Functional Variation in the Hypoxia-Inducible Factor (HIF) Pathway in Humans

A Thesis submitted for the degree of Doctor of Philosophy

Nayia Petousi

Balliol College

Michaelmas 2012

Abstract: Functional Variation in the Hypoxia-Inducible Factor (HIF) Pathway in Humans

Nayia Petousi, Balliol College **D.Phil. Thesis, Michaelmas 2012**

By undertaking a number of different experimental approaches at the genetic, cellular/ molecular and integrative physiology levels, I investigated functional variation in the Hypoxia-Inducible Factor (HIF) transcription pathway in humans. My studies focused on Tibetan natives.

Tibetan highlanders are adapted to life in a hypoxic environment and exhibit distinct physiological traits at high altitude. Recent studies identified positive selection at two genetic loci, *EPAS1* (HIF2α) and *EGLN1* (PHD2), in Tibetan highlanders and demonstrated an association of *EGLN1*/*EPAS1* genotype with haemoglobin concentration. Both are genes of the HIF pathway, which coordinates an organism's response to hypoxia. Patients living at sea level with genetic diseases of the HIF pathway have characteristic phenotypes at both the integrative physiology and cellular levels. I investigated whether Tibetans living at sea level also possess distinct phenotypic characteristics, and whether these may be related to underlying variation within the HIF pathway.

I compared Tibetans living at sea level with Han Chinese, their most closely-related major ethnic group, and found that Tibetans possess a significantly different integrative physiology phenotype. Tibetans had a lower haemoglobin concentration and haematocrit, a higher pulmonary ventilation relative to metabolism, and blunted pulmonary vascular responses to both acute (minutes) and sustained (8 hours) hypoxia. Regarding genotype-phenotype relationships within the Tibetans, I found a significant correlation between both *EPAS1* and *EGLN1* genotype and the induction of erythropoietin by systemic hypoxia. At an intermediate cellular level, the relative expression and the hypoxic induction of HIF-regulated genes were significantly lower in peripheral blood lymphocytes from Tibetans compared with Han Chinese. I also investigated whether the genetic variation in *EPAS1* selected for in Tibetans may be functional at the molecular level by affecting transcription of *EPAS1* in cells and whether certain coding variants in *EGLN1* found in Tibetans affect protein (PHD2) activity in cells and *in vitro*.

A small supplementary study was undertaken in patients with idiopathic erythrocytosis, who have elevated or inappropriately normal erythropoietin levels, to investigate if they have genetic alterations in the HIF system.

ΙΘΑΚΗ

Σα βγεις στον πηγαιμό για την Ιθάκη,
να εύχεσαι νάναι μακρύς ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον θυμωμένο Ποσειδώνα μη φοβάσαι,
τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις,
αν μέν' η σκέψις σου υψηλή, αν εκλεκτή
συγκίνησις το πνεύμα και το σώμα σου αγγίζει.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον άγριο Ποσειδώνα δεν θα συναντήσεις,
αν δεν τους κουβανείς μες στην ψυχή σου,
αν η ψυχή σου δεν τους στήνει εμπρός σου.

Η Ιθάκη σ' έδωσε τ' ωραίο ταξείδι.
Χωρίς αυτήν δεν θάβγαινες στον δρόμο.
Άλλα δεν έχει να σε δώσει πια.

Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε.
Έτσι σοφός που έγινες, με τόση πείρα,
ήδη θα το κατάλαβες οι Ιθάκες τι σημαίνουν.

Κωνσταντινος Π. Καβαφης, 1911

ITHAKA

As you set out for Ithaka
hope your road is a long one,
full of adventure, full of discovery.
Laistrygonians, Cyclops,
angry Poseidon—don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians, Cyclops,
wild Poseidon—you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

Ithaka gave you the marvelous journey.
Without her you wouldn't have set out.
She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you'll have understood by then what these Ithakas mean.

***Constantine.P. Cavafy, 1911,* source: Collected Poems (Princeton University Press, 1975)**

To Antonis

Acknowledgements

There are many people that have contributed into making my journey interesting and enjoyable.

First, I would like to thank all the volunteers whose participation in the studies made this work possible. I also thank the Tibetan Community in Britain for their warm welcoming and their help during the recruitment of participants.

I am grateful to the Wellcome Trust and the Wellcome Trust PhD Programme for Clinicians in Basic Science at the University of Oxford for giving me the opportunity to undertake this work by funding me and my research through the award of a Wellcome Trust Clinical Training Fellowship.

I was fortunate to have the chance to work in two outstanding laboratories, the Robbins human physiology lab and the Ratcliffe molecular physiology lab. I am thankful to all the members of both laboratories, past and present, for making the work environment friendly, warm and pleasant. I am particularly grateful to Quentin Croft, Federico Formenti, Nick Talbot, Vincent Cheng, Koji Ishida – who visited from Japan for a year – Mary Slingo and Annabel Nickol for their help with the physiology studies and David O' Connor for his unique technical expertise. I am deeply indebted to Lina Sciesielski and Norma Masson who "took me under their wing" and provided me with superb supervision and expert guidance in the molecular biology work that I undertook, particularly in relation to the work in Chapter 7. Lina and Norma, thank you for being great teachers and wonderful friends. I also thank Chiara Bardella for valuable help with experiments in Chapter 6, Julian Knight for advice regarding the work in Chapter 6, Gianpiero Cavalleri and Mark McCormack for help with some of the genotyping experiments, Richard Copley for undertaking the bioinformatics analysis in relation to Chapter 9, Dan Lunn for advice regarding statistical analyses and Professor Mary Frances McMullin and Professor Holger Cario for their collaboration.

Undeniably, I consider myself extremely privileged to have had Peter Robbins and Peter Ratcliffe as my DPhil supervisors and my mentors. Thank you both for "taking me on" as your student, for all the intellectually stimulating conversations, and for your continuous guidance, support and encouragement throughout my doctoral studies.

I would like to thank my brothers, Nicholas and Panayiotis, for their unconditional love and my parents for their never-ending support, especially my mum, Andri, for being my "guardian angel" from the day I was born.

And last but not least, Antonis, thank you. Thank you for your endless encouragement, infinite patience and for being there to support me, always, every step of the way, both at the "highs" and the "lows" of this journey. I couldn't have done this without you. I dedicate this thesis to you with all my heart.

Table of Contents

Chapter 1: Integrative human physiology in response to hypoxia and the Hypoxia-Inducible Factor pathway – a general introduction

1.1 Hypoxia and integrative human physiology

Oxygen is essential for life. When there is reduced oxygen availability, a condition termed hypoxia, cells, tissues and organisms as a whole have the ability to sense this and respond so as to avoid damage. Hypoxia can have many causes – e.g. reduced oxygen tension in the inspired air at high altitude, inability of efficient gas exchange in the lungs, massive haemorrhage which reduces the oxygen-carrying capacity of blood, reduced perfusion of tissue e.g. due to occlusion of arteries etc. – and can be acute (ranging from seconds to minutes) or chronic (hours to days or longer). Hypoxia can also be continuous or intermittent; the latter will not be considered in this thesis.

At the systems level, the human response to hypoxia is characterised by changes in erythropoietic, respiratory and cardiovascular physiology in an attempt to restore uptake, transport and efficient delivery of oxygen to ensure adequate oxygenation of tissues in the body. These systemic changes that occur in the human body upon exposure to hypoxia are described below.

1.1.1 Erythropoietic response to hypoxia

The hormone erythropoietin (Epo) is essential for red cell production. During embryonic life Epo is primarily produced in the liver. After birth and during adult life the main production site becomes the kidney and more specifically the peritubular fibroblasts in the renal cortex, which respond to changes in oxygen (hypoxia) by producing more erythropoietin [1]. This in turn stimulates the production of red blood cells in the bone

marrow. The location of the Epo-producing cells in the kidney is well suited for erythropoietin production regulation. First, there is a constant ratio of blood flow rate and oxygen consumption (which decreases if the glomerular filtration rate decreases) so changes in cardiac output and renal blood flow do not have a great effect on renal cortical partial pressure of oxygen (P_{O_2}) and in turn on Epo production. Second, the kidney has marked heterogeneity of tissue P_{O_2} within it, which is paramount to oxygen sensing as a small reduction of oxygen delivery to the kidney (e.g. through anaemia or hypoxaemia) leads to significant changes in local tissue P_{O_2} in the kidney cortex and the outer medulla.

When humans are exposed to hypoxia e.g. after ascent to high altitude, there is increased production of erythropoietin (Epo), which in turn leads to elevation of red cell count and haemoglobin. This was first described in 1890 by F.G. Viault who observed a rise in red blood cell numbers on a journey of up to 4500 m in Peru. Serum Epo levels rise within 90 min of the onset of hypoxia [2], peak within 2 days and thereafter decline towards sea-level values over a period of 2 weeks [3, 4]. Haemoglobin (Hb) concentration and haematocrit (Hct) are elevated within days of exposure to high-altitude hypoxia [3] and continue to rise for up to eight months, causing a total increase in red cell mass of up to 50% at 4500 m [5]. Haemoglobin concentrations greater than 21 g/dl have been reported in humans acclimatizing to extreme high altitude [6]. The above highlight the presence of complex feedback mechanisms that probably exist in the regulation of erythropoiesis by hypoxia in humans.

1.1.2 Ventilatory response to hypoxia

Pulmonary ventilation is closely coupled to energy metabolism in the body, such that provision of oxygen is balanced to the removal of carbon dioxide. It is known that

exposure to acute hypoxia in humans stimulates an immediate increase in ventilation within minutes, termed acute hypoxic ventilatory response (AHVR), which is followed – after approximately 10 minutes – by a ventilatory decline, termed hypoxic ventilatory decline (HVD) [7] [8]. At high altitude, exposure to chronic sustained hypoxia leads to time-dependent ventilatory changes. "Ventilatory acclimatization" refers to the subsequent – i.e. after AHVR and HVD – progressive rise in ventilation that occurs over hours/days following ascent to high altitude. "Hyperventilation" refers to the stabilisation of a longer-term increased ventilation over months/years. "Blunting" refers to the loss of ventilatory sensitivity with prolonged hypoxia after years of residence at high altitude. Upon return to sea level and exposure to prolonged euoxia there is possible reversal of "blunting" [9].

There is a central and a peripheral chemoreflex component in the control of ventilation. The peripheral chemoreflex is very sensitive to changes in arterial oxygen tension and is mediated primarily by the peripheral chemoreceptors situated in the carotid body. The precise oxygen-sensing mechanism in the carotid body is not yet determined. It is generally accepted that the AHVR is mediated by the Type 1 cells of the carotid body and that there is hypoxic inhibition of potassium channels leading to membrane depolarisation, calcium influx, release of neurotransmitters and stimulation of sensory nerves that activate the respiratory centre of the brain [10]. Within seconds of exposure to acute systemic hypoxia (e.g. arterial $P_{O_2} < 60$ mmHg), the sensory discharge of the carotid body is augmented in a stimulus-dependent manner and there is a rapid rise in ventilation (AHVR); the lower the P_{O_2} the greater the increase in ventilation [8]. Evidence for the role of the carotid bodies in the AHVR in humans is given by studies of patients with bilateral resection of the carotid bodies who exhibited a reduction in their AHVR [11]. The carotid body, apart from oxygen, also responds to changes in CO_2 and pH and there is indeed an interaction between

these stimuli in producing a response. For example, a raised alveolar or arterial P_{CO_2} will cause an increased ventilatory sensitivity to a given drop in P_{O_2} – and indeed ventilation begins to rise even in the presence of "non-hypoxic" P_{O_2} –, while in the presence of hypocapnia a greater drop in P_{O_2} is required to produce a rise in ventilation. Similarly, the ventilatory response to changes in P_{CO_2} becomes steeper when P_{O_2} is lowered [12]. These systemic ventilatory changes are in keeping with studies of the effects of hypoxia and hypercapnia on the chemoreceptor activity of the carotid body in vitro [13].

HVD occurs after a few minutes and during the course of minutes-to-hours ventilation decreases [7] [8]; it has been shown that HVD occurs during isocapnic hypoxia and is thus a separate phenomenon from the ventilatory effects of hypocapnia and respiratory alkalosis which also reduce ventilation [14]. The cause and site of HVD are not clear and may involve both the peripheral chemoreceptors and the central respiratory centres [15].

The peripheral chemoreflex through the carotid body plays a key role in the early ventilatory acclimatisation to sustained hypoxia, as well. Experiments on awake goats have demonstrated this involvement directly: when hypoxia was restricted to the perfusion of their carotid body, early ventilatory acclimatization was observed [16]. This did not occur when the animals were made systemically hypoxic – including the cerebral circulation – whilst maintaining the perfusion of the carotid body euoxic [17].

In humans, the progressive rise in ventilation that is observed during acclimatization to the sustained hypoxia of high altitude is accompanied by a fall in arterial/alveolar P_{CO_2}. There was long debate in the past as to what extent the ventilatory changes observed at high-altitude are a direct result of hypoxia or a result of the progressive hypocapnia and

respiratory alkalosis that occurs. Studies with human volunteers have demonstrated that at least the early stages of ventilatory acclimatization are related to hypoxia *per se*. It was shown that 8 hours of isocapnic hypoxia (end-tidal P_{O_2} of 55 mmHg) triggered a progressive rise in ventilation in human volunteers and the onset of this response was faster than that noted in studies examining the effects of poikilocapnic hypoxia of altitude [18]. In addition, it was shown that there is a significant rise in the acute hypoxic ventilatory sensitivity during exposure to 8 hours of sustained hypoxia (both in poikilocapnic and isocapnic exposures) and that this rise is a result of hypoxia *per se* and not the associated alkalosis (in the case of poikilocapnia) [19]. Finally, the persistent reduction in end-tidal P_{CO_2} following exposure to sustained hypoxia occurred after both isocapnic and poikilocapnic hypoxia [20, 21].

The peripheral chemoreflex response mediated by the carotid body is of critical importance in conditions of hypoxia such as ascent to high altitude or hypoxic lung disease – especially in patients with chronic obstructive pulmonary disease (COPD) that have chronic CO_2 retention. Under normal conditions, the most important factor in the control of ventilation is the P_{CO_2} of blood, which is tightly controlled and is mostly regulated by the action of central chemoreceptors; however the peripheral chemoreceptors also contribute and have a faster response. Note that the carotid body is also involved in setting the basal ventilation rate in humans, as evidenced by studying patients with bilateral carotid body resections who had an air-breathing end-tidal P_{CO_2} that was significantly higher by 4.6 mmHg than that of healthy controls [22].

Upon restoration of euoxia following prolonged exposure to hypoxia – this does not refer to years of hypoxic exposure during which “blunting” may have developed and which will

be discussed in later chapters – ventilation and AHVR initially remain elevated and progressively normalise over a time course similar to that of the changes occurring in ventilatory acclimatization to hypoxia [23].

1.1.3 Cardiac output response to hypoxia

Acute hypoxia triggers a reflex increase in heart rate and cardiac output, as part of a sympathetic chemoreflex response. At high altitude, an increase in cardiac output is observed over the course of several days which is associated with an elevation in heart rate rather than stroke volume. This returns to normal as individuals acclimatize, but heart rate often remains elevated with a concomitant reduction in stroke volume [24].

1.1.4 Pulmonary vascular response to hypoxia

In contrast to other vascular beds, hypoxia constricts rather than dilates the pulmonary vasculature. Regional hypoxic pulmonary vasoconstriction (HPV) may be beneficial in perfusion-ventilation matching throughout the lung (e.g. in fighting pneumonia), and is probably an adaptive mechanism that evolved at sea level; global hypoxic pulmonary vasoconstriction, however, is not. In hypoxic lung disease, hypoxia causes pulmonary hypertension and vascular remodelling worsening patient morbidity and mortality [25]. At high altitude, global hypoxic pulmonary vasoconstriction leads to elevation of pulmonary arterial pressure and vascular resistance. This is a maladaptive process and it contributes to high-altitude diseases such as chronic mountain sickness (characterized by hypoventilation, erythrocytosis and pulmonary hypertension) in high-altitude residents and high-altitude pulmonary oedema in newcomers to high altitude [26, 27].

Acute hypoxia in humans elevates pulmonary arterial pressure which reaches a plateau within 5 min [28], begins to rise again after approximately 45 min and continues to rise for at least another 2 hours [28, 29]. Pulmonary vascular acclimatization is characterised by the resultant elevation in pulmonary arterial pressure and by an accompanying increase in the sensitivity of the pulmonary vascular response to any additional acute hypoxic stimulus [29].

The mechanism of hypoxic pulmonary vasoconstriction at the molecular and cellular level is not fully known. The pulmonary arterial smooth muscle cells (PASMC) are key effectors in the process. When exposed to hypoxia there is an increase in Ca^{2+} entry, mediated by voltage-gated Ca^{2+} channels [30] and further Ca^{2+}-induced Ca^{2+} release in these cells [31], causing them to contract, producing vasoconstriction. K^+ channels that are inhibited by hypoxia causing depolarisation in these cells are believed to play an important role [32, 33]. For example, $K_v1.5$ is one of several oxygen-sensitive K^+ channels that have been implicated and mice with $K_v1.5$ deletions have reduced PASMC oxygen-sensitive K^+ conductance and impaired HPV [34].

There is increasing evidence of an important role of the endothelium in HPV. It is believed that there is a "cross-talk" between pulmonary endothelial cells and PASMC cells. HPV requires the presence of an intact endothelium, whose role appears to be primarily related to the sustainability of HPV with prolonged hypoxia [35, 36]. Indeed, the isolated pulmonary vessel response is biphasic, consisting of an initial rapid transient increase in vascular tone which is endothelium-independent, followed by a second slower sustained increase, which is endothelium-dependent [36]. It is still not clear which of the factors involved in HPV act as oxygen sensors and which as effectors of the mechanism. A

number of secreted vasoconstrictors and vasodilators have been identified, the balance of which in the pulmonary vasculature vicinity affects the process of HPV. These include vasodilators such as nitric oxide and prostaglandins and vasoconstrictors such as endothelin-1 and leukotrienes; these act as modulators but yet none of them have been found to be indispensable for HPV [10, 37-39]. It is important to note that many of the experiments investigating HPV were done in isolated pulmonary arteries, or even isolated PASMC cells, and indeed the responses may vary in these isolated preparations compared to the response of the whole human or animal. In addition, studying patients with pulmonary arterial hypertension – which in most of the cases is not driven by hypoxia-induced pulmonary vasoconstriction – can also be misleading as they have significant pulmonary vascular remodelling with different kinds of interactions between cells and humoral factors than those directly involved in pulmonary vasoconstriction induced by hypoxia.

1.2 The hypoxia-inducible factor (HIF) transcription activation pathway

Hypoxia-inducible factors (HIFs) are a family of transcription factors which coordinate oxygen sensing and intracellular responses to hypoxia through regulation of gene expression in biological pathways such as energy metabolism, angiogenesis, erythropoiesis, apoptosis and others (Figure 1). What is apparent from studying the HIF system over the last 20 years or so is that it is central in coordinating oxygen sensing and physiological responses tailored to the needs of the cell, tissue, organ system and organism as a whole.

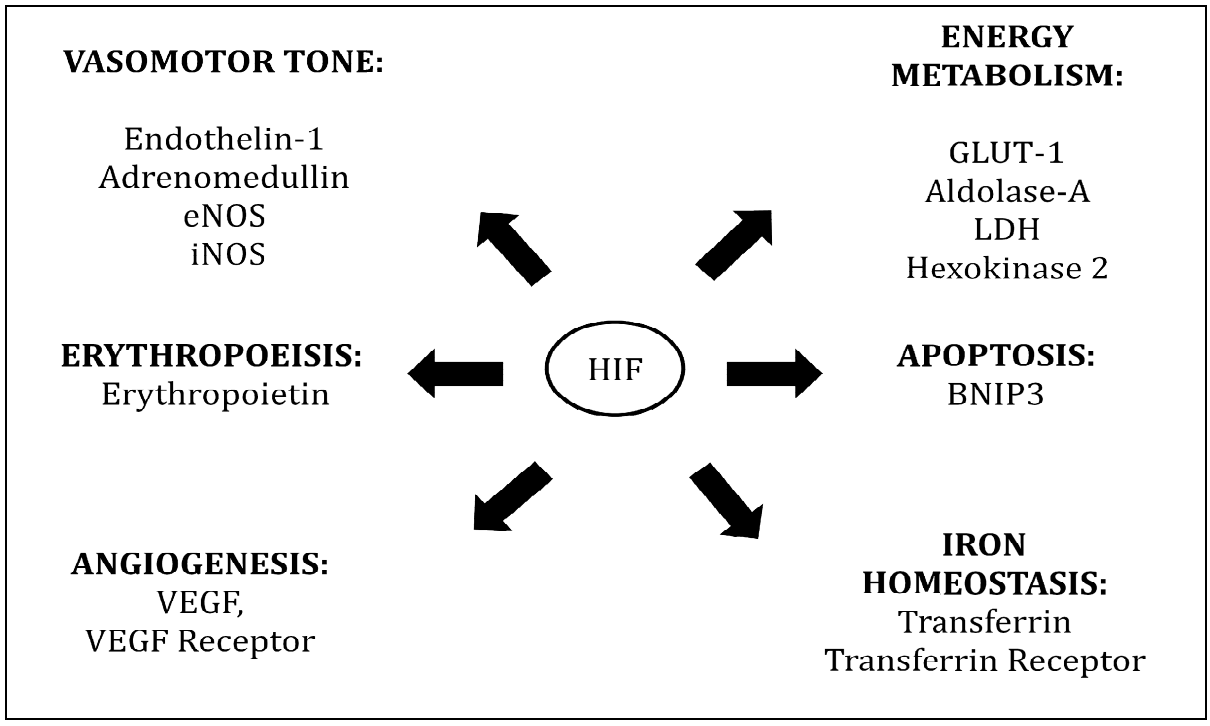


Figure 1: Examples of pathways and target genes regulated by the HIF system

HIF exists as a heterodimer with alpha (α) and beta (β) subunits. HIF-α is oxygen-regulated whilst HIF-β is constitutively expressed. In the presence of oxygen, HIF-α is inactivated and subjected to rapid and continuous proteolytic degradation; in hypoxia, it is stabilised, forms a heterodimer with HIF-β and binds to Hypoxia Response Elements (HREs) of target genes activating their transcription. The oxygen-dependent HIF-α subunit exists in 3 isoforms: HIF1α [40], HIF2α [41] and HIF3α [42]. The first two have been extensively studied and have overlapping as well as distinct transcriptional functions in response to hypoxia.

HIF1 was first discovered as a transcription factor regulating the expression of erythropoietin, by finding that it binds an oxygen responsive transcriptional enhancer (HRE) lying 3′ to the erythropoietin gene [40, 43, 44]. It was later shown that HREs – that contain a RCGTG consensus – were present and operated in a variety of cells and tissues,

even cells that did not make erythropoietin, suggesting that oxygen-sensing and gene expression via the HIF pathway is widely-spread [45].

1.2.1 Regulation of HIF by oxygen

HIF1α and HIF2α share a great degree of protein homology. They both contain two distinct oxygen-degradation domains, the N-terminal oxygen-dependent degradation domain (NODD) and the C-terminal oxygen-degradation domain (CODD), as shown in Figure 2.

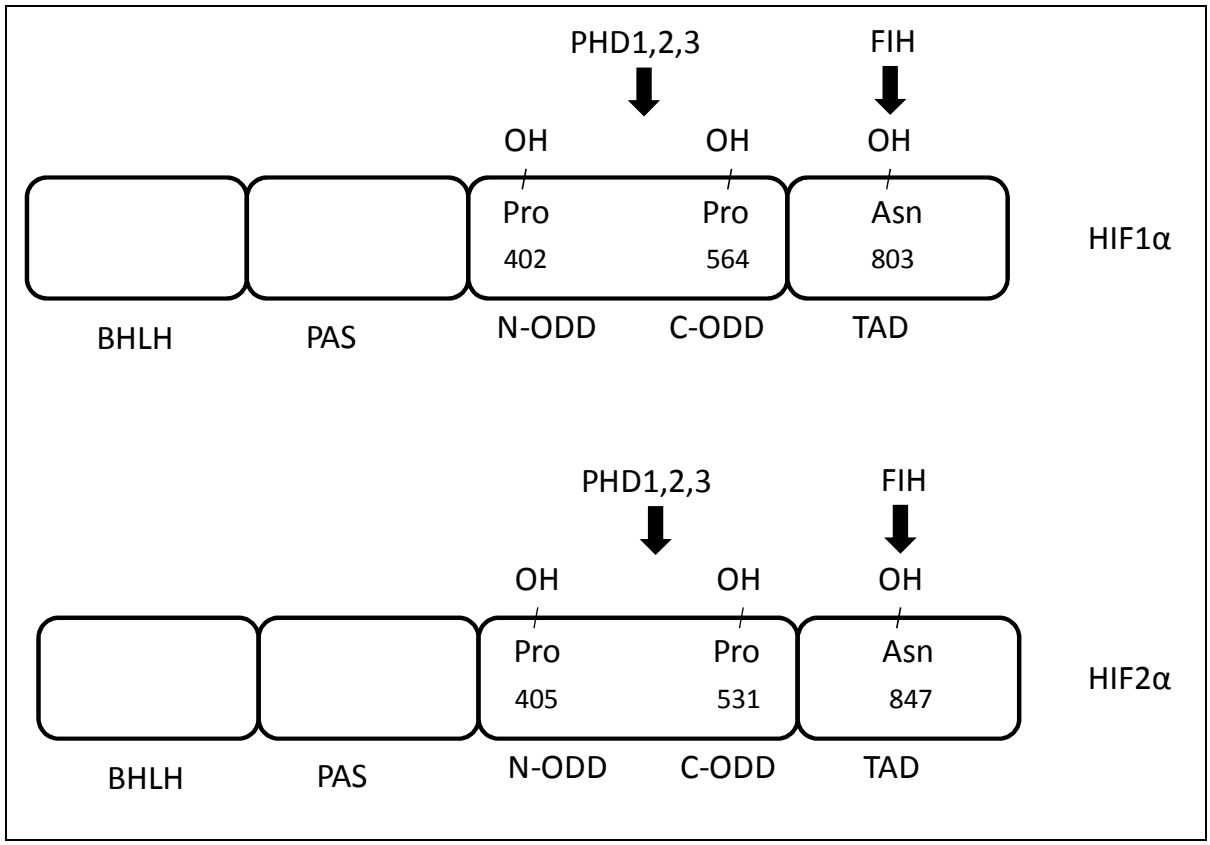


Figure 2: The domain structures of HIF-alpha subunits. The basic helix-loop-helix (BHLH) domains and the PAS domains are responsible for dimerisation and DNA binding. The amino-terrminal oxygen-degradation domain (N-ODD) and the carboxy-terminal oxygen-degradation domain (C-ODD) contain prolyl residues which are subject to hydroxylation by the prolyl-hydroxylase enzymes (PHDs) and are responsible for regulating HIFα stability. The trans-activation domain (TAD) contains an asparagines residue which is subject to hydroxylation by the enzyme factor-inhibiting HIF (FIH) and is involved in regulating transactivating ability.

Oxygen-dependent regulation of HIF predominantly occurs at the post-translational level and this is shown in Figure 3. In the presence of oxygen, enzymatic hydroxylation of 2

prolyl residues (residing in NODD and CODD, as shown in Figure 2), catalysed by prolyl hydroxylase domain (PHD) enzymes, of which there are 3 isoforms – PHD1, PHD2 and PHD3 – occurs in an oxygen-dependent manner [46-51]. In addition to oxygen, 2-oxoglutarate as co-substrate and iron (Fe^{2+}) and ascorbate as co-factors, are required by these enzymes to function [47, 52]. The hydroxylation of these residues enhances the binding of the Von-hippel Lindau (VHL) protein to HIF-α which in turn enhances HIF-α ubiquitylation, resulting in its subsequent proteasomal destruction [47, 53]. In addition, oxygen-dependent hydroxylation of a specific asparagine residue within the C-terminus of either HIF1α or HIF2α by FIH (factor inhibiting HIF), another 2-oxoglutarate dioxygenase, blocks recruitment of transcriptional co-activators (p300/CBP), thus affecting transcriptional activation of target genes [54, 55]. In the presence of hypoxia, but also iron deficiency or iron chelation, there is reduced hydroxylase activity of the oxygen-dependent hydroxylases. This results in stabilisation and accumulation of HIF-α, which then heterodimerises with HIF-β and transcriptionally activates target genes.

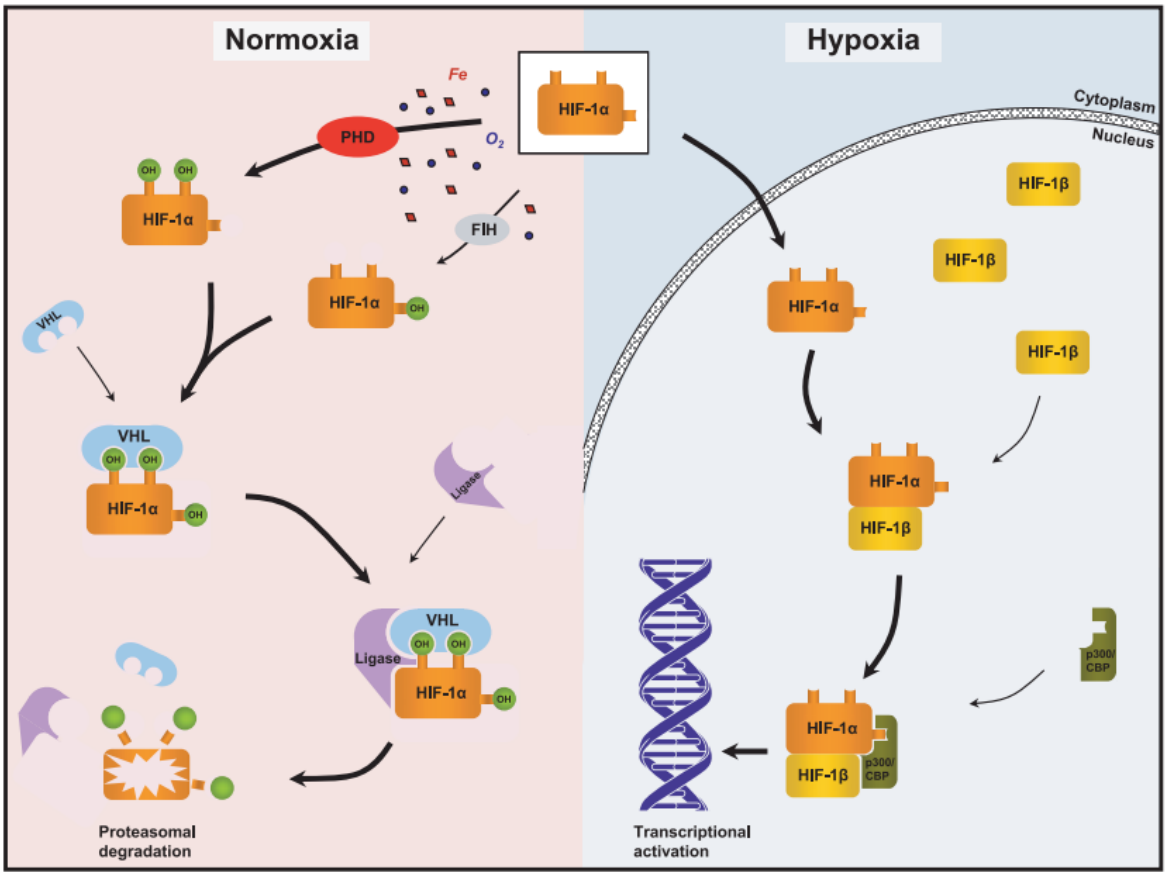


Figure 3: Regulation of HIF by hypoxia. From: The human side of hypoxia-inducible factor, Smith et al 2008 [56].

There are differences in the relative abundance of the different PHD enzymes in different tissues, adding further levels of complexity in the regulation of HIFs and oxygen sensing. PHD2 is more abundant than PHD1 and PHD3 in most tissues, and in keeping with this it has been reported to be the most dominant isoform in oxygen sensing [57]. There is also evidence that PHD2 contributes more to the regulation of HIF1α than HIF2α, and that PHD3, conversely, contributes more to the regulation of HIF2α [58]. There is also evidence that there is differential action of the various PHDs on the NODD and CODD hydroxylation domains [58]. Overall the differential patterns of tissue expression, intracellular localisation [59], response to stimuli and substrate selectivity that the PHD enzymes show suggest that they may have distinct, as well as overlapping functions and that this probably contributes to differences seen in the spatial and temporal regulation of

HIF1α versus HIF2α. Furthermore, the PHD2 and PHD3 enzymes are themselves HIF-transcriptional target genes [46].

HIF1α is expressed in all mammalian tissues and cell types analysed, whereas HIF2α has a more restricted cell and tissue-type pattern of expression; it was originally defined as an endothelial cell-specific isoform [41] but it is now recognised that its expression is more widespread [60] with high abundance in endothelium, blood vessels, lung and placenta. In fact, whereas HIF1α is present in both vertebrates and invertebrates, HIF2α is believed to have arisen during vertebrate evolution [61]. Knockout studies in mice and in cell lines, have demonstrated that HIF1α and HIF2α have non-redundant roles. Both HIF1α- and HIF2α-null mice die during embryogenesis [62, 63]. There are some HIF-target genes that have been found to be activated by both HIF1α and HIF2α, such as VEGF, whilst some others were found to be selectively HIF1α target genes, such as glycolytic genes and carbonic anhydrase 9 [64] [65] or HIF2α target genes, such as Arginase 1 [66] and Tie-2 [41], but this can largely depend on the tissue or cell line studied. In addition, genome-wide analysis by ChIP-ChIP [67] and ChIP-Seq [68] in various cell lines has demonstrated commonality as well as disparity in the binding of HIF1 and HIF2 at different target-gene loci. It is unclear whether HIF1α and HIF2α might work synergistically or antagonistically in tissues that express both upon exposure to hypoxic stimuli of different magnitude and duration.

Like other biological systems the HIF pathway is subject to negative feedback mechanisms. The effects of PHD2/3 enzymes – which are transcriptionally up-regulated by HIF – in the post-translational hydroxylation and degradation of HIFs has been mentioned. Other HIF-transcriptional targets also act to attenuate the HIF response, e.g. CITED2,

which blocks c-activator recruitment and hence decreases the HIF transcriptional effects [69]. HIF1α and HIF2α have different temporal patterns of regulation, with HIF1α protein level showing attenuation whereas HIF2α showing preferential stabilisation under conditions of prolonged hypoxia [70].

There are other mechanisms of regulation of HIFs, adding to the complexity of the system. For example, there is evidence of oxygen-independent phosphorylation pathways affecting HIF1α mRNA translation [71] but also evidence of heavy phosphorylation of HIF1α protein itself, that can affect its activity [72]. Furthermore, there is evidence of proteins that bind HIFs, affecting their stability and their transcriptional activity. One example is Heat Shock Protein 90 (HSP 90) which interacts with HIF1α affecting its stability in hypoxia [73, 74]. Another example is the hypoxia-associated factor (HAF) which binds HIF1α and promotes its proteasomal degradation; in contrast, it binds to a different site on HIF2α and promotes its transcriptional activity without affecting its stability [75].

1.2.2 The role of the HIF pathway in integrative physiology

Evidence for a central role of the HIF pathway in integrative physiology comes from studying patients with genetic abnormalities in the HIF system or from genetic mouse models, and from studies of pharmacological manipulation of the HIF system in humans.

1.2.2.1 Genetic abnormalities in the HIF pathway affect integrative physiology

Patients with mutations in VHL – Chuvash Polycythaemia:

Chuvash polycythaemia is a rare autosomal recessive disorder, in which affected individuals are homozygous for a 598C>T mutation in VHL, resulting in an Arg200Trp aminoacid change. The mutated VHL has diminished affinity for hydroxylated HIF-α, thus

inhibiting HIF-α degradation and causing up-regulation of HIF-target genes, even in euoxia. Studies in cells from these patients have shown over-expression of many HIF-regulated gene products [76, 77].

These patients have predominantly excessive erythrocytosis – due to increased up-regulation of erythropoietin – which presents at a young age. In addition, patients have abnormal cardiopulmonary physiology and metabolism. It was shown that they have pulmonary hypertension and increased resting ventilation and that upon exposure to acute systemic hypoxia they exhibit an exaggerated pulmonary vascular response and an exaggerated ventilatory sensitivity [76]. In addition, they were shown to have abnormal exercise metabolism with a reduced maximal exercise capacity, an increased exercise-induced lactate accumulation and an early and marked phosphocreatine depletion and acidosis in skeletal muscle [78]. These abnormalities highlight the involvement of the HIF pathway as a master controller of oxygen-sensing at the integrative physiology level.

Patients with mutations in HIF2α:

A gain-of-function mutation in HIF2α was first identified in 3 generations of a family with erythrocytosis [79]. The affected individuals are heterozygous for G1609T, resulting in Gly537Trp. Functional studies have shown that it leads to stabilisation of HIF2α and induction of downstream genes is increased. Further different mutations in HIF2α have also been reported [80-83]. Studies at the integrative physiology level (as in the patients with Chuvash polycythaemia) have identified a systemic phenotype of resting pulmonary hypertension, increased resting ventilation and an exaggerated pulmonary vascular response to acute hypoxia [84]. Again increased HIF-target gene expression was found in peripheral blood mononuclear cells from these patients [82].

Patients with mutations in PHD2:

A family was identified with a mutation in PHD2, a heterozygous C950G resulting in a Pro317Arg, associated with erythrocytosis [85]. Further PHD2 mutations (with the phenotype of erythrocytosis) have been identified e.g. G1112A, A1121G, frameshift mutations and others [86-88]. These mutations have been shown to cause abnormal activity of PHD2 *in vitro* including decreased HIF binding, defective HIF hydroxylase activity or truncation of the protein [85, 87]. To my knowledge, only one patient with PHD2 mutation has been studied at the integrative physiology level and was found to have increased pulmonary vascular sensitivity to hypoxia which was intermediate in terms of magnitude between healthy volunteers and Chuvash polycythaemia or patients with HIF2α mutations (unpublished data, personal communication with Dr TG Smith).

Studies in genetically-modified mice further elucidate the role of the HIF pathway in systemic physiology. Knockout mice with heterozygous loss of HIF1α had depressed carotid body oxygen sensitivity and abnormal ventilatory acclimatization to chronic hypoxia [89]. Furthermore, mice with heterozygous deficiency for functional HIF genes – either HIF1α [90] or HIF2α [91] – had impaired development of polycythaemia and pulmonary hypertension in response to hypoxia.

Mouse models of Chuvash Polycythaemia confirmed the phenotype observed in humans such as the presence of increased ventilation and pulmonary hypertension and further demonstrated that the effects seen in the disease are primarily HIF2α-driven as opposed to HIF1α-driven [92].

PHD2 conditional knockout mice were found to develop excessive erythrocytosis and increased vascular growth and double knockout of PHD1 and PHD3 in mice also led to moderate erythrocytosis [93, 94]. Furthermore, mice heterozygous for PHD2 were found to have increased ventilatory sensitivity to hypoxia and carotid hyperplasia [95]. Note that germ-line PHD2 $^{-/-}$ mice are not viable whereas PHD3 $^{-/-}$ and PHD1 $^{-/-}$ mice are [96].

Interestingly, a recent study examining the hypoxic response of the carotid body in HIF2$\alpha^{+/-}$ mice manifested heightened carotid responses to hypoxia, which contrasts previous findings in HIF1$\alpha^{+/-}$ mice, again highlighting potentially different roles of HIF1 and HIF2 in systemic physiology [97].

1.2.2.2 Manipulation of the HIF pathway by altering iron status in humans affects integrative physiology

A study of human volunteers investigated the effect of an intravenous infusion of iron – a major co-factor for the action of the oxygen-dependent hydroxylases in the HIF pathway – on the pulmonary vascular response to 8 hours of sustained hypoxia. It was found that infusion of iron greatly reduced both the elevation in baseline pulmonary systolic arterial pressure and the enhanced sensitivity of the pulmonary vasculature to acute hypoxia following 8 hours of sustained hypoxia [98]. Compatible with the above finding, an 8-hour infusion of desferrioxamine – an iron chelator – significantly elevated baseline pulmonary systolic arterial pressure and enhanced the sensitivity of the pulmonary vasculature to acute hypoxia [98]. Infusion of desferrioxamine was also shown to markedly increase plasma erythropoietin, a HIF-target gene [99]. In addition, clinical trials on healthy humans and on patients with chronic mountain sickness showed that hypoxic pulmonary hypertension may be attenuated by iron supplementation and exacerbated by iron depletion [100]. These

effects of alteration of the “HIF response” at the systemic level, i.e. down- or up-regulation by iron supplementation or iron depletion, respectively, parallel the effects seen in cultured cells in which iron status is manipulated [52, 101].

1.3 Tibetan natives – a “unique” high-altitude population

Variation in the human physiological responses to hypoxia is also seen naturally in healthy humans and this is particularly evident when studying high-altitude resident populations.

More than 140 million people live and work at high altitude [102]. The severely reduced oxygen availability at high-altitude environments, termed hypobaric hypoxia, presents a great physiological challenge to humans residing there, and is likely to have acted as an agent of natural selection. High-altitude natives show different physiological traits compared to newcomers to high altitude [103]. Moreover, amongst various high-altitude populations there are distinct physiological phenotypes, probably as a result of different patterns of adaptation.

Humans have lived on the Tibetan plateau – which has an average elevation of 4000 m, a barometric pressure ranging between 450 and 525 mmHg and an inspired PO_2 of ~ 80 mmHg – for at least 25 000 years, making Tibetan natives most likely the oldest high-altitude population [104]. This suggests that they will have had more time and opportunity for natural selection in response to a hypoxic environment than any other high-altitude human population.

Tibetans have distinct physiological traits compared with other high-altitude populations such as the Andeans, North American Rocky Mountain inhabitants or Han Chinese

migrants to Tibet. In terms of ventilation, Tibetan highlanders seem to resemble acclimatized newcomers in that they retain high levels of resting ventilation [105]. Compared to Andeans residing at similar high altitudes, Tibetans were found to have 1.5 times higher ventilation [105]. They were also found to have a hypoxic ventilatory response (HVR) as high as that of acclimatized newcomers, in contrast to Andeans, suggesting that Tibetans do not undergo the process of acquired blunting [105]. In the same study, it was also proposed that there was greater heritability in the HVR of Tibetans i.e. more of the variance seen in the HVR in Tibetans was due to genetic factors compared to the variance in HVR in Andeans.

In terms of the erythropoietic response to hypoxia, Tibetan highlanders exhibit a blunted erythropoietic response to high-altitude hypoxia. At similar high altitudes, Tibetans have significantly lower haemoglobin concentrations – typically 1 to 3.5 g/dl lower – than their Andean counterparts or Han Chinese migrants to high altitude [106-109]; at 4000 m Tibetans have haemoglobin concentrations comparable to those of US sea-level residents [106].

The prevalence of chronic mountain sickness – a high-altitude disease characterised by hypoventilation, pulmonary hypertension and excessive erythrocytosis – is very low in Tibetans compared with other high-altitude populations such as Andeans or Han Chinese migrants to high altitude [102]. One previous study of 5 Tibetans resident at 3658 m found that their pulmonary arterial pressures were within sea-level norms and were little changed by further exposure to further hypoxia [110]. Pronounced pulmonary vascular remodelling and hypertrophy accompanies pulmonary hypertension; consistent with the absence of such

effects in Tibetan highlanders is a study showing a lack of smooth muscle in the small pulmonary arteries in Tibetan men at 3600 m in Ladakh [111].

In terms of physical performance at high altitude, it has long been suggested that high altitude populations, particularly Andeans and Sherpas of Tibetan origin, have greater work capacity in hypoxic environments compared to acclimatized lowlanders. The evidence around this is not clear. Marconi et al [112] reviewed the relevant studies and concluded that overall the superior work capacity of high-altitude natives is not related to differences in maximal aerobic capacity (V_{O_2} max), in neither Tibetans nor Andeans. However, they reported that Tibetans were characterised by a better economy of locomotion (cycling, walking, running) and had a prolonged submaximal performance capacity at high altitude, and suggested that this may be a result of increased efficiency of muscle oxidative metabolism.

Another distinct characteristic of the Tibetan adaptation to high altitude is that
Tibetans are better protected from intrauterine growth retardation (IUGR). Heavier birth weights are seen in Tibetans than Andeans or Rocky mountain residents of the same altitude [113]. In addition, similar birth weights among Tibetans in Lhasa (3658 m) and Tibetans in Kathmandu (1200 m) were found [114] and in the same study it was reported that altitude-associated difference in birth weight is least among Tibetans compared with North American and South American populations.

While the physiology of Tibetan natives is well-studied at high altitude, relatively little is known as to how or whether their physiology changes upon descent and residence at sea level. Furthermore, despite the well-observed systemic physiological traits of Tibetans at

high altitude, relatively little is known about the cellular, molecular or genetic factors and mechanisms that may underlie them.

In the last 2 years, there have been a number of genomic studies investigating genetic adaptation in Tibetans. These studies used different genome-wide methodologies to compare the Tibetan with primarily the Han Chinese genome and occasionally the Japanese genome [115-119]. Tibetans and Han Chinese are thought to share a common ancestry. Thus, by comparing the Tibetan genomes with genomes of closely-related ethnic backgrounds, any differences captured in the genetic structure between the two populations are likely to have arisen through natural selection in response to the hypoxic environment that Tibetans live in. What is generally apparent from these studies is the presence of positive selective sweeps that involve the genes *EPAS1* (which codes for HIF2α) and *EGLN1* (which codes for PHD2), both key components of the HIF pathway, thus implicating these genes in the genetic adaptation of Tibetans to high altitude. Some of these studies showed a correlation of the putatively selected genetic variants in each of these two genes with haemoglobin level in Tibetans at high altitude [115, 117, 119]. The findings of the key studies are summarised below.

Beall et al [115] used a genome-wide allelic differentiation scan (GWADS) to compare ~ 500,000 single nucleotide polymorphisms (SNPs) between Tibetan high-altitude natives (from the Yunnan province) and low-land Han Chinese. They found 8 SNPs that achieved genome-wide significance (p values ranging from 2.81×10^{-7} to 1.49×10^{-9}), in terms of their divergence between the two ethnic groups, which clustered within ~235 kb on chromosome 2, just downstream of *EPAS1*. These SNPs were found to be in high linkage disequilibrium (LD) with one another forming an extended haplotype. In addition, through

candidate gene approaches in separate cohorts of Tibetans they identified 31 SNPs in *EPAS1* in high LD that correlated significantly with haemoglobin such that major allele homozygotes averaged a Hb concentration of ~ 0.8 g/dl lower than the heterozygotes. These *EPAS1* SNPs were in high LD with the signal from the GWADS and were found at higher frequencies in Tibetans than in Han Chinese, thus providing evidence for selection on *EPAS1* associated with low haemoglobin in Tibetan highlanders.

Simonson et al [119] performed a genome-wide comparison of Tibetan and Han Chinese or Japanese populations and identified chromosomal regions of positive selected sweep which were located in or near 10 genes from *a priori* selected candidate genes (based on their involvement in oxygen homeostasis). Amongst these genes were *EPAS1*, *EGLN1* and *PPARA*. In addition, variants in the 2 latter genes were shown to associate with haemoglobin concentration, with again the selected haplotype associating with low haemoglobin.

Yi et al [117] sequenced 50 Tibetan exomes and compared the sequencing data with Han Chinese and Danish populations from HapMap to identify changes in allelic frequencies consistent with adaptation to high altitude. *EPAS1* was identified as the strongest candidate for natural selection; in addition they correlated an *EPAS1* variant (an intronic SNP) that was found at very high frequencies in Tibetans (and low frequencies in Han Chinese) with erythrocyte count.

Xu et al [118] compared genome-wide allelic frequencies between Tibetan and Han Chinese and identified positive selective sweeps and dominant haplotypes in *EPAS1* and

EGLN1 genes in Tibetans. They also showed a correlation of the dominant haplotypes with altitude of residence.

Peng et al [116] SNP-typed 50 Tibetan individuals and compared the results to data from HapMap Han Chinese, identifying *EPAS1* as one of the genes that underwent a selective sweep in Tibetans. They also sequenced the *EPAS1* gene in these individuals and genotyped 3 SNPs from each of *EPAS1*, *EGLN1* and *PPARA* in a large cohort of Tibetans (n=1334). They found significant differences in allele frequencies for the *EPAS1* SNPs between Tibetan and Han Chinese/Japanese and also for one of the SNPs in *EGLN1*.

Finally, Bigham et al [120] performed a genome scan and applied four population genetic statistics to identify selection-nominated candidate genes and gene regions in two populations, Andeans and Tibetans, studied separately. They found different patterns of positive natural selection between the two populations. Interestingly, *EGLN1* showed evidence of positive selection in both Tibetans and Andeans but with different patterns of variation within the gene. This is made more interesting by noting that Tibetans and Andeans differ in terms of physiological traits at altitude, as well.

1.4 Scope of this thesis

Section 1.2 gave evidence of how certain variations (e.g. specific disease-causing genetic mutations) in the HIF pathway are associated with functional variation in the integrative physiology of humans.

Section 1.3 gave evidence that variation in human responses to hypoxia is also observed in healthy humans e.g. different high-altitude populations have distinct integrative physiology

phenotypes, and show different patterns of adaptation to life and reproduction in environments of extreme hypoxia. Moreover, evidence was presented from recent genomic studies of high-altitude populations – mostly Tibetan natives – that key genes of the HIF pathway have undergone positive selection and thus may be involved in the genetic adaptation to high altitude, although the mechanisms are unknown.

There is a great deal of variation within humans in their physiological and pathophysiological responses to hypoxia, which is unaccounted for. In the case of healthy humans, athletes, elite mountaineers and healthy volunteers studied in simulated high-altitude hypoxic environments exhibit differences in their physiological responses. In the case of disease, variation is observed in patients with hypoxic lung disease, such as COPD. Some patients with COPD develop type 1 respiratory failure (low arterial P_{O_2} and normal P_{CO_2}) whereas others – with similar degree of airflow obstruction on pulmonary function tests – develop type 2 respiratory failure (low arterial P_{O_2} and high P_{CO_2}). Moreover, some patients with mild lung disease – assessed clinically, radiologically, on pulmonary function tests and/or on arterial blood gas analysis – develop out-of-proportion pulmonary hypertension. It is possible that this phenotypic variation described above may be related to underlying variation in a patient's hypoxic sensitivity or response. The HIF pathway exhibits a multi-level complexity in the way it delicately regulates oxygen sensing and the responses to changes in oxygen at the cell, tissue, organ, system and organism level. It is conceivable that there may be a significant degree of functional variation in humans that is brought about by underlying variation – which may be more subtle than protein-coding disease-causing mutations – in the HIF pathway. While attractive, this is a generic hypothesis which is very difficult to address directly.

However, the recent reports that genes of the HIF pathway are implicated in the genetic adaptation of Tibetans to high altitude, gives an opportunity to explore how variation in the HIF pathway – other than disease-causing mutations – may be functional by focusing further studies on this human group. In this thesis, I have considered this issue by taking a number of different approaches and undertaking studies at the genetic level (genotype), at the molecular and cellular level (intermediate phenotype) and at the integrative physiology level (systems-level phenotype). The greatest part of the work in this thesis focuses on studies of Tibetan natives at sea level. I was motivated by the fact that despite the fact that the physiology of Tibetan natives is well-studied at high altitude relatively little is known on how their physiology changes upon residence at sea level, and by the fact that despite the recent genetic studies little is known regarding the cellular and molecular mechanisms that may underlie the observed physiological traits in Tibetans.

In Chapter 2 of the thesis, I set out to examine, at sea level, the integrative physiology of Tibetan natives in comparison with their most closely-related major ethnic group, the Han Chinese. In particular, I sought to determine whether Tibetans possess a distinct cardiopulmonary phenotype, as has been observed for patients with disease in the HIF system. I recruited 14 Tibetan and 13 Han Chinese volunteers living in the UK. After a set of baseline measurements, I determined their cardiopulmonary responses to acute (minutes) and sustained (8 hours) hypoxia. In view of the reports of association of putative *EGLN1* and *EPAS1* haplotypes with haemoglobin concentration in Tibetan highlanders in recent genomic studies of natural selection [115, 117, 119], I sought to define genotype-phenotype relationships for these variants in the Tibetan volunteers that participated in the study and this is described in Chapter 3. To explore the existence of an intermediate phenotype affecting the HIF system, I also compared Tibetan and Han Chinese responses

to hypoxia at the cellular level by examining the expression of HIF-regulated genes in peripheral blood lymphocytes cultured at graded oxygen tensions; this study is described in Chapter 4. Similar comparisons were also undertaken in immortalised with Epstein-Barr virus (EBV) lymphoblastoid cell lines obtained from either Tibetan or Han Chinese volunteers and these are described in Chapter 5. Finally, I investigated whether and how genetic variation in *EPAS1* and *EGLN1* may be functional at the cellular and molecular level, in Chapters 6 and 7, respectively.

Chapter 8 describes a small genetic analysis in cohorts from a South-American high-altitude population, which consist of healthy individuals and patients with chronic mountain sickness (CMS), and aims to correlate genetic variants in *EGLN1* to the phenotype of CMS. Finally, Chapter 9 is concerned with a small genetic study of patients with idiopathic erythrocytosis with inappropriately normal or elevated erythropoietin levels. I was motivated to investigate whether these patients, who have abnormally increased haemoglobin concentrations – as opposed to the Tibetans who have low haemoglobin concentrations –, have associated alterations in the HIF system causing their disease.

Chapter 10 provides concluding remarks and ideas for future work.

Chapter 2: A comparative integrative physiology study of Tibetan and Han Chinese volunteers at sea level

In this study, I investigated the integrative physiology of Tibetan volunteers in comparison with Han Chinese volunteers, at sea level, exploring in particular whether a distinct cardiopulmonary phenotype can be observed, as was seen for patients with disease in the HIF system.

2.1 METHODS

2.1.1 Participants

Fourteen healthy male Tibetan volunteers and 13 healthy male Han Chinese volunteers (control group) participated in the study. Not all volunteers participated in all protocols of the study. All Tibetan volunteers were born and had lived for a variable number of years at the Tibetan plateau before emigrating to sea level. All Han Chinese volunteers were sea-level residents. The study was explained to the volunteers both verbally and in writing, and each volunteer gave written informed consent before participating. The study was approved by the Berkshire Clinical Research Ethics Committee.

2.1.2 Description of protocols

The physiological experiments took place in our laboratory in the following order. Each volunteer reported at 7.30 am. After a period of rest (~20 min), his baseline end-tidal partial pressure of CO_2 ($P_{ET_{CO_2}}$) was recorded, while sitting quietly for 10 min. Thereafter, the volunteer undertook 3 acute hypoxic protocols, with a period of 10-15 min rest between protocols.

Protocol 1: This tests acute hypoxic ventilatory responses (AHVR). During this test, the volunteer was seated upright and breathed through a mouthpiece with his nose occluded. $P_{ET_{O_2}}$ was held at 100 mmHg (euoxia) for the first 5 min; this was then followed by six square waves in $P_{ET_{O_2}}$ varying between 1 min at 50 mmHg (hypoxia) and 1 min at 100 mmHg, as shown Figure 4A. $P_{ET_{CO2}}$ was held constant at 2 mmHg above the volunteer's resting value (isocapnia).

Protocol 2: This tests pulmonary vascular responses to acute isocapnic hypoxia. The volunteer lay down on a couch in the left lateral position and breathed through a mouthpiece with his nose occluded. The gas mixture breathed varied such that $P_{ET_{O_2}}$ was held at 100 mmHg for 5 min, 50 mmHg for 10 min and 100 mmHg for 10 min (Figure 4B). $P_{ET_{CO_2}}$ was held constant at 2 mmHg above the volunteer's resting value. Pulmonary vascular response was assessed by determining ΔPmax using non-invasive echocardiography, carried out throughout the test.

Protocol 3: This tests cardiac output responses to acute isocapnic hypoxia. This protocol is identical to Protocol 2, with echocardiography used to record cardiac output ($\dot{Q}$) measurements.

Following Protocols 1-3 and after a short period of rest (20 min), the volunteer was exposed to 8 hours of sustained isocapnic hypoxia at a $P_{ET_{O_2}}$ of 50 mmHg and $P_{ET_{CO_2}}$ at the volunteer's baseline resting level. This was achieved by asking the volunteer to stay in a purpose-built clear-sided chamber where end-tidal gases were continuously sampled and the gas composition in the chamber altered to maintain the end-tidal gas values at the required levels. Echocardiographic measurements were obtained from the volunteer at

hourly intervals. Blood samples were taken to measure plasma erythropoietin levels at baseline, 4 hours and 8 hours.

Following the chamber hypoxic exposure, and after a 30 min rest period, the volunteer repeated Protocols 1, 2 and 3. This marked the end of the experimental day. Throughout all protocols, the volunteer was continuously monitored by pulse oximetry, and a 3-lead ECG. Figure 5 shows the experimental apparatus used in an experiment in progress.

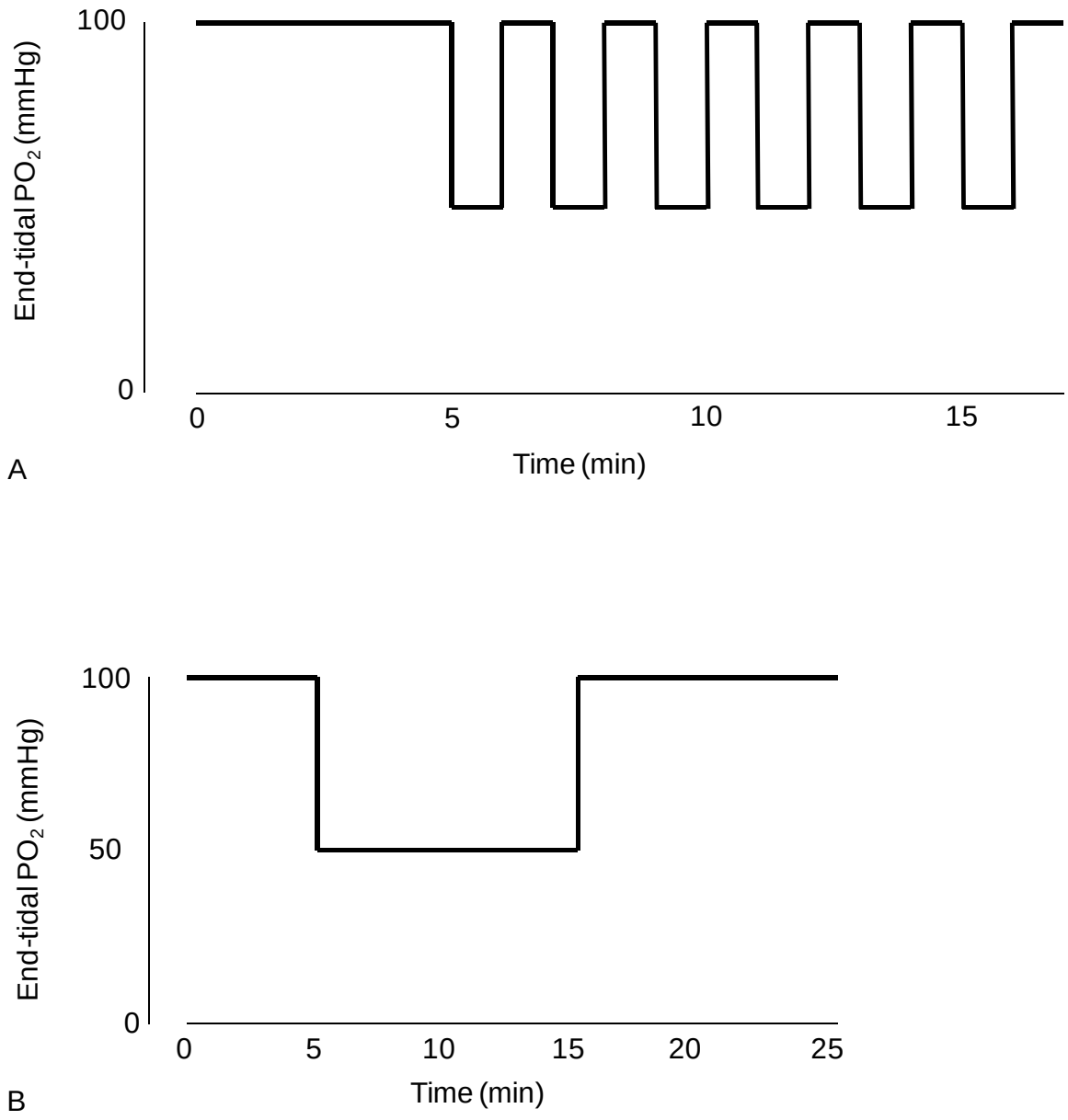


Figure 4: Schematic representation of the acute hypoxic physiology protocols. (A) Protocol 1, which tests the acute hypoxic ventilatory response. (B) Protocol 2 or 3, which test pulmonary vascular and cardiac output responses, respectively.

A

B

Figure 5: Experimental set up. (A) Apparatus used for the acute hypoxic challenges. (B) Chamber used for the 8-hr sustained hypoxic exposure.

