GENETIC ADAPTATION TO HIGH ALTITUDE IN TIBETANS

by

Tatum S. Simonson

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Human Genetics

The University of Utah

December 2011

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Tatum S. Simonson**

has been approved by the following supervisory committee members:

Lynn Jorde	, Chair	**October 5, 2011** Date Approved
Gabrielle Kardon	, Member	**October 5, 2011** Date Approved
Josef Prchal	, Member	**October 5, 2011** Date Approved
Alan Rogers	, Member	**October 5, 2011** Date Approved
Jon Seger	, Member	**October 5, 2011** Date Approved
Mark Yandell	, Member	**October 5, 2011** Date Approved

and by **Lynn Jorde** , Chair of

the Department of **Human Genetics**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Humans have lived at high altitude for hundreds of generations despite unavoidable challenges imposed by hypobaric hypoxia. The lower barometric pressure at high altitude reduces the number of oxygen molecules available in each breath of air, yet oxygen-dependent physiological processes must be maintained for survival. Cellular and system responses to hypoxic stress can result in altitude illness and may prove fatal in a small proportion of maladapted individuals.

Native high-altitude populations, however, exhibit a unique suite of heritable traits that afford tolerance to hypoxia. Compared to lowland visitors and Andean highlanders, Tibetans exhibit lower hemoglobin (Hb) levels at high altitude, which tend to be similar to those expected under sea-level conditions. Such differences suggest this population has unique adaptations to their native environment.

It has been hypothesized that genes specifically involved in the hypoxia inducible factor (HIF) pathway could underlie adaptive changes in high-altitude populations. Genome-wide analyses provide the first lines of evidence in support of genetic adaptation to high altitude. Three regions of the genome that contain genes associated with the human response to hypoxia show evidence of selection and are associated with decreased Hb levels, and two of these are also associated with metabolite levels. These phenotypic associations provide corroborative evidence for adaptive roles of genomic regions targeted by strong positive selection in Tibetans.

While many of the same selection candidate genes are reported by studies of different Tibetan populations, some signals of selection and association are unique to particular groups. The genetic makeup of Tibetan groups located throughout the plateau is therefore important to consider in studies of high-altitude adaptation.

Taken together, the data presented in this dissertation demonstrate that multiple genes are involved in Tibetan adaptation to high altitude. Some of these genes have been linked to hematological and metabolic phenotypes characterized thus far, providing further support for roles in physiological adaptation to this extreme environment. Studies aimed to identify associations between specific genetic variants, mechanisms, and phenotypes will help bridge the gap between genetic variation and organismal responses to hypoxia, and will have important implications for understanding human health and disease.

“Nothing in biology
[or medicine]
makes sense except
in the light of evolution.”
- Theodosius Dobzhansky

“Your work is to discover your world
and with all your heart
give yourself to it.”
- Buddha

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

It took not one but *many* villages to bring the research presented here to fruition. I therefore have several mentors, colleagues, participants, interdisciplinary and cross-continental collaborators in addition to supportive friends and family members to thank for the successful completion of my dissertation work.

I first thank my advisor Lynn Jorde who has provided invaluable advice and backing during my tenure as a graduate student. I was drawn to Lynn Jorde's lab because of his strong interest in human genetic variation. He also has outstanding reputation as a scientist, educator, and leader, which is reflected by his recent success as both the Chair of the Department of Human Genetics and President of the American Society of Human Genetics. He provides an upstanding example of fairness and fosters a strong sense of collegiality among members of his lab. Over the past five years, he has helped me mature as a scientist, especially by giving me freedom to explore my own research path and by helping me develop skills essential to my success as an independent investigator. I am eternally grateful to Lynn Jorde for providing such an exceptional training environment which has helped me develop into the scientist I am today.

I also am fortunate to have a diverse group of thesis committee members who bring unique perspectives to my dissertation work. I appreciate the countless hours Alan Rogers and Jon Seger spent discussing ideas and projects with me. I am grateful to Gabrielle Kardon and Mark Yandell for the feedback and different perspectives they

offered following my department presentations and committee meetings, and I am eternally grateful to Josef Prchal who established the connection with our collaborator Rili Ge, thereby enabling the research studies presented here. My committee members have been truly instrumental to my development, and it has been a pleasure having their support and guidance over the past several years.

Members of the Jorde laboratory have also been excellent mentors and friends. I have enjoyed our conversations, collaborations, and the many adventures we shared outside of the lab. I especially want to thank the extremely talented post-doctoral fellows David Witherspoon, Jinchuan Xing, and Chad Huff who each played an important role in my scientific development over the years. I am fortunate to have learned so much from them and can only hope to emulate their success as postdocs in the next stage of my career. I would also like to acknowledge the research staff in our lab, Scott Watkins and Yuhua Zhang, whom are wonderful and insightful people to work with, and medical doctors Steve Guthery and Karin Chen, whose medical expertise is always greatly appreciated. I also appreciate the support of and enjoyed working with fellow graduate students Liz Marchani and Wilfred Wu, who are great colleagues and friends.

I am also very grateful for the limited funding that has gone a long way to directly support this research, including a seed grant awarded by the Center for Clinical and Translational Science and the T32 Genetics Training Grant student research fund that covered the costs of my airline travel to China. I also thank Lynn Jorde, the Department of Human Genetics, and the Graduate School for additional financial support for my graduate training.

There are several members of the Molecular Biology/Biological Chemistry community and the Department of Human Genetics who have offered assistance over the years, and I greatly appreciate the time they have invested in student development. I thank David Stillman, David Grunwald, and the students involved with the Genetics Training Grant for their enthusiasm, insightful questions, and helpful suggestions about my research, Karen Eilbeck, David Grunwald, and Mike Shapiro for supporting my interests from the beginning of my graduate career in Utah, Mark Metzstein, Gillian Stanfield, and Carl Thummel for thoughtful discussions and invaluable advice, and Donald McClain and Janet Lindsley for insight into the metabolic aspects of my thesis work. I am grateful to Barbara Saffel, Tami Brunson, Brad Cairns, and Anthea Letsou, as well as Hope Glenn, and Kim Antry for their assistance with the Molecular Biology Program and Genetics Training Grant, respectively, and Elaine Fry, Natalie Johnson, Patty Lisieski, Jean Eden, and June Blackburn for their assistance within the Department of Human Genetics. It has been an honor to be part of such an invigorating scientific and professional community at the University of Utah.

There are people half-way across the globe whom I have had the pleasure of working with and owe special thanks. Rili Ge has provided an exceptional opportunity for me to carry out studies in China and Tibet with him and his group and partake in rich cultural experiences I never could have imagined. I owe an immense “bayarlalaa” to Uran Tanna for her assistance and sincere enthusiasm towards our research endeavors. I also thank Ga Qin for his hard work and assistance with each one of our projects, and thank his family for the kindness and generosity they showed during my second trip to China. Likewise, I say “谢谢 (xiè xie”) to Haixia Yun and Hong Mei, among many

others in Rili Ge's group, for long days and nights of sample collection, and Dorje Nanka (Bai Zhengzong) for his kindness and assistance in the lab in Xining. Finally, I want to thank the inhabitants of the high-altitude villages who opened their hearts and homes during our field work and taught me timeless lessons about the universal nature of the human race. Their willingness to participate in our studies allowed our research group to push the state of high-altitude biology and medicine forward.

I would also like to thank various mentors in my life for their support relating to my personal and professional development. Mr. Charles Beckman, who I met at my very first job, and Ms. Gloria Nielsen, a kindred spirit who happened to be my high school journalism teacher, are two people who have encouraged me to follow my dreams throughout the half of my life I have known them, and I value their wisdom and support immensely. Ken Burbridge, Mark Schmidt, and Paul Keim each heightened my academic interests during elementary, high school, and college, respectively, and I thank them for being such inspiring teachers and excellent role models.

I am fortunate to have the continued support of many childhood friends and am grateful for the strong friendships developed with classmates in the Molecular Biology/Biological Chemistry Program, Anthropology and Biology Departments, and throughout the Salt Lake City community. My friends helped to shape the path I am on today. I thank many people in Utah for their advice, encouragement, and the memories of times spent exploring the wonderful city of Salt Lake, the Wasatch Mountains, the deserts in the south, and many beautiful places in between. I especially thank Luke Evans for being my rock, and am grateful for all the encouragement he has provided

during my graduate career. We have shared many adventures and stimulating conversations, and our time together made these years of my life so special.

Finally, I thank my family. My parents Mari and James Simonson have always been supportive of my endeavors. They provide an exceptional example of hard work and selflessness, and constant, much appreciated, reminders that "anything worth doing is not easy" and that "everything will fall into place." My love and appreciation for them is beyond words. I thank my late "Granny Goose" for many life lessons that shaped the person I am, and hope to be just as wise (and feisty) as her in my later years. I am also indebted to all of my siblings for helping me evolve into the person I am today. My sisters Tami, Tris, and Tara are models of strength, hard work, and resilience, and support me without question, as does my brother Tim, whose encouragement is appreciated and progress inspiring. The next generation of my family, nephews Bret, Austin, Tim, and Tanner and niece Ashton, each inspire me in their own special way. I am so proud of all they accomplished and continue to achieve in their young lives. I have so much respect for my family and dedicate the completion of this stage of my education to them.

CHAPTER 1

INTRODUCTION

Overview of human adaptation to high altitude

The decrease in partial pressure of oxygen at high altitude reduces oxygen availability and imposes extreme physiological challenges in humans (Beall, 2007). Cellular and system responses to this stress can result in altitude illness and may prove fatal in a proportion of maladapted individuals (Hackett and Roach, 2001; MacInnis et al., 2010). Some highland populations, however, have adaptations that enable their successful existence in a hypoxic environment (Beall, 2007).

Native high-altitude Tibetans exhibit a unique suite of heritable traits that afford tolerance to hypoxia (decreased oxygen tension) (Beall, 2000; Beall, 2007; Moore, 2001). Compared to lowland visitors and high-altitude native Andean groups, Tibetans living at high altitude exhibit elevated nitric oxide levels, increased resting ventilation and hypoxic ventilatory response, and greater birth weights at high altitude (Beall, 2007; Beall et al., 1998; Moore, 2001; Wu et al., 2005). They also have hemoglobin (Hb) levels similar to those expected at sea level (Beall, 2007; Beall et al., 1998; Moore, 2001; Wu et al., 2005), which may prevent high blood viscosity caused by polycythemia (a typical response to hypoxia and a hallmark of chronic mountain sickness) (Monge, 1976; Vargas and Spielvogel, 2006), or result from changes afforded by an adaptive trait other than Hb level. These physiological and pathological differences suggest that Tibetan populations possess unique adaptations that enable survival in their native high-altitude environment.

Genomewide analysis of selection can provide unprecedented insight into the biological mechanisms of recent human evolution. By combining novel genomic techniques with decades of physiological research, we now have the ability to begin to uncover the basis of high-altitude adaptation. In addition to answering long-standing

questions in the field of high-altitude research, such studies will enhance our understanding of the human response to hypoxia, which has broad-reaching implications in medical research. Overall, the goal of this dissertation is to understand the genetic basis of Tibetan adaptation to high altitude, and to determine whether selection candidate genes are associated with a subset of phenotypic measurements that provide biological insight into mechanisms of high-altitude adaptation.

Genes related to the hypoxia inducible factor (HIF) gene pathway are likely involved in high-altitude adaptation

The mechanisms underlying high-altitude adaptation are currently unknown but may be elucidated through identification of genetic factors that provide fitness advantages at high altitude. Hypoxia-related genes are key regulators of fundamental physiological processes in mammals including red blood cell production (erythropoiesis), development, energy metabolism, vasculogenesis, iron metabolism, cardiopulmonary regulation, and tumor promotion (Majmundar et al., 2010; Semenza, 2010). The hypoxia inducible factor (HIF) signaling cascade mediates the transcription of various genes in response to oxygen level and is central to mitigating the negative effects of cellular hypoxia.

HIFs are heterodimeric transcription factors comprised of a constitutively expressed β subunit and one of three α subunits (Figure 1.1). Two of the α subunits (HIF-1α and HIF-2α) play a central role in responding to hypoxia (Hirota and Semenza, 2006; Yoon et al., 2006). Under normoxic conditions, the α subunits are hydroxylated by prolyl hydroxylases (PHDs) and targeted for degradation through interaction with the von Hippel-Lindau (VHL) complex, although other factors are also involved in this regulation

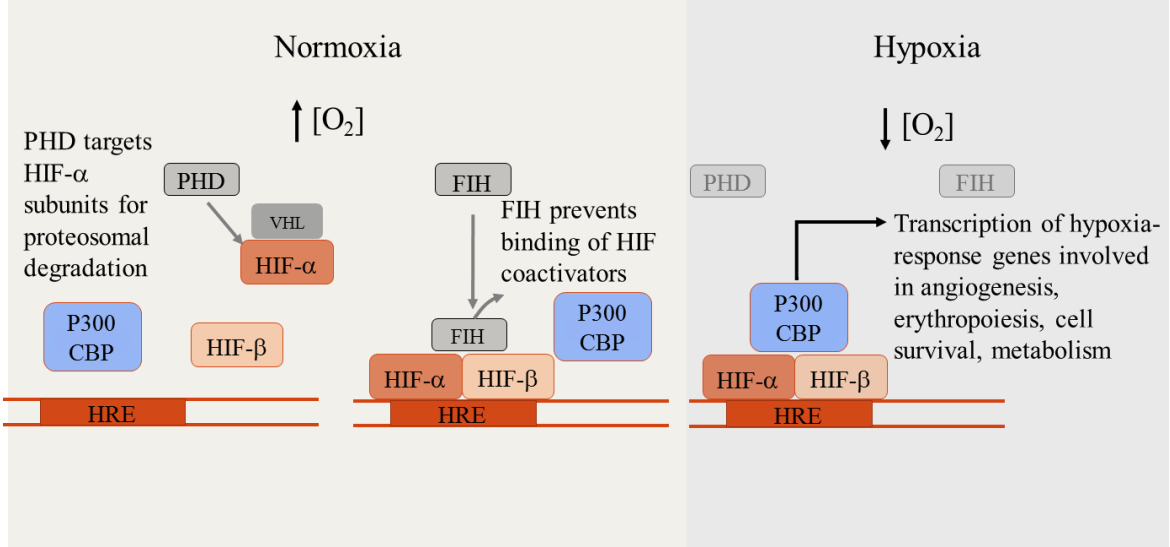


Figure 1.1: Overview of the hypoxia inducible factor (HIF) pathway. Degradation of HIF α subunits under normoxic conditions requires O_2, proline hydroxylase activity, iron, and VHL and is regulated by PHD and FIH. Under hypoxic conditions, HIF α subunits are not hydroxylated and are able to heterodimerize with the HIF β subunit, forming the active transcription factors HIF-1 and HIF-2. These factors bind with p300 and CBP coactivators, allowing for transcriptional activation of HIF-target genes.

(Kaelin and Ratcliffe, 2008; Majmundar et al., 2010). Under hypoxic conditions, the α and β subunits dimerize and transcribe HIF-regulated genes. Mutations identified in HIF-pathway genes, including *VHL*, *EGLN1*/*PHD2*, and *EPAS1*/*HIF2A*, are associated with congenital and/or hereditary erythrocytosis (Semenza, 2009), and misregulation of HIF pathway members has been shown to have metabolic consequences in humans (Formenti et al., 2010).

Genomic regions under strong selection contain genes likely involved in adaptation to high altitude in Tibetans

It has been hypothesized that genes specifically involved in the HIF pathway could underlie adaptive changes in high-altitude populations (Beall et al., 1998; Bigham, 2009; Moore, 2001; Rupert, 2010; Rupert and Koehle, 2006). However, prior to the advent of high-density genotyping platforms, it was difficult to test whether genetic factors were involved in high-altitude adaptation beyond examination of a limited set of candidate genes. It is now possible to conduct "population genomic" studies to determine informative patterns of genetic variation, such as those characteristic of a selective sweep, on a genome-wide scale (Figure 1.2) (Stinchcombe and Hoekstra, 2007). These powerful methods identify outlier regions of the genome that may be further examined to determine the exact genetic variant targeted by natural selection.

To test whether natural selection has acted on specific genes associated with high-altitude adaptation in Tibetans, we performed two genome-wide haplotype-based tests of selection: iHS (Voight et al., 2006) and XP-EHH (Sabeti et al., 2007). These statistics identify regions of the genome that exhibit a striking signal of haplotype homozygosity

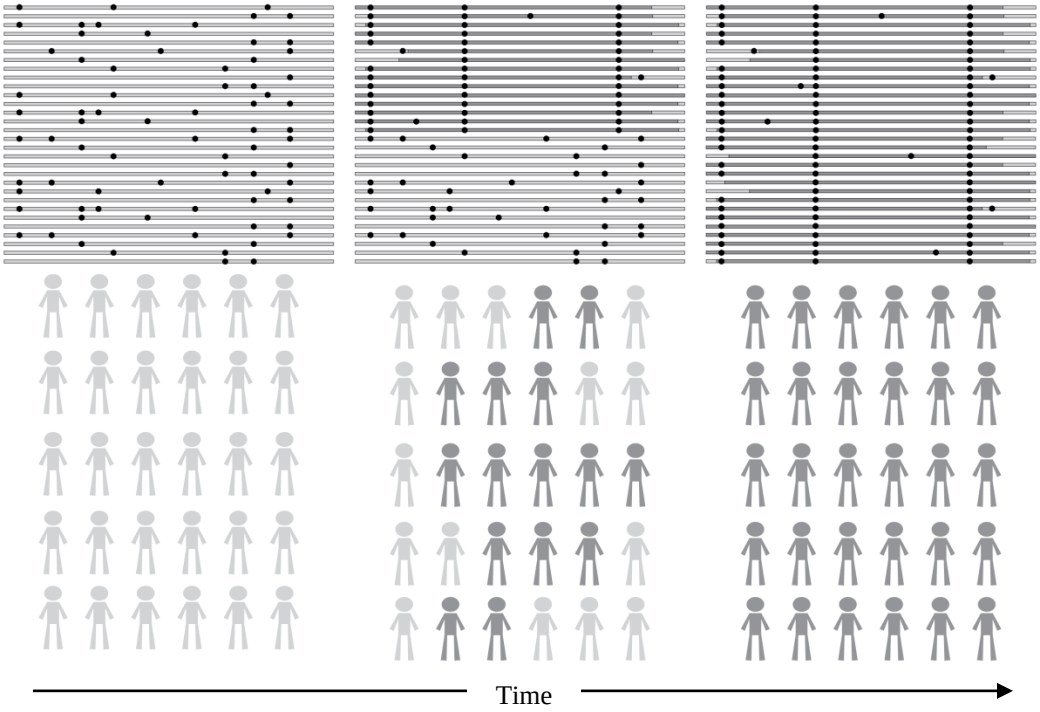


Figure 1.2: Illustration of a selective sweep. Over time, a selected variant and its surrounding linked regions will increase rapidly in a population, leaving a strong signal of haplotype homozygosity, the signature of a selective sweep. A selective sweep may be identified in the incomplete stage (middle pane) using iHS (Voight et al., 2006) and nearly fixed or fixed selection events using the cross-population comparison XP-EHH (Sabeti et al., 2007). These tests identify regions of the genome that harbor the advantageous variant targeted by selection.

that results from a rapid increase in the frequency of selected variant(s) and their neighboring polymorphisms (Figure 2).

The development of these analytical methods has shed new light on adaptation in Tibetans and the field of high-altitude research. First, our studies and many others (Bigham et al., 2010; Peng et al., 2011; Wang et al., 2011; Xu et al., 2011; Yi et al., 2010) indicate that many genes are involved in this process, suggesting that various genetic factors act in concert to provide an advantage to Tibetans. Second, the results of the research presented here and other recent studies suggest that many selection targets are similar in various Tibetan groups while others appear to be heterogeneous among populations (Chapter 5). Therefore, the idea that many genetic factors are involved in high-altitude adaptation is supported by several recently published reports of the same selection candidate genes despite different Tibetan samples and analytical strategies.

High-altitude selection candidate genes are associated with hypoxia-related phenotypes

While decades of research efforts have focused on physiological differences among native high altitude populations (Beall, 2007), evidence for an underlying genetic basis was lacking until this past year. To determine whether the selection candidate genes identified were associated with trait variation in Tibetan highlanders, we compared selected haplotypes to hematological and metabolic phenotypes. We found that the selection candidate regions containing *EGLN1*/*PHD2* and *PPARA* genes are associated with hemoglobin (Hb) concentration in a population from the Tibetan village of Maduo (4250 meters), and selection candidates *EPAS1*/*HIF2A* and *PPARA* are associated with

lactate and free fatty acid levels, respectively, in individuals from the Tuo Tuo River Tibetan village (4350 meters) (Figure 1.3).

In order to determine whether we could detect the same selection candidate regions and association signals in a population located in a different region of the Tibetan Plateau, we set out to collect samples from a second high-altitude population. While several of the same selection signals were detected, none of the selection candidate genes were associated with hemoglobin levels in the second group. We did, however, observe a nonsignificant trend between the *EPAS1*/*HIF2A* selection candidate gene haplotype and Hb level in our second population. Recent work from other groups identified an association between this particular selection candidate and hemoglobin level (Chapter 5) (Beall et al., 2010; Yi et al., 2010), suggesting similarity between these Tibetan groups and the Tuo Tuo River sample we studied (Figure 1.3). Furthermore, associations detected between selection candidates and metabolite levels in the second population are not observed in the Maduo population, although the sample size was limited for this comparison (n = 20) (Chapter 3). These results suggest there may be differences among Tibetan groups and motivate efforts to further characterize the genetic and phenotypic differences in various Tibetan groups (Chapter 5).

Despite limited sample sizes used for our two population studies, the signals are apparent. While there are various statistics that may be used to detect signals of positive selection on a genomewide scale (Akey, 2009), we chose to use the two most powerful methods to identify incomplete and complete selective sweeps for genomewide SNP data in the absence of a calibrated model of demographic history. It is unclear, however,

Figure 1.3: Sample locations for two Tibetan populations examined. The Maduo sample (~4,250 meters), collected in 2009, are Amdo Tibetan, and the Tuo Tuo River sample (~4,500 meters), collected in 2010, are Kham Tibetan.

if larger samples would afford more variation to test for genotype-phenotype relationships. Regardless, we have identified striking signals between selection candidate regions different phenotypic traits in these populations.

Population variation within Tibet

While non-overlapping selection and association signals observed in Tibetan population could reflect differences in analytic strategies and statistical power, they could also indicate that different loci have been selected in different high-altitude populations. To test whether we would observe differences among selection signals and phenotype associations using the same analytical strategies, we collected and performed the same analyses in two Tibetan populations located in different regions of the Qinghai-Tibetan Plateau.

The two populations studied are geographically distinct from each other and speak different dialects (Amdo Tibetan in the village Maduo and Kham Tibetan in the Tuo Tuo River location; Figure 1.3). We determined that many of the same loci are under selection in these distinct Tibetan populations and determined that some signals are unique, suggesting that these populations have either acquired different mechanisms to adapt to high altitude, exhibit signals presently evident in one population but potentially diminished/non-existent in the other (possibly due to timing of the selection event and consequent breakdown of linkage disequilibrium), or false positive signals.

Despite the unknown reasons for differences at this time, we were able to conclude from this work that many selection signals are similar among Tibetan populations from different regions of the Tibetan Plateau. The additional support provided by studies conducted by other research groups suggests that these selection

targets could have important relationships with as of yet unmeasured phenotypes vital to survival in populations throughout the Qinghai-Tibetan Plateau. These selection candidate regions will be prioritized in future efforts to characterize precise genetic variants and their biological roles in adaptation in highland populations (Chapter 5).

It is clear that evolutionary processes have affected specific hypoxia- and metabolism-related genetic factors in high-altitude Tibetans. The associations between several of the selected candidate genes and a decreased hemoglobin phenotype provide supporting evidence for functional roles of genetic variants specific to high-altitude Tibetans. While these studies provide a step forward in the field of high-altitude research, they also raise many questions regarding the mechanisms involved in human adaptation to hypoxia. In order to understand this process, it will be necessary to integrate these recent genetic advances with current knowledge of relevant molecular pathways and the many physiological insights afforded by decades of high-altitude research. Once the precise genetic variants are identified, functional analysis is necessary to elucidate the precise mechanisms underlying hypoxia tolerance.

Future studies will also help to elucidate whether some variants within these genes contribute to maladaptive responses that result from limited oxygen supply. These consequences extend beyond altitude sickness and are relevant to clinical pathologies such as pulmonary hypertension, heart disease, stroke, tumor hypoxia, and metabolic disorders (Majmundar et al., 2010). Further studies of high-altitude adaptation may allow interventions to reduce common, hypoxia-induced maladies suffered by lowlanders through identification of potential new targets for more common forms of clinical pathologies.

In addition to physiological and molecular aspects of high-altitude adaptation, genetic studies of native high-altitude populations will provide valuable insights into evolutionary processes. It remains to be seen how natural selection has orchestrated human biological changes and how much of this fine-tuning is based upon novel or standing genetic variation. It will also be important to determine if and how the selection candidates, some of which have been reported in more than one study (Table 5.1), are involved in high-altitude adaptation and whether differences reported in selection and association signals reflect variation in Tibetan population histories.

References

Akey, J.M. (2009). Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Research *19*, 711-722.

