, rs6544887, **rs188801636*,** rs6544889	teloHAEC	0.001	2.33E-06
		HEK293T	0.0056	0.032
ENH5	rs4953357, rs6756667, **rs375554942***	teloHAEC	0.0453	0.007
		HEK293T	1.86E-05	3.60E-05
ENH8	**rs142826801**, rs74898705*	teloHAEC	0.017	.009
		HEK293T	0.025	1.00E-05

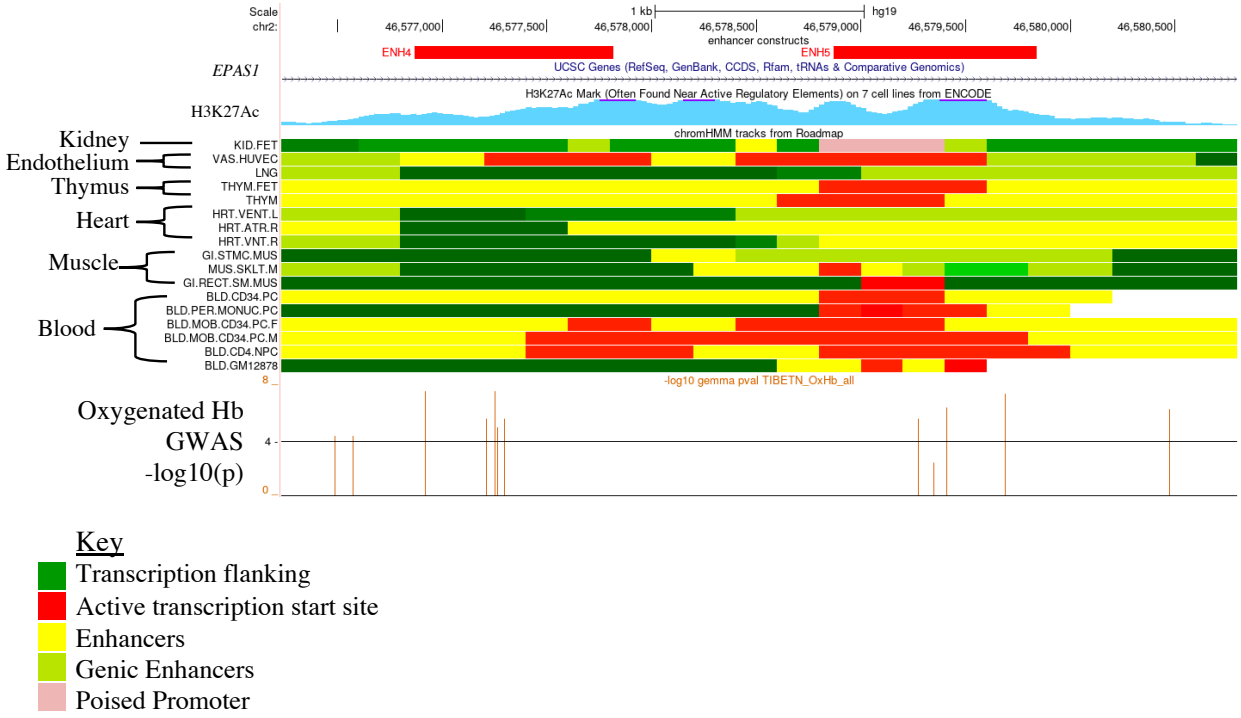
The most interesting of the three enhancers with shared activity in endothelium and kidney was ENH5. In addition to consistently showing robust allele-specific activity in both cell types, ENH5, also consistently showed higher activity in hypoxia compared to normoxia. Notably, both alleles displayed a stronger enhancer activity under conditions of hypoxia than normoxia. The presence of a hypoxia-responsive enhancer containing some of the top signals of selection in the Tibetan genome was extremely promising, and we chose to further characterize this enhancer in downstream analyses.

2.4.4 Enhancer ENH5 displays hypoxia responsive, and allele-specific activity in cardiomyocytes

A close examination of the region containing ENH5 in the comprehensive database of chromatin HMM data from the NIH Roadmap Epigenomics Mapping Consortium and the ENCODE consortium [122], [137], [148], [149] revealed evidence of strong enhancer activity in a wide variety of cell types beyond endothelium and kidney (fig 2.2). In fact, the region

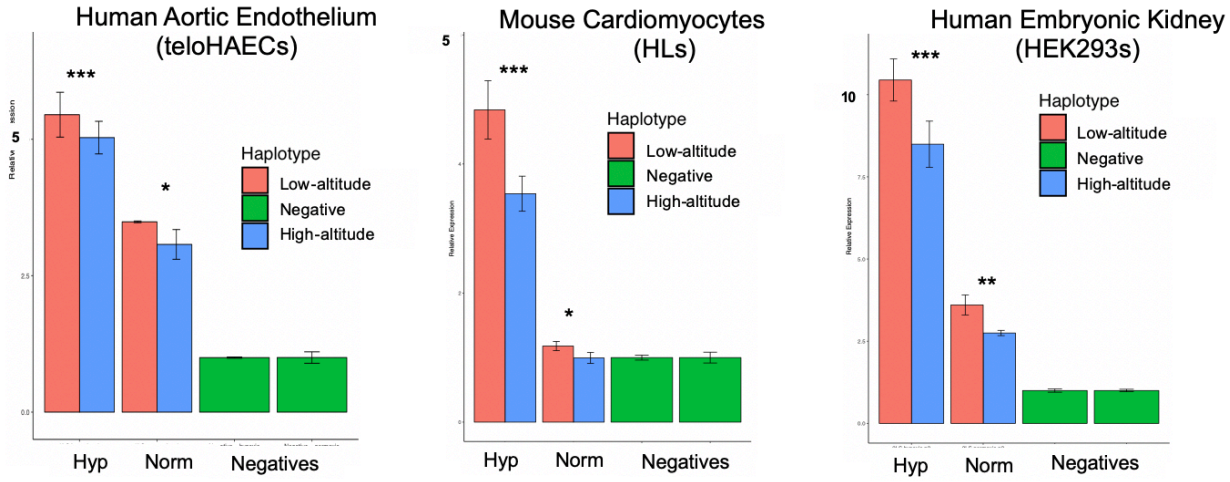
spanning ENH5 in particular is inferred to have a TSS or TSS proximal-like chromatin state, consistent with evidence that strong enhancers can be identified by active transcription at the enhancer site[150]–[153] . Interestingly, a number of cardiac tissues – both left and right ventricles, right atrium (no left atrium data was available), and fetal heart – showed strong marks of enhancer activity for ENH5. Like endothelium, cardiac cell types are important in both biological and pathological responses to hypoxia[54], [85], [154], [155]. Additionally, recent work from the Di Rienzo lab found evidence for polygenic selection favoring lower pulse in Tibetans [11]. Promoter capture Hi-C data from cardiomyocytes indicates that ENH5 loops to the *EPAS1* promoter in this cell type as well [119].

Figure 2.2: Chromatin HMM data displayed on the UCSC genome browser [146], [148]. These data come originally from the Epigenomics Roadmap Project, and display predicted chromatin state across the genome[122], [149]. The color key below indicates inferred chromatin state in multiple cell types in the region containing ENH4 and ENH5.



In order to assess ENH5 activity in the heart, we transfected the luciferase constructs into mouse cardiomyocytes (HL-1s) following the same parallel hypoxia-normoxia protocol we had previously used for kidney and endothelial cells. In HL-1s, ENH5 demonstrated the same patterns of expression as seen in kidney and endothelium (P-value normoxia = 0.092 , P-value hypoxia = 0.0016). In fact, the hypoxia-responsive activity of ENH5 was even more striking in cardiomyocytes than in kidney and endothelial cells, with significantly higher activity in hypoxia and little to no enhancer activity detected in normoxia (fig2.3).

Figure 2.3: Representative luciferase assay results in our three cell types of interest: endothelium, cardiomyocytes, and kidney. These results reflect a single plate replicate with three transfection replicated per vector. The significance indicates the fishers combined p-value of all luciferase assays performed in the indicated cell type.

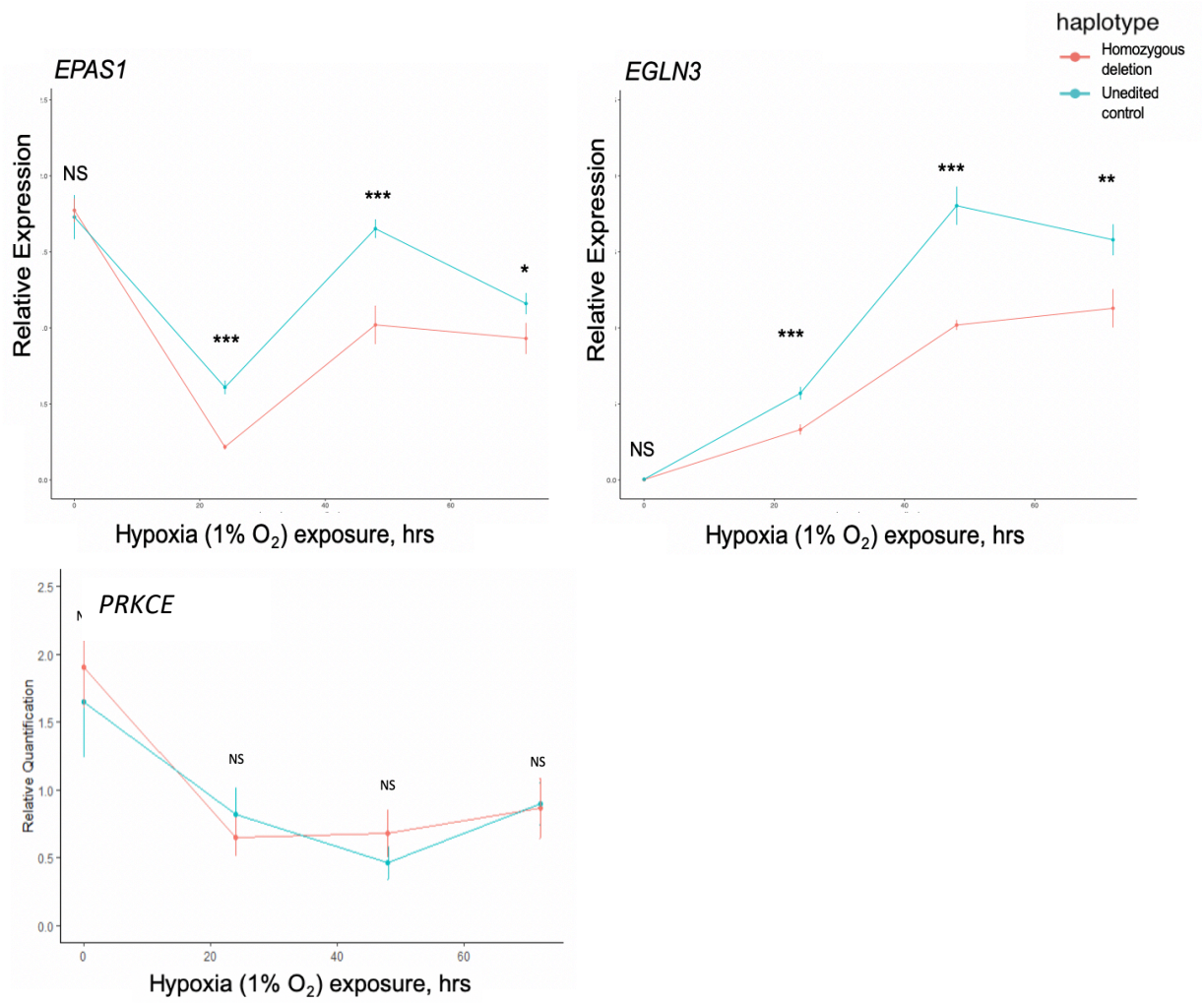


2.4.5 CRISPR-mediated deletion of ENH5 in endothelial cells results in downregulation of *EPAS1* and HIF-2α target gene *EGLN3*:

Because reporter gene assays do not assess regulatory activity in the endogenous chromatin context, we used CRISPR-mediated editing to delete the entire ENH5 region in teloHAECs and assessed the effects of the deletion on transcript levels on a hypoxia time course.

Our results, shown in figure 2.4, demonstrated significant downregulation of *EPAS1* across hypoxia timepoints. Interestingly, deletion of ENH5 did not appear to significantly alter *EPAS1* expression in normoxia. This pattern may reflect the hypoxia-responsive activity seen at ENH5 in our luciferase data.

Figure 2.4: qPCR results for three genes in the endothelial ENH5 KO clones over 72 hours in hypoxia. The red line indicates relative exprssion of the three homozygous deletion clones, while the blue line indicates the WT controls. Error bars reflect the SEM values for triplicate qPCR reactions performed in each of three clones. *EPAS1* is the predicted target of ENH5, *EGLN3* is a HIF-2α target located on a different chromosome, and *PRKCE* is a gene located just 5' of *EPAS1* and is included as a negative control.



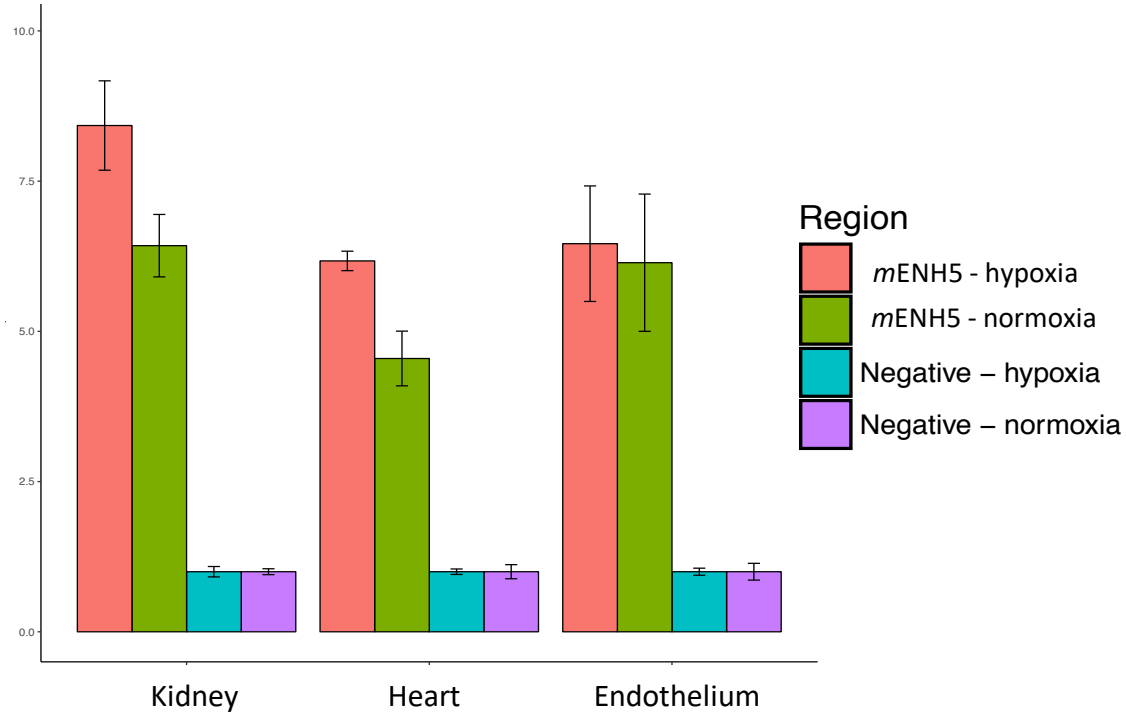
Because HIF-2α is known to be post-translationally regulated by hydroxylase activity[156], [157], it was necessary to determine if the observed downregulation of *EPAS1*

would have functional consequences on downstream HIF-2α targets. To this end, we examined the expression of a known HIF-2α target gene in the endothelium, i.e. *EGLN3* [42]. In hypoxia, *EGLN3* is known to be upregulated by HIF-2α in the endothelium, making its activity a good proxy for the abundance and activity of HIF-2α in our ENH5 deletion lines. Again, we found that deletion of ENH5 leads to significant downregulation of *EGLN3* across hypoxia time points, indicating that the transcriptional changes observed in *EPAS1* expression had downstream consequences on a HIF-2α target. We also examined the expression of *PRCKE* as a negative control. While *PRKCE* is located 5' to *EPAS1* and shares many chromatin contacts, ENH5 was not shown to loop to the *PRCKE* promoter, and was therefore not predicted to be an enhancer of *PRKCE*. No significant changes in *PRKCE* expression were detected in ENH5 knockout lines relative to WT controls.

2.4.6 musENH5 shows evidence of regulatory activity in multiple tissues *in vivo*

To better understand the degree of pleiotropic activity of ENH5 as well as its potential downstream consequences we chose to assess the effect of an *in vivo* deletion of ENH5. While the specific SNPs in ENH5 are not conserved in mice, the region itself is largely conserved at the sequence level. To test its functional conservation, this region, dubbed *mus*ENH5, was cloned into a luciferase vector and transfected into the same three cell types used to test the human ortholog. The results, shown in figure 2.5 confirmed that *mus*ENH5 functions as a strong enhancer in all three cell types of interest. Importantly, the mouse enhancer also appeared to have stronger activity in hypoxia than normoxia just like the human ortholog.

Figure 2.5: Luciferase assay results for the *mus*ENH5 allele transfected into the three cell types of interest in normoxia and hypoxia. The cell types shown here are the same as those used in the enhancer assays using human alleles. Kidney – HEK293T, Heart – HL1, Endothelium - teloHAEC



We deleted *mus*ENH5 using CRISPR RNPs microinjected into fertilized eggs and bred a stable deletion line from the resultant mice. We treated age- and cage-matched pairs of male mice overnight in hypobaric hypoxia (16 hours, 8% O2) in order to elicit an acute hypoxia response. We reasoned that if *mus*ENH5 impacted *EPAS1* expression in one or more tissues, we would be able to identify the downstream consequences of decreased HIF-2α activity only after exposure to hypoxia. We assessed the transcriptional profile by RNAseq in 7 distinct organs from each mouse: lung, kidney, left atria, right atria, left ventricle, right ventricle, and adrenal gland.

The transcriptional data were first tested for differential expression across genotypes in each tissue independently using limma. In total 201 genes were found to be differentially expressed across all tissues, the majority identified the left atrium. Many of these, including *Angpt1*, *Malat1,* and *Cox2* are known to be upregulated by HIF-2α and are all downregulated in the KOs as predicted [158], [159], [160, p. 2]. Of note, downregulated genes in this tissue are enriched for terms relating to heart rate and cardiac contraction. Given the polygenic adaptation signal for reduced heart rate seen in Tibetans, it is interesting to consider that the ENH5 allele may contribute to this phenotype. Although *EPAS1* itself was not found to be significantly differentially expressed in any one tissue at the genome-wide level, we did find it to be significantly downregulated at the single gene level in left atrium and lung of the KO mice indicating that deletion of *mus*ENH5 altered expression of *EPAS1* as predicted in at least these two tissues.

Comparison of transcriptional patterns for genes expressed across tissues revealed correlations amongst almost all tissues. An exception was observed for left ventricle where the extent of differential expression across genotypes was only loosely correlated with that of other tissues or showed opposite directions (e.g. up-regulated in left ventricle and down-regulated in right ventricle). As we were unable to identify whether this was a biologically relevant effect or an artifact in the data, we performed downstream analyses by excluding the left ventricle data. However, inclusion of data from this tissue did not substantially alter the results.

In order to leverage the shared effects across tissues, we used multivariate adaptive shrinkage (mashr) on the data from 6 of the 7 tissues. Mashr leverages the combination of many effects (in our case gene expression differences between KO and WT) ascertained across many conditions (in our case independent tissues) to increase power and model effect sharing across conditions

[130]. Significance in mashr is calculated in each tissue independently using the local false signs rate (lfsr). We identified 1007 genes that were significant in at least one tissue (lfsr < 0.05), with the greatest number of significant genes being identified in the kidney and lung (table 2.3). *EPAS1* was identified as being downregulated in the KO mice across all tissues, and it was borderline significant at the genome-wide level in the left atrium (lfsr = 0.09), lung (lfsr = 0.06), and kidney (lfsr = 0.06). This finding is consistent with our expectations a broadly pleiotropic activity for ENH5.

Table 2.3: Number and direction of differentially expressed (DE) genes identified by mashr in 6 tissues

Tissue	**# of DE Genes**	**% Up**	**% Down**
Lung	687	38%	62%
Left Atrium	596	44%	55%
Right Atrium	98	78%	22%
Right Ventricle	193	71%	29%
Adrenal Gland	201	44%	56%
Kidney	703	47%	53%

Table 2.4: Several genes of interest identified in the mashr analysis of the *mus*ENH5 KO data. The sign indicates the direction of effect, genome-wide significance is calculated by the lfsr statistic, and is indicated as follows: * < 0.01, ** < 0.05, *** < 0.01.

Gene	*RA*	*LA*	*RV*	*lung*	*KID*	*AG*
Epas1	-	- *	-	- *	- *	-
Malat1	-	- ***	- *	- ***	- ***	- ***
Peg3	- **	- ***	- **	- ***	- ***	- **
Id3	+ **	+ ***	+ ***	+ ***	+ ***	+ **
Kmt2d	-	- *	-	- ***	- **	- **
Kmt2a	-	-	-	- ***	- *	- **
Cox2	-	- ***	- *	- ***	- ***	-
Aplnr	+	- **	+	+	+ *	+
Myc	+ *	+ **	+ **	+ *	+ **	+ *

We identify many hypoxia responsive genes and HIF-2α targets among those that are significantly differentially expressed in more than one tissue, with the direction of effect typically being consistent across all tissues. Some of these genes are shown in table 2.4 including: *Malat1* – a long non-coding RNA known to be regulated by HIF-2α in certain cancers [161], *Peg3* – a hypoxia-induced gene reported to be involved in apoptosis/cell death in response to hypoxic or reperfusion injury [162, p. 3], *Id3* – a gene known to be involved in signaling and proliferation of endothelial cells which is induced in hypoxia by HIF-2α and may be involved in the pathogenesis of hypoxic pulmonary hypertension [142], Kmt2d and Kmt2a – genes which code for the MLL1/3 proteins, H3K4 methyltrasnferases responsible for depositing the H3K4me3 mark on active chromatin, also thought to be important to early hematopoiesis [163], *Cox2* – a gene activated by Hif-2α in certain cancers and is also known to be involved in cardiac rhythm and inflammation[158, p. 2], [160, p. 2], [164], *Aplnr* – the APLN receptor which, together with APLN promotes vasodilation and coveys protection against pulmonary hypertension [165], and *Myc* – a potent transcription factor and proto-oncogene reportedly suppressed by HIF-2α in pulmonary endothelial cells [166]. .