2.1.3 Control of end-tidal gases during the experiments

During the acute hypoxic exposures (Protocols 1, 2 and 3), respired gases were sampled via a fine catheter close to the mouth and analysed continuously for P_{CO_2} and P_{O_2} by mass spectrometry (*Airspec 220, Airspec Ltd, UK)*; $P_{ET_{O_2}}$ and $P_{ET_{CO_2}}$ were recorded in real time by a computer and logged breath by breath. A dynamic end-tidal forcing system [121] was used to control end-tidal gases as specified in each protocol. Before the start of each experiment, a cardiorespiratory model was used to construct a forcing function that contained the breath-by-breath values for inspiratory P_{CO_2} and P_{O_2} predicted to produce the desired end-tidal gas sequences. During the experiment, using feedback control the gas values sampled and recorded were compared with the desired values and a computer-controlled gas-mixing system delivered the appropriate concentrations of inspired O_2, CO_2 and N_2 to achieve the desired values [121]. Gases were humidified prior to final delivery to the volunteer via the mouthpiece.

During the chamber hypoxic exposure, respired gas was sampled via nasal cannulae and analysed using a commercial CO_2/O_2 gas analyser (*Datex Normocap 200 Oxy, GE Healthcare, UK)* and the values fed to a computer. $P_{ET_{O_2}}$ and $P_{ET_{CO_2}}$ were detected and recorded breath-by-breath. The computer automatically adjusted the composition of the gas in the chamber every 5 min, or at manually overridden intervals, to minimise the error between desired and actual values for the end-tidal gases. This system has been described in detail elsewhere [122].

2.1.4 Measurement and modelling of the acute hypoxic ventilatory response (AHVR) during Protocol 1

Cyclical hypoxic stimuli (square waves) were used. Ventilatory volumes were measured using a bidirectional turbine volume-measuring device (*Interface associates, UK*) and flows (phase transitions) by a pneumotachograph (*Fleisch, Switzerland*), by recording and measuring the pressure gradient across it; inspired and expired volumes were determined in real time by a computer. To quantify the ventilatory responses to acute variations in hypoxia, dynamic models relating total expired ventilation ($\dot{V}_E$) to end-tidal gas profiles were fitted to the data. Model 3 of Clement and Robbins [123] was employed. This is a single-compartment model in which the ventilatory sensitivity to hypoxia is modelled using a parameter G_P, and a parameter $\dot{V}_C$ reflects residual $\dot{V}_E$ in the absence of hypoxia.

Briefly, $\dot{V}_E = \dot{V}_P + \dot{V}_C$, where $\dot{V}_E$ represents total expired ventilation, $\dot{V}_P$ is the hypoxia-dependent ventilation (related to the peripheral chemoreflex) and $\dot{V}_C$ is the hypoxia-independent ventilation (related to the central chemoreflex). In these experiments isocapnia is maintained throughout, so in the modelling we assume that $\dot{V}_C$ is constant. In the steady-state, $\dot{V}_P = G_P \times \mathrm{Hyp}$, where G_P is the "hypoxic gain" or the ventilatory sensitivity to acute hypoxia and Hyp is the hypoxic stimulus (a non-linear function of $P_{ET_{O_2}}$). Hyp is derived from the haemoglobin saturation function described by Severinghaus, given by $SO_2 = \left(\frac{23400}{(PO_2{}^3 + 150\, PO_2)} + 1\right)^{-1}$ [124]. Conversion of $P_{ET_{O_2}}$ to O_2 desaturation (Hyp) is performed before the dynamics. If the hypoxic stimulus varies with time, the dynamic response of $\dot{V}_P$ can be represented by $\tau \frac{d\dot{V}p}{dt} + \dot{V}_P = G_P \times \mathrm{Hyp}(t - D)$ and hence, $\tau \frac{d\dot{V}_E}{dt} + \dot{V}_E - \dot{V}_C = G_P \times \mathrm{Hyp}(t - D)$, where τ is the time constant and D is the pure delay. The model was fit for each individual experiment in conjunction with a stochastic model based on a

Kalman filter to describe the correlation that is present between successive breaths [125]. Parameters τ, D, G_P and $\dot{V}_C$ were estimated for each dataset. All of the parameters were constrained to be positive and the time parameters < 30 s. The best fit parameters were estimated by nonlinear regression (*Fortran Library routine EO4FDF, Numerical Algorithms Group, Oxford*) to minimize the sum of squares of the residuals.

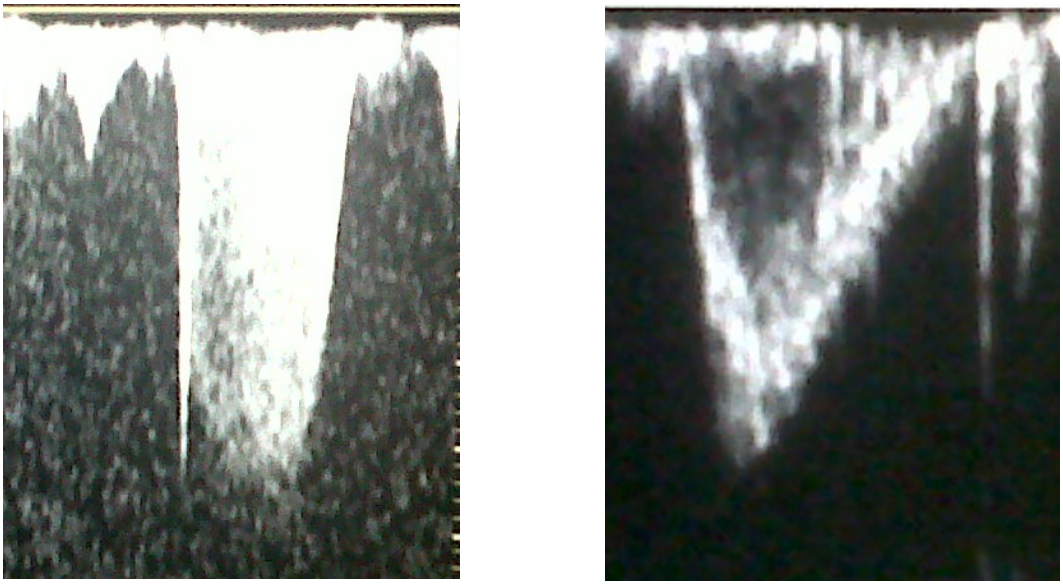
2.1.5 Echocardiography

Echocardiographic measurements were performed using Doppler echocardiography (*Vivid-i, GE Healthcare, Little Chalfont, UK*).

Most people have detectable regurgitation through their tricuspid valve during systole. The peak pressure difference (ΔPmax) across the tricuspid valve was calculated using the modified Bernoulli's equation, $\Delta\text{Pmax} = \rho \frac{v^2}{2}$, where ρ is the density of blood and v is the measured peak velocity of the tricuspid regurgitation jet. Two dimensional echocardiography was used to visualise the tricuspid valve in an apical four-chamber view; colour doppler format allowed detection of the regurgitant jet. The doppler beam was aligned and continuous-wave spectral analysis at a sweep speed of 50 mm/s was used to record the velocity profile of the jet. During analysis, the peak velocity of the jet was measured by placing an electronic calliper tool at the most distal part of the "envelope", and this velocity was transformed (using the Bernoulli's equation) by the programme's software into a ΔPmax value. Up to 5 measurements per minute of protocol were recorded. Figure 6A shows a typical echocardiographic image of the "envelope" of tricuspid regurgitation. The pulmonary arterial systolic pressure is given by $\text{PASP} = \text{RAP} + \Delta\text{Pmax}$, where RAP is the right atrial pressure. In this study, RAP was assumed constant and unaffected by hypoxia [126] and thus changes in ΔPmax mirror changes in PASP, and

hence are used here as a measure of pulmonary vascular response to hypoxia. This technique has been extensively validated [127-131]. ΔPmax under conditions of euoxia (Euoxic ΔPmax) and ΔPmax under conditions of hypoxia (Hypoxic ΔPmax) were calculated for each volunteer by averaging all the ΔPmax measurements noted in the first 5 min of isocapnic euoxia and in the last 6 min of isocapnic hypoxia of Protocol 2, respectively.

In an apical five-chamber view, the flow through the ventricular outflow tract and aortic valve during systole was identified using colour doppler mode and sampled just below the orifice of the aortic valve. The velocity profile was obtained in a pulse-wave spectral mode at a sweep speed of 100 mm/s. An automated procedure was then used to calculate the velocity time integral (VTI) of the flow, which represents the distance that blood travels in the outflow tract with each ventricular contraction. Cardiac Output ($\dot{Q}$) was calculated using $\dot{Q} = SV \times HR$ where SV is stroke volume and HR is heart rate. $SV = VTI \times CSA$, where CSA is the measured cross-sectional area of the aortic valve (in a parasternal long-axis view). Figure 6B shows a typical echocardiographic image. Up to 5 $\dot{Q}$ measurements per minute of Protocol 3 were recorded. $\dot{Q}$ under conditions of euoxia (Euoxic $\dot{Q}$) and $\dot{Q}$ under conditions of hypoxia (Hypoxic $\dot{Q}$) were calculated for each volunteer by averaging all the $\dot{Q}$ measurements noted in the first 5 min of isocapnic euoxia and in the last 6 min of isocapnic hypoxia of Protocol 3, respectively.



A B

Figure 6: Echocardiography images. (A) Continuous-wave spectral analysis showing an "envelope" of tricuspid regurgitation; (B) Pulse-wave spectral analysis showing a velocity time profile of blood flow used to calculate cardiac output.

2.1.6 Plasma erythropoietin (Epo) measurements

Venous blood samples were collected in EDTA blood tubes (*BD Vacutainer*). The plasma was separated by centrifugation and frozen at -80 ^{o}C until analysis. Plasma Epo levels were determined by a human Epo enzyme-linked immunoabsorbent assay (ELISA) (*Quantikine, R&D Systems, Abingdon, UK*), according to manufacturer's protocol, as described below.

The principle of the human Epo ELISA is based on the "double-antibody sandwich" method, as shown in Figure 7. A 96-well microplate coated with a mouse monoclonal antibody against recombinant human erythropoietin (Epo) was used. A calibration series of standard samples – of known recombinant human erythropoietin concentration, namely 0.0 mIU/ml, 2.5 mIU/ml, 5.0 mIU/ml, 20.0 mIU/ml, 50.0 mIU/ml, 100.0 mIU/ml and 200.0 mIU/ml – was arranged on the plate. 100 ml of standard (calibrator) or plasma specimen sample was added in each well, in duplicate. The plate was incubated for 1 hour at room temperature (RT) with shaking (500 rpm). During this time, the erythropoietin present in

each sample binds to the immobilised antibody on the plate. Thereafter, following thorough aspiration of the contents of each well, 200 ml of erythropoietin conjugate (rabbit polyclonal antibody against recombinant human Epo, conjugated to horseradish peroxidase) was added to each well, and the plate was incubated for a further 1 hour at RT with shaking. During this time, the conjugate binds to the immobilised Epo in each well. Following four wash steps with wash buffer and blot-drying of the plate, 200 ml of substrate solution was added to each well and the plate was incubated for 25 min at RT. The chromogen in the substrate solution is oxidised by the enzyme reaction (peroxidase) to form a blue-coloured complex. The reaction was stopped by adding 100 ml of acidic stop solution to each well, which turns the blue colour to yellow. The amount of colour generated is proportional to the amount of conjugate bound to the immobilised Epo and hence proportional to the amount of Epo in the specimen sample or standard. The optical density of each well was determined within 15 min, using a microplate reader set at 450 nm with a wavelength correction set at 600 nm. A standard curve was generated by plotting absorbance versus concentration of the Epo standard samples. The Epo concentration of the unknown plasma specimen samples was determined by comparing the optical density of the specimen samples to the standard curve. The average of the duplicates was taken as the final reading for each sample.

2.1.7 Other blood investigations

Full blood count (FBC), serum iron, ferritin, transferrin and transferrin saturation were determined in a clinical laboratory using standard procedures, for each volunteer.

2.1.8 Statistical analysis

Student's t-test (*SPSS 19),* ANOVA (*SPSS 19*) and fitting of linear mixed-effects models (*R 2.13.0*) were used to analyse the data, as explained in the results section. A number of different linear mixed-effects models were employed appropriate to the particular analysis undertaken. The usual diagnostic analyses were carried out on the model residuals, confirming the requirements of normality and stability of variance. Statistical significance was reported at $p < 0.05$.

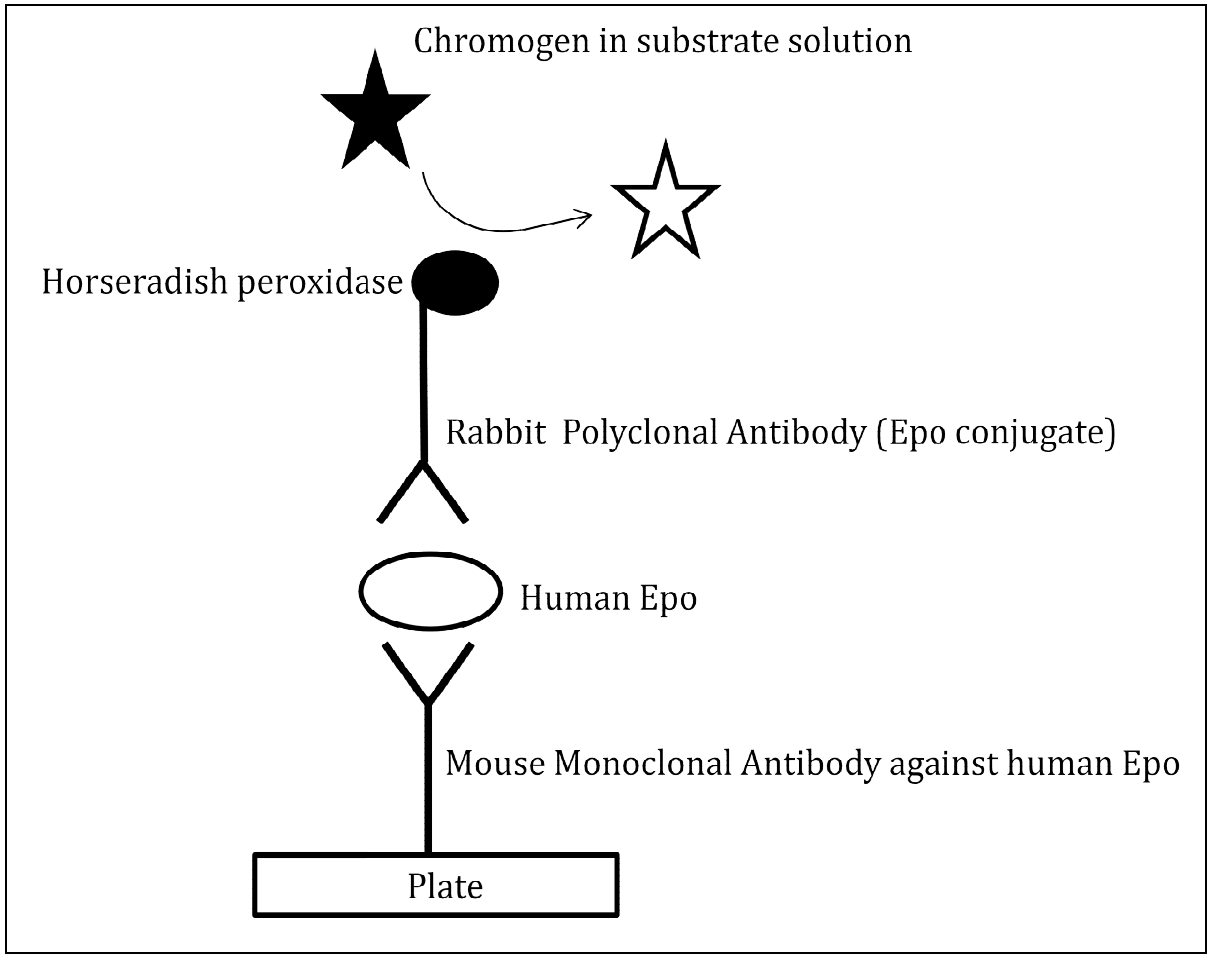


Figure 7: The principle of "sandwich" ELISA as explained in section 2.1.6.

2.2 RESULTS

2.2.1 Baseline (air-breathing) parameters

<u>Main finding</u>: ***At sea level, Tibetans have significantly lower haemoglobin concentration and haematocrit but significantly higher resting ventilation relative to metabolism, as indicated by the lower $P_{ET_{CO_2}}$, than Han Chinese.***

The baseline characteristics and physiological parameters of the participants are summarised in Table 1.

Table 1: Basic characteristics and physiological parameters during air-breathing

PARAMETER	**TIBETAN (n=14)** *MEAN ± SD*	**HAN CHINESE (n=13)** *MEAN ± SD*	**P-VALUE (unpaired t test)**
Age (years)	30.9 ± 5.5	27.6 ± 5.4	NS
Height (m)	1.73 ± 0.05	1.74 ± 0.04	NS
Weight (kg)	73.8 ± 9.1	70.5 ± 8.2	NS
Haemoglobin (g/dL)	14.2 ± 0.9	15.3 ± 1.2	**< 0.05**
Haematocrit (L/L)	0.43 ± 0.03	0.47 ± 0.04	**< 0.05**
Ferritin (μg/L)	214 ± 99	168 ± 86	NS
Transferrin Saturation (%)	35 ± 9	34 ± 8	NS
$P_{ET_{CO2}}$ (mmHg)	37.8 ± 2.4	40.7 ± 2.5	**< 0.005**

2.2.2 Pulmonary vascular response to acute hypoxia before and after 8 hours of sustained hypoxia

<u>Main finding</u>: Tibetans living at sea level have a blunted pulmonary vascular response to acute hypoxia and show a different pattern of early pulmonary vascular acclimatization to sustained hypoxia compared with Han Chinese lowlanders.

Figure 8 shows the mean ΔPmax (index of pulmonary arterial systolic pressure) at each minute of the acute hypoxic Protocol 2 for Tibetan and Han Chinese volunteers, before and after an 8-hr exposure to sustained isocapnic hypoxia in the chamber. Mean values of $G_{\Delta Pmax}$, which is the sensitivity of the pulmonary vascular response to acute hypoxia, and of "Euoxic ΔPmax" and "Hypoxic ΔPmax" are shown in Table 2. Table 2 also summarises the other physiological parameters measured in the acute hypoxic protocols before and after 8 hours of sustained hypoxia in the chamber. A detailed dataset displaying ΔPmax measurements for each individual is found in Table 3.

Table 2: Physiological parameters before and after 8 hours of sustained isocapnic hypoxia

	TIBETAN (n=14)		**HAN CHINESE (n=13)**	
PARAMETER	*Prior to 8-hour sustained hypoxia*	*Post 8-hour sustained hypoxia*	*Prior to 8-hour sustained hypoxia*	*Post 8-hour sustained hypoxia*
$\dot{V}_C$ (L/min)	10.7 ± 2.4	19.7 ± 4.9	12.6 ± 3.0	18.9 ± 9.1
G_P (L/min/%)	0.36 ± 0.14	0.98 ± 0.58	0.50 ± 0.46	0.97 ± 0.92
Euoxic ΔPmax	21.8 ± 3.9	23.2 ± 4.9	19.8 ± 1.7	24. 5 ± 2.7
Hypoxic ΔPmax	25.8 ± 3.5	28.7 ± 5.3	25.5 ± 2.7	32.6 ± 5.1
$G_{\Delta Pmax}$ (mmHg/%)	0.32 ± 0.12	0.43 ± 0.16	0.45 ± 0.11	0.62 ± 0.20
Euoxic $\dot{Q}$ (L/min)	4.8 ± 0.5	4.9 ± 0.5	4.8 ± 0.5	5.2 ± 0.7
Hypoxic $\dot{Q}$ (L/min)	5.5 ± 0.7	6.0 ± 0.7	5.8 ± 0.6	6.2 ± 0.6
G_Q(ml/min/%)	59 ± 25	82 ± 40	77 ± 18	79 ± 20
Plasma Epo	8.45 ± 3.65	16.10 ± 8.22	8.87 ± 3.20	17.16 ± 9.26

$\dot{\mathbf{V}}_\mathbf{C}$ is the residual ventilation in the absence of hypoxia (in isocapnia). $\mathbf{G_P}$ is the acute hypoxic ventilatory sensitivity. "Euoxic ΔPmax" and "Hypoxic ΔPmax" are the ΔPmax values under conditions of euoxia and hypoxia, respectively. "Euoxic $\dot{\mathbf{Q}}$" and "Hypoxic $\dot{\mathbf{Q}}$ " are the cardiac output values under conditions of euoxia and hypoxia, respectively. $G_{\Delta Pmax}$ is the sensitivity of the pulmonary vascular response to hypoxia and is calculated by $G_{\Delta Pmax}$ = (Euoxic ΔPmax – Hypoxic ΔPmax)/ ΔSa_{O_2}, where ΔSa_{O_2} is the percentage drop in haemoglobin oxygen saturation upon exposure to acute isocapnic hypoxia. Similarly, $\mathbf{G_Q}$ which is the sensitivity of the cardiac output response to hypoxia, is calculated by $\mathbf{G_Q}$ = (Euoxic $\dot{\mathbf{Q}}$ – Hypoxic $\dot{\mathbf{Q}}$)/ ΔSa_{O_2}. Plasma Epo is the concentration of erythropoietin in plasma. The data were statistically analysed by fitting linear mixed-effects models.

Note that 2 out of 14 Tibetan volunteers did not have tricuspid regurgitation and thus do not contribute to any ΔPmax parameters. A further 2 out of 14 did not complete Protocols 1, 2, 3 following the 8 hours of sustained hypoxia, and hence do not contribute to the "post 8-hour isocapnic hypoxia" column, except for plasma Epo. Two out of 13 Han Chinese volunteers did not complete the 8-hour sustained hypoxia exposure and hence do not contribute to the "post 8-hour isocapnic hypoxia" column. A further 1 out of 13 did not complete Protocols 2 and 3 following the 8-hour exposure to sustained hypoxia and hence does not contribute to any ΔPmax or $\dot{\mathbf{Q}}$ parameters in the "post 8-hour isocapnic hypoxia" column.

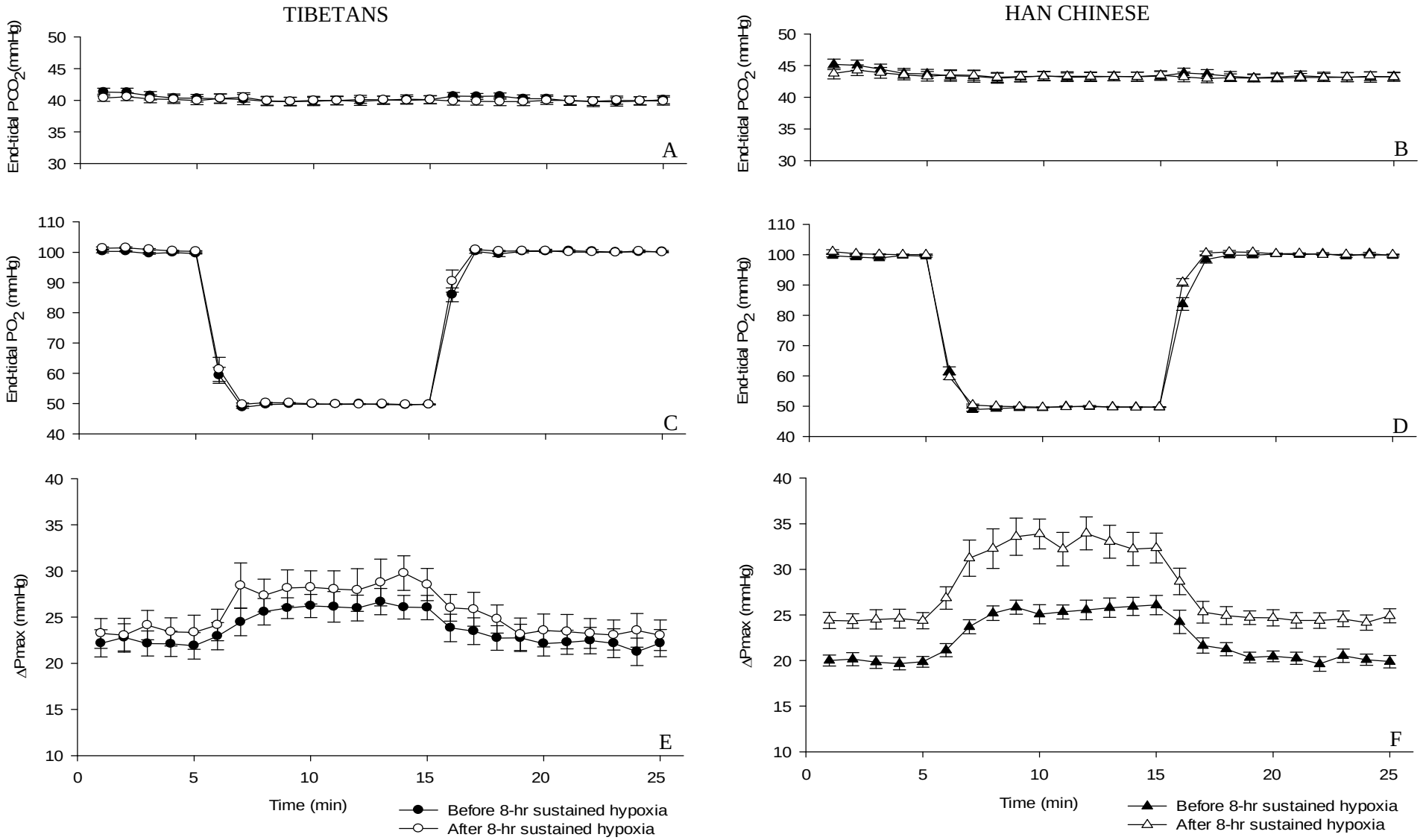


Figure 8: Pulmonary vascular response to acute isocapnic hypoxia before and after 8 hours of sustained isocapnic hypoxia. (A) and (B) show the end-tidal partial pressure of CO_2 against time for Tibetans and Han Chinese, respectively, during the acute hypoxic protocol 2. (C) and (D) show the end-tidal partial pressure of O_2 against time for Tibetans and Han Chinese, respectively. (E) and (F) show the ΔPmax values against time for 10 Tibetan and 10 Han Chinese, respectively. Values are means and error bars represent S.E.M.

Table 3: ΔPmax values during Protocol 2 before and after 8 hours of sustained hypoxia

	Prior to 8-hour isocapnic hypoxia				Post 8-hour isocapnic hypoxia			
Volunteer	**Euoxic ΔPmax (mmHg)**	**Hypoxic ΔPmax (mmHg)**	**ΔPmax change (mmHg)**	**$G_{\Delta Pmax}$ (mmHg/%)**	**Euoxic ΔPmax (mmHg)**	**Hypoxic ΔPmax (mmHg)**	**ΔPmax change (mmHg)**	**$G_{\Delta Pmax}$ (mmHg/%)**
T #1	19.73	26.37	6.64	0.52	22.10	30.70	8.60	0.66
T #3	22.63	28.51	5.88	0.45	24.06	31.93	7.88	0.60
T #4	19.12	24.63	5.51	0.43				
T #5	21.83	27.56	5.74	0.44	21.99	30.09	8.10	0.64
T #6	26.41	28.31	1.90	0.15	26.28	28.96	2.67	0.21
T #7	22.48	24.77	2.29	0.17				
T #8	25.75	28.61	2.86	0.22	28.35	34.00	5.64	0.44
T #9	21.52	24.66	3.15	0.25	21.12	25.02	3.90	0.30
T #11	16.65	20.91	4.26	0.33	17.56	22.56	5.00	0.39
T #12	17.28	21.04	3.76	0.30	19.05	23.65	4.60	0.36
T #13	29.45	32.63	3.18	0.25	33.22	37.99	4.77	0.37
T #14	18.08	21.87	3.79	0.29	18.34	22.20	3.86	0.30
T MEAN	**21.75**	**25.82**	**4.08**	**0.32**	**23.21**	**28.71**	**5.50**	**0.43**
T SD	**3.92**	**3.53**	**1.54**	**0.12**	**4.91**	**5.27**	**2.02**	**0.16**
HC #1	19.08	25.30	6.21	0.48				
HC #2	23.23	27.65	4.42	0.34	25.67	32.82	7.16	0.56
HC #3	21.43	28.29	6.86	0.54	27.13	35.11	7.98	0.62
HC #4	17.024	21.80	4.77	0.37	20.57	26.47	5.90	0.45
HC #5	20.37	27.26	6.89	0.52	24.14	32.22	8.08	0.62
HC #6	21.64	29.33	7.69	0.59	28.63	42.06	13.44	1.00
HC #7	18.55	22.92	4.38	0.34	23.25	31.73	8.47	0.64
HC #8	19.01	24.67	5.66	0.44	21.44	27.49	6.06	0.48
HC #9	20.09	24.49	4.40	0.35	26.11	32.03	5.92	0.46
HC #10	21.01	29.45	8.44	0.65	26.58	38.77	12.20	0.94
HC #11	17.92	21.85	3.93	0.31	21.68	26.78	5.10	0.40
HC #12	18.70	22.95	4.25	0.33				
HC #13	19.14	26.09	6.96	0.53				
HC MEAN	**19.78**	**25.54**	**5.76**	**0.45**	**24.52**	**32.55**	**8.03**	**0.62**
HC SD	**1.71**	**2.71**	**1.51**	**0.11**	**2.72**	**5.09**	**2.77**	**0.20**

In order to take into account repeated measures on the same subject, statistical analysis of the ΔPmax data was carried out by fitting a linear mixed-effects model (*nlme package, R2.13.0*), Model 1, shown below,

$$y_i = \tilde{\alpha} + \beta\, x_i + \gamma\, z_i + \delta\, (x \times z)_i + \varepsilon_i \qquad \text{Model 1}$$

for the i^{th} observation, where x_i is a binary fixed effect representing ethnicity, i.e. being Tibetan or Han Chinese, taking the value 0 for Tibetan and 1 for Han Chinese. z_i is a binary fixed effect representing "before or after 8 hours of sustained hypoxia", with 0 denoting "before the 8-hr sustained hypoxia", and 1 denoting "after the 8-hr sustained hypoxia". $\tilde{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ is a random variable reflecting inter-volunteer variability and $\varepsilon_i \sim N(0, \sigma^2)$ is the inter-observation error. The magnitude of the coefficient β indicates the effect of ethnicity on the dependent variable, y_i, the magnitude of the coefficient γ indicates the effect of 8 hours of sustained hypoxia in the chamber on y_i, and the magnitude of the coefficient δ indicates the effect of the interaction of 8 hours of sustained hypoxia and ethnicity. The usual diagnostic analyses were carried out on the model residuals, confirming the requirements of normality and stability of variance.

For $y = G_{\Delta Pmax}$, the results of the statistical analysis are shown in Table 4.

Table 4: Results of the statistical analysis of the $G_{\Delta Pmax}$ data

	Value	Std. Error	Df	t-value	p-value
Intercept (α)	0.3158	0.0421	23	7.4850	0.0000
β	0.1300	0.0585	23	2.2220	0.0364
γ	0.1075	0.0312	18	3.4505	0.0029
δ	0.0643	0.0440	18	1.4630	0.1607

$\sigma_\alpha = 0.1281$, $\sigma = 0.0703$

Thus, Tibetan volunteers have a significantly lower sensitivity of the pulmonary vascular response to acute hypoxia, $G_{\Delta Pmax}$, than Han Chinese ($\beta = 0.1300$, $p = 0.0364$). Both

Tibetan and Han Chinese significantly increase their $G_{\Delta Pmax}$ following 8 hours of sustained hypoxia ($\gamma = 0.1075$, $p = 0.0029$), which is a feature of pulmonary vascular acclimatization to sustained hypoxia [29]; however, ethnicity does not affect the magnitude of the change in $G_{\Delta Pmax}$ following 8 hours of sustained hypoxia in the chamber ($\delta = 0.0643$, $p = 0.1607$).

For $y = \text{Euoxic } \Delta Pmax$, the statistical analysis results are shown in Table 5.

Table 5: Results of the statistical analysis of the "Euoxic ΔPmax" data

	Value	Std. Error	Df	t-value	p-value
Intercept (α)	21.7442	0.9576	23	22.7070	0.0000
β	-1.9600	1.3280	23	-1.4760	0.1535
γ	1.2921	0.4569	18	2.8277	0.0112
δ	3.2238	0.6457	18	4.9928	0.0001

$\sigma_{\alpha} = 3.1546$, $\sigma = 1.0258$

Thus, Tibetan and Han Chinese volunteers do not differ in their baseline "Euoxic ΔPmax" ($\beta = -1.96$, $p = 0.1535$). However, the rise and persistent elevation in "Euoxic ΔPmax" following 8 hours of sustained hypoxia – another feature of early pulmonary vascular acclimatization to sustained hypoxia [29] – is significantly lower in Tibetan than in Han Chinese volunteers ($\delta = 3.2238$, $p = 0.0001$).

2.2.3 Time-dependent rise in ΔPmax during 8 hours of sustained isocapnic hypoxia

<u>Main finding</u>: Tibetans living at sea level have a blunted pulmonary vascular response to sustained isocapnic hypoxia of 8 hours.

Figure 9 shows the mean ΔPmax at each hourly interval during 8 hours of sustained isocapnic hypoxia in the chamber for Tibetan and Han Chinese volunteers. An average of at least 3 measurements (up to 5) was taken at each time interval. Table 6 shows the results

for each of the volunteers. Good gas control was achieved during the chamber protocol. The mean $P_{ET_{O_2}}$ was 53.1 ± 0.8 mmHg (mean ± sd) for Han Chinese and 52.9 ± 1.2 mmHg for Tibetans. Isocapnia was maintained throughout. The deviation of $P_{ET_{CO_2}}$ from the resting $P_{ET_{CO2}}$ was -1.0 ± 0.8 mmHg for Han Chinese and -1.1 ± 0.4 mmHg for Tibetans.

For the statistical analysis examining the influence of ethnicity on the time-dependent rise of ΔPmax during sustained hypoxia the following linear mixed-effects model was used:

$$y_i = \tilde{\alpha} + \beta x_i + \sum_{k=1}^{8} \gamma_k z_{k_i} + \sum_{k=1}^{8} \delta_k (x \times z_k)_i + \varepsilon_i$$

for the i^{th} observation, where x_i represents ethnicity and is a binary fixed effect and z_{k_i} is a categorical fixed effect representing the k^{th} hourly time point in sustained hypoxia over the 8 hour period. $\tilde{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ is a random variable reflecting inter-volunteer variability and $\varepsilon_i \sim N(0, \sigma^2)$ is the inter-observation error. This showed that the time-dependent rise in ΔPmax in response to sustained hypoxia is significantly lower in Tibetan than in Han Chinese volunteers ($p < 0.001$).

This analysis was repeated using repeated measures ANOVA (*SPSS 19)*, with ΔPmax as the "dependent variable", *time* as a "within-subject factor" and *ethnicity* as a "between-subject" factor, confirming the result that the rate of rise of ΔPmax is significantly smaller in the Tibetan group than the Han Chinese group (*time* × *ethnicity*, p = 0.001, Greenhouse Geisser). Figure 9 shows that the ΔPmax rises more steeply in the first 3 hours of hypoxia and thereafter plateaus. In fact, if pair-wise statistical analysis is carried out between time points in the repeated measures ANOVA analysis (*SPSS 19*), then ΔPmax stops rising significantly beyond 3 to 4 hours.

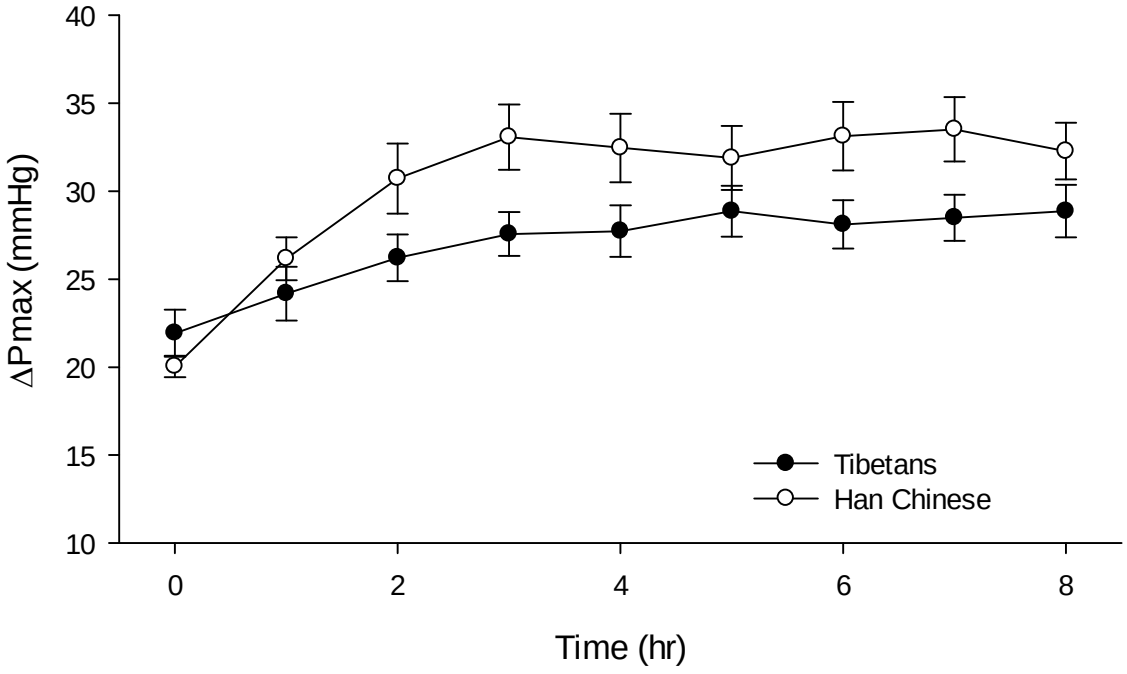


Figure 9: Pulmonary vascular response during 8 hours of sustained isocapnic hypoxia. Closed circles represent ΔPmax values from 10 Tibetan and open circles from 10 Han Chinese volunteers. Values are means and error bars indicate S.E.M.

Table 6: ΔPmax during the 8 hours of sustained isocapnic hypoxia in the chamber

	ΔPmax (mmHg)								
Volunteer	**t=0 hr**	**t=1 hr**	**t=2 hr**	**t=3 hr**	**t=4 hr**	**t=5 hr**	**t=6 hr**	**t=7 hr**	**t=8 hr**
T #1	19.73	20.52	23.29	29.56	29.15	28.74	27.32	28.55	28.55
T #3	22.63	27.82	30.85	31.37	33.21	32.74	32.81	32.41	32.96
T #4	21.83	20.97	25.60	25.30	26.69	29.53	30.43	31.82	28.64
T #6	26.41	30.15	30.55	30.24	31.13	33.65	30.64	29.04	31.76
T #8	25.75	25.74	30.85	31.29	31.17	35.23	33.54	32.78	34.82
T #9	21.52	22.44	24.20	25.30	25.53	25.75	23.99	25.82	25.82
T #11	16.65	19.28	22.00	23.58	26.24	26.27	27.99	26.43	25.87
T #12	17.28	20.93	20.62	22.77	19.99	22.88	20.91	20.79	21.94
T #13	29.45	33.42	31.10	33.20	32.92	31.85	30.77	32.93	35.20
T #14	18.08	20.49	23.07	23.06	21.23	21.93	22.70	24.28	23.16
T MEAN	**21.93**	**24.18**	**26.21**	**27.57**	**27.73**	**28.86**	**28.11**	**28.49**	**28.87**
T SD	**4.23**	**4.85**	**4.18**	**3.95**	**4.63**	**4.57**	**4.34**	**4.13**	**4.71**
HC #2	23.23	24.93	26.88	32.22	29.43	30.22	33.18	36.16	32.73
HC #3	21.43	27.85	41.08	35.47	40.09	38.73	37.82	38.64	36.37
HC #4	20.57	23.62	26.38	29.23	28.64	28.14	29.81	30.45	30.45
HC #5	20.40	28.54	29.74	36.5	32.76	33.36	35.48	37.08	35.00
HC #6	21.64	33.00	31.75	40.48	40.2	40.34	40.88	40.57	41.18
HC #7	19.41	29.17	34.71	39.41	39.7	36.61	35.97	38.77	33.92
HC #8	19.54	20.99	27.13	28.25	26.95	27.55	26.5	28.12	26.9
HC #9	20.09	23.57	25.09	27.53	25.77	24.87	26.19	25.96	26.98
HC #10	21.67	28.47	40.82	38.44	36.06	34.88	41.01	34.76	34.64
HC #11	17.92	21.5	23.59	23.15	24.95	24.08	24.33	24.57	24.57
HC MEAN	**20.03**	**26.16**	**30.72**	**33.07**	**32.46**	**31.88**	**33.12**	**33.51**	**32.27**
HC SD	**1.90**	**3.84**	**6.29**	**5.86**	**6.14**	**5.75**	**6.13**	**5.78**	**5.07**

2.2.4 Acute hypoxic ventilatory response before and after 8 hours of sustained hypoxia

<u>Main finding</u>: Tibetan and Han Chinese volunteers do not differ in their acute hypoxic ventilatory sensitivity (G_P) at sea level and exhibit similar early ventilatory acclimatization to sustained isocapnic hypoxia.

A first order model was used to fit the experimental data obtained from Protocol 1, before and after the 8-hr exposure to sustained isocapnic hypoxia and parameters τ, D, G_P, $\dot{V}_C$ were estimated for each individual. These parameters are shown for each individual in Table 7. Figure 10 shows an example of an individual's response. Mean values for G_P, which is the ventilatory sensitivity to acute hypoxia and $\dot{V}_C$, which is the residual ventilation in the absence of hypoxia (in isocapnia), before and after 8 hours of sustained hypoxia are shown in Table 2. These did not differ significantly between Tibetan and Han Chinese. Using the mixed effect linear Model 1 (*nlme package, R 2.13.0)* to analyse the data, it was found that for both Tibetan and Han Chinese, G_P significantly increased following 8 hours of sustained hypoxia (coefficient γ in the model is significant, $p < 0.001$) – a feature of early ventilatory acclimatization to sustained hypoxia – but the magnitude of this increase did not depend on ethnicity (coefficient δ in the model is not significant, $p > 0.5$). Thus, Tibetan and Han Chinese volunteers exhibit similar early ventilatory acclimatization to sustained isocapnic hypoxia.

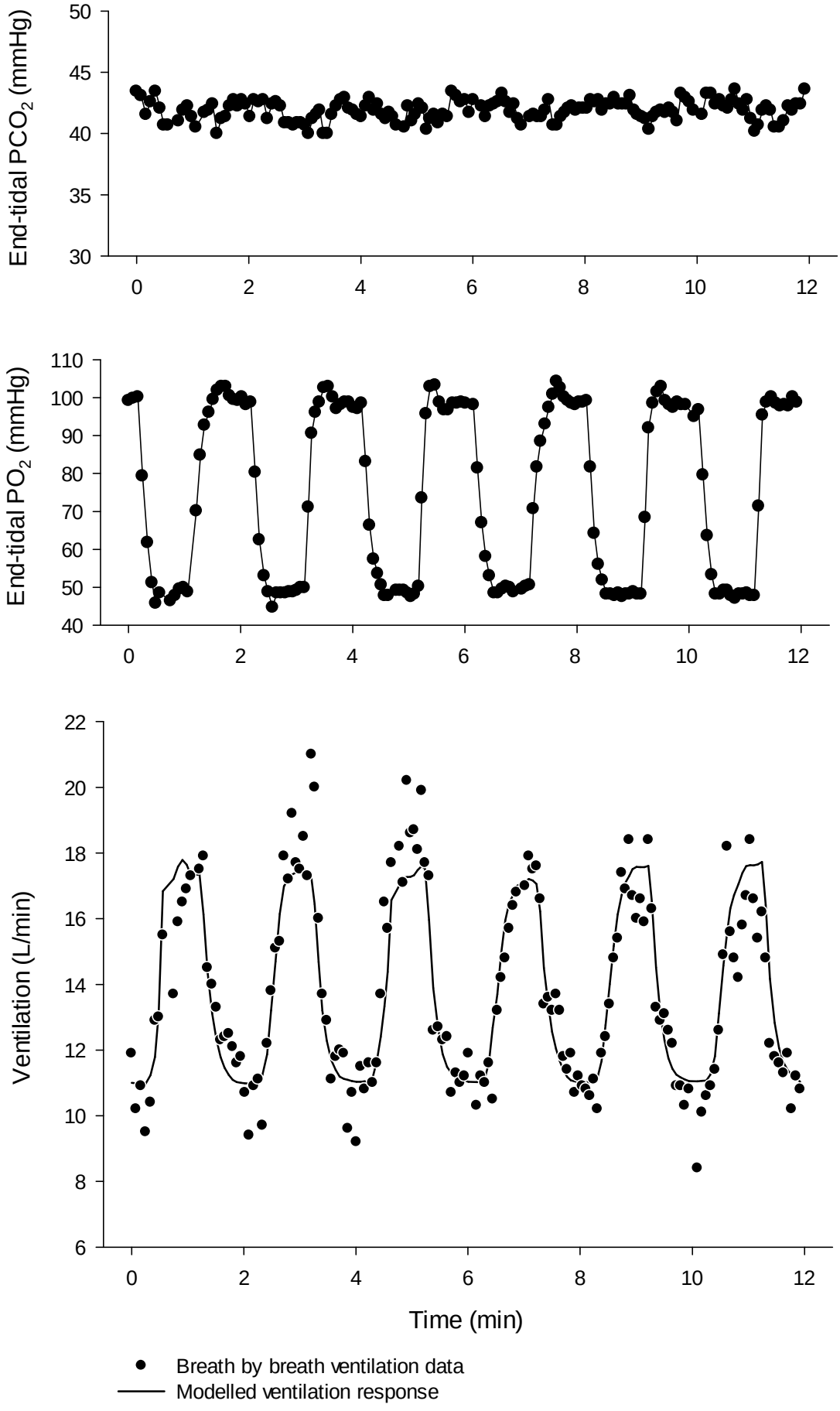


Figure 10: Example of a volunteer's ventilatory response during the acute hypoxic challenge of Protocol 1. (A) and (B) show the end-tidal partial pressure of CO_2 and O_2, respectively, against time. (C) shows the corresponding ventilation measurements (breath by breath) against time.

Table 7: Parameters estimated by modelling the AHVR

	Prior to 8-hour isocapnic hypoxia				Post 8-hour isocapnic hypoxia			
Subject	**G_P (l/min/%)**	**$\dot{V}_C$ (l/min)**	**τ (s)**	**D (s)**	**G_P (l/min/%)**	**$\dot{V}_C$ (l/min)**	**τ (s)**	**D (s)**
T #1	0.32	14.6	2.8	6.9	0.61	27.9	14.5	3.3
T #2	0.52	11.0	7.6	6.8	0.71	19.1	6.2	5.7
T #3	0.22	9.5	3.0	4.1	0.59	16.9	29.9	1.9
T #4	0.21	9.5	0.8	6.4				
T #5	0.53	12.2	8.2	4.4	0.66	19.0	2.5	6.8
T #6	0.32	8.2	9.5	5.1	0.86	14.8	18.7	4.6
T #7	0.24	12.8	11.1	5.1				
T #8	0.43	8.1	8.7	6.7	1.08	18.6	24.7	0.4
T #9	0.27	10.7	6.9	4.3	1.58	22.3	9.0	3.9
T #10	0.38	11.8	2.2	10.5	1.70	20.0	16.8	3.9
T #11	0.68	14.8	0.2	10.0	2.14	28.7	24.7	3.3
T #12	0.22	9.2	0.5	3.8	0.12	20.8	1.1	6.1
T #13	0.27	6.8	13.9	0.5	1.15	15.5	17.5	0.9
T #14	0.43	9.9	8.2	2.4	0.57	12.5	6.4	4.2
T MEAN	**0.36**	**10.7**	**6.0**	**5.5**	**0.98**	**19.7**	**14.3**	**3.8**
T SD	**0.14**	**2.4**	**4.3**	**2.7**	**0.58**	**4.9**	**9.4**	**2.0**
HC #1	0.59	9.2	20.5	2.8	1.06	16.7	23.0	2.9
HC #2	0.46	10.0	10.8	5.1	1.01	13.4	30.0	1.4
HC #3	0.06	10.8	13.0	14.0	0.19	12.0	12.1	6.4
HC #4	1.66	16.5	9.9	4.0	3.50	41.3	18.3	2.8
HC #5	0.77	14.5	20.4	3.0	1.43	15.5	23.5	2.7
HC #6	0.12	8.4	14.7	7.2	0.38	21.1	10.7	8.7
HC #7	0.13	10.3	0.5	5.6	0.35	12.0	4.9	5.0
HC #8	0.88	15.9	2.2	6.5	0.97	18.4	2.1	5.3
HC #9	0.22	13.5	5.3	9.1	0.53	11.2	23.6	6.3
HC #10	0.08	18.1	0.7	15.4	0.66	29.5	21.9	6.6
HC #11	0.20	11.9	30.0	1.5	0.56	16.8	11.0	2.8
HC #12	0.93	11.3	24.0	2.1				
HC #13	0.37	13.2	29.9	1.5				
HC MEAN	**0.50**	**12.6**	**14.0**	**6.0**	**0.97**	**18.9**	**16.5**	**4.6**
HC SD	**0.46**	**3.0**	**10.4**	**4.5**	**0.92**	**9.1**	**8.8**	**2.3**

2.2.5 Cardiac output responses to acute hypoxia before and after 8 hours of sustained isocapnic hypoxia

<u>Main finding</u>: Tibetan and Han Chinese volunteers do not significantly differ in their cardiac output response to hypoxia

Figure 11 shows mean cardiac output at each minute of the acute hypoxic Protocol 3 for Tibetan and Han Chinese volunteers, before and after an 8-hr exposure to sustained isocapnic hypoxia in the chamber. Table 2 shows the mean G_Q and "Euoxic $\dot{Q}$" values for the two groups of volunteers. Table 8 shows a complete data set for each individual. Statistical analysis using the mixed-effect linear model 1 (*R2.13.0)*, with $y = G_Q$ or $y =$ Euoxic $\dot{Q}$ showed that there was no significant difference found between Tibetan and Han Chinese volunteers in their cardiac output response to hypoxia – before or after 8 hours of sustained isocapnic hypoxia.

2.2.6 Time-dependent rise in plasma erythropoietin in response to sustained isocapnic hypoxia

<u>Main finding:</u> There was no significant difference in the time-dependent rise in plasma erythropoietin in response to 8 hours of sustained hypoxia between Tibetans and Han Chinese.

Figure 12 shows plasma erythropoietin against time in sustained hypoxia in the chamber for Tibetan and Han Chinese volunteers. Mean values are shown in Table 2. A complete dataset for each volunteer is found in Table 9. Statistical analysis of the data was undertaken by fitting the linear mixed-effect model, Model 1, with y representing plasma

erythropoietin. No statistically significant difference was observed between the two ethnic groups.

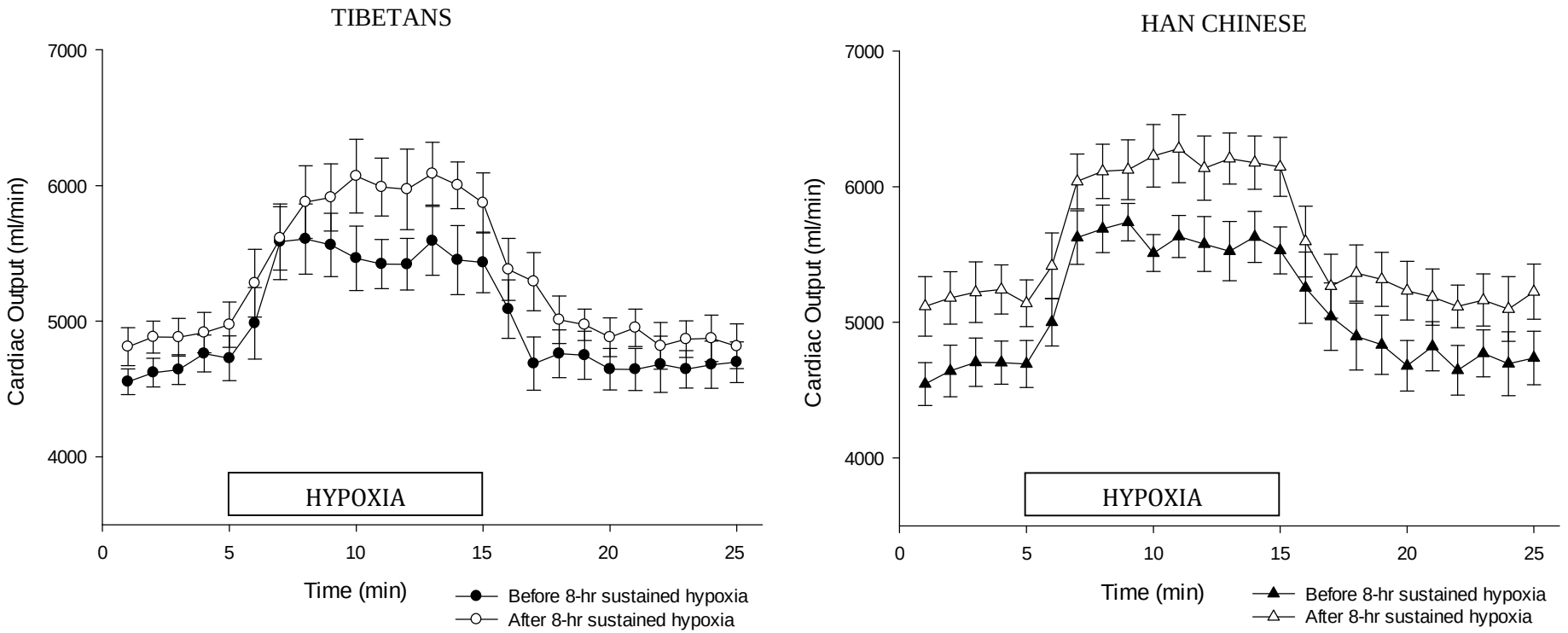


Figure 11: Cardiac output response to acute isocapnic hypoxia before and after 8 hours of sustained isocapnic hypoxia for 12 Tibetan and 10 Han Chinese volunteers. Values are means and error bars represent S.E.M.

Table 8: Cardiac output values during Protocol 3 before and after 8 hours of sustained hypoxia

	Prior to 8-hour isocapnic hypoxia				Post 8-hour isocapnic hypoxia			
Volunteer	**Euoxic $\dot{Q}$ (L/min)**	**Hypoxic $\dot{Q}$ (L/min)**	**$\dot{Q}$ change (L/min)**	**$G_{\dot{Q}}$ (ml/min/%)**	**Euoxic $\dot{Q}$ (L/min)**	**Hypoxic $\dot{Q}$ (L/min)**	**$\dot{Q}$ change (L/min)**	**$G_{\dot{Q}}$ (ml/min/%)**
T #1	4.6	5.3	0.8	61	5.1	6.4	1.4	104
T #2	4.3	4.7	0.4	31	5.1	5.3	0.2	17
T #3	4.1	4.5	0.4	32	4.4	5.0	0.6	47
T #5	5.6	7.1	1.5	116	5.7	7.5	1.9	146
T #6	4.7	5.8	1.1	84	4.6	6.1	1.5	115
T #7	4.4	4.9	4.9	37				
T #8	5.6	6.5	0.9	68				
T #9	4.8	5.3	0.5	37	5.1	5.4	0.4	27
T #10	5.1	6.2	1.1	87	5.3	6.3	0.9	72
T #11	4.4	5.1	0.6	48	4.3	5.6	1.3	100
T #12	4.8	5.4	0.6	46	5.3	6.2	0.8	65
T #13	4.4	5.1	0.7	57	4.9	6.2	1.3	99
T #14	4.9	5.8	0.8	63	4.3	5.8	1.4	110
T MEAN	**4.8**	**5.5**	**0.8**	**59**	**4.9**	**6.0**	**1.1**	**82**
T SD	**0.5**	**0.7**	**0.3**	**25**	**0.5**	**0.7**	**0.59**	**40**
HC #1	4.5	5.4	1.0	76	4.7	5.6	0.9	68
HC #2	5.4	6.6	1.2	94	6.1	7.0	0.9	73
HC #3	5.0	5.6	0.6	45	5.7	6.7	0.9	72
HC #4	5.8	7.1	1.3	102				
HC #5	4.9	5.9	1.0	73	4.8	5.7	0.9	65
HC #7	4.3	5.2	0.9	72	4.3	5.5	1.2	90
HC #8	5.2	6.0	0.7	56	5.8	7.1	1.3	100
HC #9	4.5	5.4	0.9	69	4.9	6.3	1.4	109
HC #10	4.1	5.2	1.1	85	5.0	6.0	1.1	83
HC #11	4.0	4.9	0.9	71	5.0	5.6	0.6	47
HC #12	5.0	6.4	1.4	105				
HC #13	4.9	5.8	0.9	70				
HC MEAN	**4.8**	**5.8**	**1.0**	**77**	**5.2**	**6.2**	**1.0**	**79**
HC SD	**0.5**	**0.6**	**0.2**	**18**	**0.6**	**0.6**	**0.2**	**20**

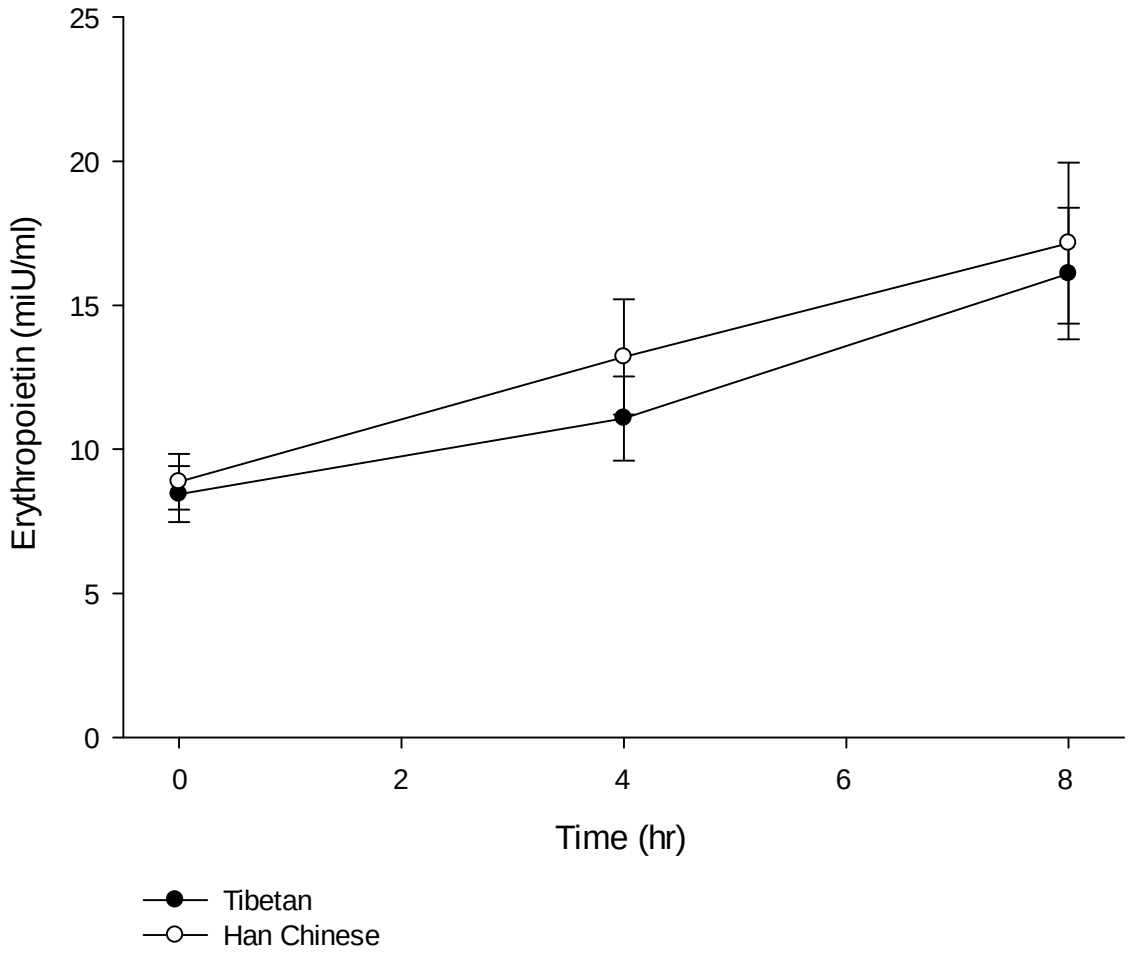


Figure 12: Plasma erythropoietin against time in sustained hypoxia in the chamber for 14 Tibetan and 11 Han Chinese volunteers. Values are means and error bars indicate S.E.M.

Table 9: Plasma erythropoietin values during 8 hours of sustained hypoxia

Volunteer	Epo at t = 0 hr (miU/ml)	Epo at t = 4 hr (miU/ml)	Epo at t = 8 hr (miU/ml)
T #1	8.94	15.17	24.42
T #2	3.15	2.84	3.54
T #3	9.26	13.50	
T #4	5.08		11.35
T #5	16.15	24.45	35.42
T #6	8.60	9.85	16.63
T #7	6.18	7.34	14.47
T #8	15.82	14.20	12.78
T #9	8.07	8.80	18.42
T #10	6.72	13.84	23.31
T #11	6.87	7.96	18.26
T #12	9.27	8.56	12.82
T #13	5.73	9.07	9.14
T #14	8.41	8.36	8.78
T MEAN	8.45	11.07	16.10
T SD	3.65	5.28	8.22
HC #1	8.42	15.02	23.14
HC #2	10.82	13.91	20.72
HC #3	8.29	14.10	17.47
HC #4	8.39	11.34	14.31
HC #5	8.20	11.77	16.56
HC #6	6.80	10.42	14.92
HC #7	16.30	31.81	40.70
HC #8	5.30	7.81	8.60
HC #9	9.81	12.03	16.12
HC #10	10.92	9.07	7.00
HC#11	4.36	8.06	9.23
HC MEAN	8.87	13.21	17.16
HC SD	3.20	6.62	9.26

2.3 DISCUSSION

This study demonstrates that Tibetan natives living at sea level exhibit a significantly different integrative-physiology phenotype compared with Han Chinese lowlanders. The differences are observed across a range of physiological variables of relevance to oxygen homeostasis. The differences relate both to basal measurements made during air-breathing conditions and measurements in response to hypoxic challenges and encompass erythropoietic, pulmonary vascular and ventilatory regulation.