Beall, C.M. (2000). Tibetan and Andean patterns of adaptation to high-altitude hypoxia. Human Biology *72*, 201-228

Beall, C.M. (2007a). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences 104, 8655-8660.

Beall, C.M., Brittenham, G.M., Strohl, K.P., Blangero, J., Williams-Blangero, S., Goldstein, M.C., Decker, M.J., Vargas, E., Villena, M., Soria, R., *et al.* (1998). Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. American Journal of Physical Anthropology *106*, 385-400.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., *et al.* (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proceedings of the National Academy of Sciences *107*, 11459-11464.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R., Scherer, S.W., Julian, C.G., Wilson, M.J., López Herráez, D., *et al.* (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics *6*, e1001116.

Bigham, A.W., X. Mao, T. Brutsaert, M. J. Wilson, C. G. Julian, E. J. Parra, J. M. Akey, L. G. Moore, M. D. Shriver (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Human Genomics *4*, 79-90.

Formenti, F., Constantin-Teodosiu, D., Emmanuel, Y., Cheeseman, J., Dorrington, K.L., Edwards, L.M., Humphreys, S.M., Lappin, T.R.J., McMullin, M.F., McNamara, C.J., *et al.* (2010). Regulation of human metabolism by hypoxia-inducible factor. Proceedings of the National Academy of Sciences *107*, 12722-12727.

Hackett, P.H., and Roach, R.C. (2001). High-altitude illness. New England Journal of Medicine *345*, 107-114.

Hirota, K., and Semenza, G.L. (2006). Regulation of angiogenesis by hypoxia-inducible factor 1. Critical Reviews in Oncology/Hematology *59*, 15-26.

Kaelin, W.G.J., and Ratcliffe, P.J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Molecular Cell *30*, 393-402.

MacInnis, M.J., Koehle, M.S., and Rupert, J.L. (2010). Evidence for a genetic basis for altitude illness: 2010 update. High Altitude Medicine & Biology *11*, 349-368.

Majmundar, A.J., Wong, W.J., and Simon, M.C. (2010). Hypoxia-Inducible Factors and the Response to Hypoxic Stress. Molecular Cell *40*, 294-309.

Monge, C.C.W., J. (1976). Chronic mountain sickness. Johns Hopkins Medical Journal, 139 SUPPL:187-139.

Moore, L.G. (2001). Human genetic adaptation to high altitude. High Altitude Medicine & Biology *2*, 257-279.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, *et al.* (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular Biology and Evolution *28*, 1075-1081.

Rupert, J. (2010). Will blood tell? Three recent articles demonstrate genetic selection in Tibetans. High Altitude Medicine & Biology *11*, 307-308.

Rupert, J.L., and Koehle, M.S. (2006). Evidence for a genetic basis for altitude-related illness. High Altitude Medicine & Biology *7*, 150-167.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., *et al.* (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913-918.

Semenza, G.L. (2009). Involvement of oxygen-sensing pathways in physiologic and pathologic erythropoiesis. Blood *114*, 2015-2019.

Semenza, G.L. (2010). Vascular responses to hypoxia and ischemia. Arteriosclerosis, Thrombosis, and Vascular Biology *30*, 648-652.

Stinchcombe, J.R., and Hoekstra, H.E. (2007). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity *100*, 158-170.

Vargas, P.E., and Spielvogel, H. (2006). Chronic mountain sickness, optimal hemoglobin, and heart disease. High Altitude Medicine & Biology *7*, 138-149.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biology *4*, e72.

Wang, B., Zhang, Y.-B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L., Li, X., *et al.* (2011). On the origin of Tibetans and their genetic basis in adapting to high-altitude environments. PLoS One *6*, e17002.

Wu, T., Wang, X., Wei, C., Cheng, H., Wang, X., Li, Y., Ge, D., Zhao, H., Young, P., Li, G., *et al.* (2005). Hemoglobin levels in Qinghai-Tibet: different effects of gender for Tibetans vs. Han. Journal of Applied Physiology *98*, 598-604.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., *et al.* (2011). A genome-wide search for signals of high altitude adaptation in Tibetans. Molecular Biology and Evolution.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., *et al.* (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75-78.

Yoon, D., Pastore, Y.D., Divoky, V., Liu, E., Mlodnicka, A.E., Rainey, K., Ponka, P., Semenza, G.L., Schumacher, A., and Prchal, J.T. (2006). Hypoxia-inducible Factor-1 deficiency results in dysregulated erythropoiesis signaling and iron homeostasis in mouse development. Journal of Biological Chemistry *281*, 25703-25711.

CHAPTER 2

GENETIC EVIDENCE FOR HIGH-ALTITUDE ADAPTATION IN TIBET

Abstract

Tibetans have lived at very high altitudes for thousands of years, and they have a distinctive suite of physiological traits that enable them to tolerate environmental hypoxia. These phenotypes are clearly the result of adaptation to this environment, but their genetic basis remains unknown. We report genome-wide scans that reveal positive selection in several regions that contain genes whose products are likely involved in high-altitude adaptation. Positively selected haplotypes of *EGLN1* and *PPARA* were significantly associated with the decreased hemoglobin phenotype that is unique to this highland population. Identification of these genes provides support for previously hypothesized mechanisms of high-altitude adaptation and illuminates the complexity of hypoxia response pathways in humans.

Introduction

The Tibetan highlands are one of the most extreme environments inhabited by humans. Many present-day Tibetan populations are thought to be descendants of people who have occupied the Tibetan Plateau since the mid-Holocene, between 7,000 and 5,000 years ago (Su et al., 2000), and possibly since the late Pleistocene, approximately 21,000 years ago (Moore, 2001; Zhao et al., 2009). Compared to Andean populations living in similar high-altitude conditions, Tibetans exhibit a distinct suite of physiologic traits: decreased arterial oxygen content, increased resting ventilation, lack of hypoxic pulmonary vasoconstriction, lower incidence of reduced birth weight, and reduced hemoglobin (Hb) concentration (on average, 3.6g/dl less for both males and females) (Beall, 2007; Chen, 1997; Ge, 1994; Ge et al., 2002; Groves et al., 1993). Neighboring Han Chinese individuals and other nonadapted lowland visitors to high-altitude regions

develop increased Hb concentration to compensate for the hypoxic high-altitude environment (Wu et al., 2005), and this response is associated with adverse effects (Mejia et al., 2005; Vargas, 2006).

High-altitude Tibetans maintain normal aerobic metabolism despite profound arterial hypoxia (Beall, 2007), perhaps through the existence of changes in the oxygen transport system. For example, elevated circulating nitric oxide (NO) levels increase vasodilation and blood flow (Beall et al., 2001), which, when combined with increased ventilation (Zhuang et al., 1993), may increase the availability of oxygen to cells (Beall, 2007). Collectively, these traits suggest strongly that Tibetans have adapted uniquely to extreme high-altitude conditions. The genetic basis of this adaptation, however, remains unknown.

Materials and methods

DNA sample collection

DNA was extracted from whole blood samples for 49 individuals (non-smokers, no chronic diseases) residing in Madou County in Qinghai province (~4,350 m). Informed consent was obtained for all participants according to guidelines approved by the Institutional Review Boards at the High Altitude Medical Research Institute (Xining, Qinghai, People's Republic of China).

SNP genotyping

Forty-nine individual DNA samples were genotyped using Affymetrix 6.0 SNP Array technology (>900,000 SNPs) at Capital Bio Corporation (Beijing, China). We used default parameters for the Birdseed algorithm (version 2) to determine genotypes for

all samples (Affymetrix, Santa Clara, CA, USA). Genotypic data were analyzed using the Affymetrix Genotyping Console 3.1 (Affymetrix) and included all autosomes but excluded the X and Y chromosomes and mitochondrial genome.

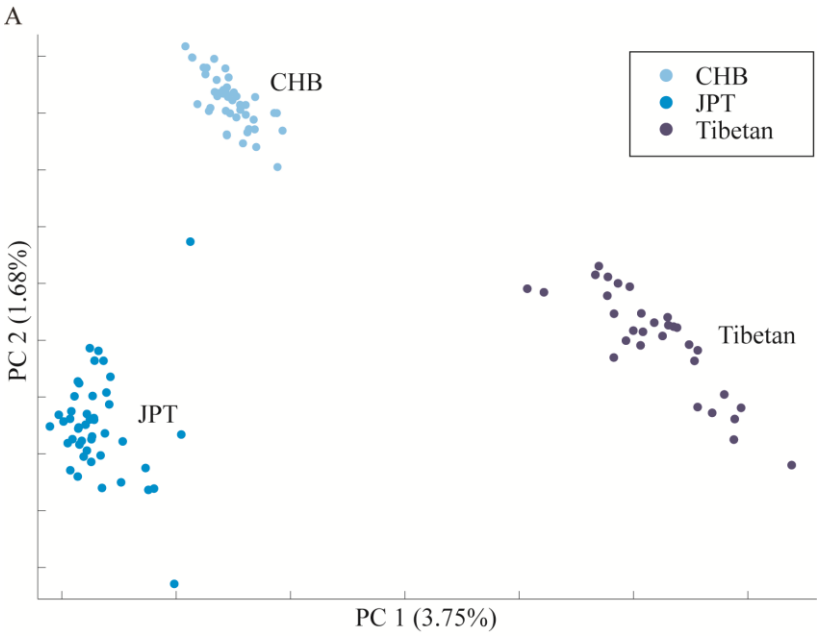
Principal components analysis

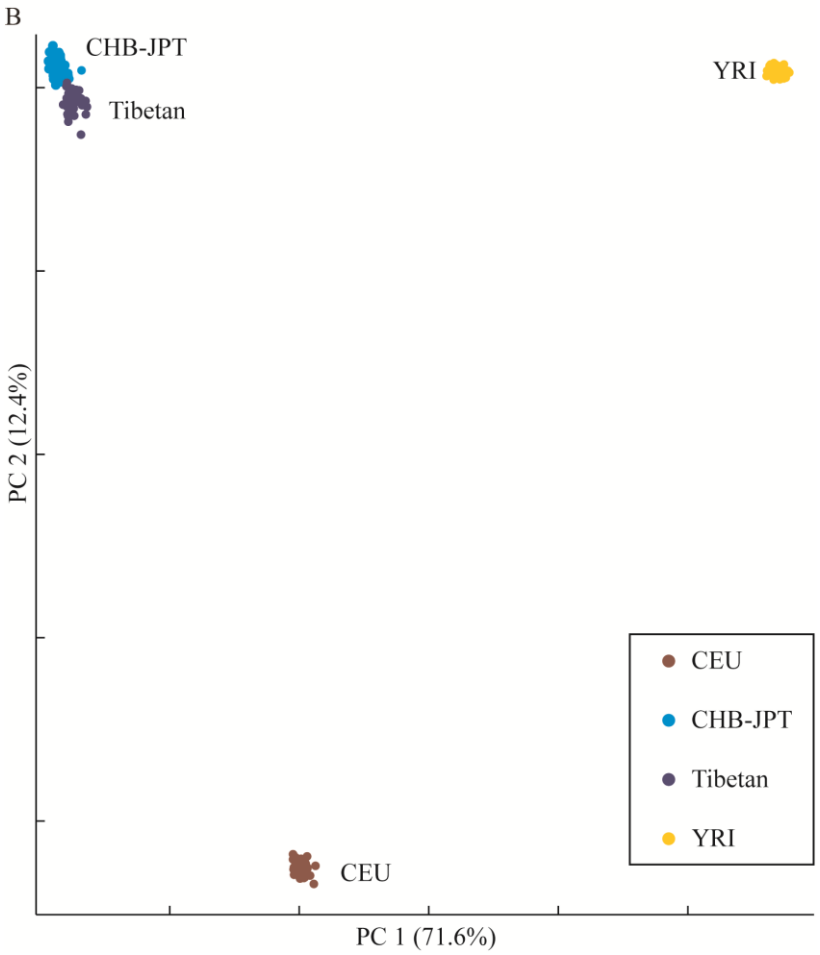
We performed principal components analysis (Fig. 2.1) based on genetic distances as previously described (Xing et al., 2009). This analysis indicates that Tibetans form a distinct group but are genetically similar to the combined CHB-JPT population (CHB = Chinese in Beijing, China; JPT = Japanese in Tokyo, Japan), which serves as an appropriate lowland East Asian comparison population (panel A of Fig. 2.1). The CEU (U.S. Utah residents with ancestry from northern and western Europe) and YRI (YRI = Yoruba in Ibadan, Nigeria) HapMap populations provide context for the patterns of variation observed among these populations (The International HapMap, 2005).

Estimates of relatedness

We collected samples at a small town health center where groups of semi-nomadic clans visit. Prior to genotyping, we excluded first-degree relatives who visited the clinic. It is possible that related individuals were examined at different times throughout the collection process; therefore, relatedness could only be determined after genotype analysis. We used pair-wise genetic distances and the proportion of shared genomic segments to determine relatedness between subjects (Gusev et al., 2009; Xing et al., 2009). When pairs of individuals exhibited genetic distances less than 4.95×10^{-2} or had genomewide identity-by-descent of greater than 400cM (minimum segment size

Figure 2.1: Principal Components Analysis (PCA) of Tibetans and the HapMap populations based on genome-wide SNP data. CHB = Chinese in Beijing, China; JPT = Japanese in Tokyo, Japan; YRI = Yoruba in Ibadan, Nigeria; CEU (U.S. Utah residents with ancestry from northern and western Europe) (10). (**A**) PCA of Tibetans and East Asian populations (CHB and JPT). (**B**) PCA of Tibetans and HapMap populations. The Tibetans are genetically most similar to (but still distinct from) the CHB-JPT populations; therefore, the CHB-JPT population comparison for XP-EHH analysis is appropriate for identifying regions of the genome subject to positive selection in Tibetans.


A
CHB
JPT
Tibetan
CHB
JPT
Tibetan
PC 2 (1.68%)
PC 1 (3.75%)


B
CHB-JPT
Tibetan
YRI
CEU
CEU
CHB-JPT
Tibetan
YRI
PC 2 (12.4%)
PC 1 (71.6%)

2.5cM), one member of the pair was excluded from the analyses. Based on these criteria, a total of 31 unrelated individuals were included in the analyses.

A priori functional candidate list

We generated a list of genes likely related to high-altitude adaptation based on categories provided in Table 2.1. We coupled genes associated with Gene Ontology (GO) categories (Gene Ontology, 2006) that may be involved in the observed high-altitude Tibetan phenotypes (481 genes), with genes listed in the "Hypoxia response via HIF activation" defined by Panther Pathways (33 genes) (Mi et al., 2007). Potential candidate genes identified in the mitochondrial genome and on the X chromosome were not considered for this study.

Although the intersection of functional and selection candidate lists is enriched for hypoxia-related signals of selection, the ten genes we identified probably do not account for all high-altitude adaptive traits in this population. For example, the genomic region containing *HIF1AN*, an inhibitor of HIF in normoxic conditions, was significant in both the XP-EHH and iHS tests, although it was not included in our *a priori* list of functional candidates (Fig. 2.2; Table 2.2). Other genes in our functional candidate list are found within regions identified in the top 2% of the selection scans, such as the human ß-globin gene cluster (Table 2.2). These post hoc findings reflect the conservative approach used to define our list of genes for high-altitude adaptation in Tibetans.

Admixture analysis

A model-based algorithm implemented in *ADMIXTURE* (Alexander et al., 2009) was used to determine the genetic ancestries of each individual in a given number of

Table 2.1

Gene pathways and ontology categories used to define an *a priori* functional candidate gene list

Source	**Accession**	**N**
Gene Ontology (GO) Categories		
Detection of oxygen	GO:0003032	14
Nitric oxide metabolic process	GO:0046209	37
Oxygen sensor activity	GO:0019826	2
Oxygen binding	GO:0019825	46
Oxygen transport	GO:0015671	29
Oxygen transporter activity	GO:0005344	14
Response to hypoxia	GO: 0001666	141
Response to oxygen levels	GO:0070482	157
Vasodilation	GO:0042311	41
Panther Pathway		
Hypoxia response via HIF activation	P00030	33
Total number of unique genes:		**247**

Each ontology category and the number of respective genes are provided above. In order to enrich for hypoxia-related genes within the top selection candidate regions of the genome, we limited our focus to these *a priori* functional candidate genes.

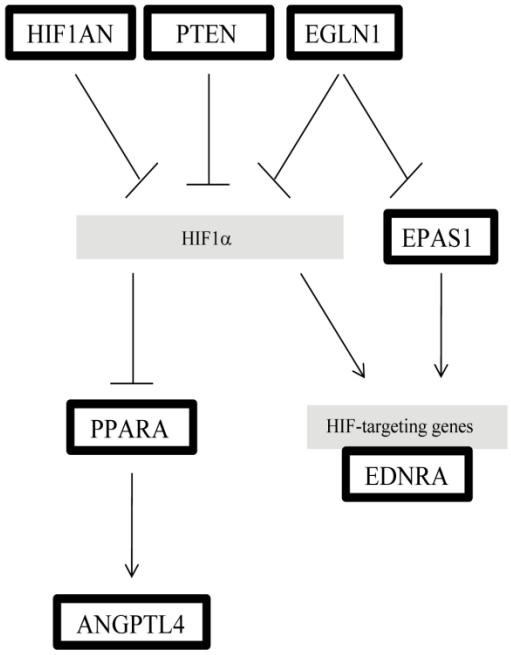


Figure 2.2: Selection candidates involved in the hypoxia-inducible factor (HIF) pathway. Genes identified as selection candidates that are related to the HIF pathway (outlined in black) are illustrated below (gene descriptions and regulation during hypoxic/normoxic conditions are provided in Table 2.3; note that *HIF1AN* was not included on the *a priori* functional candidate list but was identified by both XP-EHH and iHS analyses; Table 2.2). Genes indicated in the grey boxes are provided for reference.

Table 2.2

Genes identified in the overlap of selection candidate and functional candidate gene lists

Gene identified in 200kb region	XP-EHH score	XP-EHH empirical uncorrected p-value	iHS score	iHS empirical uncorrected p-value
EPAS1 Endothelial PAS domain protein 1	0.82	0.002	-	-
CYP2E1 Cytochrome P450, family 2, subfamily E, polypeptide 1	0.71	0.007	-	-
EDNRA Endothelin receptor type A	0.7	0.008	-	-
ANGPTL4 Angiopoietin-like 4	0.69	0.008	-	-
CAMK2D Calcium/calmodulin-dependent protein kinase II delta	0.68	0.009	-	-
EGLN1 Egl nine homolog 1	0.96†, 0.86†*	0.0002†, 0.001†*	2.68*	0.01*, -
HMOX2 Heme oxygenase (decycling) 2	-	-	3	0.001
CYP17A1 Cytochrome P450, family 17, subfamily A, polypeptide 1	-	-	4	0.007
PPARA Peroxisome proliferator-activated receptor alpha	-	-	3.58	0.009
PTEN Phosphatase and tensin homolog			3.9	0.007

populations without using information about population designation. To eliminate the effects of SNPs that are in linkage disequilibrium (LD), we first filtered out SNPs that had $r^2 > 0.2$ within 100kb using PLINK (Purcell et al., 2007), as recommended by the authors of *ADMIXTURE*. The pruned data set contains 142,888 SNPs.

While the demographic history of the Tibetan Plateau is unclear (i.e., whether modern Tibetans descended from populations who occupied this region during the mid-Holocene, the Late Pleistocene, or if they are an admixed population (Zhao et al., 2009)), this analysis indicates a distinct relationship between Tibetans and East Asian (CHB-JPT) populations. We see no strong evidence of admixture in our samples (Fig. 2.3), although signals of selection should be detectable even if Tibetans were admixed or descended from populations who occupied this region prior to or during the mid-Holocene.

Selection analyses

We used the Beagle software package to estimate phase in the 31 unrelated Tibetan individuals (Browning and Browning, 2009) and calculated all selection statistics from the phased data. To calculate iHH for each allele at each site, we integrated the expected EHH in both directions from the core SNP until expected EHH was less than or equal to 0.10 (Huff et al., 2010; Voight et al., 2006). To calculate iHS, we calculated the log of the ratio of iHH scores at each site for the derived and ancestral alleles, standardizing within each population by the derived allele frequency. We computed iHS scores in this manner for all SNPs on the Affymetrix 6.0 microarray with at least10 copies of the derived allele and the ancestral allele in a given population. For the iHS selection scan, our test statistic for each 200kb genomic region was the fraction of SNPs

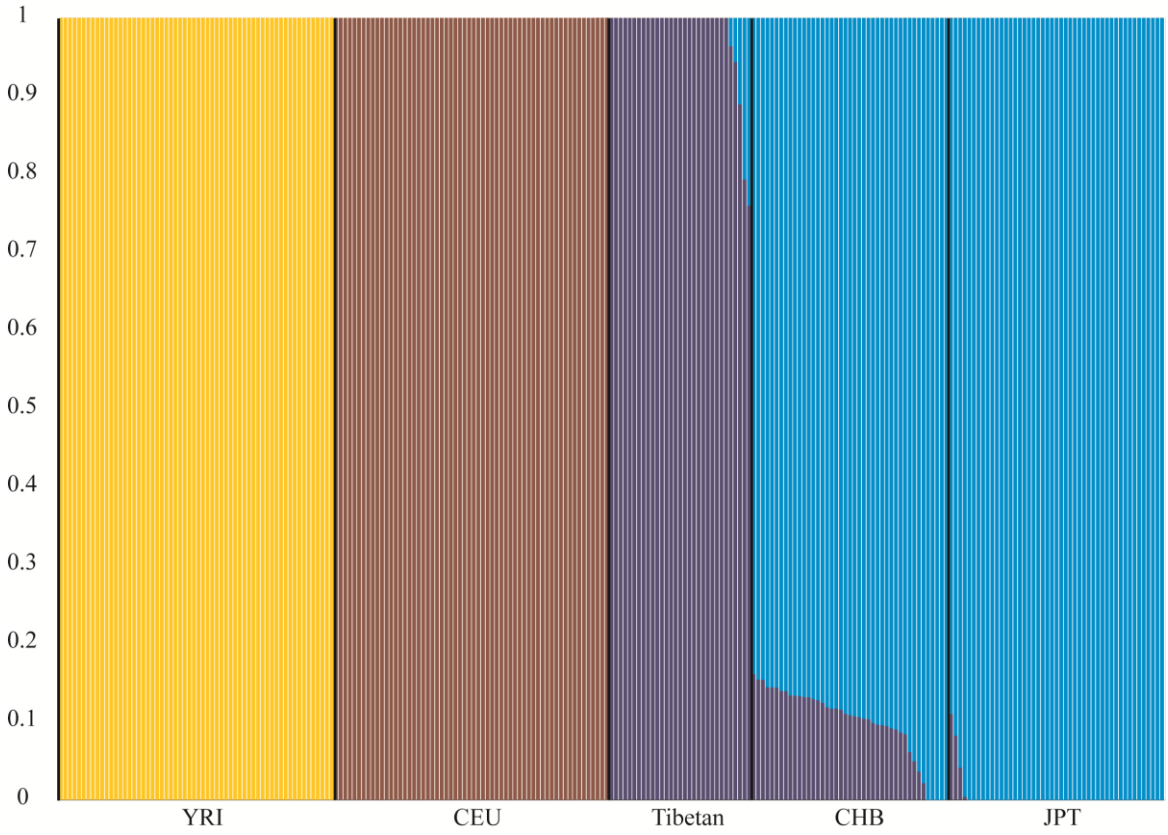


Figure 2.3: Individual grouping inferred by *ADMIXTURE* at K = 4. Each individual's genome is represented by a vertical bar composed of colored sections, where each section represents the proportion of an individual's ancestry derived from one of the K ancestral populations. Individuals are arrayed horizontally and grouped by population as indicated.

Table 2.3

Description of genes identified in the overlap of selection candidate and functional candidate gene lists

Gene	Description	Category
***EPAS1*(*HIF2A*)**	HIF-family transcription factor; up-regulated under hypoxic conditions, regulates vascular endothelial growth factor expression, erythropoietin expression in the brain and liver	GO: Response to hypoxia, Response to hypoxia levels
***CYP2E1*(*CPE1*)**	Catalyzes many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids	GO:Oxygen binding
EDNRA (ET-A)	Encodes a cell surface receptor for endothelin-1	GO: Response to hypoxia, Response to hypoxia levels
ANGPTL4(FIAF)	Directly involved in regulating glucose homeostasis, lipid metabolism, and insulin sensitivity and also acts as an apoptosis survival factor for vascular endothelial cells	GO: Response to hypoxia, Response to hypoxia levels
***CAMK2D*(*CAMKD*)**	Serine/threonine protein kinase family and to the Ca(2+)/calmodulin-dependent protein kinase subfamily; mediates nitric oxide production in response to changes in intracellular calcium	GO: Response to hypoxia, Response to oxygen levels
EGLN1(PHD2)	Catalyzes post-translational hydroxylation of the two HIF alpha proteins (HIF1a and HIF2a), targeting them for proteasomal degradation in normoxic conditions	Panther: Hypoxia response via HIF activation; GO: Response to hypoxia, Response to oxygen levels
***HMOX2* (*HO-2*)**	Involved in oxygen sensing independent of the HIF pathway; enhanced expression preserves endothelial cell viability during hypoxia; the NmrA-like family domain containing 1 (*NMRAL1*) gene, which encodes a protein involved in NO synthesis, is located within the region containing *HMOX2*	GO: Response to hypoxia, Response to oxygen levels
***CYP17A1*(*CPT7*)**	Mono-oxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids	GO: Oxygen binding
***PPARA* (*NR1C1*)**	Nuclear hormone-binding protein transcriptional regulator that controls the peroxisomal beta-oxidation pathway of fatty acids	GO: Response to hypoxia, Response to oxygen levels
***PTEN* (*TEP1*)**	Lipid phosphatase, tumor suppressor that antagonizes the PI3K-AKT/PKB signaling pathway, thereby modulating cell cycle progression and cell survival. Loss of *PTEN* increases HIF activity	Panther: Hypoxia response via HIF activation

(Mi et al., 2007; Gene Ontology Consortium, 2006)

in each region where |iHS|>2.0, excluding regions with fewer than 5 SNPs (Huff et al., 2010; Voight et al., 2006). We calculated XP-EHH at each site using the default settings of the XP-EHH software (http://hgdp.uchicago.edu/Software/). For the XP-EHH selection scan, our test statistic was the maximum XP-EHH score in each 200kb region (Sabeti et al., 2007).