Gene ontology analysis of all significant genes detected an abundance of enriched terms related to angiogenic, metabolic, and apoptotic processes, as well as several relating to hypoxia response, stress response, and hematopoiesis (table 2.5). Of interest, when the significant genes are divided into those upregulated vs downregulated across tissues, we find that the enrichment in terms relating to apoptosis and angiogenesis comes primarily from the genes identified as upregulated in the ENH5 KO mice. In addition, we find that the downregulated genes are enriched for human phenotype ontology terms HP0004890 “Elevated pulmonary artery pressure”

and HP:0002092 “Pulmonary arterial hypertension” [132], [136]. These include genes such as *Cox2, Cox3 , Notch1, Hspg2* and *Cited2* all of which have been implicated in PH pathogenesis and some of which, like *Cox2 and Cited2* are known to be directly regulated by HIF-2α[160], [167]. These findings support the hypothesis that downregulation of *EPAS1* through ENH5 may reduce the activity of genes involved in the pathogenesis of pulmonary hypertension, thus providing protection against a deleterious phenotype. Taken together, we find that deletion of the *mus*ENH5 impacts the expression of *EPAS1* and its targets in multiple hypoxia-responsive tissues, consistent human ENH5. Critically, these results implicate a wide range of possible phenotypic outcomes in different tissues resulting from a downregulation of *EPAS1* via ENH5, one or more of which could confer a fitness advantage in Tibetans.

Table 2.5: Gene ontology and pathway analysis results of all significant DE genes from the mashr analysis. Terms of interest are listed.

Source	Term ID	Term Name	Padj
GO:BP	GO:0019222	regulation of metabolic process	3.78E-22
GO:BP	GO:0031323	regulation of cellular metabolic process	4.23E-22
GO:BP	GO:0001944	vasculature development	2.46E-15
GO:BP	GO:0001568	blood vessel development	1.69E-14
GO:BP	GO:0006950	response to stres	3.50E-13
GO:BP	GO:1901700	response to oxygen-containing compouond	1.44E-10
GO:BP	GO:0048514	blood vessel morphogenesis	3.05E-10
GO:BP	GO:0012501	programmed cell death	1.00E-06
GO:BP	GO:0006915	apoptotic process	1.74E-06
GO:BP	GO:0042981	regulation of apoptotic process	2.48E-06
GO:BP	GO0048534	hematopoietic or lymphoid organ development	4.21E-06
GO:BP	GO:0030097	hemopoiesis	4.46E-06
KEGG	KEGG:05310	Asthma	1.38E-04
GO:BP	GO:0036293	response to decreased oxygen levels	2.46E-04
GO:BP	GO:0060541	respiratory system development	9.38E-04
GO:BP	GO:0080182	H3-K4 trimethylation	3.98E-03
KEGG	KEGG:04210	Apoptosis	1.31E-02
GO:BP	GO:000666	response to hypoxia	1.97E-02

2.5 Discussion

The Tibetan genome has been repeatedly scanned for signatures of selection in the hope of finding the key to the adaptive mechanisms that allow Tibetans to survive the harsh conditions of chronic exposure to high-altitude hypoxia[11], [80], [168]. These efforts have consistently identified the strongest signature of positive selection in the *EPAS1* gene, with extreme allele frequency divergence between Tibetans and closely related lowlanders consistent with strong selection [11], [79], [80], [168]. In the work described here we present a detailed analysis of the *EPAS1* locus and high-altitude Tibetan alleles in the endothelium, which is the primary site of EPAS1 expression and key tissue in the response to hypoxia. We identify multiple enhancers

spanning SNPs with strong selection signals. Through a combination of genome-wide interrogations of chromatin architecture and *in vitro* reporter assays in multiple hypoxia-responsive cell types we identify three pleiotropic enhancers whose activity is altered by Tibetan high-altitude alleles. One of these enhancers, ENH5, is distinct in that it displayed not only allele-specific, but also hypoxia-responsive activity. We confirmed the activity of ENH5 *in vitro* by deleting the enhancer in endothelial cells. Moreover, deletion of the orthologous region in mice demonstrated that this pleiotropic activity is replicated *in vivo*.

EPAS1 was found to be downregulated at the single gene level of significance in both heart and lung tissues in the individual tissue-specific analysis, and in all six tissues used in the mashr analysis (approaching genome-wide significance in three of them). Given the dynamic activity of *EPAS1* expression in the first 24 hours of hypoxia observed in both our own analysis and in the literature [138], and the fact that it is not believed to be the primary HIF involved in response to acute hypoxia [169], it is perhaps not unexpected that we do not detect *EPAS1* itself being downregulated with a genome-wide level of significance. Instead, detecting the genes and gene modules altered by a reduction of HIF-2α is a more reliable readout of the *mus*ENH5 deletion. We identify DEGs in all six tissues included in our mashr analysis, which are enriched in many terms related to HIF-2α activity such as response to low oxygen, angiogenesis, apoptosis, and metabolism. hemoglobin.

2.5.1 Pleiotropy in hypoxic adaptation

Much of the previous research into *EPAS1* variation in Tibetans has focused either on its connection to hemoglobin concentration or assesses the impact of alleles in one cell type, typically the endothelium [78, p. 1], [79, p. 1], [170]. Yet, the pleiotropy we identify in ENH5 and other enhancers affected by Tibetan alleles could provide a unique insight into the potential

mechanisms of positive natural selection acting at this locus. Several lines of evidence suggest that selective sweeps are not to be the main mode of adaptation in recent human evolution [18], [20]. However, exceptions to this rule could arise due to highly pleiotropic alleles which can confer multiple fitness effects precisely because they are pleiotropic. In essence, fitness effects due to different adaptive phenotypes could drive a very strong selection coefficient over time at a locus. This has been proposed to explain the high selection coefficient on the *EDAR* 370A coding variant in East Asian populations. An interdisciplinary study revealed that this variant seemed to affect a wide range of potentially adaptive phenotypes in a humanized mouse model including hair texture and thickness, mammary gland structure, and increased eccrine glands [32]. While Tibetan variants in ENH5 are non-coding, the fact that they appear to impact expression in so many tissues gives these alleles a comparable degree of pleiotropic activity.

For example, the Tibetan alleles at ENH5 in the kidney may be responsible for the downregulation of *EPO* in turn resulting in unelevated hemoglobin concentration and providing some protection against excessive erythrocytosis at high-altitude. In endothelial cells, where *EPAS1* is normally highly expressed, downregulation of *EPAS1* via ENH5 may result in the observed protection against pulmonary hypertension in Tibetans by altering the expression of PH-related genes involved in blood pressure and inflammation [40], [141], [142], [171]. For instance, reducing *EPAS1* has been shown to decrease expression of *EDN1*, a key hypoxia-responsive vasoconstrictor, while increasing expression of *APLN* and its receptor *APLNR,* a vasodilator, thereby reducing blood pressure [139], [142]. We see this reflected in our *in vivo* data in which *Edn1* is downregulated in the left atrium (lfsr = 0.066) while *Aplnr* is upregulated in the left atrium, right atrium, lung and kidney (lfsr = 0.04, 0.1, 0.12, and 0.05 respectively). In the heart, downregulation of *EPAS1* through ENH5 may also contribute to the polygenic

adaptation towards decreased heart rate identified in Tibetan populations. For instance, inhibition of *COX2* has been linked to decreased heart rate in humans [172]. We find *Cox2* to be significantly downregulated in multiple tissues in the ENH5 KO mice, including atrial and ventricular tissues. Finally, we also see enrichment for terms relating to angiogenesis among the upregulated genes identified by mashr, which may also be linked to a fitness effect, as some evidence suggests that Tibetans have a greater capillary density than other high altitude populations [12]. Increased angiogenesis could provide a mechanism for improved oxygen delivery to compensate for the hypoxemia resulting from unelevated Hb. While all of these phenotypes may, individually, only contribute moderate increases in fitness, the fact that a single haplotype may contribute to all of them at once may sum to a much large fitness effect, driving the strong signals of selection we see at the Tibetan *EPAS1* locus.

2.6 Acknowledgements of work performed

I would like to thank all of our collaborators for their contributions to this work.

- Promoter capture Hi-C was performed by Debora Sobreira and analyzed by Noboru Sakabe
- Luciferase assays in HEK293Ts were performed and analyzed by Jordan Jousma
- Mouse hypoxia treatments, dissection, RNA extraction, and library preparation was performed by Jennifer Yoo with help from Ying-Jie Peng

CHAPTER 3

DEVELOPMENT OF A PANEL OF TIBETAN AND HAN CHINESE IPSCS: A UNIQUE RESOURCE FOR THE STUDY OF LOCAL ADAPTATIONS

3.1 Abstract

Recent innovations now enable reliable and highly successful reprogramming of terminally differentiated cells into an induced pluripotent state. While historically used to study human disease, large population panels of induced pluripotent stem cells (iPSCs) offer an unprecedented opportunity to study the functional and molecular impact of human adaptations to local environments in a cell-type specific manner. Here, we describe the generation of a new panel of 20 human iPSCs created as a tool to study genetic adaptions to high altitude hypoxia in Tibetan populations. The panel comprises 10 Tibetan (TBC) and 10 Han Chinese (CHB) samples, reprogrammed in parallel to iPSCs. We assessed the panel for uniformity and functional pluripotency, finding that all 20 lines can spontaneously differentiate into the three germ layers and express pluripotency factors at similar levels. No genes were found to be differentially expressed between populations.

3.2 Introduction

Genetic adaptations to local environments, a uniquely important aspect of the study of human evolution, are the consequence of the expansion of human populations 'Out of Africa' and around the globe (~50,000-70,000 years ago)[173]–[175]. During this dispersal, human populations spread into a wide variety of climatic conditions, developing unique cultures and

subsistence methods. This in turn exposed them to many unique environmental pressures such as different temperatures[6], different UV levels[8], [9], and oxygen pressures [10]–[13], as well as selective pressures from the advent of novel subsistence strategies and nutrient availability[3], [5], [176]. These conditions have shaped modern human diversity, allowing for the accumulation of advantageous, heritable traits in relatively recent evolutionary history[33].

Understanding the genetic causes and phenotypic impacts of human adaptations is an important goal of human genetics research. However, although many population genetic approaches have been developed and applied to uncover beneficial alleles[28]–[31], [177], [178], elucidating the functional, phenotypic and fitness effects of these genomic regions is much more challenging[33]. While some targets of selection lend themselves more readily to predictions about functional consequences, relevant tissues, and selected phenotypes (*i.e.* lactase persistence[3]) others are much more difficult to parse. Signals of polygenic adaptations may point to adaptive traits, but pleiotropy complicates the identification of the true selected phenotype(s) among correlated ones [28], [179], [180]. Moreover, signatures of selection at non-coding variants residing in regulatory elements like enhancers present a different set of challenges in the process of deciphering the target gene(s), tissues, and phenotypic consequences of local adaption.

Accurately identifying the functional and phenotypic consequences of adaptive alleles may have important implications not just for our understanding of recent human evolution, but also from a biomedical perspective. High altitude hypoxia, for example, may be seen as analogous to the pathological hypoxia and hypoxemia that causes or complicates many cardiovascular disorders[49], [58], [104], [181], [182]. In this paradigm, adapted human populations may be viewed as the outcome of a millennia long experiment into the most

successful mechanisms for combatting this disease-associated pressure. For instance, human populations have occupied the Tibetan plateau for well over 10,000 years[110], adapting to an environment with just 60% sea level oxygen concentration[12]. Hypoxia presents a severe and systemic pressure, and the physiological responses to it involve a multitude of cell and tissue types[37], [104], [159], [183], [184]. Though strong signatures of selection have been identified in the Tibetan genome, the precise causal alleles, target genes, and molecular phenotypic output of the adaptive alleles remains largely unknown. Moreover, due to the logistic challenges of collecting deep phenotype data at high altitude in both Tibetan natives and acclimatized lowlanders, we are still far from having a complete view of the evolved phenotypes that distinguish Tibetans from populations that have not adapted to hypoxic environments.

In vitro cell culture studies may offer a complementary approach to field studies of Tibetan phenotypes. Ideally, *in vitro* phenotypes could be studied in relevant cell types in a controlled environment and compared between Tibetans and a closely related low-altitude population such as the Han Chinese (CHB). However, this too is logistically challenging, as collection of the biological material required for such an experiment would be both impractical and likely impossible for study of populations located on remote parts of the Tibetan plateau far from optimally equipped labs for harvesting and analyzing primary tissue.

Fortunately, recent technological advances in the establishment and differentiation of human induced pluripotent stem cells (iPSCs) now provide a unique method to overcome these difficulties. iPS cell model systems have become increasingly used to analyze human functional genetic variation across cell types[93]–[96]. As easily expanded cell lines that can be differentiated into a variety of different cell types, iPSCs have become widely used to study disease[99]–[102]. Yet this model system may also be leveraged to ask nuanced questions about

the effects of local human adaption on gene expression and cellular phenotypes in a given cell type. Thanks to the efforts of large consortia such as the 1000 Genomes project, large repositories of cells for many global populations now exists in the form of lymphoblastoid cell lines (LCLs). These LCLs can now be readily generated into iPSCs with a greater than 95% success rate[95], [97], [98], thus opening up the possibility of differentiating a cell type of interest from many global populations for comparison. Characterizations of iPSCs reprogrammed from LCLs have shown that these stem cells not only maintain identifiable gene expression patterns among individuals, but that they are well suited to faithfully capture inter-individual phenotypic variation in a wide variety of cell types into which they can be differentiated[94]–[96]. Thus, iPSCs provide an incredibly valuable opportunity to study molecular and phenotypic variation both within and across populations in controlled groups and environmental conditions.

Additionally, iPSCs provide a unique opportunity to perform precise genetic manipulations in a population and cell type of interest that would often not be possible in cell culture[185]. Cell biobanks with multiple populations represented are relatively scarce and maintain only select cell types. Moreover, many primary cell types are unsuited to long term cell culture or are difficult to collect and establish. Established immortalized cell lines, meanwhile represent few individuals/populations and may not faithfully replicate the biology of primary cells. iPSCs, on the other hand can be easily edited and then differentiated. This enables the study of genetic manipulations in the genetic backgrounds of a population of choice and in cell types that may otherwise be inaccessible or difficult to edit, such as cardiomyocytes[123, p. 1].

Here, we report the establishment of a panel of Tibetan (TBC) and Han Chinese (CHB) iPSCs, which provides a renewable and expandable source of stem cells from these populations.

These iPS cells can be directed down different differentiation trajectories, exposed to hypoxia, and analyzed for a variety of phenotypic outcomes including differential gene expression, protein levels, cellular physiology, and more. This will help to uncover novel gene networks and cellular and molecular functions shaped by adaptive evolution to hypoxic stress. To our knowledge this is among the first times such an approach has been used to compare molecular phenotypes of closely related populations in order to investigate human adaptations to local environments.

3.3 Methods

3.3.1 Sample Acquisition

The Tibetan samples were derived from peripheral whole blood collected with IRB consent (IRB16-1501) from Tibetans living in the Chicagoland area. All subjects were either born in Tibet or were the descendants of four Tibetan-born grandparents. Blood was collected onsite at the Tibetan Alliance of Chicago in Evanston, Illinois. The CHB samples were purchased from the Coriell Institute and were originally collected and genotyped by the HapMap Project, sequenced by the 1000 Genomes project, and contributed with consent to the NHGRI Sample Repository for Human Genetic Research[186], [187].

3.3.2 LCL Generation and Genetic Ancestry analysis

For the Tibetan lines, for which no publicly available LCLs exist, we collected whole blood, and generated LCLs in house using the same standard protocol used by the Coriell Institute for establishment of lymphoblastoid cell cultures from peripheral blood mononuclear cells (PBMCs)[187]. Briefly, approximately 9ml of blood from each Tibetan donor was collected in ACD (Solution A) vacutainer tubes, placed in Styrofoam boxes to minimize temperature change, and transported from the Tibetan Alliance of Chicago to the Di Rienzo lab at UChicago

within 3 hours. Upon arrival, PBMCs were immediately isolated and infected with Epstein Barr Virus (EBV) in RPMI 1640 culture medium containing 20% FBS and 10ul/ml phytohemagglutinin (Gibco, cat: 10576015). Cultures were established over the course of 40 days and ultimately cryopreserved at a density of 1 x 10^6 cells/ml. The newly generated lines were tested for freeze-thaw survival as well as mycoplasma contamination and were found to be stable, expandable, and uncontaminated.

A sample of LCLs were taken for DNA extraction and ancestry analysis. DNA was extracted using the QIAamp DNA mini Kit (Qiagen, cat: 51304). The DNA samples were given to the University of Chicago Genomics Core for microarray genotyping. Due to variable timing of collection and DNA extraction, the samples were genotyped across multiple platforms. Eight of the samples (TBC86, TBC252, TBC462, TBC540, TBC750, TBC801, TBC841, and TBC988) were genotyped using the Infinium Multi-Ethnic Global-8 v1.0 chip (Illumina, WG-316-1001), TBC859 was genotyped on the Infinium Multi-Ethnic AMR/AFR-8 v1.0 chip (Illumina, 20001090), and TBC569 was genotyped on the Infinium OmniExpress-24 v1.3 chip (Illumina, 20024631). TBC samples were imputed using IMPUTE2, keeping genotypes with posterior probability ≥ 0.9. Phased genotype calls from whole genome sequencing of 59 high altitude Tibetan and Sherpa individuals (reported in [11]) and phase 3 1000 Genomes Project data were used as imputation reference. The indigenous Tibetan samples (TIB) used for these analyses were previously genotyped and imputed as described in ref [11]. Two additional East Asian populations were used for comparison: JTP (Japanese in Tokyo, Japan), and KHV (Kinh in Ho Chi Minh City, Vietnam), and one South Asian population was used as an outgroup: GIH (Gujarati Indian from Houston, Texas). Individuals from these populations were randomly sampled from the 1000 Genomes database [187]. Overlapping autosomal biallelic SNPs among

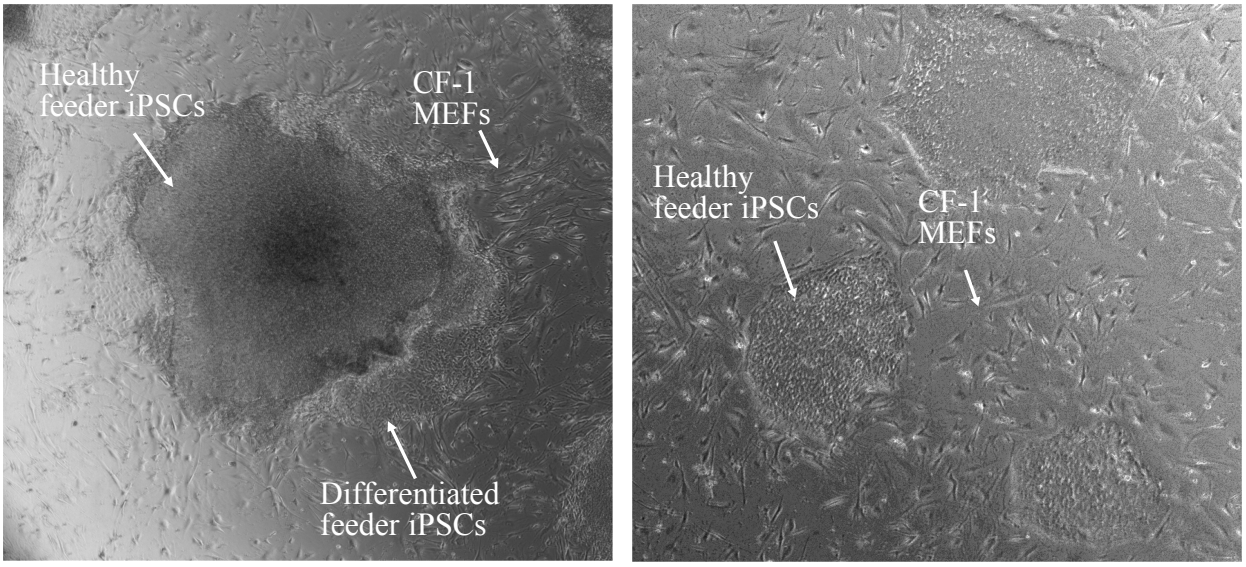
imputed TBC, imputed TIB, and 1000 Genomes samples were merged. All overlapping 400,947 SNPs with MAF>=0.05 were used to generate PCA plots with EIGENSOFT /6.1.4 smartpca [188]. For ADMIXTURE analysis [189] the 400,947 SNPs were pruned for linkage disequilibrium using plink/1.90 leaving 34,112 SNPs [190]. These were run in admixture/1.3.0 using K=2 through K=10

3.3.3 iPSC Generation

Reprogramming of iPSCs was performed in accordance with Thomas et al., 2015. Briefly, LCLs were nucleofected with episomal plasmids encoding OCT3/4, shP53, Lin28, SOX2, L-MYC, KLF4, and GFP (Addgene plasmids 27077, 27078, 27080, 27082) using the Amaxa transfection program X-005. Following transfection, cells were grown in suspension for a week before being transferred onto gelatin-coated plates containing CF-1 irradiated mouse embryonic fibroblasts (Gibco, A34180). Throughout this process cells were grown in hESC media (DMEM/F12 supplemented with 20% KOSR, 0.1mM NEAA, 2mM GlutaMAX, 1% Pen/Strep, 0.1 mM BME, and 12.5 ng/mL human bFGF) which was additionally supplemented with 0.5mM sodium butyrate on days 2-12. For each line at approximate day 24, three individual clonal colonies were chosen for expansion and continued passaging and grown separately. The colonies were visually assessed for growth rate and rates of differentiation (example shown in fig 3.1). The clone with the most consistent growth rate and lowest differentiation rate was ultimately chosen to carry forward into feeder-free conversion for each line.

Figure 3.1: Representative example of feeder iPSCs grown on CF-1 MEFs. A) CHB557 clone C, a clone that was ultimately not taken through to passage 10 due to its high propensity for differentiation as is shown here. B) CHB557 clone B, the clone that was ultimately taken through to feeder-free conversion. Visual inspection shows minimal differentiation and the typical flat, tightly-packed texture of iPSCs.