The number of Tibetan volunteers we could recruit was severely limited by the relatively small number resident in the UK. Thus, to be detected in this study, any phenotypic difference has to be large while the absence of a statistically significant effect may well simply be the result of a type II statistical error.

In this study, a complicating factor is that all Tibetans spent the first six or more years of their life at high altitude before migrating to live at sea level. For this reason, it cannot be concluded with complete certainty that the differences between Tibetans and Han Chinese arise from genetic rather than environmental factors during growth and development.

However, in the case of $P_{ET}CO_2$, which reflects the set point for the ratio of pulmonary ventilation to metabolism, there is evidence from South American populations that high-altitude residence during growth and development does not affect subsequent values for $P_{ET}CO_2$ determined at sea level. In a study of ~ 200 individuals the average difference in $P_{ET}CO_2$ between high-altitude natives that had emigrated to live at sea level and sea-level natives was 0.1 mmHg (-0.6 to 0.8, 95% C.I.) [132]. In the present study, the difference in $P_{ET}CO_2$ between Tibetan and Han-Chinese sea-level residents was 2.9 mmHg (1.0 to 4.9,

95% C.I). This suggests that this difference in resting ventilation (relative to metabolism) between Tibetans and Han Chinese might have genetic as opposed to environmental origins. The instantaneous respiratory exchange ratios associated with my observations were within the normal range and not different between the two groups, suggesting that the Tibetan volunteers were not simply hyperventilating during these measurements. It is interesting to note, as mentioned in the introduction, that Tibetans maintain higher resting ventilation at high altitude, too, in comparison to other high-altitude residents [105, 133].

In contrast with the results for $P_{ET_{CO_2}}$, Tibetans and Han Chinese at sea level do not differ in their baseline hypoxic ventilatory sensitivity and show a similar early ventilatory acclimatization; this is in keeping with studies in Andean high-altitude natives at sea level, who were found to have similar acute ventilatory responses to short-duration acute hypoxic stimuli and who – upon re-exposure to high-altitude hypoxia – were also found to undergo ventilatory acclimatization in the same manner as sea-level natives [134-136].

There is a lot of controversy in the literature regarding the process of “blunting” through which hypoxic ventilatory sensitivities are reduced following years of high-altitude residence. It has been reported that while Andeans undergo the process of “blunting” Tibetans may be resistant to it, as Tibetans were reported to have a HVR as high as newly acclimatized newcomers, whereas Andeans did not. In addition, some studies reported the persistence of “blunting” in Andeans upon descent and residence at sea level [137-139], in contrast to the studies referenced above [134-136]. These discrepancies are likely to relate to methodological differences in measuring HVR as exposing volunteers to acute hypoxic stimuli of more than 10 min, is likely to capture effects of differences in HVD rather than differences in “true” acute hypoxic ventilatory sensitivities. By employing the acute

cyclical short-duration hypoxic stimuli in Protocol 1 in my study, I avoid the effects of HVD and I estimate the acute hypoxic ventilatory sensitivity, (G_P), more accurately. There are no existing reports of how Tibetans' HVR to acute hypoxia changes after descent to sea level. The finding in this study that neither G_P nor the early acclimatization of G_P differs between Tibetan and Han Chinese at sea level, would be in keeping with Tibetans not having a blunted ventilatory response to acute or sustained hypoxia compared to lowlanders; or if they do, this is reversed after residing at sea level for many years.

It is difficult to relate the results on ventilation that I describe above to the HIF pathway. This is because the evidence from the literature is rather mixed. For example, patients with Chuvash polycythaemia and "Chuvash mice" were found to have an augmented ventilatory response to hypoxia, whereas patients with HIF2α mutations did not [76, 84, 92]; it may be that HIF1α rather than HIF2α plays a primary role in the ventilatory response to hypoxia. However, different findings are reported in HIF2$\alpha^{+/-}$ mice, which did not differ from wild-type in resting ventilation but had an augmented ventilatory response to hypoxia, attributed to an exaggerated carotid body sensitivity [97]. The involvement of the HIF pathway and the relative contribution of HIF1α and HIF2α in setting basal ventilation as well as in the "tuning" of the ventilatory sensitivity to hypoxia is not clear and is likely to be complex.

In this study, even at sea level, the mean haemoglobin concentration and mean haematocrit in the Tibetan volunteers are significantly lower than in the Han Chinese. For haemoglobin concentration, the difference between Tibetans and Han Chinese was around 1 g/dl. Three things are of note. First, this difference is almost certainly smaller than the difference observed between Tibetans and other groups at high altitude; Tibetans were found to have average haemoglobin concentrations typically 1 to 3.5 g/dl lower than their Andean

counterparts or Han Chinese migrants to high altitude [106-109]. This suggests that the differences between Tibetans and Han Chinese relate both to their set point for haemoglobin concentration in the absence of environmental hypoxia, and the induction of haemoglobin by reductions in ambient P_{O_2}.

Second, the difference in haemoglobin concentration comfortably exceeds the maximum difference observed in a number of genome-wide studies searching for genetic determinants of haemoglobin concentration at sea level – the largest effect observed was of 0.13 g/dl per copy of Hb-lowering allele for *TMPRSS6* [140-142]. However, studies of genetic selection investigating high-altitude adaptation have repeatedly identified *EPAS1* and *EGLN1* as loci which have undergone recent positive selection in Tibetans [115-120]. These genes are part of the HIF transcription activation pathway, which is well known for its regulation of erythropoietin and thus haemoglobin concentration [90, 143-146]. In three of these studies, the putatively selected "Tibetan" genetic variants were associated with lower haemoglobin concentrations at high altitude [115, 117, 119]. I did not identify this difference in my study when comparing the Hb of Tibetans with different genotype (see Chapter 3), possibly through a type II statistical error or possibly because the Tibetans are at sea level where the effects of variation within the HIF pathway may be less apparent than in hypoxic environments. Indeed, it is not clear whether genetic variation in *EPAS1* and *EGLN1* is associated with phenotypic differences in haemoglobin concentration only in conditions of extreme hypoxia. To investigate this, a larger study of Tibetan volunteers who could be recruited or segregated by genotype at *EPAS1* or *EGLN1* to determine if genotype significantly correlates with haemoglobin and haematocrit levels would be required.

Third, the iron status parameters are similar in the two ethnic groups. This is important to note as it signifies that the differences observed in haemoglobin concentration are not as a result of differences in iron status, but also because EPAS1, apart from the erythropoietic response, is implicated in iron absorption and homeostasis [147] and because iron is an important cofactor in the HIF pathway and likely to affect "HIF responses".

Perhaps the most striking differences that I observed between Tibetans and Han Chinese were the differences in their pulmonary vascular responses to hypoxia. This study is the first to assess pulmonary vascular responses to hypoxia in Tibetans living at sea level and to demonstrate the marked reduction in sensitivity to both acute and sub-acute (8-hr) hypoxia. Analogous experiments have not been performed in Andeans; however a study of 11 Andeans who descended and lived at sea level for at least 2 years showed a reversal of pulmonary hypertension but a pulmonary vascular response to exercise – hypoxia was not tested – that was exaggerated, similar to that observed at high altitude and greater than that observed in sea-level natives [148]. However, my results are consistent with a previous study of 5 Tibetans resident at 3658 m which found that their pulmonary arterial pressures were within sea-level norms and were little changed by additional exposure to further hypoxia [110].

Hypoxic pulmonary vasoconstriction is a beneficial mechanism at sea level, where it helps to maintain ventilation-perfusion matching within the lung. However, the environmental hypoxia of high altitude causes global pulmonary vasoconstriction, which leads to pulmonary hypertension and an increase in right ventricular work, all of which is likely to be maladaptive. Indeed hypoxia-induced pulmonary hypertension characterises chronic mountain sickness, a disease with high mortality and morbidity in high-altitude

populations, and quite prevalent in Andeans, but with low prevalence in Tibetans [102]. Thus, it may well be that the phenotype of blunted hypoxic pulmonary vascular response is advantageous for survival at high altitude and could have been naturally selected by genetic adaptation in Tibetans.

One attractive hypothesis is that the relative insensitivity of the pulmonary vasculature to hypoxia in Tibetans arises, at least in part, through the genetic variation within the HIF system that has already been identified in studies of genetic selection. In support of this, patients with disease caused by mutations in the HIF system (which lead to increased stabilisation of HIF1α and HIF2α) exhibit both pulmonary hypertension and increased sensitivity of the pulmonary vasculature to hypoxia [76, 84]. Furthermore, mice with heterozygous loss of HIF2α are resistant to development of hypoxic pulmonary hypertension [91].

It is indeed interesting to note that the phenotype of significantly blunted physiological response to hypoxia in Tibetans compared to Han Chinese in this study seems to be specific for the pulmonary vascular response, for which there is stronger evidence in the literature that the HIF pathway (and HIF2α in particular) plays a role in; it is not observed in ventilatory or in cardiovascular (cardiac output) responses which would instead favour a more generalised differential phenotype of the acute chemoreflex hypoxic response.

Finally, it is important to consider that while most of the genomic studies demonstrated an association of genetic variation at *EPAS1* with low haemoglobin concentration at high altitude [5-7], it is not known whether low haemoglobin is the desired phenotype that provides survival benefit and has hence driven the natural selection process in Tibetans. It

may well be that another systemic phenotype is the favourable one with the haemoglobin phenotype being simply a “side effect” of the selection process, owing to the pleiotropism of the HIF system.

Chapter 3: Is the genetic variation at *EPAS1* and *EGLN1* found in Tibetans functional at the integrative physiology level?

3.1 METHODS

Twenty-eight Tibetan DNA samples were collected from Tibetan individuals living in the UK; these included the 14 individuals that took part in the physiology study. DNA was extracted either from blood or from saliva.

3.1.1 DNA extraction from blood

Eight and a half ml of blood was collected from an individual in PAXgene Blood DNA Tubes (*Qiagen*) and the DNA was extracted using the PAXgene Blood DNA Kit (*Qiagen*), according to manufacturer's protocol, as described below.

The blood was mixed with Buffer BG1 (25 ml) and centrifuged for 5 min at 2500×g at RT in a swing-out rotor. The supernatant was discarded. The resulting pellet was washed with Buffer BG2 (5 ml), centrifuged again for 3 min at 2500×g and the supernatant was discarded. The pellet was treated with Buffer BG3/Pre-Analytix Protease mixture (5 ml) and incubated at 65 °C for 10 min. The sample changed colour from light red to light green, indicating that protein digestion had occurred. 100% isopropanol (5 ml) (*Sigma Aldrich, Poole, UK)* was mixed in with the sample until white DNA strands precipitated together. A DNA pellet was recovered following centrifugation for 3 min at 2500×g. Following a wash step with 70% ethanol (5 ml) and centrifugation at 2500×g for 3 min, the supernatant was discarded and the DNA pellet was air-dried for 10 min. The DNA pellet was dissolved in 1 ml of Buffer BG4 by incubation at 65 °C for 1 hour followed by

incubation at RT overnight. The DNA quantity and quality was determined using a spectrophotometer (*Nanodrop ND-100)* and then stored at -20 °C until further use.

3.1.2 DNA extraction from saliva

OG-500 Kits (*Oragene, DNA Genotek, Ottawa, Canada)* were used to collect saliva. To extract DNA the sample were incubated at 50 °C in a water bath for 1 hour. 500 µl of the sample was mixed with 20 µl ($1/25^{th}$ volume) of PT-L2P reagent (*DNA Genotek)* and incubated on ice for 10 min. It was then centrifuged for 5 min at 13000 rpm. The clear supernatant (~ 500 µl) was recovered, mixed with 600 µl of 100% ethanol and allowed to stand for 10 min at RT to allow DNA precipitation. Following centrifugation for 2 min at 13000 rpm at RT, the supernatant was discarded. The DNA pellet was washed with 70% ethanol and was re-suspended in 100 µl nuclease-free water (*Sigma Aldrich*).

3.1.3 Genotyping

All the collected samples were genotyped using the Illumina Infinium Human610-Quadv1 BeadChip. This was undertaken by Dr Kevin Shianna at the Duke's University, USA. TaqMan SNP genotyping assays (*Applied Biosystems*) were used to genotype 3 non-coding *EGLN1* SNPs (rs2275279, rs961154, rs1765811). This was undertaken by Dr Gianpiero Cavalleri and Dr Mark McCormack at the Royal College of Surgeons in Ireland (RCSI). PCR amplification and DNA sequencing was used to genotype an intronic *EPAS1* Hb-associating SNP (rs13006131) and 2 *EGLN1* coding SNPs (rs12097901 and rs18699510) and this was performed by me.

3.1.4 Protocol for genotyping by PCR amplification and sequencing

PCR reactions were set up using a Taq PCR Core Kit (*Qiagen)* as shown in Table 10 and the thermal cycling conditions shown in Table 11. A small portion of each PCR product was mixed with a DNA stain and loaded on a gel – made of 2% agarose dissolved in 1×TBE buffer (Tris-Borate EDTA, Sigma Aldrich) with ethidium bromide at 2 µg/ml – and subjected to electrophoresis. The gel was visualised under UV light to confirm the amplification of a single PCR product of the correct size. The PCR products were subsequently purified using a PCR purification kit (*QIAquick, Qiagen*), according to the manufacturer's protocol. The purified PCR products were quantified using a Nanodrop spectrophotometer and diluted to 1ng/µl for every 100 bp of DNA. These were submitted together with the appropriate primers to *Source Bioscience* (*Oxford*) for DNA sequencing. The primers used for each SNP were designed using the *Primer3* software and were synthesised by *Sigma Aldrich*. They are shown in Table 12.

Table 10: PCR reaction using the Taq PCR Core Kit (*Qiagen*)

Component	**Volume (µl)**
Buffer (x10)	5.0
Taq Polymerase (250 units)	0.5
Q solution	10.0
dNTP mix (10 mM)	1.0
Primer Mix *Forward & Reverse Primers (10 uM)*	1.0
MgCl2 (25 mM)	1.0 - 2.0
Nuclease-free water	29.5 – 30.5
genomic DNA (50 to 100 ng total)	1.0
Total	50.0

Table 11: Thermal cycling conditions used in the PCR reaction of Table 10

Steps	Temperature (°C)	Time	
1	95	3 min	Denature
2: Cycles (x40)	95	30 sec	Denature
	60	45 sec	Primer annealing
	72	45 sec	Primer extension
3	72	10 min	Clean up extension
4	4	∞	

Table 12: PCR primers used for genotyping

SNP ID	Gene	Position	Forward Primer (5′ to 3′)	Reverse Primer (5′ to 3′)	PCR size
rs13006131	*EPAS1*	2:46608542 C>G (intron 12)	AGAGGAGAGACATGGGCAGA	TGCTGATGCCTGTGATGACT	200 bp
rs18699510	*EGLN1*	1:231557623 C>G (exon 1)	GCACAGGCCCTATTCTCTCA	TGCAGCAGTAGAAGGAGCTG	386 bp
rs12097901	*EGLN1*	1:231557255 G>C (exon 1)	TCAGGACTGGAAGAAGCACA	TACAGGTTCGCCTTCTCCTG	345 bp

3.2 RESULTS

3.2.1 Principal Component Analysis (PCA)

Figure 13 shows a principal component analysis (PCA) amongst Asian populations. Data are from Asian populations in HGDP, Hapmap, Beall et al 2010 [115] and this study. This PCA was generated by Dr Gianpiero Cavalleri, a collaborator from the RCSI. Briefly, PCA is a statistical method commonly used in population genetics to identify a structure in the distribution of genetic variation across populations of different ethnic background. In this case, SNP data were used. Samples are projected onto a series of orthogonal axes or components, made up of a linear combination of allelic values across SNPs. The series of components is chosen such that the first principal component explains the greatest possible

variance in the data among all possible axes. The position of samples along the first two principal components is presented.

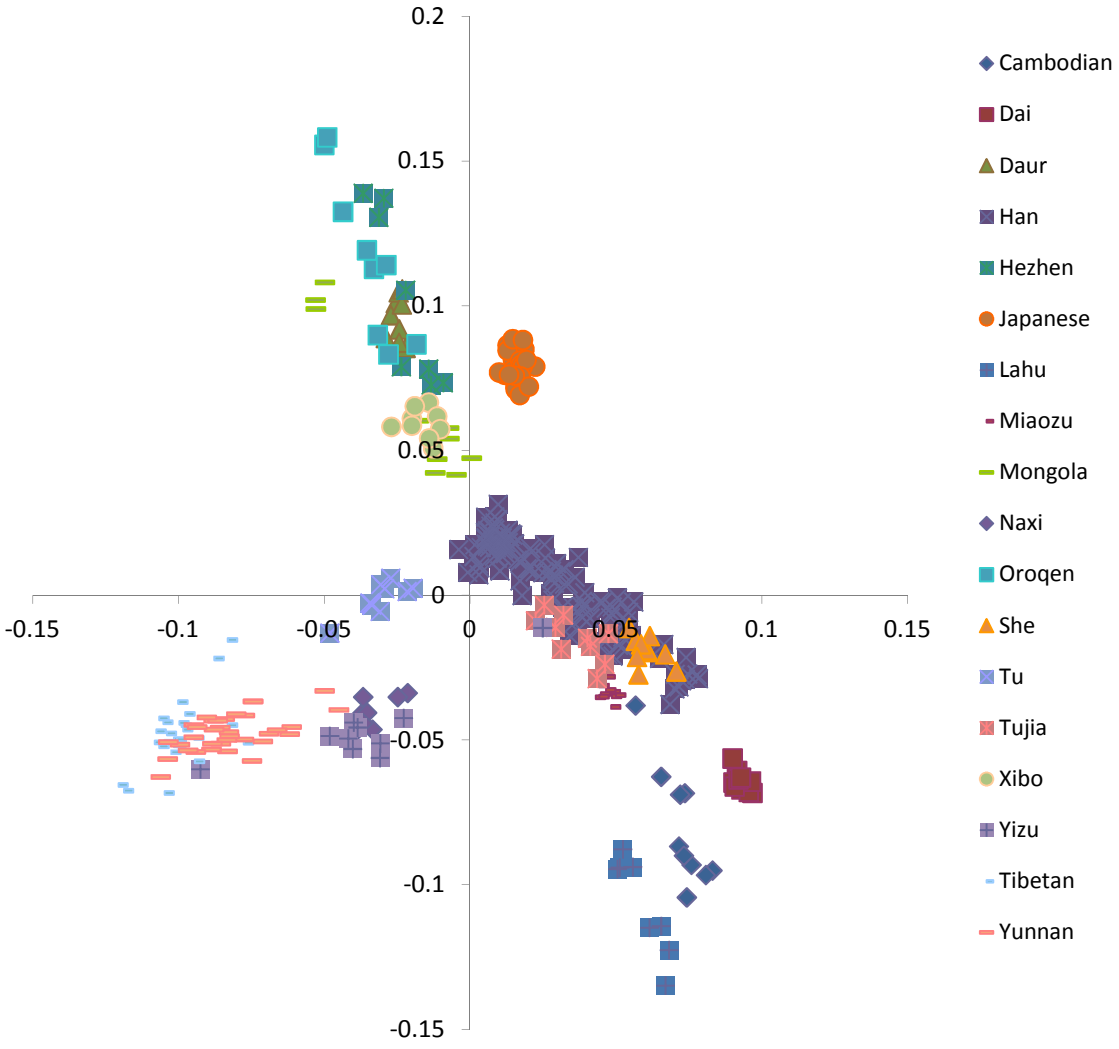


Figure 13: Principal component analysis amongst Asian populations. Data are from HGDP, Hapmap, Beall et al 2010 [115] and this study. The 28 Tibetan DNA samples from our study are represented by light blue short horizontal lines and cluster together with the Yunnan Tibetan volunteers (pink horizontal lines) from [115].

3.2.2 Genotype at the 8 GWAD significant SNPs in *EPAS1*

Table 13 shows the genotype of all the Tibetan DNA samples for the 8 SNPs that achieved genome-wide significance in [115]. These SNPs are all located downstream of the *EPAS1* gene (366 bp to 236 kb downstream of the gene). Table 14 shows the frequencies of each allele for each of the 8 SNPs in my study, in comparison with the corresponding frequencies in the Tibetan-Yunnan and the Han Chinese from [115].

Of the eight genome-wide significant SNPs defining the signal of selection at the *EPAS1* locus in the study by Beall et al [115], rs1868092 was the SNP closest in position to the *EPAS1* gene (366 bp downstream of *EPAS1*) and it was the SNP that correlated most significantly with Hb ($p = 3.39 \times 10^{-4}$). It was also in 100% linkage disequilibrium with rs13006131, which correlated with Hb in an independent sample (in which rs1868092 was not genotyped). Using rs1868092 as a marker for the selected haplotype at *EPAS1*, the Tibetan volunteers that took part in the comparative physiology study were divided into major homozygotes (AA), heterozygotes (AG) and minor homozygotes (GG). Major allele refers to the most frequent allele in Tibetans (alias “Tibetan” allele). Minor allele refers to the rare allele in Tibetans (alias “non-Tibetan” allele). To confirm the strong linkage disequilibrium between rs1868092 and rs13006131 the Tibetan volunteers that participated in the physiology study were also genotyped for rs13006131 and the results are shown in Table 15.

Table 13: Genotype at the 8 GWAD significant SNPs at the *EPAS1* locus

Tibetan volunteer	rs1868092	rs1447563	rs11125075	rs4953388	rs4953396	rs896210	rs982414	rs6735530
T #1	G G	A A	A A	G G	C C	A A	A A	T T
T #2	A G	C A	G A	A G	A C	G A	G A	C T
T #3	A A	C A	G A	A G	A C	G G	G G	C C
T #4	A G	C A	G A	A G	A C	G A	G A	C T
T #5	G G	A A	A A	G G	A C	G A	A A	T T
T #6	A A	C A	G A	A G	A C	G A	G A	C T
T #7	A G	C A	G A	A G	A C	G G	G A	C T
T #8	A G	C A	G A	A G	A C	G A	G A	C T
T#9	A G	C A	G A	A G	A C	G A	G A	C T
T #10	G/C							
T #11	G G	A A	A A	G G	C C	G A	A A	T T
T #12	A A	C C	G G	A A	A A	G G	G A	C T
T #13	G G	A A	A A	G G	C C	A A	A A	T T
T #14	A A	C C	G G	A A	A C	G A	G A	C T
T #15	A A	C C	G G	A A	A A	G G	G G	C C
T #16	A A	C C	G G	A A	A A	G G	G G	C C
T #17	A A	C C	G G	A A	A C	G A	G A	C T
T #18	A A	C C	G G	A A	A A	G G	G G	C C
T #19	A A	C C	G G	A A	A A	G G	G A	C T
T #20	A A	C C	G G	A A	A A	G G	G G	C C
T #21	A G	C A	G A	A G	A C	G G	G G	C C
T #22	A G	C A	G A	A G	A C	G A	G A	C T
T #23	A G	C A	G A	A G	A C	G A	G A	C T
T #24	A G	C A	G A	A G	A C	G G	G A	C T
T #25	A G	C A	G A	A G	A C	G A	G A	C T
T #26	A G	C A	G A	A G	A C	G A	G A	C T
T #27	A G	C A	G A	A G	A C	G A	A A	T T
T #28	G G	A A	A A	G G	C C	A A	A A	T T

homozygous for the major allele
heterozygous
homozygous for the minor allele

Table 14: Allelic frequencies of the 8 GWAD *EPAS1* SNPs

GWAD HIT SNP	Allele A1	Fr (A1) Tibetan Yunnan [115]	Fr (A1) Han Chinese [115]	Fr (A1) Tibetan (this study)	Allele A2	Fr (A2) Tibetan Yunnan [115]	Fr (A2) Han Chinese [115]	Fr (A2) Tibetan (this study)	P -value (Yunnan-Han) [115]
rs1868092	A	0.486	0.089	0.593	G	0.514	0.911	0.407	2.95E-07
rs1447563	C	0.529	0.095	0.556	A	0.471	0.905	0.444	5.15E-08
rs11125075	G	0.514	0.089	0.556	A	0.486	0.911	0.444	6.10E-08
rs4953388	A	0.529	0.065	0.556	G	0.471	0.935	0.444	1.59E-09
rs4953396	A	0.514	0.071	0.537	C	0.486	0.929	0.463	7.88E-09
rs896210	G	0.629	0.161	0.630	A	0.371	0.839	0.370	8.95E-08
rs982414	G	0.414	0.030	0.500	A	0.586	0.970	0.500	9.14E-09
rs6735530	C	0.457	0.071	0.500	T	0.543	0.500	0.929	2.04E-07

Table 15: Genotype at rs1868092 and rs130061311

Tibetan volunteer	rs1868092	rs130061311
T #1	G G	CC
T #2	A G	GC
T #3	A A	GG
T #4	A G	GC
T #5	G G	CC
T #6	A A	GG
T #7	A G	GC
T #8	A G	GC
T#9	A G	GG
T #10	GC	GC
T #11	G G	CC
T #12	A A	GG
T #13	G G	CC
T #14	A A	GG

Figure 14 shows a DNA sequencing chromatograph showing the genotype at rs13006131 for samples T#1, T#2 and T#3, as examples.

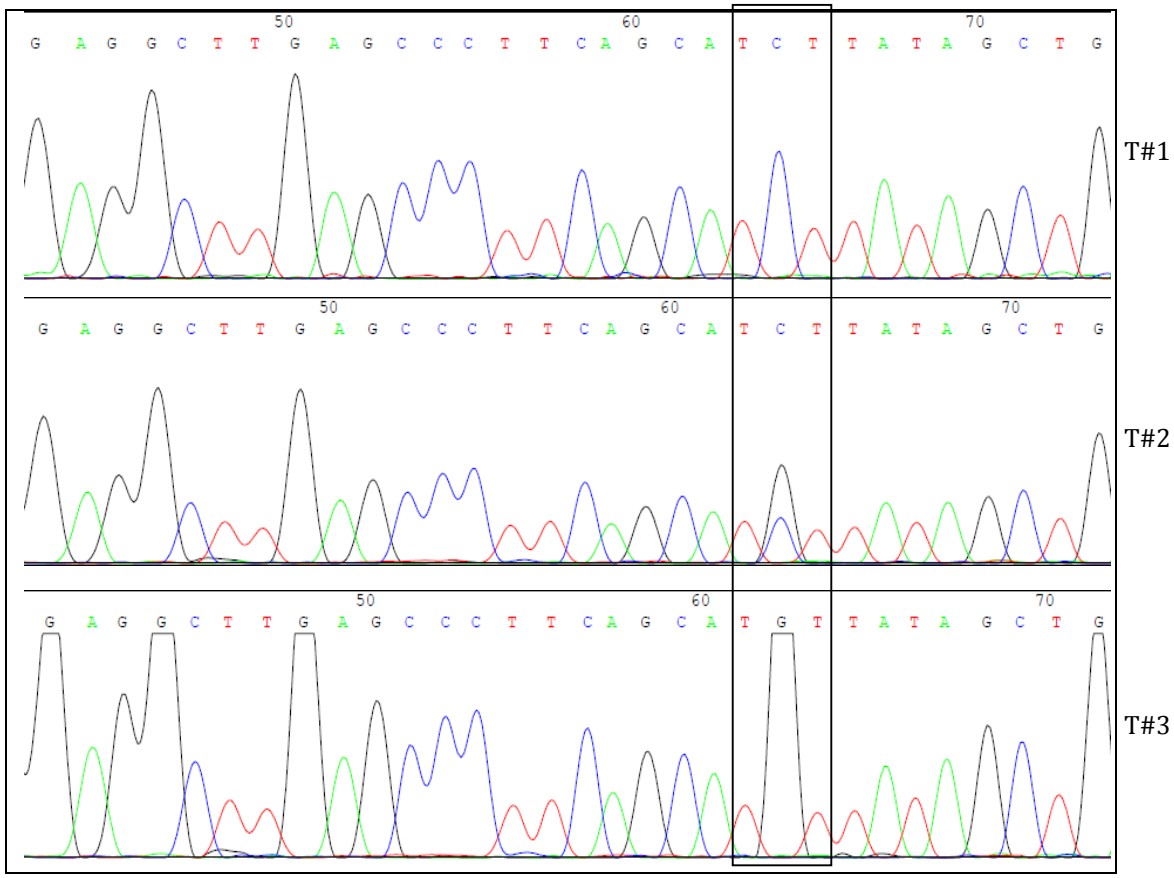


Figure 14: DNA sequencing chromatograph at the *EPAS1* locus showing the genotype for rs13006131 for volunteers T#1 (minor homozygous CC), T#2 (heterozygous CG) and T#3 (major homozygous GG).

3.2.3 Genotype at the *EGLN1* locus

Table 16 shows the results of the genotyping performed at the *EGLN1* locus. The collection of Tibetan DNA samples were genotyped for two *EGLN1* non-coding SNPs, rs2275279 and rs961154, which have frequencies that differ significantly between Tibetan and Han Chinese [116, 119] and that formed part of the positively-selected *EGLN1* haplotype that correlated with Hb in [119]. As can be inferred from Table 16, rs2275279 (genotyped by TaqMan probes) is in 100% LD with its proxy rs6655954 (part of the

Illumina SNP platform); similarly rs961154 (genotyped by TaqMan probes) is in 100 % LD with its proxy rs2790860 (Illumina SNP platform). Table 17 shows the allelic frequencies for the two non-coding *EGLN1* SNPs. In addition, the DNA samples were genotyped for 2 coding variants in exon 1 of *EGLN1*, rs12097901 (380G>C that causes a cysteine to serine substitution, C127S) and rs186996510 (12C>G that causes an aspartate to glutamate substitution, D4E) and the genotyping results are also shown in Table 16. Figure 15 shows DNA sequencing chromatographs showing the different genotypes for rs186996510 and rs12097901 for T#2 and T#3 volunteers. These coding variants were recently reported to have a high prevalence and co-occurrence in the Tibetan population [149, 150] and this was confirmed here (Table 18). I demonstrated a high degree of linkage disequilibrium between these 2 non-coding SNPs and the 2 coding variants in exon 1, suggesting the presence of a haplotype in this *EGLN1* region. In fact, rs12097901 (C127S) was highly correlated with both rs2275279 and rs961154 – which comprised the *EGLN1* selected haplotype – whereas rs186996510 was not; see Table 19. Finally, the samples were genotyped for rs1765811, an *EGLN1* SNP shown to be highly differentiated in the Andean population for comparison – but this is discussed in Chapter 8.

Table 16: Genotype at SNPs at the *EGLN1* locus

Tibetan volunteer	rs2275279	rs6655954 (proxy of rs2275279)	rs961154	rs2790860 (proxy of rs961154)	rs186996510 (C>A, D4E)	rs12097901 (G>C, 127S)	rs1765811
T #1	A T	A G	C T	A G	CC	CG	G G
T #2	T T	G G	T T	G G	GG	CC	G G
T #3	A T	A G	C T	A G	GC	CG	G G
T #4	T T	G G	T T	G G	GG	CC	G G
T #5	A T	A G	C T	A G	GC	CG	G G
T #6	A T	A G	C T	A G	GC	CG	G G
T #7	A T	A G	C T	A G	GG	CC	G G
T #8	T T	G G	T T	G G	GC	CC	G G
T#9	A T	A G	C T	AG	GC	CG	G G
T #10	T T	00	T T	00	GG	CC	G G
T #11	T T	G G	T T	G G	GC	CC	G G
T #12	T T	G G	T T	G G	GG	CC	G G
T #13	A T	A G	T T	G G	CC	CC	G G
T #14	A T	A G	C T	A G	GG	CC	G G
T #15	A T	A G	C T	AG	GC	CG	G G
T #16	T T	G G	T T	GG	GC	CC	G G
T #17	T T	G G	T T	GG	GG	CC	G G
T #18	A T	A G	T T	GG	GC	CG	A G
T #19	T T	G G	T T	GG	GG	CC	G G
T #20	A T	A G	C T	AG	CC	CG	G G
T #21	A T	A G	C T	AG	GG	CC	G G
T #22	A T	A G	T T	GG	GC	CC	G G
T #23	00	A G	00	AG	GC	CG	G G
T #24	0 0	AG	C T	AG	00	00	G G
T #25	A T	AG	C T	AG	GG	CC	G G
T #26	T T	GG	T T	GG	GG	CC	G G
T #27	T T	GG	T T	GG	00	00	G G
T #28	A A	AA	C C	AA	GC	CG	G G

Table 17: Allelic frequencies of rs961154 and rs2275279

SNP	Allele A1	Fr (A1) Tibetan Yunnan [115]	Fr (A1) Han Chinese [115]	Fr (A1) Tibetan (this study)	Allele A2	Fr (A2) Tibetan Yunnan [115]	Fr (A2) Han Chinese [115]	Fr (A2) Tibetan (this study)
rs961154	T	0.729	0.578	0.722	A	0.271	0.422	0.278
rs2275279	T	0.629	0.319	0.667	C	0.371	0.681	0.333

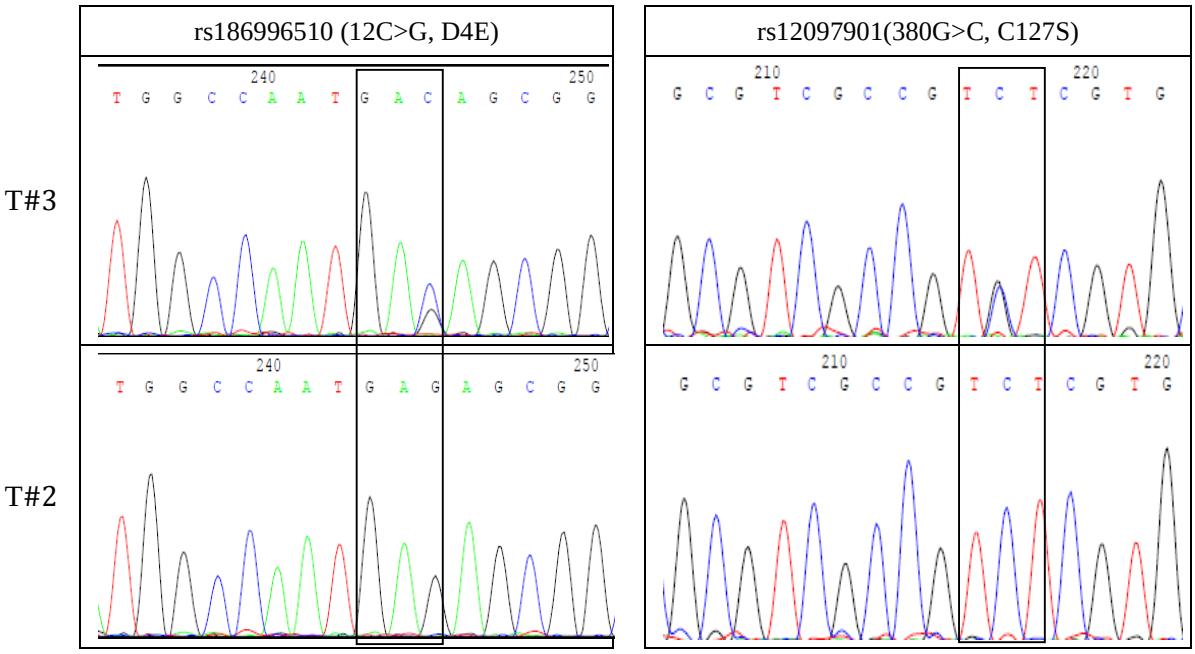


Figure 15: DNA sequencing chromatograph at the *EGLN1* locus showing the genotype for rs186996510 and rs1209701 for volunteer T#2, who is homozygous GG and CC respectively ; and for volunteer T#3, who is heterozygous GC and GC respectively.

Table 18: Allelic frequencies of the two coding variants in *EGLN1*

Coding Variant	Frequency of "Tibetan" or major allele	Frequency of "non-Tibetan" or minor allele
rs12097901 (380G>C, C127S)	0.81 (C)	0.19 (G)
rs186996510 (12C>G, D4E)	0.65 (G)	0.35 (C)

Table 19: Correlation between *EGLN1* SNPs

	rs12097901 (C127S)	**rs186996510 (D4E)**	**rs961154**	**rs2275279**
rs12097901 (C127S)		r^2= 0.613 (p=0.001)	r^2= 0.619 (p=0.002)	r^2= 0.625 (p=0.002)
rs186996510 (D4E)			r^2= 0.215 (p=0.264)	r^2= 0.428 (p=0.03)
rs961154				r^2= 0.816 (p=0.000)
rs2275279				

3.2.4 Association of integrative physiology phenotype with genotype at the *EPAS1* and *EGLN1* loci

Main findings*: *The rise in plasma erythropoietin with time in sustained hypoxia in Tibetan volunteers is significantly associated with genotype at both the EPAS1 locus and the EGLN1 locus.

Figure 16A shows the plasma erythropoietin against time in sustained hypoxia for the 3 *EPAS1* genotypes in the Tibetan volunteers. The rise in plasma erythropoietin with time in sustained hypoxia is significantly higher in the minor homozygous (GG) compared to the major homozygous (AA) Tibetan volunteers ($p < 0.05$). Figure 16B shows the plasma erythropoietin against time in sustained hypoxia for Tibetan volunteers that are homozygous (CC) or heterozygous (CG) for rs12097901 **(**C127S) in *EGLN1* – note that only these 2 genotypes are found in the Tibetan volunteers. The rise in plasma erythropoietin with time in sustained hypoxia is significantly higher in the heterozygous than in the homozygous Tibetan volunteers ($p < 0.01$). Table 20 shows the results by *EPAS1* genotype and Table 21 by *EGLN1* genotype for other physiological parameters – no significant association was shown for any other physiological parameter.

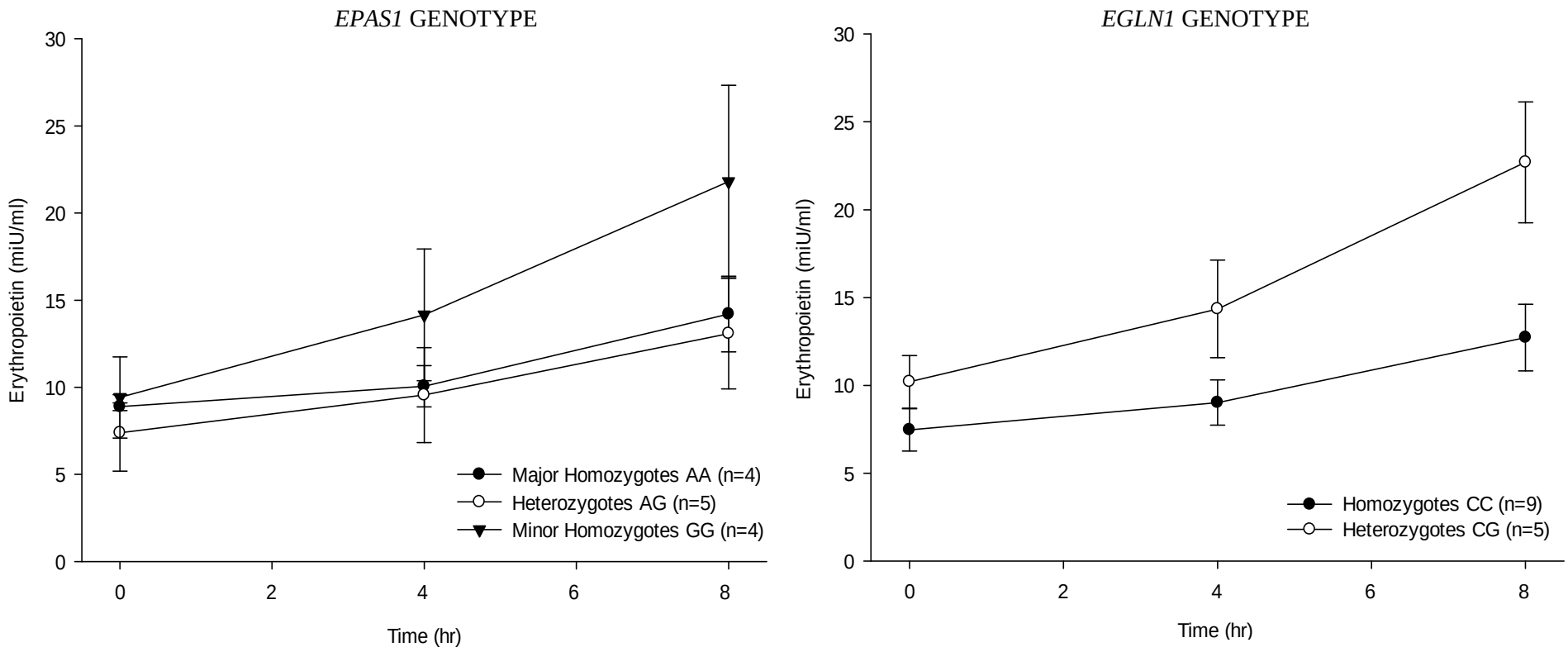


Figure 16: Plasma erythropoietin (Epo) against time in sustained hypoxia in the chamber for Tibetan volunteers according to (A) *EPAS1* genotype and (B) *EGLN1* genotype. Values are means and error bars indicate S.E.M.

Table 20: Physiological parameters before and after 8 hours of sustained isocapnic hypoxia for Tibetan volunteers according to *EPAS1* genotype

PARAMETER	MAJOR ALLELE HOMOZYGOTES (n=4)		HETEROZYGOTES (n=5)		MINOR ALLELE HOMOZYGOTES (n=4)	
	Prior to 8-hr sustained hypoxia	*Post 8-hr sustained hypoxia*	*Prior to 8-hr sustained hypoxia*	*Post 8-hr sustained hypoxia*	*Prior to 8-hr sustained hypoxia*	*Post 8-hr sustained hypoxia*
	MEAN ± SD	*MEAN ± SD*	*MEAN ± SD*	*MEAN ± SD*	*MEAN ± SD*	*MEAN ± SD*
Hb (g/dL)	13.9 ± 0.4		14.1 ±0.5		14.7 ± 1.6	
Haematocrit (L/L)	0.42 ± 0.01		0.44 ± 0.04		0.44 ± 0.05	
$P_{ET_{CO2}}$ (mmHg)	37.8 ± 2.7		37.7 ± 3.0		37.9 ± 1.4	
$\dot{V}_C$ (L/min)	9.2 ± 0.7	16.3 ± 3.5	10.7 ± 1.7	20.0 ± 1.6	12.1 ± 3.7	22.8 ± 6.5
G_P (L/min/%)	0.30 ± 0.10	0.54 ± 0.31	0.34 ± 0.12	1.27 ± 0.46	0.45 ± 0.19	1.14 ± 0.71
Euoxic ΔPmax (mmHg)	21.1 ± 4.3	21. 9 ± 3.9	22.2 ± 2.7		21.9 ± 5.5	23.7 ± 6.7
Hypoxic ΔPmax (mmHg)	24.9 ± 4.0	26.7 ± 4.5	25.7 ± 2.0		26.9 ± 4.8	30.3 ± 6.3
$G_{\Delta Pmax}$ (mmHg/%)	0.30 ± 0.12	0.37 ± 0.17	0.27 ± 0.11		0.4 ± 0.1	0.5 ± 0.2
Euoxic $\dot{Q}$ (L/min)	4.6 ± 0.4	4.7 ± 0.5	4.8 ± 0.5	5.2 ± 0.1	4.8 ± 0.6	5.0 ± 0.6
Hypoxic $\dot{Q}$ (L/min)	5.4 ± 0.6	5.8 ± 0.5	5.5 ± 0.8	5.7 ± 0.5	5.7 ± 1.0	6.4 ± 0.8
$G_{\dot{Q}}$ (ml/min/%)	57 ± 22	84 ± 34	52 ± 24	38 ± 30	70 ± 31	112 ± 23
Plasma Epo (miU/ml)	8.9 ± 0.5	14.2 ± 4.4	7.4 ± 4.9	13.1 ± 7.1	9.4 ± 4.7	21.8 ± 11.0

Major allele refers to the allele that is selected for within the Tibetan population. All the major homozygous and minor homozygous volunteers participated in all protocols of the physiology study and feature in all columns of this table. In view of the fact that 4 out of 5 heterozygous volunteers either did not have tricuspid regurgitation on echocardiography or did not complete the hypoxic protocols following 8 hours of sustained hypoxia the boxes relating to ΔPmax in the "post hypoxia" column are left blank (as only data from 1 heterozygous volunteer are available), to avoid confusion. One individual's genotype could not be classified sand this individual was thus excluded from the analysis.

Table 21: Physiological parameters before and after 8 hours of sustained isocapnic hypoxia for Tibetan volunteers according to *EGLN1* genotype

PARAMETER	HOMOZYGOTES (n=9)		HETEROZYGOTES (n=5)	
	Prior to 8-hr sustained hypoxia	*Post 8-hr sustained hypoxia*	*Prior to 8-hr sustained hypoxia*	*Post 8-hr sustained hypoxia*
	MEAN ± SD	*MEAN ± SD*	*MEAN ± SD*	*MEAN ± SD*
Hb (g/dL)	14.5 ± 1.1		13.80 ± 0.34	
Haematocrit (L/L)	0.44 ± 0.04		0.42 ± 0.01	
$P_{ET_{CO2}}$ (mmHg)	38.2 ± 2.5		36.9 ± 1.9	
$\dot{V}_C$ (L/min)	10.4 ± 2.5	19.3 ± 5.0	11.0 ± 2.5	20.2 ± 5.1
G_P (L/min/%)	0.38 ± 0.16	1.07 ± 0.69	0.33 ± 0.12	0.86 ±. 42
Euoxic ΔPmax (mmHg)	21.3 ± 4.8	23.3 ± 7.1	22.4 ± 2.5	23.1 ± 2.1
Hypoxic ΔPmax (mmHg)	24.9 ± 4.4	28.1 ± 7.4	27.1 ± 1.6	29.3 ± 2.6
$G_{\Delta Pmax}$ (mmHg/%)	0.28 ± 0.08	0.37 ± 0.05	0.36 ± 0.16	0.48 ± 0.21
Euoxic $\dot{Q}$ (L/min)	4.7 ± 0.4	4.7 ± 0.5	4.8 ± 0.6	5.0 ± 0.5
Hypoxic $\dot{Q}$ (L/min)	5.4 ± 0.6	5.8 ± 3.8	5.6 ± 1.0	6.1 ± 1.1
$G_{\dot{Q}}$ (ml/min/%)	53 ± 18	83 ± 34	66 ± 35	88 ± 50
Plasma Epo (miU/ml)	7.47 ± 3.60	12.72 ± 5.70	10.20 ± 3.35	23.72 ± 7.49

For the statistical analysis of the influence of genotype at the *EGLN1* locus on the induction of plasma erythropoietin with time in sustained hypoxia in Tibetans, the following linear mixed-effects model was fitted to the data:

$$y_i = \tilde{\alpha} + \beta x_i + \sum_{k=1}^{2} \gamma_k z_{k_i} + \sum_{k=1}^{2} \delta_k (x \times z_k)_i + \varepsilon_i$$

for the i^{th} observation, where y is plasma Epo, x_i is a binary fixed effect representing genotype – only 2 genotypes are present in the dataset, homozygous or heterozygous – and

z_{k_i} is a categorical fixed effect representing time in sustained hypoxia; when k = 1 it refers to the time point at 4 hours of sustained hypoxia and when k = 2 it refers to the time point at 8 hours of sustained hypoxia. A similar model was used for the genotype at the *EPAS1* locus where x_i was a categorical fixed effect representing the three genotypes, i.e. minor homozygote, heterozygote or major homozygote. $\tilde{\alpha} \sim N(\alpha, \sigma_{\alpha}^{2})$ is a random variable reflecting inter-volunteer variability and $\varepsilon_i \sim N(0, \sigma^{2})$ is the inter-observation error. The results are shown in the tables below; for *EGLN1* see Table 22 and for *EPAS1* see Table 23.

Table 22: Results of the statistical analysis of the Epo data (*EGLN1* genotype*)*

	Value	Std. Error	Df	t-value	p-value
Intercept (α)	10.203	2.238	22	4.559	0.0002
β (genotype)	-2.733	2.791	12	-0.979	0.3467
γ_1 (time at 4 hrs)	4.152	1.949	22	2.131	0.0445
γ_2 (time at 8 hrs)	13.344	2.111	22	6.323	0.0000
δ_1 (genotype × time at 4 hrs)	-2.781	2.468	22	-1.127	0.2720
δ_2 (genotype × time at 8 hrs)	-8.096	2.562	22	-3.160	0.0045

$\sigma_\alpha = 3.943$, $\sigma = 3.081$

Note that here the "reference genotype", i.e. x = 0 means heterozygous while x = 1 refers to homozygous.

Table 23: Results of the statistical analysis of the Epo data (*EPAS1* genotype*)*

	Value	Std. Error	Df	t-value	p-value
Intercept	9.419	2.998	18	3.142	0.0056
genotype – heterozygous	-2.029	4.022	10	-0.505	0.6248
genotype – major homozygous	-0.534	4.239	10	-0.126	0.9022
time at 4 hrs	4.742	2.488	18	1.906	0.0727
time at 8 hrs	12.389	2.488	18	4.980	0.0001
hetorozygous × time at 4 hrs	-2.976	3.465	18	-0.859	0.4017
major homozygous × time at 4 hrs	-3.558	3.518	18	-1.011	0.3253
hetorozygous × time at 8 hrs	-6.688	3.338	18	-2.004	0.0604
major homozygous × time at 8 hrs	-8.030	3.717	18	-2.160	0.0445

$\sigma_\alpha = 4.854$, $\sigma = 3.518$

3.3 DISCUSSION

The first point of note is that the principal component analysis (PCA) confirms that the Tibetan volunteers that were recruited - based on their place of birth/origin and on their grandparents' origin according to the information provided by the volunteers – were indeed of Tibetan genetic origin, as the data points cluster together with the Yunnan Tibetan population from Beall et al [115].

The second point of note is the confirmation that the frequencies of the GWAD significant SNPs at the *EPAS1* locus, the frequencies of the *EGLN1* SNPs comprising putative-selected haplotypes and the frequencies of the *EGLN1* coding variants in the Tibetan cohort in this study are similar to those reported in other Tibetan cohorts. This gives validity to the results.

Studies investigating high-altitude adaptation have repeatedly identified *EPAS1* and *EGLN1* as loci which have undergone recent positive selection in Tibetans [115-120] and in some of these studies, the putatively selected "Tibetan" genetic variants were associated with lower haemoglobin concentrations at high altitude [115, 117, 119]. Despite this, we did not identify any difference in haemoglobin concentration when comparing Tibetans of different genotypes. This could either be the result of a type II statistical error due to the small numbers of Tibetans or it could arise because the Tibetans are at sea level where the effects of variation within the HIF pathway may be less apparent than at environments of hypoxia. In keeping with the latter possibility, we did identify a statistically significant difference in erythropoietin (Epo) regulation between genotype in Tibetans for both *EPAS1* and *EGLN1* during exposure of these individuals to 8 hours of hypoxia. For both genes, the

smaller Epo response to sustained hypoxia was associated with the "Tibetan" variant, as seen in Figure 16A and Figure 16B.

One attractive hypothesis, as mentioned in the discussion in section 2.3, is that the relative insensitivity of the pulmonary vasculature to hypoxia in Tibetans – in comparison with the Han Chinese – reported in Chapter 2, arises, at least in part, through genetic variation within the HIF system. This would be in keeping with the significant divergence of allelic frequencies at the *EPAS1* locus between Tibetan and Han Chinese, with "Tibetan" alleles found at extremely low frequencies in Han Chinese, as shown in Table 14. However, in my comparisons "between genotype" in the Tibetan volunteers, I was not able to show a statistically significant association between genotype and magnitude of the hypoxic pulmonary vascular response, but this may be the result of a type II statistical error. Interestingly, the pulmonary vascular responses to hypoxia are noticeably different between Tibetans of different genotype in absolute physiological terms (Table S1 and Table S4), but these differences fall short of statistical significance and so no conclusion can be drawn. Furthermore, in the case of *EPAS1* (for which there is greater evidence in the literature that it affects integrative cardiopulmonary physiology) but not *EGLN1*, there is evidence of an emerging trend such that the "Tibetan" allele is associated with smaller hypoxic responses across most physiological parameters tested (Table S1).

A novel finding here is the demonstration of high linkage disequilibrium between the non-coding *EGLN1* SNPs – that comprised the putatively selected *EGLN1* haplotype in [119] and the coding variants in exon 1 of *EGLN1* in Tibetans. It is possible that these coding variants, which result in aminoacid substitutions, are the functional variants on the selected haplotype; this needs further investigation and is addressed later on in Chapter 7.

Chapter 4: HIF-regulated gene expression in peripheral blood lymphocytes – a comparison between Tibetans and Han Chinese

4.1 INTRODUCTION

Central to the relationship between genetic variation and observed phenotype at the integrative physiology or systems level is the hypothesis that the underlying genetic variability has functional consequence at an intermediate molecular or cellular level, while complex pathways and feedback systems are also in operation. Studying gene expression both as a qualitative and as a quantitative trait has been a useful tool in looking for associations of disease and other traits with genetic loci, despite the fact that in most cases the causative variant(s) are unknown.

Gene expression has been shown to vary within and between populations of a variety of different species, including humans [151]. There is evidence of inter-population variation, evidence of familial aggregation within a population and also evidence of inter-individual variation in gene expression; embedded in this there is also a great amount of tissue-specific variability. Peripheral blood lymphocytes are an easily accessible population of cells that circulate around the body and thus experience a great diversity of environments in the body. They show a detectable expression of a large number of genes and are a relevant cell type for many human traits. In addition, there are available data of a microarray analysis looking at transcript expression in peripheral blood lymphocytes in euoxic and hypoxic samples. This analysis reports all the genes (ranked according to a hypoxia to normoxia expression ratio) that are hypoxically regulated in PBL (performed by Jerome Brooks, deposited in the National Centre for Biotechnology Information Gene Expression Omnibus database under accession number GSE7833).

Large scale studies using both human peripheral blood lymphocytes [152] and transformed lymphoblastoid cell lines [153, 154] have demonstrated a great degree of heritability in gene expression, reinforcing the presence of a substantial genetic component in the variation observed in gene expression. In addition, distinct patterns of inter-individual variation in gene expression have also been found, e.g. in a study of peripheral blood lymphocytes from 75 healthy volunteers [155]. Population differences in gene expression have also been characterised using lymphoblastoid cell lines and significant differences were found between lines of European ancestry compared to those of Asian descent [156]. With regards to human disease, associations of disease state and gene expression [157], but also associations of genetic traits and disease-relevant gene expression [152] in peripheral blood and other tissues have been found.

There have been relatively few human studies on the variation in transcription of HIF-regulated genes. Studies in patients with known disease-causing mutations of the HIF system have shown increased levels of VEGF and Aldolase C in peripheral blood lymphocytes from patients with Chuvash polycythaemia as opposed to healthy controls [76] and increased expression levels of HIF-target genes in peripheral blood mononuclear cells in patients with HIF2α-gain of function mutations [82]. Another study has shown that hypoxic induction of VEGF in peripheral blood monocytes from patients with coronary artery disease correlates with the development of coronary artery collateral circulation [158]. Indeed, a crucial consideration when investigating functional variation in the HIF pathway in relation to gene expression is looking for variation in inducibility of HIF-target genes by hypoxia between groups of interest, as opposed to just variation in the pattern of basal HIF-target gene expression. A key study using peripheral blood lymphocytes has shown a significant inter-individual difference in the degree to which expression of HIF-

target genes was induced by graded hypoxia. It showed that if one gene was strongly induced by hypoxia in one individual then this tended to be the case for other HIF-target genes within that individual [159]. A study has looked at HIF1α mRNA, inducible nitric oxide synthase (iNOS) and VEGF in cultured umbilical venous endothelial cells from native Tibetans and immigrant Han Chinese and found no significant difference [160].

In this study, I set out to investigate whether differences can be found at an intermediate cellular level between Tibetan and Han Chinese, using a peripheral blood lymphocyte model as in [159]. The following questions were addressed: 1. Do the two ethnic groups exhibit differences in basal levels of HIF2α mRNA and EGLN1 mRNA in their PBL, given the evidence implicating these two genes in the genetic adaptation of Tibetans and 2. Is there a difference in HIF-regulated gene expression between Tibetan and Han Chinese? In other words, do Tibetans show a different degree of hypoxic induction of HIF-target genes in their cells?

4.2 METHODS

4.2.1 Isolation of peripheral blood lymphocytes (PBL) and incubation under hypoxia

Seven Tibetan and seven Han Chinese volunteers participated in this part of the study. Forty ml of blood was collected from each volunteer and centrifuged at 400×g for 10 min at room temperature (RT). The supernatant (plasma) was discarded. The remaining blood was pooled and diluted with an equal amount of RPMI 1640 medium (*Sigma Aldrich*). Eight ml of diluted blood was carefully layered on top of 4 ml of a density-gradient medium Ficoll-Paque plus (*Amersham Biosciences ,Little Chalfont, UK*) in 15 ml tubes and centrifuged for 30 min at 400×g at RT with the brake off. This resulted in 3 layers as shown in Figure 17. The intermediate layer (buffy coat) contained peripheral blood

mononuclear cells (PBMC). This was carefully recovered in 15 ml tubes and washed twice in 10 ml of RPMI 1640. The final PBMC pellets were re-suspended in 40 ml of culture medium (RPMI 1640 supplemented with 10% FBS, 1% L-glutamine 200 mM and 1% penicillin/streptomycin, *Sigma Aldrich*). The suspension was incubated at 37 °C for 45 min over a hydrophilic membrane (162 cm^2 tissue-culture flasks, *Corning Costar*) to induce cellular adhesion by the monocytes, leaving a supernatant containing the PBL. The PBL were aliquoted into gas-permeable 50 mm lumox tissue-culture dishes (*Sarsted Ltd*) and incubated at eight different oxygen tensions (20%, 10%, 5%, 2%, 1%, 0.5%, 0.2% and 0.1% O_2) for 20 hours. To achieve this, each culture dish was placed into a separate glass chamber (~ 750 ml capacity), which had gas inlet and outlet fittings. Each glass chamber was flushed with a gas mixture of pre-determined oxygen concentration (*BOC special gases, BOC Industrial Gases, UK*), for at least 10 min, until the oxygen concentration in the chamber – measured using a commercially available O_2 analyser (*Servomex*) – equilibrated with the oxygen concentration of the gas mixture. The gas was passed through a 0.45 μm bacterial filter (*Millex, Millipore, UK*), attached in series with the container inlet. Each gas mixture contained the desired oxygen concentration, 5% CO_2 and balanced with N_2. The glass containers were then sealed and placed in a 37 °C incubator for 20 hours. The oxygen concentration in each glass chamber did not significantly change in these 20 hours. Following this, the cell suspension from each dish was retrieved and centrifuged at 400×g for 5 min at 4 °C, to obtain cell pellets.

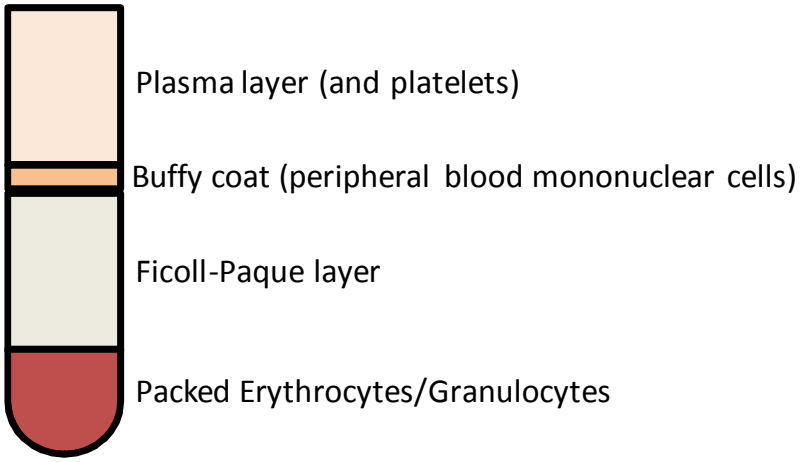


Figure 17: Resulting layers after centrifugation of blood over a density gradient medium

4.2.2 Total RNA isolation from PBL

One ml of Tri-Reagent (*Sigma Aldrich*) was added to each cell pellet to lyse the cells, followed by addition of 200 µl of chloroform (*Sigma Aldrich*) and thorough mixing by vigorous shaking. After allowing the mixture to stand at RT for 3 min, each sample was centrifuged for 30 min at 14000 rpm at 4 °C. The aqueous phase (containing the RNA) was recovered and 1 µl of glycogen was added (to assist nucleic acid precipitation). 500 µl of isopropanol (*Sigma Aldrich)* was added and mixed by inversion (15 times) and allowed to stand for 10 min at RT. This was stored at -80 °C overnight. Then, each isopropanol mixture was centrifuged at 14000 rpm at 4 °C for 30 min to precipitate the RNA into a pellet. The supernatant was carefully removed and each RNA pellet was washed with 1 ml of 75% ethanol. Following centrifugation (14000 rpm at 4 °C for 30 min) the ethanol was removed and the RNA pellet air-dried for 10 min. Each pellet was then re-suspended in 20 µl of nuclease-free water. Each sample was heated to 65 °C for 5 min and then homogenized. The quantity and quality of RNA in each sample was assessed using a Nanodrop spectrophotometer. Each sample was then diluted with nuclease-free water to achieve a concentration 50 ng/µl.