We determined statistical significance for each 200kb region from the empirical distribution of each test statistic. Our selection candidates are those genes contained in any of the 200kb regions significant at the 0.01 level in either test, excluding regions where the iHS test was significant at the 0.01 level in neighboring populations from Mongolia (n = 25), India (n = 25), Nepal (n = 25), China and Japan (CHB-JPT: n = 90), Kyrgyzstan (n = 25), and Thailand (n = 25) (unpublished data for samples provided by Scott Woodward and the Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah, USA). The goal of this exclusion step was to enrich for signals of local adaptation in the Tibetan population by filtering out signals of selection present in other Asian populations. Our exclusion criteria only included significant iHS results because the comparison population (CHB-JPT) in the XP-EHH test directly controls for genomic variation in a neighboring population. (Appendices A.1 and A.2 contain all 200kb genomic regions identified in the top 2% of each selection scan.)

Analyses for localization of selection signal

Although the composite of multiple tests (CMS) statistic is not applicable for localization of our selection signals (due to a lack of detailed information about Tibetan demographic history), we have conducted analyses for the three additional statistics

reported by Grossman et al.: F_{ST}, DDAF, and DiHH (Grossman et al., 2010). If demographic information were available, these statistics could be combined into a single test, but we present the separate results of each test for the ten gene regions described.

Phenotype collection

Hemoglobin concentration, hematocrit, and percent oxygen saturation were determined from venous blood samples using the Mindray Hematology Analyzer (BC-2300, Shenzhen, People's Republic of China) and the Pulse Oximeter (Ohmeda 3700 Pulse Oximeter, Datex-Ohmeda, Boulder, Colorado, USA), respectively. Hematocrit values are highly correlated with [Hb] ($r = 0.861$, $p < 10^{-9}$). See Table 2.4 for phenotype measurements.

Genotype-phenotype association

For the five iHS selection candidates that intersect our functional candidate list (Table 2.2), we identified the putatively advantageous haplotypes as those carrying the SNP alleles responsible for the most extreme iHS scores within the corresponding 200 kb genomic region. Ideally, we would test for a direct correlation between the advantageous genetic variants and Hb concentration. However, our selection scan results provide only indirect inferences about SNPs that are linked to the putatively advantageous variant. Because the sign of an iHS score indicates an excess of homozygosity around the ancestral (+) or derived (-) allele, the allele designated by an extreme iHS score is frequently linked closely to the advantageous allele during a selective sweep. Therefore, we selected the three alleles exhibiting the most extreme iHS scores within each 200kb

Table 2.4

Phenotype and core haplotype data for 30 Maduo Tibetan individuals

					Number of selection candidate haplotypes per individual				
Age	Gender	Hemoglobin (g/dL)	Hematocrit (%)	Oxygen Saturation (%)	*EGLN1*	*PPARA*	*HMOX2*	*CYP17A1*	*PTEN*
22	F	14.4	40.1	91	0	2	2	0	1
62	F	19.4	57.5	86	1	1	1	1	2
31	M	10.1	32.7	88	2	2	2	0	2
56	F	13.2	38.3	89	2	1	2	1	1
56	M	16.0	44.3	87	1	2	2	1	0
36	F	13.8	38.4	92	2	2	2	0	1
66	F	16.1	45.5	85	2	2	2	0	1
45	M	17.8	69.3	86	0	2	2	1	0
56	F	14.9	44.4	83	2	2	2	1	1
29	M	18.7	53.2	91	0	1	2	0	0
34	M	19.4	53.5	85	1	1	1	0	1
32	M	22.3	64.3	90	1	1	2	0	1
36	F	14.9	43.4	91	0	2	2	2	2
44	M	15.9	44.9	84	2	2	2	1	1
23	F	12.7	37.3	88	2	1	1	0	0
60	F	14.6	41.9	84	2	1	1	1	1
33	F	15.5	44.7	86	1	2	2	2	1
23	F	15.6	43.4	86	1	2	1	1	2
17	F	14.7	42.2	90	1	2	2	0	0
62	M	15.5	42.8	87	1	2	1	0	0
38	M	17.7	50.5	81	2	0	1	2	1
29	M	16.4	49.4	84	2	2	1	0	1
46	F	17.9	51.4	83	0	2	1	1	2
44	M	16.0	45.2	86	1	2	2	0	1
40	F	16.9	47	87	2	1	2	1	1
36	F	16.4	47	85	1	2	1	0	0
40	F	17.0	49	-	1	2	2	1	2
53	F	19.5	56.6	79	2	0	2	2	0
30	F	13.4	38.4	89	2	2	2	0	1
27	F	12.9	38.4	85	2	2	2	1	2
41	F	16.9	43.1	85	2	1	2	0	2

Table 2.5

SNPs used to define the selection candidate haplotypes for iHS genotype-phenotype analysis

	First Allele of Core Selection Haplotype				Second Allele of Core Selection Haplotype				Third Allele of Core Selected Haplotype			
Gene in 200kb region	HG 18 Position*	iHS Score	Selected Allele	Alternate Allele	HG 18 Position*	iHS Score	Selected Allele	Alternate Allele	HG 18 Position*	iHS Score	Selected Allele	Alternate Allele
EGLN1	Chr1: 229793717	2.68	A	T	Chr1: 229667980	2.45	T	C	Chr1: 229665156	2.34	T	C
CYP17A1	Chr10: 104568521	4.00	G	C	Chr10: 104594906	-3.54	G	T	Chr10: 104517420	3.49	C	G
PTEN	Chr10: 89770364	3.90	G	C	Chr10: 89790851	2.72	C	T	Chr10: 89778618	2.68	C	T
HMOX2	Chr16: 4456093	3.00	C	T	Chr16: 4465266	2.87	T	C	Chr16: 4442515	2.84	T	C
PPARA	Chr22: 44827140	3.58	A	G	Chr22: 44832376	-2.72	C	A	Chr22: 44842095	-2.55	T	C

genomic region to construct haplotypes that partially tag the putatively advantageous variants. We are able to test for an association between the putatively advantageous haplotypes at these loci and a phenotype. Stepwise linear regression (MATLAB R2009b) was used to detect significant relationships between these genotypes and hemoglobin concentration in 30 Tibetan individuals (after excluding one tobacco smoker) (Fig. 2.4; Tables 2.5, 2.6, 2.7) and oxygen saturation in 29 individuals (after exclusion of one missing data point).

Verification of linear multivariate regression

To check the reliability of the stepwise linear multivariate regression method that we used to test for an association of genotypes with [Hb], we used the genomewide SNP data to generate an empirical p-value distribution to compare with the theoretically expected distribution. We repeated the regression analysis for the same 30 unrelated Tibetan individuals for each SNP, retaining the original values for the sex, male age, female age, and [Hb], but changing the value of the SNP genotype variable. This value was replaced with the genotype (in the corresponding individuals) at each SNP selected from the autosomes, subject to the constraints that (1) the SNP contained no missing data (no 'No Call' genotypes); (2) the minor allele frequency of each SNP was $\geq 15\%$ (i.e., the minor allele was observed at least nine times in the 60 chromosomes). 398,020 SNPs meet these criteria. The regression method tests the variable against the null hypothesis that the effect size coefficient for that variable would be zero if it were included in the regression model and reports the resulting p-values.

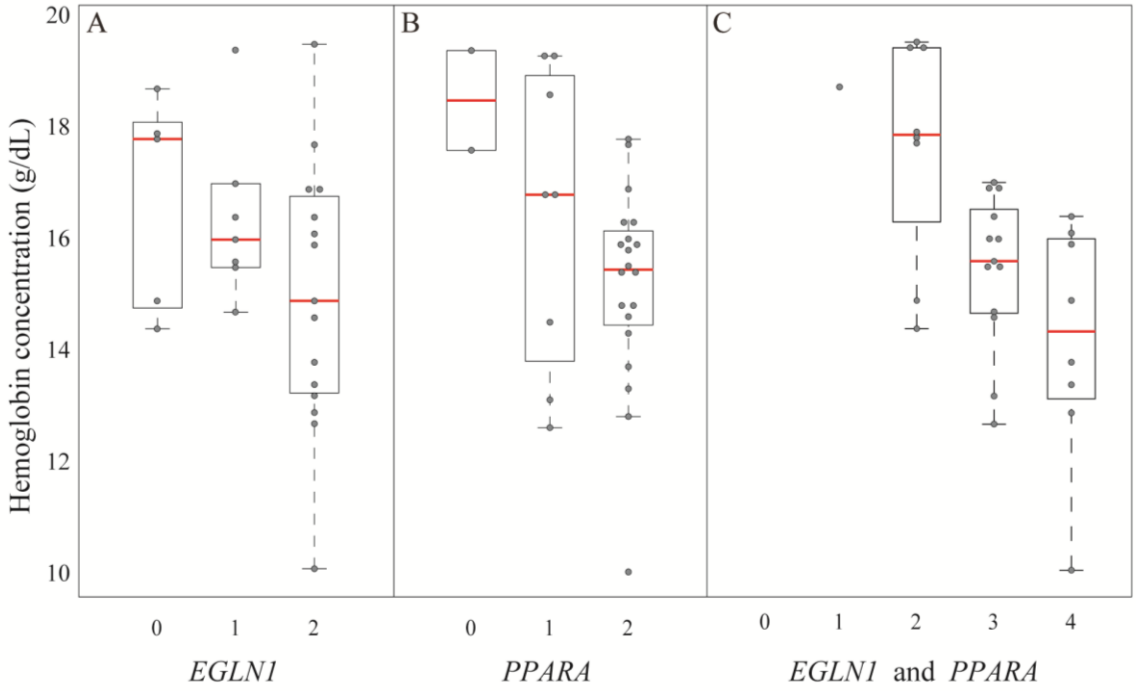


Figure 2.4: Genotype-phenotype association of the inferred adaptive *EGLN1* and *PPARA* haplotypes with hemoglobin concentration. Hemoglobin concentrations are plotted against the number of putatively advantageous haplotypes at (A) *EGLN1* and (B) *PPARA* (from 0 to 2 haplotype copies) for 30 Tibetan individuals. A box-and-whisker plot is overlaid on the data points to show the median (red line), upper and lower quartiles (box ends) and most extreme measurements (whiskers). (C) Because the effects at the two loci are similar in magnitude, direction, and significance, they are combined here for purposes of illustration. Hemoglobin concentrations are plotted against the combined number of *EGLN1* and *PPARA* haplotypes carried by an individual for both loci (from 0 to 4 copies) (Tables 2.5, 2.6, and 2.7).

Table 2.6

Regression analysis between selection candidate haplotypes and Hb phenotype

Predictor variables	**p***	**Effect size**	**Standard deviation**
Sex (0 = Male, 1 = Female)	0.44	-5.27	6.78
Male Age (0 for Females)	0.41	0.13	0.16
Female Age (0 for Males)	0.74	0.05	0.15
EGLN1	0.002	-20.06	5.39
PPARA	0.0009	-14.88	4.42
HMOX2	0.49	-4.85	6.97
CYP17A1	0.74	1.66	4.87
PTEN	0.87	-0.71	4.47

* For the regression model: $F_{(2, 27)} = 9.78$, $p < 0.0006$. Only variables with $p < 0.01$ individually were included in the final model.

Table 2.7

Regression analysis between *EGLN1* and *PPARA* selection candidate haplotypes combined and Hb phenotype

Predictor variable	**p***	**Effect size**	**Standard deviation**
Sex	0.48	-5.27	6.78
Male Age	0.47	0.11	0.16
Female Age	0.59	0.08	0.14
EGLN1+PPARA[†]	0.0002	-16.76	3.85
HMOX2	0.41	-5.67	6.83
CYP17A1	0.61	2.47	4.72
PTEN	0.82	-1.01	4.43

* For the regression model: $F_{(1, 28)} = 18.93$, $p < 0.0002$. Only variables with $p < 0.01$ individually were included in the final model.

[†] Total number of putatively selected *EGLN1* and *PPARA* alleles in an individual (0 – 4).

Results

Genes likely involved in high-altitude adaptation

We used two intersecting criteria to identify genes potentially involved in high-altitude adaptation: First, *a priori* candidates for adaptation to high-altitude hypoxia were chosen because of their known functions. Second, a genome-wide scan was conducted to identify regions that show strong evidence of local positive selection in high-altitude Tibetans (Fig. 2.5). To generate a set of *a priori* functional candidate loci, we constructed a list of Gene Ontology (GO) categories (Gene Ontology, 2006) associated with the traits discussed above (Table 1). We merged genes from this list with those in the Panther-defined pathway, "Hypoxia response via activation of hypoxia inducible factor (HIF)" (Mi et al., 2007), a major transcriptional regulator of oxygen homeostasis (Semenza, 2009) that is likely associated with high-altitude adaptation. The resulting set of 247 functional candidate loci is listed in Table 2.8.

Genes subject to recent positive selection

We next identified alleles subject to strong recent positive selection (a selective sweep) in a sample of 31 unrelated Tibetans who were genotyped for one million single nucleotide polymorphisms (SNPs) using the Affymetrix Genome-Wide Human SNP 6.0 Array. These individuals showed no evidence of admixture with Han Chinese and Japanese populations (Figure 2.3). To pinpoint loci under positive selection, we first used the cross-population extended haplotype homozygosity (XP-EHH) statistic (Sabeti et al., 2007) to make comparisons between the Tibetan highland population and the combined HapMap Chinese (CHB) and Japanese (JPT) lowland populations

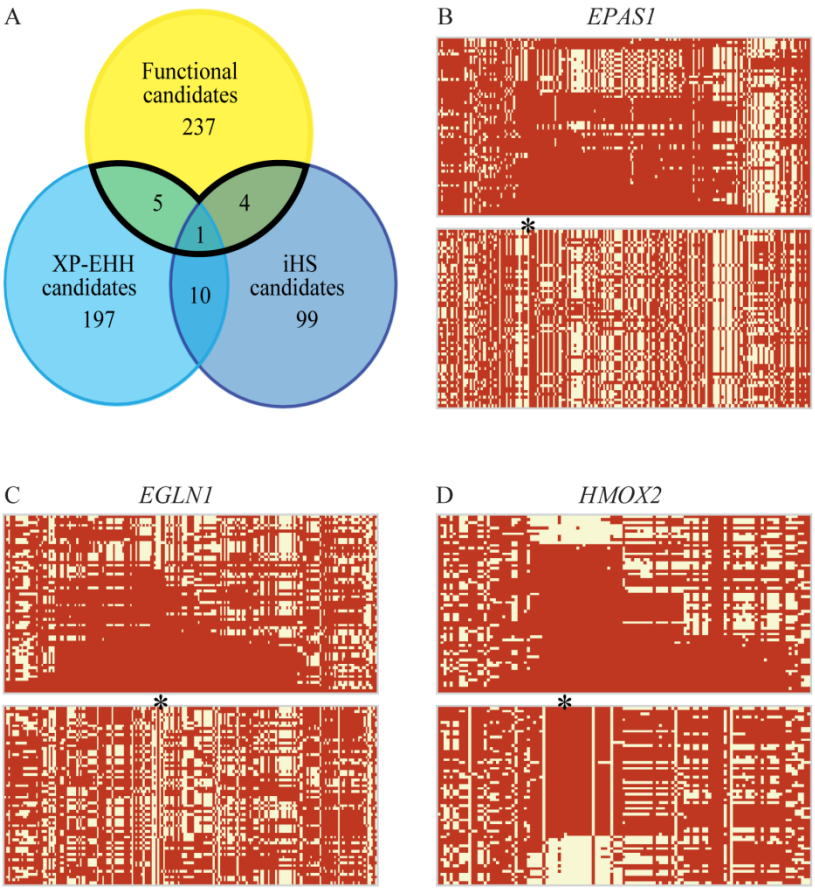


Figure 2.5: Identification of gene regions responsible for adaptation to high-altitude hypoxia in Tibetans. (A) The strategy used to identify a list of genes related to high-altitude adaptation to hypoxia relies on three sets of genes. The set of functional candidates (yellow) consists of genes associated with physiological traits related to hypoxia (see Table 2.1 for categories). The XP-EHH (light blue) and iHS (dark blue) selection candidate sets comprise genes in the top 1% of the empirical distributions of XP-EHH and iHS results, respectively, excluding those with evidence of positive selection in neighboring populations. The intersection of functional candidates with selection candidates (outlined in black) is enriched for regions containing genes that contribute to local adaptation to hypoxia in Tibetans. The genes in the intersection of functional candidates with iHS selection candidates still exhibit genetic variability in the population. (B to D), Comparison between Tibetan and CHB-JPT genomic regions identified in selection scans. The top and bottom halves of each figure represents chromosome regions in the Tibetan (number of chromosomes = 62) and CHB-JPT populations (62 randomly drawn chromosomes from 90 individuals), respectively, for the (B) *EPAS1*, (C*) EGLN1*, and (D) *HMOX2* genes identified in XPEHH, both scans, and iHS, respectively. The three SNPs with the highest iHS and XP-EHH scores (indicated by *) were designated as the core haplotype for each genomic region. All haplotypes were sorted to the horizontal midline of each panel based on the length of uninterrupted matches to the reference sequence.

Table 2.8

List of *a priori* functional candidate genes

[*] Genes within regions identified in the functional candidate list at p < 0.01 level
[†] Genes within regions identified in the functional candidate list at p < 0.02 level
[††] Genes identified in selected genomic regions of neighboring Asian populations at p < 0.01 level

ABAT	*CA9*	*CYP2E1*[*]	*GIMAP1*	*ITPR1*	*PDIA2*	*SCNN1B*	*UCN*
ACE	*CABC1*	*CYP2F1*	*GIMAP5*	*ITPR2*	*PDLIM1*	*SCNN1G*	*UCN3*
ACTN4	*CALCA*	*CYP2U1*	*GPR182*	*JAG2*	*PDPN*	*SERPINA1*	*UCP3*
ADA	*CALCB*	*CYP3A4*	*GPX1*[††]	*JAK2*	*PGF*	*SHH*	*USF1*
ADAM17[††]	*CAMK2D*[*]	*CYP3A5*	*GUCY1A3*	*KCNA5*	*PIK3C2A*	*SLC11A2*	*VEGFA*
ADIPOQ	*CAPN2*	*CYP3A7*	*HAAO*	*KCNJ8*	*PIK3C2B*	*SLC2A8*	*VHL*
ADM	*CASP1*	*CYP4A11*	*HBA1*	*KCNMA1*	*PIK3C2G*[††]	*SLC8A1*[†]	*VHLL*
ADORA1	*CAV1*	*CYP4B1*	*HBB*[†]	*KLRK1*	*PIK3C3*	*SMAD3*	*VLDLR*
ADORA2A	*CCL2*	*CYP4F12*	*HBD*[†]	*KNG1*	*PIK3CA*	*SMAD4*	*XRCC1*
ADORA2B	*CD38*	*CYP4F2*	*HBE1*[†]	*LCT*	*PIK3CB*	*SMAD9*	
ADRB1	*CDKN1A*	*CYP4F3*	*HBG2*[†]	*LONP1*	*PIK3CD*	*SOCS3*	
ADRB2	*CFTR*	*CYP8B1*	*HBM*	*MB*	*PIK3CG*	*SOD1*	
AGTR1	*CHRNA4*	*DDAH1*	*HBQ1*	*MMP14*	*PIK3R1*	*SOD2*	
AKT1	*CHRNA7*	*DDAH2*	*HBZ*	*MMP2*	*PIK3R2*	*SOD3*	
AKT2	*CHRNB2*	*DDIT4*	*HIF1A*	*MT3*	*PIK3R3*	*SPR*[††]	
AKT3	*CITED2*	*DPP4*	*HIF3A*	*NARFL*	*PIP3-E*	*STAT5B*[††]	
ALB	*CLDN3*	*ECE1*	*HMOX1*	*NF1*	*PLAT*	*TDO2*	
ALDH2	*CPS1*	*EDN1*	*HMOX2*[*]	*NGB*	*PLAU*	*TFRC*	
ALDOC	*CREBBP*	*EDNRA*	*HRH1*	*NOS1*	*PLOD1*	*TGFB1*	
ANG	*CXCR4*	*EDNRB*	*HSD11B2*	*NOS2*	*PLOD2*	*TGFB2*	
ANGPT1	*CYB5R4*	*EGFR*	*HSP90AA1*	*NOS3*	*PML*	*TGFB3*	
ANGPTL4[*]	*CYGB*	*EGLN1*[*]	*HSP90AB1*	*NOX4*[†]	*PPARA*[*]	*TGFBR1*	
APOE	*CYP17A1*[*]	*EGLN2*	*HSP90B1*	*NPPB*	*PRKAA1*	*TH*	
APOLD1	*CYP19A1*	*EGLN3*	*HYOU1*	*NPPC*	*PRKCQ*	*THBS1*[†]	
ARG2	*CYP1A1*	*ENG*	*ICAM1*	*NPR1*	*PSEN2*	*TICAM1*	
ARNT	*CYP1A2*	*EP300*	*IFNG*	*NQO1*	*PTEN*[*]	*TNF*	
ARNT2	*CYP1B1*	*EPAS1*[*]	*IL10*	*NR4A2*	*PTK2B*	*TPTE*	
ASCL2	*CYP26A1*[††]	*EPHX2*	*IL18*	*OXTR*	*PTX3*	*TPTE2*	
ATG5	*CYP2A7*	*EPO*	*IL1B*	*P2RX3*[††]	*PYGM*	*TRH*	
ATP1B1	*CYP2B6*	*ERCC3*	*INDO*	*P2RX4*	*RORA*	*TXN*	
BCL2	*CYP2C18*	*FLT1*	*INS*	*PDE5A*	*RORB*	*TXN2*	
BCL2L1	*CYP2C19*	*FRAP1*	*INSR*	*PDGFA*	*RORC*	*TXNDC2*	
BIRC2	*CYP2C8*	*GCH1*	*ITGA1*	*PDGFB*	*RYR1*	*UBE2B*	
BNIP3	*CYP2C9*	*GCHFR*	*ITGA2*	*PDGFRA*	*RYR2*	*UBQLN1*	

(Pickrell et al., 2009). The XP-EHH statistic assesses haplotype differences between two populations and is designed to detect alleles that have increased in frequency to the point of fixation or near-fixation in one of the populations (Pickrell et al., 2009; Sabeti et al., 2007). A comparison with the CHB-JPT sample is appropriate because these populations have historically lived at low altitude and exhibit a small overall genetic distance to our Tibetan samples (Figure 2.1). We also identified partial selective sweeps (in which the adaptive variant is not yet near fixation) using the integrated haplotype score (iHS), a statistic based upon the extent of decay of linkage disequilibrium surrounding a variant subjected to natural selection (Voight et al., 2006). The XP-EHH and iHS tests both have substantial statistical power to detect natural selection using genomewide SNP genotypes, even when sample sizes are limited (Huff et al., 2010; Pickrell et al., 2009).

Selection candidate genes unique to Tibetans

Many selective events are likely to be shared among multiple populations, but we wish to focus only on those that are specific to high-altitude Tibetan populations. To do this, we divided the genome into consecutive, nonoverlapping 200kb regions and excluded those that yielded significant values ($p < 0.01$) for the iHS test in neighboring Asian populations. Seven of the 247 functional candidate loci were located in these excluded regions and were therefore eliminated from subsequent analyses. In our Tibetan sample, we then calculated an XP-EHH and iHS summary statistic (see SOM) for each genomic region, using an empirical significance level of 0.01 to identify a set of regions that exhibit evidence of local positive selection.

This set of regions contained ten of the 240 genes from our functional candidate list (Table 2.8). Six of these genes were identified by the XP-EHH test, and five were identified by the iHS test (one, *EGLN1*, was identified by both). *EGLN1* and *EPAS1* are both in the HIF

pathway (Pruitt et al., 2005) and show elevated F_{ST} values. Three additional loci, *EDNRA*, *PTEN* and *PPARA*, are also associated with HIF activity (Narravula and Colgan, 2001; Zundel et al., 2000) (Fig. 2.2). *PPARA* and *ANGPTL4* are in the PPAR lipid metabolism pathway, and *CYP17A1* and *CYP2E1* are cytochrome P450 genes (Pruitt et al., 2005). The two remaining genes in Table 2.2 are *HMOX2*, involved in HIF-independent oxygen sensing (Williams et al., 2004) and hypoxic endothelial cell survival (He et al., 2010), and *CAMK2D*, which mediates nitric oxide production in response to changes in intracellular calcium (Cai et al., 2008). Because the selection signals correspond to 200kb sections of the genome rather than to individual genes, we performed additional analyses to localize the selection signal within each of the ten regions (Grossman et al., 2010). For each gene of interest, we observed multiple localization signals near the gene or bracketing the gene.