A. **B.**



After a minimum of 10 passages on CF-1 MEFs, iPSCs were converted to feeder-free maintenance conditions on Matrigel (0.1mg/ml) coated plates in MteSR1 (Stemcell Technologies, 85850) or Essential 8 media (Gibco, A1517001) Lines were fed fresh media daily and routinely passaged using Cell Release, an EDTA-based solution described in Chen et al., 2011[93], and re-plated in media supplemented with Rho-associated kinase (ROCK) inhibitor Y27632 (BioVision, 1994-1). Following 10 passages in feeder-free conditions cells were frozen and tested for freeze-thaw recovery. All generation and maintenance of iPSCs were performed in atmospheric oxygen concentration. Cells were grown in culture at at 37°C, 5% CO2 in a standard incubator. Following reprogramming, lines were assessed for ejection of episomal plasmids,

production of endogenous pluripotency factors, and ability to spontaneously differentiate into endoderm, mesoderm, and ectoderm as described below.

3.3.4 RT-qPCR to characterize expression of endogenous and exogenous pluripotency factors

Cell pellets were collected and flash frozen from each line as LCLs at day 0 prior to nucleofection, at day 8 prior to transition to CF-1 plates, at passage ten as feeder-dependent iPSCs on CF-1 plates, and at passage ten as feeder-free iPSCs on Matrigel. RNA was extracted using RNeasy Plus Mini kit (Qiagen, 74134) in a Qiagen Qiacube. cDNA was generated using the iScript cDNA synthesis kit (biorad, 1708891) and RT-qPCR performed on a QuantStudio 6 Flex machine using qPCRBIO SyGreen Blue Mix Lo-ROX (PCR Biosystems, PB20.11-05). RT-qPCR was used to identify the production of endogenous pluripotency factors (OCT3/4, SOX2, KLF4, L-MYC, NANOG, and Lin28) and the absence of plasmid-derived transcripts (Addgene plasmids 27077, 27078, 27080, 27082). Plasmid-derived transcripts can be uniquely identified from endogenous pluripotency factors based both upon their unique sequence and the EBNA-1 gene encoded in the vector backbone. EBNA-1 is also encoded by the EBV used to immortalize the LCLs. For plasmid-derived factors, expression was measured relative to a random day 8 line and for EBNA-1 expression, which measures both plasmid backbone and EBV presence, expression was measured relative to a randomly selected LCL line. For endogenous pluripotency factors, expression was measured relative to either an LCL line or a day 8 line. All expression was normalized relative to RPLP0 [125] . Primers used can be found in table 3.1. Most lines had fully ejected the plasmids by passage 10 on feeder-free, those that still showed a slight signal from the plasmid factor RT-PCR were re-tested at passage 10 post-conversion. At this point, all lines had ejected all episomal plasmids. All 20 lines were found to express similar levels of

endogenous pluripotency factor. We observed no consistent variations in pluripotency factor expression between populations.

Table 3.1: Primers used for quality control of iPSC lines. Coding sequence (CDS) primers amplify endogenous factors, while plasmid (pla) primers amplify exogenous pluripotency factors derived from the episomal plasmids nucleofected into LCLs at day 0.

Gene	Left Primer	Right Primer	Description	ID
RPLP0	CCATTCTATCATCAACGGGTACAA	TCAGCAAGTGGGAAGGTGTAATC	Normalizing gene	From [125]
OCT3/4 (CDS)	CCCCAGGGCCCCATTTTGGTACC	ACCTCAGTTTGAATGCATGGGAGAGC	Endogenous qPCR	PMID: 21460823
SOX2 (CDS)	TTCACATGTCCCAGCACTACCAGA	TCACATGTGTGAGAGGGGCAGTGTGC	Endogenous qPCR	PMID: 21460823
KLF4 (CDS)	ACCCATCCTTCCTGCCCGATCAGA	TTGGTAATGGAGCGGCGGGACTTG	Endogenous qPCR	PMID: 21460823
L-MYC (CDS)	GCGAACCCAAGACCCAGGCCTGCTCC	CAGGGGGTCTGCTCGCACCGTGATG	Endogenous qPCR	PMID: 21460823
NANOG_CDR	CAGAAGGCCTCAGCACCTAC	ATTGTTCCAGGTCTGGTTGC	Endogenous qPCR	PMID:18029452
LIN28 (CDS)	AGCCATATGGTAGCCTCATGTCCGC	TCAATTCTGTGCCTCCGGGAGCAGGGTAGG	Endogenous qPCR	PMID: 21460823
OCT3/4 (pla)	CATTCAAACTGAGGTAAGGG	TAGCGTAAAAGGAGCAACATAG	Exogenous qPCR	PMID: 21460823
SOX2 (pla)	TTCACATGTCCCAGCACTACCAGA	TTTGTTTGACAGGAGCGACAAT	Exogenous qPCR	PMID: 21460823
KLF4 (pla)	CCACCTCGCCTTACACATGAAGA	TAGCGTAAAAGGAGCAACATAG	Exogenous qPCR	PMID: 21460823
L-MYC (pla)	GGCTGAGAAGAGGATGGCTAC	TTTGTTTGACAGGAGCGACAAT	Exogenous qPCR	PMID: 21460823
EBNA-1	ATCAGGGCCAAGACATAGAGATG	GCCAATGCAACTTGGACGTT	Exogenous qPCR	PMID: 21460823

3.3.5 Embryoid body assays

Embryoid body assays were carried out in accordance with Romera et al.[191]. Briefly, feeder-dependent iPSCs grown in bFGF-containing hESC media were dethatched and manually transferred to polyhema-coated plates where they were grown in suspension for 7 days in bFGF-free hESC media. They were then moved to gelatin-coated plates and allowed to attach and grow in EB media (DMEM + 10% FBS) for an additional 2 weeks. The embryoid bodies were then fixed in PBS containing 4% paraformaldehyde and immunofluorescently stained to detect endoderm (alpha-Fetoprotein, 1:100, SC-130302, Santa Cruz Biotech; and HNF3β, 1:50 sc-374376, Santa Cruz Biotech), mesoderm (α-smooth muscle actin, 1:1500, CBL171, Millipore), and ectoderm (nestin, 1:250, SC-71665, Santa Cruz Biotech) lineages.

3.3.6 RNA-seq of iPSCs

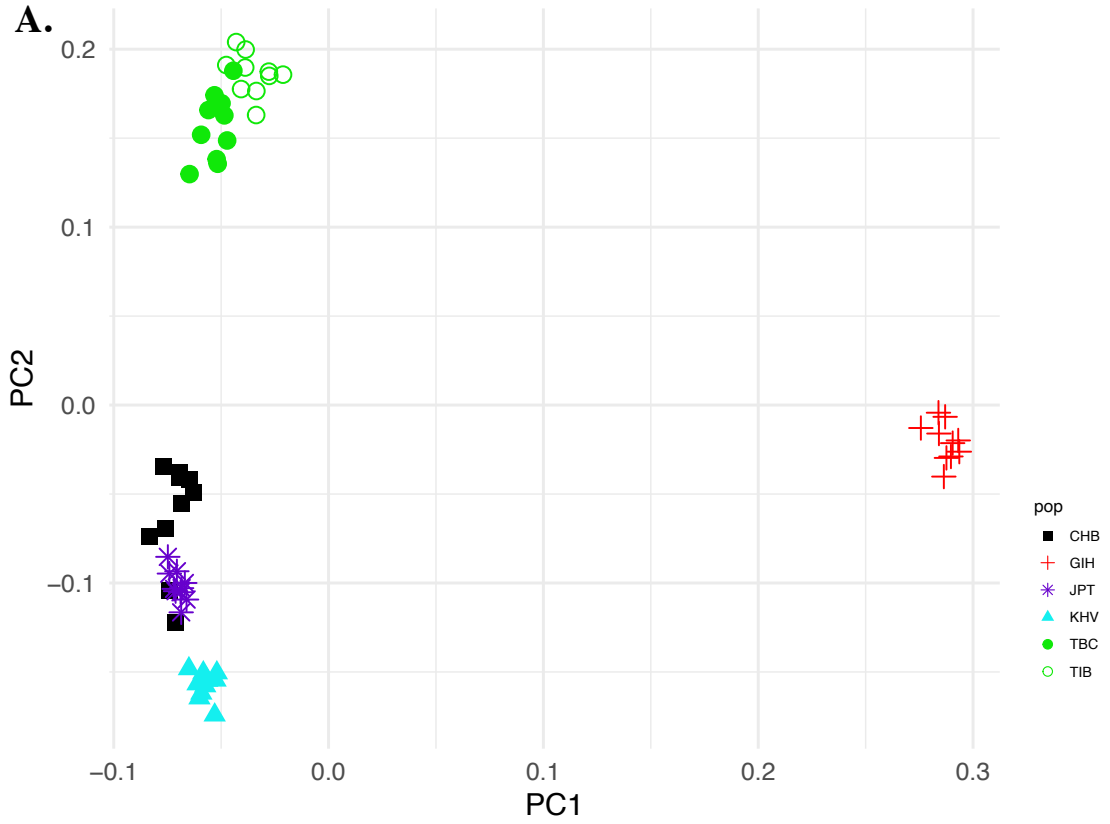
For RNA-sequencing, iPSCs at P10+10 were thawed in two batches of 10 each, balanced by sex and population, grown to confluence, expanded to p10+11 and harvested at 95% confluence. The RNA was isolated using a Qiagen RNeasy Plus Mini kit in using a Qiacube in balanced batches. Library prep and sequencing were performed at the University of Chicago Genomics core, and run together on the Illumina NovaSEQ (PE 100bp) with an average of 34M reads/sample. Analysis was performed using a standard RNA-seq pipeline. Briefly, reads were aligned using STAR and read counts performed with RSEM using standard settings [192], [193]. Differential expression analyses performed using limma-voom with standard settings [128], [129]. Data were modeled as $\mathrm{E}\left(\log_2(y_i)\right) = \beta_0 + \beta_{\mathrm{pop}} Pop_i + Sex + Conv.batch$ where Conv.batch is the feeder to feeder-free conversion batch for the iPSCs.

3.4 Results

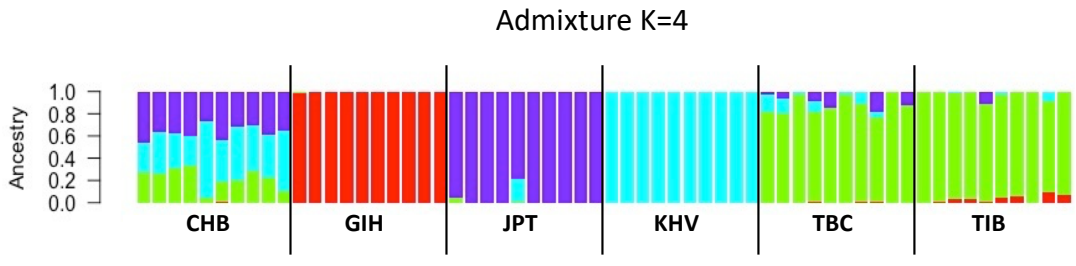
3.4.1 Population Genetic Analyses of Tibetan and Chinese samples

While CHB LCL lines were generated, sequenced, and fully characterized by the 1000 Genomes Project, no such lines exist for the Tibetan population. We therefore generated our own panel of Tibetan LCLs prior to iPSC generation. To do this, we established a relationship with a local population of Tibetans in the Chicagoland area and were able to collect blood onsite for transport and immediate use in our lab space at the University of Chicago. The ten Tibetan samples described here comprise 5 male and 5 female donors aged between 44 and 71. Self-reported ancestry and familial data were obtained via interview at the time of blood collection. DNA was isolated from all Tibetan LCL lines and genotyped in order to computationally assess genetic ancestry and relatedness among individuals using KING [194]. The analyses revealed that none of our ten individuals are closely related (kinship≤0.011). Principal Component (PC) analysis showed that our TBC Tibetans clustered tightly with indigenous Tibetan samples collected from the Mustang and Ghorka districts of Nepal in 2012[11]. PC1 distinguished east from south Asian populations, while PC2 clearly separated Tibetans (both TBC and TIB) from other east Asian populations, including our CHB samples. ADMIXTURE analysis also gave clear evidence of Tibetan ancestry in all 10 lines, which were indistinguishable from TIB samples (fig. 3.2).

Figure 3.2: Genetic ancestry analysis of TBC-CHB panel. A) PCA plots generated from 400,947 overlapping, autosomal, bi-allelic SNPs in the panel. The Tibetan samples sourced in Chicago and included in the panel (TBC) are shown in filled green circles and group tightly with Tibetan samples sourced in on the Tibetan plateau (TIB) and are easily distinguished from our CHB samples. Also included in the plots are two additional East Asian populations: JTP (Japanese in Tokyo, Japan), and KHV (Kinh in Ho Chi Minh City, Vietnam), and one South Asian outgroup: GIH (Gujarati Indian from Houston, Texas). B) ADMIXTURE results from 34,112 LD pruned SNPs for the same 6 populations at K=4.



B.



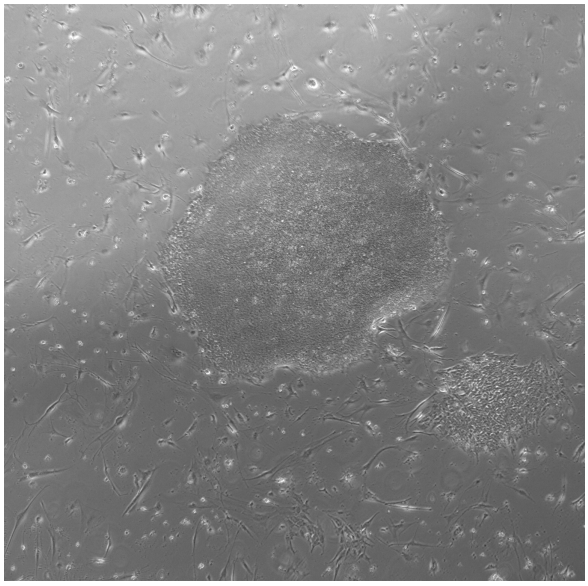
Chinese lines were chosen at random from the 1000 Genomes Project LCL bank with 5 male and 5 female samples to match the Tibetan lines. Although age information was not available for these samples, they are known to be unrelated individuals

3.4.2 Generating a 20 line Tibetan and Han Chinese iPSC panel

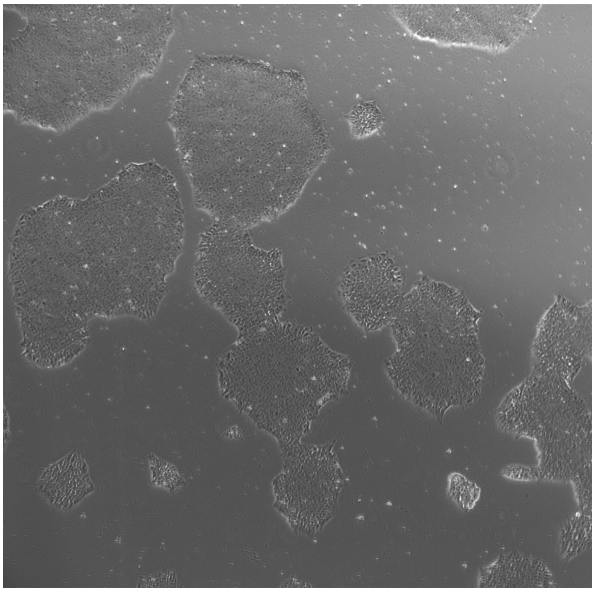
Once the Tibetan (TBC) LCL lines were successful established, they were thawed in parallel with purchased CHB LCLs (Coriell, 1000 Genomes Project) and expanded, then refrozen at high density for rapid use (10^7 cells/ml). LCL lines were reprogrammed into iPSCs in batches of 4 balanced by sex and with reprogramming start days spaced one week apart as described in Methods. In the first phase of reprogramming on CF-1 irradiated MEFs, two of the twenty lines (TBC 86B and TBC 859C) initially failed to produce colonies. These lines were re-transfected, and transferred to MEF plates at a higher density to boost cell survival. This adjustment allowed colonies to be successfully generated for both lines. For the first 3-5 passages on MEFs, multiple clonal isolates for each line were grown (three having been selected to proceed to passage 1 labeled A-C). By passage 5 one clone per line was chosen to continue forward while the other two were frozen down and cryopreserved as backups. The chosen clonal isolate was selected qualitatively based on steady growth pace, lack of spontaneous differentiation, and ease of expansion.

Figure 3.3: Representative examples of A) Feeder and B) feeder-free iPSCs in the panel. Note the tightly-packed round clusters of cells and lack of differentiated cells at the edges.

A.



B.



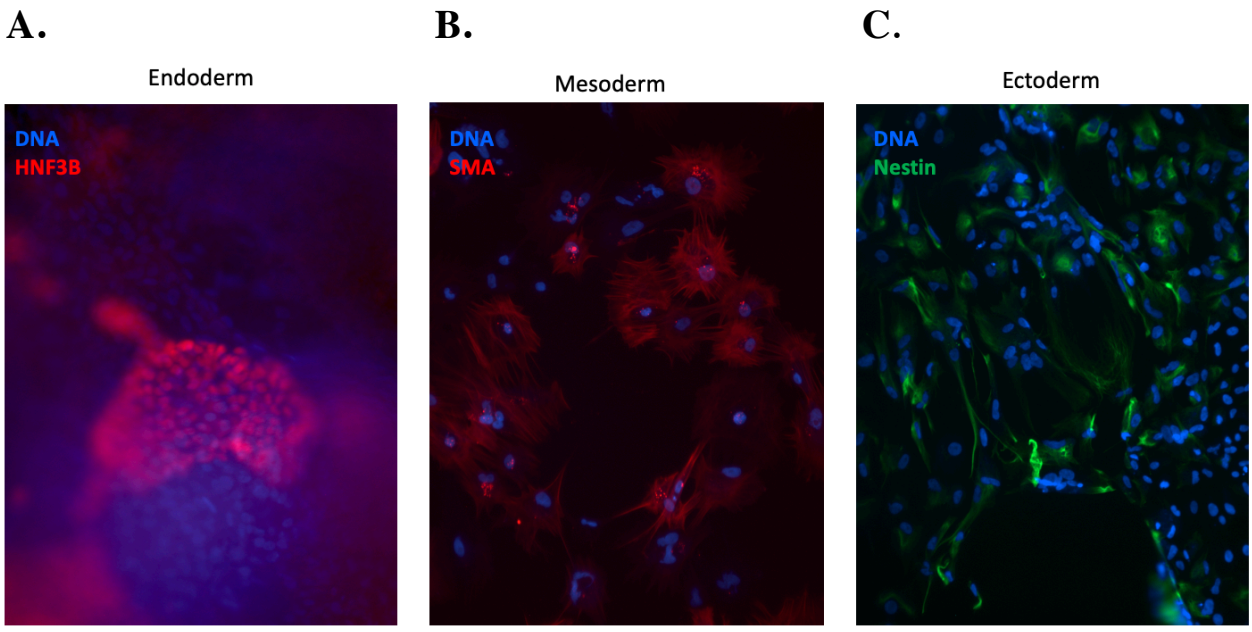
Upon reaching passage 10 on CF-1 MEFs, all lines were cryopreserved prior to conversion to feeder-free growth. Lines were thawed and re-plated on CF-1 MEFs in batches of four balanced by sex and population and converted in parallel into feeder-free iPSC lines. Each successive batch was started approximately 1 week apart and aliquots of cells for each line were cryopreserved every two passages. Most lines were converted directly from passage 10 by manually passaging portions of feeder cells into Matrigel lined wells. For one line (TBC 462C), conversion at passage 10 was not possible as the cells were too unstable and spontaneously differentiated. For this line, conversion was successful at passage 12. In early passages of conversion, cells were cultured in mTeSR1 media on 12 well plates prevented from differentiating by manually scraping off areas of visible differentiation under a dissection scope while leaving undifferentiated cells intact. By passage 5 lines were expanded to 6 well plates and

cultured from here on out in Essential 8 medium. Cells were fed daily and cleaned if areas of differentiation appeared. For most lines, spontaneous differentiation ceased at or before passage 8. At this point cells were expanded into larger dishes before finally being cryopreserved at passage 10. At this time, cell pellets were collected for RNA extraction. All lines were found to survive re-plating following cryopreservation.

3.4.3 Embryoid body assays demonstrate functional pluripotency in panel

To functionally assess the pluripotency potential of the newly generate iPSC panel, embryoid body tests were performed on all iPSC lines at passage 10 on CF-1 MEFs following one freeze-thaw cycle. Following generation of adherent embryoid bodies in minimal medium (details on the embryoid body protocol can be found in 'Methods') cells were fixed with 4% PFA and stained. For each line we plated 4 wells of adherent EBs and stained each with separate IF antibodies against proteins marking the three germ layers: endoderm: HNF3β and α-fetoprotein (AFP), mesoderm: smooth muscle actin (SMA), and ectoderm: nestin. The samples were then counterstained with various Alexa fluorophores and imaged on an EVOS fl Digital Inverted Microscope (Advanced Microscopy Group) while still *in situ* on their cell-culture plates. While the relative abundance of each lineage type differed both between lines and amongst individual EBs within lines, all lines in the panel were found to be capable of spontaneously producing cells which stained positive for the above proteins in the correct spatial pattern and location. Representative examples from TCB801A are shown in figure 3.4.

Figure 3.4: Representative image of IF staining of iPSC (line TBC801A). A) Endoderm stained with nuclear HNF3β (1:50 sc-374376, Santa Cruz Biotech) B) Mesoderm stained with cytoskeletal α-smooth muscle actin (1:1500, CBL171, Millipore). C) ectoderm stained with cytoplasmic Nestin (1:250, SC-71665, Santa Cruz Biotech).

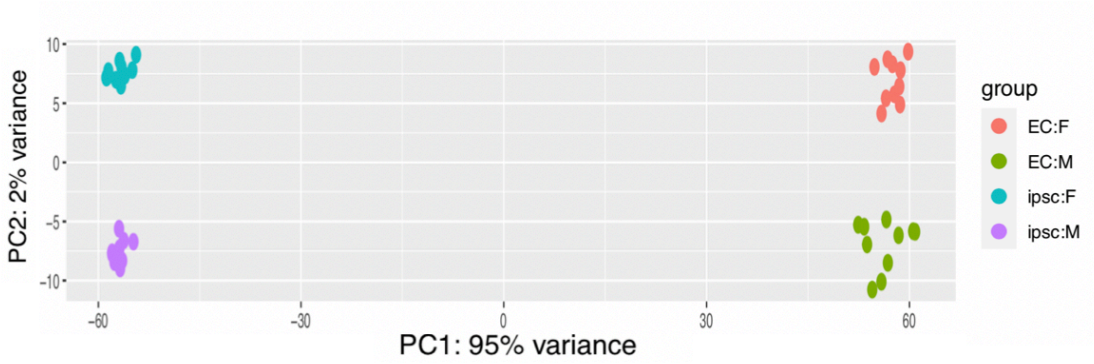


3.4.4 Analysis of gene expression in iPSC panel

Because our panel is intended to examine interpopulation differences in gene expression in differentiated cell types, it was necessary to first assess any interpopulation differences in expression in the iPSCs. Any interpopulation differences seen in the iPSCs could potentially confound differential expression analyses in differentiated cell types. Additionally, RNA-seq in the iPSC panel would allow us to more precisely examine expression of pluripotency factors. Interpopulation differences in pluripotency factors would be cause for concern as they may affect differentiation potential of the iPSC lines or alter cell-type composition of differentiated cell types.