4.2.3 Reverse transcription - cDNA synthesis

Reactions were set up using a High Capacity cDNA Reverse Transcription kit (*Applied Biosystems, Warrington, UK*), as in Table 24 using the thermal cycling conditions shown in Table 25. The resulting cDNA was diluted 4-fold before proceeding to qPCR.

Table 24: Reverse transcription reaction – cDNA synthesis

Component	**Volume (µl)**
RT Buffer (10x)	2.0
dNTP Mix (100mM) (25x)	0.8
Random primers	2.0
Multiscribe Reverse Transcriptase	1.0
RNase Inhibitor	1.0
Nuclease-free water	3.2
RNA (total 500 ng)	10.0
Total	20.0

Table 25: Thermal cycling conditions during reverse transcription of RNA

Steps	**Temperature (°C)**	**Time (min)**
1	25	10
2	37	120
3	85	5
4	4	∞

4.2.4 Creation of the calibrator sample

PBL from the buffy coat of 500 ml of a donor's blood (*National Blood Bank*) were obtained, using the experimental procedures described in 4.2.1. Aliquots of PBL were incubated for 20 hrs under different oxygen concentrations (5%, 10% and 20% O_2). RNA was harvested, pooled together and used to synthesise cDNA, using the methods described in 4.2.2 and 4.2.3.

4.2.5 Measurement of mRNA transcript expression with quantitative (real time) PCR

TaqMan gene expression assays *(Applied Biosystems, Warrington, UK)* were used. In the TaqMan assay, a fluorescent TaqMan probe is designed to complementarily bind the cDNA template at a position between a forward and a reverse primer (see schematic in Figure 18). HIF2α mRNA, EGLN1 mRNA and 4 HIF-target gene transcripts – Vascular Endothelial Growth Factor A (VEGFA), Aldolase C (ALDC), Adrenomedullin (ADM) and Prolyl-4-Hydroxylase-Alpha-1 (P4HA1) – which were chosen based on a microarray analysis on oxygen-regulated gene expression in PBL [159] –, were assayed. B2 microglobulin (B2M) – which is not affected by hypoxia in PBL – was used as a reference gene (endogenous control). Details of the TaqMan assays used are shown in the Appendix. 10 µl PCR reactions were set up in triplicate on a 96-well plate, as shown in Table 26. The reference and target gene probes were multiplexed in the same well, as the reference gene is more abundant than the target gene. On every plate there was always a cDNA calibrator sample in triplicate. A StepOnePlus ™ thermal cycler (*Applied Biosystems, UK)* was used, with the conditions shown in Table 27.

A fixed fluorescence threshold was set above the baseline in the linear part of the curve and the cycle threshold (CT) was determined as the cycle number at which fluorescence crosses the threshold. CT values were recorded for each well for the reference gene and the target gene. Average values for the triplicate wells were taken. ΔΔCT values were then calculated for each sample using $\Delta\Delta CT = (CT_{\text{target gene}} - CT_{\text{reference gene}})_{\text{experimental sample}} - (CT_{\text{target gene}} - CT_{\text{reference gene}})_{\text{calibrator sample}}$. Expression (relative to the calibrator sample) was calculated using $(1+E)^{-\Delta\Delta CT}$, assuming that E (the amplification efficiency) is the same for the target gene and the reference gene (this was verified using a dilution series). E is taken as 1 [161].

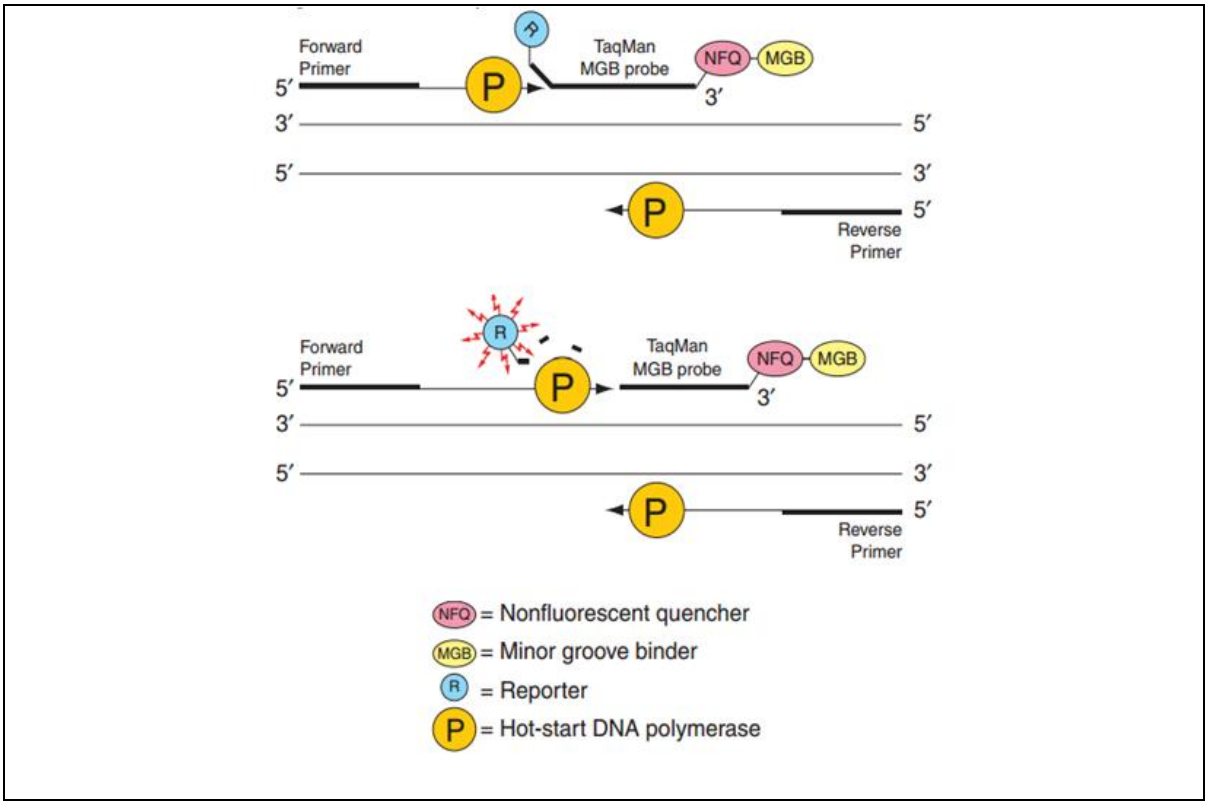


Figure 18: TaqMan qPCR chemistry. From Applied Biosystems online protocol "TaqMan Gene Expression Assays Protocol" [162]. Tha TaqMan probe binds between the forward and reverse primer sites. During PCR, the DNA polymerase cleaves only probes that are bound to the target. This cleavage seperates the reporter dye (R) from the quencher dye (NFQ), which results in increased fluorescence by the reporter.

Table 26: qPCR reactions using TaqMan gene expression assays

Component	**Volume (µl)**
Fast Universal PCR Master Mix (2×) *includes DNA polymerase*	5.0
TaqMan target gene expression assay mix (20×) *includes forward & reverse primers and Taqman Probe (FAM, MGB)*	0.5
TaqMan endogenous control gene mix (20×) *includes forward & reverse primers and Taqman Probe (VIC, TAMRA)*	0.5
Nuclease-free water	2.0
cDNA template	2.0
Total	10.0

Table 27: Thermal cycling conditions used in the qPCR reactions

Steps	Temperature (°C)	Time	
Hold	95	10 min	Polymerase activation
Cycles	95	15 sec	Denature
×40	60	1 min	Anneal and extend

4.3 RESULTS

4.3.1 Expression of HIF2α and EGLN1 mRNA in peripheral blood lymphocytes

<u>Main Finding:</u> The HIF2α mRNA but not EGLN1 mRNA levels are significantly lower in peripheral blood lymphocytes from Tibetan than from Han Chinese volunteers.

Figure 19 shows the baseline expression of HIF2α and EGLN1 mRNA in peripheral blood lymphocytes in Tibetan and Han Chinese volunteers. The expression of HIF2α mRNA is significantly lower in Tibetan than in Han Chinese volunteers ($p < 0.05$, Student's unpaired t-test, *SPSS 19.0*), while the expression of EGLN1 is similar in the two ethnic groups.

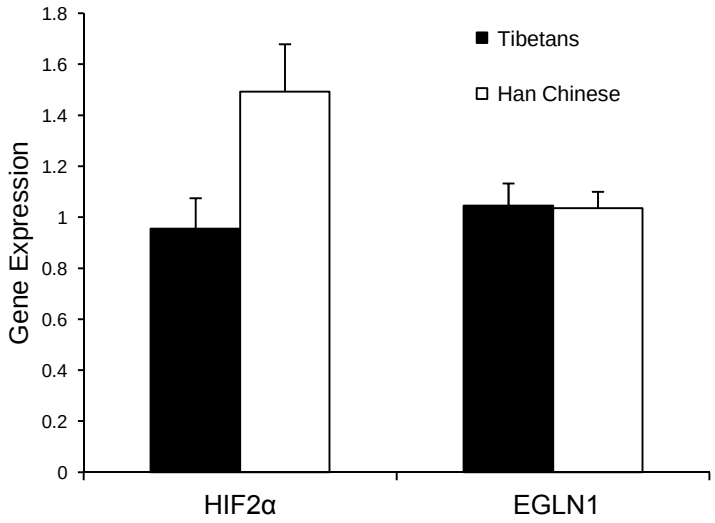


Figure 19: Baseline expression of HIF2α and EGLN1 mRNA (relative to a standard calibrator sample) in Tibetan and Han Chinese volunteers. Filled bars represent results from 7 Tibetan and open bars from 7 Han Chinese volunteers. Values are means and error bars indicate S.E.M.

4.3.2 Induction of HIF-target genes by hypoxia in peripheral blood lymphocytes

<u>Main finding:</u> Both the relative expression and the hypoxic induction of HIF-target genes are significantly lower in peripheral blood lymphocytes from Tibetan than from Han Chinese volunteers.

ΔΔCT appears to be a cubic function of log (oxygen concentration) as shown for transcript Aldolase C, as an example, in Figure 20 and this is true for each of the four gene transcripts. Figure 21 shows the expression levels of 4 HIF-target gene transcripts at 8 different oxygen concentrations for 7 Tibetan and 7 Han Chinese volunteers.

For the statistical analysis examining the influence of ethnicity on the induction of HIF-target genes by graded hypoxia in PBL, the following linear mixed-effects model was used:

$$y_i = \tilde{\alpha} + \beta x_i + \gamma z_i + \delta z_i^3 + \zeta(x \times z)_i + \sum_{k=2}^{4} \tilde{\eta}_k w_{k_i} + \sum_{k=2}^{4} \theta_k (z \times w_k)_i + \varepsilon_i$$

for the i^{th} observation, where x_i represents ethnicity and is a binary fixed effect, z_i is a continuous variable represent log(oxygen concentration) and w_{k_i} is a categorical fixed effect, representing gene; k = 2 refers to gene ADM, k = 3 refers to gene ALDC, k = 4 refers to gene P4HA1 with the reference gene being VEGF. $\varepsilon_i \sim N(0, \sigma^2)$ is the inter-observation error. $\tilde{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ and $\tilde{\eta}_k \sim N(\eta_k, \sigma_{\eta_k}^2)$ are random variables reflecting inter-volunteer variability for different genes. In this analysis, the dependent variable y_i represents ΔΔCT, which is taken as the measure of gene expression. The results – displayed in Table 28 – show that log(oxygen concentration) significantly affects ΔΔCT ($p < 0.001$), as expected for hypoxically-regulated HIF-target genes; both baseline gene

expression ($p < 0.05$) and the degree of induction of genes by hypoxia is significantly affected by ethnicity ($p < 0.05$).

Table 28: Statistical analysis of HIF-regulated gene expression

	Value	Std. Error	Df	t-value	p-value
Intercept	1.6169	0.1883	424	8.5878	0.0000
Logoxygen	2.3552	0.0642	424	36.6784	0.0000
I(logoxygen^3)	-0.3508	0.0421	424	-8.3244	0.0000
Ethnicity	-0.5099	0.1787	12	-2.8533	0.0145
$Gene_2$	-0.8244	0.0763	424	-10.8059	0.0000
$Gene_3$	-1.6914	0.1317	424	-13.3814	0.0000
$Gene_4$	-1.5581	0.1164	424	-13.3814	0.0000
Logoxygen × Ethnicity	0.0859	0.0406	424	2.1153	0.0350
Logoxygen × $Gene_2$	-0.7297	0.0575	424	-12.6993	0.0000
Logoxygen × $Gene_3$	-0.3720	0.05748	424	-6.4815	0.0000
Logoxygen × $Gene_4$	-0.01230	0.0574	424	-0.2264	0.8210

$\sigma = 0.327$

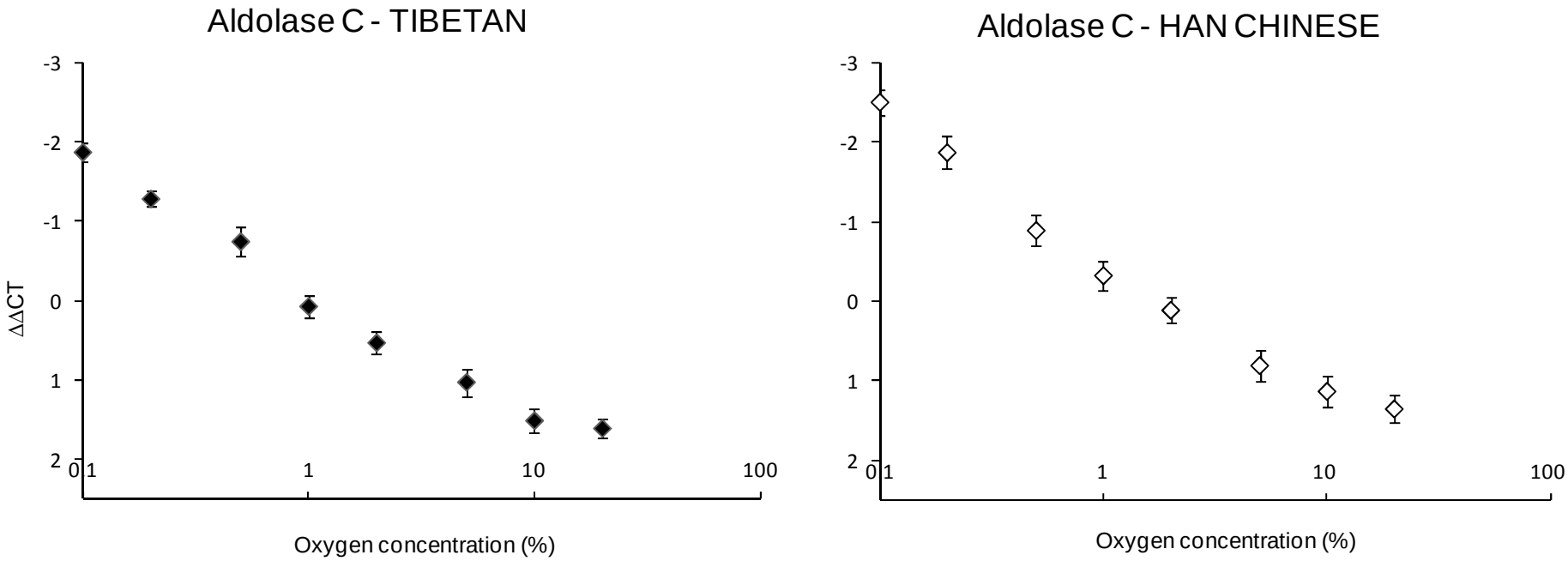


Figure 20: ΔΔCT (for Aldolase C) against oxygen concentration (%) on a logarithmic scale (x-axis) for Tibetan and Han Chinese volunteers. Values are means and error bars indicate S.E.M.

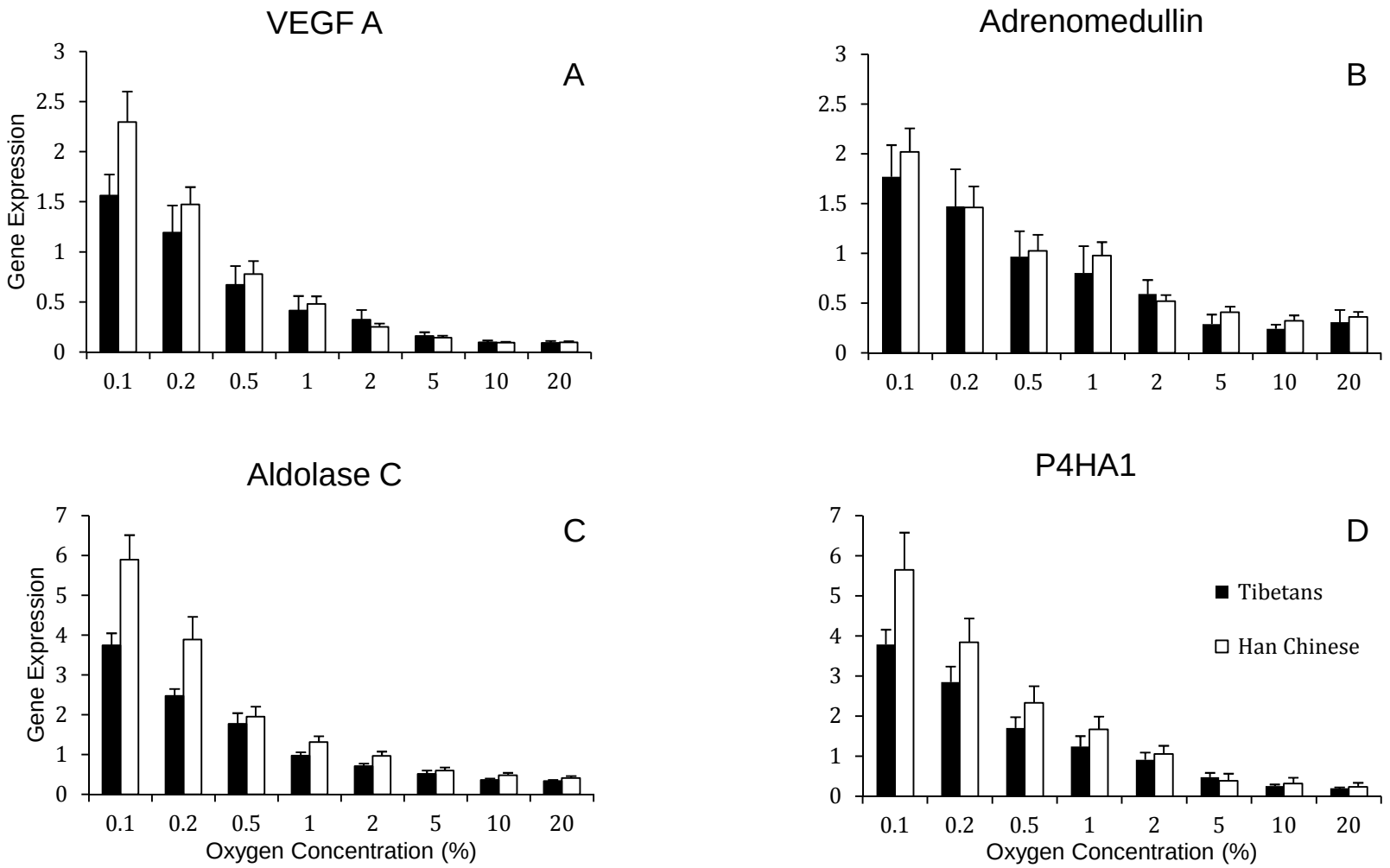


Figure 21: Expression of HIF-target genes (relative to a standard calibrator sample) at graded oxygen tension for Tibetan and Han Chinese volunteers. (A) Vascular Endothelial Growth Factor A (VEGFA), (B) Adrenomedullin, (C) Aldolase C and (D) Prolyl-4-Hydroxylase-alpha-1. Filled bars represent results from 7 Tibetan and open bars from 7 Han Chinese volunteers. Values are means and error bars indicate S.E.M.

4.4 DISCUSSION

Genome-wide studies comparing Tibetan and Han Chinese genomes have found evidence of positive selection at the *EPAS1* locus in Tibetans and have shown evidence of markedly different allelic and haplotype differences in the two ethnic groups at *EPAS1*. What is interesting is that all the genetic variants described in Tibetans are non-coding and are found either within introns of *EPAS1* or downstream of *EPAS1*; while this may often be related to the design of the genomic studies – e.g. genome-wide allelic differentiation scans, genome-wide association studies or candidate gene studies – whole exome sequencing [117] or targeted sequencing of the full-length *EPAS1* gene [116] failed to reveal any coding genetic variants in Tibetans. Thus it is likely that "functional" genetic variation in Tibetans might lie in regulatory regions and actually affect transcription of the *EPAS1* gene itself, as opposed to downstream HIF2α structure and function.

In keeping with the above possibility, in my study, I found lower HIF2α mRNA expression in peripheral blood lymphocytes from Tibetan than from Han Chinese volunteers. This suggests a lower level of transcription of *EPAS1* – or alternatively reduced mRNA stability – and may indeed be related to underlying genetic variation between Tibetans and Han Chinese. It is interesting that the "sign" of the difference between Tibetans and Han Chinese is in the direction that one might hypothesize, i.e. lower in Tibetans, given the observed phenotypes in Tibetans e.g. lower Hb and smaller physiological responses to hypoxia.

Lower levels of transcription of *EPAS1* might be one way whereby regulatory genetic variation exerts its transcriptional effects. Another possibility is that it may be associated

with alternative splicing in producing either novel splice variants or splice variants of different abundance. Such phenomena may be both tissue-specific and context-specific.

In contrast to EPAS1, EGLN1 was not different in expression at the mRNA level between Tibetans and Han Chinese. Although this finding is limited to PBL and may be different in other cell types, it could suggest that any difference between Tibetan and Han Chinese arises from differences at the post-transcriptional level affecting protein stability or function. In support of this, two coding variants have been described of particularly high frequencies in Tibetans – as discussed in Chapter 3 – although the functional significance of these is not yet known. This is addressed in Chapter 7.

Perhaps the most interesting finding in this study is that both the relative expression and the hypoxic induction of HIF-regulated genes were significantly lower in PBL from Tibetan volunteers. This gives evidence of variation in the HIF transcription activation pathway, between Tibetan and Han Chinese at the cellular level. This is consistent with a reduced sensitivity of the HIF response in Tibetans and parallels the observations of reduced hypoxic sensitivities in Tibetans at the integrative physiology level.

Of course, there may be reasons other than genetic variation – which is one hypothesis – as to why this difference in the hypoxic sensitivity of the HIF system at the cellular level is observed between the two ethnic groups. For example, differences in epigenetic phenomena may be producing the observed variation. Epigenetics or epigenomics may refer to a wide range of phenomena related to cellular state including DNA and histone modifications which can have significant effects on gene transcription. A few things are of relevance. One is that while epigenetic phenomena may be indeed driven by underlying

DNA sequence variation either acting in *cis* or in *trans*, they can also be driven by environmental factors determined by a cell's or an individual's physical, biochemical, physiological or behavioural context. Of interest are studies of monozygotic twins in whom variation in DNA methylation has been largely attributed to environmental changes during an individual's life time [163] and studies of monozygotic and dizygotic twins in whom it was attributed to early developmental changes *in utero* [164]. Second, epigenetic phenomena – irrespective of origin – can be inherited and passed from generation to generation [165]. Third, there is increasing evidence that epigenetics play a crucial role in the cellular response to hypoxia, with hypoxia not only having an impact on global patterns of histone modification and methylation but with epigenetics also having an impact in the regulation of HIF transcriptional activity [166].

Chapter 5: HIF-regulated gene expression in EBV-transformed lymphoblastoid cell lines from Tibetan and Han Chinese

Epstein-Barr virus immortalised lymphoblastoid cell lines have proved a useful resource for genetic analysis of gene expression, as discussed in the introduction of the previous chapter. The rationale for using them in this study to look for potential differences in Tibetan and Han Chinese is that: (a) they provide an ongoing source of a single subpopulation cell type in which growth, culture and experimental conditions can be standardised to minimise environmental variation; and (b) they allow the harvesting of large number of cells for analysis and give the opportunity of greater and more precise experimental replication. Thus, the aim was to investigate if differences can be observed in transformed cell lines of Tibetan versus Han Chinese origin as was seen in peripheral blood lymphocytes.

5.1 METHODS

5.1.1 Acquisition of cell lines

Ten ml of blood was obtained from each of 11 Tibetan volunteers. This was sent to *ECACC* Genetic Support Services within 24 hours of collection. Peripheral blood lymphocytes were isolated from the blood and underwent immortalisation by EBV infection, according to the protocols of the company. This process was complete within 6 weeks and the EBV-transformed B lymphoblastoid cell lines were stored in frozen aliquots in a liquid nitrogen tank upon receipt from ECACC until use. In addition, 5 EBV-transformed lymphoblastoid cell lines, each from a Han Chinese individual (*HapMap*) were purchased from the *Coriell Cell Institute,* USA.

5.1.2 Resuscitation of frozen cell lines

The frozen ampoule of cells was thawed quickly (1-2 min) at 37 °C in a water bath and then added to 10 ml of pre-warmed culture medium (RPMI 1640 supplemented with 10% FBS, 1% L-glutamine and 1% penicillin/streptomycin, *Sigma Aldrich*). To eliminate the freezing medium, the cell suspension was centrifuged at 200×g for 5 min at RT. The cell pellet was re-suspended in 7 ml of fresh medium (supplemented with 20% FBS) overnight. The cell suspension was cultured with medium supplemented with 20% FBS for the first few days (~ 4) and thereafter with medium supplemented with 10% FBS.

5.1.3 Subculture of cell lines

EBV-transformed B lymphocytes grow in suspension (and in large clumps) and were cultured in a 75 cm^2 tissue tissue-culture flask (*Sarsted Ltd, UK*) with a vented cap in an upright position at 37 °C in a humidified 5% CO_2 standard tissue-culture incubator. Fresh medium was added to the cell suspension 3 times a week and a cell count was obtained approximately once a week. The optimal density for cell growth is 200,000 to 1,000,000 cells/ml (linear phase of growth) and thus this cell density range was maintained throughout. The cell lines were initially expanded and a proportion of them cryo-preserved and stored in frozen aliquots in a liquid nitrogen tank, while the rest were used for experiments.

5.1.4 Cell quantification

The cell suspension was pipetted gently to break clumps. 20 µl of cell suspension was removed and mixed with 20 µl of Trypan blue (*Sigma Aldrich)* i.e. diluted by a factor of 2. A haemocytometer was used to count the viable number of cells.

5.1.5 Cryopreservation of cell lines

Ten ml of cell suspension (density of 5×10^5 cells/ml) was centrifuged at 100×g at 4°C for 3 min. The supernatant was discarded and the pellet re-suspended in 1 ml of cold freezing medium (RPMI 1640, 20% FBS and 6% DMSO) (*Sigma Aldrich)* to yield 5×10^6/ml. The ampoule was frozen at a rate of -1°C/min down to -80°C in an isopropanol bath placed in -80°C freezer overnight and then stored in a liquid nitrogen tank.

5.1.6 Experimental protocol

The expression levels of 5 HIF-regulated gene transcripts – VEGFA, ADM, ALDC, P4HA1 and N-myc downstream regulated 1 (NDRG1) – and of EGLN1 and HIF2α were compared between 5 Tibetan and 5 Han Chinese cell lines. Cell suspensions from each cell line were cultured in gas-permeable 50 mm lumox dishes at a density of 4 $x10^5$ cells/ml (5 ml per dish) at 4 different oxygen concentrations – 21%, 5%, 1% and 0.1% O_2 – for 24 hours. These hypoxic incubations were performed in parallel using 3 hypoxia work stations (*Invivo2 400, Ruskin Technology*) set at 5%, 1% and 0.1% O_2 and a tissue-culture incubator (21% O_2). Following the 24-hr incubation, the cell suspension from each dish was retrieved and centrifuged at 400×g for 5 min at 4 °C, to obtain cell pellets. Three independent experimental repeats (on 3 separate days) were undertaken for each cell line. Total RNA was isolated from each cell pellet and reverse transcriptase quantitative PCR (RTqPCR) was used to determine gene expression as described in section 4.2. A calibrator sample was used as reference (as was done in the experiments with peripheral blood lymphocytes), and this was created from RNA pooled from a mixture of Tibetan and Han Chinese cells cultured at various oxygen concentrations.

5.2 RESULTS

5.2.1 Expression of HIF-target genes

<u>Main findings:</u> ***Ethnicity (Tibetan or Han Chinese) does not significantly affect expression of HIF-target genes or their induction by hypoxia in these cell lines. Inter-cell line variability significantly determines the hypoxic induction of HIF-target genes.***

Figure 22 shows the mean expression of each of the HIF-target genes investigated in Tibetan and Han Chinese cell lines. The results suggest that ethnicity (i.e. whether the cell line comes from a Tibetan or a Han Chinese individual) does not affect expression of HIF-target genes in these cell lines, nor does it affect induction of HIF-target genes by hypoxia.

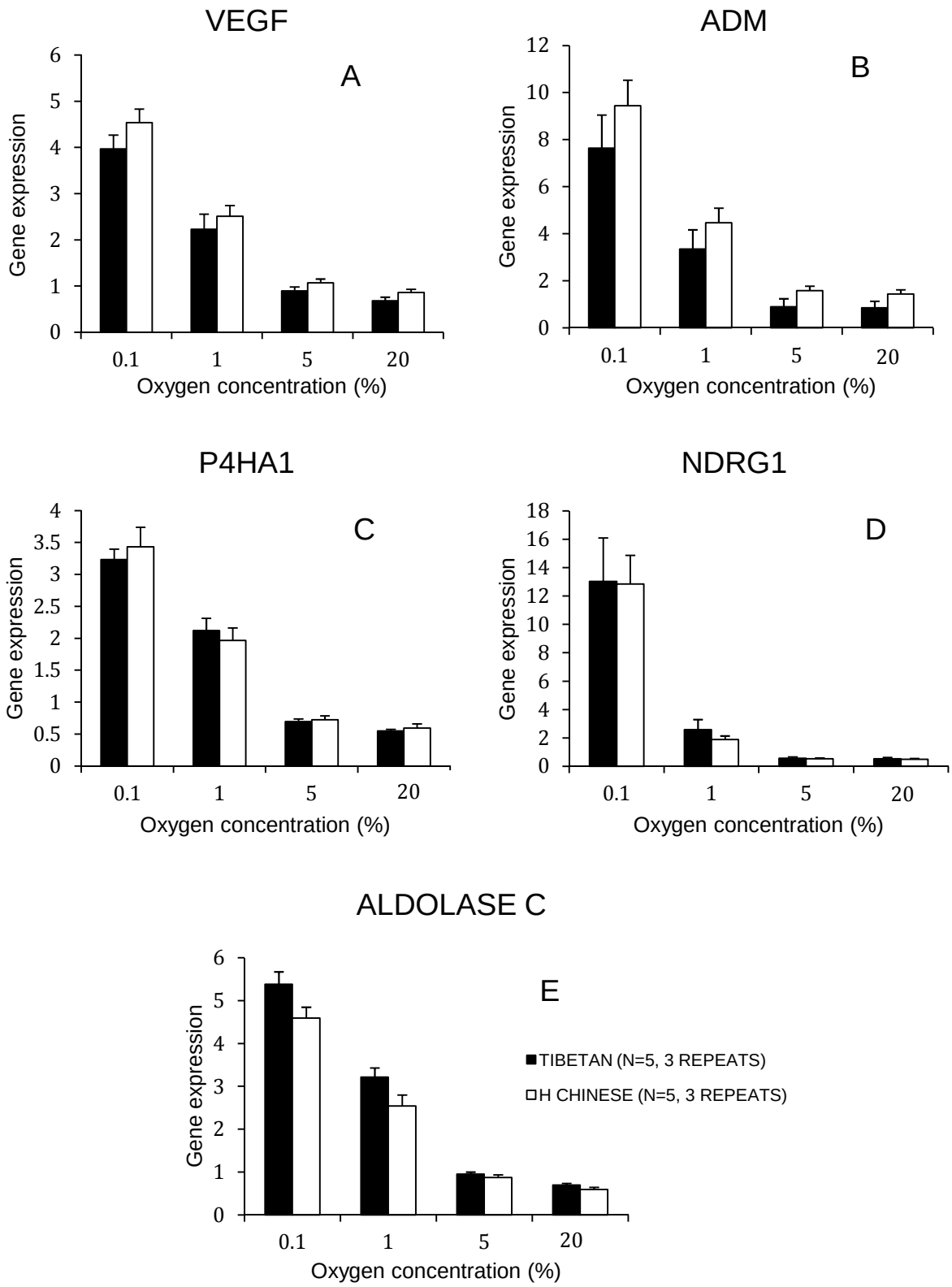


Figure 22: Expression of HIF-target genes in EBV-transformed lymphoblastoid cell lines (relative to a standard calibrator sample) at 4 different oxygen tensions (20%, 5%, 1%, 0.1% O_2). (A) VEGF; (B) ADM; (C) P4HA1; (D) NDRG1; and (E) Aldolase C. Filled bars represent results from 3 experimental repeats from 5 Tibetan and open bars from 3 experimental repeats from 5 Han Chinese cell lines. Values are means and error bars indicate S.E.M.

The statistical analysis was carried out by fitting a linear mixed-effects model to the data (*R 2.13.0*). The analysis was carried out for each gene separately, first, using the model:

$$y_i = \tilde{\alpha} + \beta x_i + \gamma z_i + \zeta(x \times z)_i + \varepsilon_i$$

and also for all genes considered together using the model:

$$y_i = \tilde{\alpha} + \beta x_i + \gamma z_i + \zeta(x \times z)_i + \sum_{k=2}^{5} \eta_k w_{k_i} + \sum_{k=2}^{5} \theta_k (z \times w_k)_i + \varepsilon_i$$

for the i^{th} observation, where y_i represents $\Delta\Delta CT$, x_i represents ethnicity and is a binary fixed effect, z_i is a continuous variable represent log (oxygen concentration) and w_{k_i} is a categorical fixed effect, representing gene; k = 2 refers to gene ADM, k = 3 refers to gene ALDC, k = 4 refers to gene P4HA1 and k = 5 refers to gene NRDG1 with the reference gene being VEGF. $\varepsilon_i \sim N(0, \sigma^2)$ is the inter-observation error. $\tilde{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ is a random variable reflecting variability between experimental repeats and cell lines. Note that *cell line* is treated as a random effect and *experimental repeat* is nested within *cell line.* It was found that ethnicity did not significantly affect gene expression nor did it affect the effect of oxygen tension on gene expression, ie. the induction of genes by hypoxia.

Figure 23 shows, as an example, the expression of ADM in 3 separate cell lines. The y-axis represents $\Delta\Delta CT$ and the x-axis represents oxygen on a logarithmic scale. If one parameterises the response using a gradient and an intercept, it appears that the gradient of the response is similar within a cell line and less similar between cell lines. To test this, the gradient and intercept for each experimental repeat for each gene were calculated. Ignoring ethnicity as a factor, as I have already shown that it does not play a significant role, I used ANOVA (*SPSS 19*) and tested if "cell line" is a significant factor affecting either the "gradient" or the "intercept" of the response.

In the analysis for “gradient”, *gradient* was the dependent variable, *cell line* was a fixed factor and *gene* was a random factor. The term *gene* was significant ($p < 0.001$), suggesting that each gene is induced differently by hypoxia; the term *cell line* was significant ($p < 0.05$), suggesting that each cell line behaves differently in hypoxia and the interactive term *cell line* × *gene* was significant ($p < 0.001$), suggesting that within a gene cell lines differ in their sensitivity to hypoxia. The results are shown in Table 29.

In the analysis for “intercept”, *intercept* was the dependent variable, *cell line* was a fixed factor and *gene* was a random factor. The term *gene* was not significant. The term *cell line* was not significant suggesting that different cell lines do not differ in their general baseline expression of genes, but the interactive term *cell line* × *gene* was significant ($p < 0.001$), suggesting the expression differs between cell lines for different genes – see Table 30.

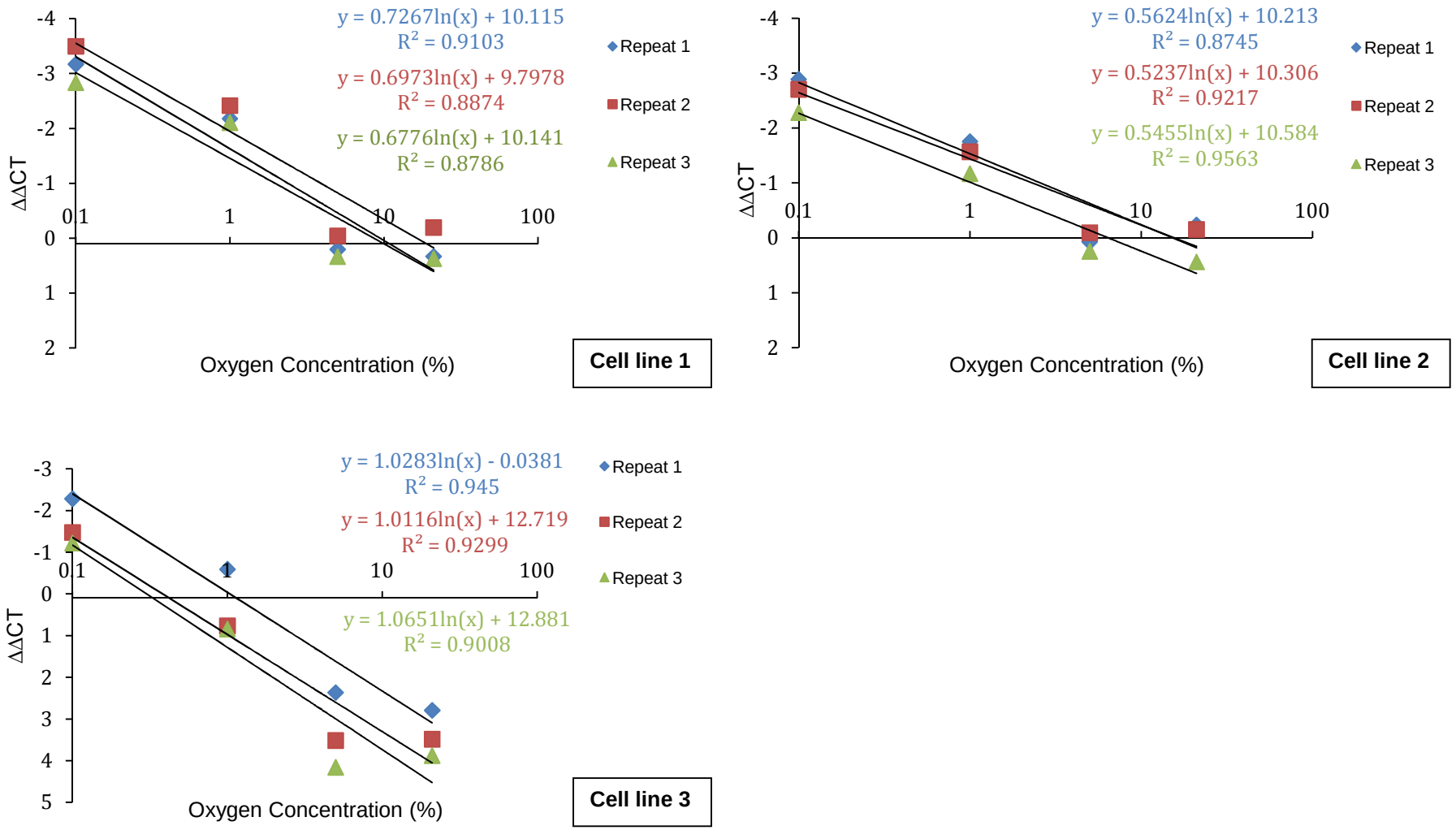


Figure 23: ΔΔCT against oxygen concentration (%) on a logarithmic scale (x-axis) for each experimental repeat for three different cell lines for ADM, as examples. The equation of the line of best fit is shown on the graphs. A gradient and intercept (at oxygen concentration of 10%) is derived from this equation.

Table 29: Results of ANOVA for the "gradient" of response

Dependent Variable: *Gradient*

Source		**Type III Sum of Squares**	**Df**	**Mean Square**	**F**	**P-value**
Intercept	Hypothesis	57.478	1	57.478	88.818	0.001
	Error	2.589	4.000	0.647		
Cell line	Hypothesis	0.581	9	0.065	2.326	0.035
	Error	0.999	36.010	0.028		
Gene	Hypothesis	2.589	4	0.647	23.531	0.000
	Error	0.993	36.080	0.028		
Cell line × *Gene*	Hypothesis	1.000	36	0.028	9.438	0.000
	Error	0.282	96	0.003		

Table 30 : Results of ANOVA for the "intercept" of response

Dependent Variable: *Intercept*

Source		**Type III Sum of Squares**	**Df**	**Mean Square**	**F**	**P-value**
Intercept	Hypothesis	34.027	1	34.027	11.215	0.029
	Error	12.136	4.000	3.034		
Cell line	Hypothesis	22.668	9	2.519	1.285	0.279
	Error	70.569	36.006	1.960		
Gene	Hypothesis	12.139	4	3.035	1.562	0.205
	Error	70.044	36.045	1.943		
Cell line × *Gene*	Hypothesis	70.644	36	1.962	16.865	0.000
	Error	11.170	96	.116		

5.2.2 Expression of HIF2α and EGLN1 mRNA in EBV-transformed lymphoblastoid cell lines

<u>Main Finding:</u> There is no significant difference in the baseline expression of EGLN1 and HIF2α mRNA in Tibetan versus Han Chinese cell lines.

Results are shown in Figure 24. A student's t-test (*SPSS 19*) was used to analyze the results.

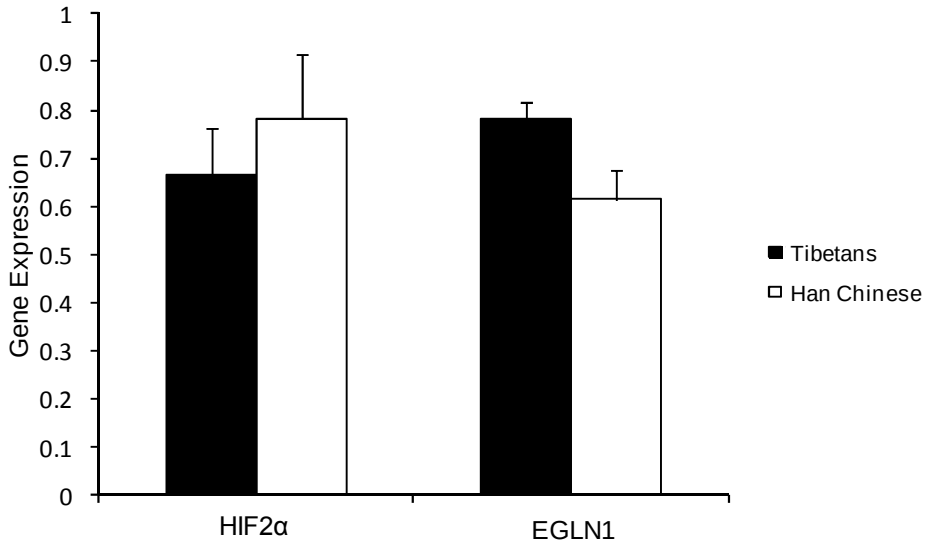


Figure 24: Baseline expression of HIF2α and EGLN1 mRNA in EBV-transformed lymphoblastoid cell lines (relative to a standard calibrator sample). Filled bars represent results from 3 experimental repeats from 5 Tibetan and open bars from 3 experimental repeats from 5 Han Chinese cell lines. Values are means and error bars indicate S.E.M.

5.2.3 Relative expression of HIF2α and HIF1α in EBV-transformed lymphoblastoid cell lines

Figure 25 shows the expression of HIF2α and HIF1α mRNA at different oxygen concentrations, in Tibetan and Han Chinese cell lines. There is a trend in this figure that as oxygen concentration decreases HIF1α mRNA decreases whereas HIF2α mRNA is relatively unchanged but increases at 0.1% O_2. Thus, in Figure 26, the relative change (from baseline euoxic value) in HIF2α and HIF1α mRNA levels was plotted against oxygen concentration, for Tibetan and Han Chinese. There is evidence of a reciprocal relationship, such that when HIF1α mRNA levels increase, HIF2α mRNA levels decrease and *vice versa*.

Statistical analysis using repeated measures ANOVA (*SPSS 19*) was used, separately for each of HIF1α and HIF2α, where ΔΔCT was the dependent variable, *oxygen* was a "within-subject" factor and *ethnicity* was a "between-subject" factor. For HIF1α mRNA it was shown that HIF1α mRNA level changes (decreases) as oxygen concentration decreases (term *oxygen* in the analysis is significant, $p < 0.001$) but there is no difference between Tibetan and Han Chinese cell lines in how this occurs (interactive term *oxygen* × *ethnicity* is not significant). Also, no difference was found in the overall level of HIF1α between cell lines of different ethnic origin. For HIF2α, it was shown that HIF2α mRNA level changes with oxygen concentration and that Tibetan and Han Chinese cell lines differ in the way this occurs; however the overall level of HIF2α mRNA is not significantly different between cell lines of different ethnic origin.

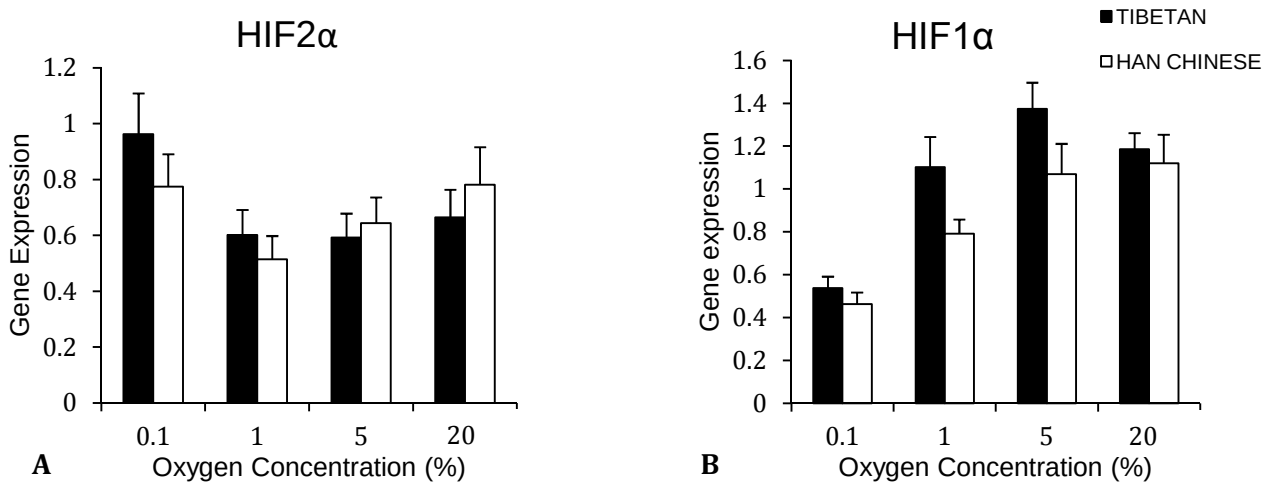


Figure 25: Expression of (A) HIF2α and (B) HIF1α in EBV-transformed lymphoblastoid cell lines (relative to a standard calibrator sample) at 4 different oxygen concentrations (20%, 5%, 1%, 0.1% O_2). Filled bars represent results from 3 experimental repeats from 5 Tibetan and open bars results from 3 experimental repeats from 5 Han Chinese cell lines. Values are means and error bars indicate S.E.M.

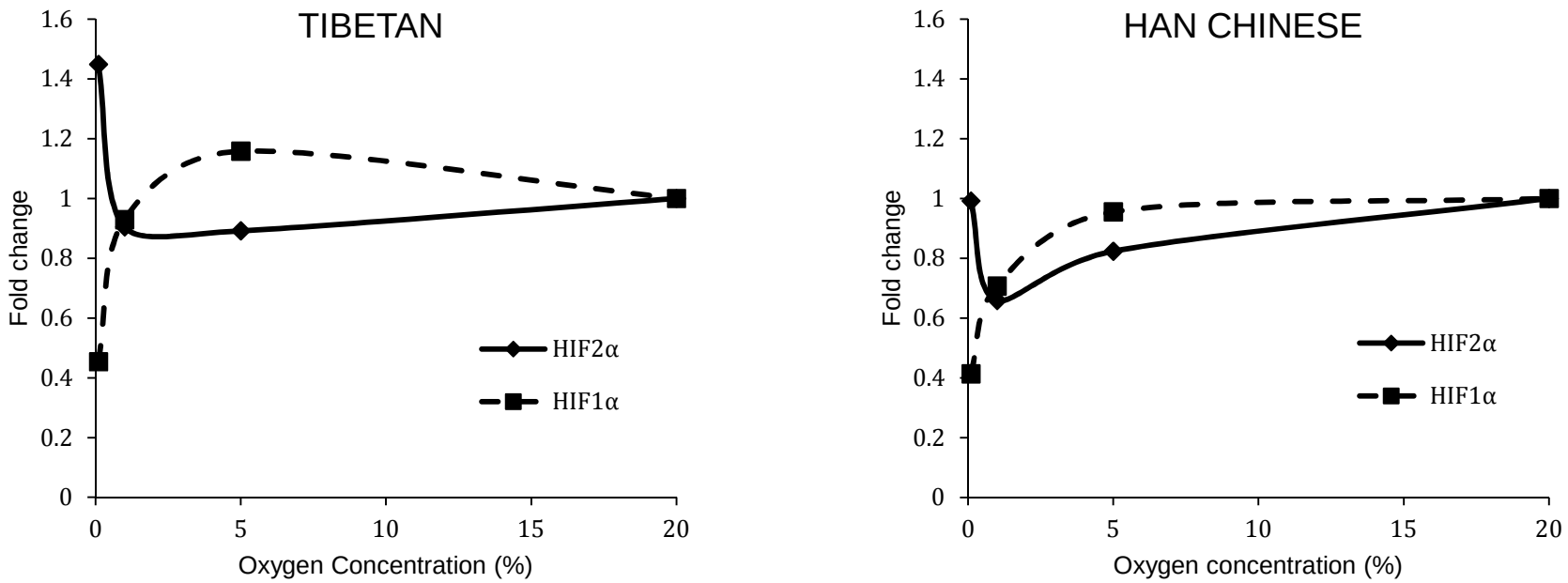


Figure 26: Relative change from baseline euoxic value in HIF2α and HIF1α mRNA against oxygen concentration, in EBV-transformed lymphoblastoid cell lines. (A) Tibetan (5 cell lines) and (B) Han Chinese (5 cell lines). Data are from 3 experimental repeats for each cell line.

5.3 DISCUSSION

I would like to focus the discussion on two key points.

The first is that the inter-cell line variability significantly determines the hypoxic induction of HIF-target genes, and that within genes there is significant inter-cell line variability in basal expression of each gene. This is an example reminding one that indeed, use of EBV-transformed lymphocytes is not without its caveats. These cell lines are immortalised with Epstein-Barr virus which itself can modulate expression of genes and may show heterogeneity between individual cell lines with viral load and other factors. Another problem is that prolonged cell culture and passage may lead to the establishment of clonality and differential DNA methylation [167] which may significantly affect gene expression. While it is possible that the inter-cell line variability observed may relate to actual inter-individual variability in the volunteers that the cell line was obtained from, there is evidence that the transformation process itself can have significant effects on gene expression in the cell line; indeed experiments have shown that the degree of variation in gene expression between cell lines originating from the same individual but produced by transformation separately (on different occasions) was greater than the variation across different individuals (DPhil thesis of Jerome TS Brooks 2006 [168]). This may be one of the reasons for the discrepancy between the findings here and the findings in peripheral blood lymphocytes. Thus, while it may well be that transformed lymphobastoid cell lines are useful in high-throughput large-scale analysis of gene expression and genetic fine mapping, primary cells may be more suitable in looking for differences between human individuals or groups in cellular gene expression, as they better represent natural gene expression *in vivo* and are closer to the “true physiological” state.

The second point relates to the changes in HIF1α and HIF2α mRNA as oxygen decreases. It is evident here that in these cell lines cultured for 24 hours, HIF1α mRNA decreases as oxygen concentration decreases. Like most other biological systems the HIF system is subject to a variety of negative feedback loops; most of the described negative feedback mechanisms relate to post-translational effects and indeed HIF1α protein levels are attenuated in prolonged hypoxia [70]. At the mRNA level, a mechanism of negative feedback on HIF1α mRNA has been proposed whereby hypoxia induces a naturally-occurring antisense transcript of HIF1α (aHIF) which might affect the stability and lead to degradation of HIF1α mRNA [169]. Indeed antisense HIF1α mRNA was found to be over-expressed in clear cell renal cancers and in EBV-transformed lymphocytes it was induced by graded hypoxia while HIF1α was concomitantly decreased [170]. Although the duration of hypoxia used in that study was significantly lower, i.e. 2 hours, as opposed to 24 hours in my study, the range of oxygen tensions used was similar, and it is indeed possible that the HIF1α mRNA down-regulation I observed might be the result of the same mechanism.

The observation that is most interesting here is that as HIF1α mRNA decreases, HIF2α mRNA increases and *vice versa* (Figure 26). This observation of the reciprocal effect of hypoxia on HIF1α and HIF2α mRNA is in fact significantly tighter in the Tibetan versus the Han Chinese cell lines. Differential regulation of HIF1α and HIF2α by hypoxia has been the focus of many studies, with most of them examining this at the post-translational level and indeed finding different temporal patterns of HIF1α and HIF2α expression with prolonged hypoxia. Surprisingly, relatively little is known about the transcriptional regulation of HIF1α and HIF2α, or indeed regulation at the mRNA level. A study investigating the differential regulation of HIF1α and HIF2α in lung epithelial cells (A549 cell line) also demonstrated amongst other findings that HIF1α mRNA decreases with time

in hypoxia (0.5% oxygen, time course up to 12 hours) while HIF2α mRNA increases [70] and discussed the notion that trans-regulation between HIF1α and HIF2α exists.

In any case, the findings of my study in which there is a reciprocal effect in graded prolonged hypoxia such that HIF2α mRNA increases when HIF1α mRNA decreases and vice versa, raises the question of whether the balance or the ratio between the two isoforms might be important physiologically; preservation of such balance or a given ratio might be important in cells, and indeed deviation from a given ratio might be more critical in producing effects rather than absolute levels of HIF1α or HIF2α in cells or tissues. An example of where this antiphase regulation and balance of the two isoforms – i.e. HIF1α versus HIF2α – is physiologically relevant is demonstrated in a study in macrophages where it acts to, respectively, increase or suppress nitric oxide synthesis [66].

Chapter 6: Is the genetic variation at the *EPAS1* locus found in Tibetans functional at the cellular/molecular level?

6.1 INTRODUCTION

One important question to address is whether, and if so, how the genetic variation at the *EPAS1* locus that is selected for in Tibetans – and which correlates with phenotypic differences at the systems level – is functional at the cellular or molecular level. This is a difficult question to address primarily because the causal variant in *EPAS1* is unknown. Furthermore, there are regions of high linkage disequilibrium in *EPAS1*, which makes it even harder to identify the casual variant within them. The hypothesis I decided to test was whether genetic variation in *EPAS1* acts in *cis* to affect or regulate transcription of *EPAS1* itself. The reasons for addressing this hypothesis are: 1) The genome-wide significant SNPs identified in Tibetans in all the published studies are non-coding (intronic or downstream of the gene); 2) All the *EPAS1* Hb-associating SNPs identified in these studies are intronic; 3) Sequencing of the *EPAS1* gene in Tibetans did not reveal any coding variation of potential functional significance [116, 117]; and 4) There is evidence from the peripheral blood lymphocyte study in Chapter 4 that HIF2α mRNA is lower in expression in Tibetan than in Han Chinese primary cells.

I could only obtain peripheral blood lymphocytes from 7 Tibetan volunteers, and thus addressing this question using these primary cells would be difficult due to the small sample size. Similarly, an approach whereby the level of HIF2α and HIF-target gene expression is compared between major homozygous, heterozygous and minor homozygous EBV-transformed cell lines (of Tibetan origin) would also likely to be ineffective in view of the significant inter-cell line variability between EBV-transformed cell lines.

Thus, I decided to test the hypothesis by exploring whether there is allele-specific expression of *EPAS1* in Tibetan cells that are heterozygous for the genetic variant of interest, as defined in Chapter 3. By comparing relative allelic expression within a cellular sample, much of the potential heterogeneity associated with the comparison between multiple samples is avoided. Hence, this approach provides good internal experimental control. Below I explain the methodology and rationale used to investigate this.

6.2 METHODS

To investigate allele-specific expression, three separate approaches were taken, described in sections 6.2.2, 6.2.3 and 6.2.4, respectively. The first two employed methodologies which involved assaying relative allelic expression in both genomic DNA and cDNA samples; this required the identification of a transcribed SNP to serve as a marker for the genetic variant of interest. For this, the coding region of *EPAS1* was sequenced, as described in section 6.2.1.

6.2.1 Sequencing of the *EPAS1* coding region

As all the known genetic variants in *EPAS1* in Tibetans were non-coding, the coding region of *EPAS1* in 13 Tibetan DNA samples was sequenced in search for coding/transcribed SNPs that could be used as markers for allele-specific expression of *EPAS1*. Essentially, the aim was to identify heterozygous (for the "high altitude adaptation" selected genetic variant) Tibetan samples that were also heterozygous for a coding or transcribed marker SNP. One would also need to identify Tibetan samples that were homozygous for the "high-altitude adaptation" selected haplotype but heterozygous for the coding/transcribed marker SNP, so that they can be used as negative controls in the experiments.

The complete DNA sequence of human *EPAS1* is available on on-line databases. Primer pairs were designed to PCR-amplify regions of approximately 300 bp (Table 31). 25 µl PCR reactions were set up for each set of primers, according to the protocol described in section 3.1.4, with adjustments in the melting temperature and the $MgCl_2$ content, as required. The quality of the PCR products was checked by agarose gel (2%) electrophoresis, as previously described. The product of the PCR reactions was purified by treatment with Illustra Exostar (*GE Healthcare*) which contains alkaline phosphatase and exonuclease 1. Briefly, 5 µl of the completed PCR reaction was mixed with 2 µl Illustra Exostar; this was incubated at 37 °C for 15 min and then at 80 °C for 15 min.

Each purified PCR product was diluted 4-fold before proceeding to the sequencing reaction. Big Dye Terminator (BDT) (*Invitrogen*) sequencing reactions were set up, as shown in Table 32 and processed on a thermal cycler using the conditions shown in Table 33. The products of the sequencing reaction were purified by a plate and spin column purification method using the DyeEx Kit (*Qiagen*) and were submitted to the Department of Zoology, University of Oxford, for processing onto a sequencer.

Table 31: Primer pairs used for Sanger sequencing of the *EPAS1* coding region

Coding Region	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)
5′ UTR	CCGACGGAGTTTTTAAAGTGG	GTGGTGAAGGTCCGAGGTG
5′ UTR	AGGTTCCTCCCAGTCACCTT	TTCGGACTCGTTTTCAGAGC
Exon 1	CACCTCGGACCTTCACCAC	CTCCGAAGACCTCTCAGCAC
Exon 2	ATGTCCTGGCCAGAGGTATG	GGGTCCCCTGCATATAGGA
Exon 3	GTGTCCTCTGTGCAGTGGTG	GTGGCTAGCACCTTCCACTC
Exon 4	AAATCTGGAAGGTGGCTCAG	TCCATTTTTCCTGCATCACA
Exon 5	GGGCCCCTCTCATGAATATC	GTCAGACATGGAGCCGAGAG
Exon 6	GTGCAGGGGATGCCTAAG	TCTCTGCCCACAGCATACAC
Exon 7	GGGTTAGCTCCTGCCCACAT	CACCAGTCACACCCCACTC
Exon 8	ATCTGCTGAGCCTGTGGTG	CCTAGGCCTGTCATCCTGAG
Exon 9	GCCTGGAGTCCTACCCATTT	GCCAGGGAGGCCTGTTATAG
Exon 10	GCCTTCCAGACCCTCTTAGG	AATGTCCCTTCAGCCATCTG
Exon 11	GACTTCTGGGGAGACCAGTG	AATGGATGTGATCCTTTGCT
Exon 12	TGCAGGAGCTGAGTTGGAAT	ATGGAAGAGAGAGGGGTGCT
Exon 12	TGGAGACACTGGCACCCTAT	CTGGGGGTCCAGCTATCTTA
Exon 13	AAGGGCCCCTAAGATGAGAA	CCCTAAGGGCTATGGAGAGG
Exon 14	GTCTGGGGACCTGGTTCTCT	CCCAGTCCTTCCACCAGAAC
Exon 15	GCCAATGCTACCGTCACTCT	AAAAAGCAGACGTGGGCTTCC
Exon 16	TCCCCCAGTCACAAAGAAGT	ACGGAGAGAGTGAAGCTGGT
3′ UTR	CTGCCCGAACTGACCAGATA	CAGAAGAACAGACATGCACCA
3′ UTR	TTTTCCTCCCCACCTTCAAT	TTTAGGAAAAGCCACGCTGT
3′ UTR	AAATGTGAAGGGTCAACTCC	GACTCCACTGCTCGGATTGT
3′ UTR	CCTGTACGGACACTGTGGAA	AGCCTCCTCCTCAGTCCAG
3′ UTR	GGGAATTCATGCAAATGAGG	TAGTTGGCCCAATGTGCTTT
3′ UTR	TTCTGTAATGGGTACAATGAAGAAG	CGTTACGTTGACAGGTAGGG
3′ UTR	GGGTTACTGACGTGTAAATGCT	CCATGATAAATCCACAGG

Table 32: BDT sequencing reaction

Component	Volume (µl)
BDT mix	8.0
Primer (forward or reverse) (10 µM)	1.0
DNA template (PCR product)	4.0
Nuclease-free water	7.0
Total	**20.0**

Table 33: Thermal cycling conditions for BDT sequencing reaction

Steps	Temperature (°C)	Time
1	96	5 min
2: Cycles (x35)	96 50 60	30 sec 45 sec 4 min
3	4	∞

6.2.2 Investigation of allele-specific expression using a TaqMan SNP genotyping assay

By sequencing the *EPAS1* coding region described in 6.2.1, four Tibetan samples were found to be heterozygous for the coding SNP rs59901247 (2806A>C) (see the results section 6.3 that follows). A TaqMan SNP genotyping assay (*Applied Biosystems)* was designed for rs59901247. This consisted of two primers (forward and reverse) amplifying the region of interest (containing the SNP) and two TaqMan probes, for distinguishing between the two alleles, one with a VIC reporter dye at the 5′ end (specific for allele "A") and one with a FAM reporter dye linked to the 5′ end (specific for allele "C"). Each probe had a minor groove binder (MGB) and a non-fluorescent quencher (NFQ) at the 3′ end. The principle of the assay is shown in Figure 27. During PCR, each TaqMan probe anneals specifically to its complementary sequence between the forward and reverse primer sites. The DNA polymerase extends the primers bound to the template DNA and cleaves probes that are hybridized to the target. This cleavage separates the reporter dye from the quencher dye, which results in increased fluorescence by the reporter. A mismatch between a probe and the target sequence (even a single nucleotide mismatch) reduces the efficiency of probe hybridization and the amount of reporter dye cleavage. In addition, the polymerase displaces a mismatched probe without cleaving it. The results are read using the sequence detection system (SDS) software (*Applied Biosystems*) that provides

fluorescence (Rn) values based on the signal from each well. The relative fluorescence signals of the two dyes indicate which alleles are present in each sample.