These results suggest that high-altitude adaptation in Tibetans has resulted from local positive selection on several distinct genes. However, even in the absence of selection, some loci will appear in the intersection of our functional candidate list and selection screens just by chance. To test the statistical significance of the observed pattern, we carried out a randomization test by resampling sets of 240 loci one million times from a list of all known autosomal genes (Kent et al., 2002). For each set of randomly chosen loci, we tabulated the number of loci that intersected the genomic regions identified by our selection scans. On average, the XP-EHH and iHS intersections contained 2.7 and 1.4 genes, respectively, which is significantly fewer than the six ($p < 0.05$) and five ($p< 0.01$) genes actually observed.

Genotype-phenotype analysis

In contrast to the fixed or nearly fixed alleles detected by XP-EHH, the iHS test identifies regions affected by incomplete selective sweeps. Therefore, interindividual genetic variation

should be observable at these loci, allowing us to test for their association with Hb concentration, which is highly heritable and has distinctively low levels in high-altitude Tibetan populations (Beall, 2007). In each of the five 200kb regions identified as targets of selection by iHS, we defined the selected core haplotype as the one containing the three SNP alleles that exhibit the most extreme iHS scores (Table 2.5).

We then used stepwise linear regression to test for relationships between Hb concentration and these five core haplotypes. The five independent predictor variables are the numbers of putatively advantageous haplotypes (0, 1 or 2) present in each 200kb region. Since sex and age affect Hb concentration in Tibetans and Han Chinese, and the age effect differs between males and females (Wu et al., 2005), covariates are included in the regression analysis to allow an independent age effect in males and females (Tables 2.4, 2.6, and 2.7).

The putatively advantageous haplotypes of *EGLN1* and *PPARA* both show significant negative correlations with Hb concentration ($p < 0.002$ and 0.0009, respectively; Tables 2.6 and 2.7). A genomewide regression analysis showed no excess of associations (398,020 SNPs with minor allele frequency ≥ 0.15 and no missing genotypes). The phenotypic effects are substantial: each additional copy of an advantageous haplotype at either locus decreases Hb concentration by ~1.7 g/dl on average (Fig. 2.4). This effect is greater than the well-established sex-related difference in high-altitude Tibetans (~1.1g/dl) (Wu et al., 2005). The strong and significant association between Hb concentration and haplotype variation at *EGLN1* and *PPARA* provides evidence of a genetic contribution to a form of high-altitude adaptation that appears to be unique to Tibetan populations.

A potential explanation for the relationship between *EGLN1* and decreased Hb concentration lies in the regulation of HIF and its target genes. EGLN1 targets two HIFα

proteins for degradation during normoxia, decreasing the transcription of HIF-regulated targets such as *EPO*, the erythropoietin gene whose product induces red blood cell (RBC) production (Fig. 2.2). Furthermore, mutations in *EGLN1* prevent targeted degradation of HIF, leading to polycythemia (excessive RBC production) in mice and humans (Percy et al., 2006).

Although *PPARA* has not previously been considered as a candidate gene for high-altitude adaptation, it interacts with components of the HIF pathway. *PPARA* expression is inhibited by HIF1 during hypoxia in mice (Fig. 2) (Zundel et al., 2000), and genes targeted by HIF are regulated by a HIF-independent mechanism involving PPARG coactivator-1α (Arany et al., 2008). In addition, a PPARA agonist, the anti-diabetic agent tesaglitazar, resulted in decreased Hb levels during human clinical trials (Wilding, 2007). This effect is consistent with the relationship between the putatively advantageous *PPARA* haplotype and Hb concentration found here.

Discussion

It is plausible that the diminished Hb levels found in Tibetans offset complications associated with sustained high Hb levels (e.g., hyper-viscosity) seen in non-Tibetans exposed to high-altitude conditions (Mejia et al., 2005; Vargas, 2006). Alternatively, decreased Hb levels could be a side effect of other phenotypes which are the actual targets of natural selection. Functional analysis of genes such as *EGLN1* and *PPARA*, which have undergone positive selection and are associated with Hb levels in this Tibetan sample, will further increase our understanding of genetic adaptation to high-altitude environments. In addition, we will better understand the human response to hypoxia, which has important implications for treating mountain sickness, high-altitude pulmonary and cerebral edema, and other hypoxia-related diseases.

References

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research *19*, 1655-1664.

Arany, Z., Foo, S.-Y., Ma, Y., Ruas, J.L., Bommi-Reddy, A., Girnun, G., Cooper, M., Laznik, D., Chinsomboon, J., Rangwala, S.M., *et al.* (2008). HIF-independent regulation of VEGF and angiogenesis by the transcriptional coactivator PGC-1 alpha. Nature *451*, 1008-1012.

Beall, C.M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences *104*, 8655-8660.

Beall, C.M., Laskowski, D., Strohl, K.P., Soria, R., Villena, M., Vargas, E., Alarcon, A.M., Gonzales, C., and Erzurum, S.C. (2001). Pulmonary nitric oxide in mountain dwellers. Nature *414*, 411-412.

Browning, B.L., and Browning, S.R. (2009). A unified approach to genotype imputation and haplotype-phaseiInference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics *84*, 210-223.

Cai, H., Liu, D., and Garcia, J.G.N. (2008). CaM Kinase II-dependent pathophysiological signalling in endothelial cells. Cardiovascular Research *77*, 30-34.

Chen, Q. (1997). Exercise performance of Tibetan and Han adolescents at altitudes of 3,417 and 4,300 m. Journal of Applied Physiology *83*, 661-667.

Ge, R. (1994). Higher exercise performance and lower V0 2 max in Tibetan than Han residents at 4,700 m altitude. Journal of Applied Physiology *77*, 684-691.

Ge, R., Witkowski, S., Zhang, Y., Alfrey, C., Sivieri, M., Karlsen, T., Resaland, G.K., Harber, M., Stray-Gundersen, J., and Levine, B.D. (2002). Determinants of erythropoietin release in response to short-term hypobaric hypoxia. Journal of Applied Physiology *92*, 2361-2367.

Gene Ontology Consortium. (2006). The Gene Ontology (GO) project in 2006. Nucleic Acids Research *34*, D322-326.

Grossman, S.R., Shylakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., *et al.* (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. Science *327*, 883-886.

Groves, B.M., Droma, T., Sutton, J.R., McCullough, R.G., McCullough, R.E., Zhuang, J., Rapmund, G., Sun, S., Janes, C., and Moore, L.G. (1993). Minimal hypoxic pulmonary hypertension in normal Tibetans at 3,658 m. Journal of Applied Physiology *74*, 312-318.

Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M., and Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. Genome Research *19*, 318-326.

He, J.Z., Ho, J.J.D., Gingerich, S., Courtman, D.W., Marsden, P.A., and Ward, M.E. (2010). Enhanced translation of heme oxygenase-2 preserves human endothelial cell viability during hypoxia. Journal of Biological Chemistry *285*, 10.

http://hgdp.uchicago.edu/Software/.

Huff, C., Harpending, H., and Rogers, A. (2010). Detecting positive selection from genome scans of linkage disequilibrium. BMC Genomics *11*, 8.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, a.D. (2002). The Human Genome Browser at UCSC. Genome Research *12*, 996-1006.

Mejia, O.M., Prchal, J.T., Leon-Velarde, F., Hurtado, A., and Stockton, D.W. (2005). Genetic association analysis of chronic mountain sickness in an Andean high-altitude population. Haematologica *90*, 13-19.

Mi, H., Guo, N., Kejariwal, A., and Thomas, P.D. (2007). PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Research *35*, D247-252.

Moore, L.G. (2001). Human genetic adaptation to high altitude. High Altitude Medicine & Biology *2*, 257-279.

Narravula, S., and Colgan, S.P. (2001). Hypoxia-Inducible Factor 1-mediated inhibition of peroxisome proliferator-activated receptor alpha expression during hypoxia. Journal of Immunology *166*, 7543-7548.

Percy, M.J., Zhao, Q., Flores, A., Harrison, C., Lappin, T.R.J., Maxwell, P.H., McMullin, M.F., and Lee, F.S. (2006). A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. Proceedings of the National Academy of Sciences of the United States of America *103*, 654-659.

Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., *et al.* (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Research *19*, 826-837.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research *33*, D501-504.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., *et al.* (2007). PLINK: A tool set for whole-genome

association and population-based linkage analyses. The American Journal of Human Genetics *81*, 559-575.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., *et al.* (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913-918.

Semenza, G.L. (2009). Oxygen homeostasis. Wiley Interdisciplinary Reviews: Systems Biology and Medicine.

Su, B., Xiao, C., Deka, R., Seielstad, M., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L., Chakraborty, R., *et al.* (2000). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Human Genetics *107*, 582-590.

The International HapMap, C. (2005). A haplotype map of the human genome. Nature *437*, 1299-1320.

Vargas, E.P., H. Spielvogel (2006). Chronic mountain sickness, optimal hemoglobin, and heart disease. High Altitude Medicine & Biology *7*, 138-149.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biology *4*, e72.

Wilding, J.P.H. (2007). Tesaglitazar, as add-on therapy to sulphonylurea, dose-dependently improves glucose and lipid abnormalities in patients with type 2 diabetes. Diabetes and Vascular Disease Research *4*.

Williams, S.E.J., Wootton, P., Mason, H.S., Bould, J., Iles, D.E., Riccardi, D., Peers, C., and Kemp, P.J. (2004). Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. Science *306*, 2093-2097.

Wu, T., Wang, X., Wei, C., Cheng, H., Wang, X., Li, Y., Ge, D., Zhao, H., Young, P., Li, G., *et al.* (2005). Hemoglobin levels in Qinghai-Tibet: different effects of gender for Tibetans vs. Han. Journal of Applied Physiology *98*, 598-604.

Xing, J., Watkins, W.S., Witherspoon, D.J., Zhang, Y., Guthery, S.L., Thara, R., Mowry, B.J., Bulayeva, K., Weiss, R.B., and Jorde, L.B. (2009). Fine-scaled human genetic structure revealed by SNP microarrays. Genome Research *19*, 815-825.

Zhao, M., Kong, Q.-P., Wang, H.-W., Peng, M.-S., Xie, X.-D., Wang, W.-Z., Jiayang, Duan, J.-G., Cai, M.-C., Zhao, S.-N., *et al.* (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. Proceedings of the National Academy of Sciences *106*, 21230-21235.

Zhuang, J., Droma, T., Sun, S., Janes, C., McCullough, R.E., McCullough, R.G., Cymerman, A., Huang, S.Y., Reeves, J.T., and Moore, L.G. (1993). Hypoxic ventilatory responsiveness in Tibetan compared with Han residents of 3,658 m. Journal of Applied Physiology *74*, 303-311.

Zundel, W., Schindler, C., Haas-Kogan, D., Koong, A., Kaper, F., Chen, E., Gottschalk, A.R., Ryan, H.E., Johnson, R.S., Jefferson, A.B., *et al.* (2000). Loss of PTEN facilitates HIF-1-mediated gene expression. Genes & Development *14*, 391-396.

CHAPTER 3

METABOLIC INSIGHT INTO MECHANISMS OF HIGH-ALTITUDE ADAPTATION IN TIBET

Abstract

Recent studies have identified genes involved in high-altitude adaptation in Tibetans. Three of these genes (*EPAS1*, *EGLN1*, and *PPARA*) are associated with relatively decreased hemoglobin levels observed in Tibetans at high altitude, providing corroborative evidence for genetic adaptation to this extreme environment. The mechanisms that afford adaptation to high-altitude hypoxia, however, remain unclear. Considering the strong metabolic demands imposed by hypoxia, we hypothesized that a shift in fuel preference to glucose oxidation and glycolysis at the expense of fatty acid oxidation would provide adaptation to decreased oxygen availability. Measurements of serum metabolites collected from Tibetan groups living at high altitude are consistent with this hypothesis: an *EPAS1* haplotype that exhibits a signal of positive selection is significantly associated with increased lactate levels, and there is a positive relationship between a *PPARA* haplotype and serum free fatty acid levels. Although further studies are required to assess the molecular mechanisms underlying these patterns, these changes suggest that genetic adaptation to high altitude involves alteration in energy utilization pathways.

Introduction

Members of the hypoxia inducible factor (HIF) pathway help to orchestrate molecular responses during hypoxic stress through a complex series of cellular metabolites (Majmundar et al., 2010). In the absence of adequate oxygen, energy production from oxidative metabolism may be diminished. At the same time, if oxidative metabolism proceeds in hypoxia, reactive oxidative intermediates will accumulate in

mitochondria. Either of these conditions can result in cell death. Recent work on the whole-organism level has revealed that HIF plays a major role in regulating metabolism, highlighting a strong relationship between HIF and metabolic demands in humans (Formenti et al., 2010).

Native high-altitude populations have persisted for hundreds of generations in an oxygen-deprived (hypoxic) environment. Recent genomewide scans of positive selection in Tibetans have identified hypoxia-sensing and -regulated genes as candidates for high-altitude adaptation (Beall et al., 2010; Bigham et al., 2010; Peng et al., 2011; Simonson et al., 2010; Wang et al., 2011; Xu et al., 2011; Yi et al., 2010a). A few of the candidate regions thought to underlie Tibetans' successful existence were previously shown to be associated with their hemoglobin (Hb) levels, which are lower relative to those of native highland groups from the Andes or visitors to high-altitude (Beall, 2007; Beall et al., 2010; Simonson et al., 2010; Yi et al., 2010a). Positively selected regions containing the *EPAS1* gene, which encodes the HIF-2α subunit, were associated with Hb concentration in Tibetan populations examined in two separate studies (Beall et al., 2010; Yi et al., 2010b). Two additional genomic regions were associated with Hb concentration in a Qinghai Tibetan population. One contains *EGLN1*, which encodes the proline hydroxylase, PHD2, that negatively regulates HIFs' α subunits in an oxygen-dependent manner (Storz et al., 2010). The second is the genomic region containing *PPARA*, which encodes the nuclear peroxisome proliferator activated receptor alpha (PPARα) that regulates fatty acid metabolism and is regulated by HIF (Simonson et al., 2010).

It is unclear whether these previously identified associations reflect direct control of Hb level or are consequences of natural selection acting on other advantageous traits

(Storz et al., 2010; Tissot van Patot and Gassmann, 2011). The hypoxia signaling system triggers a pleiotropic response that increases tissue oxygenation by increasing vascularization and oxygen carrying capacity of the blood (Formenti et al., 2010; Majmundar et al., 2010). Simultaneously, HIF signaling triggers alterations in metabolism to decrease tissue oxygen demand (Denko, 2008; Semenza, 1999). Oxidation of fatty acids yields less ATP per molecule of oxygen consumed than oxidation of carbohydrates, suggesting that decreased fatty acid oxidation should also be a favorable adaptation to hypoxia (Holden et al., 1995). Several studies have demonstrated decreased reliance on fat metabolism at high altitude, both in people living habitually at high altitude (Holden et al., 1995) and in those not dwelling at high altitude but acclimatized (Brooks et al., 1991; Roberts et al., 1996). Another metabolic change induced by hypoxia is a conversion from oxidative glucose metabolism to glycolysis in order to maintain energy production. This occurs through up regulation of glucose uptake and glycolysis and down regulation of mitochondrial glucose oxidation (Kim et al., 2006; Papandreou et al., 2006).

Considering their previous association with a Tibetan-specific hypoxia-related phenotype and that HIF pathway members and PPARα play major roles in metabolism (Majmundar et al., 2010; Pruitt et al., 2005), we examined relationships between these genes and metabolic phenotypes in another Tibetan population. We show here that the *EPAS1* selected haplotype previously associated with Hb concentration is also highly associated with changes in serum lactate. In addition, *PPARA* haplotypes exhibit an association with serum free fatty acids in this population.

Materials and methods

DNA sample collection

DNA was extracted from whole blood samples for individuals (non-smokers, no chronic diseases). This population, who speak the Kham dialect, is from the Tuo Tuo River area in the Qinghai-Tibetan Plateau (~4,500 m), People's Republic of China. Informed consent was obtained for all participants according to guidelines approved by the Institutional Review Boards at the High Altitude Medical Research Institute (Xining, Qinghai, People's Republic of China).

SNP genotyping

DNA samples were genotyped using Affymetrix 6.0 SNP Array technology (>900,000 SNPs) at Capital Bio Corporation (Beijing, China). We used default parameters for the Birdseed algorithm (version 2) to determine genotypes for all samples (Affymetrix, Santa Clara, CA, USA). Genotypic data were analyzed using the Affymetrix Genotyping Console 3.1 (Affymetrix).

Estimates of relatedness

We used unrelated samples as described in Simonson et al. (Simonson et al., 2010) and the program ERSA (Huff et al., 2011) to exclude related individuals in the second set of Tibetan samples, excluding one member of the pair if their relationship exceeded that expected for first cousins. Based on these criteria, a total of 36 unrelated individuals were included in the genotype-phenotype analyses.

Selection analyses

We performed selection scans using the iHS (Voight et al. 2006) and XP-EHH (Sabeti et al. 2007) test statistics, and we used iHS scores to define selected haplotypes in the second population as previously described (Simonson et al., 2010).

Phenotype collection

We measured metabolites in a cohort of Tibetans, excluding related individuals (N=36) using kit protocols for lactate (Point Scientific, Inc., Lincoln Park, MI), β-hydroxybutyrate (WAKO Diagnostics, Richmond, VA), free fatty acids (Half Micro Test, Roche, Penzberg Germany) and triglycerides (Sigma Chemical, St Louis, MI).

Genotype-phenotype association

We determined whether the selection candidate regions that were previously associated with [Hb] (Table 3.1) are also associated with metabolites. We defined three-SNP haplotypes as previously described (Simonson et al., 2010): SNPs within 200 kb genomic region with the most extreme iHS scores comprise each of the putatively advantageous haplotypes (Hg build haplotype-defining SNPs: chromosome 1 positions 229601735, 229604075, and 229793717 for *EGLN1*; chromosome 2 positions 46490868, 46590298, and 46592661 for *EPAS1*; chromosome 22 positions 44807657, 44827140, and 44866192 for *PPARA*). We tested for an association between these haplotypes and free fatty acids, triglycerides, and 3-hydroxybutyrate in 36 Tibetan individuals.).

Selection analyses

We performed selection scans using the his (Voight et al.,, 2006) and XP-EHH

Table 3.1

Haplotype-phenotype significance values for Spearman rank-order correlation analysis of metabolites measured in Tibetans

	EGLN1		*EPAS1*		*PPARA*	
	P value	Rho	P value	Rho	P value	Rho
Triglycerides	0.150	0.245	0.307	0.175	0.973	-0.006
Free fatty acids	0.505	-0.115	0.860	-0.031	**0.014**	0.406
Three hydroxybutyrate	0.230	-0.205	0.590	0.093	0.182	0.227
Lactate	0.070	0.305	**0.003**	0.482	0.424	0.137

(Sabeti et al., 2007) statistics) used in our previous study of Tibetan high-altitude adaptation (Simonson et al., 2010) to identify common targets of natural selection in a second and distinct Tibetan population. Both *EPAS1* and *EGLN1* gene regions exhibit significant signals of selection for the iHS test (Voight et al., 2006) ($p < 0.004$ and $p < 0.01$, respectively), although the genomic region containing *PPARA* does not reach significance ($p < 0.14$) for tests of selection performed.

In order to determine whether these regions were also associated with variation in metabolites, we measured serum levels of triglycerides, free fatty acids, β-hydroxybutyrate, and lactate in non-fasted serum samples from 36 individuals. We performed Spearman rank-order correlation analysis between the serum levels and the selected haplotypes (0, 1, or 2 copies; see Table 3.1). Lactate was positively associated with the adaptive *EPAS1* haplotype ($p<0.003$; Fig. 3.1, Table 3.1). Additionally, the previously identified *PPARA* haplotype exhibits a positive relationship with serum free fatty acids ($p < 0.01$; Figure 3.1, Table 3.1). The *EGLN1* gene region was not associated with any of the serum metabolite levels measured.

Discussion

Both *EPAS1* and *PPARA* are involved in hypoxia signaling (Aragones et al., 2008). Humans acutely exposed to hypoxia consequently exhibit increased anaerobic glucose metabolism (Kelly et al., 2010). We observed that the adaptive *EPAS1* haplotypes were associated with increased lactate levels, consistent with decreased

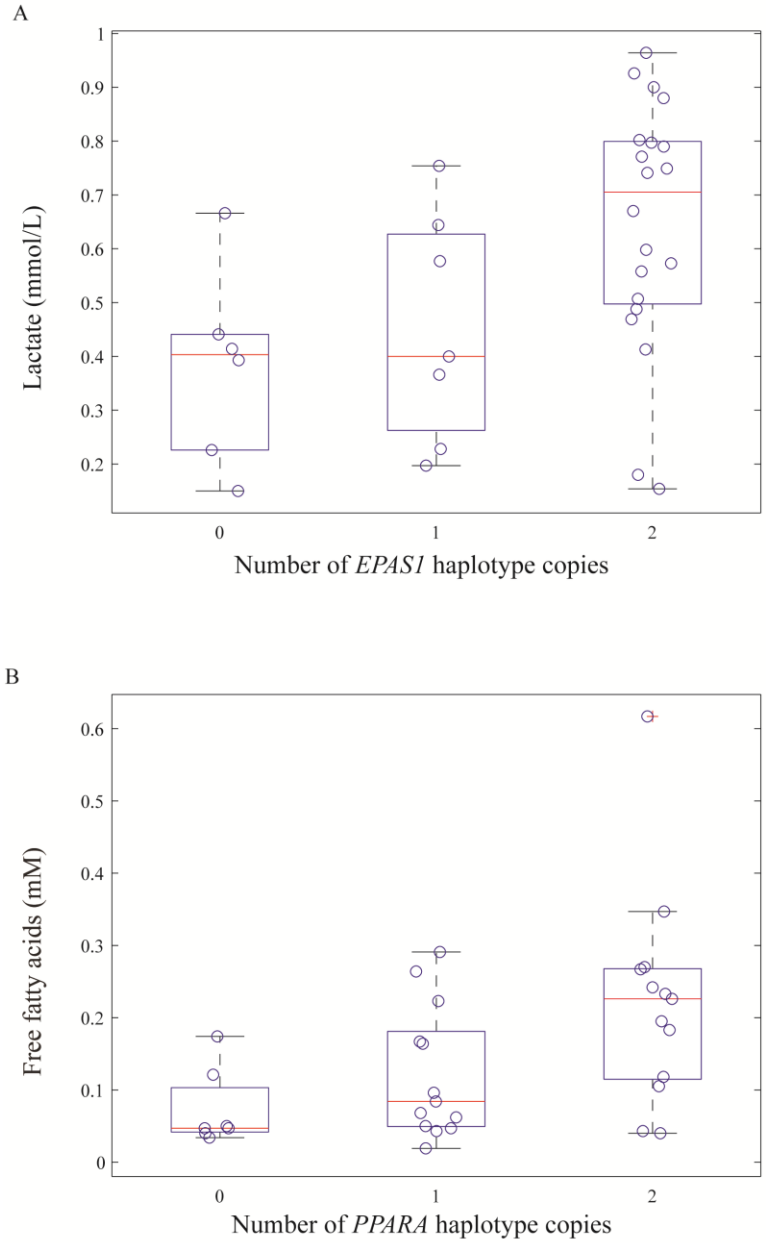


Figure 3.1. Association of previously identified adaptive haplotypes and metabolites. A) Lactate levels are plotted against the group of putatively advantageous haplotypes (0, 1, or 2) at the *EPAS1* locus. B) Serum free fatty acids (FFA) levels are plotted against the number of putatively advantageous *PPARA* haplotype copies (0 to 2).

glucose oxidation. HIF-2α's metabolic activity has been shown to be required for the shift to anaerobic metabolism that facilitates adaptation to hypoxia in skeletal muscle of mice (Majmundar et al., 2010). Additionally, individuals with Chuvash polycythemia, an autosomal recessive disorder in which HIF degradation is impaired, exhibit higher lactate levels during exercise than do normal individuals (Formenti et al., 2010). Mice lacking *EPAS1*, however, also exhibit lactic acidosis (Scortegagna et al., 2003), implying that the adaptive Tibetan polymorphism could be associated with either increased or decreased HIF-2α activity.

A previous report demonstrates a high prevalence of hypertriglyceridemia in Tibetan highlanders (Sherpa et al., 2011). We measured fasting triglyceride levels in a separate cohort of Tibetans and Han Chinese living in an urban environment near sea level, and confirmed this observation of significantly higher triglycerides in Tibetans (298±23 compared to 139±8 mg/dl in the Han, $p = 0.006$; data not shown). We therefore examined the relationship of the adapted haplotypes with serum free fatty acids and triglycerides in the high-altitude Tibetan samples. Free fatty acids were associated with the *PPARA* haplotype, although serum triglycerides were not. It should be pointed out, however, that the sera were not collected in the fasting state, and serum triglycerides vary acutely with fasting, feeding, and composition of the diet.