We first assessed how iPSCs clustered when compared to their differentiation products (the differentiation and results are described in Chapter 4). We found that compared to iPSC-derived vascular endothelium, iPSCs clustered tightly together, and separated from the iPSC-derived vascular endothelial cells (ECs) along PC1 which explained 95% of the variance (fig. 3.5).

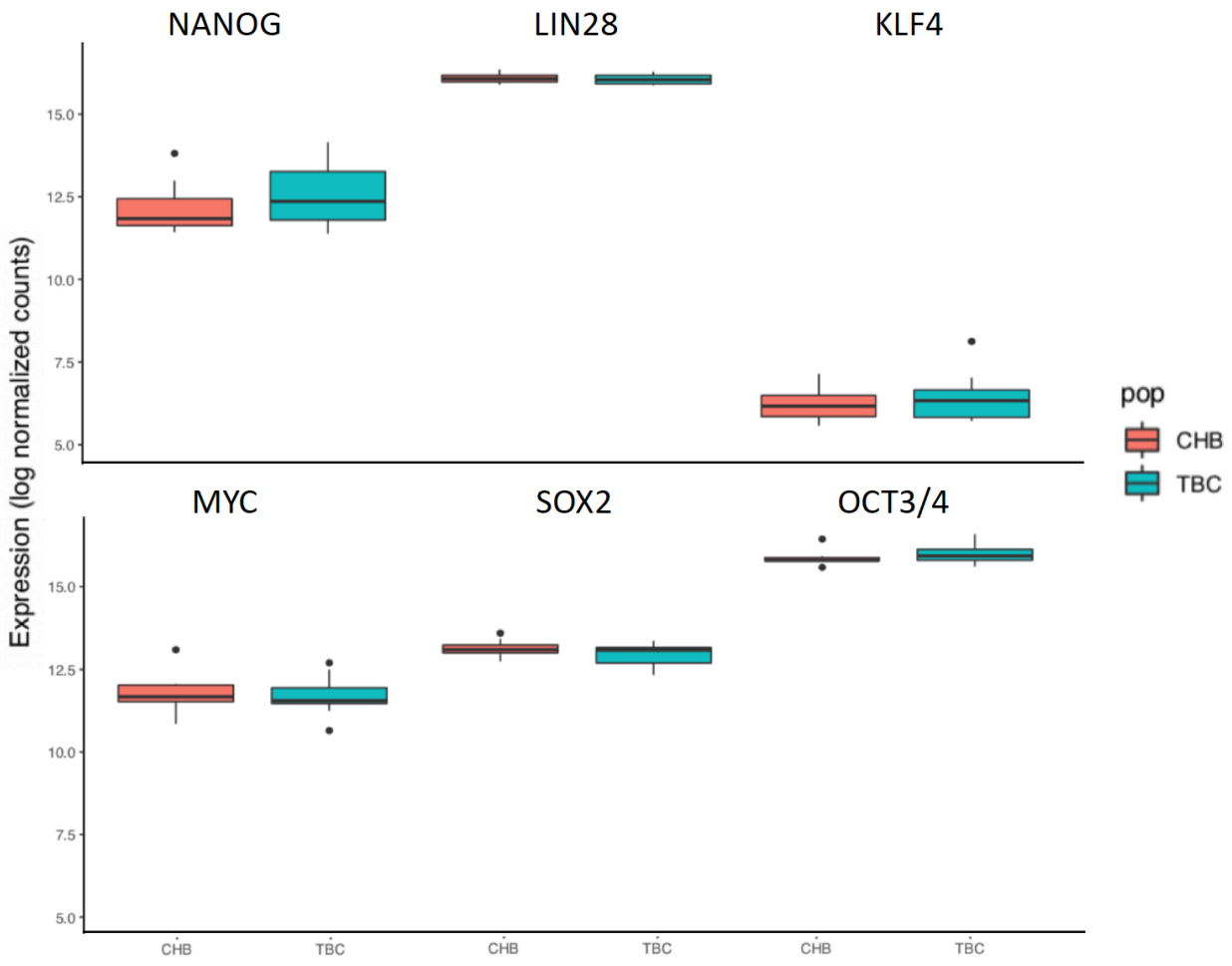
Figure 3.5: Principal components analysis shows the clustering of iPSCs (blue – female and purple – male) and differentiation products, iPSC-derived vascular endothelium (red – female and green – male). PC1 separates the cell types and explains 95% of the variance, which PC2 captures sex, and explains 2% of the variance.



PC2 capture sex differences and explained 2% of the variance. Notably, the iPSCs clustered more tightly together than did the differentiated cells, indicating the uniformity of the panel. Next, we examined differential expression between populations. In particular, we looked at expression of endogenous pluripotency factors *NANOG*, *LIN28*, *KLF4*, *MYC*, *SOX2*, and *OCT3/4* (*POU5F1*). While these factors were examined via qPCR before conversion from feeder to feeder-free iPSCs, we wanted to more directly compare their expression in the complete feeder-free panel. Population-specific differences in pluripotency factors might be expected to influence differentiation potential or even cell-type composition in downstream experiments. However, we

find remarkable consistency in expression levels of all six pluripotency factors, and no significant population differences were identified (p.adj > 0.998) (fig. 3.6).

Figure 3.6: Log-normalized expression counts by population for six pluripotency factors in the TBC-CHB iPSC panel. No significant differential expression was detected between populations for any pluripotency genes. *NANOG* (P = 0.33 P.adj = 0.9998), *LIN28A* (P = 0.35, P.adj = 0.9998), *KLF4* (P = 0.32, P.adj = 0.998), *MYC* (P = 0.97, P.adj = 0.998), *SOX2* (P = .038, P.adj = 0.9998), *OCT3/4* (P = 0.93, P.adj = 0.9998).



Indeed, we see no significant population differences in the expression of any gene in our TBC-CHB panel, and a uniform distribution of P-values. These results indicate not only similar pluripotency among iPSC lines and populations, but also no population-specific differences in gene expression that might be expected to persist in differentiated cell types. Therefore, DE genes between populations in the differentiated cell types are likely to derive from cell-type specific differences in gene expression among populations.

3.5 Discussion

In addition to proving a unique opportunity to study human disease states and functional genetics[95], [98], [100], [195], iPSCs represent a novel resource for examining the impact of local adaptation in relevant cell types. We have therefore generated a panel of 10 Tibetan and 10 Han Chinese iPSCs, the first of its kind, in order to assess the functional impact of Tibetan adaptation to high-altitude hypoxia. These lines have all been fully characterized and assessed for quality control metrics. All twenty lines have ejected exogenous reprogramming vectors, express endogenous pluripotency factors at a similar level, and are capable of spontaneously generating endoderm, mesoderm, and ectoderm. We have further characterized the lines for interpopulation differences in transcript levels and found none, indicating the uniformity of this iPSC panel. Based on these results, we expect this panel to be well-behaved and to differentiate without confounders between populations. Importantly, because we see no differentially expressed genes on the basis of population in this iPSC panel, any differentially expressed genes identified in differentiated cells can be interpreted as cell-type specific interpopulation differences.

In recent years, large population cohorts of iPSCs from healthy individuals have begun to be developed. These cohorts have largely been use to functionally validate GWAS hits and identify tissue specific eQTLs [196], [197]. Contemporaneously, panels of iPSCs from non-human primates have been developed for the study of human evolution and regulatory adaptation [191], [198]. The panel of iPSCs described here will enable yet another novel use of these of iPSCs for population-based functional genetic studies. By creating an iPSC panel from a high-altitude population, we will be able to ask nuanced questions about the effects of local adaptation in a variety of cell types under hypoxic conditions.

The panel described here offers exciting avenues of investigation into the functional molecular impacts of local adaptation. The ability to make precise comparisons in a cell type of choice in well-balanced, controlled conditions would previously have been impossible for many cell types and populations due to the logistical, technical, and even ethical challenges to acquiring and culturing the requisite biological material. iPSC-differentiated cells, on the other hand, can be generated at will and subjected to treatments or conditions of interest in order to identify differential genes or gene expression modules between populations in a biologically relevant condition, or to ask more complex questions such as differential responses to treatment between populations. The genetic variability inherent to a large multi-population panel such as ours also provides the opportunity to map regulatory variants by expression quantitative trait loci mapping and allele specific expression[199]–[201].

In addition to expression phenotypes, iPSC-derived cells provide the opportunity to explore both molecular and physiological difference between populations or disease cohorts by examining cellular phenotypes. For example, iPSC-derived neurons from schizophrenia patients have been found to recapitulate reduced neural connectivity in culture [202], while iPSC-derived

endothelial cells can form vascular structures in endothelial tube forming assays [203]. Metabolism, protein levels and localization, electrophysiological traits and more have all been studied using iPSC-derived terminally differentiated cell types [100], [195], [203], [204]. While these explorations are often performed in the context of healthy and diseased states, it is exciting to imagine how they may be applied to population-based iPSC panels such as ours, to detect and infer physiological impacts of adaptive evolution.

Finally, in addition to their myriad uses in investigating interpopulation differences in expression and cellular phenotypes, iPSCs are very permissive to genome editing using CRISPR/Cas9 [185]. Primary cells often do not survive long in culture, have low transfection efficiency, and may be poorly suited to the single-cell clonal isolation necessary to derive pure CRISPR-edited lines [205], [206]. Editing iPSCs, therefore, offers an appealing alternative to removing, inserting, or making single base pair alterations in an isogenic background and a cell type of interest in a reproducible manner. Our panel, for instance, offers the tantalizing possibility of making alterations to one of the two main loci implicated in Tibetan adaptation, *EPAS1* and *EGLN1* which encode HIF-2α and PHD2 respectively[11], [80]. These two proteins both fall into the hypoxia inducible factors pathway. Indeed, PHD2 is epistatic to HIF-2α and regulates it post-translationally based on oxygen availability[41]. Without access to primary biological samples of varying genotypes, it would be incredibly difficult to disentangle how these two loci interact and contribute to phenotype. Previous studies, for example, have included simple knockdowns of *EPAS1* in a humanized *EGLN1* mouse model[207, p. 2]. While informative, this cannot answer the question of how high-altitude variants at these loci interact in different cell types. Using CRISPR in our panel, however, alterations could be made to either of

these genes in an isogenic background in iPSCs, then differentiated into a cell type of interest and compared.

3.6 Acknowledgements of work performed

I would like to thank my lab mates for their contributions to this work.

- Population genetic analyses of Tibetan and Han Chinese samples was performed by David Witonsky
- Embryoid body assays were performed with assistance from Xander Van Alstyne and Jennifer Yoo.

CHAPTER 4

DIFFERENTIATION OF A PANEL OF TIBETAN AND HAN CHINESE IPSCS INTO VASCULAR ENDOTHELIUM

4.1 Introduction

All animal life depends absolutely on mitochondrial respiration and therefore access to oxygen in every cell. Hypoxia, therefore, represents a serious threat and must be sensed and reacted to immediately. For this reason all mammals have a highly conserved pathway of genes for sensing and responding to low oxygen, known as the hypoxia inducible factors (HIF) pathway. The HIF pathway is driven by a heterodimeric transcription factor comprised of a HIF-α and HIF-β subunit which activates a variety of response genes across cell types which are broadly involved in angiogenesis, inflammation, metabolism, vasoconstriction/dilation, hematopoiesis, and immune response [39]–[42]. Due to its location lining the body's blood vessels, vascular endothelium is the first point of contact for oxygenated blood, and must act as one of the body's primary mediators of the physiological response to hypoxia and hypoxemia[138], [208]. The endothelium therefore plays a key role in in a multitude of human diseases that are complicated by hypoxia. For example, in asthma, neovascularization and vascular remodeling driven by activation of the HIF pathway in the endothelium worsens prognoses [53], [58]–[60]. The Endothelium and the activity of one HIF-α isoform, HIF-2α, has been causally linked to hypoxia-induced pulmonary hypertension (PH). A large number of recent studies have shown that HIF-2α activity is essential to the onset and pathogenesis of this disease, and that suppression of HIF-2α is protective against PH [40], [141], [142], [171] .

The importance of endothelial response to hypoxia both within the tissue itself and within the broader acclimatization response, is also implicated in human adaptation to high-altitude hypoxia. Tibetans, for example have lived on the plateau for over 10,000 years, adapting to an altitude with just 60% the oxygen concentration available at sea level [12], [41], [109], [110]. As a result of this severe selective pressure, Tibetans have evolved a number of unique phenotypes that distinguish them from closely related acclimatized lowlanders such as the Han Chinese. For example, Tibetans display a lack of hypoxic vasoconstriction at high altitude, and consequently suffer lower rates of chronic mountain sickness (CMS) and pulmonary hypertension (PH) – diseases characterized by excessive hypoxic vasoconstriction and subsequent elevation of pulmonary arterial pressure (PAP) – than do acclimatized Chinese lowlanders living at the same elevation [49], [56], [103], [111]. Additionally, Tibetans have been reported to have higher nitric oxide (NO) concentrations in plasma as well and elevated exhaled NO [209]. Their calf muscle capillary density has also been reported to be high compared to that of other high altitude populations [12]. Finally, Tibetans display a blunted erythropoietic acclimatization response to hypoxia. This is a highly heritable phenotype and has been mapped to an extended haplotype carrying strong signals of positive selection in the endothelial PAS domain 1 (*EPAS1*) gene, which codes for HIF-2α and is highly expressed in the endothelium [11], [41], [74], [77]–[80], [11], [41, p. 2], [77], [78, p. 1], [79], [80]. The 17kb fragment in which the strongest signals of selection and highest GWAS association co-localize is decorated with H3K27Ac marks in endothelial cells, which points once more to the endothelium as a crucial tissue in Tibetan adaptation (A deeper investigation in the *EPAS1* locus in particular can be found in Chapter 2).

Although the endothelium is a prime target of interest in both biomedical inquiries into hypoxia response and in Tibetan genetics, there are many practical difficulties in acquiring

enough of this cell type for a well-controlled population genetics study. Publicly available endothelial lines are typically not available from enough donors to capture population diversity from any population of interest. For this reason, to this point, all studies of Tibetan endothelial cells have been carried out on umbilical vein endothelial cells (HUVECs)[170], [210], a cell type preferred for its relative ease of acquisition and stability in culture[211]. However, this approach requires not only access to Tibetans near a lab in which the cells can be isolated, but also requires the cords to be donated upon giving birth. This alone limits the number of samples that can be collected simultaneously. Fortunately, recent innovations in induced stem cell technology enable the creation of population panels of induced pluripotent stem cells (iPSCs) which can be differentiated into cell types of interest in a controlled environment. The generation of such a panel of Tibetan and Han Chinese iPSCs is described and characterized in Chapter 3. Here we will report the differentiation of this panel into vascular endothelium following the protocol first defined in Patsch et al., 2015 [203].

4.2 Methods

4.2.1 Sample acquisition and Generation of iPSC library

The sample acquisition and generation of our TBC-CHB iPSC panel is described in depth in Chapter 2 – METHODS.

4.2.2 Differentiation of vascular endothelial cells

iPSCs were differentiated into vascular endothelium according to the protocol defined in [203] with small adjustments. A step-by-step protocol from [203]can be found at Nature Protocol Exchange (https://protocolexchange.researchsquare.com/article/nprot-3985/v1). Our panel of TBC-CHB iPSCs were grown to confluence in balanced batches of 4-8 lines (balanced for

population and sex). On day 0 cells were detached using Cell Release, an EDTA-based solution described in [93], counted, and replated at a density of ~50,000 cells/cm^2 in Essential 8 media supplemented with Rho-associated kinase (ROCK) inhibitor Y27632 (BioVision, 1994-1). On day 1, medium was replaced with N2B27 medium [1:1 mixture of DMEM:F12 (Life technologies, 11320-033) and Neurobasal media (Life technologies, 21103049) with glutaMAX (Life technologies, 35050038), β-Mercaptoethanol (0.097%)(Life technologies, 21985023), N2(Life technologies, 17502048), and B27 minus vitamin A (Life technologies, 12587010)] this was supplemented with 8uM CHIR (Tocris, 4953) and 25ng/ml hBMP4 (PeproTech, 120-05ET) and added to the cells at high volume (ie 15 mL in 22.1cm dish). The cells were left undisturbed for 3 days. On day 4, media was exchanged for EC Induction Medium [StemPro-34 SFM medium (Life technologies, 10639-011) supplemented with 200 ng/mL VEGF165 (PeproTech, 100-20) and 2μM forskolin (Abcam, ab120058), and penicillin-streptomycin]. EC Induction Medium was refreshed on day 5, and on day 6 CD144+ cells were isolated via pulldown. ECs were dissociated using Accutase (STEMCELL Technologies, 7920) and pulldown performed with the quadroMACS Separator (Miltenyi Biotec, 130-091-051) using CD144 MicroBeads (Miltenyi Biotec, 130-097-857). Isolated cells were seeded at ~26,000 cellls/cm^2 on plates coated in 2μg/cm^2 fibronectin (Fisher, 356008) in Expansion Medium [StemPro-34 SFM medium (Life technologies, 10639-011) supplemented with 50 ng/mL VEGF165 (PeproTech, 100-20). Cells were allowed to expand for 4 days, with Expansion Media refreshed every 24hrs, then switched in EGM-2 for further expansion and cryopreservation, while the remaining cells were placed into either hypoxia (1% O_2, 5% CO_2 – Coy Labs Hypoxia Glove Box) or normoxia (~20% O_2, 5% CO_2 – lab incubator) for 48 hours.

4.2.3 Immunofluorescent staining of iPSC-derived vascular endothelium

Two iPSC-derived endothelial lines, one Tibetan (TBC801) and one Chinese (CHB608) were selected to assess the presence of additional endothelial markers. Following pulldown and expansion, the lines were fixed in the wells of a 6-well cell culture plate with PBS containing 4% paraformaldehyde. They were stained with antibodies against CD31 (JC/70A) (Invitrogen, MA5-1318) and VWF (Invitrogen, PA5-16634). The samples were then counterstained with various Alexa fluorophores visually assessed on an EVOS FL Digital Inverted Microscope (Advanced Microscopy Group) for presence and localization of the two proteins. In both lines CD31 was visible on the cell surface of all cells visualized, while VWF was seen in the cytoplasm clustered into prototypical storage granules known as Weibel–Palade bodies [212].

4.2.4 iPSC-derived EC harvest and bulk sequencing

iPSC-derived endothelial cells were washed and lysed in-plate using RLT Plus buffer from the Qiagen RNeasy Plus Mini Kit (74134). Cell culture replicates in hypoxia were lysed and harvested in the hypoxia box with no exposure to normoxia prior to lysis. Lysates were flash frozen and stored at -80C degrees until all batches were complete. RNA extraction was performed in 5 balanced batched of 8 (balanced for population, sex, and treatment) in a Qiagen Qiacube (9001292) using the RNeasy Plus Mini Kit. RNA was assessed for concentration prior to library preparation. 500ng of each sample was used for library prep using Illumina TruSeq2 RNA-seq library prep kit and set A indexes (RS-122-2001). All libraries were assessed via bioanalyzer to be of sufficient quality, concentration, and fragment size to proceed with processing. Samples were run across 5 lanes of the Illumina HiSeq4000 at the University of Chicago Genomics Core, with eight balanced samples per lane resulting in an average of 51M single-end reads per sample.

4.2.5 Data Analysis: iPSC-derived endothelial cells

PCA analysis was performed using plotPCA() function from DESEeq2 following variance stabilizing transformation [213]. Analysis was performed using a standard RNA-seq pipeline. Briefly, reads were aligned using STAR and read counts performed with RSEM using standard settings [192], [193]. Gene filtering was p[erformed with edgeR (v 3.32.1) filterByExpr with design = ~pop+condition+sex [214]. Differential expression analysis was performed with dream (variancePartition v. 1.20.0 R package) analysis following normalization using the the voomWithDreamWeights()[129], [215]. The model: $\sim 0 + (1|\mathrm{ID}) + \mathrm{pop} + \mathrm{condition} + \mathrm{pop{:}\,condition} + \mathrm{sex} + \mathrm{differentiation.batch}$ was used, separately setting normoxia and hypoxia as baseline to retrieve pop DE in both conditions. PLIER analysis was performed following minimal filtering (rowSums(counts(dds)) > 1), normalization with DESeq2's estimateSizeFactors() and variance stabilizing transformation (vst)[213]. Pathways included in analysis were downloaded from https://github.com/wgmao/PLIER/tree/master/data and include canonicalPathways and bloodCellMarkersIRISDMAP[216]. Differences between populations were assessed by comparing all CHB and all TBC values for each LV with a Wilcoxon rank-sum test followed by FDR multiple testing correction.