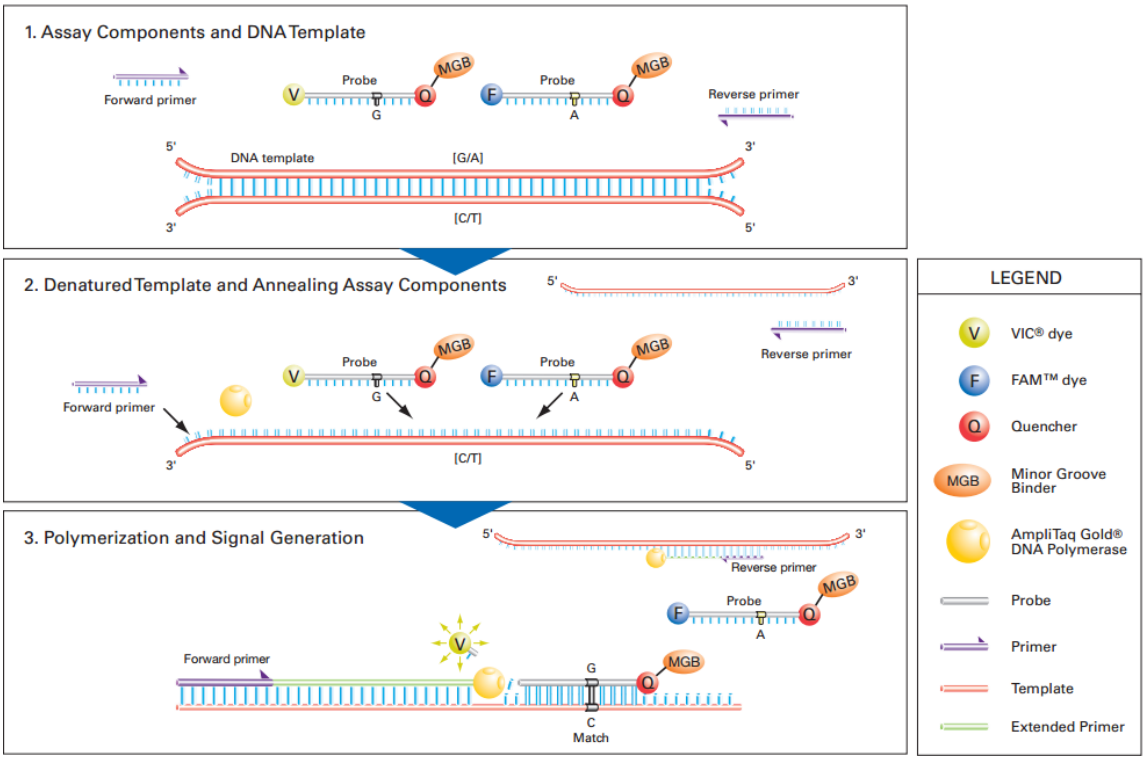


Figure 27: TaqMan SNP genotyping technology, illustrating the principles described in section 6.2.2 above. From Applied Biosystems online protocol "TaqMan SNP Genotyping Assays Protocol" [171].

The PCR reactions were performed in triplicate on a 96-well plate, as shown in Table 34 and were ran (on a real-time PCR system) using the thermal cycling conditions mentioned in Table 11. These experiments were performed in collaboration with Dr Chiara Bardella (post-doctoral research scientist).

Table 34: PCR reactions used using the TaqMan SNP genotyping assay

Component	Volume (µl)
TaqMan Universal PCR Master Mix, No AmpErase UNG (2x) (contains the AmpliTaq DNA Polymerase)	5.0
Primer and Taqman probe dye mix (20x) (contains probes for both alleles,VIC and FAM)	0.5
genomic DNA or cDNA (~20 ng in total)	1.0
Nuclease-free water	3.5
Total	**10.0**

6.2.2.1 Cloning of DNA fragments harbouring either allele "C" or allele "A"

In order to interpret the assay described in Section 6.2.2, samples with all three genotypes AA, AC, and CC should be available for a TaqMan SNP genotyping experimental run. However, the homozygous state CC (for rs59901247) was not found in any of the genomic DNA samples screened. For this reason this was cloned using the pCR8/GW/TOPO TA Cloning Kit (*Invitrogen*), by Dr Chiara Bardella, as described below. Genomic DNA from a heterozygous sample (AC) was used as a template for a PCR reaction using the forward primer (5′ to 3′): GCGATGTCCTCTGTCTCTCC and the reverse primer (5′ to 3′): ACACCCACCTGACACCTTGT. The (linearized) vector in the kit has 3′-T overhangs for direct ligation of the Taq-amplified PCR products (which have A′ at their 3′ end). Thus, after ligation each vector contains an insert with either one allele (A) or the other allele (C). Following bacterial transformation and culture, 16 colonies were picked (each colony has either DNA with allele "A" or allele "C"); DNA was extracted from each (*Lyse and GO PCR reagent, Thermoscientific)* – by applying 10 µl of reagent at 95 °C for 2 min – and genotyped to obtain DNA which is homozygous AA and DNA which is homozygous CC.

6.2.3 Investigation of allele-specific expression using a primer extension assay followed by mass spectrometry

Genomic DNA and DNase-treated cDNA samples were prepared from Tibetan cell lines with informative genotypes, cultured in either euoxia or hypoxia.

The principle of the methodology is displayed in Figure 28 and described below.

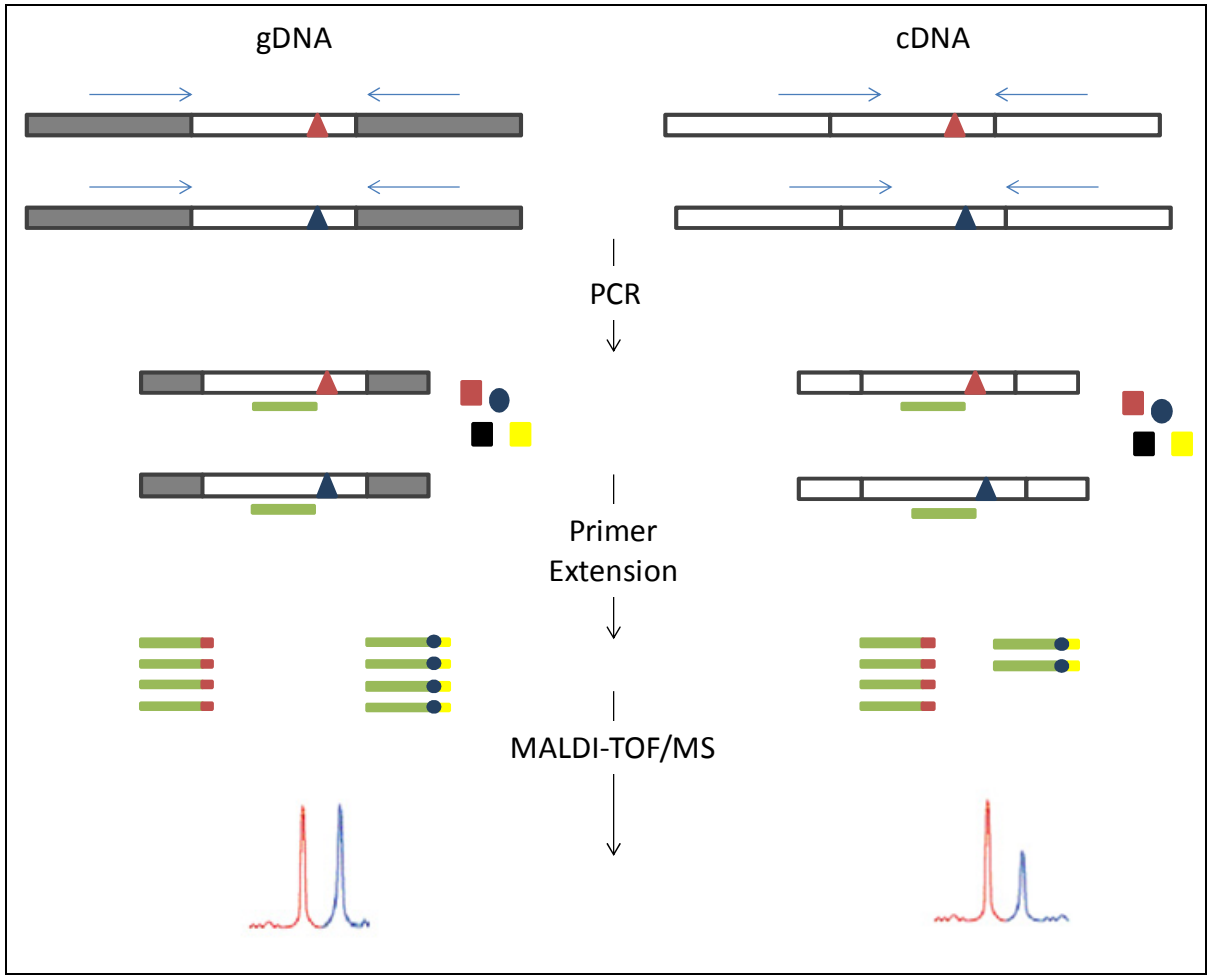


Figure 28: Principle of using the MassArray System (*Sequenom*) to quantify allele-specific expression. PCR primers (blue horizontal arrows) are designed to amplify a heterozygous coding polymorphism (triangle) on both gDNA and cDNA, as shown. The PCR product is then used as a template for a primer extension reaction carried out using an extension primer (green rectangle) annealing one base next to the polymorphism and an appropriate mixture of dideoxynucleotide (ddNTPs) terminators (squares) and one deoxynucleotide (dNTP) (circle). The resulting primer extension products have different lengths (and hence masses) which allows separation and quantification by MALDI-TOF mass spectrometry. The smaller peak of the "blue" allele relative to the "red" allele in the cDNA indicates relatively lower expression of the "blue" allele.

Two pairs of PCR primers – one pair for genomic DNA amplification and one pair for cDNA amplification of the region that the coding SNP rs59901247 lies – and one extension primer were designed using the Assay Design Software (*Sequenom*) and manufactured by *Metabion*. The extension primer was designed so that it is the same for genomic and cDNA PCR products and it anneals adjacent to the SNP of interest (one base away). They are shown in Table 35.

Table 35: PCR primers and extension primer

	Forward Primer (5' to 3')	**Reverse Primer (5' to 3')**	**Extension Primer**
gDNA	ACGTTGGATGCTCTCTGACTTTGGTCTTTC	ACGTTGGATGTGGCAGCGGCAGATGTCTCA	CCTCATGGGGTTTTGGG
cDNA	ACGTTGGATGGAGCGCAAATGTACCCAATG	ACGTTGGATGTGGCAGCGGCAGATGTCTCA	CCTCATGGGGTTTTGGG

Seven µl PCR reactions were set up in duplicate or triplicate for each sample (gDNA or cDNA). The PCR product was purified by incubating 5 µl of the completed PCR reaction with 2 µl Illustra Exostar (*GE Healthcare*) at 37 °C for 20 minutes. Figure 29 shows a picture of the gel electrophoresis of the PCR products for genomic DNA and cDNA samples that were produced and used for the subsequent primer extension reactions. Note that the two lanes on the right show the PCR products from cDNA samples that were not treated with DNase, suggesting that they are contaminated with genomic DNA (hence two PCR products, one at ~ 400 bp and one at ~ 100 bp). Only DNase-treated samples were used in the assay.

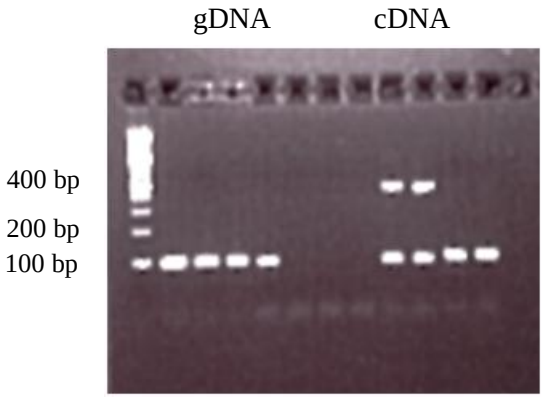


Figure 29: Gel showing bands from PCR reactions for 4 genomic DNA and 4 cDNA samples. Note the presence of 2 bands in two of the cDNA samples (at ~100 bp and ~ 400 bp) indicating that those two samples are contaminated with genomic DNA. The size of the correct PCR product is 100 bp.

The purified PCR products were then submitted to the Oxford Genomics Sequenom Service at the Wellcome Trust Centre for Human Genetics, where further processing of the samples using the MassARRAY system (*Sequenom*) took place. A primer extension reaction was set up for each "clean" PCR product and run using the MassARRAY system according to the manufacturer's instructions. A typical primer extension reaction includes a starting step at 94 °C for 2 min, followed by 40 cycles of 94 °C for 5 s, 52 °C for 5 s and 72 °C for 5 s and includes: the PCR product, the extension primer, the enzyme (Thermosequenase) and an appropriate mixture of ddNTP/dNTPs. The extended products were cleaned with SpectroCLEAN resin and then spotted in quadruplicate on a 384-well microarray Chip (SpectroCHIP). The extended oligonucleotides were quantified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a SpectroREADER mass spectrometer. MALDI-TOF MS can easily distinguish two oligonucleotides if they differ by more than 10 Da. Following mass spectrometry, an allelotyping report was generated that provided peak area data, i.e. area-under-peak for allele "A" and area-under-peak for allele "C", obtained from the MS profiles produced. The peak area (allelic) ratio was calculated. The cDNA allelic ratio was compared to the mean gDNA allelic ratio to assess deviation from the expected 1:1 ratio.

6.2.4 Investigation of allele-specific expression by assessing differential RNA Polymerase II loading at the *EPAS1* locus with TaqMan SNP genotyping PCR assay

RNA Polymerase II (Pol II) Chromatin Immunoprecipitation (ChIP) was performed in heterozygous – for the putative selected "Tibetan" haplotype – Tibetan EBV-transformed lymphocytes. The aim was to assess whether there is a difference in the amount of Pol II binding at the *EPAS1* locus at one or the other allele, which might in turn suggest that there is increased or decreased transcription of one allele compared the other. The protocol for ChIP is described in 6.2.4.1.

6.2.4.1 Chromatin Immunoprecipitation (ChIP) protocol

The principle of the ChIP method is shown in Figure 30.

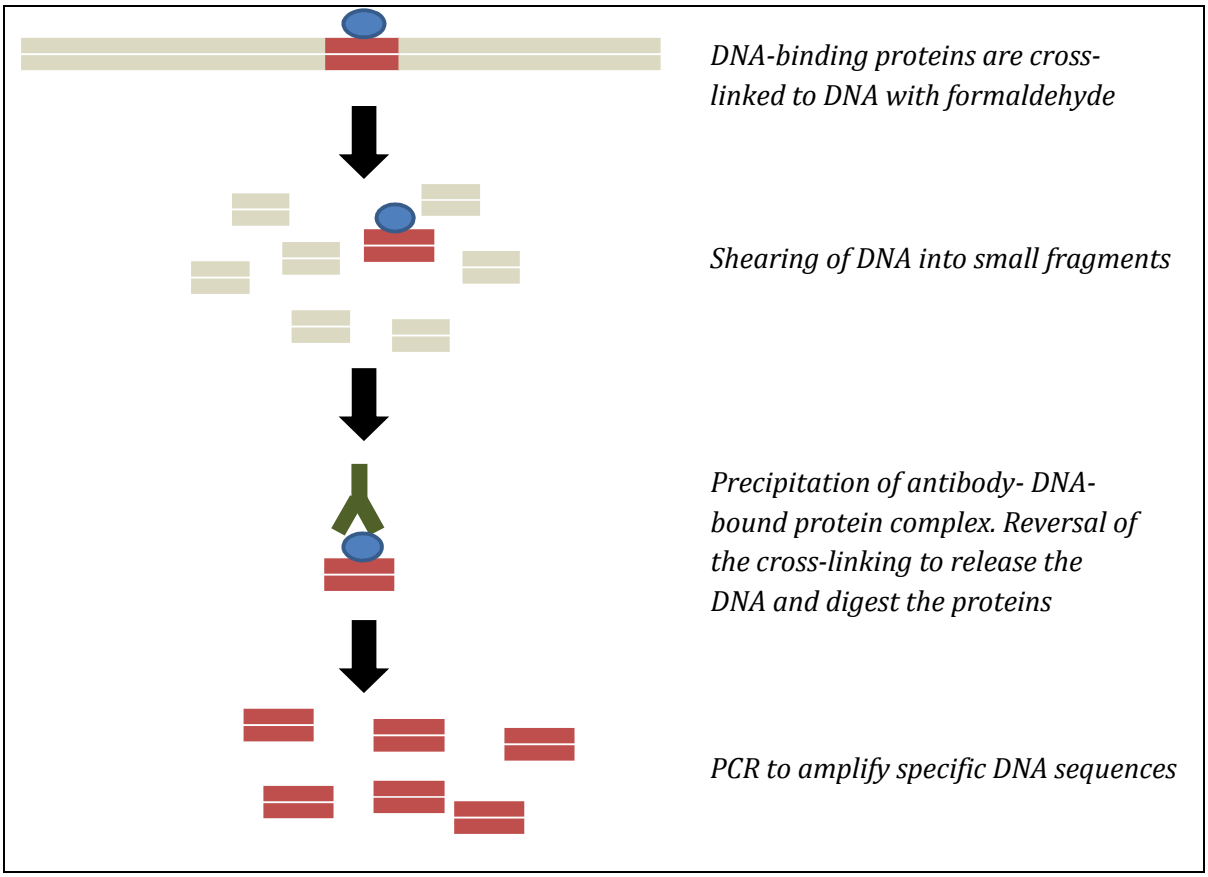


Figure 30: Basic steps of the Chromatin Immunoprecipitation procedure.

For each cell line, cells were cultured at a cell density of 5×10^5 cells/ml for 24 hours in euoxia (using a standard tissue culture incubator) and in hypoxia of 0.2% O_2 using a hypoxia workstation (*Invivo2 400, Ruskin technology*). The total number of cells used per experimental condition was 1×10^7.

Following incubation, chromatin was cross-linked by adding formaldehyde (*Sigma Aldrich)* to a final concentration of 1% for 12 min, at 4 °C; the fixation was stopped by adding glycine (*Sigma Aldrich*) to a final concentration of 125 mM for 5 min, at 4 °C. The cells were then washed twice with cold PBS[*1] and centrifuged at 500×g for 5 min. Each resulting cell pellet was lysed with 600 µl SDS Lysis buffer[*] and incubated on ice for 10 min. Each lysate was then sonicated 3 times at full power for 5 min in an ice waterbath to shear the DNA into small fragments. Following centrifugation at high speed for 10 min, at 4 °C, the supernatant was retrieved and diluted by adding 10 volumes of Dilution buffer[*].

The degree of DNA shearing was assessed by extracting DNA from 200 µl of diluted chromatin – after incubation with 8 µl of NaCl solution (5 M) at 65 °C for 3 hours – using phenol:chloroform:isoamyl alcohol (25:24:1, Tris-buffered) (*Sigma Aldrich*) and subjecting it to agarose gel (2%) electrophoresis.

The rest of the diluted chromatin was incubated with Protein A agarose beads (*Millipore*), which were added at a concentration of 10 µl/ml, for 1 hour at 4 °C. The beads were subsequently collected by centrifugation (3 min, 2000 rpm). 100 µl of the chromatin – the "input chromatin" – was removed and stored at -20 °C.

[1] the composition of all buffers marked with [*] are given in the Appendix

Two µl of chromatin was incubated with 2 µg of non-specific rabbit (Rb) IgG antibody (*Millipore*) and 2 µl of chromatin was incubated with 2 µg of RNA Pol II Antibody (*Santa Cruz)* at 4 °C overnight.

Then, fresh Protein A agarose beads were added to each sample (10 µl/ml) and incubated (on an end-on-end rotor) for 1 hour at 4 °C. The agarose beads were collected by centrifugation (500×g for 5 min at 4 °C), washed sequentially with 1 ml of the following buffers: Low Salt buffer*, High Salt buffer*, Lithium buffer* and TE buffer* (twice), and then collected with a final centrifugation step at 300×g for 5 min. The DNA was eluted by washing the beads twice with 250 µl fresh Elution buffer* for 15 min at RT and then collecting the eluate (total of 500 µl) by centrifugation at high speed for 3 min. "Input Chromatin" (100 µl) was diluted with 400 µl Elution buffer and was processed together with the eluted samples as follows: 20 µl of NaCl solution (5 M) was added per sample and incubated for at least 4 hours at 65 °C, to reverse the chromatin cross-linking. 10 µl of EDTA (0.5 M) (*Sigma Aldrich*), 20 µl Tris-Hcl (1 M, pH 6.5) (*Sigma Aldrich)* and 2 µl of Proteinase K (20 ug/ml) (*Roche*) were added and the mixture was incubated for 1 hour at 45 °C. Four hundered µl of DNA was recovered by phenol:chloroform:isoamyl alcohol (25:24:1, Tris-buffered) extraction. One µl of glycogen (20 µg/µl) (*Sigma Alrich)*, 50 µl NaAc (3 M) and 1ml of 100% ethanol were mixed in. The samples were incubated on dry ice for 15 min and were subsequently centrifuged at full speed for 45 min at 4 °C. The resulting DNA pellets were washed twice with 70% ethanol, dried and re-suspended in 30 µl nuclease-free water. The resulting DNA was diluted 1:4 for subsequent qPCR.

6.2.4.2 qPCR to assess "enrichment"

qPCR (*SyBR Green Technology, Applied Biosystems*) was used to assess "enrichment" at DNA loci of interest. The aim of this is to quantify the abundance of DNA fragments (containing the DNA sequence of interest) in the Pol II antibody-treated samples compared to the Rb Ig treated samples (control). If there is increased abundance, this means that there is "enrichment", i.e. RNA Polymerase II selectively associates with these DNA fragments. To ensure that indeed the chromatin immunoprecipitation with RNA Polymerase II was successful, PCR primers were designed at a "positive control" locus such as the gene β-actin, and at a "negative control" locus, a non-significant inter-genic region, as well as at the loci of interest. The inclusion of a "negative control" locus at which enrichment is also quantified, allows a further level of control in the experimental design, such that any potential technical differences in processing between Pol II Antibody-treated samples and Rb Ig- treated samples (control) can be seen and accounted for. The qPCR reaction was set up as shown in Table 36. Briefly, SyBR green binds to all newly synthesised double-stranded DNA and fluoresces. The fluorescence accumulates as cycling PCR continues and is measured at the end of each cycle, recording a CT value.

Table 36: qPCR reaction of the immunoprecipated DNA

Component	**Volume (µl)**
SyBR Green Mastermix (contains Taq Polymerase)	5.0
Primer mix (forward & reverse primers at 10 µM each)	0.5
(immunoprecipitated) DNA	2.0
Nuclease-free water	2.5
Total	10.0

6.2.4.3 Taqman SNP genotyping PCR assay

Once enrichment at the DNA locus of interest at *EPAS1* was confirmed, a TaqMan SNP assay was designed for a heterozygous SNP (rs1447563) at the locus of interest – this is explained in the results section – to assess allelic imbalance. The methodological steps of this assay have already been described in section 6.2.2.

6.3 RESULTS

6.3.1 Results from the Sanger sequencing of the *EPAS1* coding region

***Main finding:* In the samples sequenced, there is a high degree of conservation in the coding/transcribed region of EPAS1. Only 1 coding SNP was found.**

Coding SNP rs59901247 was identified in 4 Tibetan samples, as shown in Table 37. This was the only transcribed variant identified from sequencing the *EPAS1* coding region in the Tibetan DNA samples.

Table 37: Tibetan samples heterozygous for rs59901247

Tibetan volunteer/sample	Genotype for the *EPAS1* selected haplotype	Genotype for rs59901247
T #1 (1778)	Minor homozygous	AC
T #4 (1781)	Heterozygous	AC
T #5 (1782)	Minor homozygous	AC
T #13 (1825)	Minor homozygous	AC

In addition, this coding SNP was identified in 2 samples from Han Chinese volunteers and in a number of cell lines (by TaqMan genotyping) as shown in Table 38.

Table 38: Other DNA samples heterozygous for rs59901247

Sample	Cell line	rs59901247
1819	Han Chinese primary lymphocytes	AC
1808	Han Chinese primary lymphocytes	AC
HEC1A	human endometrial cell cancer	AC
RCCL15	renal cell carcinoma	AC
HEB3B	hepatocellular carcinoma	AC

This coding SNP rs59901247 (2806A>C) is found in exon 15 and causes a threonine to proline substitution at codon 766. There is no evidence of functional significance in this aminoacid substitution in the existing literature. The overall expected average heterozygosity for this SNP is 0.274. Population allelic frequencies (as reported in dbSNP) are shown in Table 39.

Table 39: Population allelic frequencies for rs59901247 as reported in dbSNP

Population	Frequency of allele "A"	Frequency of allele "C"
Han Chinese/Japanese (1000 Genome project) 120 chromosomes	0.825	0.175
YRI (African); 118 chromosomes	0.424	0.576
ESP Cohort Population (Europeans)	0.837	0.163

Cells from the heterozygous Tibetan T#4 (1781) were used to look for allele-specific expression of HIF2α. Cells from minor homozygous Tibetans T#5 and T#13 (1782, 1825) were used as negative controls.

6.3.2 Results from the TaqMan allele-specific expression PCR assay

<u>Main finding</u>: Using this assay there is evidence of allelic imbalance in the cDNA samples tested but this is not related to the putative "high-altitude" selected genetic variant in Tibetans.

Figure 31 shows the results of a TaqMan allele-specific assay assessing allele-specific expression in cDNA from Tibetan primary lymphocytes and Figure 32 shows the results of the same assay using cDNA from EBV-transformed lymphoblastoid cells from the same Tibetan volunteers. The fluorescent signals from cDNA from both the primary cells and the lymphoblastoid cells form a "cluster" which is shifted towards the minor allele "C" in comparison with the cluster formed by fluorescent signals from heterozygous gDNA. This shift suggests allelic imbalance in cDNA and an increased expression of HIF2α from allele "C". However, it is evident that this allelic shift is present in all Tibetan cDNA samples, including the negative controls.

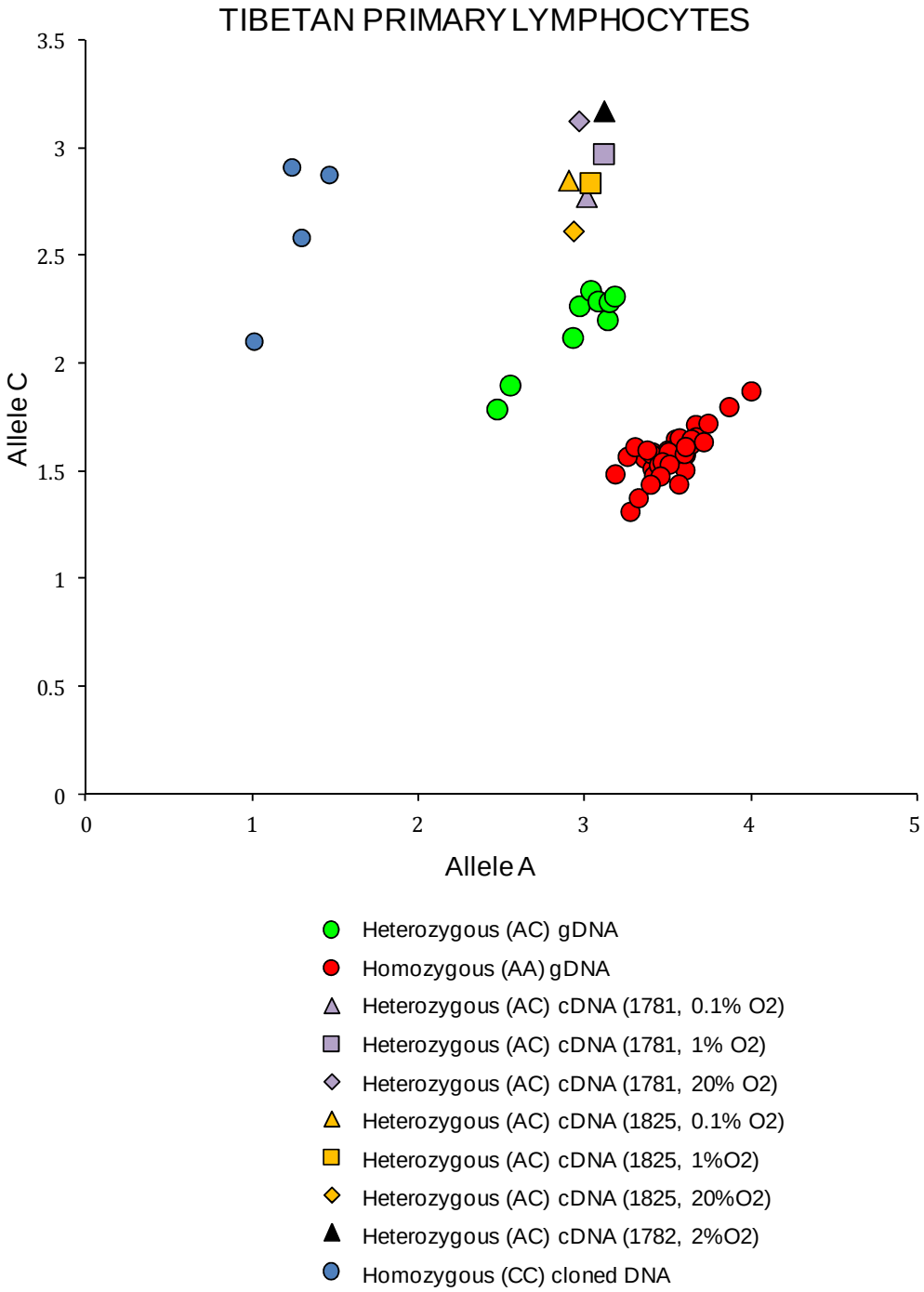


Figure 31: TaqMan genotype-specific PCR assay assessing allele-specific expression in Tibetan primary lymphocytes. The y-axis represents fluorescent intensity (normalised to background) for allele C (FAM dye) and the x-axis fluorescent intensity (normalised to background) for allele A (VIC dye). Homozygous AA, heterozygous AC and homozygous CC refer to the genotypes with respect to the marker coding SNP rs59901247. Sample 1781 (T#4) is also heterozygous for the putatively selected Tibetan *EPAS1* haplotype. Samples 1782 (T#5) and 1825 (T#13) are minor homozygous for the *EPAS1* haplotype, i.e. they are negative controls. Homozygous (AA) genomic DNA samples, heterozygous (AC) genomic DNA samples and homozygous (CC) "cloned" DNA samples are represented by red, green and blue circles, respectively. Heterozygous (AC) cDNA samples from 1781 (T#4) are represented by purple markers. Heterozygous (AC) cDNA samples from 1782 (T#5) and 1825 (T#13) – the negative controls – are marked by yellow and black markers, respectively. In some cases, the cDNA is derived from RNA extracted from cells cultured in hypoxia (0.1%, 1% or 2% O_2) as shown in the figure key.

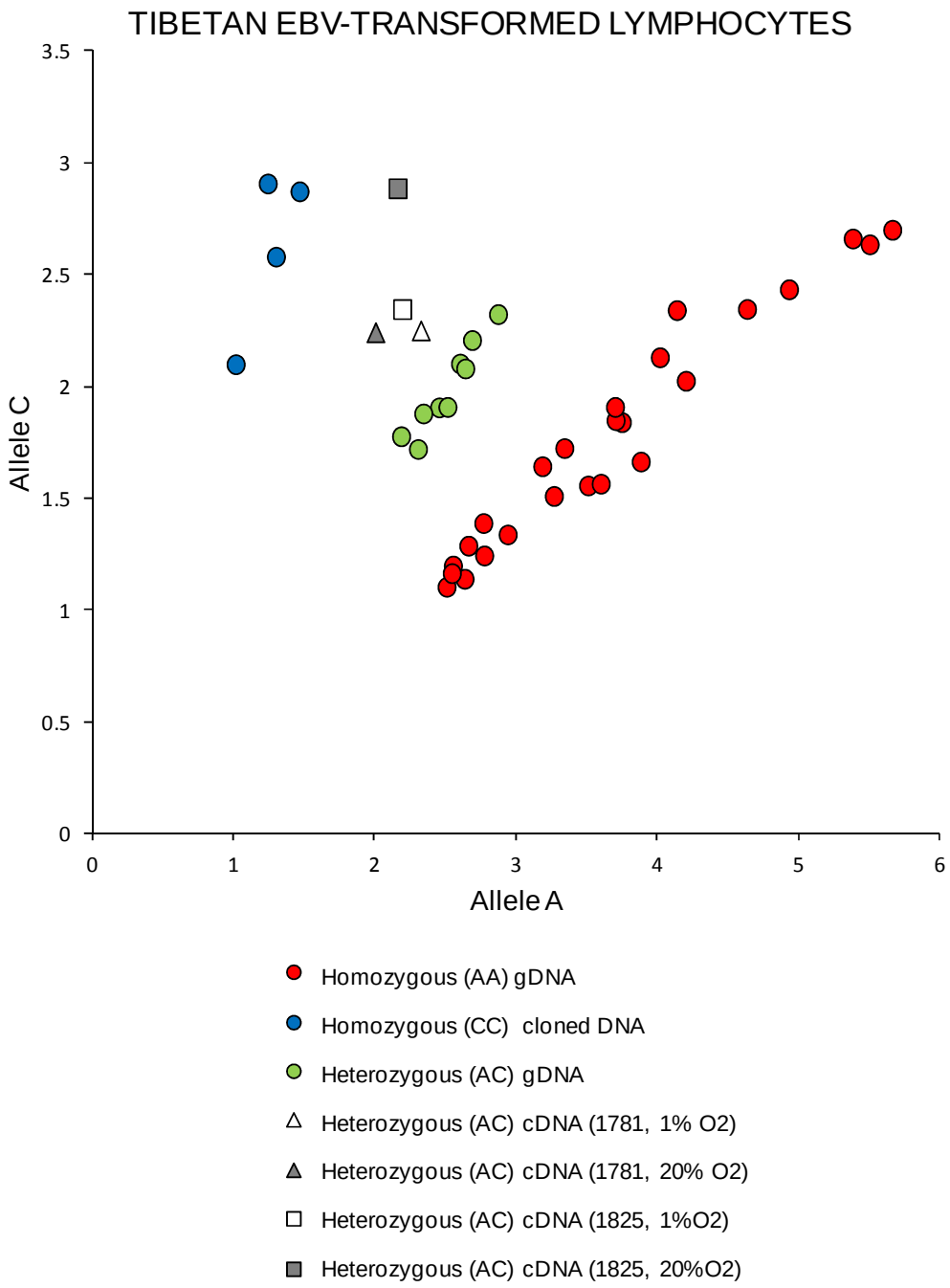


Figure 32: TaqMan genotype-specific PCR assay assessing allele-specific expression in Tibetan EBV-transformed lymphocytes. The y-axis represents fluorescent intensity (normalised to background) for allele C (FAM dye) and the x-axis fluorescent intensity (normalised to background) for allele A (VIC dye). Homozygous AA, heterozygous AC and homozygous CC refer to the genotypes with respect to the marker coding SNP rs59901247. Sample 1781 (T#4) is also heterozygous for the putatively selected Tibetan *EPAS1* haplotype. Sample 1825 (T#13) is minor homozygous for the *EPAS1* haplotype, i.e negative control. Homozygous (AA) genomic DNA samples, heterozygous (AC) genomic DNA samples and homozygous (CC) "cloned" DNA samples are represented by red, green and blue circles, respectively. Heterozygous (AC) cDNA samples from 1781 (T#4) are represented by triangular markers. Heterozygous (AC) cDNA samples from 1825 (T#13) – the negative control – are marked by square markers. In some cases, the cDNA is derived from RNA extracted from cells cultured in hypoxia (1% O_2) as shown in the figure key.

To test the hypothesis that this observation is not specific to Tibetan cells or to cells of lymphoid origin, cDNA from primary lymphocytes originating from Han Chinese volunteers – see results in Figure 33 – and a number of heterozygous cancer cell lines (for rs59901247) – see results in Figure 34A and B – were also tested using the same assay. It is evident that the allelic shift towards the "C" allele is found in all the heterozygous (for rs59901247) cDNA samples tested.

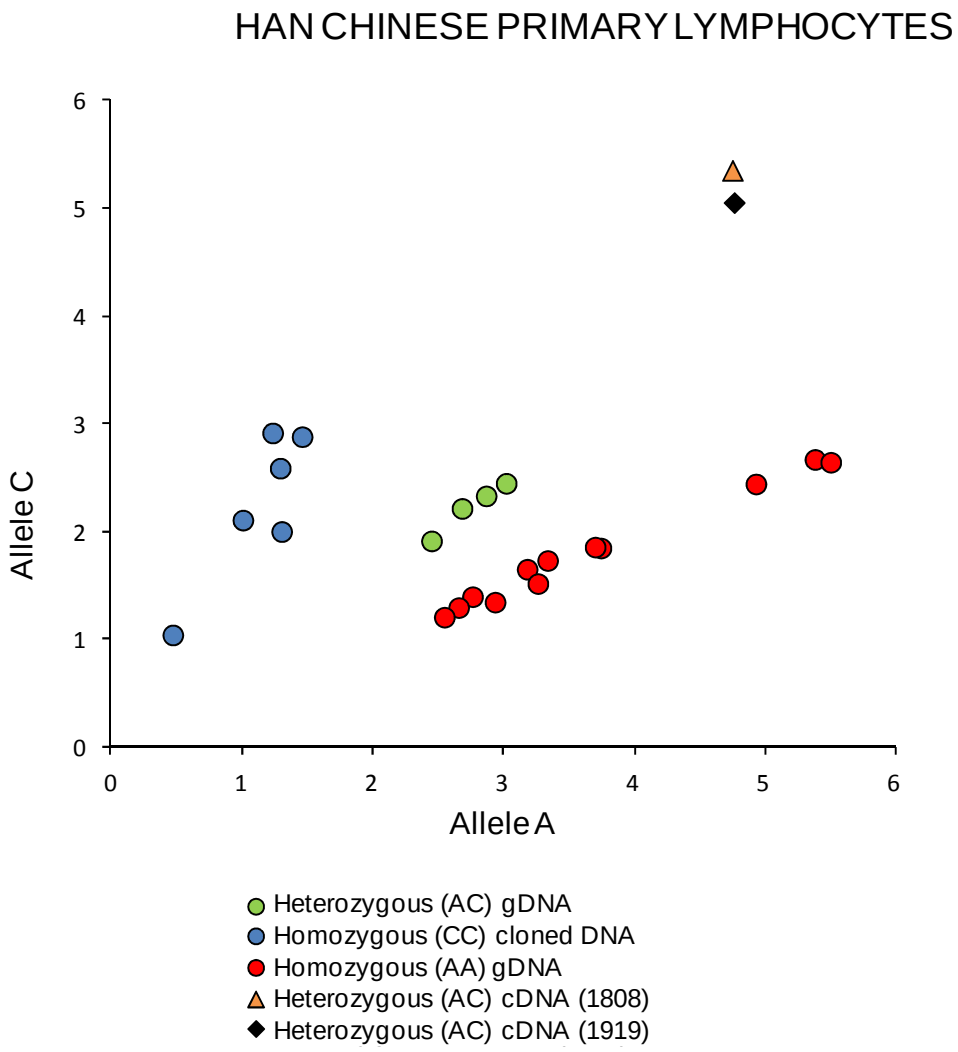


Figure 33: TaqMan genotype-specific PCR assay assessing allele-specific expression in Han Chinese primary lymphocytes. The y-axis represents fluorescent intensity (normalised to background) for allele C (FAM dye) and the x-axis fluorescent intensity (normalised to background) for allele A (VIC dye). Homozygous AA, heterozygous AC and homozygous CC refer to the genotypes with respect to the marker coding SNP rs59901247. Homozygous (AA) genomic DNA samples, heterozygous (AC) genomic DNA smples and homozygous (CC) "cloned" DNA samples are represented by red, green and blue circles, respectively. Heterozygous (AC) cDNA samples from Han Chinese 1808 and 1918 are represented by yellow triangles and black rhomboids, respectively.

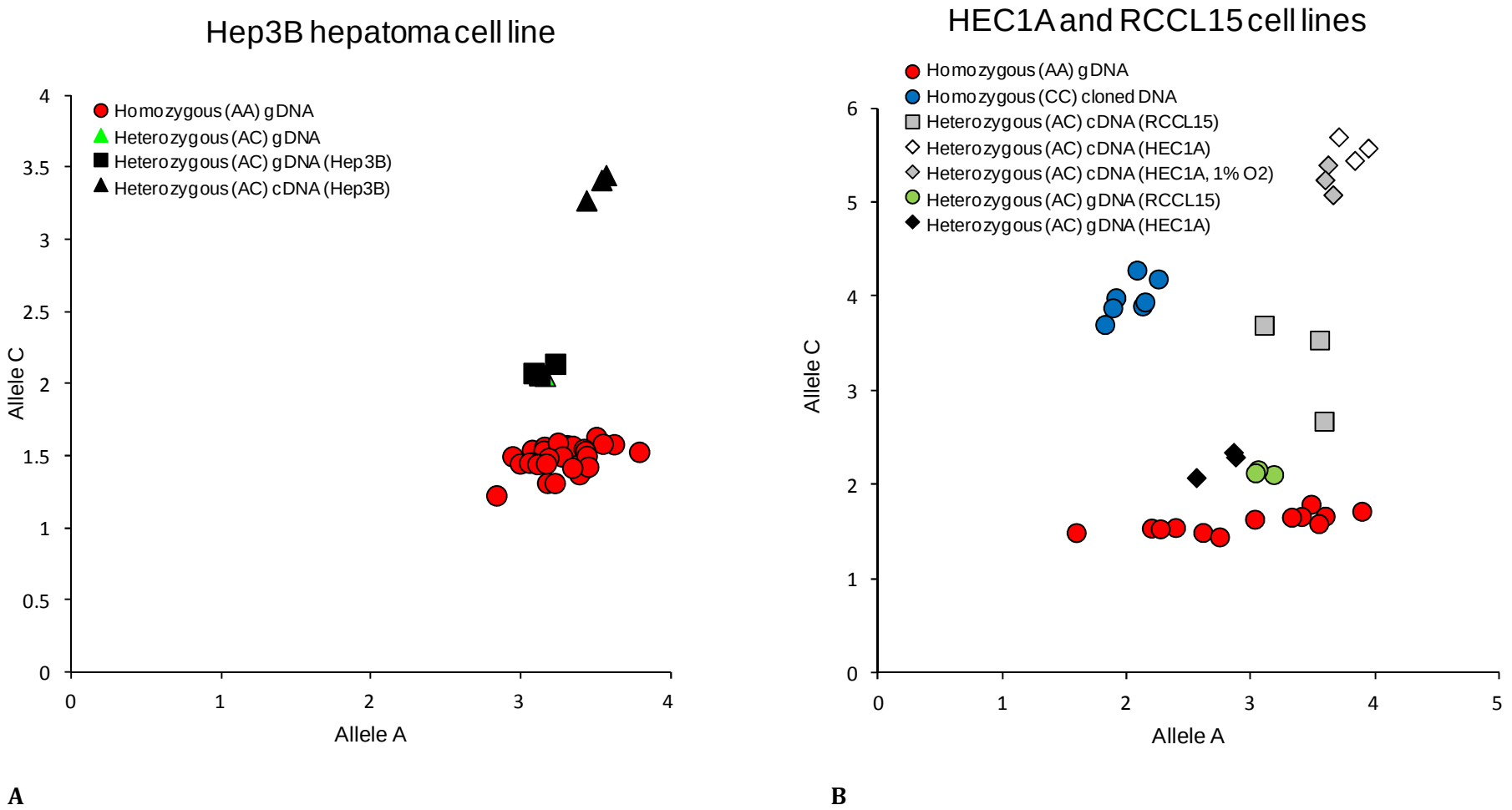


Figure 34: TaqMan PCR assay assessing allele-specific expression in (A) Heterozygous Hep3B hepatoma cell line; (B) heterozygous RCCL15 renal cell cancer cell line and HEC1A endometrial cell cancer cell line. Experiment and figure provided by Dr Chiara Bardella.

6.3.3 Results from primer extension and mass spectrometry assay

To investigate whether the allelic shift observed in section 6.3.2 is the result of technical bias in the TaqMan assay or whether it is the result of a real "functional effect" of the coding SNP rs59901247 the methodology described in section 6.2.3 was used.

<u>Main finding:</u> The experiments using this methodology confirm that: 1. The results of the Taqman assay in section 6.3.2 are most likely a result of technical bias; and 2. There is no evidence of allele-specific expression of EPAS1 in Tibetan lymphoblastoid cells lines.

Along with the Tibetan samples, a calibration series was included which consisted of samples made up by mixing known quantities of cloned DNA carrying allele "C" and cloned DNA carrying allele "A" at various ratios. All the samples tested are shown in Table 40.

Table 40: Samples used in the MassArray experiment

Calibration curve C:A ratio	gDNA	cDNA
8:1	1781 gDNA (heterozygous Tibetan)	1781 cDNA (1% O_2)
4:1	1825 gDNA (heterozygous Tibetan)	1781 cDNA (20% O_2)
2:1	1824 gDNA (homozygous AA Tibetan)	1825 cDNA (1% O_2)
1.5:1		1825 cDNA (20% O_2)
1:1		
1:1.5		
1:2		
1:4		
1:8		

Figure 35 and Figure 36 show examples of the spectra derived for some of the samples analysed. Figure 37 shows the calibration curve produced when the measured peak area ratio (allele C : allele A) is plotted against the expected peak area ratio (allele C : allele A) on double logarithmic scale. This confirms that indeed the assay is able to differentiate between the relative abundance of the two alleles in a quantitative way.

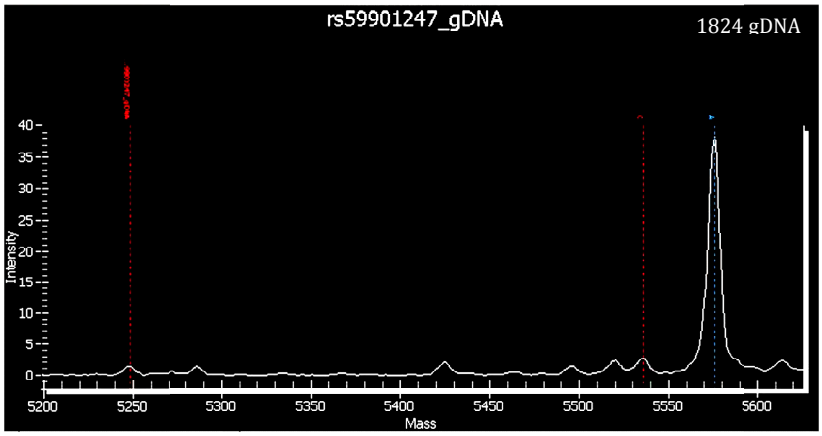


Figure 35: Screen shot showing spectrum derived for a homozygous (AA) gDNA sample (1824) showing one peak only.

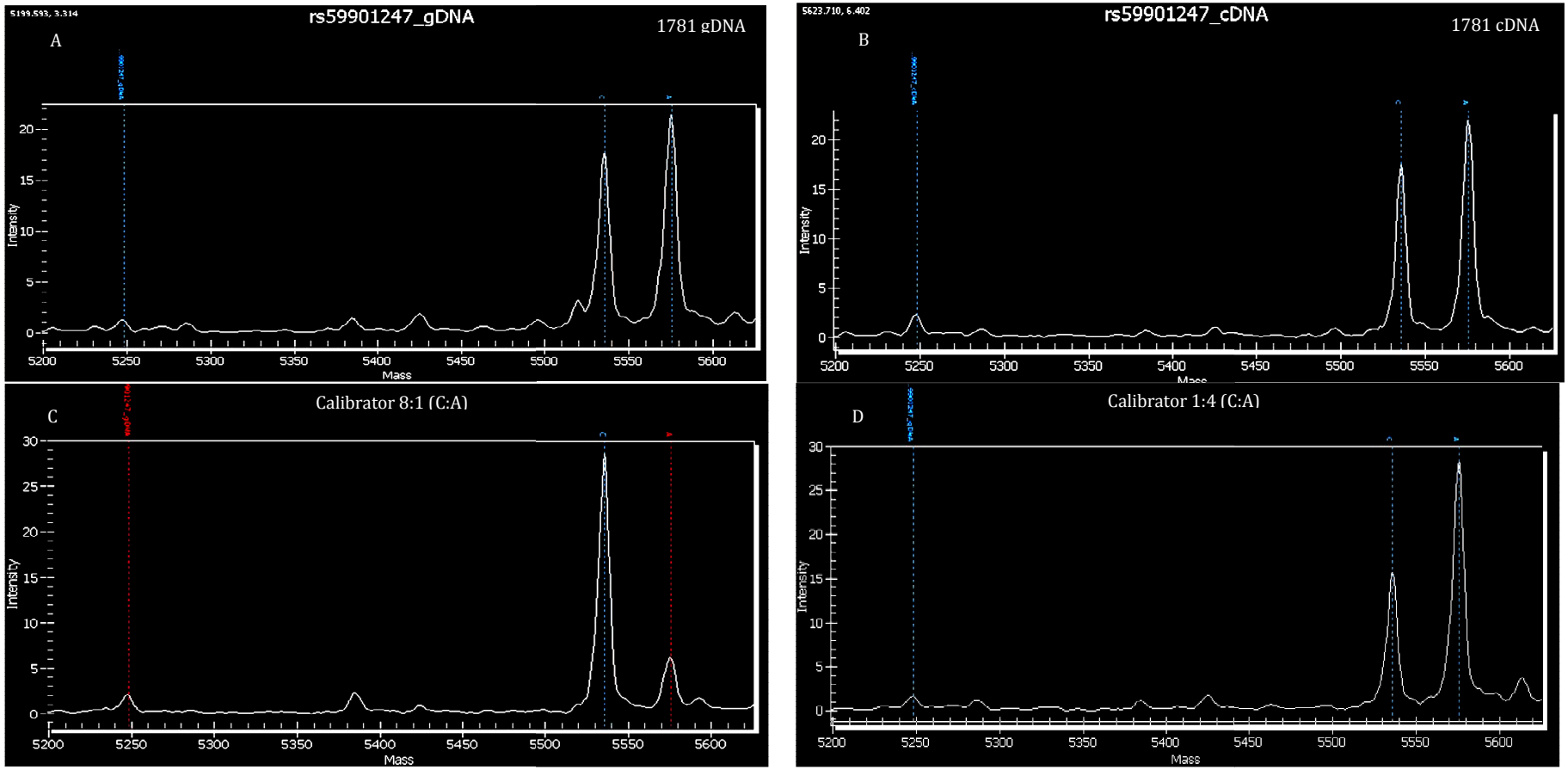


Figure 36: Screen shot showing spectra derived from (A) a heterozygous (AC) gDNA sample and (B) a heterozygous (AC) cDNA sample from 1781; (C) a "calibrator sample with a 8:1 ratio of allele "C" to allele "A" and (D) a calibrator sample with a 1:4 ratio of allele "C" to allele "A".

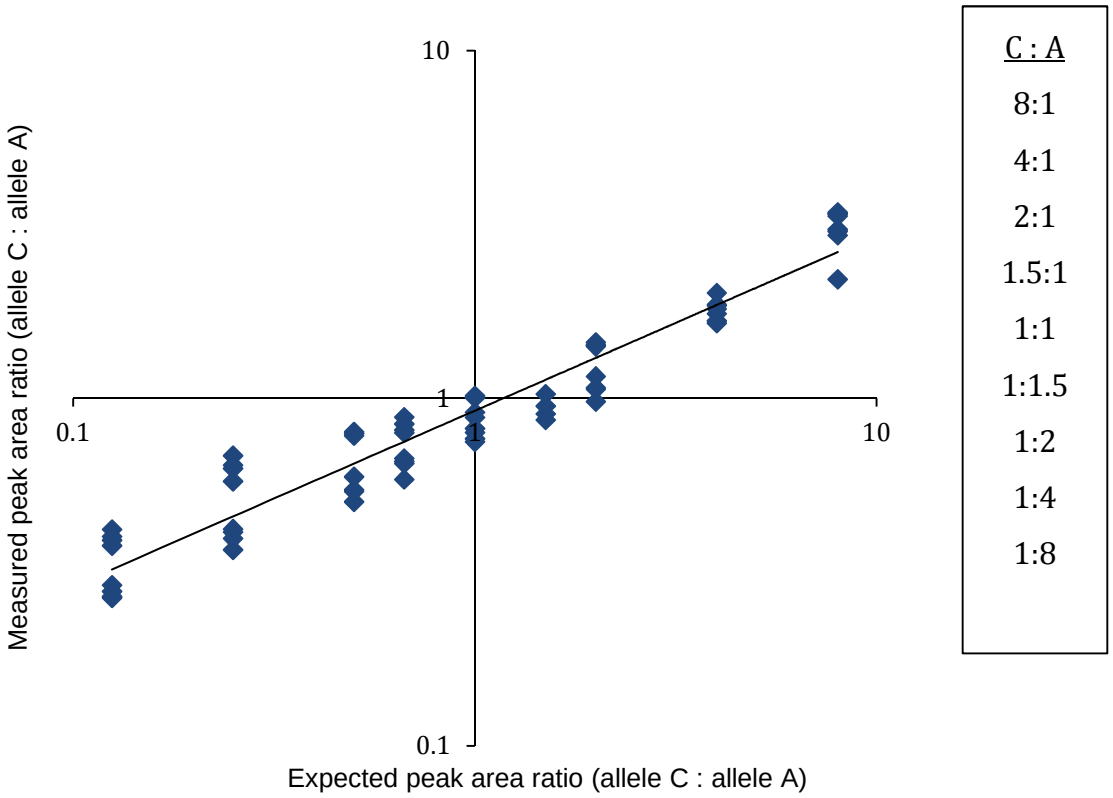


Figure 37: Calibration curve showing "Measured allele C: allele A peak area ratio" plotted against "Expected allele C: allele A peak area ratio" on a double logarithmic scale

The following method was used to analyse the data produced. The ratio $\frac{\text{Area under peak for Allele C}}{\text{Area under peak for Allele A}}$ was derived for each spectrum produced. The mean $\frac{\text{Area under peak for Allele C}}{\text{Area under peak for Allele A}}$ ratio in genomic DNA ($Ratio_g$) was calculated. For each cDNA replicate sample, a $\frac{\text{Area under peak for Allele C}}{\text{Area under peak for Allele A}}$ ratio ($Ratio_c$) was calculated – this represents the mean ratio of the 4 spectra produced for each cDNA sample. Mean ratio values are shown in Table 41. $Ratio_c$ was compared to $Ratio_g$ using a paired t-test *(SPSS 19)*. This test statistic was not significant for either Tibetan cell line 1781 (heterozygous for the selected "Tibetan" haplotype) or for Tibetan cell line 1825 (minor homozygous for selected "Tibetan haplotype i.e. the negative control), as shown in Table 42. In addition, to assess whether indeed 1781 and 1825 differ in terms of allele-specific expression from one another, an unpaired t-test was used to compare the difference in $Ratio_c$ - $Ratio_g$ (which was

calculated for each cDNA sample) between 1781 samples and 1825 samples. This difference was statistically significant as shown in Table 43. Figure 38 shows a histogram displaying the expression of allele "C" relative to allele "A" (normalised to genomic ratio), highlighting the subtle but statistically significant difference between 1781 and 1825. This is discussed in the discussion section 7.4.

Table 41: Mean genomic and cDNA ratios (Allele C : Allele A)

	Mean	N	St Deviation	St Error Mean
$Ratio_g$	0.81377	12	0.06401	0.01848
$Ratio_c$ (all samples)	0.80383	12	0.07291	0.02105
$Ratio_c$ (1781)	0.74105	6	0.02663	0.01087
$Ratio_c$ (1825)	0.86660	6	0.03911	0.01597

Table 42: Paired t-test comparing the cDNA ratio in each sample with the gDNA ratio

$Ratio_c$ - $Ratio_g$	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval	t	Df	p-value
All samples	-0.00994	0.09343	0.02697	-0.06930 to 0.04942	-0.369	11	0.719
1781	-0.07272	0.07497	0.03061	-0.15139 to 0.00596	-2.376	5	0.063
1825	0.05283	0.06422	0.02622	-0.01457 to 0.12023	2.015	5	0.100

Table 43:Unpaired t-test comparing the difference ($Ratio_c$ - $Ratio_g$) between 1781 and 1825

Difference: $Ratio_c$ - $Ratio_g$					
Mean	**Std. Error Mean**	**95% Confidence Interval**	**t**	**Df**	**p-value**
-0.12007	0.04031	-0.20989 to -0.30248	-2.979	10	0.014

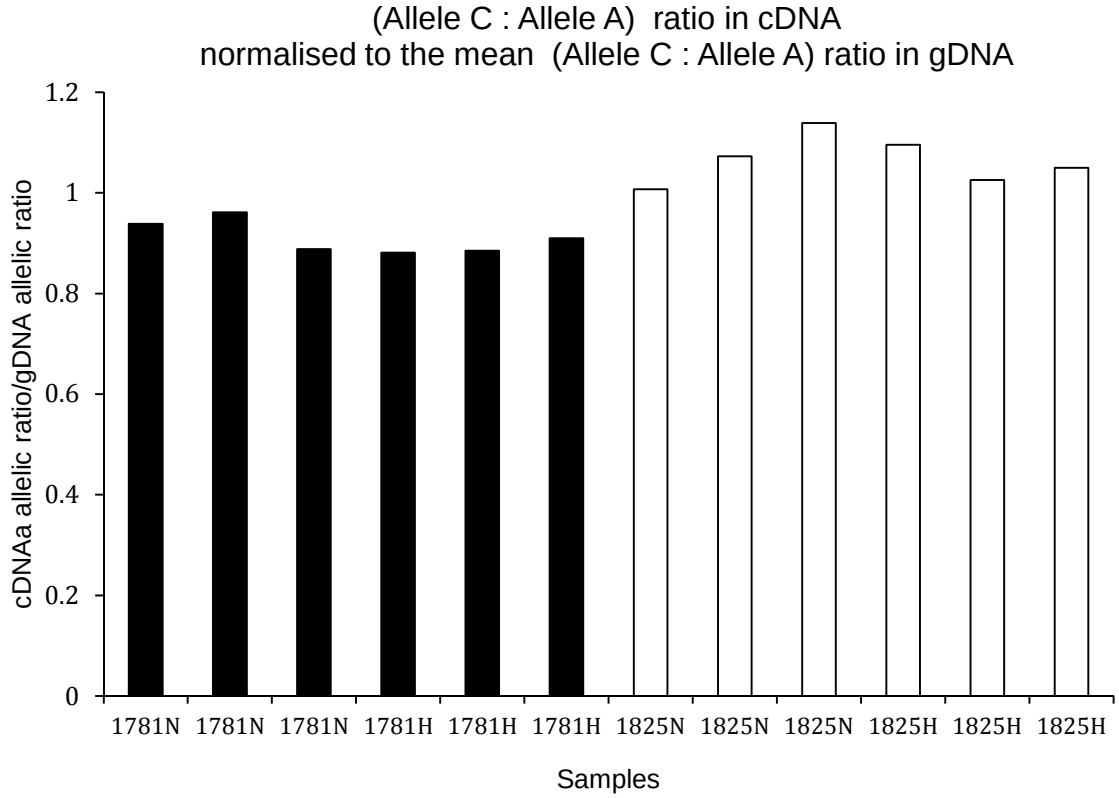


Figure 38: Histogram showing the expression of allele "C" relative to allele "A" (normalised to the mean gDNA allelic ratio). N refers to cDNA from cells cultured in normoxia (euoxia). H refers to cDNA from cells cultured in hypoxia (1% O_2). For both 1781 and 1825 the expression of allele "C" relative to allele "A" is very close to 1, when normalised to the allelic ratio in gDNA.