Down regulation of several genes, including *PPARA*, involved in fatty acid oxidation has been observed in rats exposed to hypoxia (Kennedy et al., 2001). *PPARA* encodes the nuclear receptor protein PPARα, a major regulator of fatty acid oxidation (Narravula and Colgan, 2001; Piguet et al., 2010). Previous studies have shown that PPARα activation is associated with lower serum free fatty acids and triglycerides

(Barbier et al., 2002); hence the adaptive genotype is consistent wtih decreased expression or activity of PPARα. We did not, however, observe decreased β-hydroxybutyrate with the adaptive *PPARA* haplotype (Table 3.1), so there is no direct evidence of decreased fatty acid oxidation. Another explanation for increased lipids would be increased lipid synthesis. Fat anabolic pathways are up regulated in hypoxia, mediated at least in part by up regulation of PPARα (Krishnan et al., 2009; Piguet et al., 2010) and SREBP-1 (Li et al., 2006), but we have no data from these studies to specifically implicate either increased synthesis or decreased degradation of fatty acids to the observed phenotype.

Recent studies suggest that some of the PPARα-dependent effects of hypoxia on fat metabolism may be mediated through HIF-2α (Aragones et al., 2008). Paradoxically, mice with either deletion of *Epas1* or liver-specific over expression of *Epas1* exhibit hepatic steatosis and both models show evidence of decreased fatty acid oxidation (Rankin et al., 2009; Scortegagna et al., 2003). Thus, the interrelationships among the status of the primary hypoxia signaling pathways, their downstream metabolic effectors, and the final metabolic phenotype of the organism are highly complex. In addition, environmental factors are crucial in determining metabolic status. Thus, determining the specific roles of changes in *EPAS1*, *EGLN1*, and *PPARA* on the observed changes in metabolites will require further study.

Hypoxia-induced regulation of metabolism and its alteration in adapted populations may carry implications for the risks of diabetes and obesity. The selected haplotypes may result in a relative inability to shift between fat and glucose oxidation, so-called metabolic inflexibility (Storlien et al., 2004). Such inflexibility and fatty acid

oxidation capacities (Holland et al., 2007; Koves et al., 2008) are both implicated in the pathogenesis of type 2 diabetes mellitus. Tibetan highlanders have a relatively low prevalence of diabetes (Matsubayashi et al., 2009), but the diet is also relatively low-calorie (Beall et al., 2010) and high altitudes are associated with lower body weights among Tibetans (Sherpa et al., 2011). As populations move to lower altitudes and encounter a more industrialized lifestyle and higher calorie diets, however, the metabolic adaptations to altitude could have health implications. For example, increasing total fat and calorie consumption with a metabolic profile that will not support fat oxidation could result in accumulation of lipid intermediates thought to play a role in diabetes pathogenesis (Holland et al., 2007; Koves et al., 2008). Further study of the metabolic implications of high altitude adaptation may allow interventions to ameliorate this risk and also identify potential new targets to treat obesity and diabetes.

Our results demonstrate increased lactate and free fatty acids in Tibetans with selected haplotypes. This pattern is consistent with the hypothesis that anaerobic glucose metabolism is increased and fatty acid oxidation may be decreased in the adapted Tibetans compared to Tibetans without the adapted haplotypes living at the same altitude. Controlled studies including more dynamic metabolic analyses and studies at different altitudes will be required to better understand the physiological significance of these patterns.

References

Aragones, J., Schneider, M., Van Geyte, K., Fraisl, P., Dresselaers, T., Mazzone, M., Dirkx, R., Zacchigna, S., Lemieux, H., Jeoung, N.H., *et al.* (2008). Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. Nature Genetics *40*, 170-180.

Barbier, O., Torra, I.P., Duguay, Y., Blanquart, C., Fruchart, J.C., Glineur, C., and Staels, B. (2002). Pleiotropic actions of peroxisome proliferator-activated receptors in lipid metabolism and atherosclerosis. Arterioscleroaia, Thrombosis, and Vascular Biology *22*, 717-726.

Beall, C.M. (2007a). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences 104, 8655-8660.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., *et al.* (2010). Natural selection on EPAS1 (HIF2 alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proceedings of the National Academy of Sciences *107*, 11459-11464.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R., Scherer, S.W., Julian, C.G., Wilson, M.J., López Herráez, D., *et al.* (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics *6*, e1001116.

Brooks, G.A., Butterfield, G.E., Wolfe, R.R., Groves, B.M., Mazzeo, R.S., Sutton, J.R., Wolfel, E.E., and Reeves, J.T. (1991). Increased dependence on blood glucose after acclimatization to 4,300 m. Journal of Applied Physiology *70*, 919-927.

Denko, N.C. (2008). Hypoxia, HIF1 and glucose metabolism in the solid tumour. Nature Reviews Cancer *8*, 705-713.

Formenti, F., Constantin-Teodosiu, D., Emmanuel, Y., Cheeseman, J., Dorrington, K.L., Edwards, L.M., Humphreys, S.M., Lappin, T.R.J., McMullin, M.F., McNamara, C.J., *et al.* (2010). Regulation of human metabolism by hypoxia-inducible factor. Proceedings of the National Academy of Sciences *107*, 12722-12727.

Holden, J.E., Stone, C.K., Clark, C.M., Brown, W.D., Nickles, R.J., Stanley, C., and Hochachka, P.W. (1995). Enhanced cardiac metabolism of plasma glucose in high-altitude natives: adaptation against chronic hypoxia. Journal of Applied Physiology *79*, 222-228.

Holland, W.L., Brozinick, J.T., Wang, L.P., Hawkins, E.D., Sargent, K.M., Liu, Y., Narra, K., Hoehn, K.L., Knotts, T.A., Siesky, A., *et al.* (2007). Inhibition of ceramide synthesis ameliorates glucocorticoid-, saturated-fat-, and obesity-induced insulin resistance. Cell Metabolism *5*, 167-179.

Huff, C.D., Witherspoon, D.J., Simonson, T.S., Xing, J., Watkins, W.S., Zhang, Y., Tuohy, T.M., Neklason, D.W., Burt, R.W., Guthery, S.L., *et al.* (2011). Maximum-likelihood estimation of recent shared ancestry (ERSA). Genome Research *21*, 768-774.

Kennedy, S.L., Stanley, W.C., Panchal, A.R., and Mazzeo, R.S. (2001). Alterations in enzymes involved in fat metabolism after acute and chronic altitude exposure. Journal of Applied Physiology *90*, 17-22.

Kim, J.W., Tchernyshyov, I., Semenza , G.L., and Dang , C.V. (2006). HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. Cell Metabolism *3*, 177-185.

Koves, T.R., Ussher, J.R., Noland, R.C., Slentz, D., Mosedale, M., Ilkayeva, O., Bain, J., Stevens, R., Dyck, J.R., Newgard, C.B., *et al.* (2008). Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. Cell Metabolism *7*, 45-56.

Krishnan, J., Suter, M., Windak, R., Krebs, T., Felley, A., Montessuit, C., Tokarska-Schlattner, M., Aasum, E., Bogdanova, A., Perriard, E., *et al.* (2009). Activation of a HIF1alpha-PPARgamma axis underlies the integration of glycolytic and lipid anabolic pathways in pathologic cardiac hypertrophy. Cell Metabolism *9*, 512-524.

Li, J., Bosch-Marce, M., Nanayakkara, A., Savransky, V., Fried, S.K., Semenza, G.L., and Polotsky, V.Y. (2006). Altered metabolic responses to intermittent hypoxia in mice with partial deficiency of hypoxia-inducible factor-1alpha. Physiological Genomics *25*, 450-457.

Majmundar, A.J., Wong, W.J., and Simon, M.C. (2010). Hypoxia-inducible factors and the response to hypoxic stress. Molecular Cell *40*, 294-309.

Matsubayashi, K., Kimura, Y., Sakamoto, R., Wada, T., Ishimoto, Y., Hirosaki, M., Konno, A., Chen, W., Ishine, M., Kosaka, Y., *et al.* (2009). Comprehensive geriatric assessment of elderly highlanders in Qinghai, China I: activities of daily living, quality of life and metabolic syndrome. Geriatric Gerontology International *9*, 333-341.

Narravula, S., and Colgan, S.P. (2001). Hypoxia-inducible factor 1-mediated inhibition of peroxisome proliferator-activated receptor alpha expression during hypoxia. Journal of Immunology *166*, 7543-7548.

Papandreou, I., Cairns, R.A., Fontana, L., Lim, A.L., and Denko, N.C. (2006). HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. Cell Metabolism *3*, 187-197.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, *et al.* (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular Biology and Evolution *28*, 1075-1081.

Piguet, A.C., Stroka, D., Zimmermann, A., and Dufour, J.F. (2010). Hypoxia aggravates non-alcoholic steatohepatitis in mice lacking hepatocellular PTEN. Clinical Science (Lond) *118*, 401-410.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research *33*, D501-504.

Rankin, E.B., Rha, J., Selak, M.A., Unger, T.L., Keith, B., Liu, Q., and Haase, V.H. (2009). Hypoxia-Inducible Factor 2 regulates hepatic lipid metabolism. Molecular Cell Biology *29*, 4527-4538.

Roberts, A.C., Butterfield, G.E., Cymerman, A., Reeves, J.T., Wolfel, E.E., and Brooks, G.A. (1996). Acclimatization to 4,300-m altitude decreases reliance on fat as a substrate. Journal of Applied Physiology *81*, 1762-1771.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., *et al.* (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913-918.

Scortegagna, M., Ding, K., Oktay, Y., Gaur, A., Thurmond, F., Yan, L.J., Marck, B.T., Matsumoto, A.M., Shelton, J.M., Richardson, J.A., *et al.* (2003). Multiple organ pathology, metabolic abnormalities and impaired homeostasis of reactive oxygen species in Epas1-/- mice. Nature Genetics *35*, 331-340.

Semenza, G.L. (1999). Perspectives on oxygen sensing. Cell *98*, 281-284.
Sherpa, L.Y., Deji, Stigum, H., Chongsuvivatwong, V., Luobu, O., Thelle, D.S., Nafstad, P., and Bjertness, E. (2011). Lipid profile and its association with risk factors for coronary heart disease in the highlanders of Lhasa, Tibet. High Altitude Medicine & Biology *12*, 57-63.

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., *et al.* (2010). Genetic evidence for high-altitude adaptation in Tibet. Science *329*, 72-75.

Storlien, L., Oakes, N.D., and Kelley, D.E. (2004). Metabolic flexibility. Proceedings of the Nutrition Society *63*, 363-368.

Storz, J.F., Scott, G.R., and Cheviron, Z.A. (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. Journal of Expirmental Biology *213*, 4125-4136.

Tissot van Patot, M.C., and Gassmann, M. (2011). Hypoxia: Adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2alpha. High Altitude Medicine and Biology *12*, 157-167.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biology *4*, e72.

Wang, B., Zhang, Y.-B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L., Li, X., *et al.* (2011). On the origin of Tibetans and their genetic basis in adapting to high-altitude environments. PLoS ONE *6*, e17002.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., *et al.* (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. Molecular Biology and Evolution *28*, 1003-1011.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., *et al.* (2010a). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75-78.

Yi, Z., Jun, W., XingBo, S., XiaoJun, L., Liu, D., and BinWu, Y. (2010b). Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group. Leg Med (Tokyo) *12*, 108-111.

CHAPTER 4

SHARED AND UNIQUE SIGNALS OF HIGH-ALTITUDE ADAPTATION AND PHENOTYPE ASSOCIATION IN DISTINCT TIBETAN POPULATIONS

Abstract

Recent studies have used several different analytical methods to identify genes targeted by natural selection in high-altitude populations located throughout the Tibetan Plateau. Despite differences in analytic strategies and population samples, several hypoxia-related genes, including *EPAS1* and *EGLN1*, were identified in multiple studies. By employing the same analytic strategies (iHS and XP-EHH statistics) used in our previous study of Tibetan high-altitude adaptation and through further comparison with a neighboring Mongolian population, we have identified common targets of natural selection in a second distinct Tibetan population from a different region of the Tibetan Plateau. Our selection analysis provides evidence for adaptation in two Tibetan populations at the $p < 0.01$ significance level for *EPAS1*, *EGLN1*, *HMOX2*, and *CYP17A1* and $p < 0.02$ for *PKLR*, *HFE*, and *HBB* and *HBG2*, which have been identified in other studies. By excluding the possibility of differences attributed to analytical methods and highlighting the relevance of admixture in studies of Tibetan populations, we suggest that sample location should be considered in replication studies of adaptation and phenotype association in Tibet.

Introduction

Native high-altitude populations have specific traits that afford survival to hypoxic (low oxygen) conditions. Several regions of the genome, many which include hypoxia-sensing and -regulated genes, were recently identified as candidates for Tibetan high-altitude adaptation (Beall et al., 2010; Bigham et al., 2010; Peng et al., 2011; Simonson et al., 2010; Wang et al., 2011; Xu et al., 2011; Yi et al., 2010). Two regions

that exhibit strong signals of positive selection contain genes involved in hypoxia sensing and response: the *EPAS1* gene, which encodes the hypoxia inducible factor (HIF)-2α subunit, and the *EGLN1* gene, which encodes the proline hydroxylase PHD2 that regulates the alpha subunits of hypoxia induced factors in an oxygen-dependent manner. *EPAS1* is associated with hemoglobin concentration ([Hb]) in Tibetan populations examined in two separate studies (Beall et al., 2010; Yi et al., 2010), and *EGLN1* is associated with [Hb] in a Qinghai Tibetan population (Simonson et al., 2010). In this same population, *PPARA*, which encodes the nuclear peroxisome proliferator activated receptor alpha that regulates fatty acid metabolism, is also significantly associated with Hb level (Simonson et al., 2010). In addition to these phenotype-associated selection targets, many other genes have been reported by more than one study as strong targets of selection.

Tibetans inhabit vast regions of the Qinghai-Tibetan Plateau, which spans approximately one million square miles. At least three major Tibetan dialects are spoken among the geographically distinct groups located on the Plateau (Encyclopædia Britannica, 2011) and various groups appear to have distinct cultural characteristics. These differences suggest isolation of different Tibetan groups, although the demographic history of high-altitude populations is highly debated. Archaeological evidence suggests that the ancestors of present-day Tibetan groups migrated to the Qinghai-Tibetan Plateau at various times, ranging from 25,000 to 5,000 years ago (Aldenderfer, 2011). Patterns of genomewide SNP variation support single-route migration into this region (Wang et al., 2011), although analyses of mitochondrial DNA

variation suggest that different migrations, dating to pre- and post- Last Glacial Maximum, contributed to genetic variation observed among present-day inhabitants throughout the Plateau (Qin et al., 2010). It is also possible that neighboring groups have mixed with various Tibetan groups, although this mixture is thought to be much less substantial than the prevalence of admixture among high-altitude counterparts in the Andes (Beall, 2007). It is therefore plausible that nonoverlapping targets of selection may be attributed to genetic variation among distinct Tibetan populations.

While no specific genetic variants have yet been identified as direct targets of selection, the genomic regions identified thus far likely harbor functional variants. It is unknown whether the regions exhibiting a signal of selection harbor the same selected variant, or whether the same regions of the genome are under strong selection in the various different Tibetan groups. It is also unknown whether phenotype associations may be affected by the genetic variation in each of the groups examined. In order to evaluate whether our previously reported selection targets are identified in a different Tibetan group, we carried out analyses in a second distinct population from a different region of the Qinghai-Tibetan Plateau. We performed genomewide selection scans and determined that many selected haplotypes are also found in second population, highlighting signals that warrant further investigation as important factors of genetic adaptation to high altitude in Tibet. We also highlight differences in the prevalence of admixture among the two Tibetan groups and the value of studying particular groups from a mixed genetic background.

Materials and methods

DNA sample collection

DNA was extracted from whole blood samples for 85 individuals (nonsmokers, no chronic diseases) residing in the Tuo Tuo River region in Qinghai province (~4,500 m). Informed consent was obtained for all participants according to guidelines approved by the Institutional Review Boards at the High Altitude Medical Research Institute (Xining, Qinghai, People's Republic of China).

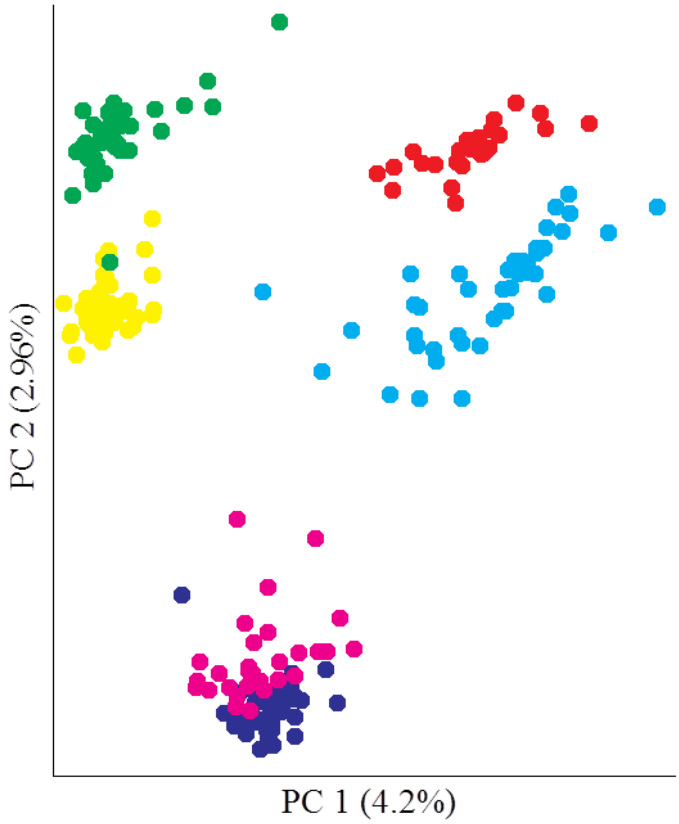
SNP genotyping

Seventy individual DNA samples were genotyped using Affymetrix 6.0 SNP Array technology (>900,000 SNPs) at Capital Bio Corporation (Beijing, China). We used default parameters for the Birdseed algorithm (version 2) to determine genotypes for all samples (Affymetrix, Santa Clara, CA, USA). Genotypic data were analyzed using the Affymetrix Genotyping Console 3.1 (Affymetrix), excluding the X and Y chromosomes and mitochondrial genome.

Principal components analysis

We performed principal components analysis (Figure 4.1) based on genetic distances as previously described (Xing et al., 2009). This analysis indicates that the two Tibetans form a distinct group but are genetically similar to each other when compared in the context of other Asian groups such as the HapMap CHB and JPT populations (CHB = Chinese in Beijing, China; JPT = Japanese in Tokyo, Japan). The CEU (U.S. Utah residents with ancestry from northern and western Europe) and YRI (YRI = Yoruba in

Figure 4.1: Principal Components Analysis of two Tibetan groups from Maduo (pink) and Tuo Tuo River (dark blue), neighboring Monogolian populations Buryat (red) and Qinghai Mongolians (light blue), and HapMap populations JPT = Japanese (green) and CHB = Chinese (yellow).



Ibadan, Nigeria) HapMap populations provide context for the patterns of variation observed among these populations (HapMap Consortium, 2007).

Estimates of relatedness

We collected samples at a small town health center where groups of semi-nomadic clans visit. We used the ERSA program to determine degrees of relationship between subjects and excluded one of any pair more closely related than second cousins (Huff et al., 2011).

***A priori* functional candidate gene list**

We generated a list of genes likely related to high-altitude adaptation based on categories provided as described in Simonson et al. (2010). Potential candidate genes identified in the mitochondrial genome and on the X chromosome were not considered for this study.

Admixture analysis

A model-based algorithm implemented in *ADMIXTURE* (Alexander et al., 2009) was used to determine the genetic ancestries of each individual in a given number of populations without using information about population designation. To eliminate the effects of SNPs that are in linkage disequilibrium (LD), we first filtered out SNPs that had $r^2 > 0.2$ within 100kb using PLINK (Purcell et al., 2007), as recommended by the authors of *ADMIXTURE*.

Selection scan analysis

We performed XP-EHH and iHS tests of selection on phased data estimated by the Beagle software package (Browning and Browning, 2009) as previously described (Simonson et al., 2010). Rather than focus on the selection candidate genes contained within 200kb regions significant at the 0.01 level in either test, we expanded our analysis to regions identified at the $p < 0.02$ level. We further excluded regions where the iHS test was significant at this level in neighboring populations as in (Simonson et al., 2010).

Genotype-phenotype association

We identified the putatively advantageous haplotypes as previously described (Simonson et al., 2010) and tested whether the three alleles exhibiting the most extreme iHS scores within each 200kb genomic region were associated with Hb in each population and both populations combined. Stepwise linear regression (MATLAB R2010a) was used to detect significant relationships between these genotypes and hemoglobin concentration.

Results

Patterns of selection in two Tibetan groups

For a direct comparison of selection signals, we employed the same analytic strategies (iHS and XP-EHH statistics) used in our previous study of Tibetan high-altitude adaptation to identify common targets of natural selection in a second linguistically distinct Tibetan population from a different region of the Tibetan Plateau (Simonson et al., 2010). The first population described in Simonson et al. (2010) is a group of Qinghai-Tibetans who speak the Amdo Tibetan dialect (one of three major

Tibetan dialects spoken on the Qinghai-Tibetan Plateau); the second population, who speak the Kham dialect, are from the Tuo Tuo River area (Figure 1.1).

Stratification in two Tibetan populations

We used pair-wise allele sharing distances to calculate F_{ST}. The two Tibetan populations exhibit the least amount of genetic differentiation from each other (F_{ST} = 0.006), and both are equally differentiated from the Han Chinese population (F_{ST} = 0.006) which has commonly been used for genetic comparisons (Table 4.1). To determine the extent of population structure among Tibetans and neighboring populations, we used the program *frappe* to examine the group-specific proportion of each individual's genome studied (Figure 4.2). When these six populations are compared, it is apparent that some of the Qinghai-Tibetans exhibit components associated with Han Chinese, suggesting recent admixture with this group. Individuals from the Tuo Tuo River population, examined here for the first time, do not, however, exhibit a signal of admixture.

Genotype-phenotype analysis for hemoglobin level

In our previous study of Qinghai-Tibetans, we identified associations between *EGLN1* and *PPARA* selected regions and [Hb], although *EPAS1* lacked enough genetic variation to test for an association signal considering the limited sample size of our first study. The relationship between [Hb] and the *EPAS1* in the second population (Tuo Tuo River) is also non-significant, although there is a trend between *EPAS1* and decreased [Hb] in this group ($p < 0.14$). There is no association signal between the [Hb] and

Table 4.1

F_{ST} between two Tibetan populations from distinct regions of the Qinghai-Tibetan Plateau and Mongolian and East Asian populations

F_{ST}						
	Buryat Mongolian	HapMap Chinese (CHB)	HapMap Japanese (JPT)	Qinghai Mongolian	Kham Tibetan	Amdo Tibetan
Buryat Mongolian	0	0.011	0.011	0.008	0.012	0.012
CHB	0.011	0	0.006	0.008	0.008	0.008
JPT	0.011	0.006	0	0.009	0.01	0.011
Qinghai Mongolian	0.008	0.008	0.009	0	0.008	0.009
Kham Tibetan	0.012	0.008	0.01	0.008	0	0.006
Amdo Tibetan	0.012	0.008	0.011	0.009	0.006	0

Figure 4.2. Admixture analysis of Tibetan groups (from Maduo and Tuo Tuo River), neighboring Monogolian populations (Buryat and Qinghai Mongolians), and HapMAp populations (JPT = Japanese; CHB = Chinese). A vertical bar is used to represent the proportion of ancestry derived from one of the K ancestral populations. Individuals are arrayed along the horizontal axis and grouped by population (JPT component: green; CHB component: yellow; Maduo Tibetan component: pink; Tuo Tuo River Tibetan component: dark blue; Qinghai Mongolian component: light blue; Buryat Mongolian component: red).

EGLN1 or *PPARA* in the second population or when the two groups are combined, suggesting that the signals detected in our first study could have been detected based on genetic variation afforded by Han Chinese admixture in the Maduo group (Figure 4.2).

Overlapping, nonprotein coding regions of the genome in both Tibetan groups

Our selection scan analyses provide evidence for adaptation in two Tibetan populations at $p < 0.02$ for seven gene regions previously identified in our study of Tibetans from Maduo (Simonson et al., 2010) or other recently published studies. The list of regions identified in both Tuo Tuo River and Maduo Tibetans is provided in Table 4.2.

In addition to the candidate list of genes we thought would likely be involved in high-altitude adaptation (Simonson et al., 2010), we were also interested in determining whether the same regions of the genome that do not contain *a priori* functional candidates exhibited a strong signal of selection in both populations using the same analytical methods. Although it is difficult to make conclusions about these signals at this time, we highlight the regions identified in both populations here (Table 4.2).