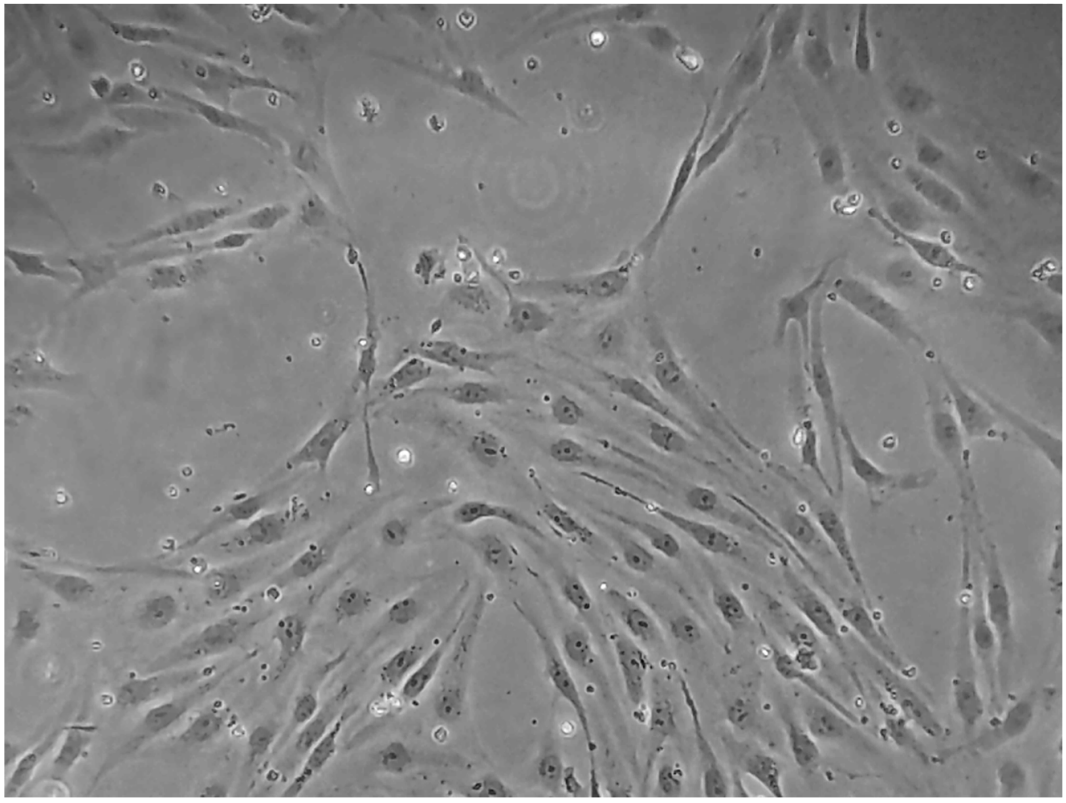
4.3 Results

4.3.1 All 20 lines successfully differentiate and are pulled down with high success rate

We differentiated our panel of 10 Tibetan (TBC) and 10 Han Chinese (CHB) iPSCs – described in Chapter 3 – into vascular endothelium in accordance with the protocol described in [203]. Following the first 5 days of culture in which first mesoderm and then vascular endothelium were induced, CD144+ positive cells were isolated via pulldown. CD144 is also

known as vascular endothelial cadherin (VE-cadherin). It is a cell-cell adhesion glycoprotein that is a highly specific marker for vascular endothelial cells. This pulldown yielded a high number of positive cells for each line. While the proportion of cells pulled down was variable, the median number of positive cells isolated across all lines was 6.05 million. The cells demonstrated typical endothelial morphology (fig. 4.1) and grew well in both normoxia and hypoxia.

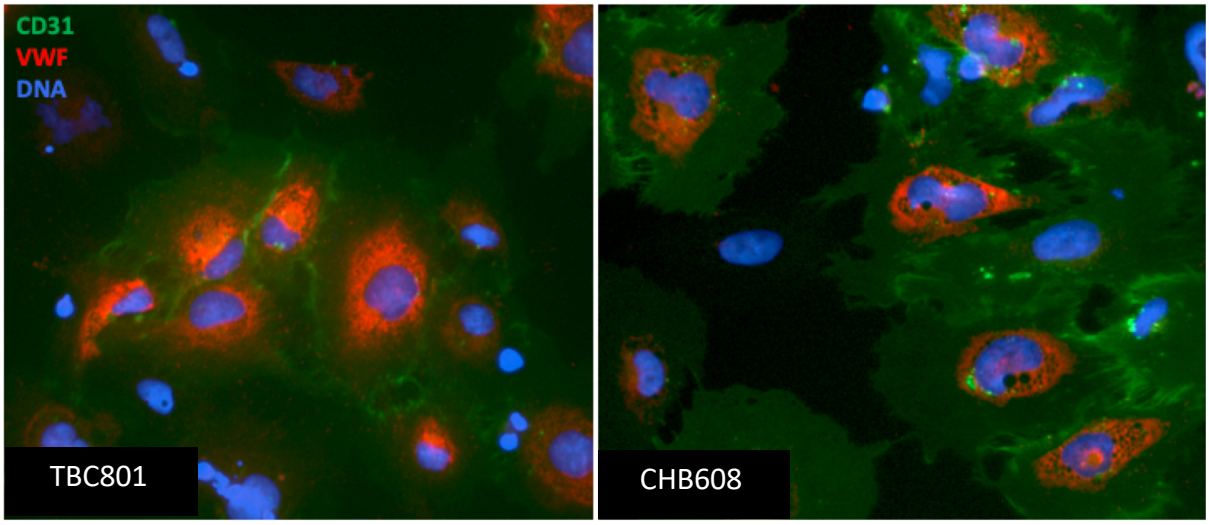
Figure 4.1: Light microscopy image of P0 iPSC-derived vascular endothelial cells, 3 days after pulldown. Cells demonstrate stereotypical endothelial morphology and cobblestone texture where confluent. The black dots seen on the cells are attached CD144 magnetic beads from pulldown.



iPSC-derived ECs were inferred to be a fairly pure population based on the CD144+ MicroBead purification. The resulting population of cells was quantified by FACs in ref [203] to be 99.6% positive for CD144. However, in order to further characterize endothelial markers in

these cells, two lines were selected for additional staining. These lines were stained for two additional endothelial markers and were qualitatively assessed for presence and localization of the factors. Endothelial cells should express Platelet endothelial cell adhesion molecule (PECAM-1, also called CD31) on the cell surface. Von Willebrand factor (VWF), meanwhile is known to cluster into endothelial storage granules in the cytoplasm called Weibel Palade bodies [212]. Our staining of iPSC-derived endothelium on day 10 (4 days after pulldown) is shown in figure 4.2. We see that all cells display stereotypical localization of CD31 and VWF, showing that the differentiated cells not only were marked by CD144, but also carried additional endothelial-specific markers.

Figure 4.2: Immunofluorescent staining of two iPSC-derived endothelial lines (TBC801 and CHB608) for two additional endothelial markers: CD31 – green and VWF - red. DNA stained with Hoechst and marks the nucleus – blue.



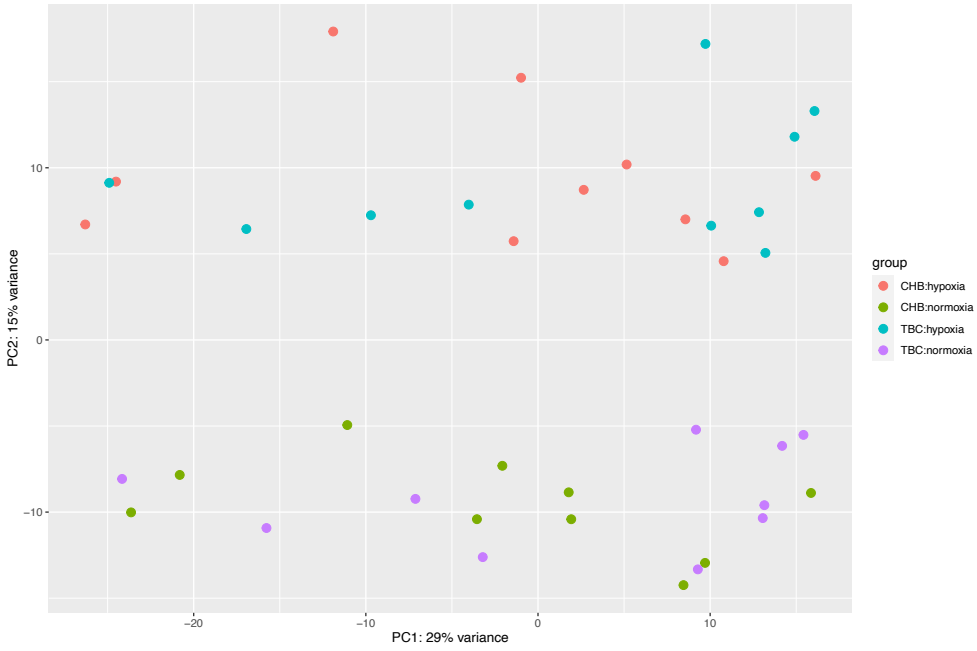
4.3.2 Differential expression (DE) analyses of endothelial cells recapitulate downregulation of *EPAS1* in Tibetans but are highly noisy

We next sought to characterize differential gene expression in vascular iPSC-derived endothelium cultured in normoxia or hypoxia for 48 hours in parallel. This would enable us to

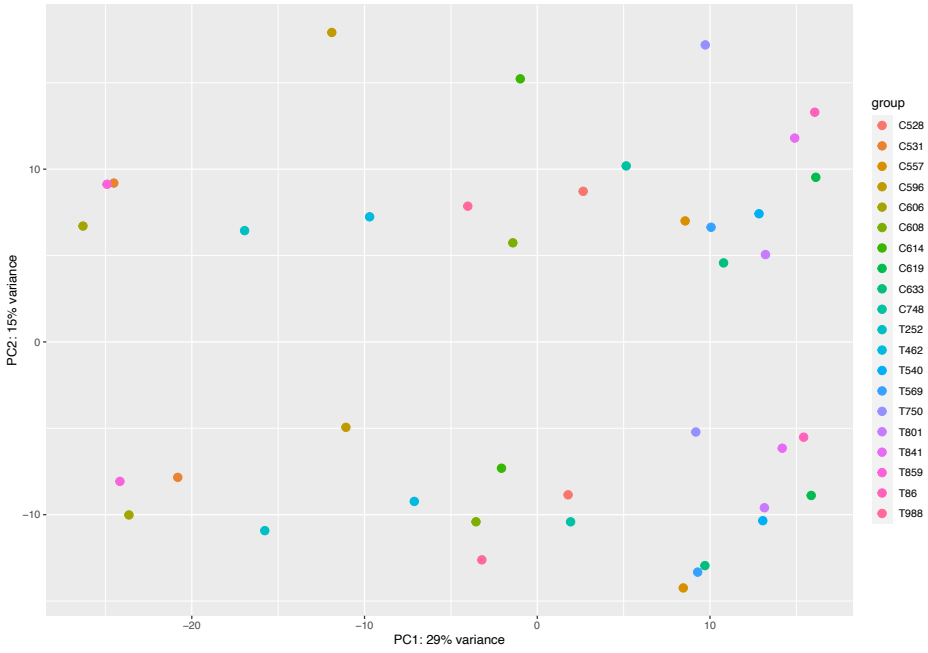
detect differentially expressed genes between populations in both normoxia and hypoxia, and potentially allow for the detection of differential response to hypoxia between populations. PCA analysis indicated a high degree of heterogeneity among lines. PC2 rather than PC1 captured the effect of hypoxia (fig 4.3). This was surprising as hypoxia is expected to have a profound effect on transcription and would generally be expected to contribute to the majority of variance within a single cell type. PC1, in this case, appeared to capture inter-line variability, with individual iPSC lines tending to align along PC1 between normoxia and hypoxia replicates. While not unexpected [96], [217], this result indicates a level of inter-line variation in the data which could overwhelm true biological signal.

Figure 4.3: PCA graphs of iPSC-derived vascular endothelial RNA-seq data. A) PCA colored by population and treatment condition. B) PCA colored by iPSC line. PC1 separates individual lines while PC2 is correlated with treatment.

A.



B.



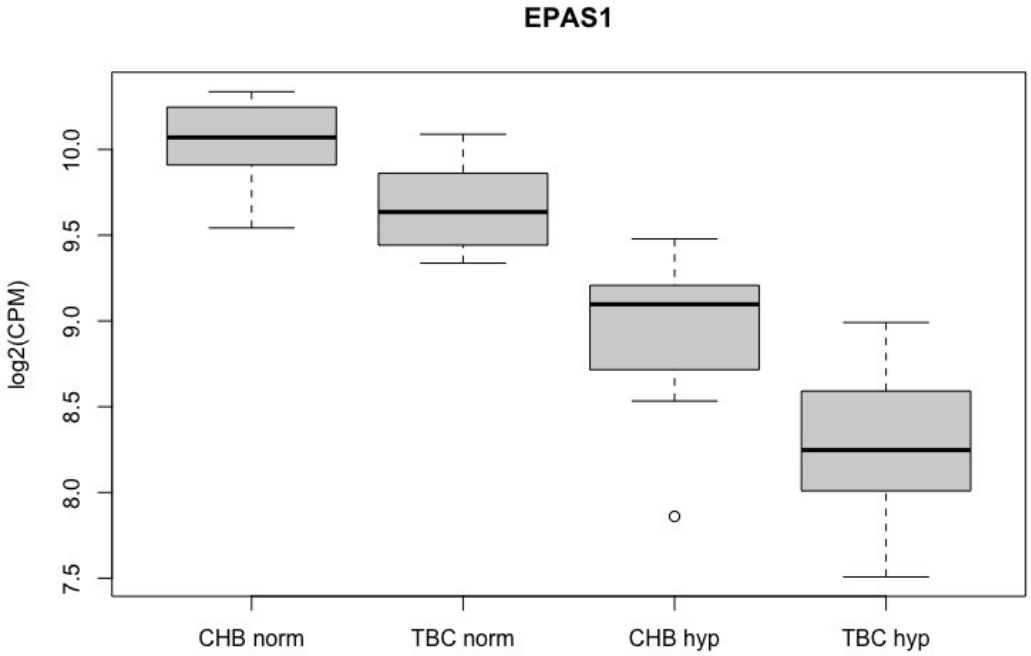
Differential expression analysis was performed using DREAM, which enabled us to model all of our data – normoxia and hypoxia – together in a paired linear model (see Methods) [215], and produced no significant results. Indeed, we see many DE genes at single-gene significance level (but not transcriptome-wide significance), which may further indicate a very low signal to noise ratio. Multiple methods to reduced noise by identifying unmeasured variance and remove batch effects were employed to address this issue including surrogate variable analysis (SVA) [218], removal of unwanted variation (RUVseq-g) [219], and ComBat-seq [220]. All of these methods were individually and collectively capable of shifting the p value distribution and identifying transcriptome-wide significant differentially expressed genes (DEGs). However, upon careful inspection of permutation analyses (in which population labels were permuted while leaving other covariates consistent) we discovered that, in this dataset, these methods will consistently produce an excess of low p-values and identify DEGs in permuted samples similar to the real data, indicating that these methods result in false positives in our data.

We therefore proceeded with a typical linear modeling approach using DREAM including only known covariates [215]. Although this approach yields no genome-wide DEGs, we were interested to examine the single-gene significance of *EPAS1* expression, which may be viewed as a positive control in this cell type. Previous studies have identified *EPAS1* as being downregulated in Tibetan relative to Chinese endothelium [78], and both our work presented in chapter 2, and other studies [170] have indicated the presence of multiple endothelial enhancer of *EPAS1* whose activity is altered by the Tibetan adaptive haplotype. At the single-gene significance level, we find that in both normoxia and hypoxia, *EPAS1* expression is expressed at significantly lower levels in TBC samples relative to CHB samples (normoxia p-value = 3.37E-

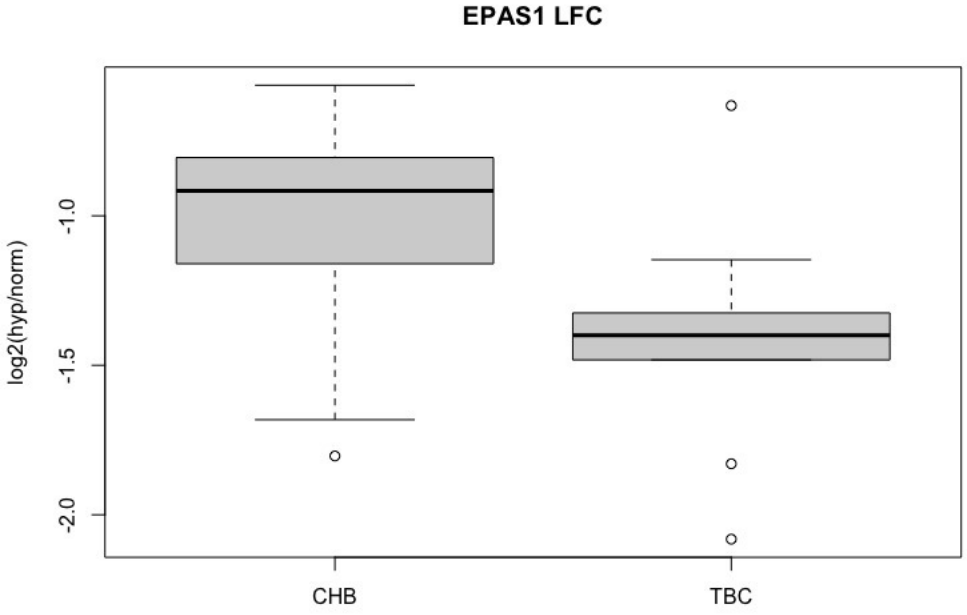
02, hypoxia p-value = 1.19E-04) (fig 4.4). Moreover, we find that the response to hypoxia also differs between populations at borderline significance levels (response p-value = 6.41E-02). In both populations, *EPAS1* expression decreases between normoxia and hypoxia. However, in the TBC lines, the decrease is more profound resulting in a more significant difference between populations in hypoxia . This is particularly interesting in relation to our findings in Chapter 2 of a hypoxia-responsive endothelial enhancer of *EPAS1* that is impacted by Tibetan variation. This enhancer, among other factors, could help explain the differential response seen in Tibetan *EPAS1*.

Figure 4.4: Graphs showing the between population differences in *EPAS1* log-CPM expression in A) normoxia and Hypoxia and C) response.

A.



B.



4.3.3 PLIER analysis of differential gene modules between populations

In order to try to glean more data from the transcriptional data in the iPSC-derived ECs, we used the pathway level information extractor (PLIER) package [216]. This package takes RNA-sequencing data, and cell type marker and pathway data and performs unsupervised clustering to identify latent variables (LVs) in the data, which are similar to eigengenes, and correlate them to pathways. We were interested to see if at a broader pathway level, we could identify LVs that differed between populations. This analysis identified one LV (LV16) that was significantly different between TBC and CHB endothelial cells following multiple testing correction. This LV is associated with both the Reactome pathway "Cytokine Signaling in the Immune System" (R-HSA-1280215) [221] and with the resting neutrophils cell type from the IRIS database [222]. Although our cells were endothelial and not immune in nature, there are many shared pathways that may result in this signal. For example, endothelial cells 'activate' in response to endotoxin challenge and amplify immune responses by activating cytokine signaling inflammatory responses, inducing cytokine production in immune cells, and recruiting neutrophils[208], [223]. Acute hypoxia treatment is also known to activate endothelial cells and induce an inflammatory response[208]. Consistently, A number of known hypoxia response and immune response genes are found in the gene set which defines LV16 (loading > 0), include both *EPAS1, EGLN1,* the genes containing the top two signals of selection in the Tibetan genome, as well as *EPO*, and a number of *HLA* genes [79], [224]. When we run gene ontology analysis on the top 25% of genes with loadings > 0 in LV16, we find a number of significant terms and pathways that may be related to hypoxia response including many related metabolism, apoptosis, response to oxygen levels, TNF signaling, MAPK signaling and others. Additional work is needed to fully unravel and follow-up on these results, but they suggest that, while we are unable

to detect single gene differences in our data, differences in gene modules are detectable between populations.

4.4 Discussion

Although we were unable to conclusively detect DEGs between populations, we did demonstrate the viability of our panel for differentiation into vascular endothelium, opening up doors for future improvements and experiments. Our failure to detect differentially expressed genes with transcriptome-wide significance speaks to the inherent challenges of working with iPSC-derived cell types. iPSC panels are notable for heterogeneity among lines and potential confounders in downstream data [217]. One particularly difficult hurdle to overcome in adapting iPSC models to studies of human adaptation, is that the substantial impact of variability derived from the generation of iPSCs and the differentiation process itself can readily subsume real biological signal, particularly if that signal is subtle and driven by small effect sizes [217], [225], [226]. Additionally inter-donor variability is reportedly much higher in iPSC-derived cells than in their primary counterparts[226] . In the case of our panel these features may be particularly problematic. Inter-line variability is captured by PC1(29% of the variance) while hypoxia effect (arguably one of the strongest drivers of gene expression changes) is captured by PC2 (15% of the variance). As two closely related populations, whose genomes only significantly differ at a few highly differentiated locations (such as the *EPAS1* locus), there might not be an expectation of profound differences in gene expression with large effect sizes, even in primary cell types. There have been few studies that report RNA-sequencing of primary Tibetan and CHB tissue by which to gauge expectations, and none have been reported in vascular endothelium. One report of RNA-seq comparison of 34 Tibetan and 14 CHB placentas found only 276 DEGs, none with

an adjusted p-value below 0.015 . Moreover, unlike our panel, this investigation specifically chose Tibetans with the adaptive haplotype at *EPAS1,* which might be expected to boost differences derived from trans effects of the variation this locus [210].

On the other hand, the fact that despite the multitude difficulties and complications of this method, *EPAS1* consistently rises to the top of any DE genes list generated, is always significantly differentially expressed on the single-gene level, and demonstrates tantalizing evidence of a differential response at the single-gene level is evidence to the strength of that signal and demonstrates the potential viability of this method. There are a number of methods we believe may be leveraged to extract valuable biological information from our panel of iPSC-derived vascular endothelial cells. The first is simply to increase panel size, while this may also increase noise, increasing sample size should improve our power to detect DEGs between populations. Second, we will investigate methods of quantifying and accounting for any cell-type heterogeneity between samples which may be contributing to the inter-individual noise. This may involve additional pulldowns for CD144 or another marker or estimation of heterogeneity using computational methods [227], [228]. We could also investigate phenotypes beyond RNA-seq that may differ in endothelium between Tibetan and Han Chinese. For instance, differences in metabolism could be measured with the Seahorse system [229], or differences in angiogenesis using a tube forming or sprouting assay[230], [231]. Because these phenotypes are driven by large gene modules, it is possible that differences too subtle to be observed in the RNA-seq data may be combined to produce a measurable difference in phenotype. Finally, we may investigate loci of interest (such as *EPAS1*) in isogenic line pairs which have been modified by CRISPR-cas9 [232]. For instance, we may be interested in deleting or modifying one of the enhancers identified in Chapter 2 in different Tibetan or Chinese backgrounds and compare expression and

phenotypic impacts. While this will not give us population-wide information, it would allow us to make comparisons that would otherwise be impossible without our panel, as no publicly available endothelial lines exist for Tibetan populations.

4.5 Acknowledgements of work performed

I would like to thank my lab member, for his contribution to this work.

- DREAM differential expression analysis was performed by David Witonsky.