6.3.4 Results from the experiments assaying allele-specific RNA Polymerase II loading at the *EPAS1* locus

This is an alternative approach of looking for allele-specific expression by assessing whether there is a difference in the amount of Pol II binding at the *EPAS1* locus at one versus the other allele. The advantage of this approach is that it allows assessing allele-specific expression using intronic or non-coding SNPs as markers, when there are no informative transcribed SNPs that can be used. RNA Polymerase II Chromatin Immunoprecipitation (ChIP) was performed in heterozygous – for the putative high-altitude selected genetic haplotype – Tibetan EBV-transformed lymphocytes. It was also

performed in a cell line which was minor homozygous for the putative high-altitude haplotype.

The level of enrichment obtained at the *EPAS1* locus was considerably low compared to enrichment at the B-actin locus (which ranged between 13-fold to 200-fold in the samples tested). Fold enrichment is defined as the abundance of the DNA fragment (amplified by PCR) in the sample treated with RNA Pol II antibody relative to the sample treated by the non-specific RbIgG antibody. At the *EPAS1* locus, there was no enrichment at a DNA locus in intron 1 containing SNP rs6758592 – which is also a marker SNP for the putatively selected haplotype. This suggests that there is very low "run-through" of RNA Polymerase II across the *EPAS1* gene in lymphoblastoid cells, which is consistent with low HIF2α expression in this cell type. At the *EPAS1* transcriptional start site (TSS), there was a 4-fold enrichment in euoxia and an 11-fold enrichment in hypoxia, which confirms that indeed there is Pol II binding at the transcriptional start site. However, no informative SNPs were present in the available cell lines at the genomic region near the TSS so as to allow exploration of allele-specific binding of Pol II at this locus. Nevertheless, available data sets give evidence of a second Polymerase II binding site ("peak") at the EPAS1 locus, found downstream of the gene, as shown in Figure 39. Interestingly, rs1447563, which is one of the 8 GWADS SNPs that achieved genome-wide significance in [115], is found in close proximity to this locus. Enrichment at rs1447563 locus was tested (and also at rs12615428, another SNP found in close proximity to the Pol II peak) and it was found that there is some enrichment (~ 5-fold). Hence, I decided to explore allele-specific Pol II loading at the locus of rs1447563 using a TaqMan PCR assay in the heterozygous Tibetan cell lines. The results are shown in Figure 40. In (A) there is evidence of increased "expression" of the minor allele "A" (alias "non-Tibetan") suggesting increased Pol II

binding at minor allele “A”, both in euoxia and in hypoxia. In (B) there is no evidence of differential allelic Pol II binding. The experiment was repeated a third time and the results were similar to (B).

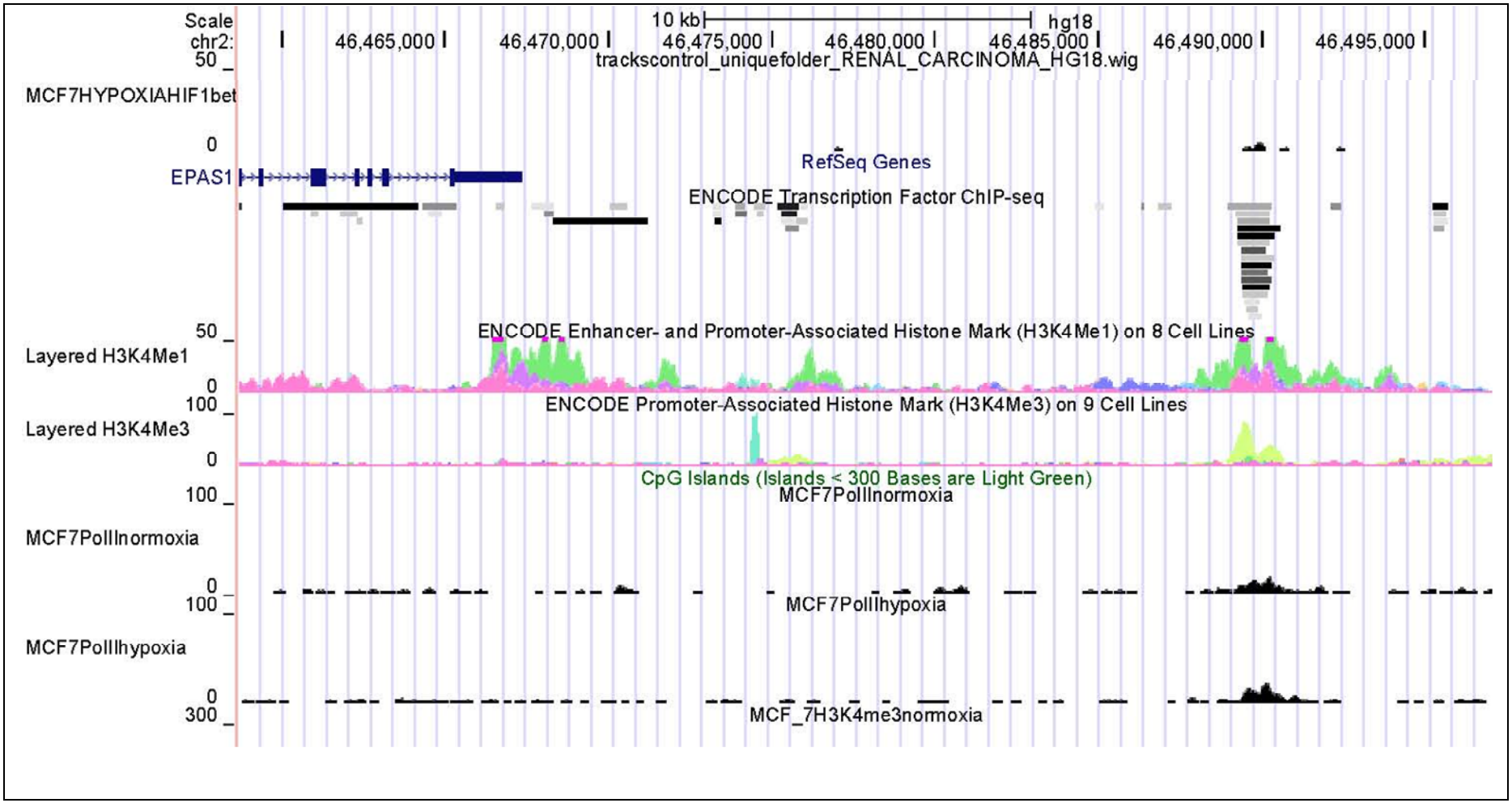


Figure 39: Data from UCS genome browser, shared by Dr David Mole. Approximately 20 kb downstream of the *EPAS1* gene there is a locus of increased transcription factor binding (in many different cell lines) and Pol II binding in euoxia and hypoxia in MCF7 cell lines. There is also increased H3K4Me1, a marker of enhancer activity. These suggest that this is likely a regulatory region.

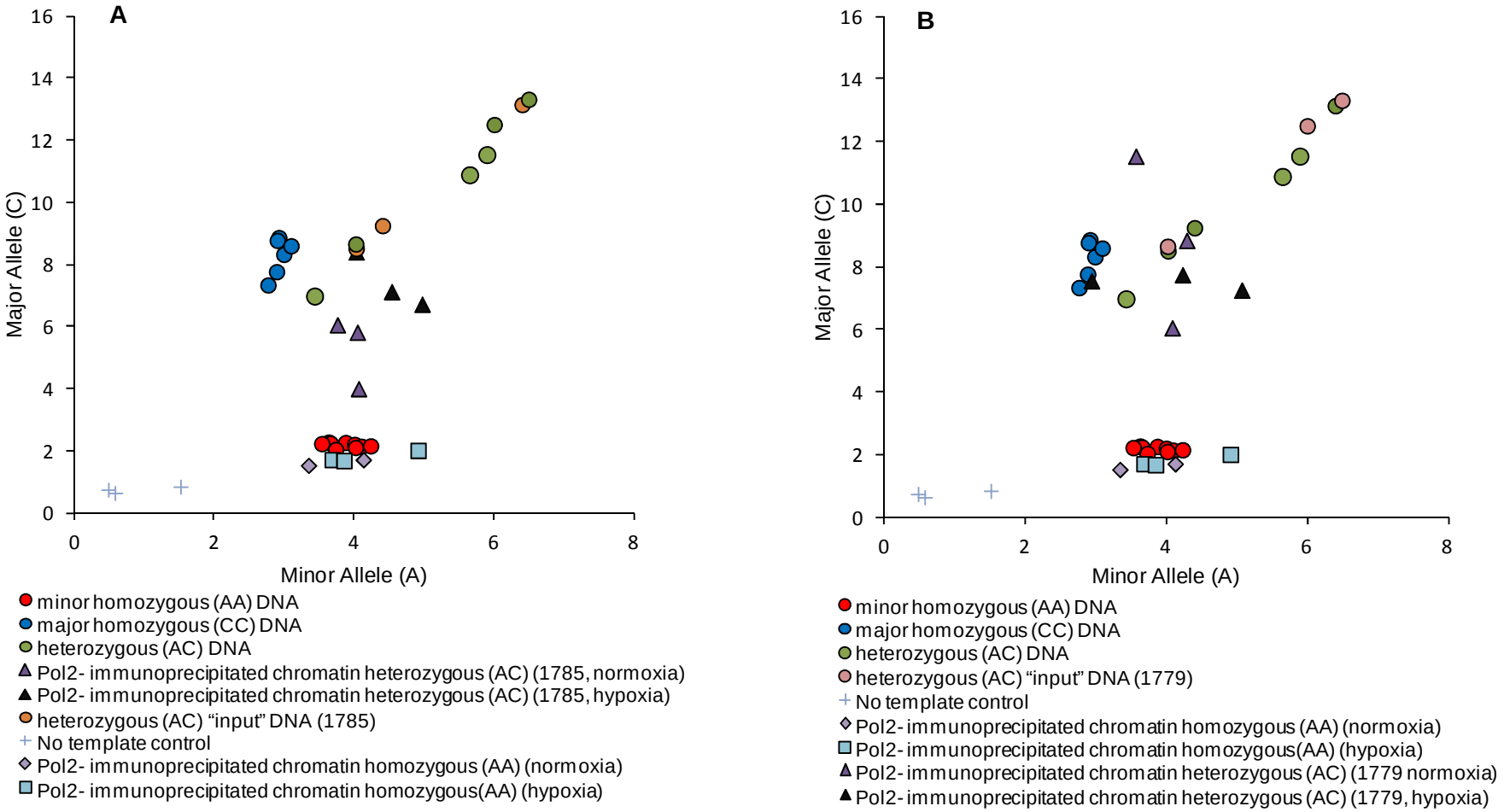


Figure 40: TaqMan allele-specific qPCR assay assessing allele-specific Pol II binding at rs1447563 for (A) heterozygous Tibetan cell line 1785, Pol II immunoprecipitated DNA from euoxia and hypoxia, denoted with purple and blue triangles, respectively; and (B) heterozygous Tibetan cell line 1779, Pol II immunoprecipitated DNA from euoxia and hypoxia, denoted with purple and blue triangles, respectively. In both figures circles represent the same experimental data points from genomic DNA. Blue squares and purple rhomboids represent the same experimental data points denoting Pol II immunoprecipitated DNA from euoxia and hypoxia experiments for a minor homozygous cell line. The y-axis represents normalized fluorescent intensity for allele C and the x-axis fluorescent intensity for allele A.

6.4 DISCUSSION

6.4.1 Allele-specific gene expression

Allele-specific gene expression refers to the different levels of transcription of a gene from one relative to another allele. Extreme forms of allele-specific expression are found when expression is restricted to a single allele (monoallelic expression) as seen in X-chromosome inactivation in females [172] and in genomic imprinting where expression of the imprinted gene occurs in a parent-of-origin-specific manner [173]. However, there is increasing evidence of differential allelic expression in autosomal non-imprinted genes [174]. The magnitude of this can be variable and often relatively modest, but can have significant biological significance. For the same gene, allele-specific expression can be tissue- and/or context-specific. Allele-specific quantification techniques have allowed detection of differences in allelic expression as low as 20%. There is increasing evidence that variation in non-coding DNA is probably as important as coding variation in producing functionally biological effects, through regulation of gene expression. Allele-specific analysis of gene expression provides a useful functional approach to defining regulatory variants.

6.4.2 In search of allele-specific expression of *EPAS1* in Tibetans using a transcribed SNP as a marker

Despite a number of genomic studies identifying *EPAS1* as undergoing selection in Tibetans, the causal or functional variant(s) are not known. There is high degree of linkage disequilibrium at the *EPAS1* putatively selected genomic region. On one hand, this can be of help in defining individuals as major homozygotes, heterozygotes or minor homozygotes for the *EPAS1* locus; this is especially so as the functional variant is unknown and so the high degree of linkage disequilibrium gives increased confidence or

likelihood that the functional variant will be "linked" to the SNPs used to define the haplotype/genotype of an individual. On the other hand, if one were able to demonstrate allele-specific expression, while this would be valuable in confirming that genetic variation is functional by affecting *EPAS1* transcription, the high linkage disequilibrium can cause great difficulty in resolving the true functionally important variants.

The TaqMan SNP genotyping assay illustrates the importance of having negative control cell lines in the assay for allele-specific expression. There is evidence of allelic imbalance in cDNA towards the same allele "C" of the coding marker SNP in both the Tibetan sample of interest (1781) and in the negative control Tibetan samples (1782, 1825) but also in the Han Chinese samples and in a number of cell lines. This suggests that either the TaqMan assay for this coding marker SNP is technically biased or that this coding marker SNP is functional in its own right, e.g. through affecting mRNA stability.

To resolve the possibility of technical bias associated with TaqMan probes and primers for this particular SNP, one approach would be to use a TaqMan assay for an alternative transcribed marker SNP. Unfortunately, no other transcribed SNP was identified in any of the Tibetan samples that were Sanger sequenced in this study. Thus, I decided to use an alternative assay altogether which would resolve both possibilities. This is the methodology that uses primer extension and mass spectrometry (*Sequenom*) described in section 6.2.3. Using this method, it became evident that the most likely explanation for the observations in the TaqMan assay is a technical bias. In addition, at least in the Tibetan lymphocytes, there is no evidence that the coding SNP rs59901247 has a functional effect on HIF2α mRNA expression. This is because, not only is the relative allele expression in "opposite directions" in 1781 and 1825 samples, as seen in Figure 38, but also for both

1781 and 1825 the allelic imbalance of alleles “A” and “C” in cDNA relative to genomic DNA is of extremely small magnitude (< 10%), too small to be of biological significance, and is not statistically significant. While the comparison of imbalance in allelic expression between 1781 and 1825 is statistically significant, this is unlikely to have any biological meaning.

I have not been able to show allele-specific expression of *EPAS1* in lymphocytes of individual 1781, neither in euoxia nor in hypoxia. As I mention in section 6.4.1, allele-specific expression can be tissue-specific. The expression of HIF2α in lymphocytes is relatively low, especially in comparison to HIF1α. Thus, it is possible that while allele-specific expression of *EPAS1* in Tibetans may not be found in lymphocytes, it may still occur in other tissues where HIF2α is in greater abundance and where it may play a weightier role in regulating gene expression, e.g. in endothelial cells or in erythropoietin-producing cells.

6.4.3 In search of allele-specific expression of *EPAS1* in Tibetan by assessing RNA Polymerase II loading at the *EPAS1* locus using Pol II ChIP.

This approach was undertaken in parallel with the MassArray/MALDI-TOF/MS (*Sequenom*) approach, following the equivocal results of the initial TaqMan SNP PCR assay using the coding marker SNP rs59901247. The advantage of this approach is that in the lack of informative transcribed SNPs, intronic non-transcribed SNPs can be used as markers to assess allele-specific Pol II binding at *EPAS1*, which is taken as a surrogate of allele-specific expression. In lymphoblastoid cell lines it was found that there was very low run-through of RNA Polymerase II across the *EPAS1* gene, consistent with low HIF2α expression in this cell type. At the transcriptional start site, where there is evidence of Pol

II binding, there are no informative SNPs. Consequently, I was not able to use intronic SNPs as markers to assess allele-specific Pol II loading at *EPAS1* – and hence allele-specific expression – using this method.

However, at a genomic region ~ 20 kb downstream of the *EPAS1* gene there is a Pol II binding "peak", which is also related to increased transcription factor binding (as shown in Figure 39) and increased H3K4Me1 (histone-3 lysine-4 monomethylation) signal, which has been associated with transcriptional enhancers. This suggests that this is a regulatory region. The data obtained from the UCS genome browser and displayed in Figure 39 are amalgamated from studies in a number of cell lines and the Polymerase II enrichment data are generated from Pol II ChIP-Seq experiments in a MCF7 cell line (shared by Dr David Mole). It is difficult to know to what extent this genomic region also has regulatory function in lymphoblastoid cell lines. In lymphoblastoid cell lines there is some enrichment at this locus as I show in my Pol II ChIP experiments, but this is low. The results from the TaqMan allele-specific assay shown Figure 40A suggest increased Pol II binding at the minor (alias "non-Tibetan") allele but this result is reproduced in other cell lines as shown in Figure 40B.

EPAS1 is the closest transcribed gene to the presumed regulatory region referred to above and so it is tempting to hypothesize that this region is involved in the transcriptional regulation of *EPAS1*. In addition, transcription factors that bind this site, such as HNF-4α (hepatocyte nuclear factor 4α) are known to be involved in hypoxic responses and in the transcriptional activation of genes regulated by the HIF pathway, e.g. erythropoietin [175]. At any rate, spatial and temporal regulation of genes is complex. There is increasing evidence that regulatory elements can be located far from their target genes, sometimes

even several megabases [176-178]. There is also evidence that human disease is associated with mutations in distant regulatory elements and not in the actual disease-relevant gene [179, 180]. Furthermore, regulatory elements can directly associate with their target gene promoters through chromatin looping [176-178, 181].

In [115] 8 GWAD significant SNPs were identified to be in high pairwise linkage disequilibrium in Tibetan samples and to form an extended haplotype found in high frequencies in Tibetans. These SNPs lie between 366 bp and 235 kb downstream of *EPAS1*; thus this region of linkage disequilibrium, which extends into the *EPAS1* gene itself, spans the putative regulatory region located at ~ 20 kb downstream of *EPAS1* referred to earlier. One attractive hypothesis is that there is a regulatory element in this region that affects transcription of *EPAS1* – this may involve chromatin looping to allow association with the gene's promoter – and that in Tibetans the functional or causal variant at the *EPAS1* locus, which is naturally selected for high-altitude adaptation, lies within this region and affects regulation of transcription of *EPAS1*. Further functional studies are required to investigate this hypothesis. Bearing in mind that regulation of gene expression is also tissue-specific, these studies should be performed in cells or tissues that express high amounts of HIF2α and that are phenotype-relevant e.g. endothelial cells or erythrocyte progenitor cells.

Chapter 7: Is the genetic variation at *EGLN1* found in Tibetans functional at the cellular/molecular level?

Genome-wide studies investigating high altitude adaptation in Tibetans have identified positive selection at the *EGLN1* locus, too. In the peripheral blood lymphocyte study in Chapter 5, unlike HIF2α, I found no difference in expression of PHD2 at the mRNA level between Tibetans and Han Chinese. It may well be that genetic variation in *EGLN1* is functional at the post-transcriptional level affecting protein translation, stability or function. In support of this, two coding variants have been described of particularly high frequency in Tibetans – which co-occur and are in high linkage-disequilibrium with SNPs identified in genetic adaptation studies in Tibetans – although their functional significance is not yet known.

Thus, I decided to investigate directly whether the mis-sense coding variants in exon 1 have functional significance at the molecular level by affecting the enzymatic activity of PHD2. To do this, I implemented three experimental approaches: 1) Determination of PHD2 protein level and HIF1α protein stabilisation by hypoxia in lymphoblastoid cell lines of different genotypes (for *EGLN1*); 2) Assessment of the hydroxylation activity of wild-type PHD2 and of "mutant" PHD2, i.e. PHD2 D4E C127S, *in vitro*; and 3) Determination of HIF1α protein level and of HIF1α hydroxylation in Phd-null mouse embryonic fibroblasts (MEF) infected with a viral vector carrying the coding sequence of either wild-type PHD2 or PHD2 D4E C127S.

7.1 METHODS

7.1.1 Protein immunoblot (western blot) protocol

The method of western blotting is described below, as this is relevant to a number of sections in this chapter.

7.1.1.1 SDS-PAGE (sodium dodecyl sulphate polyacramide gel electrophoresis)

Protein samples (in Laemmli buffer*) are loaded into wells in a SDS-PAGE gel* alongside a protein marker (*PageRuler Plus Prestained Protein Ladder, Thermo Scientific)* and are subjected to electrophoresis at a voltage of 100 V in SDS-PAGE Running buffer *(1×).

7.1.1.2 Immunoblotting ("Transfer")

Proteins resolved by SDS-PAGE are electroblotted onto a polyvinylidene difluoride membrane (PVDF) (*Millipore).* This is done by placing the membrane on top of the SDS-PAGE gel and "sandwiching" these membrane-gel stacks between chromatography paper sheets (*Whatman*). The entire stack is placed in a cassette in a tank filled with Transfer buffer* (1×) and an electrical current is applied through it (180 mA for 25 min). This causes the proteins to "transfer" from the gel onto the membrane, maintaining the position they had on the gel.

7.1.1.3 Blocking

After transfer, the membrane is placed in a dilute protein solution, typically 5% non-fat dry milk (*Sigma Aldrich*) in Phosphate Buffered Saline with Tween (1× PBST*) and incubated for 30 min to 1 hour at room temperature (RT) with gentle rocking. This aims to prevent non-specific binding of the primary antibody that is subsequently added.

7.1.1.4 Probing with antibody

A primary antibody solution (in 5% milk in PBST) is added to the membrane for a minimum of 90 min at RT or overnight at 4 °C, with gentle rocking. The antibody solution is retrieved and the membrane is washed four times with PBST solution (for 4 min with gentle rocking) in order to remove any unbound primary antibody. The membrane is then placed in a secondary antibody solution (in 5% milk in PBST) and incubated at RT for 1 hour, with gentle rocking. The secondary antibody is linked to a reporter enzyme (horseradish peroxidase) and binds in a species-specific manner to the constant region of the primary antibody.

7.1.1.5 Detection

Following four wash steps (4 min each) with PBST solution, to remove unbound antibody, a chemiluminescent agent (*SuperSignal West Dura Chemiluminescent Substrate, Thermo Scientific)* is added to the membrane; this is cleaved by the horseradish peroxidase enzyme that is conjugated to the secondary antibody producing luminescence. This luminescence is proportional to the amount of protein present. The light is then detected by photographic film (*Kodak*) or by a CCD camera which captures a digital image of the western blot.

7.1.2 Experimental procedure for assessing HIF1α protein stabilisation in EBV-transformed lymphocytes of different genotypes

Cell lines were cultured as described in section 5.1 and incubated in euoxia (in a 5% CO_2 tissue-culture incubator) and in hypoxia (1% O_2 and 5% CO_2 in a hypoxia workstation), for 2 hours, 4 hours and 24 hours and then harvested in the following way: the cell suspension was centrifuged, the medium (supernatant) was removed, the cell pellet was washed with PBS and then lysed and re-suspended in Igepal Lysis buffer[*]. Following incubation on ice

for 5 min and centrifugation at 13000 rpm for 4 min at 4 °C to remove the cell debris and DNA, the supernatant, which contained the soluble protein extracts, was recovered. In the case of hypoxia, the cells were harvested and lysed in the hypoxia workstation. Cell pellets and protein extracts were kept on ice at all times. The protein extracts were then denatured by mixing with Laemmli buffer (3×) and heating at 90 °C for 3 min. The extracts were stored at -20 °C if not used immediately.

Extracts were then resolved by SDS-PAGE, blotted onto PVDF membranes and probed with primary antibodies followed by HRP-conjugated secondary antibodies, as described in section 7.1.1. The antibodies used were: mouse monoclonal anti-HIF1α (*BD Transduction Laboratories)* at 1:1000 dilution, mouse monoclonal anti-PHD2 (clone 76a, *Millipore)* at 1:100 dilution and HRP-conjugated anti-β-actin *(Abcam)* at 1:5000 dilution, to assess HIF1α, PHD2 and β-actin levels, respectively.

7.1.3 Construction of recombinant plasmids harbouring the coding region of either wild-type or mutant PHD2

In order to undertake experiments assessing the hydroxylation of wild-type and mutant human PHD2 proteins *in vitro*, recombinant plasmids harbouring the coding region of either wild-type or mutant versions of PHD2 were constructed. The following eight plasmid vectors were created: wild-type pcDNA3-PHD2, pcDNA3-PHD2 C127S, pcDNA3-PHD2 D4E, pcDNA3-PHD2 D4E C127S, wild-type pRRL-PHD2, pRRL-PHD2 C127S, pRRL-PHD2 D4E, pRRL-PHD2 D4E C127S. Maps of the two plasmid vectors used, pcDNA3 and pRRL, are shown in Figure 41. The methodology used is described in the sub-sections that follow.

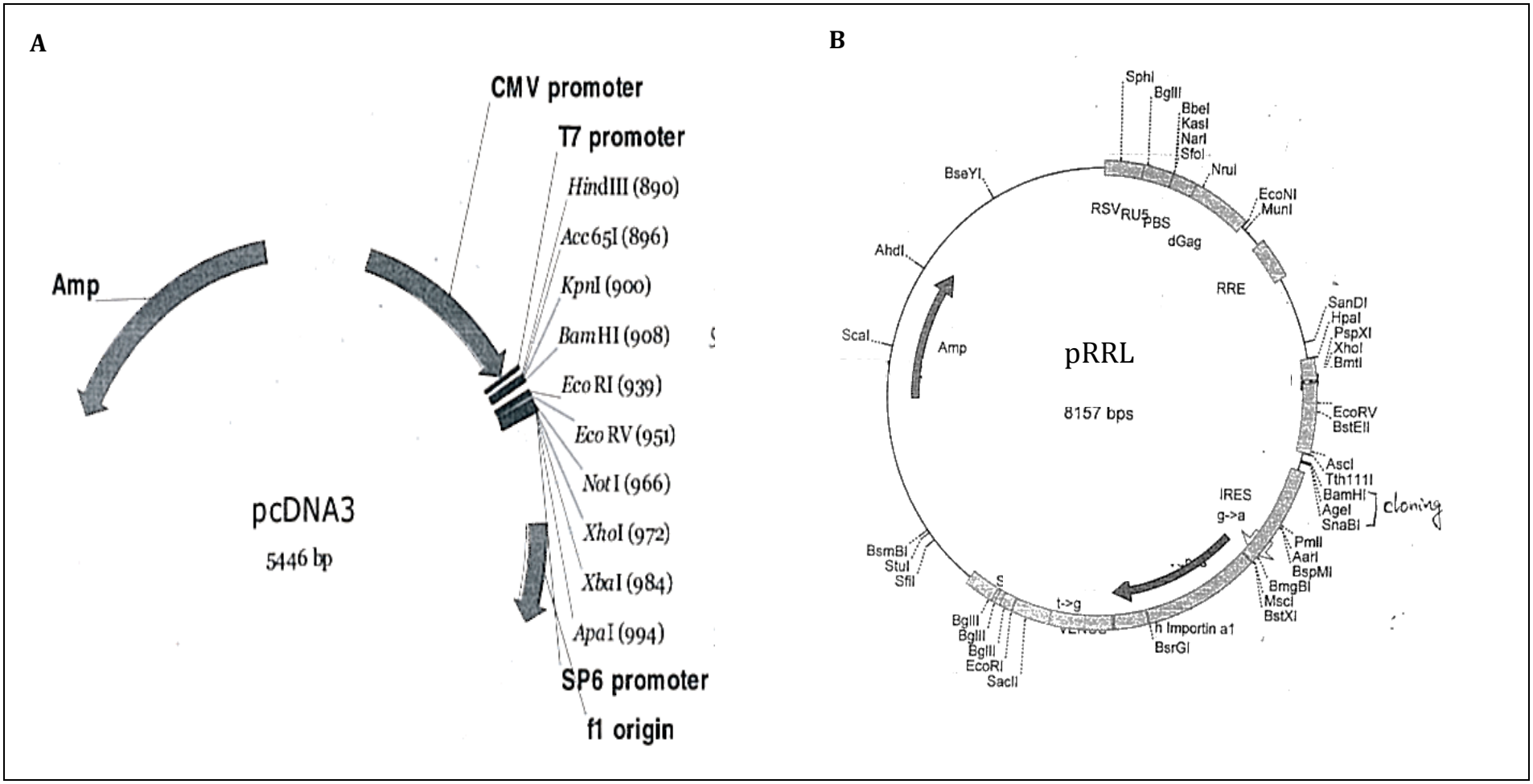


Figure 41: (A) Map of pcDNA3 and (B) Map of pRRL vectors

7.1.3.1 PCR reactions

A pcDNA3-PHD2 clone (*36.29, Ratcliffe group database, submitted by Dr Jonathan Gleadle*) was used as a "template" of the coding region of human PDH2. Primers were designed (*Primer3 software)* and a number of different PCR reactions (1 to 6, described below) were carried out such that PCR products of the coding region of EGLN1 (PHD2) were produced, which could subsequently be cloned into either pcDNA3 or pRRL. The two plasmids (pcDNA3 and pRRL) have different cloning sites. Thus, for cloning into pcDNA3, the forward primer was designed to have a tag sequence (5′ GGATCC 3′) recognised by restriction enzyme BamHI and the reverse primer was designed to have a tag sequence (5′ GAATTC 3′) recognised by EcoRI. For cloning into pRRL, the forward primer was the same as for cloning into pcDNA3 but the reverse primer was designed to have a tag sequence (5′ ACCGGT 3′) recognised by AgeI. A model PCR reaction is shown in Table 44 and the PCR thermal cycling conditions used in Table 45.

Table 44: PCR reaction used for cloning

Component	**Volume (µl)**
DNA template (~ 10 ng plasmid DNA)	2.5
Forward Primer (10 uM) *(Sigma)*	2.0
Reverse Primer (10 uM) *(Sigma*)	2.0
10× PFU Buffer with $MgSO_4$*(Promega)*	5.0
PFU Polymerase (2 unit/ul) (*Promega*)	0.5
dNTP mix (10 mM) (*Qiagen*)	4.0
DMSO (*Sigma*)	1.0
Nuclease-free water	33.0
Total	**50.0**

Table 45: Thermal cycling conditions used in the PCR reaction shown in Table 44

Steps	Temperature (°C)	Time (min)
1	97	10
2: Cycles (×30)	97 55 72	1 3 5
3	4	∞

PCR Reaction 1 was performed in order to produce a PCR product representing the coding region of wild-type PHD2. The primer pairs used were:

Forward (5′ to 3′): WT4_F: GCG-GGATCC-ATGGCCAATGACAGCGGCGGG

Reverse (5′ to 3′): WT_R_1: GCGGCG-ACCGGT-CTAGAAGACGTCTTTACCGACCG (for cloning into pRRL); and

WT_R_2: GCGGCG-GAATTC-CTAGAAGACGTCTTTACCGACCG (for cloning into pcDNA3)

PCR Reaction 2 was performed to produce a PCR product representing the coding region of PHD2 D4E, i.e. aims to introduce a point mutation (C>G) at codon 4. Thus a forward primer was designed that contains the mismatch of interest, MUT4_F, as shown below:

MUT4_F: GCG-GGATCC-ATGGCCAATGA**G**AGCGGCGGG (where C is substituted by G). The reverse primer used was either WT_R_1 or WT_R_2, as in ***PCR Reaction 1***. See Figure 42A.

To produce a PCR product representing the coding regions of PHD2 C127S, a two-step approach was used. First, two separate PCR reactions were set up in parallel, ***PCR Reaction 3*** and ***PCR Reaction 4.*** Each produced a PCR product that overlaps with the region in which the mutation is introduced. Thereafter, the PCR products of these two separate reactions were diluted 100-fold and mixed to set up ***PCR Reaction 5.*** Here, the

PCR products denatured and re-annealed, such that some of the single strands from ***PCR Reaction 3*** annealed to strands from ***PCR Reaction 4*** in the region where they overlap. See Figure 42B.

In ***PCR Reaction 3***, the primer pair used was:

Forward (5' to 3'): WT4_F

Reverse (5' to 3'): MUT127_R: CCGCACGA**G**ACGGCGACGC

In ***PCR Reaction 4***, the primer pair used was:

Forward (5' to 3'): MUT127_F: GCGTCGCCGT**C**TCGTGCGGCCGCCGGCGG

Reverse (5' to 3'): WT_R_1 (for cloning into pRRL)

WT_R_2 (for cloning into pcDNA3)

In ***PCR Reaction 5***, the primer pair used was:

Forward (5' to 3'): WT4_F

Reverse (5' to 3'): WT_R_1 (for cloning into pRRL)

WT_R_2 (for cloning into pcDNA3)

To produce a PCR product representing PHD2 D4E C127S, a similar approach was used as for PHD2 C127S, described above, and ***PCR Reaction 6*** was performed, with the forward primer WT4_F substituted by MUT4_F in all the relevant reactions.

The final PCR products were loaded on a gel (1% agarose in TBE buffer plus ethidium bromide to 2 μg/ml). Following electrophoresis, bands of the correct size were visualised using a UV transilluminator (*UV products*), excised from the gel using a scalpel and

purified using a column-based PCR purification kit (*Qiagen*). Figure 43 shows a typical gel image displaying bands corresponding to the products of the various PCR reactions.

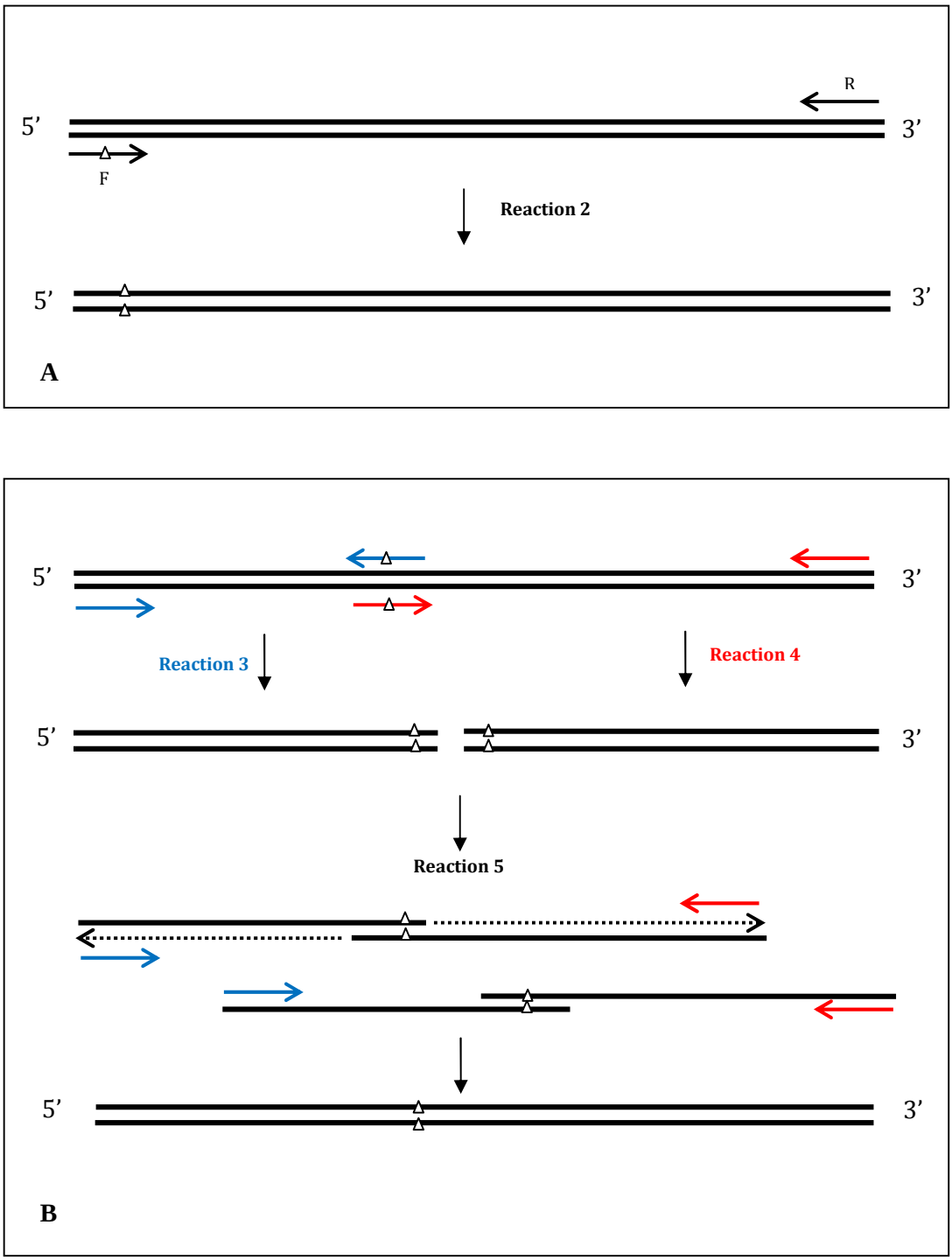


Figure 42: PCR reaction steps in the cloning methodology. (A) shows *PCR Reaction 2* and (B) shows *PCR reactions 3, 4* and *5*.

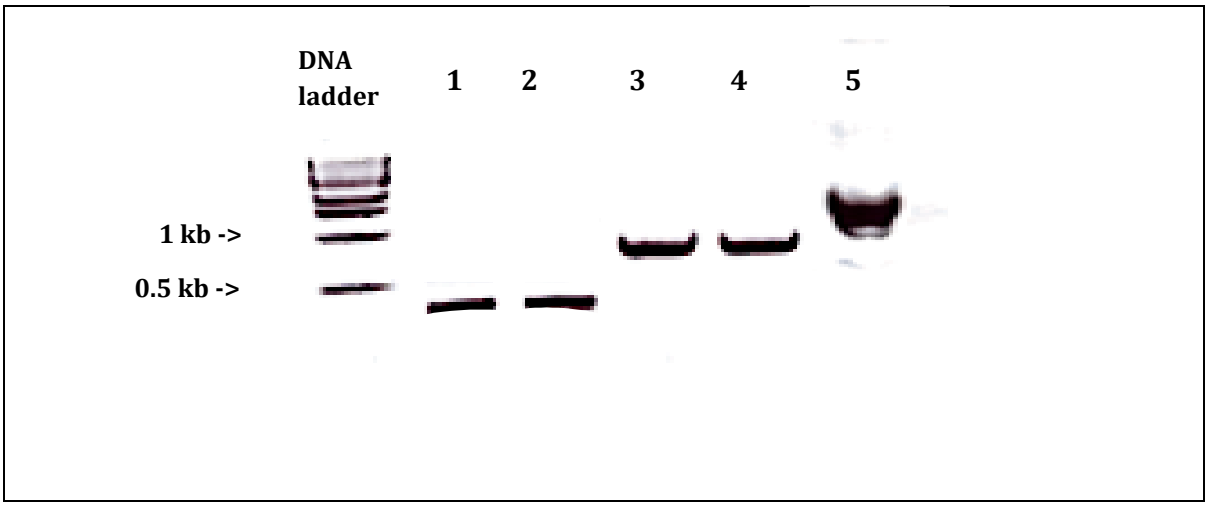


Figure 43: PCR products. The first lane represents a DNA marker ladder. 1 and 2 represents PCR products from *Reaction 3* (predicted size of 400 bp); 3 and 4 represent PCR products from *Reaction 4* (predicted size of 900 bp); 5 represents the PCR product from *Reaction 5* (predicted size of 1300 bp).

7.1.3.2 Restriction enzyme digestion of the PCR products and of the plasmid

The digestion reactions shown in Table 46 were set up and incubated at 37 °C overnight.

Table 46 : Restriction digestion for PCR products (left) and plasmid (right)

Component	**Volume (µl)**
DNA (PCR product) (~ 10 µg)	45.0
BamHI enzyme	2.5
AgeI or EcoRI enzyme	2.5
NEB Buffer 4 (x10)	7.0
BSA (x10)	7.0
Nuclease-free water	6.0
Total	70.0

Component	**Volume (µl)**
DNA (plasmid) (~ 2 µg)	3.75
BamHI enzyme	2.5
AgeI or EcoRI enzyme	2.5
NEB Buffer 4 (x10)	6.0
BSA (x10)	6.0
Nuclease-free water	39.25
Total	60.0

All reagents are from *New Engand Biolabs*

The digested PCR products were then purified (*Qiagen PCR purification kit).* Following the digestion, the plasmid vector was treated with 1 unit of calf intestinal alkaline phosphatase (*New Engand Biolabs*) for 25 min at 37 °C and then loaded onto a 1% agarose gel and subjected to electrophoresis. It was subsequently purified by cutting it out of the gel (*Qiagen gel extraction purification kit*).

7.1.3.3 Ligation reaction of the PCR product into the plasmid

The ligation reaction shown in Table 47 was set up and incubated at RT for 2 hours. The purified digested PCR product and plasmid concentrations were determined and an approximate molar ratio of insert to vector of 10:1 was used for the reaction. A control ligation reaction with plasmid alone (i.e. no insert) was also set up in parallel.

Table 47: Ligation reaction

Component	**Volume (µl)**
PCR product	6.0
Plasmid vector	1.0
T4 DNA Ligase (*Fermentas)*	1.0
Ligation Buffer Mix (x10) (*Fermentas*)	1.0
Nuclease-free water	1.0
Total	**10.0**

7.1.3.4 Transformation of bacterial cells

To transform bacterial cells to take up pcDNA3-PHD2 vectors, DH5α competent bacterial cells were thawed (from -80 °C) on ice and 2 µl of ligation reaction was mixed in with 50 µl bacterial suspension. Following incubation at 4 °C for 30 min, the bacterial cells were subjected to heatshock (42 °C for 2 min) and immediately put back on ice for 2 min. Five hundred µl of Luria-Bertani (LB) medium (*Sigma)* was added to each cell suspension and following incubation at 37 °C for one hour, 100 µl of the resulting culture was spread on a pre-warmed LB/ampicillin plate (LB containing 100 µg/ml ampicillin) and grown overnight in a 37 °C incubator. For pRRL-PHD2 vectors, XL10-Gold ultracompetent cells (*XL10-Gold, Agilent Technologies*) were used and processed as DH5α cells but were subjected to heat shock for only 30 sec. Colonies were picked ~ 16 hours later from each plate. The plates were then stored at 4 °C.

7.1.3.5 “Miniprep”: Rapid, small-scale isolation of plasmid DNA from bacteria.

Colonies were picked from freshly transformed bacterial plates. Each colony was placed in ~ 3 ml of LB/ampicillin medium and incubated at 37 °C overnight, with shaking at 200 rpm. One ml of the resulting culture was centrifuged at full speed for 1 min and the supernatant discarded. A “miniprep kit” (*Qiagen*) was used to isolate DNA from each colony, according to the manufacturer’s protocol as described in QIAprep Miniprep Handbook (*Qiagen*).

7.1.3.6 Miniprep “diagnostic” restriction digest

To confirm the integrity of the isolated DNA from each colony, a portion of each isolated DNA sample was analysed by restriction digest. For pcDNA3-PHD2 vectors, the restriction enzymes BamHI and SnaBI were used, whereas for pRRL-PHD2 vectors BamHI and XhaI were used. These digestion reactions are expected to give 3 fragments: the plasmid backbone and 2 PHD2 fragments (~ 600 and 700 bp). The digested products were subjected to agarose gel electrophoresis for diagnostic purposes – this is shown in Figure 44. “Positive” miniprep DNA samples were sequenced (*Source Bioscience, Oxford*) using primers T7 and SP6 for pcDNA3 and primers IRES_F and IRES_R for pRRL, in order to confirm the introduction of the “desired” mutations and the absence of any unwanted mis-matches in the sequence.

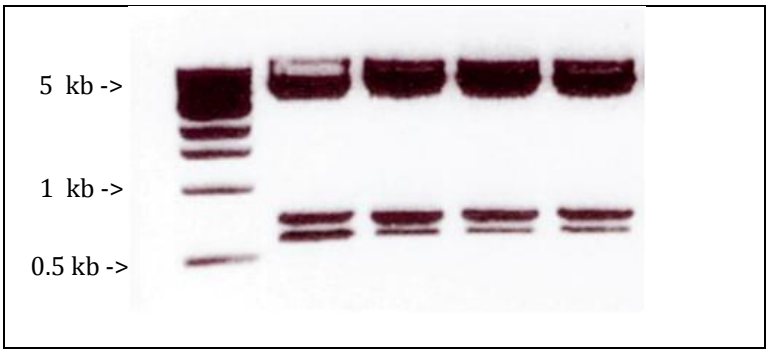


Figure 44: Miniprep "diagnostic" restriction digest. "Positive miniprep" DNA samples. The first lane represents a DNA marker ladder. In each of the other lanes, the large fragment represents the pcDNA3 plasmid backbone, while the other two bands represent two PHD2 fragments (600 and 700 bp).

7.1.3.7 "Midiprep": Intermediate-scale DNA isolation from bacteria

This procedure aims to increase the yield of purified plasmid-PHD2 DNA. Three ml of a miniprep culture was diluted into a larger volume (100 ml) of LB/ampicillin medium and incubated at 37 °C with shaking at 220 rpm overnight, to allow growth of the bacteria. Following this, a "midiprep kit" (Qiagen) was used to isolate DNA according to the manufacturer's protocol as described in QIAprep Midiprep handbook (*Qiagen*).

7.1.4 *In vitro* transcription and translation of PHD2 using pcDNA3-PHD2 vectors

Wild-type pcDNA3-PHD2 and pcDNA3-PHD2 D4E-C127S plasmids were linearised by enzyme restriction digestion using BamHI restriction enzyme, as described in section 7.1.3.2. DNA was subsequently extracted by phenol:choloroform and precipitated by ethanol. Briefly, an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) (*Sigma*), was added to the linearised plasmid, mixed and centrifuged at 13000 rpm for 3 min. The aqueous layer was retrieved, 1/10 volume of sodium acetate (3 M) and 5/2 volumes of 100% ethanol were added. After incubation on dry ice for 10 min, the mixture was centrifuged at 13000 rpm for 10 min and the resulting DNA pellet was re-suspended in 30 µl of nuclease-free water and quantified.

A TNT Coupled Wheat Germ Extract System (*Promega*) was used to carry out *in vitro* transcription and translation of: GAL-HIF1α peptide (aminoacids: 28 – 826), "wild-type" PHD2 and PHD2 D4E-C127S (this may be referred to as "double-mutant" PHD2 in the following sections for convenience). The reaction was set up (Table 48) and incubated at 30 °C for 2 hours.

Table 48: *In vitro* transcription-translation reaction

Component	**Volume (µl)**
TNT wheat germ extract	75.0
TNT reaction buffer	6.0
T7 RNA polymerase	3.0
Aminoacid mix (minus leucine)	1.5
Aminoacid mix (minus methionine)	1.5
RNase Inhibitor (20 units/µl)	3.0
Linearised DNA (total 3 µg)	3.0 (variable)
Nuclease-free water	54 (variable)
Total	**150.0**

This was immediately followed by an *in vitro* hydroxylation assay.

7.1.5 *In vitro* hydroxylation assay

The *in vitro* hydroxylation reactions (Table 49) were performed at 37 °C in euoxia or in hypoxia (1% O_2).

Table 49: *In vitro* hydroxylation reaction

Component	Volume (µl)	
WGE GAL-HIF1α (aa 28-826)	6.0	
Ascorbate (100 mM)	4.0	at a final concentration of 4 mM
2-oxoglutarate (2 mM)	8.0	at a final concentration of 160 µM
Fe^{2+} (4 mM)	2.0	at a final concentration of 80 µM
WGE PHD2 enzyme	40.0	
Hypotonic extraction buffer (HEB)*	40.0	
Total	100.0	

A control reaction was set up without cofactors (i.e. ascorbate and Fe^{2+}) and without co-substrate (i.e. 2-oxoglutarate). Note that the 4 mM ferrous solution was freshly prepared. This was done by diluting a $FeSO_4$ solution (500 mM) in water. The $FeSO_4$ solution (500 mM) was itself formed by dissolving 1.39g of $FeSO_4.7H_2O$ powder in 10 ml of 2 M HCl solution. The reactions were stopped at time intervals 5 min, 10 min, 20 min, 30 min, 40 min and 60 min, by retrieving 15 µl of reaction and mixing it with Laemmli buffer (6×) (which denatures the proteins) and placing it on ice. The resulting samples were heated at 95 °C for 3 min and then stored at -20 °C.

The samples were processed by immunoblotting as described in section 7.1.1 and total HIF1α, hydroxylated HIF1α (at proline 564) and total PHD2 protein levels in each sample were assayed. Mouse anti-HIF1α (*BD Transduction Laboratories*) at 1:1000 dilution, rabbit anti-Pro564-OH (*New England Biolabs*) at 1:1000 dilution and mouse anti-PHD2 (*clone 76a, Millipore*) at 1:100 dilution were used. The signal intensities of the

immunoblots were measured by densitometry and quantified using the AutoChemi System with Labworks 4.6 Image Acquisition and Analysis Software (*UVP Inc. USA*).

7.1.6 Analysis of results of the in-vitro hydroxylation assay

Five experimental repeats were performed in euoxia and five experimental repeats in hypoxia (1% O_2).

Following quantitation of the signal intensities by densitometry, the intensity of the band at 60 min (in each experimental repeat) was taken to represent 100% hydroxylation. The percentage hydroxylation at each time point was calculated and was corrected to account for any difference in PHD2 protein levels between "wild type" and "double mutant" in each paired experiment. This is because of the assumption that the rate of the reaction is dependent on the amount of PHD2 enzyme present. This correction was carried out based on the assumption that the following model applies:

$$[\mathrm{HIF1\alpha}_{\,\mathrm{P564-OH}}] = [\mathrm{HIF1\alpha_{total}}] - [\mathrm{HIF1\alpha_{total}}]\mathrm{e}^{-\mathrm{k}\,[\mathrm{PHD2}]\mathrm{t}}$$

For each paired experiment (i.e. hydroxylation reactions with "wild-type" and "double-mutant" PHD2), the % hydroxylation was plotted against time; the area under each curve was obtained using *Microsoft Excel*. Using a paired t-test (*SPSS 19*), the area under curve for the "wild-type" plot was compared to that of the "double-mutant" plot, to determine if there is a significant difference.

7.1.7 Infection of Phd-null MEF cells with pRRL-PHD2 viral vectors

7.1.7.1 Lentivirus production in 293T cells

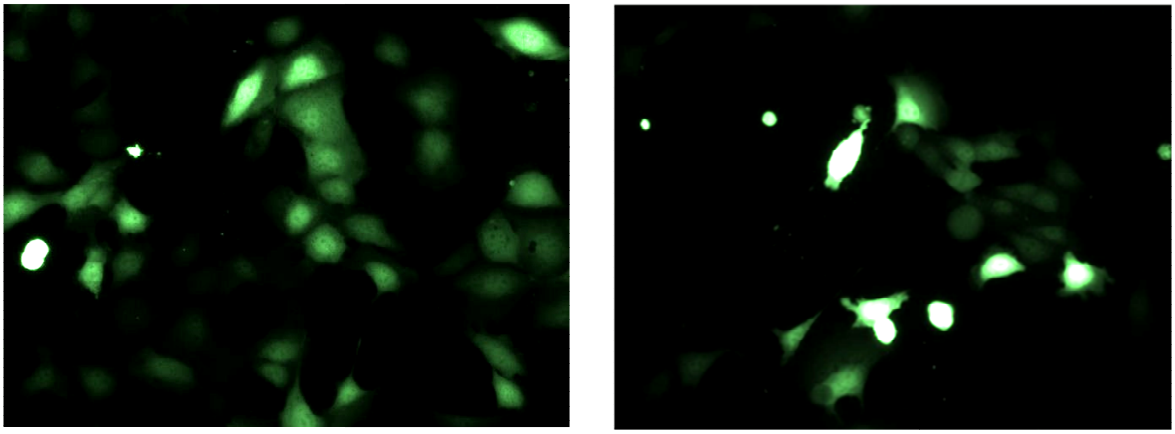
293T cells were cultured in a 10 cm dish in 8 ml of Dulbecco's Modified Eagle Medium (DMEM, *Sigma*) supplemented with 10 % FBS (*Sigma*) at a confluence of 40 to 70 % and were transfected using a transfection mix comprised of: OPTIMEM medium (*Gibco*), a set of lentiviral packaging and envelope plasmids – pCMVΔR8.91 and pCMV-VSV-G (8454) (*Addgene*) – polyethylenimine (PEI) (*Sigma Aldrich)* at 1mg/ml and pRRL vector ("empty vector" or wild-type pRRL-PHD2 or pRRL-PHD2 D4E C127S). The transfection mix was incubated at RT for 20 min and then added to the 293T cells. The virus-containing supernatant was harvested after culturing the cells for 48 hours at 37°C and filtered through a 0.45 μm filter (*Millex, Millipore*). Fresh virus was used to subsequently infect Phd-null MEF cells while aliquots were also frozen at -80 °C for future use.

7.1.7.2 Infection and analysis of Phd-null MEF cells

Phd-null immortalised MEF cells – also referred to as triple-knock-out (TKO) cells – lack all three types of Phd enzymes (1, 2 and 3). These cells were generated and validated (in terms of genotype and upregulation of HIF1α and HIF-target genes) by Dr Julie Adam [182]. These cells were grown at 40 to 50% confluence and infected with either empty pRRL, wild-type pRRL-PHD2 or pRRL-PHD2 D4E C127S PHD2 derived lentivirus, with the addition of polybrene (1 mg/ml). Three separate dilutions for each vector were set up with 1:1, 1:3 and 1:10 virus to medium volume ratios. The medium was changed 24 hours later. Fluorescent microscopy was performed at 48 hours to check the efficiency of the infection by visualising the expression of green fluorescent protein (GFP) in the cells (Figure 45). At this point, expression of PHD2 was also checked in all of the cell preparations by western blot. The cells infected with the concentrated virus (1:1 and 1:3

dilutions) were expanded by continuous culture, frozen down and stored in a liquid nitrogen tank. The cells infected with the viral preparation that was diluted 1:10 were expanded by continuous culture and were used for experiments.

The stably infected cells were cultured for 4 hours in either euoxia or hypoxia (1% O_2), with and without the addition of a proteasome inhibitor, MG132 (25 μM) (*Enzo Life Sciences*). Expression of HIF1α, hydroxylated (at the NODD proline residue) HIF1α, PHD2 and β-actin proteins was assayed by western blot, using rabbit anti-HIF1α NB100-479 (*Novus*) at 1:1000 dilution, rabbit anti-Pro564-OH (*New England Biolabs*) at 1:500 dilution, mouse monoclonal anti-PHD2 (clone 76a, *Millipore)* at 1:100 dilution and HRP-conjugated anti-β-actin *(Abcam)* at 1:5000 dilution, respectively.



A B

Figure 45: Expression of GFP protein in cells infected with (A) wild-type pRRL-PHD2 and (B) double-mutant pRRL-PHD2 derived virus.

7.2 RESULTS

7.2.1 Induction of HIF1α protein by hypoxia in EBV-transformed lymphoblastoid cell lines

***Main finding**: There is no obvious difference in the level of HIF1α protein induction by hypoxia between cell lines of different EGLN1 (PHD2) genotypes.*

The experimental procedure described in section 7.1.2 was performed a number of times, with a different combination of cell lines (shown in Table 50). Results from 3 separate experiments are shown in Figure 46. Repetition of these experiments showed that the degree of between-cell line and between-experimental repeat variability (even within the same cell line) was considerably large and outweighed any potential difference that might exist as a result of a different *EGLN1* genotype.

Table 50: Cell lines

Cell line	Genotype
WT1	wild-type PHD2 (Caucasian) rs18699510: CC, rs12097901: GG
WT2	wild-type PHD2 (Han Chinese) rs18699510: CC, rs12097901: GG
HET1	heterozygous PHD2 (Tibetan) rs18699510: CG, rs12097901: GC
HOM1	homozygous “double mutant” D4E C127S PHD2 (Tibetan) rs18699510: GG, rs12097901: CC
HOM2	homozygous “double mutant” D4E C127S PHD2 (Tibetan) rs18699510: GG, rs12097901: CC

In this section and the rest of this chapter the terms euoxia and normoxia are used interchangeably.

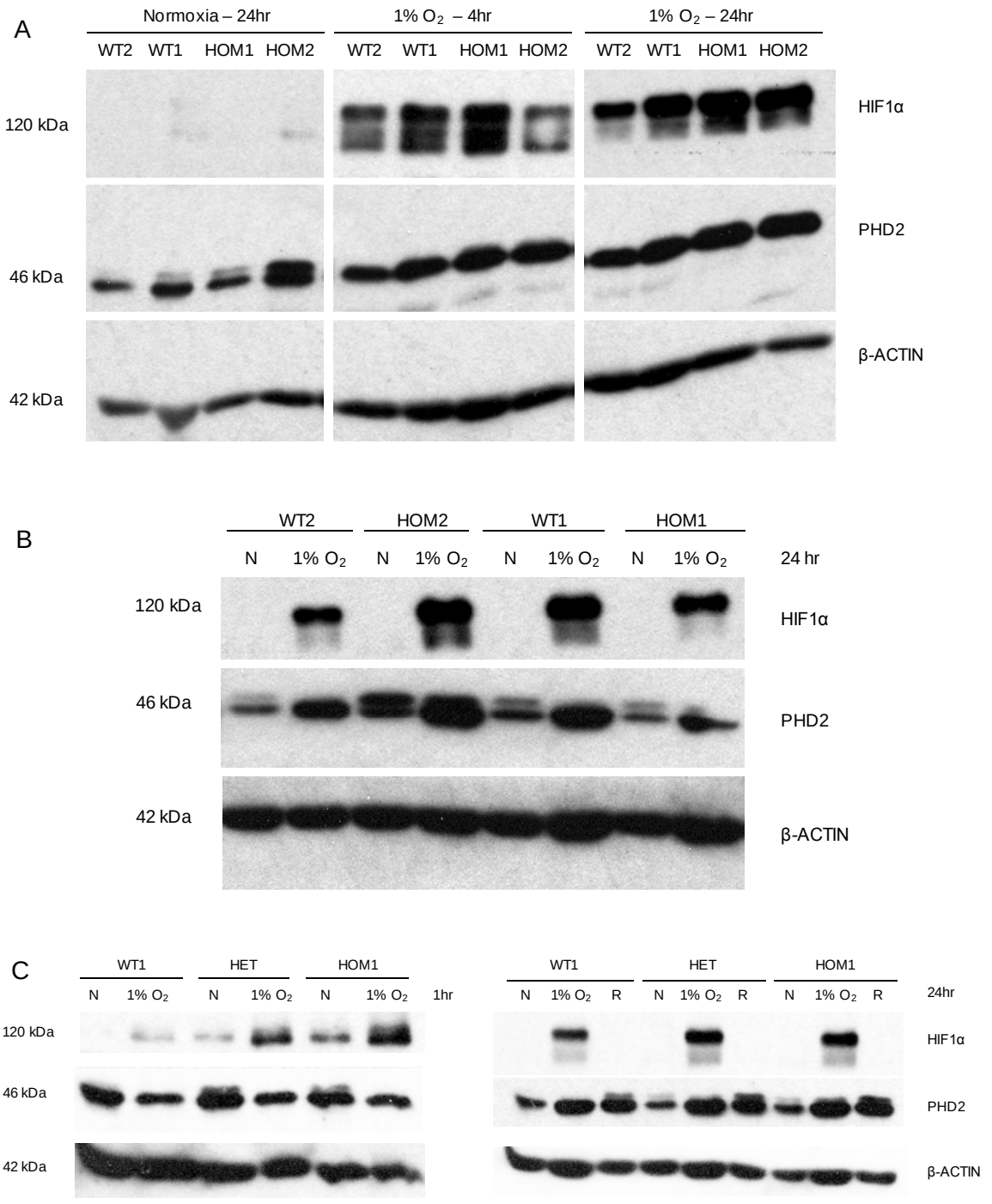


Figure 46: Immunoblots demonstrating HIF1α, PHD2 and β-actin protein levels in EBV-transformed lymphoblastoid cell lines of different genotypes cultured in normoxia (denoted also by N) or hypoxia (1% O_2) for different time durations. (A), (B) and (C) show results from 3 independent experiments (biological replicates). R signifies re-oxygenation (at 30 min).