Considering the lack of differentiation detected through analysis of the protein-coding regions of Han Chinese and a Tibetan group (Yi et al., 2010), it is also possible that many genetic targets of selection are in non-coding, regulatory regions of the genome. A recent study has shown that two polymorphisms within the first intron of

Table 4.2

List of selection candidate loci
identified in Tibetans from Tuo Tuo River

Gene	Chromosome	200 kb Bin	P value	Selection scan
EGLN1	Chr1	1147	1.23E-03	XP-EHH 2009 (Han)
			7.67E-04	XP-EHH 2009 (Mongolian)
			7.68E-04	XP-EHH 2010 (Mongolian)
		1148	1.54E-04	XP-EHH 2009 (Han)
			3.07E-04	XP-EHH 2009 (Mongolian)
			6.14E-04	XP-EHH 2010 (Mongolian)
			9.83E-03	iHS 2009
			9.45E-03	iHS 2010
EPAS1	Chr2	231	1.54E-03	XP-EHH 2009 (Han)
			1.34E-02	XP-EHH 2010 (Mongolian)
		232	1.03E-02	XP-EHH 2009 (Han)
			1.15E-03	XP-EHH 2009 (Mongolian)
			7.68E-05	XP-EHH 2010 (Mongolian)
			4.09E-03	iHS 2010
PPARA	Chr22	224	9.00E-03	iHS 2009
CYP17A1	Chr10	522	7.09E-03	iHS 2009
			1.40E-02	iHS 2010
HMOX2	Chr16	22	1.29E-03	iHS 2009
			7.08E-04	iHS 2010
			1.24E-03	XP-EHH 2010 (Han)
PKLR	Chr1	767	4.33E-03	iHS 2010
			2.00E-02	iHS 2009
HFE	Chr6	130	1.43E-02	XP-EHH 2009 (Mongolian)
			1.13E-02	XP-EHH 2010 (Mongolian)
		131	1.42E-02	XP-EHH 2009 (Mongolian)
			1.37E-02	iHS 2010

EGLN1 are found at elevated frequency (71%) in a high-altitude population from India and are associated with *EGLN1* expression and HAPE (Aggarwal et al., 2010). Although functional data are not available at this time, these SNPs are in linkage disequilibrium linked with the selected haplotypes identified in both of the populations we have examined thus far.

To investigate whether noncoding regions of the genome could be associated with high-altitude adaptation, we also report shared signals of strong selection in non-protein coding regions of the genome and determined whether the selection candidate regions we identified contain hypoxia-regulated micro RNAs (miRNAs) (Table 4.3). Of the overlapping noncoding regions, one of the ten contains regions that are highly conserved among mammals (ENCODE Consortium). One of the regions identified in our first study, within the *PPARA* selection candidate region, contains a hypoxia-associated miRNA (Table 4.3) and may be involved in the hypoxia response at high altitude.

Discussion

Each of the previous reports of high-altitude adaptation in Tibet was based on a range of analytical methods and different Tibetan populations located throughout the Qinghai-Tibetan Plateau. While many of the selection candidate genes reported by these various studies are the same, there are also selection targets unique to each study. By excluding the possibility of differences attributed to analytical methods, we highlight further evidence for positive selection in each of the two populations studied here at the $p < 0.01$ significance level for the iHS test of selection for *EPAS1*, *EGLN1*, *HMOX2*, and

Table 4.3

Selection candidate regions that contain hypoxia-related micro RNAs

Genes in region	Chromosome	200kb bin	P value	miRNA	Population
C21orf34	chr21	84	0.05	mir125b2; let7c	Maduo
PPARA,C22orf26	chr22	224	0.009	let7a3	Maduo

CYP17A1 (Table 4.2). While *EPAS1 EGLN1* (Bigham et al., 2010; Peng et al., 2011; Simonson et al., 2010; Wang et al., 2011; Xu et al., 2011; Yi et al., 2010) and *EGLN1* have been identified in various other studies, it is interesting to note that *HMOX2* was also reported as a selection candidate in a separate study of high-altitude adaptation in a sample of 50 pooled Tibetan samples collected from various regions of the Plateau (Peng et al., 2011). Furthermore, *PKLR*, and *HBB* and *HBG2*, identified in one of the populations studied here at the $p < 0.02$ level, were also reported as top selection candidates in an independent assessment of protein-coding regions of Tibetan genomes (Yi et al., 2010).

The *EPAS1* and *EGLN1* genes are HIF pathway members that are associated with hemoglobin levels in different Tibetan populations. Previous studies report associations with *CYP17A1* and hypertension in European and Asian populations (Liu et al., 2011). *HMOX2*, a heme oxygenase involved in HIF-independent hypoxia sensing, may also play an adaptive role with respect to hypoxia tolerance. While the other two selection candidates, *CYP17A1* and *HMOX2*, are not associated with the phenotypes measured here, they should be considered as strong candidates for future studies of genetic adaptation. Our analysis of strong selection in nonprotein coding regions of the genome also warrant further investigation because of the lack of functional evidence within coding regions of the Tibetan genome (Yi et al., 2010).

These differences suggest that background genetic variation could influence the extent to which selection signals are captured (the early stages of a selection sweep versus those that are fixed or nearly fixed in the population, see Fig. 1.2, Table 4.1) and possibly influence the potential for selection events to occur in different groups. It will

be necessary to fully characterize the functional variants in these regions and better understand population history parameters and differences in phenotypic relationships in order to address such issues. Furthermore, it should be determined whether variation caused by admixture allows for detection of genotype-phenotype relationships observed in the Maduo population.

Tibetan population differences may reflect isolation of the Tibetan groups and/or gene flow from neighboring populations. These differences influence the amount of genetic variation available to test for association between genotype-phenotype, and may underlie the ability to detect associations with phenotypes. Furthermore, the regions identified in many of the recent selection studies may be associated with various unmeasured traits and perhaps secondary to the actual phenotype that affords a selective advantage. Further characterization of other physiological traits is required to determine if the remaining selection targets highlighted here are of significance.

References

Aggarwal, S., Negi, S., Jha, P., Singh, P.K., Stobdan, T., Pasha, M.A.Q., Ghosh, S., Agrawal, A., Consortium, I.G.V., Prasher, B., *et al.* (2010). EGLN1 involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. Proceedings of the National Academy of Sciences *107*, 18961-18966.

Aldenderfer, M. (2011). Peopling the Tibetan Plateau: Insights from archaeology. High Altitude Medicine & Biology *12*, 141-147.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research *19*, 1655-1664.

Beall, C.M. (2007a). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences 104, 8655-8660.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., *et al.* (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proceedings of the National Academy of Sciences *107*, 11459-11464.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R., Scherer, S.W., Julian, C.G., Wilson, M.J., López Herráez, D., *et al.* (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics *6*, e1001116.

Browning, B.L., and Browning, S.R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics *84*, 210-223.

Encyclopædia Britannica (2011). Tibetan language.

HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

Huff, C.D., Witherspoon, D.J., Simonson, T.S., Xing, J., Watkins, W.S., Zhang, Y., Tuohy, T.M., Neklason, D.W., Burt, R.W., Guthery, S.L., *et al.* (2011). Maximum-likelihood estimation of recent shared ancestry (ERSA). Genome Research *21*, 768-774.

Liu, C., Li, H., Qi, Q., Lu, L., Gan, W., Loos, R.J., and Lin, X. (2011). Common variants in or near FGF5, CYP17A1 and MTHFR genes are associated with blood pressure and hypertension in Chinese Hans. Journal of Hypertension *29*, 70-75.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, *et al.* (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular Biology and Evolution *28*, 1075-1081.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., *et al.* (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics *81*, 559-575.

Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., Lu, Y., Zheng, H., Zhu, D., Fei, D., *et al.* (2010). A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. American Journal of Physical Anthropology *143*, 555-569.

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., *et al.* (2010). Genetic evidence for high-altitude adaptation in Tibet. Science *329*, 72-75.

Wang, B., Zhang, Y.-B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L., Li, X., *et al.* (2011). On the origin of Tibetans and their genetic basis in adapting high-altitude environments. PLoS One *6*, e17002.

Xing, J., Watkins, W.S., Witherspoon, D.J., Zhang, Y., Guthery, S.L., Thara, R., Mowry, B.J., Bulayeva, K., Weiss, R.B., and Jorde, L.B. (2009). Fine-scaled human genetic structure revealed by SNP microarrays. Genome Research *19*, 815-825.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., *et al.* (2011). A genome-wide search for signals of high altitude adaptation in Tibetans. Molecular Biology and Evolution.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., *et al.* (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75-78.

CHAPTER 5

CONCLUSIONS AND PERSPECTIVES

Thesis summary

Native highland populations provide an excellent opportunity to study genetic differentiation and adaptation. Tibetan, Ethiopian, and Andean populations appear to have developed adaptive mechanisms that ensure survival despite the decreased availability of oxygen due to hypobaric hypoxia (Beall, 2007a). The long-standing hypothesis that genetic adaptations enhance their survival is now supported by the research presented here and various recent genomewide studies of selection in Tibetans.

Many genomic regions appear to afford an advantage in this extreme environment, and some of these loci are associated with variation in hypoxia-related phenotypes. Three genes within regions targeted by selection, *EPAS1*/*HIF2A* (Beall et al., 2010; Yi et al., 2010), *EGLN1*/*PHD2*, and *PPARA* (Simonson et al., 2010), are significantly associated with a Tibetan-specific decreased-hemoglobin phenotype in different Tibetan groups, providing corroborative evidence for their role in adaptation. Further evidence for functional roles of *EPAS1*/*HIF2A* and *PPARA* is given by associations with lactate and free fatty acids levels, respectively, suggesting complex interplay of genes linked to various physiological systems. While many signals of selection and phenotypic associations are shared among Tibetan groups located throughout the Qinghai-Tibetan Plateau, some are not. Furthermore, associations detected between selection candidate genes and phenotypes appear unique to particular Tibetan groups, suggesting differences among Tibetan populations located throughout the plateau.

Taken together, this work provides insight into the genetic factors that underlie adaptation to the physiological challenges imposed by hypoxia in Tibetans. The nature of

these adaptive mechanisms, however, remains largely unknown. Efforts to characterize the precise variants and molecular mechanisms underlying this adaptation will have implications for understanding human biological processes, including hypoxia tolerance, which involves a variety of cardiovascular, pulmonary and metabolic functions.

By coupling information about regions of the genome subject to natural selection with phenotypic data collected from Tibetan populations, we have started to bridge the gap between genetic factors and phenotypic traits exhibited in high-altitude populations. Studies of populations that have historically lived in extreme conditions, such as hypobaric hypoxia at high altitude, will enhance the current understanding of adaptation in our species and help elucidate medically relevant mechanisms of biological homeostasis in humans.

Summary and interpretation of results

We hypothesized that genes subject to natural selection that are specifically related to hypoxia sensing would be associated with phenotypes in native high-altitude Tibetan populations. To test this, we performed selection scans and compared putatively advantageous haplotypes to phenotypic data collected in two Tibetan populations. We identified many hypoxia-related genes in our selection candidate regions, and relationships between genotype and phenotype for two of these regions.

Recent genome-wide analyses, including the research presented here, provide the first lines of evidence in support of genetic adaptation to high altitude (Beall et al., 2010; Bigham et al., 2010; Peng et al., 2011; Simonson et al., 2010; Wang et al., 2011; Xu et al., 2011; Yi et al., 2010). The analytical tools used for these analyses pinpoint: 1) extreme allele frequency differences between highland and lowland populations, and/or

2) regions of the genome that exhibit a striking signal of haplotype homozygosity which results from a rapid increase in the frequency of selected variant(s) and their neighboring polymorphisms (Figure 1.2).

Extreme signals of selection are apparent among regions of Tibetan genomes that contain key members of the HIF pathway. Furthermore, three of these genes are significantly associated with decreased hemoglobin (Hb) levels in Tibetans. Single nucleotide polymorphisms (SNPs) in the haplotype regions containing *EGLN1/PHD2* (a negative regulator of HIF) and *PPARA* (a transcriptional target of HIF) exhibited relationships with lower Hb levels in our first study (Simonson et al., 2010) (Table 4.2); and SNPs in the *EPAS1/HIF2A* locus (which encodes the oxygen-regulated a subunit of the HIF-2 transcription factor) showed a significant relationship with Hb in two studies that tested for an association between this gene and Hb levels (Beall et al., 2010; Yi et al., 2010). These association signals provide corroborative evidence for adaptive roles of genes that appear to have undergone strong positive selection. Several other genes, which have as of yet not been associated with phenotype, were independently identified in more than one study (Table 5.1; Figure 5.1), suggesting that there are shared signals of selection in populations located throughout the Qinghai-Tibetan Plateau (Chapter 4).

Additional studies of high-altitude adaptation

Several studies of Tibetan high-altitude adaptation report selection signals for the *EPAS1/HIF2A* gene (Table 1.1). Our first study identified *EPAS1/HIF2A* as a selection candidate gene by coupling *a priori* functional candidate gene information with a list of genes identified by haplotype-based tests of selection. These statistics measure the decay

Table 5.1: Selection candidate genes identified in more than one genome-wide study of high-altitude adaptation or associated with hemoglobin levels in different Tibetan populations. The list of all selection candidate genes reported thus far is not provided here as it is likely that many are false positives, which are inherent to genome-wide tests of selection. Studies are numbered 1: (Simonson et al., 2010); 2: (Beall et al., 2010); 3: (Yi et al., 2010); 4: (Bigham et al., 2010); 5: (Peng et al., 2011); 6: (Xu et al., 2011); 7: (Peng et al., 2011)*Identified in the top two percent of Simonson et al. selection candidates (see (Simonson et al., 2010)).

	Study								
Gene region	1	2	3	4	5	6	7	Associated with Hb level in Tibetans	Target of selection in Tibetans and Andeans
EPAS1/HIF2A	x	x	x	x	x	x	x	Studies 2 and 3	
EGLN1/PHD2	x		x	x	x	x	x	Study 1	Study 4
HBB, HBG2	x*		x						
HMOX2	x				x				
PKLR	x*		x						
PPARA	x							Study1	

Figure 5.1: Map of sample locations for each of the seven published studies of high-altitude adaptation in Tibet (location for Study 4 (Bigham et al. 2010) reported as Tibet). Numbers correspond to the study number listed in Table 5.1.

of linkage disequilibrium (LD) from polymorphic sites across the genome and identify extended haplotype blocks of homozygous single nucleotide polymorphisms (SNPs) that result from a rapid increase in the frequency of selected variant(s) and neighboring polymorphisms (Figure 1.2). The *EPAS1/HIF2A* region exhibited a significant signal for a cross-population test for natural selection (XP-EHH (Sabeti et al., 2007)), which is particularly powerful for detecting selection events that have produced high-frequency (fixed or nearly fixed) variants, between a set of Han Chinese and Japanese individuals (Consortium, 2007) and Tibetan high-altitude dwellers from a small village in the northeastern region of the Qinghai-Tibetan Plateau (Maduo) and a second village approximately 1000 kilometers away in a village referred to as Tuo Tuo River (Figure 1.3).

In addition to the strong selection signals identified in the research presented here (Chapters 2 and 4), two additional studies identified *EPAS1/HIF2A* as a target of selection and determined that variants in this gene are associated with lower Hb concentration in different Tibetan samples (Beall et al. 2010; Yi et al. 2010). Beall et al. first compared more than 500,000 SNPs in a Yunnan Tibetan population (Figure 1.3) to publicly available Han Chinese SNP data (Hap Map Consortium, 2007) in a genome-wide allele differentiation scan (GWADS) and identified outlier SNPs in the *EPAS1/HIF2A* gene (Beall et al., 2010). These SNPs were genotyped in two additional Tibetan groups, and a subset of eight SNPs was highly associated with Hb concentration in both groups (Beall et al., 2010). Rather than examining SNPs dispersed throughout the entire genome, Yi et al. performed exome sequencing to capture sequence information for 92% of protein-coding genes (exomes) in 50 Tibetan subjects (Yi et al., 2010). Using a

population branch statistic (PBS) based on pairwise comparisons of allele frequencies between two Tibetan groups, Han Chinese, and a Danish outgroup, Yi and colleagues also identified a signal at the *EPAS1/HIF2A* locus. The largest allele-frequency difference detected between Han Chinese (9%) and Tibetans (87%) in this study was an intronic SNP within *EPAS1/HIF2A*, captured on the exome-targeted array (Yi et al., 2010). Congruent with the Beall et al. study, Yi and colleagues examined SNPs in the *EPAS1/HIF2A* locus and identified an association between SNPs in this region and lower Hb concentration (Beall et al., 2010; Yi et al., 2010). Based on further analysis of the exome data, Yi et al. concluded that selection at the *EPAS1/HIF2A* locus occurred rapidly in Tibetans. The Tibetan-Han divergence time (2750 years) reported by Yi and colleagues (2010) is, however, highly debated based on archaeological data collected from the Tibetan Plateau (Aldenderfer, 2011; Brantingham et al., 2010).

We did not detect an association between the *EPAS1*/*HIF2A* locus and Hb level in the second population we studied from Tuo Tuo River (Figure 1.3), although it is possible that there is not enough power to detect this signal based on the sample size. We did, however, identify an association between this selection candidate gene and lactate levels in the Tuo Tuo River population (Chapter 3). Further studies of this population, and direct comparisons with the as of yet unidentified selected variant, may be used to determine whether an association with the Hb phenotype exists in this population.

Additional cross-population studies performed by Bigham et al., Xu et al., Peng et al., and Wang et al. also indicate a selection signal for *EPAS1/HIF2A* in Tibetan populations, although no phenotypic data were analyzed. Bigham et al. (2010) examined more than one million SNPs across the genomes of both Tibetan and Andean populations

and performed a battery of genome-wide selection tests in both high-altitude groups. They determined that a genomic region containing *EPAS1/HIF2A* exhibited significant variation between Tibetans and HapMap Asians but not Andeans (Bigham et al., 2010). Xu and colleagues determined that several SNPs within the *EPAS1*/*HIF2A* region in 50 Tibetans collected from five different areas in Tibet (Figure 5.1) exhibit significant differences compared to Han Chinese using two other genome-wide tests for selection: F_{ST} (Cockerham and Weir, 1984) and the cross-population statistic XP-CLR (Chen et al., 2010), which detects selective sweeps in one of two populations based on allele frequency differences at multiple linked SNP loci (Xu et al., 2011). Xu et al. also reprocessed and analyzed Qinghai-Tibetan SNP data from our first study (Chapter 2) using these statistics and identified a signal of selection at the *EPAS1*/*HIF2A* locus. Xu and colleagues (2011) also identified a significant allele-frequency-based signal of differentiation at this locus, further supporting a strong signal of selection at *EPAS1/HIF2A* in the Qinghai-Tibetan population.

Peng et al. used the same statistics employed by Xu et al. (2011) and also identified the *EPAS1/HIF2A* selection candidate by comparing Han Chinese and 50 individuals randomly selected from seven different Tibetan populations located throughout the Qinghai-Tibetan Plateau (Figure 1.3). The Peng et al. sequence analysis of the *EPAS1*/*HIF2A* gene showed that the frequencies of more than half of the SNPs compared in this region were significantly different between Han Chinese and Tibetans (Peng et al., 2011). Using this sequence information, Peng et al. also estimated a divergence time between these populations that is six-fold greater than the estimate provided by the exome study (Yi et al., 2010). It is interesting to note that despite major differences in

allele frequencies between the Tibetan and Chinese at the *EPAS1/HIF2A* locus, neither the exome study (Yi et al., 2010) nor the *EPAS1/HIF2A* sequence analysis (Peng et al., 2011) identified evidence for functional coding variants, suggesting regulatory variation could underlie adaptive function at this locus.

The most recently published Tibetan genomewide selection study (Wang et al. 2011) provided results from ten tests of selection using genomewide SNP data collected from a Tibetan region located near Lhasa: three haplotype-based tests (iHS, and XP-EHH based on comparisons with HapMap Chinese and Japanese populations (Consortium, 2007)) and seven F_{ST} analyses based on comparisons with HapMap and East Asian HGDP populations (Li et al., 2008; Peng et al., 2011). *EPAS1/HIF2A* was identified among more than 50 selection candidate genes reported in this study and exhibited a significant selection signal for each of the tests performed (Peng et al., 2011).

Extreme patterns of genetic differentiation between Tibetans and other populations at the *EPAS1/HIF2A* locus provide strong evidence for selection in this genomic region (Beall, 2011; MacInnis and Rupert, 2011; Scheinfeldt and Tishkoff, 2010). The identification of *EPAS1/HIF2A* in seven different studies suggests it is common throughout the Qinghai-Tibetan Plateau and amenable to detection by various genome-wide tests of selection.

While the *EPAS1/HIF2A* statistical signal is clear, the precise genetic variant(s) and functional role(s) remain unknown. Future studies will be necessary to determine how *EPAS1/HIF2A* contributes to adaptation to high altitude. Such studies will help determine whether adaptive changes are related to the varied roles , HIF2α is known to

play, including regulation of erythropoietin, in the human response to hypoxia (Tissot van Patot and Gassmann, 2011).

A second selection candidate identified in both our Maduo and Tuo Tuo River studies (Simonson et al., 2010) and others (Bigham et al., 2010; Peng et al., 2011) is *EGLN1/PHD2*, which encodes prolyl hydroxylase 2 (PHD2), an oxygen-dependent modulator of HIFα subunits. The *EGLN1/PHD2* selection candidate region was significantly associated with differences in Hb concentration in our study of Tibetans from Maduo (Simonson et al., 2010) (Chapter 2), but not Tuo Tuo River Tibetans (Chapter 4) (Figure 1.3), and was not examined for a genotype-phenotype association with Hb levels in other studies. While not all studies of Tibetan adaptation identified *EGLN1*/*PHD2* as a selection candidate gene, reports of genetic adaptation in New World high-altitude populations suggest that an *EGLN1/PHD2* haplotype, distinct from that identified as a selection target in Tibetans, exhibits a signature of selection in high-altitude Andeans (Table 1.1) (Bigham et al., 2010; Bigham, 2009).

No functional variants have been published for *EGLN1*/*PHD2* thus far; however, polymorphisms in the first intron of this gene are found at elevated frequency in a high-altitude population in India and are associated with *EGLN1*/*PHD2* expression levels and incidence of high-altitude pulmonary edema (HAPE) in this population (Aggarwal et al., 2010). However, our recent preliminary studies identified a high frequency variant in the first exon of *EGLN1*/*PHD2* in Tibetans (Lorenzo et al. abstract submitted). Efforts to characterize the genetic and functional underpinnings at the *EGLN1*/*PHD2* locus and to determine whether there are precise genetic similarities or differences among Tibetan and

Andean high-altitude populations will help to answer long-standing questions in the field of high-altitude research (Beall, 2000; Beall, 2007b; Bigham et al., 2010; Moore, 2001). While *EGLN1/PHD2* and *EPAS1/HIF2A* regulate various aspects of tissue and cellular homeostasis, the PHD/VHL/HIF axis is directly involved in erythropoiesis (Furlow et al., 2009; Lee, 2008). It is unclear whether coordinated signaling mediated by these adaptive HIF pathway members directly or indirectly affects Hb level (Beall, 2011; Storz et al., 2010), as the precise variants and their functions remain to be determined. Attempts to understand molecular mechanisms of these genes are further complicated, as hypoxia is not the only regulator of HIF. Iron, oxygen radicals, nitric oxide, heat shock proteins (HSP), spermidine/spermine-N-acetyltransferase-2 (SSAT2) (Semenza, 2010), in addition to other factors such as epigenetic modifications may also play an important role in HIFs' regulation. Furthermore, selection imposed by environmental extremes (other than hypoxia) at high altitude may also influence, or be influenced by, altered HIF pathway activity. While a discussion of these factors is outside of the scope of the research presented here, they may also contribute to the complexity of HIF regulation that appears to be uniquely coordinated in Tibetan populations.

Metabolic challenges to high-altitude hypoxia

In addition to selection targets directly involved in the hypoxia-sensing pathway, selection studies have identified genes that encode members of metabolic pathways that respond to HIF signaling (Semenza, 2011). *PPARA* (the nuclear receptor peroxisome proliferator activated receptor alpha), a master regulator of fatty acid oxidation, is identified as a selection candidate and is associated with decreased hemoglobin levels in our first study of Tibetans from Maduo (Chapter 2) (Simonson et al., 2010). It is possible

that inherited alterations in genes that regulate metabolism, such as *PPARA* and other downstream targets, may compensate for wide-ranging changes inherent to global alterations in the HIF pathway.

The most obvious adaptive metabolic change induced by hypoxia is a conversion from oxidative glucose metabolism to glycolysis in order to maintain energy production, which has been well described and shown to depend largely upon HIF signaling (Denko, 2008; Semenza, 1999). Certain adaptive changes, such as interrupting normal HIF-mediated hypoxic signaling in order to limit possibly maladaptive increases in erythrocyte mass (Prchal, 2010), may need to be balanced by other changes (e.g., rescuing downstream hypoxia-mediated regulatory changes that would otherwise have been abrogated by any global changes in hypoxia signaling). Alternatively, orchestrated changes in the hypoxia-sensing pathway may allow adaptation through various aspects of tissue-specific and cellular homeostasis whereby Hb level is a secondary outcome (Storz et al., 2010).

Shared and unique selection candidate genes identified in studies of Tibetan adaptation

Additional selection targets that have not yet been associated with phenotypes may also be biologically relevant for Tibetan adaptation. We report selection candidate genes associated with phenotypes in addition to those identified in our studies and at least one additional study in Table 5.1, and prioritize these genes for future investigation. The genomic region containing *HMOX2*, a heme oxygenase, reported as a top selection candidate in both of our studies, was also identified by Peng et al. (2011). Two exome candidate regions reported by Yi et al. (2010) also contain genes identified within the top

two percent of the selection candidate list in our first study: *HBB* and *HBG2*, in the hemoglobin beta locus, and *PKLR*, which is involved in red blood cell maintenance. It should be noted, however, that no physiologically associated variants have yet been reported among Tibetans for any of the selection candidate genes discussed here.