CHAPTER 5: CONCLUSIONS

Tibetan populations have lived and adapted to the extreme conditions of high-altitude hypoxia for over 10,000 years, surviving with just 60% the oxygen concentration available at sea-level. Understanding local adaptations in human populations is a major motivation of human genetics research. Thus, understanding the genetic mechanisms by which Tibetans have evolved to deal with such a profound local environmental pressure has been the focus of decades of research[12], [66], [75], [80], [108], [112], [233]. These studies have revealed not only a strong signal of positive natural selection at the *EPAS1* locus[10, p. 1], [11], [80], [210, p. 1], but also a bevy of distinct phenotypes that distinguish Tibetans from both closely-related, acclimatized lowlanders and other high-altitude adapted populations. These phenotypes include not only unelevated hemoglobin at high altitude – which has been linked the selected *EPAS1* alleles – but also a number of other complex traits implicating a range of cardiovascular adaptations to hypoxia[11], [12], [41], [66], [68]. Polygenic selection scans performed by the Di Rienzo lab also indicate the presence of selection towards lower pulse and identifies a correlation between lower pulse and improved reproductive outcomes, linking cardiac function with reproductive success in low oxygen environments[11]. These crucial findings motivate further research into the causal genetic variants and functional molecular outcomes of Tibetan adaptation both at the *EPAS1* locus and genome-wide.

5.1 Tibetan genetic adaptations have downregulated *EPAS1* in multiple hypoxia-responsive cell and tissue types

Although decades of research have identified Tibetan *EPAS1* alleles as carrying strong selective signal, and being associated with lower hemoglobin levels at high altitude, the

functional impacts and fitness effects of this locus remain poorly understood [10, p. 1], [11], [80], [210, p. 1]. While the selected alleles are associated with unelevated hemoglobin it is unclear precisely how unelevated hemoglobin would be strongly adaptive. Outside of *EPAS1*, there is no evidence of polygenic selection towards reducing hemoglobin levels in Tibetans, suggesting that this phenotype might be a pleiotropic effect of selection on a different advantageous trait also influenced by *EPAS1* variation, rather than the selective target. Even if unelevated hemoglobin does contribute fitness effects in the form of reducing blood viscosity or preeclampsia at high-altitude [82], [85], [234], this leaves unanswered the question of how Tibetans cope with the resultant hypoxemia without detrimental health effects. The strong LD at this locus limits the resolution of population genetic studies in identifying causal alleles or regulatory elements, and motivated the functional genetic studies described here. Research in our lab had previously noted that the coincidence of strong Hb GWAS signal and selection signal in an H3K27Ac peak in endothelial cells, implicating this cell type in Tibetan adaptation.

In Chapter 2, I describe our comprehensive examination of the Tibetan *EPAS1* locus and its activity in the endothelium. As a critical hypoxia-response tissue, the endothelium represents an important site of hypoxic response and adaptation [138], [210], [235, p. 1]. Through a combination of genome-wide interrogations of chromatin architecture, and luciferase reporter assays I discovered multiple Tibetan alleles at the *EPAS1* locus than reduce enhancer activity in the endothelium. In order to assess the degree and nature of pleiotropy in the Tibetan haplotype and how it may contribute to unelevated hemoglobin, I re-examined enhancer activity in embryonic kidney cells, finding that all three endothelial enhancers demonstrated the same activity in this cell type. One enhancer, ENH5, was particularly of interest as it demonstrated hypoxia-responsive activity and downregulated both *EPAS1* and HIF-2α target *EGLN3* when

deleted *in vitro* in endothelial cells. Moreover, this enhancer was found to have the same activity in cardiomyocytes as seen in endothelium and kidney and displayed epigenetic evidence of activity in an even wider range of cell types.

The increasing evidence of pleiotropy of alleles at this enhancer prompted us to investigate its impacts in an *in vivo* knockout mouse model. This model provided crucial evidence of the broad activity of ENH5 and its potential downstream effects. Differentially expressed genes were identified in heart, lung, kidney and adrenal gland, with gene ontology enrichment analyses hinting at potentially adaptive outcomes including protection against pulmonary hypertension, lower pulse, and angiogenesis. The pleiotropy observed at ENH5 provides a major contribution to our understanding of the selective signature at *EPAS1*. There are few examples of strong selection signals in recent human evolution, and the evidence presented here may provide a mechanism through which these rare selective events can occur. Pleiotropic activity across tissues contributing to several smaller fitness effects, has been proposed as a mechanism for the selective sweeps observed for coding variants such as EDAR370A [32]. Here we may see a similar pattern recapitulated in non-coding variants in a highly pleiotropic enhancer element, ENH5. The many individual fitness effects contributed by these alleles in different tissues may ultimately combine to the observed strong selection signals at *EPAS1*.

5.2 Establishing an iPSC panel for the investigation of local adaptations

There are major hurdles to gaining a full understanding of tissue-specific effects of local adaptations. These include challenges to accessing material, short-term viability of primary cells in culture, limitations on the ability to collect and analyze samples in a controlled and balanced manner, and inabilities to perform specific genetic manipulations in many cell-types of interest. In recent years, a multitude of new technologies in molecular biology have come online, among

them, the ability to reliably generate human iPSCs from a variety of terminally differentiated or immortalized cell types. The high success-rate of reprogramming LCLs into iPSCs opens up the possibility of creating iPSC panels from many global populations [95]. In Chapter 3 I described a novel application of this technology to create one of the first iPSC panels designed to study local adaptation by comparing differentiated cell types from a locally adapted population and a closely related control. In this case, I reprogrammed 10 Tibetan and 10 Han Chinese (CHB, 1000 Genomes) LCLs into iPSCs and assessed their quality control metrics.

In Chapter 4, I differentiated this panel into vascular endothelial cells. While the differentiation protocol was broadly successful in generating CD144+ cells in all lines, our conclusions were limited by noise in the data. We believe based on previous studies in primary cells [210] that expression differences between Tibetan and Han Chinese populations are likely to be relatively subtle making it difficult to detect at genome-wide significance level. These true biological differences may therefore be overwhelmed by the inherent variation introduced in the generation and differentiation of iPSC lines [217], [225]. Nevertheless, our iPSC-derived vascular endothelium does capture differential expression of *EPAS1* between populations at the single-gene significance level. This finding underlines both the strength of the signal and the viability of the iPSC panel. We believe that several strategies may power us to gain insights from this panel in future differentiation experiments including i) increasing the size of the panel, ii) comparing isogenic lines with targeted CRISPR perturbations [232], and iii) employing additional strategies to correct for cell-type heterogeneity in the samples [227], [228].

All told, this dissertation not only provides novel insights into functional variation in Tibetan *EPAS1*, but also some of the first evidence of the pleiotropy of this variation and how it

may contribute to changes in gene expression in multiple tissues *in vivo*. I have also created to our knowledge the first panel of Tibetan iPSCs. When differentiated, the panel which captures expression differences in *EPAS1* between Tibetans and Han Chinese in the endothelium, and will hopefully enable us to investigate cell-type specific Tibetan phenotypes and expression patterns.

5.3 Future Directions

The results as well as the resources presented here open up many new avenues of investigation into Tibetan adaptations and their effects on not only molecular function, but organismal phenotype. With regards to our iPSC panel, we foresee a number of ways to increase power and make use of the cells to investigate genome-wide changes in expression patterns between Tibetans and CHB cells. As described above, we aim to increase the panel size, investigate methods of noise reduction, and identify and correct for cell type heterogeneity if possible [227], [228]. In addition, I am already involved in a project to knock out enhancer ENH5 in Tibetan iPSCs followed by differentiation. Isogenic backgrounds and specific single-locus perturbations may help overcome some of the inherent inter-line noise of iPSC-derived cells [232].

With regards to characterizing selected alleles at *EPAS1,* the results of RNA-sequencing in our *mus*ENH5 knockout line after an acute hypoxia exposure point to the importance of this enhancer for HIF-2α activity in many hypoxia-responsive tissues. However, due to the constraints of sample size and lack of equipment and protocols designed for longer hypoxia exposure, we have only scratched the surface of characterizing the phenotypes of this mouse line and the physiological impacts of ENH5 deletion. In the short term, I am already involved in talks to establish a protocol for chronic hypoxia in order to assess protection against pulmonary hypertension in our KO mouse line and perform additional RNA-seq experiments following this

long-term exposure. However, given the enrichment results of the DEGs detected in our mice, one can also envision more directed phenotypic inquiries into these mice exposed to either chronic or acute hypoxia. For example, many assays exist to interrogate whole body or tissue specific glucose metabolism in mice [236], [237]. Assessing metabolic differences between the KO and WT mice may be of interest to us given the multitude of metabolic term enrichments we identified. Additionally we may be interested in ascertaining if certain Tibetan phenotypes are recapitulated in these mice. For example, it would be interesting to determine if the KO mice have increased capillary density relative to WT, which could speak towards how Tibetans manage efficient oxygen delivery despite severe hypoxemia [12], [238]. Can we measure cardiac differences such as a blunted increase in heart rate in response to hypoxia [13]? More directly, we may look to detect differences in erythropoietin or hemoglobin in the blood of KO mice. Identifying phenotypes in our mouse line that can be connected with or used to infer Tibetan adaptive phenotypes will be crucial in translating the observed molecular differences to a deeper understanding of Tibetan adaptation.

BIBLIOGRAPHY

[1] O. Savolainen, M. Lascoux, and J. Merilä, "Ecological genomics of local adaptation," *Nat. Rev. Genet.*, vol. 14, no. 11, Art. no. 11, Nov. 2013, doi: 10.1038/nrg3522.

[2] I. Tattersall, "Why Was Human Evolution So Rapid?," in *Human Paleontology and Prehistory: Contributions in Honor of Yoel Rak*, A. Marom and E. Hovers, Eds. Cham: Springer International Publishing, 2017, pp. 1–9.

[3] S. A. Tishkoff *et al.*, "Convergent adaptation of human lactase persistence in Africa and Europe," *Nat. Genet.*, vol. 39, no. 1, Art. no. 1, Jan. 2007, doi: 10.1038/ng1946.

[4] D. M. Swallow, "Genetics of lactase persistence and lactose intolerance," *Annu. Rev. Genet.*, vol. 37, pp. 197–219, 2003, doi: 10.1146/annurev.genet.37.110801.143820.

[5] C. J. Adler *et al.*, "Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions," *Nat. Genet.*, vol. 45, no. 4, Art. no. 4, Apr. 2013, doi: 10.1038/ng.2536.

[6] A. Cardona *et al.*, "Genome-Wide Analysis of Cold Adaptation in Indigenous Siberian Populations," *PLOS ONE*, vol. 9, no. 5, p. e98076, May 2014, doi: 10.1371/journal.pone.0098076.

[7] N. G. Jablonski and G. Chaplin, "Human skin pigmentation as an adaptation to UV radiation," *Proc. Natl. Acad. Sci.*, vol. 107, no. Supplement 2, pp. 8962–8968, May 2010, doi: 10.1073/pnas.0914628107.

[8] K. Nakayama *et al.*, "Identification of novel functional variants of the melanocortin 1 receptor gene originated from Asians," *Hum. Genet.*, vol. 119, no. 3, pp. 322–330, Apr. 2006, doi: 10.1007/s00439-006-0141-1.

[9] M. Edwards *et al.*, "Association of the OCA2 Polymorphism His615Arg with Melanin Content in East Asian Populations: Further Evidence of Convergent Evolution of Skin Pigmentation," *PLOS Genet.*, vol. 6, no. 3, p. e1000867, Mar. 2010, doi: 10.1371/journal.pgen.1000867.

[10] C. M. Beall *et al*., "Natural selection on EPAS1 (HIF2) associated with low hemoglobin concentration in Tibetan highlanders," *Proc. Natl. Acad. Sci*., vol. 107, no. 25, pp. 11459–11464, Jun. 2010, doi: 10.1073/pnas.1002443107.

[11] C. Jeong *et al*., "Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal," *PLOS Genet*., vol. 14, no. 9, p. e1007650, Sep. 2018, doi: 10.1371/journal.pgen.1007650.

[12] C. M. Beall, "Two routes to functional adaptation: Tibetan and Andean high-altitude natives," *Proc. Natl. Acad. Sci*., vol. 104, no. Supplement 1, pp. 8655–8660, May 2007, doi: 10.1073/pnas.0701985104.

[13] N. Petousi *et al*., "Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor system and blunted physiological responses to hypoxia," *J. Appl. Physiol*., vol. 116, no. 7, pp. 893–904, Apr. 2014, doi: 10.1152/japplphysiol.00535.2013.

[14] L. B. Barreiro and L. Quintana-Murci, "Evolutionary and population (epi)genetics of immunity to infection," *Hum. Genet*., vol. 139, no. 6–7, pp. 723–732, Jun. 2020, doi: 10.1007/s00439-020-02167-x.

[15] M. Lopez *et al*., "Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest," *Curr. Biol*., vol. 29, no. 17, pp. 2926-2935.e4, Sep. 2019, doi: 10.1016/j.cub.2019.07.013.

[16] F. Racimo, S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez, "Evidence for archaic adaptive introgression in humans," *Nat. Rev. Genet*., vol. 16, no. 6, Art. no. 6, Jun. 2015, doi: 10.1038/nrg3936.

[17] J. M. Smith and J. Haigh, "The hitch-hiking effect of a favourable gene," *Genet. Res*., vol. 23, no. 1, pp. 23–35, Feb. 1974.

[18] J. K. Pritchard and A. Di Rienzo, "Adaptation – not by sweeps alone," *Nat. Rev. Genet*., vol. 11, no. 10, Art. no. 10, Oct. 2010, doi: 10.1038/nrg2880.

[19] J. K. Pritchard, J. K. Pickrell, and G. Coop, "The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation," *Curr. Biol. CB*, vol. 20, no. 4, pp. R208–R215, Feb. 2010, doi: 10.1016/j.cub.2009.11.055.

[20] R. D. Hernandez *et al.*, "Classic selective sweeps were rare in recent human evolution," *Science*, vol. 331, no. 6019, pp. 920–924, Feb. 2011, doi: 10.1126/science.1198878.

[21] P. W. Hedrick, "Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation," *Mol. Ecol.*, vol. 22, no. 18, pp. 4606–4618, 2013, doi: https://doi.org/10.1111/mec.12415.

[22] "UK Biobank," *Neale lab*. http://www.nealelab.is/uk-biobank (accessed Jan. 31, 2021).

[23] A. Nagai *et al.*, "Overview of the BioBank Japan Project: Study design and profile," *J. Epidemiol.*, vol. 27, no. 3 Suppl, pp. S2–S8, Feb. 2017, doi: 10.1016/j.je.2016.12.005.

[24] G. Lettre, C. Lange, and J. N. Hirschhorn, "Genetic model testing and statistical power in population-based association studies of quantitative traits," *Genet. Epidemiol.*, vol. 31, no. 4, pp. 358–362, May 2007, doi: 10.1002/gepi.20217.

[25] W. S. Bush and J. H. Moore, "Chapter 11: Genome-Wide Association Studies," *PLoS Comput. Biol.*, vol. 8, no. 12, Dec. 2012, doi: 10.1371/journal.pcbi.1002822.

[26] N. G. Crawford *et al.*, "Loci associated with skin pigmentation identified in African populations," *Science*, vol. 358, no. 6365, Nov. 2017, doi: 10.1126/science.aan8433.

[27] J. T. Daub *et al.*, "Evidence for polygenic adaptation to pathogens in the human genome," *Mol. Biol. Evol.*, vol. 30, no. 7, pp. 1544–1558, Jul. 2013, doi: 10.1093/molbev/mst080.

[28] J. J. Berg and G. Coop, "A population genetic signal of polygenic adaptation," *PLoS Genet*, vol. 10, p. 1004412, 2014.

[29] J. J. Berg, *Reduced signal for polygenic adaptation of height in UK Biobank*. 2018.

[30] P. C. Sabeti *et al*., "Genome-wide detection and characterization of positive selection in human populations," *Nature*, vol. 449, no. 7164, pp. 913–918, Oct. 2007, doi: 10.1038/nature06250.

[31] S. R. Grossman *et al*., "A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection," *Science*, vol. 327, no. 5967, pp. 883–886, Feb. 2010, doi: 10.1126/science.1183863.

[32] Y. G. Kamberov *et al*., "Modeling recent human evolution in mice by expression of a selected EDAR variant," *Cell*, vol. 152, no. 4, pp. 691–702, Feb. 2013, doi: 10.1016/j.cell.2013.01.016.

[33] J. S. Rees, S. Castellano, and A. M. Andrés, "The Genomics of Human Local Adaptation," *Trends Genet*., vol. 36, no. 6, pp. 415–428, Jun. 2020, doi: 10.1016/j.tig.2020.03.006.

[34] T. Goebel, "The 'Microblade Adaptation' and Recolonization of Siberia during the Late Upper Pleistocene," *Archaeol. Pap. Am. Anthropol. Assoc*., vol. 12, no. 1, pp. 117–131, 2002, doi: https://doi.org/10.1525/ap3a.2002.12.1.117.

[35] A. J. Osbora, "Eye of the Needle: Cold Stress, Clothing, and Sewing Technology During the Younger Dryas Cold Event in North America," *Am. Antiq*., vol. 79, no. 1, pp. 45–68, Jan. 2014, doi: 10.7183/0002-7316.79.1.45.

[36] A. J. Majmundar, W. J. Wong, and M. C. Simon, "Hypoxia-Inducible Factors and the Response to Hypoxic Stress," *Mol. Cell*, vol. 40, no. 2, pp. 294–309, Oct. 2010, doi: 10.1016/j.molcel.2010.09.022.

[37] G. L. Semenza, "Hypoxia-Inducible Factors in Physiology and Medicine," *Cell*, vol. 148, no. 3, pp. 399–408, Feb. 2012, doi: 10.1016/j.cell.2012.01.021.

[38] G.-H. Fong and K. Takeda, "Role and regulation of prolyl hydroxylase domain proteins," *Cell Death Differ*., vol. 15, no. 4, Art. no. 4, Apr. 2008, doi: 10.1038/cdd.2008.10.

[39] C.-J. Hu, L.-Y. Wang, L. A. Chodosh, B. Keith, and M. C. Simon, "Differential Roles of Hypoxia-Inducible Factor 1α (HIF-1α) and HIF-2α in Hypoxic Gene Regulation," *Mol.*

Cell. Biol., vol. 23, no. 24, pp. 9361–9374, Dec. 2003, doi: 10.1128/MCB.23.24.9361-9374.2003.

[40] A. S. Cowburn *et al.*, "HIF2α–arginase axis is essential for the development of pulmonary hypertension," *Proc. Natl. Acad. Sci.*, vol. 113, no. 31, pp. 8801–8806, Aug. 2016, doi: 10.1073/pnas.1602978113.

[41] A. W. Bigham and F. S. Lee, "Human high-altitude adaptation: forward genetics meets the HIF pathway," *Genes Dev.*, vol. 28, no. 20, pp. 2189–2204, Oct. 2014, doi: 10.1101/gad.250167.114.

[42] N. L. Downes, N. Laham-Karam, M. U. Kaikkonen, and S. Ylä-Herttuala, "Differential but Complementary HIF1α and HIF2α Transcriptional Regulation," *Mol. Ther.*, vol. 26, no. 7, pp. 1735–1745, Jul. 2018, doi: 10.1016/j.ymthe.2018.05.004.

[43] D. L. Ramírez-Bergeron, A. Runge, D. M. Adelman, M. Gohil, and M. C. Simon, "HIF-Dependent Hematopoietic Factors Regulate the Development of the Embryonic Vasculature," *Dev. Cell*, vol. 11, no. 1, pp. 81–92, Jul. 2006, doi: 10.1016/j.devcel.2006.04.018.

[44] J. R. Walls, L. Coultas, J. Rossant, and R. M. Henkelman, "Three-Dimensional Analysis of Vascular Development in the Mouse Embryo," *PLoS ONE*, vol. 3, no. 8, p. e2853, Aug. 2008, doi: 10.1371/journal.pone.0002853.

[45] Y. Yang *et al.*, "Transcriptome dynamics during human erythroid differentiation and development," *Genomics*, vol. 102, no. 5–6, pp. 431–441, Nov. 2013, doi: 10.1016/j.ygeno.2013.09.005.

[46] C. C. Zhang and H. A. Sadek, "Hypoxia and Metabolic Properties of Hematopoietic Stem Cells," *Antioxid. Redox Signal.*, vol. 20, no. 12, pp. 1891–1901, Apr. 2014, doi: 10.1089/ars.2012.5019.

[47] X. An, V. P. Schulz, N. Mohandas, and P. G. Gallagher, "Human and murine erythropoiesis:," *Curr. Opin. Hematol.*, vol. 22, no. 3, pp. 206–211, May 2015, doi: 10.1097/MOH.0000000000000134.

[48] A. V. Guitart *et al.*, "Hif-2α is not essential for cell-autonomous hematopoietic stem cell maintenance," *Blood*, vol. 122, no. 10, pp. 1741–1745, Sep. 2013, doi: 10.1182/blood-2013-02-484923.

[49] R. Naeije and R. Vanderpool, "Pulmonary Hypertension and Chronic Mountain Sickness," *High Alt. Med. Biol.*, vol. 14, no. 2, pp. 117–125, Jun. 2013, doi: 10.1089/ham.2012.1124.

[50] H. T. Duong, S. C. Erzurum, and K. Asosingh, "Pro-angiogenic hematopoietic progenitor cells and endothelial colony-forming cells in pathological angiogenesis of bronchial and pulmonary circulation," *Angiogenesis*, vol. 14, no. 4, pp. 411–422, Dec. 2011, doi: 10.1007/s10456-011-9228-y.

[51] S. Farha *et al.*, "Hypoxia-inducible factors in human pulmonary arterial hypertension: a link to the intrinsic myeloid abnormalities," *Blood*, vol. 117, no. 13, pp. 3485–3493, Mar. 2011, doi: 10.1182/blood-2010-09-306357.

[52] T. Löfstedt *et al.*, "Hypoxia Inducible Factor-2α in Cancer," *Cell Cycle*, vol. 6, no. 8, pp. 919–926, Apr. 2007, doi: 10.4161/cc.6.8.4133.

[53] C. E. Green and A. M. Turner, "The role of the endothelium in asthma and chronic obstructive pulmonary disease (COPD)," *Respir. Res.*, vol. 18, no. 1, p. 20, Dec. 2017, doi: 10.1186/s12931-017-0505-1.

[54] F. León-Velarde, F. C. Villafuerte, and J.-P. Richalet, "Chronic Mountain Sickness and the Heart," *Prog. Cardiovasc. Dis.*, vol. 52, no. 6, pp. 540–549, May 2010, doi: 10.1016/j.pcad.2010.02.012.

[55] M. Maggiorini and F. Leon-Velarde, "High-altitude pulmonary hypertension: a pathophysiological entity to different diseases," *Eur. Respir. J.*, vol. 22, no. 6, pp. 1019–1025, Dec. 2003, doi: 10.1183/09031936.03.00052403.

[56] M. Stembridge *et al.*, "Ventricular structure, function, and mechanics at high altitude: chronic remodeling in Sherpa vs. short-term lowlander adaptation," *J. Appl. Physiol.*, vol. 117, no. 3, pp. 334–343, Aug. 2014, doi: 10.1152/japplphysiol.00233.2014.