Although not directly relevant to the aim of these experiments, an interesting observation was made here. There is evidence of 2 separate bands for PHD2 on the western blot in euoxia, with the lower band being in greater abundance. The ratio of these 2 bands seems to drastically change in hypoxia – as seen most clearly in Figure 46C – with the lower band being in abundance compared to the upper band. This phenomenon was recapitulated in another cell line, A547, a lung epithelium cancer cell line, as shown in Figure 47A. Using siRNA to PHD2 it was confirmed that the upper band indeed represents PHD2 and is not simply a non-specific band (Figure 47B). Interestingly, it was shown that this phenomenon is selectively observed in hypoxia but not in cells treated with dioxygenase inhibitors such as dimethyloxaloyglycine (DMOG) and the iron chelator desferioxamine (DFO). This suggests that it is oxygen-dependent but not HIF hydroxylase-dependent, as shown in Figure 48. The nature of this observation is not clear, and determining its origin is beyond the scope of this thesis. One hypothesis is that this is a transcriptional effect. For example, two PHD2 isoforms may exist in euoxia which may represent alternatively spliced variants and hypoxia may selectively induce one isoform (represented by the lower band) in preference to the other. An alternative hypothesis is that this phenomenon represents a post-translational modification of PHD2 such as phosphorylation (which would lead to a mobility shift) that is oxygen-dependent; for example, the upper band may represent phosphorylated PHD2 that it is predominantly present in euoxia but not hypoxia.

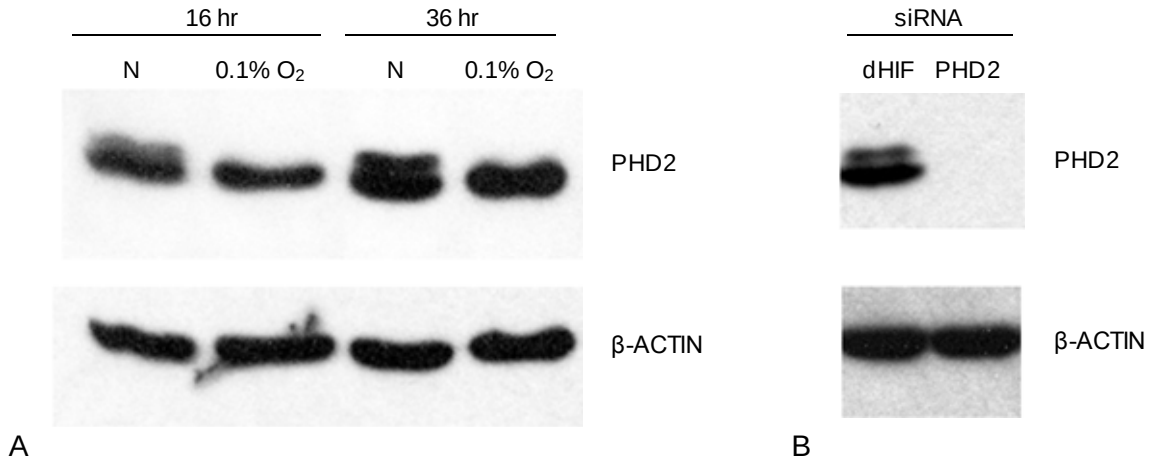


Figure 47: Immunoblots illustrating PHD2 and β-actin protein levels in a A549 cell line in (A) under normoxia (N) or hypoxia (either 0.1% O_2 or 1% O_2) and in (B) following treatment with dHIF (drosophila HIF siRNA as a control) or PHD2 siRNA.

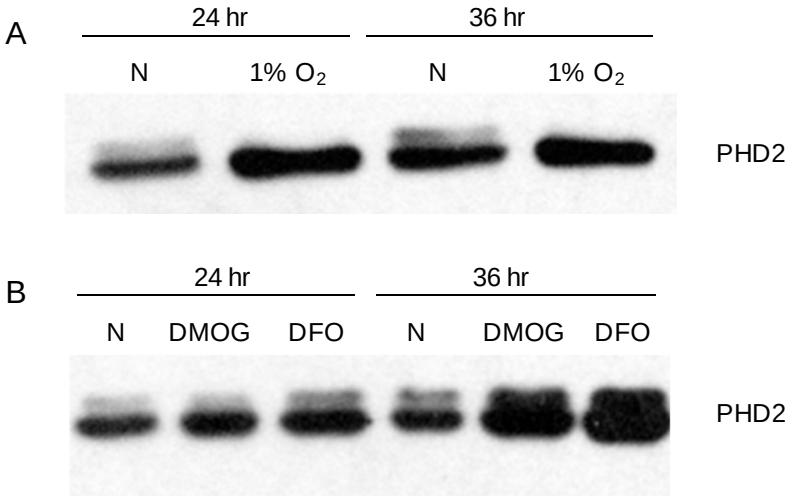


Figure 48: Immunoblot illustrating PHD2 protein levels in a A549 cell line under (A) normoxia (N) or hypoxia of 1 % O_2 and (B) euoxia or treatment with DMOG (1 mM) or DFO (100 µM).

7.2.2 Analysis of the DNA sequence of the constructed wild-type pcDNA3-PHD2 and pcDNA3-PHD2 D4E C127S.

Figure 49 displays the chromatographs confirming the sequence of each pcDNA3- PHD2 variant.

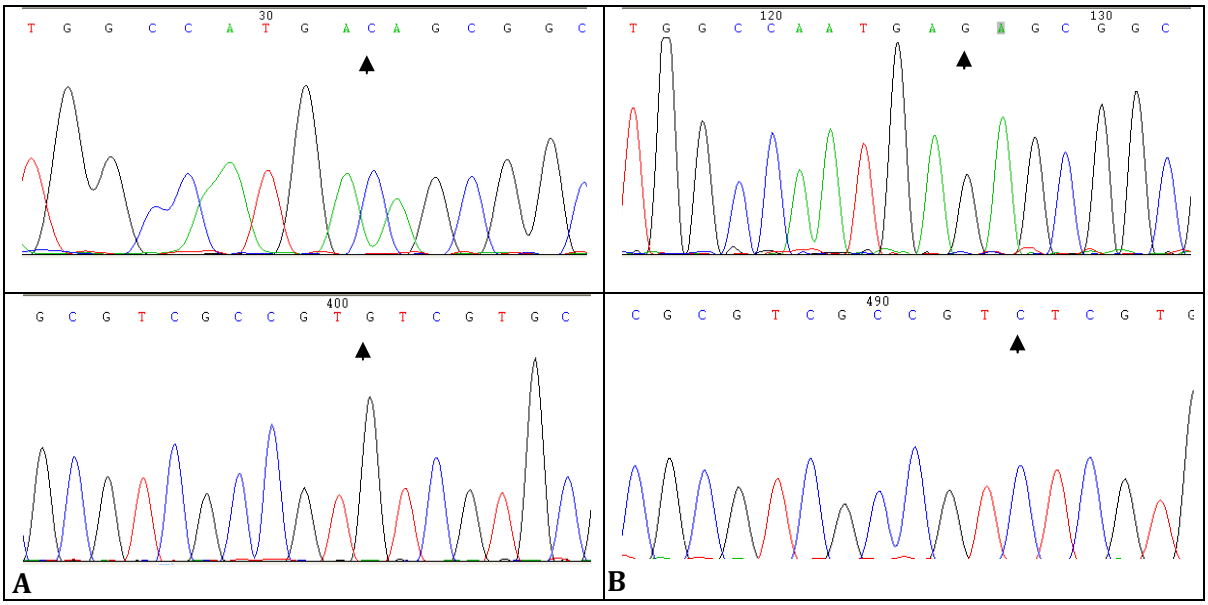


Figure 49: Chromatograph showing the sequence at codon 4 (top panels) and codon 127 (bottom panels) for A) wild-type pcDNA3- PHD2, (B) "double-mutant" D4E-C127S pcDNA3- PHD2. Arrows mark the base that is mutated.

7.2.3 *In vitro* hydroxylation activity of wild-type PHD2 and PHD2 D4E C127S

***Main findings*: I was able to show – by using a western blot readout – progressive hydroxylation of HIF1α for both the wild-type and the "double-mutant" PHD2 enzymes, in vitro. I was also able to show that this in vitro hydroxylation reaction is significantly slower in hypoxia than in euoxia. There is some indication that the "double mutant" PHD2 enzyme is more active than the wild-type at early time points in hypoxia but this was not statistically significant.**

Five experimental repeats were performed in euoxia and 5 experimental repeats were performed in hypoxia (1% O_2). Figure 50 shows protein immunoblots from 2 independent experiments in euoxia and Figure 51 shows protein immunoblots from 2 independent experiments in hypoxia. The immunoblots in hypoxia suggest that the "double-mutant" PHD2 has higher hydroxylation activity at early time points (see the 5 min time point in the long-exposure blot).

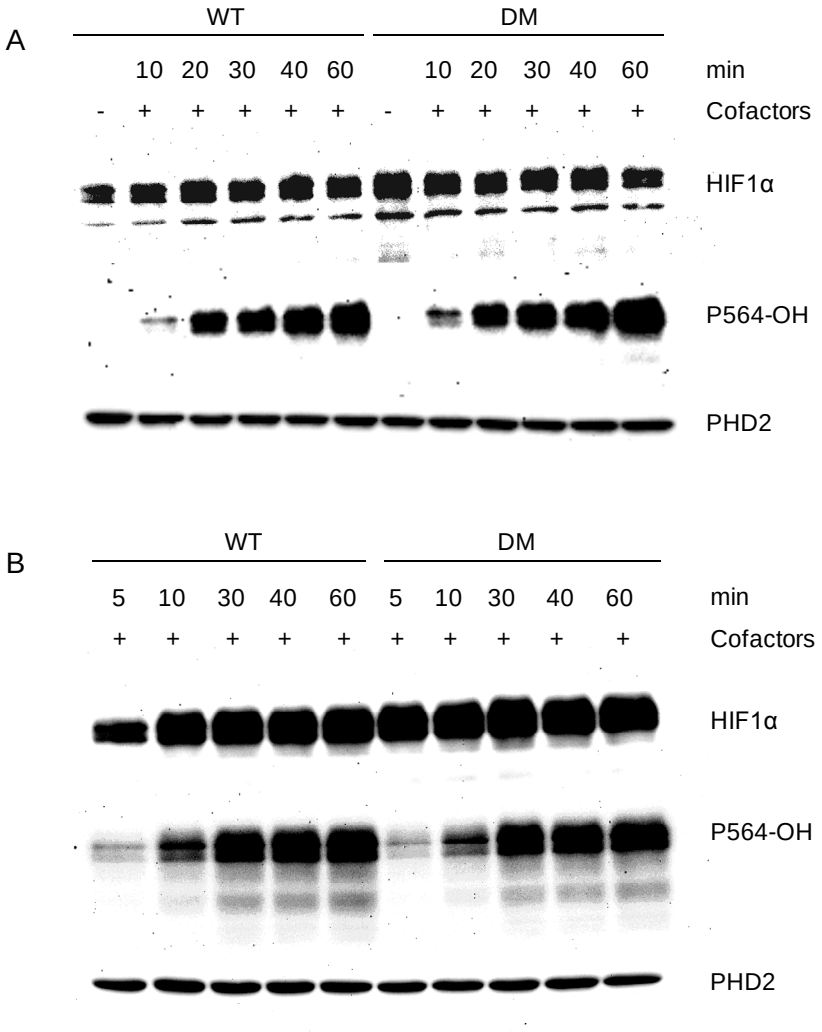


Figure 50: Immunoblot displaying the protein levels of HIF1α and of PHD2 and the degree of HIF1α hydroxylation at Pro^{564} (denoted by P564-OH) in an *in vitro* hydroxylation assay in euoxia. (A) and (B) represent two independent experimental repeats. WT represents wild-type and DM represents double-mutant PHD2.

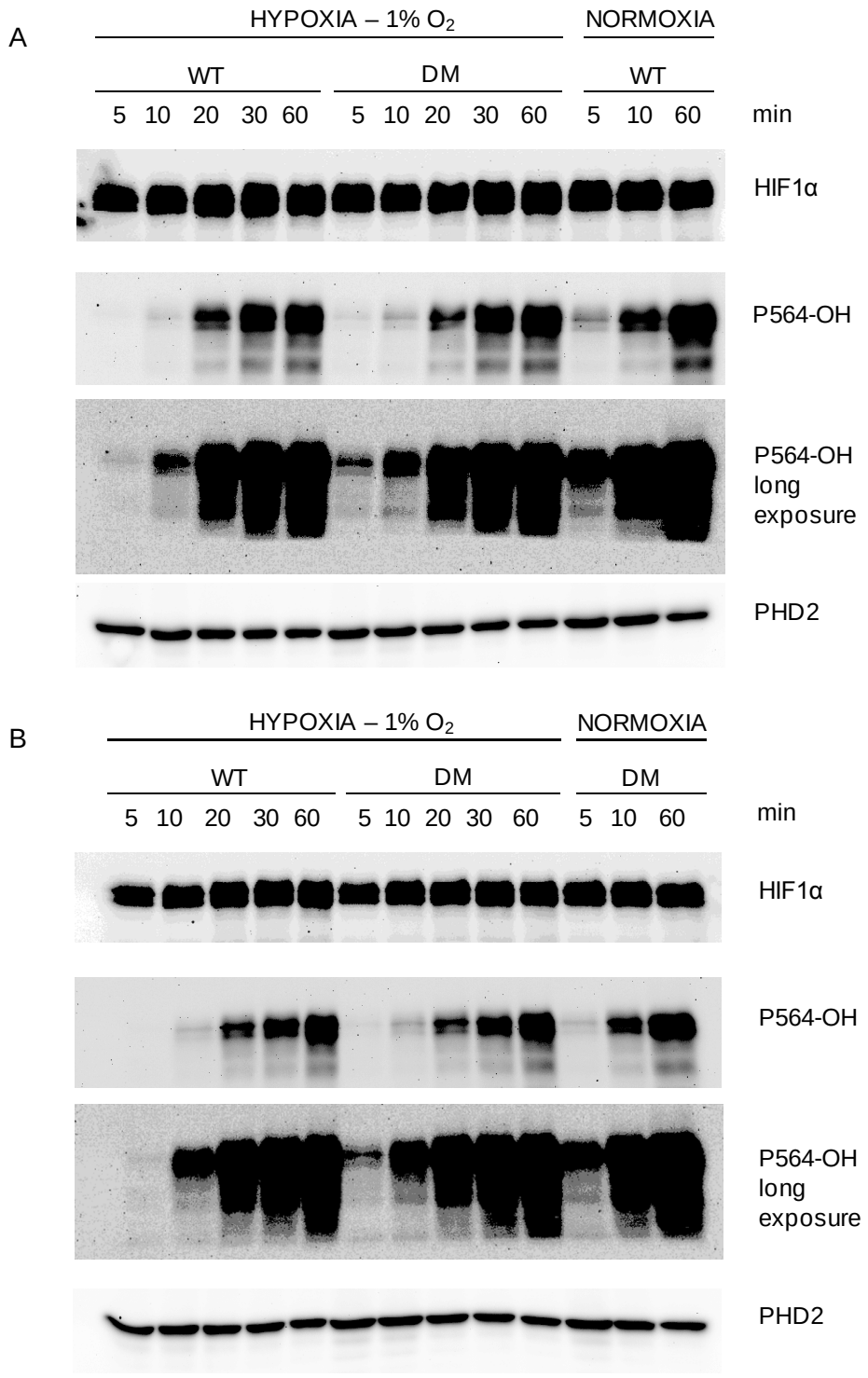


Figure 51: Immunoblot displaying the protein levels of HIF1α and of PHD2 and the degree of HIF1α hydroxylation at Pro^{564} (P564-OH) in an in-vitro hydroxylation assay in hypoxia. (A) and (B) represent two separate experimental repeats. WT represents wild-type, DM represents "double-mutant".

Figure 52 shows the mean percentage hydroxylation against time in for wild-type and "double-mutant" PHD2 in (A) euoxia and (B) hypoxia (1% O_2). Figure 53 shows a comparison between hydroxylation activity in euoxia and hypoxia for (A) wild-type PHD2 and (B) "double-mutant" PHD2.

Analysis of the data was undertaken as explained in section 7.1.6. It was found that there was no statistical difference between the hydroxylation activity of wild-type or "double-mutant" PHD2 in neither euoxia nor hypoxia (paired t-test, *SPSS 19*). The results of the analysis giving the area under curve for each experiment are shown in Table 51.

Table 51: Calculated area-under-curve for each experimental repeat

	Area under curve (Euoxia) (min)					**Area under curve (Hypoxia 1% O_2) (min)**				
Repeat	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
WT	3509	4050	4417	4908	4509	3750	3613	3477	3903	4400
DM	4790	4369	4374	4275	4564	3145	3871	4393	3912	3739

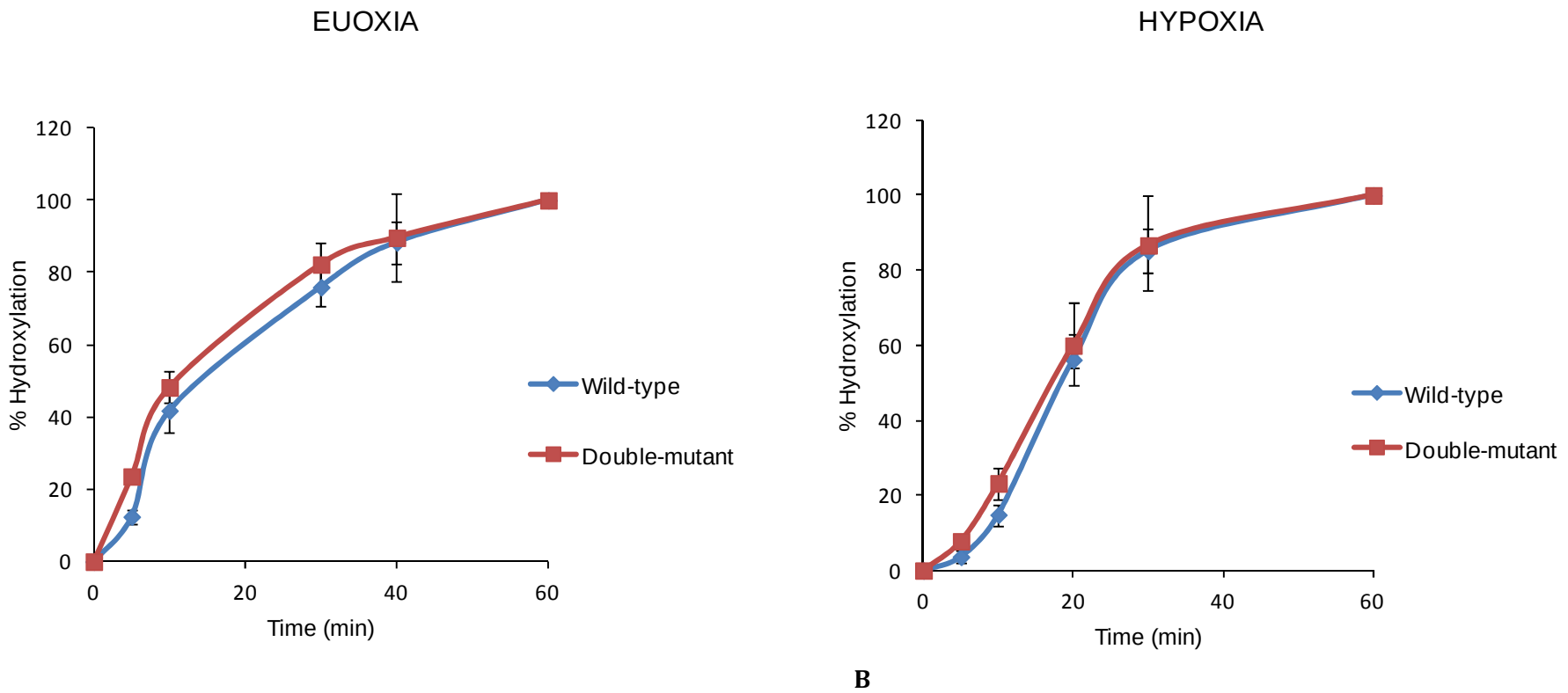


Figure 52: Percentage hydroxylation of HIF1α against time for wild-type and "double-mutant" PHD2 in the *in vitro* hydroxylation assay in (A) euoxia and (B) hypoxia (1% O_2).

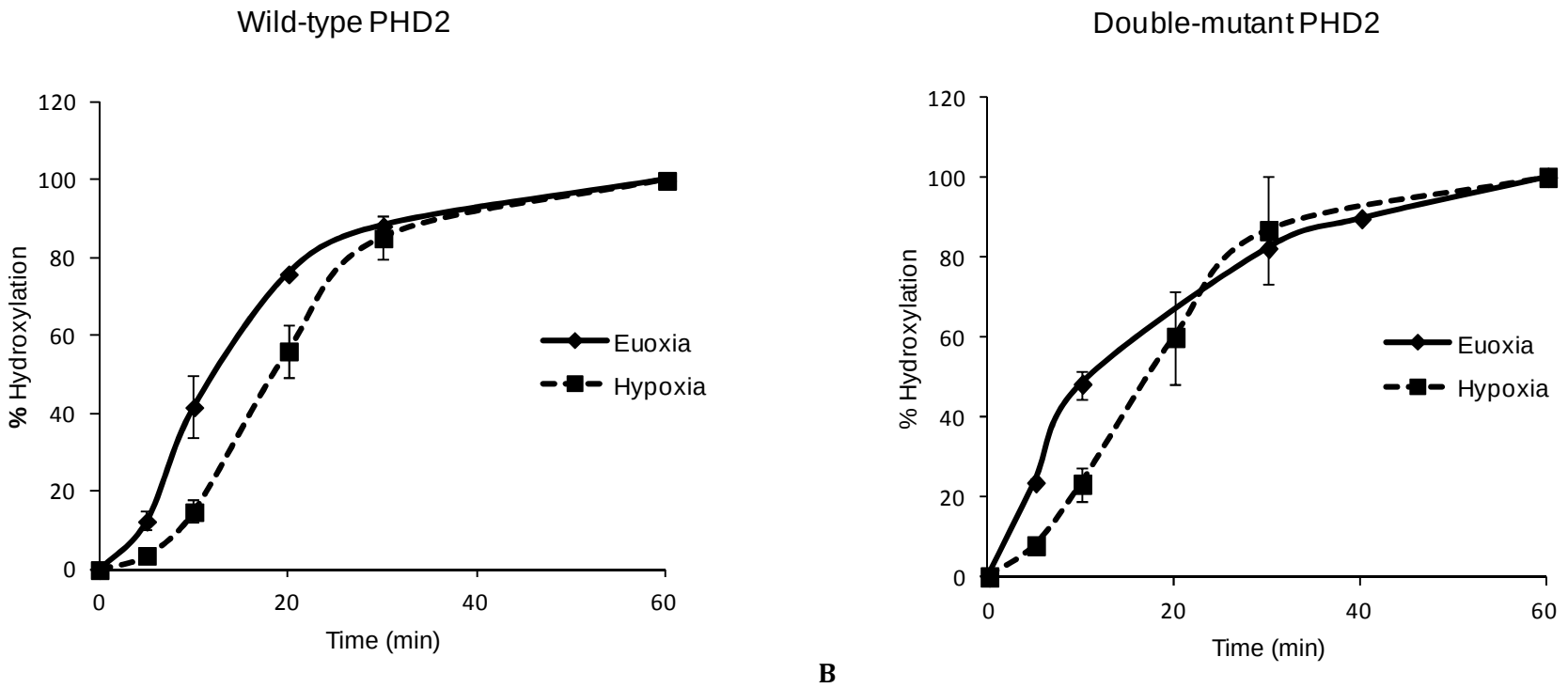


Figure 53: Percentage hydroxylation of HIF1α against time. (A) Wild-type PHD2 in euoxia and hypoxia (1% O_2); (B) "Double-mutant" PHD2 in euoxia and hypoxia (1% O_2).

For completion, I also investigated whether there is a difference between wild-type and "double-mutant" PHD2 in their ability to hydroxylate HIF1α at the Pro402 residue. The results of this are shown in Figure 54, which suggests that there is no clear difference between the two enzymes.

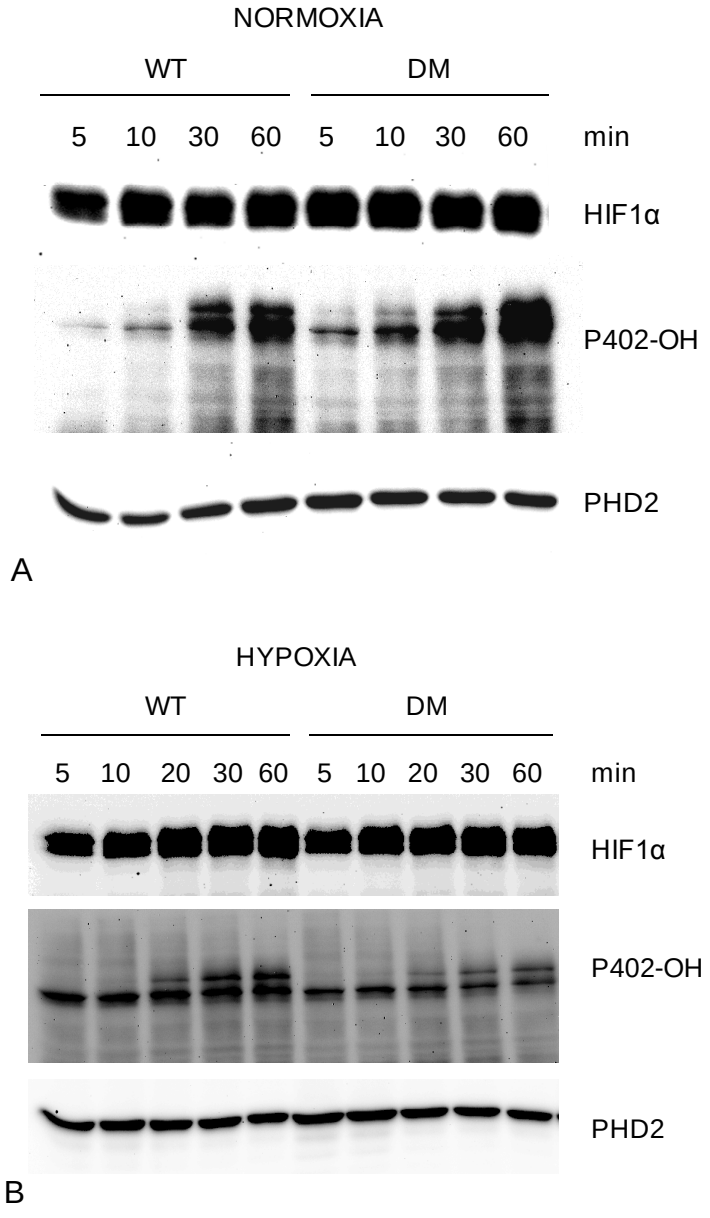


Figure 54: Immunoblot displaying the protein levels of HIF1α and of PHD2 and the degree of HIF1α hydroxylation at Pro^{402} (P402-OH) in an *in vitro* hydroxylation assay in (A) normoxia and (B) hypoxia, for wild-type (WT) and double mutant (DM) PHD2 enzymes.

7.2.4 Infection of Phd-null MEF cells with pRRL-PHD2 variants and assessment of HIF1α induction by hypoxia and hydroxylation

Figure 55 confirms the infection of Phd-null (TKO) cells with pRRL-PHD2 derived virus. There is evidence of over-expression of PHD2 in the cells in all 3 "viral dilutions", compared to the levels of PHD2 in other cell lines and compared to levels of β-actin in the TKO cells. Figure 56 confirms that both the wild-type and the "double-mutant" human PHD2 are active in the TKO MEF cells as there is significant reduction in euoxic levels of mouse HIF1α in the cells, and there is also induction of HIF1α upon exposure to 4 hours of hypoxia (1% O_2). Figure 57 illustrates HIF1α hydroxylation (NODD) in TKO cells infected with pRRL-PHD2 treated with MG132 in euoxia and in hypoxia. It appears that there is no difference in the level of HIF1α hydroxylation between wild-type and double-mutant PHD2 in the degree of HIF1α hydroxylation, in this experimental setting. This is discussed in the discussion section.

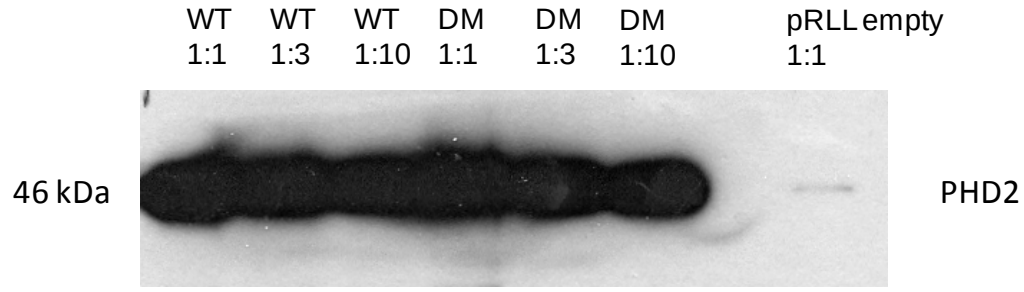


Figure 55: Immunoblot illustrating the expression of human PHD2 protein, wild-type (WT) or double mutant (DM), following infection of Phd-null MEF cells with pRRL-PHD2 viral vectors.

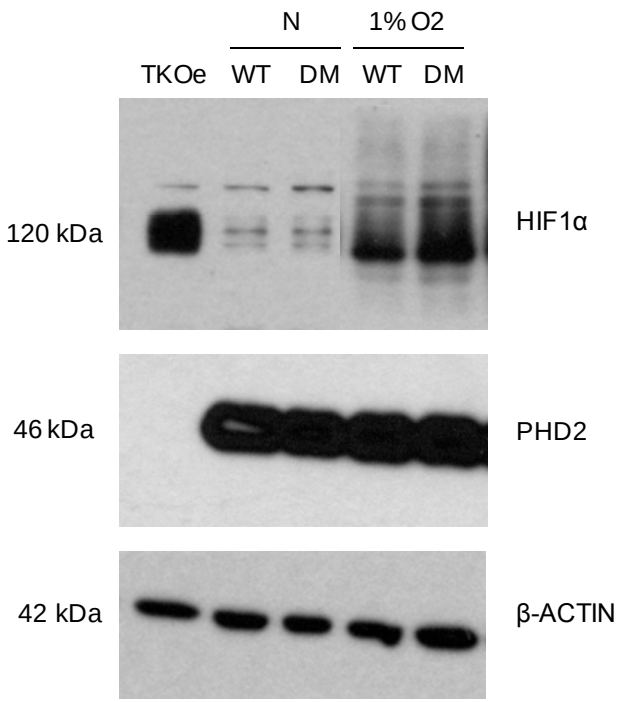


Figure 56: Immunoblot displaying HIF1α, PHD2 and β-actin protein levels in Phd-null MEF cells infected with "empty" pRRL vector (TKOe), wild-type pRRL-PHD2 (WT) and double-mutant pRRL- PHD2 (DM) derived virus cultured for 4 hours in conditions of normoxia (N) and hypoxia (1% O_2).

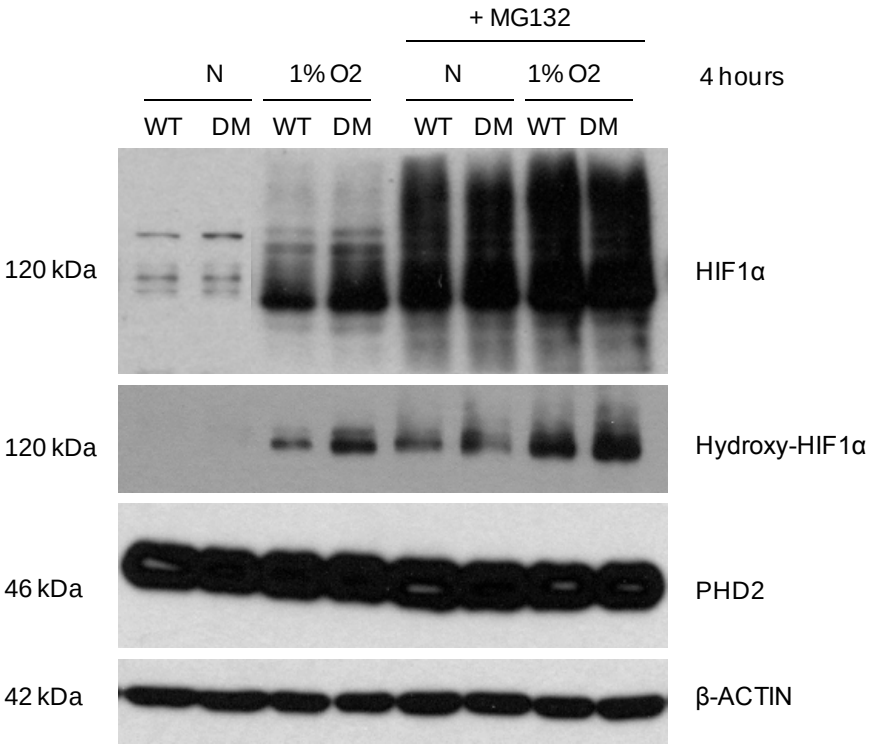


Figure 57: Immunoblot displaying HIF1α, hydroxylated HIF1α, PHD2 and β-actin protein levels in Phd-null MEF cells infected with wild-type pRRL-PHD2 (WT) and double-mutant pRRL-PHD2 (DM) derived virus cultured for 4 hours in conditions of normoxia (N) and hypoxia (1% O_2) with and without MG132 treatment.

7.3 DISCUSSION

From the experiments in lymphoblastoid cell lines of different genotypes (section 7.2.1), the first observation of note is that the degree of HIF1α protein stabilisation by hypoxia exhibits a great degree of variability between cell lines and also between experimental repeats (even within the same cell line). Thus, in this experimental system there is no evidence of a difference in the action of PHD2 enzymes of different genotype. Alternatively, if there is a difference, it is too subtle to be detected using this experimental setting. For example, as discussed in Chapter 5, there is a significant inter-cell line variability between lymphoblastoid cell lines related to factors such as transformation process, EBV-viral load, subculture conditions and pauciclonality. These factors may be exerting a greater effect on the "behaviour" of the cell line in hypoxia than any difference between activities of PHD2 enzymes of different genotypes and may account for the variability in the results I observed. In addition, HIF1α is a very labile protein with a short half-life in cells. As a consequence, assay-to-assay variability in the quantification of HIF1α protein by immunoblot can be considerable and may exceed a small difference occurring because of potentially different activities of the PHD2 variants. This assay-to-assay variability is indeed evident in my experimental repeats. Furthermore, there are complex mechanisms of feedback regulation when considering the HIF system in cells, and it may not be possible to capture differences in the activity of the PHD2 variants by assaying HIF1α levels at fixed timepoints ("snapshots").

In order to have a more direct experimental approach which is well-controlled and which avoids potentially confounding variation in cellular background, I then used an *in vitro* hydroxylation assay to compare the activities of wild-type PHD2 and PHD2 D4E C127S.

The first point I would like to discuss is the validation of this assay and of the hydroxy residue-specific antibody, in particular, in terms of sensitivity and specificity. The specificity of the antibody against the epitope that encompasses the hydroxyproline residue in HIF1α (P564-OH) was validated in [183]. This antibody showed strong reactivity to HIF1α produced in the presence of $FeCl_2$, but little or no reactivity to that produced in the presence of DFO, suggesting that it is highly specific for the hydroxylated form. It was also shown to be highly specific for HIF1α in comparison with HIF2α. In addition, substitution of the relevant proline residue abolished the reactivity of the antibody. In the same study, comparison with mass spectrometry analysis confirmed the high-degree of the hydroxyl-residue specificity of the antibody and demonstrated a complete absence of activity against the unhydroxylated epitope. Finally, the authors demonstrated a graded reduction in the hydroxylated form of H1F1α, quantified by densitometry, with graded hypoxia. Ladroue et al [87] used a similar *in vitro* hydroxylation assay to the one I used in this study and demonstrated a difference in the capacity of different PHD2 mutants – that are found in congenital polycythaemia, such as H374R and D254H – and wild-type PHD2 to hydroxylate HIF1α. Thus, the *in vitro* hydroxylation assay and the detection of hydroxylation by immunoblotting appear specific and sensitive to allow comparison of activities of different PHD2 enzymes.

Using this *in vitro* hydroxylation assay, I was able to demonstrate a progressive increase in hydroxylation of HIF1α with time in euoxia, for both PHD2 enzyme variants (wild-type and "double-mutant"), as shown in the immunoblots in Figure 50. In addition, I demonstrated that in hypoxia (1% O_2) there is also progressive increase in HIF1α hydroxylation with time (Figure 51), but this occurs at a much slower rate, for both the wild-type and the mutant PHD2 enzyme (Figure 53). This is in keeping with earlier studies

which have used VHL-capture assays that demonstrated that PHD enzymes are highly sensitive to *in vitro* hypoxia [46].

This experimental approach aimed to investigate whether PHD2 D4E C127S has a different hydroxylation activity than wild-type PHD2, in euoxia and hypoxia. If indeed the two coding variant leading to D4E and C127S substitutions affected PHD2 enzymatic function then one might predict PHD2 D4E C127S to have a higher enzymatic activity than the wild-type PHD2. This would be in keeping with the evidence of a hyporesponsive HIF system at both the integrative physiology level and at the peripheral blood lymphocyte level in Tibetans and also the genotype-phenotype association in terms of erythropoietin induction by hypoxia in the integrative physiology study in Chapter 3.

In euoxia, I could not detect any difference in activity between the two enzymes. However, in hypoxia there was an indication of increased activity in the mutant form of PHD2 at early time points, although this was not statistically significant in the subsequent analysis. One question that arises here, is why might the hydroxylation activities of the two PHD2 variants in this *in vitro* setting be different from one another in hypoxia, but not in euoxia. One potential explanation might lie in the affinity of PHD2 for oxygen. The Km value for O_2 for the PHDs has been initially reported to be between 230 and 250 μM, which is higher than other dioxygenases including FIH (90 μM) [184] [185]. It was later reported that it may be much lower and was found to depend on the length of the HIF1α peptide used in the reaction [186]. In any case, the Km value for the PHDs is higher than the physiological oxygen concentrations in the body (reports refer to 10 to 30 μM, and alveolar oxygen concentration is ~ 100 μM), making the activity of PHDs extremely sensitive at low oxygen concentrations (such as 1% O_2 *in vitro* or the physiological range of oxygen

concentration found in cells). Furthermore, compared to other studied 2-oxoglutarate dioxygenases, PHD2 was shown to react relatively slowly with oxygen, in a combined kinetic analyses of purified PHD2 and its preferred HIF-CODD substrate [187]; this property may be critical for its presumed function as an oxygen sensor. It is possible that the "double-mutant" PHD2 varies in its Km value for O_2 compared to the wild-type, and this affects the rate of hydroxylation in hypoxia.

In any case, however, the conclusion that can be drawn from the results using this experimental system is that there is not strong evidence that the two coding variants significantly affect the hydroxylase activity of PHD2 on HIF1α, at least *in vitro. In vitro* assays, such the one I used in this study, have their own caveats and there is always a danger of correlating findings *in vitro* with the *in vivo* situation. In cells, the kinetics of the PHD2 enzyme and the relative concentrations of substrate, co-substrate and co-factors may be substantially different than in this *in vitro* assay. Thus, it is still possible that the coding mutations may affect PHD2 hydroxylase activity *in vivo;* or they may be "functional" by other means and not through affecting hydroxylase activity. Or, they may not be functional at all, but are rather simply in linkage with an "unknown" variant at the *EGLN1* locus, which may itself be functional in the genetic adaptation in Tibetans.

The aim of using the last experimental approach, i.e. infection of Phd-null MEF cells with either of the two pRRL-PHD2 variants, was to have an experimental setting where: 1) There is a uniform cellular background that avoids the heterogeneity associated with different cell lines and 2) it allows examining the effects of PHD2 in isolation by avoiding interference by the other PHDs (PHD1 and PHD2) that are normally found in cells. This approach, however, has a number of methodological caveats. First, there is massive over-

expression of PHD2 in the cells, which is supra-physiological and likely to mask any small differences in activity between the two PHD2 enzymes, at least in euoxia. Second, while there was massive over-expression of PHD2, there was not 100% infection efficiency in terms of the percentage of cells infected, but approximately 80% (as evaluated by visualising the expression of GFP under fluorescent microscopy). A small difference in the percentage of cells effectively infected with wild-type compared with "double-mutant" pRRL-PHD2 could not only affect the result, but could even become bigger with subsequent subculture of the cells. Third, the anti-hydroxy Pro564 antibody does not seem to be specific enough for the hydroxylated form of mouse HIF1α (as it was for human); rather there is cross-reactivity and the antibody also detects the un-hydroxylated HIF too (Figure 57). Thus, this system does not appear to have the sensitivity to detect small differences in PHD2 action on HIF.

Aside from the above results, I would like to discuss a few key points regarding the sequence, structure and function of the PHD2 enzyme, and try and relate these to the two "Tibetan" coding variants, the role of which I investigated in my experiments.

The first point to mention is that crystallisation of the full-length PHD2 has been unsuccessful to date, but truncated forms of PHD2 including the catalytic domain have been crystallised [188]. Secondary structure prediction shows that PHD2 contains two structural domains: the N-terminal domain (aminoacids 21-58), which has homology to the MYND (myeloid, Nervy, and DEAF-1) zinc finger domains, and the catalytic C-terminal domain (aminoacids 181-426), which has homology to the 2-OG dioxygenase and which includes the 2-OG and Fe^{2+} binding sites. Of note is that the mutations associated with congenital erythrocytosis that have been reported thus far in the literature are found in the

catalytic domain, and in those cases this can confidently explain why they are functional. On the contrary, the "Tibetan" coding variants that I tested here involve aminoacids 4 and 127. These parts of the protein have not been crystallised and thus it is uncertain what their contribution to protein function is. However, one interesting observation is that while the three human PHDs have different substrate specificity (both in terms of HIF1α, HIF2α and oxygen, but also in terms of specific proline residues) the catalytic domain is highly conserved between them, suggesting that these specificity differences may be determined by regions remote from the catalytic domain and may involve the variable N- and C-terminal regions. Thus, equally, it may be that the regions in which the two coding variants I am testing lie may have a significant role to play in the structure and function of the PHD2 protein.

Interestingly, the sequence of *EGLN1* amongst species is conserved at codon 4, but not at codon 127. In addition, the variation at codon 4 is found at high frequencies in Tibetans but is nearly absent in other populations whereas the variation at codon 127 is found at high frequencies in Tibetans, moderate frequencies in Han Chinese and very low frequencies in Caucasian. This might suggest that the variation at codon 4 is more likely to affect protein function. However, while mutations in genes in conserved regions are likely to cause disease, it may be that mutations in genes in non-conserved regions – e.g. codon 127 – might be responsible for functional human variation.

It is important to bear in mind that variation in *EGLN1* may be functional by means other than producing differences in the hydroxylase activity of PHD2 on HIF1α. PHD2 substrates other than HIF have been proposed. One example is RNA Polymerase 2. Both PHD1 and PHD2 were shown to regulate Rpb1, the large subunit of RNA Polymerase 2 in

response to oxidative stress [189, 190]. The regulation of HIF and of PHD enzymes *in vivo* are complex, as mentioned in Chapter 1, and thus the mechanisms by which genetic variants of PHD2 may be producing functional phenotypic effects may not be captured by simply assessing PHD2 hydroxylation activity.

Chapter 8: Andean high-altitude populations – Chronic mountain sickness

The Tibetan plateau was first populated approximately 25000 years ago whereas the Andean Altiplano approximately 11000 years ago [191, 192]. The different physiological phenotypes in Tibetan versus Andeans highlanders have been extensively discussed in Chapter 2. To what extent underlying genetic variation is responsible for the observed phenotypic variation is not yet known. One hypothesis is that different genes have evolved in the two populations producing different physiological phenotypes, as in the case of malaria where multiple genes confer resistance in different populations [193, 194]. Another hypothesis is that the same candidate genes have evolved but with different functional changes thus producing different physiological phenotypes; or, there may be some common genetic variation that has evolved independently in the two populations, as in the case of lactase [195]. A study by Bigham et al [120] has shed some light into this. They performed genome scans aiming to identify gene regions that may be involved in the adaptation to high altitude in both Tibetans and Andeans. Their results suggest that gene regions and candidate genes that were identified as having undergone positive selection in Andeans and Tibetans are largely distinct from one another. However, they show that *EGLN1* shows evidence of positive directional selection in both populations, but the pattern differs for each population, with different allelic frequencies of SNPs in the extended regions of *EGLN1* and different dominant haplotypes in each. In their study, rs1765811 is one the six most significant SNPs in *EGLN1* in Andeans but does not feature at all amongst the significant SNPs in Tibetans.

In this chapter, a small genetic analysis is undertaken attempting to correlate genetic variation in *EGLN1* with the phenotype of chronic mountain sickness (CMS) in cohorts from Andean high-altitude natives. This work is collaborative between me and Dr McCormack and Dr Cavalleri from the Royal College of Surgeons in Ireland. The samples were collected from 3 different expeditions in the Peruvian Andes and were provided to us for genotyping. There are 3 cohorts.

Cohort 1 consists of 100 samples from mountain Cerro de Pasco, Peru: 50 CMS patients and 50 healthy volunteers. These samples were all from male volunteers. The patients with CMS were diagnosed according to standard diagnostic criteria: excessive erythrocytosis (haemoglobin $\geq$ 21 g/dl) occurring in high-altitude residents of more than 2500 m above sea level, hypoxaemia, and absence of chronic pulmonary disease or concomitant medical conditions that worsen hypoxaemia [196].

Cohort 2 consists of 89 samples from mountain Puno, Peru: 29 CMS patients and 60 healthy volunteers. The same criteria were applied to diagnose patients with CMS, but this cohort included women, in whom Hb $\geq$ 19 g/dl is required to make the diagnosis. Personal communication with Dr William Chuckley gives an estimated CMS prevalence in Puno of 5 % (from a sample size of 901).

Cohort 3 consists of 130 samples: 41 CMS patients and 89 healthy volunteers (a mixture of Aymara and Quechua natives).

All three cohorts were genotyped by Dr Mark McCormack for SNP rs1765811 and the analysis of the data presented here was performed by me. The aim was to investigate whether there was correlation between genotype at this SNP and CMS.

The first 2 cohorts were also genotyped for the coding variant rs12097901 in *EGLN1* (that causes C127S); this genotyping and the data analysis were performed by me. The rationale for genotyping the Andean samples for this coding variant is that: 1) it is found in high frequencies in Tibetans, who have low prevalence of chronic mountain sickness. 2) There is a suggestion that it correlates with haemoglobin at high altitude, and I also showed a correlation of this variant with erythropoietic response to sustained hypoxia in Tibetans in the physiology study. 3) It is found at moderate frequencies in Han Chinese, and thus it is likely that it probably has high enough minor allele frequency (MAF) to be present in other populations around the world, too. Thus, I wanted to investigate: 1) What the frequency of this coding SNP is in Andeans and 2) Whether there is any correlation of this SNP with CMS phenotype.

8.1 Rs1765811

Table 52 and Table 53 show the number of individuals for each genotype and the allele frequencies, respectively, for rs1765811.

Table 52: SNP rs1765811

	CMS Patients				**Healthy Volunteers**			
Cohort	**AA (N)**	**AG (N)**	**GG (N)**	**Total (N)**	**AA (N)**	**AG (N)**	**GG (N)**	**Total (N)**
1	17	11	22	50	26	19	5	50
2	12	12	5	29	26	27	7	60
3	7	17	17	41	11	44	34	89

Table 53: Allele Counts and allele frequencies for SNP rs1765811

	CMS Patients		Healthy Volunteers		Total	
Cohort	**Allele A**	**Allele G**	**Allele A**	**Allele G**	**Allele A**	**Allele G**
1	45 (0.45)	55 (0.55)	71 (0.71)	29 (0.29)	116 (0.58)	84 (0.42)
2	36 (0.62)	22 (0.38)	79 (0.66)	41 (0.34)	115 (0.65)	63 (0.35)
3	31 (0.38)	51 (0.62)	66 (0.37)	112 (0.63)	97 (0.37)	163 (0.63)

The first observation of note is that the frequency of alleles "A" and "G" (for rs1765811) are significantly different between the Andean cohorts here (Table 53) and the Tibetan cohort in my study (Table 16) in which the frequency for allele "G" is 0.98 and for allele "A" is 0.02. The second observation of note is that the three different cohorts vary between them in the allelic frequencies for rs1765811.

Statistical analysis for each Andean cohort was performed separately. This was done using a chi-square test asking the question whether the percentage of allele "A" or allele "G" is significantly different between CMS patients and healthy volunteers. For Cohort 1, $\chi^2 = 13.88$, Df = 1, $p < 0.001$, suggesting that the allelic frequency differences between CMS patients and healthy controls are significant, with allele "G" found at higher frequency in the CMS patients. For Cohorts 2 and 3, which were analysed separately, there was no statistical difference between CMS patients and healthy controls in the allelic frequencies.

The discrepancies between the cohorts may be explained by:

1. Different cohorts come from different regions and represent subpopulations of Andeans with possible variations in their genetic "make-up". It is thus possible that this SNP is linked to an un-identified "causal" or "protective" variant in *EGLN1* in the subpopulation of Cohort 1.

2. The selection of patients between cohorts may vary in an operator-dependent manner.

3. The result in cohort 1 is may be spurious.

At this point I tested whether the genotypes obtained in each cohort deviate from the Hardy-Weinberg equilibrium. This states that both allele and genotype frequencies in a population remain constant – i.e. are in equilibrium – from generation to generation. This was tested using a χ^2 goodness-of-fit test. For cohorts 2 and 3 there was no deviation from the Hardy-Weinberg equilibrium. For Cohort 1, there was, as shown and explained below:

Table 54: All participants in Cohort 1- Genotypes

Cohort 1	AA	AG	GG	TOTAL
OBSERVED	43	30	27	100
EXPECTED	33.64	48.72	17.64	100

The frequency of allele A is given by $\mathrm{p(A)} = 0.58$. If the Hardy-Weinberg equilibrium holds, then the frequency of allele G, $\mathrm{p(G)} = 1 - \mathrm{p(A)}$, i.e. $\mathrm{p(G)} = 0.42$. Hence expected genotype numbers are calculated by:

$\mathrm{N\ (AA)} = 100 \times \mathrm{p(A)}^2$, i.e. $\mathrm{N\ (AA)} = 33.64$

$\mathrm{N\ (AG)} = 100 \times 2\ \times\ \mathrm{p(A)}\ \times\ \mathrm{p(G)}$, i.e. $\mathrm{N\ (AG)} = 48.72$

$\mathrm{N\ (GG)} = 100 \times \mathrm{p(G)}^2$, i.e. $\mathrm{N\ (GG)} = 17.64$

The χ^2 test was calculated at 14.76 with 2 degrees of freedom and p value < 0.001, suggesting that there is significant deviation from the Hardy-Weinberg equilibrium. This deviation might have occurred because of a genotyping error. However, when the CMS individuals are considered separately from the healthy individuals, it was found that there was deviation from the Hardy-Weinberg equilibrium only in the CMS cases but not in the

healthy controls, as shown in Table 55 and Table 56, and thus it is less likely to constitute a genotyping error as all the samples from Cohort 1 were genotyped together. It still remains possible that this SNP is linked to an un-identified "causal" disease variant in the CMS patients in this cohort or that this result has arisen spuriously e.g. small sample size, relatedness between individuals or by chance.

Table 55: Cohort 1 (Healthy controls) - Genotypes

Cohort 1 (Healthy)	AA	AG	GG	TOTAL
OBSERVED	26	19	5	50
EXPECTED	25.205	20.59	4.205	50

Table 56: Cohort 1 (CMS patients)- Genotypes

Cohort 1 (CMS)	AA	AG	GG	TOTAL
OBSERVED	17	11	22	50
EXPECTED	10.125	24.75	15.125	50

8.2 Rs12097901

Table 57 and Table 58 show the number of individuals for each genotype and the allele frequencies, respectively, for rs12097901.

Table 57: SNP rs12097901 C127S

	CMS Patients				Healthy Volunteers			
Cohort	**GG (N)**	**GC (N)**	**CC (N)**	**Total (N)**	**GG (N)**	**GC (N)**	**CC (N)**	**Total (N)**
1	31	12	0	43	38	9	0	47
2	22	5	2	29	50	8	2	60

Table 58: Allele Counts and allele frequencies for SNP rs12097901

	CMS Patients		Healthy Volunteers		Total	
Cohort	**Allele G**	**Allele C**	**Allele G**	**Allele C**	**Allele G**	**Allele C**
1	74 (0.86)	12 (0.14)	85 (0.90)	9 (0.10)	159 (0.88)	21 (0.12)
2	49 (0.84)	9 (0.16)	108 (0.90)	12 (0.10)	157 (0.88)	21 (0.12)

The thing to note here is that the frequency of allele C (which causes C127S substitution) in both the Andean cohorts (Table 58) is significantly lower than in the Tibetan cohort (Table 16) in which it was 0.81. Statistical analysis was performed separately for each cohort (using a χ^2 test as for rs1765811) and revealed that there is no significant difference in the allelic frequencies of rs12097901 in the CMS versus healthy volunteers, in either cohort.

I also went on to determine any correlation between SNP rs12097901 and SNP rs17655811 in each Andean cohort. The results show that there is a significant Spearman's correlation coefficient between the genotypes at the two SNPs, as shown in Table 59.

Table 59: Spearman's correlation coefficient

	CMS patients	Healthy Controls	Total
Cohort 1	$r^2 = 0.381$ ($p = 0.01$)	$r^2 = 0.593$ ($p = 0.00$)	$r^2 = 0.498$ ($p = 0.00$)
Cohort 2	$r^2 = 0.541$ ($p = 0.02$)	$r^2 = 0.472$ ($p = 0.00$)	$r^2 = 0.498$ ($p = 0.00$)

This suggests that – as was the case in Tibetans, see Chapter 3 – there may be a high degree of linkage disequilibrium in this region of *EGLN1* in Andean populations as well. This is of potential interest as the genomic area between the two SNPs in the Andeans, includes the 5'UTR of *EGLN1* in which there is a known HIF binding site and where there are markers associated with regulatory activity in this genomic region. Fine genomic mapping of this area in both the Andeans and Tibetans is likely to be very informative. For

example, what would be interesting to determine is whether correlation between these 2 SNPs and indeed linkage disequilibrium in this genomic region is present in sea-level populations as well or whether it is only observed in high-altitude populations.

Chapter 9: Supplementary study examining patients with idiopathic erythrocytosis by whole genome sequencing of patients in search for potential functional variation in the HIF pathway

9.1 INTRODUCTION

While most of my thesis focused on the study of Tibetan natives who appear to have an altered ("hyporesponsive") HIF system and have reduced haematocrit and haemoglobin concentrations, it was attractive to undertake a small supplementary study examining patients who have increased haematocrit and haemoglobin concentrations with evidence that these may result from an altered ("upregulated") HIF system, i.e. elevated or inappropriately normal erythropoietin levels.

Erythrocytosis is a clinical condition characterised by increased red cell mass (and typically elevated haematocrit and haemoglobin concentration) [197]. It is divided into: *Primary*, where there is an intrinsic defect of the erythropoietic pathway in the bone marrow (e.g. polycythaemia *rubra vera*) and typically there are low levels of serum erythropoietin (Epo) and *secondary*, where the increased red cell production is driven by factors external to the erythroid compartment, e.g. increased erythropoietin production (for any reason) with typically elevated or inappropriately normal levels of serum Epo. Secondary erythrocytoses can be further classified into congenital, e.g. genetic defects of the oxygen-sensing (HIF) pathway or *acquired,* e.g. hypoxic conditions triggering Epo production (such as congenital heart disease). Table 60 shows a classification of erythrocytosis (adapted from [198]).

Table 60: Classification of erythrocytoses

PRIMARY ERYTHROCYTOSIS	
Congenital Epo Receptor Mutations	**Acquired** Polycythaemia vera (including JAK2 mutations)
SECONDARY ERYTHROCYTOSIS	
Congenital (inherited) 1. Defects in the oxygen-sensing pathway: • VHL mutations • PHD2 mutations • HIF2α mutations 2. Other congenital defects: • High-oxygen affinity haemoglobin • Bisphosphoglycerate mutase deficiency	**Acquired (Epo-mediated)** 1. Central hypoxia: • Chronic lung disease • Right-to-left cardiopulmonary vascular shunts • Smoker's erythrocytosis • Hypoventilation syndromes (e.g. obstructive sleep apnoea) • High altitude 2. Local hypoxia: • Renal artery stenosis • Renal cysts (Polycystic kidney disease) • Post-renal transplant erythrocytosis 3. Pathological Epo production (tumours): • Cerebellar haemangioblastoma • Meningioma • Parathyroid carcinoma/adenoma • Hepatocellular carcinoma • Renal cell cancer • Phaechromocytoma • Uterine leiomyoma 4. Drug-associated: • Erythropoietin administration • Androgen administration

As discussed in Chapter 1, a number of coding mutations in key genes of the HIF pathway have been identified in patients with erythrocytosis such as: 1) Mutations in VHL (Chuvash Polycythaemia), 2) Mutations in PHD2 and 3) Gain-of-function mutations in HIF2α. It was also discussed in Chapter 1 that many of these patients have, apart from erythrocytosis, other abnormalities in their integrative physiology. There still remains a group of patients with erythrocytosis in whom – despite thorough investigation to exclude the causes mentioned in Table 60 – the cause of their disease is not yet known. These

patients are said to have idiopathic erythrocytosis. It is estimated that 1/3 of these patients have Epo levels below the normal range, suggesting that they have primary erythrocytosis. The other 2/3s have normal or (inappropriately) elevated Epo levels, and so have a secondary idiopathic erythrocytosis [198].

A conceivable hypothesis that merits investigation is that idiopathic erythrocytosis with inappropriately normal or elevated serum Epo may be caused by genetic variation and abnormality in the HIF pathway and that patients with this disease may also exhibit other integrative physiology phenotypes consisted with an “up-regulated HIF response”.

One approach to explore the above hypothesis would be to undertake an integrative physiology study of a selected group of patients with idiopathic erythrocytosis who have extreme phenotypes, of early age-onset and who have family history of the disease. If evidence of an integrative physiology phenotype was found such as abnormal cardiopulmonary responses to hypoxia or abnormal metabolism in relation to exercise, then this would imply a high probability for a defect in the oxygen-sensing HIF pathway. This could then be followed by a more targeted approach in which candidate genes of the HIF pathway were sequenced in individuals who exhibited an abnormal integrative physiology, looking for genetic mutations.

An alternative approach would be to vary the sequence of steps in the above approach and start by undertaking sequencing of candidate genes in the HIF pathway in a selected group of patients with idiopathic erythrocytosis (extreme phenotype, early-onset disease and/or family history) looking for genetic mutations. If potential causative mutations were found

in certain patients, then these patients could be investigated at the integrative physiology level to look for a systemic phenotype characterised by an "upregulated HIF response".

A collaborative inter-departmental initiative at the University of Oxford (WGS500) in partnership with Illumina, which aims to evaluate the clinical utility of whole genome sequencing by sequencing a total of 500 clinical samples where the results could have immediate clinical translational relevance, gave me the opportunity to take a somewhat varied approach. I contributed to this initiative by putting forward a proposal for undertaking whole genome sequencing in a small collection of patients with idiopathic erythrocytosis with inappropriately normal or raised Epo levels. The scientific rationale for this is elucidated below.

First, studying patients with extreme phenotypes where there is good understanding of basic physiology has more chance of success in the effort to resolve causal or functional mutations. Second, the oxygen-sensing HIF pathway is a well-studied system and there is capacity for downstream follow-up of findings with functional assays at the cellular and molecular level (in Professor Ratcliffe's laboratory) in order to elucidate functional causality and significance. Third, there is capacity for the individuals in whom mutations are identified to be studied at the integrative physiology level for other functional consequences (in Professor Robbins's laboratory). The latter is not only of scientific importance by reinforcing functional significance of genetic variation but can also have significant clinical implications for the patients. Fourth, the vast number of genes that are "key players" in the oxygen-sensing pathway which drives erythropoiesis (e.g HIF1α, HIF3α, HIF1β, PHD1, PHD2, PHD3, FIH, and also EPO and EPOR (Epo Receptor gene) together with growing evidence of new genes interacting with the HIF factors directly or

indirectly, make whole genome sequencing an appropriate – and most informative – next investigation of choice for this group of patients at the genomics/genetics level. In addition, WGS enables the "interrogation" of other genes that may be involved in producing this phenotype (e.g. the α- and β- globin genes given that high-affinity haemoglobin variants can also lead to erythrocytosis, and the bisphosphoglycerate mutase gene, deficiency of which has also been described in erythrocytosis [199]).

WGS can play an important diagnostic role in clarifying the aetiology of the disease, guiding treatment strategies and in some cases for family screening and counselling. Currently, the treatment in these polycythaemic patients is empirical (venesection and low-dose aspirin) and there is little evidence to guide it. Further understanding of the aetiology may prove helpful, particularly at a time when pharmacological manipulation of the HIF pathway is on its way.

Finally, if genetic loci in which mutations lead to functional anomaly in the small collection of patients in this project were discovered, then a more targeted screening using gene-selective Sanger sequencing of other patients with erythrocytosis (and their families) on the already established UK and European databases could be carried out.

9.2 METHODS

The patients included in the whole genome sequencing project were chosen from two registries, a UK-based registry (Professor Mary Frances McMullin, collaborator) and a Germany-based registry (Professor Holger Cario, collaborator). The rationale employed for selecting them was: extreme phenotype (very high haemoglobin), inappropriately normal or raised serum Epo level, early onset of disease (childhood, adolescence or early

adulthood), exclusion of potentially confounding co-existing disease, exclusion of alternative aetiology through clinical and laboratory investigation (including screening for already known erythrocytosis-causing mutation in candidate genes) and presence of family history. In the case of family pedigrees, some unaffected members were also sequenced.