While many of the selection signals converge, there are also selection targets unique to each study. Some of these differences may be attributed to genetic variation among Tibetan populations. At least four major Tibetan dialects are spoken (Britannica, 2011), suggesting long-term population isolation and/or influence from neighboring populations that may result in genetic differentiation among groups. While an analysis of genome-wide SNP variation supports a model in which the Tibetan Plateau was colonized via a single migration route, it is recognized that populations located throughout the Tibetan Plateau are likely admixed with neighboring populations (Peng et al., 2011). Recent analyses of mitochondrial DNA suggest that significant differentiation, dating to pre- and post- Last Glacial Maximum, can be found among present-day inhabitants of the Tibetan Plateau (Beall et al., 2010). It is therefore likely that populations located throughout this vast region have experienced different population histories and various degrees admixture of different neighboring populations (Zhao et al. 2009; Qin et al., 2010; Wang et al., 2011; Aldenderfer, 2011). Such differences could account for the ability to detect selection signals in some groups but not others (Wills 2011).

Differences in analytical methods, and their power to identify specific signals (Holsinger and Weir, 2009; Sabeti et al., 2007), could also help to account for some of the variation among different studies. Tests based on extreme differences in allele

frequencies are useful for identifying major single-site differences between populations, in contrast to within-population, haplotype-based sweeps that may be present and strongly detectable at intermediate allele frequencies (Huff et al., 2010; Voight et al., 2006). One example is the *PPARA* haplotype, which exhibits a significant iHS signal and is correlated with hemoglobin concentration in the Qinghai-Tibetan population (Simonson et al., 2010) but has not been reported as a selection candidate in other studies. It is possible that natural selection on *PPARA* is either unique to the Qinghai-Tibetan population or has not been identified in other populations due to methodological differences.

Peng and colleagues (2011) genotyped three *EPAS1*/*HIF2* outlier SNPs identified in their Tibetan samples and three haplotype SNPs reported by Simonson et al. (2010) for *EGLN1*/*PHD2* and *PPARA* in more than 1334 individuals from seven Tibetan groups, and conclude that three, one, and none of the respective SNPs exhibit highly differentiated allele frequencies between Tibetans and Han Chinese and Japanese populations (Peng et al., 2011). These results should, however, be interpreted with caution. Positive selection may not always result in extreme allele frequency differences (Holsinger and Weir, 2009; Sabeti et al., 2007). Haplotype tests of selection, for example, are based upon patterns of extended homozygosity across a genomic region rather than differences in allele frequency. The within-population integrated haplotype score (iHS, Voight et al., 2006), which was used by Simonson et al. (2010) to determine *EGLN1*/*PHD2* and *PPARA* haplotype SNPS, has the greatest power to identify selective sweeps when the selected variant is present at intermediate (20%-80%) frequency in a single population (Huff et al., 2010; Voight et al., 2006). Extreme allele frequency

differences are, therefore, not necessarily expected to be observed among a subset of SNPs used to tag a region that exhibits an intermediate selective sweep signal. While it remains to be seen whether signals identified in a signal study are true signals of selection or false positives, the Peng et al. (2011) analysis highlights important concepts about differences in analytical strategies that have been employed in various studies.

Future studies

Identification of specific genetic variants targeted by selection

It is clear that evolutionary processes have favored specific hypoxia- and metabolism-related genetic factors in high-altitude Tibetans. The associations between several of the selected candidate genes and a decreased hemoglobin phenotype provide supporting evidence for functional roles of genetic variants specific to high-altitude Tibetans.

Although analysis of the protein-coding regions of the genome has been successful in identifying disease-causing mutations for a number of Mendelian conditions (Ng et al., 2010; Lupski et al., 2010; Ku et al., 2011), there is good evidence that many or most variants associated with complex traits and diseases may lie in noncoding regions (Hindorff 2009) that would be missed by examination of protein-coding regions alone. Such noncoding regulatory variants are increasingly recognized as important factors of adaptive evolution (Wray et al., 2007; Kudaravalli et al., 2009; Lappalainen et al., 2010). For example, strong positive selection for hereditary lactase persistence has been attributed to regulatory variants located ~10kb upstream of the *LCT* gene in two separate human populations (Tishkoff et al., 2007). A recent study of 50 high-altitude Tibetan exomes failed to identify potential causal variants in genes exhibiting patterns of highly

differentiated allele frequencies in Tibetans and Han Chinese: the SNP with the greatest allelic-frequency difference between these groups is an intronic variant in the *EPAS1*/*HIF2A* locus (Yi et al., 2010). To avoid missing functional variants that may be located outside coding portions of the genome, it would be ideal to carry out whole-genome sequencing for a more complete picture of potentially functional variation.

With the advent of high-throughput sequencing technology, it is now possible to sequence a single genome for less than 5,000 dollars, and this price is estimated to rapidly decline in the next few years (Mardis, 2011). Even with this present estimate in mind, it is conceivable that we one could obtain important variant information through careful selection of only a few samples, chosen to maximize the information obtained from many putatively advantageous regions. Using haplotype information gathered from the chip analysis, samples that in combination provide the most information about selected and nonselected haplotypes could be selected for sequence analysis. The regions of greatest interest that should be prioritized for this selection process include those identified in more than one of our studies (Table 5.1).

Further consideration will be given to potential regulatory variants that may be involved in adaptive evolution. To filter these targets for functional significance, the variants identified in sequence analysis will be compared with a subset of microRNAs (miRNA) that have been shown to be associated with hypoxia response (Kulshreshtha et al., 2008). The entire putatively advantageous haplotype should be compared with specific nongenic ENCyclopedia of DNA Elements (ENCODE) regions that have been well-characterized in human cell lines and are highly conserved across species or exhibit methylation/chromatin modifications (ENCODE Consortium, 2007). This comparison

will provide a list of potentially functional noncoding variants important for high-altitude adaptation.

Sequence data will also provide valuable insight into the population history of Tibetans and may be used to define a parameter set that is required to perform selection scans on whole-genome sequence data. The CMS test (Grossman et al., 2010) is a newly developed statistical method designed specifically to identify genetic variants that have been targeted by recent positive selection. CMS relies on a well-calibrated model of demographic history to incorporate information from five or more individual tests of selection in a composite likelihood framework. Previous results predict that the CMS test will reduce our search space from approximately 1,000 to approximately 100 candidate variants in each region and, in half of the cases, the true causal variant is found among the top 10 candidate variants (Grossman et al., 2010), providing valuable prioritization information within each candidate list.

Once a handful of highly informative genomes or a sample size conducive to demographic analysis and subsequent genomewide selection scans is obtained, it will be possible to identify which coding and noncoding variant(s) are harbored within regions exhibiting a strong signal of positive selection in Tibetans. This information is crucial for follow-up studies necessary to understand the functional mechanisms of this process.

Perspectives

Teasing apart mechanisms of high-altitude adaptation Tibetans

While the work presented here and others' studies provide a step forward in the field of high-altitude research, these data raise many questions regarding the mechanisms involved in human adaptation to hypoxia. In order to understand this process, it will be

necessary to integrate these recent genetic advances with current knowledge of relevant molecular pathways and the many physiological insights afforded by decades of high-altitude research. Once the precise genetic variants are identified, functional analysis will be necessary to elucidate the precise mechanisms underlying hypoxia tolerance.

Our analyses indicate that *PPARA* is also associated with increased triglyceride levels and may be related to shifts in energy fuel choice (fatty acid to carbohydrate oxidation) in order to compensate for decreased oxygen availability (Holden et al., 1995; McClelland et al., 1998; Roberts et al., 1996). Misregulation of a key HIF pathway member has recently been shown to influence metabolic phenotypes (blood lactate levels, phosphocreatine levels, and acidosis in muscle) and exercise capability (Formenti et al., 2010), suggesting a strong connection between hypoxia sensing and metabolism, although further research will be necessary to directly test these relationships.

While our study and others' studies have focused specifically on hemoglobin (Hb) levels, which are uniquely lower among Tibetans compared to other populations at high altitude, the biological mechanisms underlying this characteristic are unknown. It is unclear whether this trait is directly targeted by selection and is itself advantageous or, alternatively, whether it is a side effect of a different adaptive trait. An obvious physiological change that may provide an advantage in a low-oxygen environment is efficiency at some point of the oxygen transport pathway. Characterizing each step of the oxygen-transport system of Tibetans located in their native high-altitude environment will provide insight to both genetic associations and relationships among various physiological traits. By combining our genotype and additional phenotype data, we hope to begin to identify some of the major biological underpinnings of high-altitude

adaptation. Future efforts will be focused on investigating whether adaptation is a function of hemoglobin through examination of the oxygen transport pathway in these individuals and determine, for the first time, the precise genetic factors underlying some of the other traits unique to the Tibetan population. Thorough, complementary examinations of Tibetan genetics and physiology will provide unprecedented insight into mechanisms of high-altitude adaptation and have significant implications for understanding the human response to hypoxia tolerance and stress.

Maladaptive responses to hypoxia

Future studies will also help to elucidate whether some variants within these genes contribute to maladaptive responses that result from limited oxygen supply. It has long been observed that red blood cell production increases in response to reduced atmospheric oxygen pressure (Bert, 1878; Viault, 1890), and many visitors to high altitude develop polycythemia and concomitantly elevated hemoglobin (Hb) levels. This response causes higher blood viscosity and is associated with pulmonary hypertension, increased risk of stroke, high-altitude pulmonary edema (HAPE), and chronic mountain sickness, and it may influence *in utero* development (Beall et al., 2007; Moore, 2001). These consequences extend beyond altitude sickness and are relevant to clinical pathologies such as pulmonary hypertension, heart disease, stroke, tumor hypoxia, and metabolic disorders (Majmundar et al., 2010). Further studies of high-altitude adaptation may allow interventions to reduce common, hypoxia-induced maladies suffered by lowlanders by identifying potential new targets for more common forms of clinical pathologies.

Insight into recent human evolution

In addition to physiological and molecular aspects of high-altitude adaptation, genetic studies of native high-altitude populations will provide valuable insights into evolutionary processes. It remains to be seen how natural selection has orchestrated human biological changes and how much of this fine-tuning is based upon novel or standing genetic variation. It will also be important to determine if and how the selection candidates, some of which have been reported in more than one study (Table 5.1), are involved in high-altitude adaptation and whether differences reported in selection and association signals reflect variation in Tibetan population histories.

References

Aggarwal, S., Negi, S., Jha, P., Singh, P.K., Stobdan, T., Pasha, M.A.Q., Ghosh, S., Agrawal, A., Consortium, I.G.V., Prasher, B., et al. (2010). EGLN1 involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. Proceedings of the National Academy of Sciences 107, 18961-18966.

Aldenderfer, M. (2011). Peopling the Tibetan Plateau: Insights from archaeology. High Altitude Medicine & Biology 12, 141-147.

Beall, C.M. (2000). Tibetan and Andean patterns of adaptation to high-altitude hypoxia. Human Biology 72, 201-228.

Beall, C.M. (2007a). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences 104, 8655-8660.

Beall, C.M. (2011). Genetic changes in Tibet. High Altitude Medicine & Biology 12, 101-102.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., et al. (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proceedings of the National Academy of Sciences 107, 11459-11464.

Bert, P. (1878). La pression barometrique. G Masson, 1133-1138.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R., Scherer, S.W., Julian, C.G., Wilson, M.J., López Herráez, D., et al. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics 6, e1001116.

Bigham, A.W., X. Mao, T. Brutsaert, M. J. Wilson, C. G. Julian, E. J. Parra, J. M. Akey, L. G. Moore, M. D. Shriver (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Human Genomics 4, 79-90.

Brantingham, P.J., Rhode, D., and Madsen, D.B. (2010). Archaeology augments Tibet's genetic history. Science 329, 1467-1467.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Research 20, 393-402.

Cockerham, C.C., and Weir, B.S. (1984). Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics 40, 157-164.

Denko, N.C. (2008). Hypoxia, HIF1 and glucose metabolism in the solid tumour. Nature Reviews Cancer 8, 705-713.

ENCODE Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816.

Encyclopedia Britannica. (2011). Tibetan language.

Formenti, F., Constantin-Teodosiu, D., Emmanuel, Y., Cheeseman, J., Dorrington, K.L., Edwards, L.M., Humphreys, S.M., Lappin, T.R.J., McMullin, M.F., McNamara, C.J., et al. (2010). Regulation of human metabolism by hypoxia-inducible factor. Proceedings of the National Academy of Sciences 107, 12722-12727.

Furlow, P.W., Percy, M.J., Sutherland, S., Bierl, C., McMullin, M.F., Master, S.R., Lappin, T.R.J., and Lee, F.S. (2009). Erythrocytosis-associated HIF-2α mutations demonstrate a critical role for residues C-terminal to the hydroxylacceptor proline. Journal of Biological Chemistry 284, 9050-9058.

Grossman, S.R., Shylakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327, 883-886.

Hap Map Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-861.

Holden, J.E., Stone, C.K., Clark, C.M., Brown, W.D., Nickles, R.J., Stanley, C., and Hochachka, P.W. (1995). Enhanced cardiac metabolism of plasma glucose in high-altitude natives: adaptation against chronic hypoxia. Journal of Applied Physiology 79, 222-228.

Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. Nature Reviews Genetics 10, 639-650.

Huff, C., Harpending, H., and Rogers, A. (2010). Detecting positive selection from genome scans of linkage disequilibrium. BMC Genomics 11, 8.

Kulshreshtha, R., Davuluri, R.V., Calin, G.A., and Ivan, M. (2008). A microRNA component of the hypoxic response. Cell Death and Differentiation 15, 667-671.

Lee, F.S. (2008). Genetic causes of erythrocytosis and the oxygen-sensing pathway. Blood Reviews 22, 321-332.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100-1104.

MacInnis, M.J., and Rupert, J.L. (2011). 'ome on the range: Altitude adaptation, positive selection, and Himalayan genomics. High Altitude Medicine & Biology 12, 133-139.

Majmundar, A.J., Wong, W.J., and Simon, M.C. (2010). Hypoxia-Inducible Factors and the response to hypoxic stress. Molecular Cell 40, 294-309.

McClelland, G.B., Hochachka, P.W., and Weber, J.-M. (1998). Carbohydrate utilization during exercise after high-altitude acclimation: A new perspective. Proceedings of the National Academy of Sciences of the United States of America 95, 10288-10293.

Moore, L.G. (2001). Human genetic adaptation to high altitude. High Altitude Medicine & Biology 2, 257-279.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, et al. (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular Biology and Evolution 28, 1075-1081.

Prchal, J.T. (2010). Clinical manifestations and classification of erythrocyte disorders. In Williams Hematology, N. Kaushansky, E. Beutler, U. Seligsohn, M.A. Litchman, T.J. Kipps, and J.T. Prchal, eds. (New York: McGraw-Hill), pp. 455-462.

Roberts, A.C., Butterfield, G.E., Cymerman, A., Reeves, J.T., Wolfel, E.E., and Brooks, G.A. (1996). Acclimatization to 4,300-m altitude decreases reliance on fat as a substrate. Journal of Applied Physiology 81, 1762-1771.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913-918.

Scheinfeldt, L., and Tishkoff, S. (2010). Living the high life: high-altitude adaptation. Genome Biology 11, 133.

Semenza, G.L. (1999). Perspectives on oxygen sensing. Cell 98, 281-284.

Semenza, G.L. (2010). Vascular responses to hypoxia and ischemia. Arteriosclerosis, Thrombosis, and Vascular Biology 30, 648-652.

Semenza, G.L. (2011). Hypoxia-inducible factor 1: Regulator of mitochondrial metabolism and mediator of ischemic preconditioning. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1813, 1263-1268.

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. Science 329, 72-75.

Storz, J.F., Scott, G.R., and Cheviron, Z.A. (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. Journal of Experimental Biology 213, 4125-4136.

Tissot van Patot, M.C., and Gassmann, M. (2011). Hypoxia: Adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2α. High Altitude Medicine & Biology 12, 157-167.

Viault, F. (1890). Sur l'augmentation considérable du nombre des globules rouges dans le sang chez les habitants des hauts plateaux de l'Amérique du Sud. Comptes Rendus Hebdomaires des Seances de l'Académie des Sciences, Paris 111, 917–918.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biology 4, e72.

Wang, B., Zhang, Y.-B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L., Li, X., et al. (2011). On the origin of Tibetans and their genetic basis in adapting to high-altitude environments. PLoS One 6, e17002.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., et al. (2011). A genome-wide search for signals of high altitude adaptation in Tibetans. Molecular Biology and Evolution.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75-78.

APPENDIX A

REGIONS OF THE GENOME THAT EXHIBIT STRONG SIGNALS OF SELECTION IN NATIVE HIGH-ALTITUDE TIBETANS FROM MADUO

Table A.1

200kb genomic regions identified in the top two percent of the XP-EHH selection scan

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
PRDM2	Chr1	70	0.60	0.0196
PAX7	Chr1	94	0.65	0.0132
CLIC4,RUNX3	Chr1	125	0.65	0.0127
No genes in this region	Chr1	153	0.63	0.0163
DAB1	Chr1	288	0.69	0.0082
No genes in this region	Chr1	403	0.66	0.0114
OR10J5,OR10J1	Chr1	788	0.69	0.0082
No genes in this region	Chr1	1060	0.80	0.0023
USH2A	Chr1	1070	0.65	0.0131
CNIH3	Chr1	1114	0.64	0.0146
No genes in this region	Chr1	1117	0.69	0.0084
ITPKB,C1orf95	Chr1	1124	0.72	0.0064
TRIM67,C1orf124,GNPAT,C1orf131, ***EGLN1****,*				
EXOC8	Chr1	1147	0.86	0.0012
*TSNAX,**EGLN1***	Chr1	1148	0.96	0.0002
DISC1	Chr1	1149	0.92	0.0002
AKT3	Chr1	1210	0.65	0.0134
KLF6	Chr10	19	0.79	0.0028
DHTKD1,NUDT5,SEC61A2,CDC123	Chr10	61	0.61	0.0185
SEPHS1,BEND7	Chr10	67	0.74	0.0054
RSU1,CUBN	Chr10	84	0.65	0.0125
No genes in this region	Chr10	147	0.66	0.0110
MBL2	Chr10	271	0.60	0.0198
No genes in this region	Chr10	272	0.76	0.0039
No genes in this region	Chr10	273	0.69	0.0086
ANK3	Chr10	308	0.73	0.0055
LRRC20,NPFFR1,PPA1,SAR1A	Chr10	358	0.64	0.0149
CDH23,C10orf54	Chr10	365	0.61	0.0183
RPS24,POLR3A	Chr10	397	0.64	0.0148
CHUK,ERLIN1,CWF19L1,SNORA12,CPN1	Chr10	509	0.86	0.0011
BLOC1S2,SCD,PKD2L1,CWF19L1	Chr10	510	0.89	0.0007
*WNT8B,**HIF1AN**,SEC31B,NDUFB8*	Chr10	511	0.81	0.0017
PAX2	Chr10	512	0.60	0.0199
NCRNA00081,PDCD4,SHOC2	Chr10	563	0.78	0.0035
No genes in this region	Chr10	609	0.62	0.0170
ADAM12	Chr10	639	0.62	0.0169
*FRG2B,SYCE1,DUX4,**CYP2E1**,LOC728410,*				
LOC653544,LOC653545,LOC653543	Chr10	676	0.71	0.0074

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
HBD,OR51B2,OR51B5,HBG2,HBBP1, OR51M1,HBE1,HBG1,HBB,OR51B6, OR51B4	Chr11	26	0.63	0.0158
TRIM34,UBQLNL,TRIM6-TRIM34,OR52H1,OR52B6,TRIM6,OR52D1, UBQLN3,OR51I1,OR51Q1,OR51I2	Chr11	27	0.64	0.0150
TEAD1,RASSF10	Chr11	64	0.62	0.0169
No genes in this region	Chr11	65	0.66	0.0109
TMEM86A,SPTY2D1,PTPN5,IGSF22	Chr11	93	0.66	0.0109
PRMT3,SLC6A5	Chr11	102	0.84	0.0013
NELL1,SLC6A5	Chr11	103	0.67	0.0100
LUZP2	Chr11	124	0.71	0.0075
ALX4,EXT2	Chr11	221	0.62	0.0178
APLNR,LRRC55	Chr11	283	0.66	0.0112
KCNE3,POLD3	Chr11	369	0.74	0.0052
RNF169,CHRDL2,POLD3	Chr11	370	0.61	0.0189
C11orf87	Chr11	544	0.67	0.0102
SC5DL	Chr11	603	0.79	0.0032
BLID,LOC399959	Chr11	607	0.76	0.0042
OPCML	Chr11	659	0.62	0.0174
CCDC77,SLC6A13,JARID1A	Chr12	1	0.74	0.0052
CCDC77,NINJ2,B4GALNT3	Chr12	2	0.64	0.0140
CACNA1C	Chr12	10	0.73	0.0060
CACNA1C	Chr12	11	0.66	0.0119
BCL2L14,LOH12CR2,LRP6,MANSC1	Chr12	61	0.66	0.0118
No genes in this region	Chr12	143	0.67	0.0101
SYT10	Chr12	167	0.82	0.0015
OR10AD1,PFKM,C12orf68,ASB8	Chr12	234	0.73	0.0058
DYRK2	Chr12	331	0.80	0.0025
FGD6,VEZT	Chr12	470	0.69	0.0079
NUAK1	Chr12	524	0.62	0.0179
No genes in this region	Chr12	562	0.65	0.0136
No genes in this region	Chr12	587	0.67	0.0105
CCDC64	Chr12	594	0.75	0.0048
TMEM120B,LOC338799,HPD,SETD1B, RHOF	Chr12	603	0.63	0.0156
PSMD9,WDR66,BCL7A	Chr12	604	0.72	0.0063
No genes in this region	Chr12	631	0.62	0.0179
No genes in this region	Chr13	429	0.61	0.0181
NALCN,ITGBL1	Chr13	504	0.61	0.0195
ING1,CARKD,RAB20,CARS2	Chr13	550	0.78	0.0034
ANKRD10	Chr13	551	0.63	0.0165
MAX	Chr14	323	0.61	0.0194
RAD51L1	Chr14	340	0.68	0.0089

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
WDR21A,DPF3,RBM25,ZFYVE1	Chr14	362	0.62	0.0180
TMEM63C,KIAA1737,ZDHHC22	Chr14	383	0.66	0.0120
TTC8	Chr14	442	0.63	0.0157
FSIP1,THBS1	Chr15	188	0.64	0.0138
No genes in this region	Chr15	219	0.75	0.0044
KIAA0256,SHC4	Chr15	235	0.63	0.0159
WDR72	Chr15	257	0.62	0.0171
RFX7	Chr15	271	0.77	0.0036
MCTP2	Chr15	464	0.61	0.0183
No genes in this region	Chr15	465	0.69	0.0086
TELO2,C16orf38,UNKL,LOC283951, TMEM204,IFT140,CLCN7	Chr16	7	0.90	0.0006
HN1L,IGFALS,SPSB3,EME2,NUBP2, MRPS34, NME3,IFT140,HAGH,MAPK8IP3, CRAMP1L	Chr16	8	1.04	0.0001
NDE1,MYH11,KIAA0430	Chr16	78	0.77	0.0035
GSG1L	Chr16	139	0.74	0.0055
No genes in this region	Chr16	266	0.74	0.0049
No genes in this region	Chr16	311	0.76	0.0039
WWOX	Chr16	385	0.69	0.0080
CDH13	Chr16	408	0.62	0.0173
MLYCD,HSBP1,OSGIN1,NECAB2	Chr16	412	0.65	0.0126
MAP1LC3B,ZCCHC14,FBXO31	Chr16	429	0.65	0.0124
DHX40P,HEATR6,CA4	Chr17	277	0.74	0.0051
WIPI1,ARSG	Chr17	319	0.66	0.0115
ACTG1,BAHCC1,FSCN2,C17orf70,NPLOC 4	Chr17	385	0.66	0.0112
No genes in this region	Chr18	7	0.63	0.0153
LAMA1,ARHGAP28	Chr18	34	0.67	0.0097
No genes in this region	Chr18	54	0.62	0.0173
No genes in this region	Chr18	56	0.87	0.0009
TCF4	Chr18	255	0.74	0.0050
DSEL	Chr18	316	0.70	0.0076
No genes in this region	Chr18	348	0.71	0.0071
BRUNOL5,NFIC	Chr19	16	0.75	0.0043
C19orf29,FZR1,TBXA2R,C19orf71,GIPC3, DOHH,HMG20B,NFIC,LOC284422, C19orf28, PIP5K1C	Chr19	17	0.72	0.0065
FBN3,LASS4,CCL25	Chr19	40	0.67	0.0104
***ANGPTL4**,KANK3,RPS28,MARCH2, NDUFA7, LASS4,CD320,RAB11B*	Chr19	41	0.69	0.0083