[57] R. L. Ge and G. Helun, "Current concept of chronic mountain sickness: pulmonary hypertension-related high-altitude heart disease," *Wilderness Environ. Med.*, vol. 12, no. 3, pp. 190–194, 2001, doi: 10.1580/1080-6032(2001)012[0190:ccocms]2.0.co;2.

[58] X. Li and J. W. Wilson, "Increased Vascularity of the Bronchial Mucosa in Mild Asthma," *Am. J. Respir. Crit. Care Med.*, vol. 156, no. 1, pp. 229–233, Jul. 1997, doi: 10.1164/ajrccm.156.1.9607066.

[59] B. Vrugt, S. Wilson, A. Bron, S. t Holgate, R. Djukanovic, and R. Aalbers, "Bronchial angiogenesis in severe glucocorticoid-dependent asthma," *Eur. Respir. J.*, vol. 15, no. 6, p. 1014, Jun. 2000, doi: 10.1034/j.1399-3003.2000.01507.x.

[60] G. Salvato, "Quantitative and morphological analysis of the vascular bed in bronchial biopsy specimens from asthmatic and non-asthmatic subjects," *Thorax*, vol. 56, no. 12, pp. 902–906, Dec. 2001, doi: 10.1136/thorax.56.12.902.

[61] T. D. Price, A. Qvarnström, and D. E. Irwin, "The role of phenotypic plasticity in driving genetic evolution," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 270, no. 1523, pp. 1433–1440, Jul. 2003, doi: 10.1098/rspb.2003.2372.

[62] J. F. Storz, G. R. Scott, and Z. A. Cheviron, "Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates," *J. Exp. Biol.*, vol. 213, no. 24, pp. 4125–4136, Dec. 2010, doi: 10.1242/jeb.048181.

[63] J. Merilä, "Perplexing effects of phenotypic plasticity," *Nature*, vol. 525, no. 7569, pp. 326–327, Sep. 2015, doi: 10.1038/nature15214.

[64] C. K. Ghalambor, J. K. McKAY, S. P. Carroll, and D. N. Reznick, "Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments," *Funct. Ecol.*, vol. 21, no. 3, pp. 394–407, 2007, doi: https://doi.org/10.1111/j.1365-2435.2007.01283.x.

[65] C. K. Ghalambor, K. L. Hoke, E. W. Ruell, E. K. Fischer, D. N. Reznick, and K. A. Hughes, "Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature," *Nature*, vol. 525, no. 7569, pp. 372–375, Sep. 2015, doi: 10.1038/nature15256.

[66] Y. He *et al*., "Blunted nitric oxide regulation in Tibetans under high-altitude hypoxia," *Natl. Sci. Rev*., vol. 5, no. 4, pp. 516–529, Jul. 2018, doi: 10.1093/nsr/nwy037.

[67] H. V. Carey, "Elucidating Nature's Solutions to Heart, Lung, and Blood Diseases and Sleep Disorders," *Circ Res*, vol. 110, pp. 915–921, 2012.

[68] C. M. Beall *et al*., "Ventilation and hypoxic ventilatory response of Tibetan and Aymara high altitude natives," *Am. J. Phys. Anthropol*., vol. 104, no. 4, pp. 427–447, Dec. 1997, doi: 10.1002/(SICI)1096-8644(199712)104:4<427::AID-AJPA1>3.0.CO;2-P.

[69] C. M. Beall and M. C. Goldstein, "Hemoglobin concentration of pastoral nomads permanently resident at 4,850-5,450 meters in Tibet," *Am. J. Phys. Anthropol*., vol. 73, no. 4, pp. 433–438, Aug. 1987, doi: 10.1002/ajpa.1330730404.

[70] L. G. Moore, D. Young, R. E. McCullough, T. Droma, and S. Zamudio, "Tibetan protection from intrauterine growth restriction (IUGR) and reproductive loss at high altitude," *Am. J. Hum. Biol*., vol. 13, no. 5, pp. 635–644, 2001, doi: 10.1002/ajhb.1102.

[71] G. Alkorta-Aranburu, C. M. Beall, D. B. Witonsky, A. Gebremedhin, J. K. Pritchard, and A. Di Rienzo, "The Genetic Architecture of Adaptations to High Altitude in Ethiopia," *PLoS Genet*., vol. 8, no. 12, p. e1003110, Dec. 2012, doi: 10.1371/journal.pgen.1003110.

[72] C. M. Beall *et al*., "Percent of oxygen saturation of arterial hemoglobin among Bolivian Aymara at 3,900-4,000 m," *Am. J. Phys. Anthropol*., vol. 108, no. 1, pp. 41–51, Jan. 1999, doi: 10.1002/(SICI)1096-8644(199901)108:1<41::AID-AJPA2>3.0.CO;2-K.

[73] C. M. Beall *et al*., "Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara," *Am. J. Phys. Anthropol*., vol. 106, no. 3, pp. 385–400, Jul. 1998, doi: 10.1002/(SICI)1096-8644(199807)106:3<385::AID-AJPA10>3.0.CO;2-X.

[74] C. Jeong *et al*., "Admixture facilitates genetic adaptations to high altitude in Tibet," *Nat. Commun*., vol. 5, no. 1, Art. no. 1, Feb. 2014, doi: 10.1038/ncomms4281.

[75] C. M. Beall and A. B. Reichsman, "Hemoglobin levels in a Himalayan high altitude population," *Am. J. Phys. Anthropol*., vol. 63, no. 3, pp. 301–306, Mar. 1984, doi: 10.1002/ajpa.1330630306.

[76] L. G. Moore, S. Zamudio, J. Zhuang, S. Sun, and T. Droma, "Oxygen transport in tibetan women during pregnancy at 3,658 m," *Am. J. Phys. Anthropol.*, vol. 114, no. 1, pp. 42–53, Jan. 2001, doi: 10.1002/1096-8644(200101)114:1<42::AID-AJPA1004>3.0.CO;2-B.

[77] J. I. Cho *et al.*, "Ethnically Tibetan women in Nepal with low hemoglobin concentration have better reproductive outcomes," *Evol. Med. Public Health*, vol. 2017, no. 1, pp. 82–96, 2017, doi: 10.1093/emph/eox008.

[78] Y. Peng, "Down-Regulation of EPAS1 Transcription and Genetic Adaptation of Tibetans to High-Altitude Hypoxia," *Mol Biol Evol*, vol. 34, pp. 818–830, 2017.

[79] C. M. Beall, "Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders," *Proc Natl Acad Sci U A*, vol. 107, pp. 11459–11464, 2010.

[80] J. Yang *et al.*, "Genetic signatures of high-altitude adaptation in Tibetans," *Proc. Natl. Acad. Sci.*, vol. 114, no. 16, pp. 4189–4194, Apr. 2017, doi: 10.1073/pnas.1617042114.

[81] C. T. Noguchi, "Where the Epo cells are," *Blood*, vol. 111, no. 10, pp. 4836–4837, May 2008, doi: 10.1182/blood-2008-02-135988.

[82] Ö. A. Amburgey, E. Ing, G. J. Badger, and I. M. Bernstein, "Maternal Hemoglobin Concentration and its Association with Birth Weight in Newborns of Mothers with Preeclampsia," *J. Matern.-Fetal Neonatal Med. Off. J. Eur. Assoc. Perinat. Med. Fed. Asia Ocean. Perinat. Soc. Int. Soc. Perinat. Obstet.*, vol. 22, no. 9, pp. 740–744, Sep. 2009, doi: 10.3109/14767050902926947.

[83] S. Maghsoudlou *et al.*, "Maternal haemoglobin concentrations before and during pregnancy and stillbirth risk: a population-based case-control study," *BMC Pregnancy Childbirth*, vol. 16, no. 1, p. 135, Jun. 2016, doi: 10.1186/s12884-016-0924-x.

[84] S. K. Palmer, L. G. Moore, D. Young, B. Cregger, J. C. Berman, and S. Zamudio, "Altered blood pressure course during normal pregnancy and increased preeclampsia at high altitude (3100 meters) in Colorado," *Am. J. Obstet. Gynecol.*, vol. 180, no. 5, pp. 1161–1168, May 1999, doi: 10.1016/s0002-9378(99)70611-3.

[85] R. Naeije, "Physiological Adaptation of the Cardiovascular System to High Altitude," *Prog. Cardiovasc. Dis.*, vol. 52, no. 6, pp. 456–466, May 2010, doi: 10.1016/j.pcad.2010.03.004.

[86] S. K. Ganesh *et al.*, "Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium," *Nat. Genet.*, vol. 41, no. 11, pp. 1191–1198, Nov. 2009, doi: 10.1038/ng.466.

[87] S. Hackinger *et al.*, "Wide distribution and altitude correlation of an archaic high-altitude-adaptive EPAS1 haplotype in the Himalayas," *Hum. Genet.*, vol. 135, no. 4, pp. 393–402, Apr. 2016, doi: 10.1007/s00439-016-1641-2.

[88] E. Huerta-Sánchez *et al.*, "Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA," *Nature*, vol. 512, no. 7513, pp. 194–197, Aug. 2014, doi: 10.1038/nature13408.

[89] D. Reich *et al.*, "Genetic history of an archaic hominin group from Denisova Cave in Siberia," *Nature*, vol. 468, no. 7327, Art. no. 7327, Dec. 2010, doi: 10.1038/nature09710.

[90] P. Skoglund and M. Jakobsson, "Archaic human ancestry in East Asia," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 45, pp. 18301–18306, Nov. 2011, doi: 10.1073/pnas.1108181108.

[91] D. Reich *et al.*, "Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania," *Am. J. Hum. Genet.*, vol. 89, no. 4, pp. 516–528, Oct. 2011, doi: 10.1016/j.ajhg.2011.09.005.

[92] F. Chen *et al.*, "A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau," *Nature*, vol. 569, no. 7756, Art. no. 7756, May 2019, doi: 10.1038/s41586-019-1139-x.

[93] G. Chen *et al.*, "Chemically defined conditions for human iPSC derivation and culture," *Nat. Methods*, vol. 8, no. 5, Art. no. 5, May 2011, doi: 10.1038/nmeth.1593.

[94] C. K. Burrows *et al.*, "Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs," *PLoS Genet.*, vol. 12, no. 1, p. e1005793, Jan. 2016, doi: 10.1371/journal.pgen.1005793.

[95] S. M. Thomas *et al*., "Reprogramming LCLs to iPSCs Results in Recovery of Donor-Specific Gene Expression Signature," *PLOS Genet*., vol. 11, no. 5, p. e1005216, May 2015, doi: 10.1371/journal.pgen.1005216.

[96] H. Kilpinen *et al*., "Common genetic variation drives molecular heterogeneity in human iPSCs," *Nature*, vol. 546, no. 7658, Art. no. 7658, Jun. 2017, doi: 10.1038/nature22403.

[97] S. M. Choi, "Reprogramming of EBV-immortalized B-lymphocyte cell lines into induced pluripotent stem cells," *Blood*, vol. 118, pp. 1801–1805, 2011.

[98] N. E. Banovich, "Impact of regulatory variation across human iPSCs and differentiated cells," *Genome Res*, vol. 28, pp. 122–131, 2018.

[99] M. A. Israel *et al*., "Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells," *Nature*, vol. 482, no. 7384, Art. no. 7384, Feb. 2012, doi: 10.1038/nature10821.

[100] C. Kim, "iPSC technology-Powerful hand for disease modeling and therapeutic screen," *BMB Rep*., vol. 48, no. 5, pp. 256–265, May 2015, doi: 10.5483/BMBRep.2015.48.5.100.

[101] M. D. Davidson, B. R. Ware, and S. R. Khetani, "Stem cell-derived liver cells for drug testing and disease modeling," *Discov Med*, vol. 19, pp. 349–358, 2015.

[102] F. T. Merkle and K. Eggan, "Modeling Human Disease with Pluripotent Stem Cells: from Genome Association to Function," *Cell Stem Cell*, vol. 12, no. 6, pp. 656–668, Jun. 2013, doi: 10.1016/j.stem.2013.05.016.

[103] B. M. Groves *et al*., "Minimal hypoxic pulmonary hypertension in normal Tibetans at 3,658 m," *J. Appl. Physiol. Bethesda Md 1985*, vol. 74, no. 1, pp. 312–318, Jan. 1993, doi: 10.1152/jappl.1993.74.1.312.

[104] G. L. Semenza, "Hypoxia-Inducible Factor 1 and Cardiovascular Disease," *Annu. Rev. Physiol*., vol. 76, no. 1, pp. 39–56, Feb. 2014, doi: 10.1146/annurev-physiol-021113-170322.

[105] A. Franovic, C. E. Holterman, J. Payette, and S. Lee, “Human cancers converge at the HIF-2␣ oncogenic axis,” *Med. Sci.*, p. 6.

[106] W. Gao, Y. Gao, G. Zhang, L. Song, B. Sun, and J. Shi, “Hypoxia-induced expression of HIF-1α and its target genes in umbilical venous endothelial cells of Tibetans and immigrant Han,” *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.*, vol. 141, no. 1, pp. 93–100, May 2005, doi: 10.1016/j.cca.2005.05.006.

[107] R. M. Garruto, C.-T. Chin, C. A. Weitz, J.-C. Liu, R.-L. Liu, and X. He, “Hematological differences during growth among Tibetans and Han Chinese born and raised at high altitude in Qinghai, China,” *Am. J. Phys. Anthropol.*, vol. 122, no. 2, pp. 171–183, Oct. 2003, doi: 10.1002/ajpa.10283.

[108] C. M. Beall, “Quantitative genetic analysis of arterial oxygen saturation in Tibetan highlanders,” *Hum Biol*, vol. 69, pp. 597–604, 1997.

[109] X. Qi, C. Cui, Ouzhuluobu, T. Wu, and B. Su, “Prehistoric Colonization and Demographic History of Modern Humans on the Tibetan Plateau,” in *eLS*, John Wiley & Sons Ltd, Ed. Chichester, UK: John Wiley & Sons, Ltd, 2014, p. a0025527.

[110] M. Aldenderfer, “Peopling the Tibetan Plateau: Insights from Archaeology,” *High Alt. Med. Biol.*, vol. 12, no. 2, pp. 141–147, Jun. 2011, doi: 10.1089/ham.2010.1094.

[111] P. I. Aaronson *et al.*, “Hypoxic pulmonary vasoconstriction: mechanisms and controversies: Hypoxic pulmonary vasoconstriction,” *J. Physiol.*, vol. 570, no. 1, pp. 53–58, Jan. 2006, doi: 10.1113/jphysiol.2005.098855.

[112] S. Zamudio *et al.*, “Protection from intrauterine growth retardation in Tibetans at high altitude,” *Am. J. Phys. Anthropol.*, vol. 91, no. 2, pp. 215–224, 1993, doi: https://doi.org/10.1002/ajpa.1330910207.

[113] R. M. Winslow *et al.*, “Different hematologic responses to hypoxia in Sherpas and Quechua Indians,” *J. Appl. Physiol. Bethesda Md 1985*, vol. 66, no. 4, pp. 1561–1569, Apr. 1989, doi: 10.1152/jappl.1989.66.4.1561.

[114] T. S. Simonson, D. A. McClain, L. B. Jorde, and J. T. Prchal, "Genetic determinants of Tibetan high-altitude adaptation," *Hum. Genet.*, vol. 131, no. 4, pp. 527–533, Apr. 2012, doi: 10.1007/s00439-011-1109-3.

[115] J. Li *et al.*, "GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children," *Hum. Mol. Genet.*, vol. 22, no. 7, pp. 1457–1464, Apr. 2013, doi: 10.1093/hmg/dds534.

[116] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide," *Curr. Protoc. Mol. Biol.*, vol. 109, no. 1, Jan. 2015, doi: 10.1002/0471142727.mb2129s109.

[117] X. Bao *et al.*, "A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63," *Genome Biol.*, vol. 16, no. 1, p. 284, Dec. 2015, doi: 10.1186/s13059-015-0840-9.

[118] S. S. P. Rao *et al.*, "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, Dec. 2014, doi: 10.1016/j.cell.2014.11.021.

[119] L. E. Montefiori *et al.*, "A promoter interaction map for cardiovascular disease genetics," *eLife*, vol. 7, p. e35788, Jul. 2018, doi: 10.7554/eLife.35788.

[120] J. Cairns *et al.*, "CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data," *Genome Biol.*, vol. 17, no. 1, p. 127, Dec. 2016, doi: 10.1186/s13059-016-0992-2.

[121] S. Heinz *et al.*, "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities," *Mol. Cell*, vol. 38, no. 4, pp. 576–589, May 2010, doi: 10.1016/j.molcel.2010.05.004.

[122] X. Zhou *et al.*, "Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser," *Nat. Biotechnol.*, vol. 33, no. 4, Art. no. 4, Apr. 2015, doi: 10.1038/nbt.3158.

[123] W. C. Claycomb *et al.*, "HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte," *Proc. Natl. Acad. Sci.*, vol. 95, no. 6, pp. 2979–2984, Mar. 1998, doi: 10.1073/pnas.95.6.2979.

[124] M. D. Krause *et al.*, "Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 48, pp. E11349–E11358, 27 2018, doi: 10.1073/pnas.1810568115.

[125] S. Bakhashab *et al.*, "Reference Genes for Expression Studies in Hypoxia and Hyperglycemia Models in Human Umbilical Vein Endothelial Cells," *G3amp58 GenesGenomesGenetics*, vol. 4, no. 11, pp. 2159–2165, Nov. 2014, doi: 10.1534/g3.114.013102.

[126] Y.-J. Peng, J. L. Overholt, D. Kline, G. K. Kumar, and N. R. Prabhakar, "Induction of sensory long-term facilitation in the carotid body by intermittent hypoxia: Implications for recurrent apneas," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 17, pp. 10073–10078, Aug. 2003, doi: 10.1073/pnas.1734109100.

[127] D. D. Kline, T. Yang, P. L. Huang, and N. R. Prabhakar, "Altered respiratory responses to hypoxia in mutant mice deficient in neuronal nitric oxide synthase," *J. Physiol.*, vol. 511, no. Pt 1, pp. 273–287, Aug. 1998, doi: 10.1111/j.1469-7793.1998.273bi.x.

[128] B. Phipson, S. Lee, I. J. Majewski, W. S. Alexander, and G. K. Smyth, "ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION," *Ann. Appl. Stat.*, vol. 10, no. 2, pp. 946–963, Jun. 2016, doi: 10.1214/16-AOAS920.

[129] C. W. Law, Y. Chen, W. Shi, and G. K. Smyth, "voom: precision weights unlock linear model analysis tools for RNA-seq read counts," *Genome Biol.*, vol. 15, no. 2, p. R29, Feb. 2014, doi: 10.1186/gb-2014-15-2-r29.

[130] S. M. Urbut, G. Wang, P. Carbonetto, and M. Stephens, "Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions," *Nat. Genet.*, vol. 51, no. 1, pp. 187–195, Jan. 2019, doi: 10.1038/s41588-018-0268-8.

[131] M. D. Young, M. J. Wakefield, G. K. Smyth, and A. Oshlack, "Gene ontology analysis for RNA-seq: accounting for selection bias," *Genome Biol.*, vol. 11, no. 2, p. R14, Feb. 2010, doi: 10.1186/gb-2010-11-2-r14.

[132] U. Raudvere *et al.*, "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W191–W198, Jul. 2019, doi: 10.1093/nar/gkz369.

[133] F. Cunningham *et al.*, "Ensembl 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D745–D751, Jan. 2019, doi: 10.1093/nar/gky1113.

[134] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe, "New approach for understanding genome variations in KEGG," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D590–D595, Jan. 2019, doi: 10.1093/nar/gky962.

[135] D. N. Slenter *et al.*, "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D661–D667, Jan. 2018, doi: 10.1093/nar/gkx1064.

[136] P. N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos, "The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease," *Am. J. Hum. Genet.*, vol. 83, no. 5, pp. 610–615, Nov. 2008, doi: 10.1016/j.ajhg.2008.09.017.

[137] The ENCODE Project Consortium, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012, doi: 10.1038/nature11247.

[138] R. Bartoszewski *et al.*, "Primary endothelial cell–specific regulation of hypoxia-inducible factor (HIF)-1 and HIF-2 and their target gene expression profiles during hypoxia," *FASEB J.*, vol. 33, no. 7, pp. 7929–7941, Jul. 2019, doi: 10.1096/fj.201802650RR.

[139] P. P. Kapitsinou *et al.*, "The Endothelial Prolyl-4-Hydroxylase Domain 2/Hypoxia-Inducible Factor 2 Axis Regulates Pulmonary Artery Pressure in Mice," *Mol. Cell. Biol.*, vol. 36, no. 10, pp. 1584–1594, May 2016, doi: 10.1128/MCB.01055-15.

[140] F. Coulet, S. Nadaud, M. Agrapart, and F. Soubrier, "Identification of Hypoxia-response Element in the Human Endothelial Nitric-oxide Synthase Gene Promoter," *J. Biol. Chem.*, vol. 278, no. 47, pp. 46230–46240, Nov. 2003, doi: 10.1074/jbc.M305420200.

[141] K. Brusselmans *et al.*, "Heterozygous deficiency of hypoxia-inducible factor–2α protects mice against pulmonary hypertension and right ventricular dysfunction during prolonged hypoxia," *J. Clin. Invest.*, vol. 111, no. 10, pp. 1519–1527, May 2003, doi: 10.1172/JCI15496.

[142] C.-J. Hu *et al.*, "Suppression of HIF2 signalling attenuates the initiation of hypoxia-induced pulmonary hypertension," *Eur. Respir. J.*, vol. 54, no. 6, Dec. 2019, doi: 10.1183/13993003.00378-2019.

[143] M. Batie, L. del Peso, and S. Rocha, "Hypoxia and Chromatin: A Focus on Transcriptional Repression Mechanisms," *Biomedicines*, vol. 6, no. 2, Apr. 2018, doi: 10.3390/biomedicines6020047.

[144] A. B. Johnson and M. C. Barton, "Hypoxia-induced and stress-specific changes in chromatin structure and function," *Mutat. Res.*, vol. 618, no. 1–2, pp. 149–162, May 2007, doi: 10.1016/j.mrfmmm.2006.10.007.

[145] A. Buniello *et al.*, "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1005–D1012, Jan. 2019, doi: 10.1093/nar/gky1120.

[146] W. J. Kent *et al.*, "The human genome browser at UCSC," *Genome Res.*, vol. 12, no. 6, pp. 996–1006, Jun. 2002, doi: 10.1101/gr.229102.