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
RHOB	Chr2	102	0.64	0.0139
ALK	Chr2	149	0.64	0.0147
XDH	Chr2	157	0.80	0.0024
SRD5A2	Chr2	158	0.74	0.0049
BIRC6	Chr2	162	0.65	0.0122
BIRC6,TTC27	Chr2	163	0.72	0.0062
SLC8A1	Chr2	202	0.64	0.0146
No genes in this region	Chr2	215	0.79	0.0033
PRKCE	Chr2	229	0.67	0.0096
*PRKCE,**EPAS1***	Chr2	231	0.82	0.0015
ATP6V1E2,EPAS1	Chr2	232	0.67	0.0103
C2orf61,CALM2	Chr2	236	0.70	0.0079
FSHR	Chr2	246	0.64	0.0144
AFTPH,SERTAD2	Chr2	323	0.63	0.0162
MCEE,PAIP2B,MPHOSPH10	Chr2	356	0.81	0.0018
ZNF638,DYSF	Chr2	357	0.81	0.0018
LOC150568	Chr2	522	0.72	0.0066
No genes in this region	Chr2	646	0.63	0.0154
No genes in this region	Chr2	686	0.61	0.0184
GTDC1,ZEB2	Chr2	724	0.65	0.0133
ARL5A,NEB	Chr2	761	0.68	0.0096
CACNB4	Chr2	762	0.72	0.0067
XIRP2	Chr2	837	0.67	0.0098
XIRP2	Chr2	838	0.68	0.0087
DNAJC10	Chr2	916	0.80	0.0022
CYP20A1,ABI2	Chr2	1019	0.65	0.0130
PARD3B	Chr2	1028	0.61	0.0191
PARD3B	Chr2	1029	0.80	0.0022
LOC643905,MYEOV2,NDUFA10,OR6B3, OR6B2,OTOS	Chr2	1203	0.66	0.0117
No genes in this region	Chr20	21	0.63	0.0163
PLCB4	Chr20	46	0.73	0.0061
B4GALT5,PTGIS	Chr20	238	0.79	0.0029
SPATA2,SLC9A8,RNF114	Chr20	239	0.72	0.0068
TMEM189,TMEM189-UBE2V1,SNAI1,UBE2V1,RNF114	Chr20	240	0.67	0.0099
ERG	Chr21	193	0.79	0.0029
DGCR11,TSSK2,CLTCL1,DGCR2,SLC25A1, DGCR14,GSC2	Chr22	87	0.63	0.0159
CLTCL1,HIRA	Chr22	88	0.79	0.0031
KIAA1644	Chr22	215	0.73	0.0056
CELSR1	Chr22	226	0.92	0.0003
EDEM1	Chr3	26	0.67	0.0108
No genes in this region	Chr3	29	0.70	0.0077
No genes in this region	Chr3	31	0.61	0.0187
SRGAP3	Chr3	45	0.67	0.0106

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
THUMPD3,SRGAP3	Chr3	46	0.68	0.0095
SETD5,THUMPD3,LHFPL4	Chr3	47	0.64	0.0152
VGLL4	Chr3	58	0.81	0.0019
SATB1	Chr3	92	0.88	0.0008
No genes in this region	Chr3	93	0.92	0.0004
CADPS	Chr3	313	0.64	0.0141
ADAMTS9	Chr3	323	0.68	0.0093
RYBP	Chr3	362	0.63	0.0156
KIAA1407,QTRTD1,DRD3	Chr3	576	0.68	0.0088
ADPRH,TMEM39A,KTELC1,C3orf1,CD80, PLA1A,CDGAP	Chr3	603	0.76	0.0038
H1FOO,IFT122,RHO,RPL32P3,PLXND1, C3orf25,MBD4	Chr3	653	0.62	0.0166
TRIM42	Chr3	709	0.64	0.0149
No genes in this region	Chr3	726	0.64	0.0143
VEPH1,PTX3,C3orf55	Chr3	793	0.67	0.0107
DGKG,ETV5	Chr3	936	0.80	0.0021
RTP1,RPL39L,ST6GAL1	Chr3	941	0.65	0.0126
No genes in this region	Chr4	150	0.88	0.0008
PCDH7	Chr4	152	0.64	0.0137
No genes in this region	Chr4	158	0.61	0.0182
TBC1D1	Chr4	189	0.67	0.0099
SLC4A4	Chr4	362	0.64	0.0145
DMP1,MEPE,IBSP	Chr4	444	0.71	0.0069
UNC5C,BMPR1B	Chr4	481	0.66	0.0111
CENPE,BDH2,NHEDC2	Chr4	521	0.79	0.0030
No genes in this region	Chr4	522	0.62	0.0166
No genes in this region	Chr4	559	0.64	0.0153
***CAMK2D**,ANK2*	Chr4	572	0.68	0.0092
TMEM155,CCNA2,EXOSC9,ANXA5, LOC100192379,BBS7	Chr4	614	0.76	0.0040
TRPC3,BBS7	Chr4	615	0.79	0.0027
No genes in this region	Chr4	637	0.68	0.0091
No genes in this region	Chr4	742	0.68	0.0089
***EDNRA**,TMEM184C,LOC90826*	Chr4	743	0.70	0.0078
No genes in this region	Chr4	817	0.62	0.0175
C4orf43,MARCH1,TKTL2	Chr4	823	0.62	0.0172
No genes in this region	Chr4	825	0.62	0.0177
No genes in this region	Chr5	16	0.83	0.0014
No genes in this region	Chr5	23	0.61	0.0193
No genes in this region	Chr5	95	0.64	0.0136
RNASEN,C5orf22	Chr5	157	0.65	0.0129
THBS4,SERINC5	Chr5	397	0.71	0.0072
MCC	Chr5	562	0.63	0.0164
IL12B,UBLCP1	Chr5	793	0.66	0.0116
No genes in this region	Chr5	823	0.71	0.0070

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
JARID2	Chr6	77	0.64	0.0139
HLA-DRB6,HLA-DQA1,HLA-DRB1, HLA-DRB5,HLA-DQB1	Chr6	163	0.65	0.0123
IL17A,PKHD1	Chr6	260	0.64	0.0142
PRIM2	Chr6	286	0.61	0.0189
PRIM2	Chr6	287	0.81	0.0020
B3GAT2,SMAP1	Chr6	358	0.75	0.0045
No genes in this region	Chr6	434	0.63	0.0161
No genes in this region	Chr6	493	0.68	0.0092
No genes in this region	Chr6	494	0.69	0.0081
SOBP,SCML4	Chr6	540	0.61	0.0186
NKAIN2	Chr6	624	0.86	0.0012
AKAP7,EPB41L2	Chr6	657	0.75	0.0045
No genes in this region	Chr6	707	0.69	0.0085
No genes in this region	Chr6	708	0.63	0.0160
UTRN	Chr6	724	0.65	0.0132
UTRN	Chr6	725	0.90	0.0005
SAMD5	Chr6	739	0.62	0.0168
ULBP3,PPP1R14C	Chr6	752	0.65	0.0135
NOX3	Chr6	779	0.65	0.0129
MAD1L1	Chr7	9	0.64	0.0143
CYTH3,EIF2AK1,USP42,PMS2,JTV1	Chr7	30	0.61	0.0188
MACC1,ITGB8	Chr7	101	0.76	0.0042
ITGB8	Chr7	102	0.67	0.0106
BBS9	Chr7	167	0.60	0.0197
No genes in this region	Chr7	172	0.65	0.0128
HERPUD2	Chr7	178	0.66	0.0116
ELMO1	Chr7	186	0.71	0.0073
EPDR1,TXNDC3,SFRP4	Chr7	189	0.72	0.0069
ADCY1	Chr7	227	0.60	0.0200
PCLO	Chr7	411	0.82	0.0016
PCLO	Chr7	412	0.79	0.0032
SSPO,KRBA1,ZNF862,ZNF467	Chr7	745	0.63	0.0155
C7orf29,REPIN1,RARRES2,LRRC61, GIMAP8, ZNF775	Chr7	748	0.62	0.0167
CSMD1	Chr8	18	0.60	0.0196
No genes in this region	Chr8	79	0.66	0.0113
No genes in this region	Chr8	104	0.62	0.0176
INTS9,EXTL3	Chr8	143	0.71	0.0075
INTS9,KIF13B,HMBOX1	Chr8	144	0.68	0.0090
LOC286135	Chr8	149	0.76	0.0041
No genes in this region	Chr8	246	0.90	0.0005
No genes in this region	Chr8	247	0.86	0.0010
EFCAB1	Chr8	248	0.76	0.0037
No genes in this region	Chr8	327	0.66	0.0121

Table A.1 Continued

Genes in XP-EHH regions	Chromosome	200kb Region	XP-EHH value	P value
MYBL1,VCPIP1,C8orf44,SGK3	Chr8	338	0.61	0.0190
ARFGEF1,CSPP1	Chr8	341	0.75	0.0046
ARFGEF1,CPA6	Chr8	342	0.79	0.0026
CPA6	Chr8	344	0.73	0.0057
No genes in this region	Chr8	384	0.67	0.0102
No genes in this region	Chr8	639	0.73	0.0059
No genes in this region	Chr8	640	0.66	0.0119
No genes in this region	Chr8	646	0.61	0.0192
No genes in this region	Chr9	124	0.73	0.0059
TMEM215,TAF1L	Chr9	163	0.75	0.0047
MAMDC2	Chr9	359	0.80	0.0025
KLF9,SMC5,MAMDC2	Chr9	360	0.73	0.0062
SLC28A3	Chr9	430	0.62	0.0176
IARS,OGN,SNORA84,NOL8,CENPP	Chr9	470	0.64	0.0151
BARX1	Chr9	478	0.71	0.0072
PTPDC1	Chr9	479	0.74	0.0053
No genes in this region	Chr9	542	0.61	0.0193
HSDL2,KIAA1958	Chr9	571	0.66	0.0122
ZNF618	Chr9	578	0.68	0.0094
ASTN2,SNORA70C	Chr9	594	0.72	0.0065
No genes in this region	Chr9	598	0.61	0.0186

Table A.2

200kb genomic regions identified in the top two percent of the iHS selection scan

Genes in iHS regions	**Chromosome**	**200kb Region**	**P value**
XPR1	Chr1	894	0.0135
ATXN7L2,CYB561D1,AMIGO1,GPR61,AMPD2,GNAI3,SYPL2,GNAT2	Chr1	549	0.0106
No genes in this region	Chr1	964	0.0159
TCEA3,ZNF436,HNRNPR,C1orf213	Chr1	117	0.0141
No genes in this region	Chr1	959	0.0122
CACNA1E	Chr1	898	0.0095
No genes in this region	Chr1	1117	0.0135
PTPRF,ST3GAL3,JMJD2A	Chr1	219	0.0176
DISC1	Chr1	1149	0.0142
SGIP1,PDE4B	Chr1	333	0.0097
PFKFB2,C4BPB,C4BPA,FCAMR,YOD1,C1orf116	Chr1	1026	0.0115
No genes in this region	Chr1	151	0.0094
FAM163A,TOR1AIP2,IFRG15,TOR1AIP1,CEP350	Chr1	890	0.0035
*TSNAX,**EGLN1***	Chr1	1148	0.0098
No genes in this region	Chr1	963	0.0063
AGBL4,ELAVL4	Chr1	251	0.0083
No genes in this region	Chr1	344	0.0093
FCER1A,DARC,CADM3,OR10J3	Chr1	787	0.0091
GBAP,PKLR,C1orf104,SCAMP3,MTX1,KRTCAP2,TRIM46,GBA,HCN3,MUC1,FDPS,C1orf2,CLK2,THBS3,RUSC1,ASH1L	Chr1	767	0.0200
No genes in this region	Chr10	295	0.0197
No genes in this region	Chr10	286	0.0187
ATRNL1	Chr10	584	0.0009
*WNT8B,**HIF1AN**,SEC31B,NDUFB8*	Chr10	511	0.0020
ENTPD1,CCNJ,LOC100127889,CC2D2B	Chr10	488	0.0042

Table A.2 Continued

Genes in iHS regions	Chromosome	200kb Region	P value
*TRIM8,**CYP17A1**,ARL3,SFXN2,C10orf26*	Chr10	522	0.0071
MYOF	Chr10	475	0.0143
No genes in this region	Chr10	550	0.0008
FRMPD2	Chr10	245	0.0026
***PTEN**,KILLIN*	Chr10	448	0.0072
CUEDC2,C10orf95,GBF1,NFKB2,PSD, FBXL15	Chr10	520	0.0159
C10orf32,AS3MT,CNNM2	Chr10	523	0.0164
PAX2	Chr10	512	0.0001
ZNF37A,LOC100129055	Chr10	192	0.0070
ATRNL1	Chr10	585	0.0051
C10orf125,PRAP1,CALY,ECHS1,CYP2E1, SPRN,PAOX,MTG1,LOC619207	Chr10	675	0.0148
PCDH15	Chr10	278	0.0002
KIAA1217	Chr10	121	0.0039
BLOC1S2,SCD,PKD2L1,CWF19L1	Chr10	510	0.0075
ATRNL1	Chr10	586	0.0185
No genes in this region	Chr11	486	0.0166
APLNR,LRRC55	Chr11	283	0.0183
TTC17,API5	Chr11	216	0.0149
No genes in this region	Chr11	126	0.0016
QSER1,DEPDC7,PRRG4	Chr11	164	0.0089
P2RX3,PRG3,SLC43A3,RTN4RL2,SSRP1, TNKS1BP1,PRG2	Chr11	284	0.0101
NOX4	Chr11	444	0.0170
GYS2,LDHB	Chr12	108	0.0081
No genes in this region	Chr12	364	0.0104
KITLG	Chr12	437	0.0006
SOX5	Chr12	122	0.0108
BCL2L14,LOH12CR2,LRP6,MANSC1	Chr12	61	0.0028
SOX5,C12orf67	Chr12	123	0.0198
ETNK1	Chr12	113	0.0039
ANKS1B	Chr12	489	0.0173
PZP,LOC642846	Chr12	46	0.0107
No genes in this region	Chr12	362	0.0085

Table A.2 Continued

Genes in iHS regions	**Chromosome**	**200kb Region**	**P value**
No genes in this region	Chr13	336	0.0120
No genes in this region	Chr13	371	0.0080
No genes in this region	Chr13	325	0.0048
SLITRK6	Chr13	426	0.0147
No genes in this region	Chr13	357	0.0184
KLF12	Chr13	367	0.0019
No genes in this region	Chr14	434	0.0043
SYT16,FLJ43390	Chr14	308	0.0160
No genes in this region	Chr14	221	0.0110
ERH,SLC39A9,GALNTL1	Chr14	344	0.0151
RTN1,GPR135,C14orf149,C14orf100	Chr14	295	0.0089
SV2B	Chr15	447	0.0161
LBXCOR1,MAP2K5	Chr15	329	0.0018
SH2D7,CIB2,TBC1D2B	Chr15	380	0.0139
RGMA,CHD2	Chr15	456	0.0036
PRDM7,GAS8,DBNDD1,C16orf3	Chr16	443	0.0060
No genes in this region	Chr16	321	0.0038
NMRAL1,CORO7,DNAJA3,C16orf5, ***HMOX2****,FAM100A*	Chr16	22	0.0013
No genes in this region	Chr16	127	0.0176
BLMH,TMIGD1,CPD	Chr17	128	0.0052
DBF4B,CCDC43,ADAM11	Chr17	200	0.0111
GHDC,STAT5B,STAT5A,STAT3	Chr17	188	0.0137
No genes in this region	Chr17	336	0.0194
BLMH,EFCAB5,CCDC55,SLC6A4	Chr17	127	0.0186
TAF4B,PSMA8	Chr18	110	0.0034
FAM59A,MEP1B	Chr18	140	0.0191
No genes in this region	Chr18	73	0.0084
No genes in this region	Chr18	332	0.0133
No genes in this region	Chr18	89	0.0150
TCF4	Chr18	256	0.0060
GALNT13	Chr2	772	0.0073
CYP20A1,ABI2	Chr2	1019	0.0056
No genes in this region	Chr2	86	0.0012

Table A.2 Continued

Genes in iHS regions	Chromosome	200kb Region	P value
GPBAR1,C2orf62,SLC11A1,TMBIM1, PNKD,CTDSP1,ARPC2,AAMP,VIL1	Chr2	1094	0.0056
CCDC148	Chr2	794	0.0195
No genes in this region	Chr2	627	0.0124
MKI67IP,TSN	Chr2	611	0.0145
No genes in this region	Chr2	112	0.0061
ANKRD44	Chr2	988	0.0085
No genes in this region	Chr2	109	0.0079
ORMDL1,PMS1,ANKAR,OSGEPL1, ASNSD1	Chr2	951	0.0118
LYPD6	Chr2	749	0.0169
No genes in this region	Chr2	977	0.0177
MBOAT2	Chr2	45	0.0010
EFEMP1	Chr2	280	0.0154
XIRP2	Chr2	837	0.0077
No genes in this region	Chr2	632	0.0119
VSNL1,SMC6,GEN1	Chr2	88	0.0023
RHOB	Chr2	102	0.0181
No genes in this region	Chr2	286	0.0096
SULT1C3,SULT1C2,SULT1C4	Chr2	541	0.0105
ARL6IP6,FMNL2,PRPF40A	Chr2	766	0.0130
EPC2,KIF5C	Chr2	746	0.0174
ASAP2	Chr2	46	0.0078
No genes in this region	Chr2	84	0.0180
RAPH1,ABI2	Chr2	1020	0.0117
FOXA2	Chr20	112	0.0099
PARD6B,ADNP,BCAS4,DPM1	Chr20	244	0.0168
LOC100130264,SLC24A3	Chr20	95	0.0068
FAM83C,EIF6,UQCC,PROCR,MMP24	Chr20	166	0.0019
SCAND1,PHF20,EPB41L1,C20orf152	Chr20	170	0.0024
DHX35,FAM83D	Chr20	185	0.0143
UQCC,GDF5,CEP250,ERGIC3	Chr20	167	0.0022
C21orf33,AGPAT3,TRAPPC10,PWP2	Chr21	221	0.0172
C21orf7,BACH1	Chr21	147	0.0128
C21orf34	Chr21	81	0.0041
APOL4,APOL2,APOL3,APOL1	Chr22	174	0.0172

Table A.2 Continued

Genes in iHS regions	Chromosome	200kb Region	P value
PPARA,*C22orf26*	Chr22	224	0.0092
PLXNB1,CCDC51,ATRIP,TREX1,PFKFB4, FBXW12,UCN2,CCDC72,SHISA5,COL7A1	Chr3	242	0.0031
OTOL1	Chr3	813	0.0126
NAALADL2	Chr3	883	0.0101
MORC1	Chr3	551	0.0121
PA2G4P4,LEKR1	Chr3	790	0.0146
ABHD5,ANO10	Chr3	218	0.0153
SLC15A2,EAF2,ILDR1,IQCB1	Chr3	615	0.0068
No genes in this region	Chr3	374	0.0064
FLJ46210,LOC389151,LOC729627	Chr3	701	0.0125
SLC25A26,LRIG1	Chr3	332	0.0157
ZBTB20	Chr3	578	0.0155
TIPARP	Chr3	789	0.0130
No genes in this region	Chr3	220	0.0055
PPARG,TSEN2,MKRN2	Chr3	62	0.0178
WDR49,SERPINI1,PDCD10	Chr3	844	0.0196
FSTL1,NDUFB4	Chr3	608	0.0162
EPHA6	Chr3	490	0.0168
No genes in this region	Chr3	839	0.0015
No genes in this region	Chr3	695	0.0062
CADPS	Chr3	313	0.0047
ZBTB20	Chr3	579	0.0037
PAK2,SENP5,PIGZ,LOC152217,NCBP2	Chr3	990	0.0131
No genes in this region	Chr3	219	0.0014
FOXP1	Chr3	357	0.0193
MAGI1	Chr3	327	0.0114
DPPA2,DPPA4	Chr3	552	0.0082
INPP4B	Chr4	718	0.0175
TRPC3,BBS7	Chr4	615	0.0029
GRSF1,MOBKL1A,RUFY3	Chr4	359	0.0193
GAB1	Chr4	722	0.0192
No genes in this region	Chr4	492	0.0064
TMEM155,EXOSC9,CCNA2,ANXA5, LOC100192379,BBS7	Chr4	614	0.0044
No genes in this region	Chr4	742	0.0114
KCNIP4	Chr4	106	0.0058

Table A.2 Continued

Genes in iHS regions	**Chromosome**	**200kb Region**	**P value**
No genes in this region	Chr4	795	0.0134
GRID2	Chr4	468	0.0003
No genes in this region	Chr4	171	0.0158
RAPGEF2	Chr4	802	0.0005
No genes in this region	Chr4	172	0.0190
No genes in this region	Chr4	140	0.0136
SCRG1,SAP30,GALNT7,HMGB2	Chr4	872	0.0087
USP38	Chr4	721	0.0025
No genes in this region	Chr4	803	0.0004
No genes in this region	Chr4	526	0.0011
GRID2	Chr4	469	0.0032
CPEB2	Chr4	73	0.0072
INPP4B	Chr4	719	0.0053
SLIT2	Chr4	99	0.0086
No genes in this region	Chr4	593	0.0103
CXCL5,CXCL3,PPBP,CXCL2,PF4	Chr4	375	0.0189
NCRNA00099,KCNIP4	Chr4	107	0.0007
EDNRA,TMEM184C,LOC90826	Chr4	743	0.0118
No genes in this region	Chr4	170	0.0067
MARCH11	Chr5	81	0.0093
PDE4D	Chr5	294	0.0122
RNF180	Chr5	318	0.0188
RASA1	Chr5	432	0.0033
HAVCR1,TIMD4,PPP1R2P3	Chr5	781	0.0043
MCC	Chr5	562	0.0027
No genes in this region	Chr5	823	0.0021
MCC,YTHDC2	Chr5	564	0.0050
No genes in this region	Chr5	485	0.0123
CCT5,FAM173B,CMBL	Chr5	51	0.0065
No genes in this region	Chr5	586	0.0100
APC,SRP19,MCC,REEP5,DCP2	Chr5	561	0.0035
No genes in this region	Chr5	588	0.0002
ADAMTS6	Chr5	323	0.0155
CLINT1	Chr5	786	0.0045
No genes in this region	Chr5	680	0.0110
CDH9	Chr5	135	0.0197
No genes in this region	Chr5	585	0.0147
ZNF131,HMGCS1,MGC42105	Chr5	216	0.0156

Table A.2 Continued

Genes in iHS regions	Chromosome	200kb Region	P value
HCP5,MICA,HLA-B,MICB,HCG26	Chr6	157	0.0074
COL21A1	Chr6	281	0.0126
No genes in this region	Chr6	629	0.0113
No genes in this region	Chr6	417	0.0027
HMGN3	Chr6	400	0.0199
BMP5	Chr6	278	0.0144
ASF1A,FAM184A,MCM9	Chr6	596	0.0030
No genes in this region	Chr6	416	0.0127
CDYL,RPP40	Chr6	24	0.0066
GABRR1,PNRC1,SRrp35,PM20D2	Chr6	449	0.0069
SOBP,SCML4	Chr6	540	0.0180
PHACTR1	Chr6	65	0.0052
OPN5,C6orf138	Chr6	239	0.0152
CD2AP,GPR111,GPR115	Chr6	238	0.0010
PRIM2	Chr6	286	0.0017
No genes in this region	Chr6	521	0.0163
GCNT2	Chr6	53	0.0088
PRIM2	Chr6	287	0.0006
PKHD1	Chr6	258	0.0182
No genes in this region	Chr7	677	0.0138
NRF1	Chr7	645	0.0164
TRPV6,C7orf34,EPHB6,KEL,TRPV5	Chr7	711	0.0054
ZNF800	Chr7	634	0.0023
GUSB,VKORC1L1,ASL	Chr7	325	0.0057
ZPBP	Chr7	250	0.0014
PRSS1,TRY6,PRSS2	Chr7	710	0.0040
No genes in this region	Chr7	99	0.0102
No genes in this region	Chr7	425	0.0165
SDK1	Chr7	20	0.0109
GCC1,FSCN3,ARF5,PAX4,SND1	Chr7	635	0.0081
LOC401397,GPR85	Chr7	562	0.0059
EXOC4	Chr7	664	0.0139
No genes in this region	Chr7	274	0.0129
IFRD1,C7orf53	Chr7	559	0.0077
C7orf58,FAM3C,WNT16	Chr7	603	0.0106
MAD1L1	Chr7	10	0.0151

Table A.2 Continued

Genes in iHS regions	Chromosome	200kb Region	P value
No genes in this region	Chr8	247	0.0046
No genes in this region	Chr8	103	0.0184
No genes in this region	Chr8	385	0.0048
No genes in this region	Chr8	634	0.0097
No genes in this region	Chr8	384	0.0031
No genes in this region	Chr8	246	0.0049
No genes in this region	Chr8	102	0.0140
HNF4G	Chr8	383	0.0112
CHRNA6,CHRNB3	Chr8	213	0.0132
MAMDC2	Chr9	359	0.0116
LINGO2	Chr9	139	0.0171
SDCCAG3,C9orf163,SNAPC4,SEC16A, PMPCA,NOTCH1,INPP5E	Chr9	692	0.0076
NXNL2,SPIN1	Chr9	451	0.0167
WNK2,C9orf129	Chr9	475	0.0179
HSPA5,GAPVD1,RABEPK	Chr9	635	0.0188

Appendix A lists genes and identified in the top two present of genomic regions by XP-EHH and iHS selection scans in the Maduo Tibetan population.