[147] P. Imanirad and E. Dzierzak, "Hypoxia and HIFs in regulating the development of the hematopoietic system," *Blood Cells. Mol. Dis.*, vol. 51, no. 4, pp. 256–263, Dec. 2013, doi: 10.1016/j.bcmd.2013.08.005.

[148] J. Ernst and M. Kellis, "ChromHMM: automating chromatin-state discovery and characterization," *Nat. Methods*, vol. 9, no. 3, Art. no. 3, Mar. 2012, doi: 10.1038/nmeth.1906.

[149] B. E. Bernstein *et al*., "The NIH Roadmap Epigenomics Mapping Consortium," *Nat. Biotechnol*., vol. 28, no. 10, pp. 1045–1048, Oct. 2010, doi: 10.1038/nbt1010-1045.

[150] M. F. Melgar, F. S. Collins, and P. Sethupathy, "Discovery of active enhancers through bidirectional expression of short transcripts," *Genome Biol*., vol. 12, no. 11, p. R113, Nov. 2011, doi: 10.1186/gb-2011-12-11-r113.

[151] N. D. Tippens, A. Vihervaara, and J. T. Lis, "Enhancer transcription: what, where, when, and why?," *Genes Dev*., vol. 32, no. 1, pp. 1–3, Jan. 2018, doi: 10.1101/gad.311605.118.

[152] C. Perez-Cervantes *et al*., "Enhancer transcription identifies cis-regulatory elements for photoreceptor cell types," *Development*, vol. 147, no. 3, Feb. 2020, doi: 10.1242/dev.184432.

[153] X. H. Yang *et al*., "Transcription-factor-dependent enhancer transcription defines a gene regulatory network for cardiac rhythm," *eLife*, vol. 6, p. e31683, Dec. 2017, doi: 10.7554/eLife.31683.

[154] M. C. Simon, L. Liu, B. C. Barnhart, and R. M. Young, "Hypoxia-Induced Signaling in the Cardiovascular System," *Annu. Rev. Physiol*., vol. 70, pp. 51–71, 2008, doi: 10.1146/annurev.physiol.70.113006.100526.

[155] L. Bernardi, "Heart Rate and Cardiovascular Variability at High Altitude," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, Aug. 2007, pp. 6678–6680, doi: 10.1109/IEMBS.2007.4353892.

[156] E. Metzen and P. J. Ratcliffe, "HIF hydroxylation and cellular oxygen sensing," *Biol. Chem*., vol. 385, no. 3–4, pp. 223–230, Apr. 2004, doi: 10.1515/BC.2004.016.

[157] E. Metzen, D. P. Stiehl, K. Doege, J. H. Marxsen, T. Hellwig-Bürgel, and W. Jelkmann, "Regulation of the prolyl hydroxylase domain protein 2 (phd2/egln-1) gene: identification of a functional hypoxia-responsive element," *Biochem. J*., vol. 387, no. 3, pp. 711–717, May 2005, doi: 10.1042/BJ20041736.

[158] X. Li, Z.-N. Han, Y. Liu, L. Hong, B.-R. Cui, and X. Cui, "Endogenous ET-1 promotes ANP secretion through activation of COX2-L-PGDS-PPARγ signaling in hypoxic beating rat atria," *Peptides*, vol. 122, p. 170150, Dec. 2019, doi: 10.1016/j.peptides.2019.170150.

[159] V. L. Dengler, M. Galbraith, and J. M. Espinosa, "Transcriptional Regulation by Hypoxia Inducible Factors," *Crit. Rev. Biochem. Mol. Biol.*, vol. 49, no. 1, pp. 1–15, 2014, doi: 10.3109/10409238.2013.838205.

[160] X. Xue and Y. M. Shah, "Hypoxia-inducible factor-2α is essential in activating the COX2/mPGES-1/PGE2 signaling axis in colon cancer," *Carcinogenesis*, vol. 34, no. 1, pp. 163–169, Jan. 2013, doi: 10.1093/carcin/bgs313.

[161] F. Luo *et al.*, "A MALAT1/HIF-2α feedback loop contributes to arsenite carcinogenesis," *Oncotarget*, vol. 7, no. 5, pp. 5769–5787, Dec. 2015, doi: 10.18632/oncotarget.6806.

[162] A. Yamaguchi *et al.*, "Peg3/Pw1 is involved in p53-mediated cell death pathway in brain ischemia/hypoxia," *J. Biol. Chem.*, vol. 277, no. 1, pp. 623–629, Jan. 2002, doi: 10.1074/jbc.M107435200.

[163] Y. Chen and P. Ernst, "Hematopoietic transformation in the absence of MLL1/KMT2A: distinctions in target gene reactivation," *Cell Cycle Georget. Tex*, vol. 18, no. 14, pp. 1525–1531, Jul. 2019, doi: 10.1080/15384101.2019.1618642.

[164] D. Wang *et al.*, "Cardiomyocyte cyclooxygenase-2 influences cardiac rhythm and function," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 18, pp. 7548–7552, May 2009, doi: 10.1073/pnas.0805806106.

[165] C. U. Andersen, O. Hilberg, S. Mellemkjær, J. E. Nielsen-Kudsk, and U. Simonsen, "Apelin and pulmonary hypertension," *Pulm. Circ.*, vol. 1, no. 3, p. 334, Sep. 2011, doi: 10.4103/2045-8932.87299.

[166] A. J. Zarrabi, D. Kao, D. T. Nguyen, J. Loscalzo, and D. E. Handy, "Hypoxia-induced suppression of c-Myc by HIF-2α in human pulmonary endothelial cells attenuates TFAM expression," *Cell. Signal.*, vol. 38, pp. 230–237, Oct. 2017, doi: 10.1016/j.cellsig.2017.07.008.

[167] A. Agrawal *et al*., "Cited2 modulates hypoxia-inducible factor-dependent expression of vascular endothelial growth factor in nucleus pulposus cells of the rat intervertebral disc," *Arthritis Rheum*., vol. 58, no. 12, pp. 3798–3808, Dec. 2008, doi: 10.1002/art.24073.

[168] X. Yi *et al*., "Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude," *Science*, vol. 329, no. 5987, pp. 75–78, Jul. 2010, doi: 10.1126/science.1190371.

[169] M. Y. Koh and G. Powis, "Passing the baton: The HIF switch," *Trends Biochem. Sci*., vol. 37, no. 9, pp. 364–372, Sep. 2012, doi: 10.1016/j.tibs.2012.06.004.

[170] J. Xin *et al*., "Chromatin accessibility landscape and regulatory network of high-altitude hypoxia adaptation," *Nat. Commun*., vol. 11, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41467-020-18638-8.

[171] H. Tang *et al*., "Endothelial HIF-2α Contributes to Severe Pulmonary Hypertension by Inducing Endothelial-to-Mesenchymal Transition," *Am. J. Physiol.-Lung Cell. Mol. Physiol*., p. ajplung.00096.2, Oct. 2017, doi: 10.1152/ajplung.00096.2017.

[172] C. Lamprecht *et al*., "Effect of selective cyclooxygenase-2 inhibitors on heart rate in healthy young men," *Am. J. Cardiol*., vol. 95, no. 12, pp. 1531–1532, Jun. 2005, doi: 10.1016/j.amjcard.2005.02.033.

[173] P. Soares *et al*., "The Expansion of mtDNA Haplogroup L3 within and out of Africa," *Mol. Biol. Evol*., vol. 29, no. 3, pp. 915–927, Mar. 2012, doi: 10.1093/molbev/msr245.

[174] C. Jeong and A. Di Rienzo, "Adaptations to local environments in modern human populations," *Curr. Opin. Genet. Dev*., vol. 29, pp. 1–8, Dec. 2014, doi: 10.1016/j.gde.2014.06.011.

[175] M. Haber *et al*., "A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans Out of Africa," *Genetics*, vol. 212, no. 4, pp. 1421–1428, Aug. 2019, doi: 10.1534/genetics.119.302368.

[176] M. O'Connell, B. Ghilardi, and L. Morrison, "A 7000-year record of environmental change, including early farming impact, based on lake-sediment geochemistry and pollen

data from County Sligo, western Ireland," *Quat. Res.*, vol. 81, no. 1, pp. 35–49, Jan. 2014, doi: 10.1016/j.yqres.2013.10.004.

[177] B. F. Voight, S. Kudaravalli, X. Wen, and J. K. Pritchard, "A Map of Recent Positive Selection in the Human Genome," *PLOS Biol.*, vol. 4, no. 3, p. e72, Mar. 2006, doi: 10.1371/journal.pbio.0040072.

[178] J. K. Pickrell *et al.*, "Signals of recent positive selection in a worldwide sample of human populations," *Genome Res.*, vol. 19, no. 5, pp. 826–837, May 2009, doi: 10.1101/gr.087577.108.

[179] A. J. Stern, L. Speidel, N. A. Zaitlen, and R. Nielsen, "Disentangling selection on genetically correlated polygenic traits using whole-genome genealogies," *bioRxiv*, p. 2020.05.07.083402, May 2020, doi: 10.1101/2020.05.07.083402.

[180] J. J. Berg, X. Zhang, and G. Coop, "Polygenic Adaptation has Impacted Multiple Anthropometric Traits," *bioRxiv*, p. 167551, Apr. 2019, doi: 10.1101/167551.

[181] M. Alidoosti *et al.*, "Study on the role of environmental parameters and HIF-1A gene polymorphism in coronary collateral formation among patients with ischemic heart disease," *Clin. Biochem.*, vol. 44, no. 17–18, pp. 1421–1424, Dec. 2011, doi: 10.1016/j.clinbiochem.2011.09.001.

[182] H. V. Carey *et al.*, "Elucidating Nature's Solutions to Heart, Lung, and Blood Diseases and Sleep Disorders," *Circ. Res.*, vol. 110, no. 7, pp. 915–921, Mar. 2012, doi: 10.1161/CIRCRESAHA.111.255398.

[183] V. H. Haase, "Regulation of erythropoiesis by hypoxia-inducible factors," *Blood Rev.*, vol. 27, no. 1, pp. 41–53, Jan. 2013, doi: 10.1016/j.blre.2012.12.003.

[184] F. J. Giordano, "Oxygen, oxidative stress, hypoxia, and heart failure," *J. Clin. Invest.*, vol. 115, no. 3, pp. 500–508, Mar. 2005, doi: 10.1172/JCI200524408.

[185] R. Ben Jehuda, Y. Shemer, and O. Binah, "Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9," *Stem Cell Rev. Rep.*, vol. 14, no. 3, pp. 323–336, Jun. 2018, doi: 10.1007/s12015-018-9811-3.

[186] International HapMap Consortium, "The International HapMap Project," *Nature*, vol. 426, no. 6968, pp. 789–796, Dec. 2003, doi: 10.1038/nature02168.

[187] 1000 Genomes Project Consortium *et al*., "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, pp. 68–74, Oct. 2015, doi: 10.1038/nature15393.

[188] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nat. Genet.*, vol. 38, no. 8, pp. 904–909, Aug. 2006, doi: 10.1038/ng1847.

[189] D. H. Alexander and K. Lange, "Enhancements to the ADMIXTURE algorithm for individual ancestry estimation," *BMC Bioinformatics*, vol. 12, no. 1, p. 246, Jun. 2011, doi: 10.1186/1471-2105-12-246.

[190] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, "Second-generation PLINK: rising to the challenge of larger and richer datasets," *GigaScience*, vol. 4, no. s13742-015-0047–8, Dec. 2015, doi: 10.1186/s13742-015-0047-8.

[191] I. Gallego Romero *et al*., "A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics," *eLife*, vol. 4, p. e07103, Jun. 2015, doi: 10.7554/eLife.07103.

[192] A. Dobin *et al*., "STAR: ultrafast universal RNA-seq aligner," *Bioinforma. Oxf. Engl.*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.

[193] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, p. 323, Aug. 2011, doi: 10.1186/1471-2105-12-323.

[194] A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen, "Robust relationship inference in genome-wide association studies," *Bioinforma. Oxf. Engl.*, vol. 26, no. 22, pp. 2867–2873, Nov. 2010, doi: 10.1093/bioinformatics/btq559.

[195] L. D'Aiuto, "Large-scale generation of human iPSC-derived neural stem cells/early neural progenitor cells and their neuronal differentiation," *Organogenesis*, vol. 10, pp. 365–377, 2014.

[196] E. E. Pashos *et al.*, "Large, Diverse Population Cohorts of hiPSCs and Derived Hepatocyte-like Cells Reveal Functional Genetic Variation at Blood Lipid-Associated Loci," *Cell Stem Cell*, vol. 20, no. 4, pp. 558-570.e10, Apr. 2017, doi: 10.1016/j.stem.2017.03.017.

[197] C. R. Warren *et al.*, "Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease," *Cell Stem Cell*, vol. 20, no. 4, pp. 547-557.e7, Apr. 2017, doi: 10.1016/j.stem.2017.01.010.

[198] M. C. Ward and Y. Gilad, "A generally conserved response to hypoxia in iPSC-derived cardiomyocytes from humans and chimpanzees," *eLife*, vol. 8, p. e42374, Apr. 2019, doi: 10.7554/eLife.42374.

[199] B. van de Geijn, G. McVicker, Y. Gilad, and J. K. Pritchard, "WASP: allele-specific software for robust molecular quantitative trait locus discovery," *Nat. Methods*, vol. 12, no. 11, pp. 1061–1063, Nov. 2015, doi: 10.1038/nmeth.3582.

[200] N. Kumasaka, A. J. Knights, and D. J. Gaffney, "Fine-mapping cellular QTLs with RASQUAL and ATAC-seq," *Nat. Genet.*, vol. 48, no. 2, pp. 206–213, Feb. 2016, doi: 10.1038/ng.3467.

[201] G. McVicker *et al.*, "Identification of genetic variants that affect histone modifications in human cells," *Science*, vol. 342, no. 6159, pp. 747–749, Nov. 2013, doi: 10.1126/science.1242429.

[202] K. J. Brennand *et al.*, "Modelling schizophrenia using human induced pluripotent stem cells," *Nature*, vol. 473, no. 7346, pp. 221–225, May 2011, doi: 10.1038/nature09915.

[203] C. Patsch *et al.*, "Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells," *Nat. Cell Biol.*, vol. 17, no. 8, Art. no. 8, Aug. 2015, doi: 10.1038/ncb3205.

[204] P. Rana, B. Anson, S. Engle, and Y. Will, "Characterization of human-induced pluripotent stem cell-derived cardiomyocytes: bioenergetics and utilization in safety screening," *Toxicol. Sci. Off. J. Soc. Toxicol.*, vol. 130, no. 1, pp. 117–131, Nov. 2012, doi: 10.1093/toxsci/kfs233.

[205] A. Seki and S. Rutz, "Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells," *J. Exp. Med.*, vol. 215, no. 3, pp. 985–997, Mar. 2018, doi: 10.1084/jem.20171626.

[206] P. G. Kamble *et al.*, "Proof-of-concept for CRISPR/Cas9 gene editing in human preadipocytes: Deletion of FKBP5 and PPARG and effects on adipocyte differentiation and metabolism," *Sci. Rep.*, vol. 10, no. 1, p. 10565, Jun. 2020, doi: 10.1038/s41598-020-67293-y.

[207] D. Song *et al.*, "Tibetan PHD2, an allele with loss-of-function properties," *Proc. Natl. Acad. Sci.*, vol. 117, no. 22, pp. 12230–12238, Jun. 2020, doi: 10.1073/pnas.1920546117.

[208] C. Michiels, T. Arnould, and J. Remacle, "Endothelial cell responses to hypoxia: initiation of a cascade of cellular interactions," *Biochim. Biophys. Acta BBA - Mol. Cell Res.*, vol. 1497, no. 1, pp. 1–10, Jun. 2000, doi: 10.1016/S0167-4889(00)00041-0.

[209] C. M. Beall *et al.*, "Pulmonary nitric oxide in mountain dwellers," *Nature*, vol. 414, no. 6862, pp. 411–412, Nov. 2001, doi: 10.1038/35106641.

[210] Y. Peng *et al.*, "Down-Regulation of EPAS1 Transcription and Genetic Adaptation of Tibetans to High-Altitude Hypoxia," *Mol. Biol. Evol.*, vol. 34, no. 4, pp. 818–830, 01 2017, doi: 10.1093/molbev/msw280.

[211] S. P. Crampton, J. Davis, and C. C. W. Hughes, "Isolation of Human Umbilical Vein Endothelial Cells (HUVEC)," *J. Vis. Exp. JoVE*, no. 3, Apr. 2007, doi: 10.3791/183.

[212] D. P. Kaufman, T. Sanvictores, and M. Costanza, "Weibel Palade Bodies," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2020.

[213] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, p. 550, 2014.

[214] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, Jan. 2010, doi: 10.1093/bioinformatics/btp616.

[215] G. E. Hoffman and P. Roussos, "Dream: powerful differential expression analysis for repeated measures designs," *Bioinformatics*, no. btaa687, Jul. 2020, doi: 10.1093/bioinformatics/btaa687.

[216] W. Mao, E. Zaslavsky, B. M. Hartmann, S. C. Sealfon, and M. Chikina, "Pathway-level information extractor (PLIER) for gene expression data," *Nat. Methods*, vol. 16, no. 7, pp. 607–610, Jul. 2019, doi: 10.1038/s41592-019-0456-1.

[217] V. Volpato and C. Webber, "Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility," *Dis. Model. Mech.*, vol. 13, no. 1, Jan. 2020, doi: 10.1242/dmm.042317.

[218] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, "The sva package for removing batch effects and other unwanted variation in high-throughput experiments," *Bioinformatics*, vol. 28, no. 6, pp. 882–883, Mar. 2012, doi: 10.1093/bioinformatics/bts034.

[219] J. A. Gagnon-Bartsch and T. P. Speed, "Using control genes to correct for unwanted variation in microarray data," *Biostat. Oxf. Engl.*, vol. 13, no. 3, pp. 539–552, Jul. 2012, doi: 10.1093/biostatistics/kxr034.

[220] Y. Zhang, G. Parmigiani, and W. E. Johnson, "ComBat-seq: batch effect adjustment for RNA-seq count data," *NAR Genomics Bioinforma.*, vol. 2, no. 3, p. lqaa078, Sep. 2020, doi: 10.1093/nargab/lqaa078.

[221] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649–D655, Jan. 2018, doi: 10.1093/nar/gkx1132.

[222] A. R. Abbas *et al.*, "Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data," *Genes Immun.*, vol. 6, no. 4, Art. no. 4, Jun. 2005, doi: 10.1038/sj.gene.6364173.

[223] J. Mai, A. Virtue, J. Shen, H. Wang, and X.-F. Yang, "An evolving new paradigm: endothelial cells – conditional innate immune cells," *J. Hematol. Oncol.J Hematol Oncol*, vol. 6, p. 61, Aug. 2013, doi: 10.1186/1756-8722-6-61.

[224] K. Xiang *et al.*, "Identification of a Tibetan-Specific Mutation in the Hypoxic Gene EGLN1 and Its Contribution to High-Altitude Adaptation," *Mol. Biol. Evol.*, vol. 30, no. 8, pp. 1889–1898, Aug. 2013, doi: 10.1093/molbev/mst090.

[225] L. T. Ghaffari, A. Starr, A. T. Nelson, and R. Sattler, "Representing Diversity in the Dish: Using Patient-Derived in Vitro Models to Recreate the Heterogeneity of Neurological Disease," *Front. Neurosci.*, vol. 12, 2018, doi: 10.3389/fnins.2018.00056.

[226] J. Schwartzentruber *et al.*, "Molecular and functional variation in iPSC-derived sensory neurons," *Nat. Genet.*, vol. 50, no. 1, pp. 54–61, Jan. 2018, doi: 10.1038/s41588-017-0005-8.

[227] K. K. Dey, C. J. Hsiao, and M. Stephens, "Visualizing the structure of RNA-seq expression data using grade of membership models," *PLOS Genet.*, vol. 13, no. 3, p. e1006599, Mar. 2017, doi: 10.1371/journal.pgen.1006599.

[228] *stephenslab/fastTopics*. Stephens lab, 2021.

[229] M. B. de Moura and B. Van Houten, "Bioenergetic Analysis of Intact Mammalian Cells Using the Seahorse XF24 Extracellular Flux Analyzer and a Luciferase ATP Assay," in *Molecular Toxicology Protocols*, P. Keohavong and S. G. Grant, Eds. Totowa, NJ: Humana Press, 2014, pp. 589–602.

[230] K. J. Bayless, H.-I. Kwak, and S.-C. Su, "Investigating endothelial invasion and sprouting behavior in three-dimensional collagen matrices," *Nat. Protoc.*, vol. 4, no. 12, pp. 1888–1898, Dec. 2009, doi: 10.1038/nprot.2009.221.

[231] I. Arnaoutova and H. K. Kleinman, "In vitro angiogenesis: endothelial cell tube formation on gelled basement membrane extract," *Nat. Protoc.*, vol. 5, no. 4, pp. 628–635, Apr. 2010, doi: 10.1038/nprot.2010.6.

[232] S. Yumlu *et al.*, "Gene editing and clonal isolation of human induced pluripotent stem cells using CRISPR/Cas9," *Methods*, vol. 121–122, pp. 29–44, May 2017, doi: 10.1016/j.ymeth.2017.05.009.

[233] A. Bigham, "Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data," *PLoS Genet*, vol. 6, p. 1001116, 2010.

[234] J. Danesh, R. Collins, R. Peto, and G. D. Lowe, "Haematocrit, viscosity, erythrocyte sedimentation rate: meta-analyses of prospective studies of coronary heart disease," *Eur Heart J*, vol. 21, pp. 515–520, 2000.

[235] H. Tian, S. L. McKnight, and D. W. Russell, "Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells," p. 12.

[236] E. D. Berglund *et al.*, "Glucose Metabolism In Vivo in Four Commonly Used Inbred Mouse Strains," *Diabetes*, vol. 57, no. 7, pp. 1790–1799, Jul. 2008, doi: 10.2337/db07-1615.

[237] N. D. Oakes *et al.*, "Cardiac metabolism in mice: tracer method developments and in vivo application revealing profound metabolic inflexibility in diabetes," *Am. J. Physiol.-Endocrinol. Metab.*, vol. 290, no. 5, pp. E870–E881, May 2006, doi: 10.1152/ajpendo.00233.2005.

[238] H. Nagahisa and H. Miyata, "Influence of hypoxic stimulation on angiogenesis and satellite cells in mouse skeletal muscle," *PLoS ONE*, vol. 13, no. 11, Nov. 2018, doi: 10.1371/journal.pone.0207040.