Blood-extracted DNA was provided from the relevant databases. The quality and quantity of DNA was assessed by gel electrophoresis. Whole genome sequencing was undertaken by the Genomics Services at the Wellcome Trust Centre in Human Genetics, in collaboration with Illumina, using Illumina HiSeq 2000 high-throughput sequencing technology. The detailed methodologies involved in the technique are beyond the scope of this thesis and hence are not going to be described here.

The bioinformatic analysis of the output data was performed by Dr Richard Copley, Reader in Bioinformatics at the Wellcome Trust Centre of Human Genetics. This project is ongoing with more analysis being undertaken as new information arises, but at the time of writing the analysis done employed the following broad criteria:

1. Search for coding (heterozygous or homozygous) variants (mutations, copy number variation etc.) in candidate genes.
2. Search for non-coding (heterozygous or homozygous) variants (mutations, copy number variations etc) in candidate genes
3. Search for rare coding SNPs in the erythrocytosis samples that are: heterozygous, found in “1000 genomes” at a frequency lower than 5%, not seen in the WGS500 project as a whole more than 5 times, are not in a segmental duplication, and are found in genes that are common across a number of erythrocytosis samples (i.e. the same gene harbours a SNP/mutation in more than one erythrocytosis sample).

4. Search for rare coding homozygous mutations in sporadic cases.

9.3 RESULTS

9.3.1 Patients selected for whole genome sequencing

9.3.1.2 Sporadic cases

The details of these are found in Table 61.

Table 61: Sporadic cases

Sample ID	Demographic Characteristics	Hb (g/dl)	Epo (miU/ml)	Clinical Information
PAR03	Male; Caucasian	19.3	15.9	Diagnosed age 27; HCT: 0.58 l/l, Red cell mass: 146% predicted, Normal Hb electrophoresis Non-smoker Normal abdominal ultrasound Severe varicose veins
PAR02	Male; Pakistani	20.5	14	Diagnosed age 28 HCT: 0.6 l/l, platelets: 99 x 10^9/l, Red cell mass: 110% predicted LDH: 456 normal ABG, normal P50 Liver haemangiomas
PAR11	Male; Caucasian	22.4	21	Diagnosed age 3

Normal range for Epo: 3.3 – 15.8 miU/ml

9.3.1.2 Families

Samples from two families, Family M and Family S were included in the WGS. The family trees are shown in Figure 58. DNA samples from PAR15 and PAR16 (from family M) were whole-genome sequenced whereas DNA from PAR17 was used in follow-up genotyping. DNA samples from PAR09, PAR07 and PAR18 (from family S) were whole-genome sequenced whereas DNA from PAR08 and PAR19 were used in follow-up genotyping. The clinical characteristics of the affected patients are shown in Table 62.

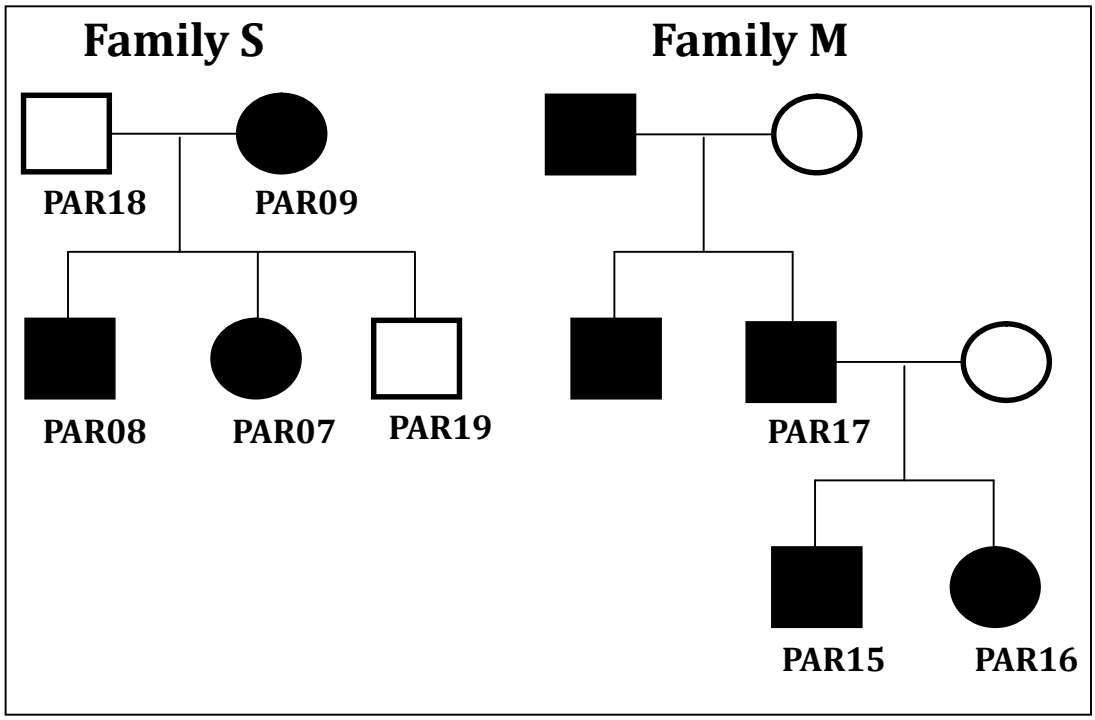


Figure 58: Family trees for Family S and Family M. Black denotes affected with erythrocytosis members.

Table 62

Sample ID	Demographic Characteristics	Hb (g/dl)	Epo (miU/ml)	Clinical Information
PAR17	Male	U/A	U/A	Known to have polycythaemia; Treated with venesection
PAR16	Female	U/A	U/A	Diagnosed age 15; Hct:0.6 l/l
PAR15	Male	U/A	U/A	Diagnosed age 12; Hct:0.55 l/l
PAR09	Female, Caucasian	16.4	15.1	Diagnosed age 29; Hct: 0.54 l/l
PAR07	Female, Caucasian	18.2	24.5	Diagnosed age 2; Hct: 0.59 l/l
PAR08	Male, Caucasian	17.7	10.0	Diagnosed age 2; Hct: 0.59 l/l

U/A: data unavailable.

9.3.1.3 Distant relatives

PAR04: Male, Pakistani, Hb: 21.6 g/dl, Epo: 17.4 miU/ml, diagnosed as a teenager

PAR12: Female, Pakistani, diagnosed age 30 (currently 54)

Relationship: PAR12 is the first cousin of PAR04's father

9.3.2 Coding mutations identified in candidate genes that are likely to be functional

Patient PAR03 was found to have a heterozygous mutation in BPGM (bisphosphoglycerate mutase) gene. This is a G269A non-synonymous SNV leading to R90H.

9.3.3 Non-coding mutations identified in candidate genes that are likely to be functional

There were a few non-coding variants (SNPs) identified in candidate genes. One such variant is a heterozygous single nucleotide variant (SNV) in the 5′ UTR of the EPO gene (chr7: 100318468 G>A). Affected members (sequenced or genotyped) of families M and S share this SNV which is absent from the unaffected members.

9.3.4 Rare homozygous coding (or transcribed) mutations

There was a number of rare homozygous coding or non-coding mutations. The ones thought to be worth pursuing further in terms of likelihood of functional relevance are the ones shown in Table 63.

Table 63

Sample	Gene	Mutation	Other notes
PAR02	*GFI1b*	chr 9: 135863848 G>T C168F	Extended homozygosity across the genome
PAR11	*KDM6A*	chr X: 44920641 T>C C468R	X-linked, not homozygous
PAR04 and PAR12	*SHARP1* (DEC2)	rs76268917 (C>T) rs76306214 (T>C) both in the 3'UTR	Extended homozygosity in both these samples

9.4 DISCUSSION

9.4.1 Mutation in BPGM

Bisphosphoglycerate Mutase (BPGM) is an enzyme responsible for the synthesis of 2,3-DPG (from 1,3-DPG) in red blood cells. In fact, BPGM is found only in red blood cells and in the placenta [200]. It binds to haemoglobin and reduces its affinity to oxygen. Thus, deficiency of BPGM leads to reduced level of 2,3-DPG which in turn leads to a leftward shift of the haemoglobin-oxygen dissociation curve, which is similar to having a high oxygen affinity haemoglobin. This could lead to either increased erythropoiesis or haemolysis.

Defects in BPGM have been associated with haemolytic anaemia [201] [202] but also erythrocytosis, as discussed below. In 1978, Rosa et al reported the first case of a 42-year old man that had complete deficiency of BPGM. His erythrocyte 2,3 DPG was below 3% of normal values, he had a low P_{50} (partial pressure of oxygen in the blood at which the haemoglobin is 50% saturated) and had erythrocytosis with a Hb concentration of 19 g/dl [203]. It was later shown that this patient belonged to a family in which a number of members had erythrocytosis; it was found that the cases with complete BPGM deficiency –

who had excessive erythrocytosis – were in fact compound heterozygous with a SNV on one allele causing a R90C substitution resulting in low levels of an abnormal (inactive) BPGM protein and a deletion of nucleotide 205 or 206 (codon 19) on the other allele, translating an aberrant product. Some members of this family had partial BPGM deficiency (but were also polycythaemic) and only had the R90C mutation [199] [204]. Garel et al [205] reproduced the same functional defect in BPGM as the original arginine to cysteine mutation by site-directed mutagenesis producing Ar90Gly and Arg90Ser, giving evidence that indeed the arginine at codon 90 is functionally important in the enzyme's function. Note also that a male patient with concurrent G6-phosphate dehydrogenase deficiency and erythrocytosis – with low P_{50} and low erythrocyte 2,3-DPG – was also described; he had a homozygous Arg62Gln, and there was evidence of consanguinity in his family [206].

The heterozygous coding mutation identified in PAR03 is at the same codon (90) as in the previously reported cases; this makes it highly likely that the mutation is the causal variant for erythrocytosis in this patient. It is interesting to note that it not the same base substitution that occurs and the aminoacid substitution is also different, arginine to histidine as opposed to arginine to cysteine in previous cases. Thus, this is a novel mutation and it probably reinforces the findings of Garel et al that the arginine residue is important for the function of the enzyme.

At the time of write up, this case is followed up by: 1. Screening both parents of the patient (who are healthy) to find out if this is a *de novo* mutation in the patient, or a mutation with variable penetrance in producing an erythrocytosis phenotype. (Note that we know already that there is no evidence of compound heterozygosity in this patient from the whole genome sequencing data); 2. Arranging to measure P_{50} and red cell 2,3-DPG to prove

functionality. (This is organised locally by the haematologist that is looking after the patient).

9.4.2 SNV in the 5′ UTR of Epo

From the family trees, there is a clear pattern of an apparent multigenerational autosomal dominant inheritance, which favours a heterozygous variant as a candidate. The SNV identified in the 5′ UTR of EPO segregates appropriately in both families such that affected members carry it whereas those unaffected don't. It is rare and has not been reported before, neither is it found in any other samples in the entire WGS500 project. Unfortunately two affected male members (grandfather and uncle of PAR15 and PAR16) are not available for us to study. Confirming the presence of the SNV in these individuals would increase our confidence for functional candidacy of this SNV.

In what way could this SNV in the 5′ UTR be affecting erythropoietin production and thus contribute to the development of erythrocytosis? To try and speculate on this, I will first discuss some key points on the regulation of erythropoietin and then discuss some key points on how variation in the un-translated regions of genes can affect gene expression and/or cause disease.

9.4.2.1 Erythropoietin regulation

Erythropoietin is regulated in an oxygen-dependent manner under the transcriptional regulation of HIF. In fact HIF1 was initially discovered as it was found to bind a Hypoxia-Response Element (HRE) in an enhancer 3′ of the Epo gene [44]. There are 2 further interesting aspects of erythropoietin regulation. One is it tissue-specificity; only very few cell types in the mammalian body have the ability to produce Epo – brain, renal epithelial

cells and hepatocytes –, suggesting that there are different activating or repressing regulatory mechanisms in different tissues. This is interesting as the hypoxic signalling cascade is universally present in many different cell lines [45]. The second is the temporal or developmental aspect, with evidence that the liver produces Epo during embryonic life while this is repressed after birth when the kidney becomes the primary site of Epo production. Experiments with transgenic mice mapped regions of the Epo gene that regulate Epo expression in a tissue-specific or developmental manner [207-210]. In addition transient transfections of cultured cell lines have shed light into the DNA sequences that are critical for the hypoxic response [43]. In brief, elements found to be important in the regulation of the Epo gene are: the promoter – situated 5′ stream of the gene –, a 3′ enhancer (containing a HRE) along with other additional regulatory DNA elements such as the kidney-inducible element (KIE), the negative regulatory element (NRE), the liver-inducible element (LIE) and the negative regulatory liver element (NRLE) [175]. It is also thought that different HREs are operative in the kidney and the liver. Apart from HIF, there are other nuclear factors and hypoxia-associated factors involved in the transcriptional regulation of Epo [175].

Yi et al showed that the promoter and the 5′ UTR of the Epo gene comprise a CpG island, methylation of which correlates inversely with expression [211]. Methylation represses the expression of the erythropoietin gene in 2 ways: high-density methylation of the 5′ UTR recruits a methyl-CpG binding protein to the promoter, and methylation of CpGs in the proximal promoter blocks the association of nuclear proteins. They showed less methylation of CpGs in Epo-producing cells compared to non-Epo producing cells. There is also some evidence that binding of an Epo mRNA binding protein on the 3’ UTR is believed to affect Epo mRNA stability, although this mechanism is not fully elucidated

[212]. Finally, there is a report that removal of the 3′ and 5′ untranslated regions amplifies expression of the erythropoietin gene in transfected mammalian cells (COS-7) [213].

9.4.2.2 Translational pathophysiology in human disease related to variation in ′UTR regions

There is evidence of negative regulatory elements in the 5′ UTR of genes that can interfere with the initiation of protein translation. Examples of such regulatory elements are:

1. Formation of stable secondary stem-loop structures of mRNA: Self complementary sequences form stable stem-loop structures that interfere with the assembly of the pre-initiation complex and/or ribosomal scanning, and that often interact with RNA-binding proteins. One such loop structure is iron-responsive elements (IREs) which interact with iron regulatory proteins (IRPs) [214, 215]. A functional IRE is found in the 5′ UTR of ferritin and is important in cellular iron homeostasis [216]. Hereditary hyperferritinemia/cataract syndrome arises from various point mutations or deletions within a protein-binding sequence in the 5′ UTR of the L-ferritin mRNA [217]. IREs have also been found in the 5′ UTR of genes such as HIF2α and DMT1 gene [218] [219].

2. Presence of upstream open reading frames in the 5′ UTR: AUG codons upstream to the physiological start site can cause premature initiation and inhibit translation by preventing the ribosome from reaching the physiological start codon. Mutations interfering with these open reading frames and with translational repression have been found in the thrombopoietin gene that causes the disease hereditary thrombocythaemia, in which there is elevated thrombopoietin expression [220] [221, 222].

In the Epo gene, a 14-amino-acid open reading frame (ORF) located upstream of the initiator codon in the human (and mouse) gene was found [223] . No evidence of a functional role for this ORF has been observed and the sequence is not required for the expression of a biologically active product [224]. The SNV identified in the family pedigrees in the WGS does not lie in this ORF. However, it does lie in the 5′ UTR region comprising the CpG island.

Current and future work involves screening the maternal grandparents in family S (who are healthy) for the mutation to investigate whether it is a *de novo* mutation in individual PAR09 or whether it is inherited. If it is indeed inherited, further bioinformatic analysis will also be targeted in finding out whether there is an extended haplotype within this region that is present in a number of different individuals and whether the affected individuals in these two families share a haplotype that carries this rare variant; if so, this will suggest that this variant arose once in the past and that the families share a common ancestor. It is also possible that this variant arose independently in the two families and there are also plans to screen a number of patients (familial or sporadic) for this SNV from available databases.

9.4.3 Rare homozygous coding or transcribed mutations

Rare homozygous mutations have been found in genes GFI1b, KDM6A and DEC2. Below, I give some information regarding these genes to explain why they may be potentially relevant in the phenotype of erythrocytosis and why they merit further follow-up.

GFI1b encodes a 330-amino acid zing finger protein. It is expressed in the bone marrow and it is thought to be important in erythropoiesis. It was shown that it is an essential transcriptional regulator of erythroid and magakaryocyte development through catalysing serial histone modification of target genes leading to their graded silencing [225]. In addition it was shown that GFI1b$^{-/-}$ embryos appear to die during the transition from primitive to definitive haematopoiesis, with failure to produce enucleated erythrocytes [226].

KDM6A is JmJC-domain-containing demethylase that catalyses demethylation of tri/dimethylated histone at H3 lysine 27 loci. There is evidence that mutations in KDM6A (11 somatic and one germline) are found in renal cancer [227, 228] and are also associated with a congenital syndrome, Kabuki Syndrome [229, 230].

DEC2 is a transcription factor that is hypoxically regulated and it has recently been shown to interact with HIF; it was shown to cause proteasomal inhibition of HIF and transcriptional suppression of HIF-target genes [231].

Whole genome sequencing as a clinical diagnostic tool is at its infancy with the methodology continuously evolving. More analysis could be possible as new information regarding genes, gene function and regulation becomes available. The candidate variations that have been identified so far could be used for targeted screening of a larger number of patients with idiopathic erythrocytosis from existing registries.

Chapter 10: Discussion, conclusions and future work

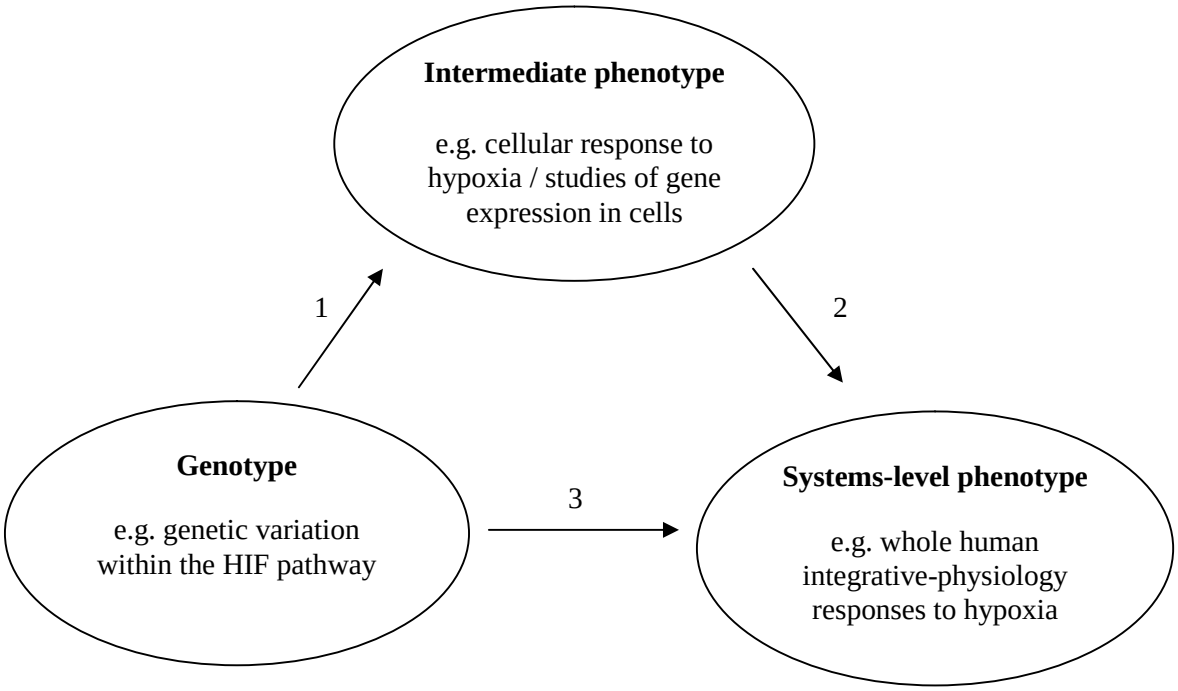


Figure 59

The diagram in Figure 59 displays the general "philosophy" underlying the work in this thesis. By taking a number of different experimental approaches and employing a number of different methodologies addressing links 1, 2 and 3, I sought to explore functional variation in the Hypoxia-Inducible Factor pathway in humans.

Using a validated and well-controlled experimental methodology, I described the integrative physiology – ventilatory, cardiopulmonary and erythropoietic – of Tibetan natives during air-breathing and in response to hypoxia, at sea level. I showed that even at sea level, Tibetan natives have a distinct integrative physiology phenotype when compared to their most closely-related major ethnic group, the Han Chinese. In particular, I found that Tibetans had a lower haemoglobin concentration and haematocrit, a higher pulmonary ventilation relative to metabolism, and blunted pulmonary vascular responses to both acute

(min) and sustained (8 hours) hypoxia. This distinct phenotype may arise – at least in part – from genetic differences between Tibetans and Han Chinese in major genes of the HIF pathway – namely *EGLN1* and *EPAS1* – in which positive selection has previously been reported in genomic studies of Tibetans. For one phenotypic characteristic, the induction of systemic erythropoietin by sustained hypoxia, I was able to show that this correlates significantly with genotype at both the *EPAS1* and *EGLN1* loci. I was able to reinforce the findings of the physiology study with findings at the intermediate cellular level using a peripheral blood lymphocyte model. Here, I showed that Tibetans and Han Chinese differ in their relative HIF-target gene expression and in their HIF-regulated cellular hypoxic response.

In conclusion, this study provides further evidence that Tibetans respond less vigorously to hypoxic challenge, that this is also the case at sea level and that, at least in part, this appears to arise from a hypo-responsive HIF transcription activation system.

One limitation to the study was the small number of Tibetan volunteers I was able to recruit to participate in the physiology study which was limited by the small number of Tibetans resident in the UK. This means that the phenotypic differences that were detected in the study had to be large whereas the absence of statistically significant differences could be the result of a type II statistical error, owing to a small sample size. For example, a number of the physiological traits correlated with *EPAS1* genotype in the “predicted” or hypothesised “direction”, but this fell short of statistical significance. This, together with the other significant results of the study, would support the recruitment-by-genotype of more Tibetan volunteers in a potential larger-scale study, to investigate specifically whether other physiological traits – e.g. pulmonary vascular response – associate

significantly with genotype. This merits investigation as it is still unknown whether Hb concentration at high altitude is the advantageous phenotype or whether – given the pleiotropism of the HIF pathway – another phenotype is the naturally selected one.

Peripheral blood lymphocytes were used as an intermediate cellular model as they are cells that are easily accessible and obtainable from human volunteers and also give the opportunity to produce stable cell lines by EBV transformation. The limitations of the latter were discussed in Chapter 5. In the peripheral blood lymphocytes there is evidence of a smaller hypoxic induction of HIF-regulated gene in Tibetans compared to Han Chinese, as was seen in the responses at the integrative physiology level. To follow up on this result and to investigate the oxygen-regulated gene expression in Tibetans in more detail in the future, one approach would be to employ Next Generation RNA Sequencing in Tibetan cells cultured in euoxia and in hypoxia. This would generate a vast amount of information and at the same time would allow increased internal control of experiments and give more power to the results. It could provide an opportunity for unbiased gene expression profiling and for identifying expression patterns e.g. of hypoxically-regulated genes, metabolic genes or genes belonging to certain relevant biological pathways. Furthermore, RNASeq would make it possible to identify novel transcripts and alternative splice variants (that may differ between euoxia and hypoxia), to investigate non-coding RNAs and to look for allele-specific expression (in heterozygous cells).

One factor to consider in the profiling of gene expression is tissue specificity. Lymphocytes are a useful tool in examining hypoxically-regulated genes; however, it may be more appropriate to embark on a complex, costly and time-consuming procedure such

as RNASeq using a cell type that is more disease- or phenotype-relevant. One such cell type is endothelial progenitor cells, which I will refer to later on.

The functional sequence variants in either *EPAS1* or *EGLN1,* as well as their mechanism of action are still unknown. In Chapter 6, I investigated whether the mechanism of action of the variation in *EPAS1* is regulation of *EPAS1* transcription and in Chapter 7 I investigated whether coding variants identified in Tibetans are functional. I was not able to show differential allelic expression – which would infer differential allelic transcription – of EPAS1 in lymphocyte cells in Tibetans in Chapter 6. As I discussed in Chapter 6, the fact that this may not be present in lymphocytes does not preclude that it occurs in other cell types. Cells of endothelial lineage have high abundance in HIF2α in contrast with lymphocytes. Hence it is worth investigating allele-specific expression of EPAS1 in endothelial progenitor cells.

Endothelial progenitor cells (EPC) are a population of rare cells that circulate in the blood and contribute to angiogenesis. There is controversy in the field over how to accurately identify circulating endothelial progenitors. When I refer to endothelial progenitor cells, I refer to a highly proliferative, polyclonally expanded EPC type, termed endothelial colony-forming cells or late-outgrowth EPCs [232]. In addition to the high abundance of HIF2α, these EPC constitute a cell type that is relevant to the pulmonary vasculature as they were shown to be involved in the vascular remodelling associated with pulmonary arterial hypertension [233]. These cells can be obtained from peripheral blood mononuclear cells (isolated from as little as 50 ml of blood), selected and cultured via a specialised protocol and can remain stable as a primary cell line for a few passages (personal communication with Mark Toshner, University of Cambridge). Thus, they may represent an ideal cell type

to obtain from Tibetan individuals; this would allow not only examination of gene expression but also the potential to look for phenotypic characteristics in these cells in Tibetan (e.g. proliferative potential in euoxia versus hypoxia, ability to form vascular networks etc).

Furthermore, the methodology and the considerations around the topic of allele-specific expression of EPAS1 in Chapter 6 have sparked interest in investigating how transcription of *EPAS1* is regulated. While there is increasing evidence that non-coding variation in *EPAS1* is associated with "functional effects" eg. genetic adaptation in Tibetans and genetic pre-dispositions in renal cancer [234] [235], little is known about how EPAS1 expression is regulated. This is something worth investigating further, in the appropriate tissue or cells, both in general and specifically in Tibetans. For example, is the genomic region downstream of *EPAS1* a functional regulatory element affecting the transcription of EPAS1? Is there evidence of association of the *EPAS1* promoter through chromatin looping with regulatory elements on the same chromosome near or distant to *EPAS1*? What role does hypoxia have to play in the transcriptional regulation of *EPAS1*? In EPAS1-rich cells what is the mapping of DNase hypersensitivity sites at the *EPAS1* locus in euoxia or hypoxia?

There is a lot to learn about and from the genetic structure of Tibetans. A collaboration that has already begun is a project in which 15 Tibetan genomes are whole-genome sequenced – 10 of these are from Tibetan volunteers in my study. Data from the whole-genome sequencing – that was performed in the Institute of Systems Biology in Seattle – have already been generated and are currently analysed by a consortium of investigators. This can lead to fine genomic mapping of the *EPAS1* and *EGLN1* regions, can allow

investigation of the whole genome in an unbiased way and can potentially identify other genetic signatures from whole genome scans generated from these data.

Furthermore, the possibility of coupling or integrating results from whole-genome sequencing – that has already been undertaken – with results from analyses such a RNASeq, analysis of epigenetic effects e.g. histone modifications at a genome-wide level or analysis of patterns of DNA accessibility e.g. using a FAIRE-Seq assay (FAIRE stands for formaldehyde-assisted isolation of regulatory elements) is likely to have a powerful impact in aiding our understanding of functional variation in the genetic adaptation to hypoxia in Tibetans. These are avenues worth pursuing.

In Chapter 9, the approach of whole genome sequencing in investigating idiopathic erythrocytosis was explored. The application of whole genome sequencing (WGS) to clinical medicine has the potential to revolutionise patient care. However, this approach is currently at its infancy. In addition to aiming to resolve causal mutations for diseases with clearly defined phenotypes, one another important aim of the WGS500 project, as a whole, is to evaluate its clinical utility and feasibility at this time, as a "proof of principle". From the experience of the WGS of a small number of cases of idiopathic erythrocytosis it is evident that some candidate variants are found in candidate genes (e.g. *BPGM* and *EPO* genes) selected *a priori* whereas others are found in genes that are associated with the HIF pathway, but were not selected as candidate genes prior to analysis, e.g. *KDM6A* and *SHARP1* (DEC2). This information has helped define a strategy to take this forward. In collaboration with haematologists in charge of registries with patients that have idiopathic erythrocytosis across Europe and with geneticists within the Wellcome Trust Centre of Human Genetics in Oxford, we have plans to set up a diagnostic platform whereby a panel

of candidate genes for idiopathic erythrocytosis (including both key genes and variants previously established and genes/variants of potential candidacy identified through WGS500) can be sequenced – using the AmpliSeq technology (*Life technology*) – in a large cohort of patients (~ 100 or more) with either sporadic or familial idiopathic erythrocytosis. The overall process will involve a two-step approach whereby a large number of cases can be screened through the AmpliSeq panel followed by whole-genome sequencing of a smaller selection of patients.

There is a great deal of variation within humans in their physiological and pathophysiological responses to hypoxia e.g. in patients with hypoxic lung disease, which is unaccounted for. Such phenotypic variation may be related to underlying variation in oxygen sensing and the adaptive response to hypoxia and may be related – amongst other things or in part – to underlying variation in the complex HIF pathway. I believe that an approach whereby tools are utilised to relate experimental findings or medical observations upwards to a systems-level phenotype and downwards to genotype through an intermediate cellular or molecular phenotype can be extremely valuable. Not only can it aid exploration of the HIF system in relation to high altitude adaptation, to human disease and to natural human variation, but lessons learnt from such an approach in one system can be useful and applicable in other domains of medicine and biology.

REFERENCES

1. **Jelkmann, W., *Regulation of erythropoietin production.* Journal of Physiology-London, 2011. 589(6): p. 1251-1258.**
2. **Eckardt, K.U., et al., *Rate of erythropoietin formation in humans in response to acute hypobaric hypoxia.* J Appl Physiol, 1989. 66(4): p. 1785-8.**
3. **Richalet, J.P., et al., *Control of erythropoiesis in humans during prolonged exposure to the altitude of 6,542 m.* Am J Physiol, 1994. 266(3 Pt 2): p. R756-64.**
4. **Milledge, J.S. and P.M. Cotes, *Serum erythropoietin in humans at high altitude and its relation to plasma renin.* J Appl Physiol, 1985. 59(2): p. 360-4.**
5. **Reynafarje, C., R. Lozano, and J. Valdivieso, *The polycythemia of high altitudes: iron metabolism and related aspects.* Blood, 1959. 14(4): p. 433-55.**
6. **Winslow, R.M., M. Samaja, and J.B. West, *Red cell function at extreme altitude on Mount Everest.* J Appl Physiol, 1984. 56(1): p. 109-16.**
7. **Weil, J.V. and C.W. Zwillich, *Assessment of ventilatory response to hypoxia: methods and interpretation.* Chest, 1976. 70(1 Suppl): p. 124-8.**
8. **Powell, F.L., W.K. Milsom, and G.S. Mitchell, *Time domains of the hypoxic ventilatory response.* Respir Physiol, 1998. 112(2): p. 123-34.**
9. **Moore, L.G., *Comparative human ventilatory adaptation to high altitude.* Respir Physiol, 2000. 121(2-3): p. 257-76.**
10. **Weir, E.K., et al., *Acute oxygen-sensing mechanisms.* N Engl J Med, 2005. 353(19): p. 2042-55.**
11. **Honda, Y., et al., *Hypoxic chemosensitivity in asthmatic patients two decades after carotid body resection.* J Appl Physiol, 1979. 46(4): p. 632-8.**
12. **West, J.B., *Respiratory Physiology: The Essentials*. 2005: Lippincott Williams & Wilkins.**
13. **Kumar, P. and N.R. Prabhakar, *Peripheral Chemoreceptors: Function and Plasticity of the Carotid Body.* Comprehensive Physiology, 2012. 2: p. 141-219.**
14. **Pedersen, M.E., K.L. Dorrington, and P.A. Robbins, *Effects of haloperidol on ventilation during isocapnic hypoxia in humans.* J Appl Physiol, 1997. 83(4): p. 1110-5.**
15. **Robbins, P.A., *Hypoxic ventilatory decline: site of action.* J Appl Physiol, 1995. 79(2): p. 373-4.**
16. **Busch, M.A., G.E. Bisgard, and H.V. Forster, *Ventilatory acclimatization to hypoxia is not dependent on arterial hypoxemia.* J Appl Physiol, 1985. 58(6): p. 1874-80.**
17. **Weizhen, N., et al., *Ventilatory effects of prolonged systemic (CNS) hypoxia in awake goats.* Respir Physiol, 1992. 87(1): p. 37-48.**
18. **Howard, L.S. and P.A. Robbins, *Ventilatory response to 8 h of isocapnic and poikilocapnic hypoxia in humans.* J Appl Physiol, 1995. 78(3): p. 1092-7.**
19. **Howard, L.S. and P.A. Robbins, *Alterations in respiratory control during 8 h of isocapnic and poikilocapnic hypoxia in humans.* J Appl Physiol, 1995. 78(3): p. 1098-107.**
20. **Fatemian, M. and P.A. Robbins, *Human ventilatory response to CO2 after 8 h of isocapnic or poikilocapnic hypoxia.* J Appl Physiol, 1998. 85(5): p. 1922-8.**
21. **Fatemian, M. and P.A. Robbins, *Selected contribution: chemoreflex responses to CO2 before and after an 8-h exposure to hypoxia in humans.* J Appl Physiol, 2001. 90(4): p. 1607-14; discussion 1606.**
22. **Fatemian, M., et al., *The respiratory response to carbon dioxide in humans with unilateral and bilateral resections of the carotid bodies.* J Physiol, 2003. 549(Pt 3): p. 965-73.**
23. **Clar, C., et al., *Effects of 8 h of isocapnic hypoxia with and without muscarinic blockade on ventilation and heart rate in humans.* Exp Physiol, 2001. 86(4): p. 529-38.**

24. **Vogel, J.A. and C.W. Harris, *Cardiopulmonary responses of resting man during early exposure to high altitude.* J Appl Physiol, 1967. 22(6): p. 1124-8.**
25. **Barbera, J.A., V.I. Peinado, and S. Santos, *Pulmonary hypertension in chronic obstructive pulmonary disease.* Eur Respir J, 2003. 21(5): p. 892-905.**
26. **Hackett, P.H. and R.C. Roach, *High-altitude illness.* N Engl J Med, 2001. 345(2): p. 107-14.**
27. **Penaloza, D. and J. Arias-Stella, *The heart and pulmonary circulation at high altitudes: healthy highlanders and chronic mountain sickness.* Circulation, 2007. 115(9): p. 1132-46.**
28. **Talbot, N.P., et al., *Two temporal components within the human pulmonary vascular response to approximately 2 h of isocapnic hypoxia.* J Appl Physiol, 2005. 98(3): p. 1125-39.**
29. **Dorrington, K.L., et al., *Time course of the human pulmonary vascular response to 8 hours of isocapnic hypoxia.* Am J Physiol, 1997. 273(3 Pt 2): p. H1126-34.**
30. **McMurtry, I.F., et al., *Inhibition of hypoxic pulmonary vasoconstriction by calcium antagonists in isolated rat lungs.* Circ Res, 1976. 38(2): p. 99-104.**
31. **Jabr, R.I., et al., *Prominent role of intracellular Ca2+ release in hypoxic vasoconstriction of canine pulmonary artery.* Br J Pharmacol, 1997. 122(1): p. 21-30.**
32. **Post, J.M., et al., *Direct role for potassium channel inhibition in hypoxic pulmonary vasoconstriction.* Am J Physiol, 1992. 262(4 Pt 1): p. C882-90.**
33. **Yuan, X.J., et al., *Hypoxia reduces potassium currents in cultured rat pulmonary but not mesenteric arterial myocytes.* Am J Physiol, 1993. 264(2 Pt 1): p. L116-23.**
34. **Archer, S.L., et al., *Impairment of hypoxic pulmonary vasoconstriction in mice lacking the voltage-gated potassium channel Kv1.5.* FASEB J, 2001. 15(10): p. 1801-3.**
35. **Robertson, T.P., P.I. Aaronson, and J.P. Ward, *Ca2+ sensitization during sustained hypoxic pulmonary vasoconstriction is endothelium dependent.* Am J Physiol Lung Cell Mol Physiol, 2003. 284(6): p. L1121-6.**
36. **Ward, J.P. and T.P. Robertson, *The role of the endothelium in hypoxic pulmonary vasoconstriction.* Exp Physiol, 1995. 80(5): p. 793-801.**
37. **Shimoda, L.A., et al., *Acute and chronic hypoxic pulmonary vasoconstriction: a central role for endothelin-1?* Respir Physiol Neurobiol, 2002. 132(1): p. 93-106.**
38. **Aaronson, P.I., T.P. Robertson, and J.P. Ward, *Endothelium-derived mediators and hypoxic pulmonary vasoconstriction.* Respir Physiol Neurobiol, 2002. 132(1): p. 107-20.**
39. **Archer, S.L., et al., *Hypoxic pulmonary vasoconstriction is enhanced by inhibition of the synthesis of an endothelium derived relaxing factor.* Biochem Biophys Res Commun, 1989. 164(3): p. 1198-205.**
40. **Wang, G.L., et al., *Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension.* Proc Natl Acad Sci U S A, 1995. 92(12): p. 5510-4.**
41. **Tian, H., S.L. McKnight, and D.W. Russell, *Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells.* Genes Dev, 1997. 11(1): p. 72-82.**
42. **Hara, S., et al., *Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha.* Biochem Biophys Res Commun, 2001. 287(4): p. 808-13.**
43. **Pugh, C.W., et al., *Functional analysis of an oxygen-regulated transcriptional enhancer lying 3' to the mouse erythropoietin gene.* Proc Natl Acad Sci U S A, 1991. 88(23): p. 10553-7.**

44. **Semenza, G.L., et al.,** ***Hypoxia-inducible nuclear factors bind to an enhancer element located 3' to the human erythropoietin gene.*** **Proc Natl Acad Sci U S A, 1991. 88(13): p. 5680-4.**
45. **Maxwell, P.H., C.W. Pugh, and P.J. Ratcliffe,** ***Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism.*** **Proc Natl Acad Sci U S A, 1993. 90(6): p. 2423-7.**
46. **Epstein, A.C., et al.,** ***C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation.*** **Cell, 2001. 107(1): p. 43-54.**
47. **Jaakkola, P., et al.,** ***Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation.*** **Science, 2001. 292(5516): p. 468-72.**
48. **Bruick, R.K. and S.L. McKnight,** ***A conserved family of prolyl-4-hydroxylases that modify HIF.*** **Science, 2001. 294(5545): p. 1337-40.**
49. **Ivan, M., et al.,** ***Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor.*** **Proc Natl Acad Sci U S A, 2002. 99(21): p. 13459-64.**
50. **McNeill, L.A., et al.,** ***The use of dioxygen by HIF prolyl hydroxylase (PHD1).*** **Bioorg Med Chem Lett, 2002. 12(12): p. 1547-50.**
51. **Masson, N., et al.,** ***Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation.*** **EMBO J, 2001. 20(18): p. 5197-206.**
52. **Knowles, H.J., et al.,** ***Effect of ascorbate on the activity of hypoxia-inducible factor in cancer cells.*** **Cancer Res, 2003. 63(8): p. 1764-8.**
53. **Cockman, M.E., et al.,** ***Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein.*** **J Biol Chem, 2000. 275(33): p. 25733-41.**
54. **Lando, D., et al.,** ***Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch.*** **Science, 2002. 295(5556): p. 858-61.**
55. **Lando, D., et al.,** ***FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor.*** **Genes Dev, 2002. 16(12): p. 1466-71.**
56. **Smith, T.G., P.A. Robbins, and P.J. Ratcliffe,** ***The human side of hypoxia-inducible factor.*** **Br J Haematol, 2008. 141(3): p. 325-34.**
57. **Berra, E., et al.,** ***HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia.*** **EMBO J, 2003. 22(16): p. 4082-90.**
58. **Appelhoff, R.J., et al.,** ***Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor.*** **J Biol Chem, 2004. 279(37): p. 38458-65.**
59. **Metzen, E., et al.,** ***Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing.*** **J Cell Sci, 2003. 116(Pt 7): p. 1319-26.**
60. **Wiesener, M.S., et al.,** ***Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs.*** **FASEB J, 2003. 17(2): p. 271-3.**
61. **Loenarz, C., et al.,** ***The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens.*** **EMBO Rep, 2011. 12(1): p. 63-70.**
62. **Ryan, H.E., J. Lo, and R.S. Johnson,** ***HIF-1 alpha is required for solid tumor formation and embryonic vascularization.*** **EMBO J, 1998. 17(11): p. 3005-15.**
63. **Peng, J., et al.,** ***The transcription factor EPAS-1/hypoxia-inducible factor 2alpha plays an important role in vascular remodeling.*** **Proc Natl Acad Sci U S A, 2000. 97(15): p. 8386-91.**
64. **Sowter, H.M., et al.,** ***Predominant role of hypoxia-inducible transcription factor (Hif)-1alpha versus Hif-2alpha in regulation of the transcriptional response to hypoxia.*** **Cancer Res, 2003. 63(19): p. 6130-4.**

65. **Raval, R.R., et al., *Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma.* Mol Cell Biol, 2005. 25(13): p. 5675-86.**
66. **Takeda, N., et al., *Differential activation and antagonistic function of HIF-{alpha} isoforms in macrophages are essential for NO homeostasis.* Genes Dev, 2010. 24(5): p. 491-501.**
67. **Mole, D.R., et al., *Genome-wide association of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha DNA binding with expression profiling of hypoxia-inducible transcripts.* J Biol Chem, 2009. 284(25): p. 16767-75.**
68. **Schodel, J., et al., *High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq.* Blood, 2011. 117(23): p. e207-17.**
69. **Bhattacharya, S., et al., *Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1.* Genes Dev, 1999. 13(1): p. 64-75.**
70. **Uchida, T., et al., *Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha.* J Biol Chem, 2004. 279(15): p. 14871-8.**
71. **Prabhakar, N.R. and G.L. Semenza, *Adaptive and maladaptive cardiorespiratory responses to continuous and intermittent hypoxia mediated by hypoxia-inducible factors 1 and 2.* Physiol Rev, 2012. 92(3): p. 967-1003.**
72. **Webb, J.D., M.L. Coleman, and C.W. Pugh, *Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing.* Cell Mol Life Sci, 2009. 66(22): p. 3539-54.**
73. **Minet, E., et al., *Hypoxia-induced activation of HIF-1: role of HIF-1alpha-Hsp90 interaction.* FEBS Lett, 1999. 460(2): p. 251-6.**
74. **Isaacs, J.S., Y.J. Jung, and L. Neckers, *Aryl hydrocarbon nuclear translocator (ARNT) promotes oxygen-independent stabilization of hypoxia-inducible factor-1alpha by modulating an Hsp90-dependent regulatory pathway.* J Biol Chem, 2004. 279(16): p. 16128-35.**
75. **Koh, M.Y., et al., *The hypoxia-associated factor switches cells from HIF-1alpha- to HIF-2alpha-dependent signaling promoting stem cell characteristics, aggressive tumor growth and invasion.* Cancer Res, 2011. 71(11): p. 4015-27.**
76. **Smith, T.G., et al., *Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology.* PLoS Med, 2006. 3(7): p. e290.**
77. **Ang, S.O., et al., *Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia.* Nat Genet, 2002. 32(4): p. 614-21.**
78. **Formenti, F., et al., *Regulation of human metabolism by hypoxia-inducible factor.* Proc Natl Acad Sci U S A, 2010. 107(28): p. 12722-7.**
79. **Percy, M.J., et al., *A gain-of-function mutation in the HIF2A gene in familial erythrocytosis.* N Engl J Med, 2008. 358(2): p. 162-8.**
80. **Percy, M.J., et al., *Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis.* Blood, 2008. 111(11): p. 5400-2.**
81. **Furlow, P.W., et al., *Erythrocytosis-associated HIF-2alpha mutations demonstrate a critical role for residues C-terminal to the hydroxylacceptor proline.* J Biol Chem, 2009. 284(14): p. 9050-8.**
82. **van Wijk, R., et al., *Erythrocytosis associated with a novel missense mutation in the HIF2A gene.* Haematologica, 2010. 95(5): p. 829-32.**
83. **Percy, M.J., et al., *Two new mutations in the HIF2A gene associated with erythrocytosis.* Am J Hematol, 2012. 87(4): p. 439-42.**
84. **Formenti, F., et al., *Cardiopulmonary function in two human disorders of the hypoxia-inducible factor (HIF) pathway: von Hippel-Lindau disease and HIF-2{alpha} gain-of-function mutation.* FASEB J, 2011. 25(6): p. 2001-11.**
85. **Percy, M.J., et al., *A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove.* Blood, 2007. 110(6): p. 2193-6.**

86. **Ladroue, C., et al.,** ***PHD2 mutation and congenital erythrocytosis with paraganglioma.*** **N Engl J Med, 2008. 359(25): p. 2685-92.**
87. **Ladroue, C., et al.,** ***Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia.*** **Haematologica, 2012. 97(1): p. 9-14.**
88. **Albiero, E., et al.,** ***Isolated erythrocytosis: study of 67 patients and identification of three novel germ-line mutations in the prolyl hydroxylase domain protein 2 (PHD2) gene.*** **Haematologica, 2012. 97(1): p. 123-7.**
89. **Kline, D.D., et al.,** ***Defective carotid body function and impaired ventilatory responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1 alpha.*** **Proc Natl Acad Sci U S A, 2002. 99(2): p. 821-6.**
90. **Yu, A.Y., et al.,** ***Impaired physiological responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1alpha.*** **J Clin Invest, 1999. 103(5): p. 691-6.**
91. **Brusselmans, K., et al.,** ***Heterozygous deficiency of hypoxia-inducible factor-2alpha protects mice against pulmonary hypertension and right ventricular dysfunction during prolonged hypoxia.*** **J Clin Invest, 2003. 111(10): p. 1519-27.**
92. **Hickey, M.M., et al.,** ***The von Hippel-Lindau Chuvash mutation promotes pulmonary hypertension and fibrosis in mice.*** **J Clin Invest, 2010. 120(3): p. 827-39.**
93. **Takeda, K., et al.,** ***Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins.*** **Blood, 2008. 111(6): p. 3229-35.**
94. **Takeda, K., A. Cowan, and G.H. Fong,** ***Essential role for prolyl hydroxylase domain protein 2 in oxygen homeostasis of the adult vascular system.*** **Circulation, 2007. 116(7): p. 774-81.**
95. **Talbot, N.P., et al.** ***A Role For Prolyl Hydroxylase Domain (phd) Proteins In Respiratory Control*****. in** ***American Thoracic Society 2011 International Conference*****. 2011.**
96. **Takeda, K., et al.,** ***Placental but not heart defects are associated with elevated hypoxia-inducible factor alpha levels in mice lacking prolyl hydroxylase domain protein 2.*** **Mol Cell Biol, 2006. 26(22): p. 8336-46.**
97. **Peng, Y.J., et al.,** ***Hypoxia-inducible factor 2alpha (HIF-2alpha) heterozygous-null mice exhibit exaggerated carotid body sensitivity to hypoxia, breathing instability, and hypertension.*** **Proc Natl Acad Sci U S A, 2011. 108(7): p. 3065-70.**
98. **Smith, T.G., et al.,** ***The increase in pulmonary arterial pressure caused by hypoxia depends on iron status.*** **J Physiol, 2008. 586(Pt 24): p. 5999-6005.**
99. **Ren, X., et al.,** ***Effects of desferrioxamine on serum erythropoietin and ventilatory sensitivity to hypoxia in humans.*** **J Appl Physiol, 2000. 89(2): p. 680-6.**
100. **Smith, T.G., et al.,** ***Effects of iron supplementation and depletion on hypoxic pulmonary hypertension: two randomized controlled trials.*** **JAMA, 2009. 302(13): p. 1444-50.**
101. **Wang, G.L. and G.L. Semenza,** ***Desferrioxamine induces erythropoietin gene expression and hypoxia-inducible factor 1 DNA-binding activity: implications for models of hypoxia signal transduction.*** **Blood, 1993. 82(12): p. 3610-5.**
102. **Moore, L.G., et al.,** ***Comparative aspects of high-altitude adaptation in human populations.*** **Adv Exp Med Biol, 2000. 475: p. 45-62.**
103. **Severinghaus, J.W., C.R. Bainton, and A. Carcelen,** ***Respiratory insensitivity to hypoxia in chronically hypoxic man.*** **Respir Physiol, 1966. 1(3): p. 308-34.**
104. **Moore, L.G., S. Niermeyer, and S. Zamudio,** ***Human adaptation to high altitude: regional and life-cycle perspectives.*** **Am J Phys Anthropol, 1998. Suppl 27: p. 25-64.**
105. **Beall, C.M., et al.,** ***Ventilation and hypoxic ventilatory response of Tibetan and Aymara high altitude natives.*** **Am J Phys Anthropol, 1997. 104(4): p. 427-47.**

106. **Beall, C.M., et al.,** ***Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara.*** **Am J Phys Anthropol, 1998. 106(3): p. 385-400.**
107. **Winslow, R.M., et al.,** ***Different hematologic responses to hypoxia in Sherpas and Quechua Indians.*** **J Appl Physiol, 1989. 66(4): p. 1561-9.**
108. **Garruto, R.M., et al.,** ***Hematological differences during growth among Tibetans and Han Chinese born and raised at high altitude in Qinghai, China.*** **Am J Phys Anthropol, 2003. 122(2): p. 171-83.**
109. **Moore, L.G., et al.,** ***Are Tibetans better adapted?*** **Int J Sports Med, 1992. 13 Suppl 1: p. S86-8.**
110. **Groves, B.M., et al.,** ***Minimal hypoxic pulmonary hypertension in normal Tibetans at 3,658 m.*** **J Appl Physiol, 1993. 74(1): p. 312-8.**
111. **Gupta, M.L., et al.,** ***Lack of smooth muscle in the small pulmonary arteries of the native Ladakhi. Is the Himalayan highlander adapted?*** **Am Rev Respir Dis, 1992. 145(5): p. 1201-4.**
112. **Marconi, C., M. Marzorati, and P. Cerretelli,** ***Work capacity of permanent residents of high altitude.*** **High Altitude Medicine & Biology, 2006. 7(2): p. 105-115.**
113. **Moore, L.G.,** ***Maternal O-2 Transport and Fetal Growth in Colorado, Peru, and Tibet High-Altitude Residents.*** **American Journal of Human Biology, 1990. 2(6): p. 627-637.**
114. **Zamudio, S., et al.,** ***Protection from intrauterine growth retardation in Tibetans at high altitude.*** **Am J Phys Anthropol, 1993. 91(2): p. 215-24.**
115. **Beall, C.M., et al.,** ***Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders.*** **Proc Natl Acad Sci U S A, 2010. 107(25): p. 11459-64.**
116. **Peng, Y., et al.,** ***Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas.*** **Mol Biol Evol, 2011. 28(2): p. 1075-81.**
117. **Yi, X., et al.,** ***Sequencing of 50 human exomes reveals adaptation to high altitude.*** **Science, 2010. 329(5987): p. 75-8.**
118. **Xu, S., et al.,** ***A genome-wide search for signals of high-altitude adaptation in Tibetans.*** **Mol Biol Evol, 2011. 28(2): p. 1003-11.**
119. **Simonson, T.S., et al.,** ***Genetic evidence for high-altitude adaptation in Tibet.*** **Science, 2010. 329(5987): p. 72-5.**
120. **Bigham, A., et al.,** ***Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data.*** **PLoS Genet, 2010. 6(9).**
121. **Robbins, P.A., G.D. Swanson, and M.G. Howson,** ***A prediction-correction scheme for forcing alveolar gases along certain time courses.*** **J Appl Physiol, 1982. 52(5): p. 1353-7.**
122. **Howard, L.S., et al.,** ***Chamber for controlling end-tidal gas tensions over sustained periods in humans.*** **J Appl Physiol, 1995. 78(3): p. 1088-91.**
123. **Clement, I.D. and P.A. Robbins,** ***Dynamics of the ventilatory response to hypoxia in humans.*** **Respir Physiol, 1993. 92(3): p. 253-75.**
124. **Severinghaus, J.W.,** ***Simple, accurate equations for human blood O2 dissociation computations.*** **J Appl Physiol, 1979. 46(3): p. 599-602.**
125. **Liang, P.J., J.J. Pandit, and P.A. Robbins,** ***Statistical properties of breath-to-breath variations in ventilation at constant PETCO2 and PETO2 in humans.*** **J Appl Physiol, 1996. 81(5): p. 2274-86.**
126. **Groves, B.M., et al.,** ***Operation Everest II: elevated high-altitude pulmonary resistance unresponsive to oxygen.*** **J Appl Physiol, 1987. 63(2): p. 521-30.**
127. **Yock, P.G. and R.L. Popp,** ***Noninvasive estimation of right ventricular systolic pressure by Doppler ultrasound in patients with tricuspid regurgitation.*** **Circulation, 1984. 70(4): p. 657-62.**

128. Skjaerpe, T. and L. Hatle, *Noninvasive estimation of systolic pressure in the right ventricle in patients with tricuspid regurgitation.* Eur Heart J, 1986. 7(8): p. 704-10.
129. Peacock, A.J., V. Challenor, and G. Sutherland, *Estimation of pulmonary artery pressure by Doppler echocardiography in normal subjects made hypoxic.* Respir Med, 1990. 84(4): p. 335-7.
130. Chemla, D., et al., *New formula for predicting mean pulmonary artery pressure using systolic pulmonary artery pressure.* Chest, 2004. 126(4): p. 1313-7.
131. Allemann, Y., et al., *Echocardiographic and invasive measurements of pulmonary artery pressure correlate closely at high altitude.* Am J Physiol Heart Circ Physiol, 2000. 279(4): p. H2013-6.
132. Leon-Velarde, F., et al., *Alveolar Pco2 and Po2 of high-altitude natives living at sea level.* J Appl Physiol, 1996. 81(4): p. 1605-9.
133. Zhuang, J., et al., *Hypoxic ventilatory responsiveness in Tibetan compared with Han residents of 3,658 m.* J Appl Physiol, 1993. 74(1): p. 303-11.
134. Vargas, M., et al., *Similar hypoxic ventilatory responses in sea-level natives and high-altitude Andean natives living at sea level.* J Appl Physiol, 1998. 84(3): p. 1024-9.
135. Gamboa, A., et al., *Selected contribution: Acute and sustained ventilatory responses to hypoxia in high-altitude natives living at sea level.* J Appl Physiol, 2003. 94(3): p. 1255-62; discussion 1253-4.
136. Rivera-Ch, M., et al., *Selected contribution: High-altitude natives living at sea level acclimatize to high altitude like sea-level natives.* J Appl Physiol, 2003. 94(3): p. 1263-8; discussion 1253-4.
137. Sorensen, S.C. and Severing.Jw, *Irreversible Respiratory Insensitivity to Acute Hypoxia in Man Born at High Altitude.* Journal of Applied Physiology, 1968. 25(3): p. 217-&.
138. Velasque.T, et al., *Ventilatory Effects of Oxygen in High Altitude Natives.* Respiration Physiology, 1968. 5(1): p. 211-&.
139. Lahiri, S., et al., *Irreversible Blunted Respiratory Sensitivity to Hypoxia in High Altitude Natives.* Respiration Physiology, 1969. 6(3): p. 360-&.
140. Chambers, J.C., et al., *Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels.* Nat Genet, 2009. 41(11): p. 1170-2.
141. Soranzo, N., et al., *A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium.* Nat Genet, 2009. 41(11): p. 1182-90.
142. Benyamin, B., et al., *Common variants in TMPRSS6 are associated with iron status and erythrocyte volume.* Nat Genet, 2009. 41(11): p. 1173-5.
143. Rankin, E.B., et al., *Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo.* J Clin Invest, 2007. 117(4): p. 1068-77.
144. Semenza, G.L. and G.L. Wang, *A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation.* Mol Cell Biol, 1992. 12(12): p. 5447-54.
145. Gruber, M., et al., *Acute postnatal ablation of Hif-2alpha results in anemia.* Proc Natl Acad Sci U S A, 2007. 104(7): p. 2301-6.
146. Scortegagna, M., et al., *The HIF family member EPAS1/HIF-2alpha is required for normal hematopoiesis in mice.* Blood, 2003. 102(5): p. 1634-40.
147. Mastrogiannaki, M., et al., *HIF-2alpha, but not HIF-1alpha, promotes iron absorption in mice.* J Clin Invest, 2009. 119(5): p. 1159-66.
148. Sime, F., D. Penaloza, and L. Ruiz, *Bradycardia, increased cardiac output, and reversal of pulmonary hypertension in altitude natives living at sea level.* Br Heart J, 1971. 33(5): p. 647-57.
149. Lorenzo, F.P.V., et al., *The Tibetan PHD2 Polymorphism Asp4Glu Is Associated with Hypersensitivity of Erythroid Progenitors to EPO and Upregulation of HIF-1*

Regulated Genes Hexokinase (HK1) and Glucose Transporter 1 (SLC2A/GLUT1), in *53rd ASH Annual Meeting and Exposition*. 2011.
150. Lorenzo, F.R.V., et al., *A Novel PHD2 Mutation Associated with Tibetan Genetic Adaptation to High Altitude Hypoxia*, in *52nd ASH Annual Meeting and Exposition*. 2010.
151. Enard, W., et al., *Intra- and interspecific variation in primate gene expression patterns.* Science, 2002. 296(5566): p. 340-3.
152. Goring, H.H., et al., *Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes.* Nat Genet, 2007. 39(10): p. 1208-16.
153. Cheung, V.G., et al., *Natural variation in human gene expression assessed in lymphoblastoid cells.* Nat Genet, 2003. 33(3): p. 422-5.
154. Monks, S.A., et al., *Genetic inheritance of gene expression in human cell lines.* Am J Hum Genet, 2004. 75(6): p. 1094-105.
155. Whitney, A.R., et al., *Individuality and variation in gene expression patterns in human blood.* Proc Natl Acad Sci U S A, 2003. 100(4): p. 1896-901.
156. Spielman, R.S., et al., *Common genetic variants account for differences in gene expression among ethnic groups.* Nat Genet, 2007. 39(2): p. 226-31.
157. Emilsson, V., et al., *Genetics of gene expression and its effect on disease.* Nature, 2008. 452(7186): p. 423-8.
158. Schultz, A., et al., *Interindividual heterogeneity in the hypoxic regulation of VEGF: significance for the development of the coronary artery collateral circulation.* Circulation, 1999. 100(5): p. 547-52.
159. Brooks, J.T., et al., *Variations within oxygen-regulated gene expression in humans.* J Appl Physiol, 2009. 106(1): p. 212-20.
160. Gao, W., et al., *Hypoxia-induced expression of HIF-1alpha and its target genes in umbilical venous endothelial cells of Tibetans and immigrant Han.* Comp Biochem Physiol C Toxicol Pharmacol, 2005. 141(1): p. 93-100.
161. *Application Note: Amplification Efficiency of Taqman Expression Assays*. Available from: docs.appliedbiosystems.com/pebiodocs/00113186.pdf.
162. Biosystems, A., *TaqMan Gene Expression Assays Protocol.*
163. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. 102(30): p. 10604-9.
164. Kaminsky, Z.A., et al., *DNA methylation profiles in monozygotic and dizygotic twins.* Nat Genet, 2009. 41(2): p. 240-5.
165. Youngson, N.A. and E. Whitelaw, *Transgenerational epigenetic effects.* Annu Rev Genomics Hum Genet, 2008. 9: p. 233-57.
166. Watson, J.A., et al., *Epigenetics, the epicenter of the hypoxic response.* Epigenetics, 2010. 5(4): p. 293-6.
167. Pastinen, T., B. Ge, and T.J. Hudson, *Influence of human genome polymorphism on gene expression.* Hum Mol Genet, 2006. 15 Spec No 1: p. R9-16.
168. Brooks, J.T.S., *Variations within Oxygen-Regulated Gene Expression in Humans*, in *Department of Physiology, Anatomy and genetics*. 2006.
169. Rossignol, F., C. Vache, and E. Clottes, *Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues.* Gene, 2002. 299(1-2): p. 135-40.
170. Thrash-Bingham, C.A. and K.D. Tartof, *aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia.* J Natl Cancer Inst, 1999. 91(2): p. 143-51.
171. Biosystems, A., *TaqMan SNP Genotyping Assays Protocol.* docs.appliedbiosystems.com/pebiodocs/04332856.pdf.
172. Plath, K., et al., *Xist RNA and the mechanism of X chromosome inactivation.* Annu Rev Genet, 2002. 36: p. 233-78.
173. Luedi, P.P., et al., *Computational and experimental identification of novel human imprinted genes.* Genome Res, 2007. 17(12): p. 1723-30.

174. **Knight, J.C., *Allele-specific gene expression uncovered.* Trends Genet, 2004. 20(3): p. 113-6.**

175. **Stockmann, C. and J. Fandrey, *Hypoxia-induced erythropoietin production: a paradigm for oxygen-regulated gene expression.* Clin Exp Pharmacol Physiol, 2006. 33(10): p. 968-79.**

176. **Kleinjan, D.A. and V. van Heyningen, *Long-range control of gene expression: emerging mechanisms and disruption in disease.* Am J Hum Genet, 2005. 76(1): p. 8-32.**

177. **West, A.G. and P. Fraser, *Remote control of gene transcription.* Hum Mol Genet, 2005. 14 Spec No 1: p. R101-11.**

178. **Dekker, J., *Gene regulation in the third dimension.* Science, 2008. 319(5871): p. 1793-4.**

179. **D'Haene, B., et al., *Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening.* PLoS Genet, 2009. 5(6): p. e1000522.**

180. **Kleinjan, D.A., et al., *Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6.* Hum Mol Genet, 2001. 10(19): p. 2049-59.**

181. **Gheldof, N., et al., *Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene.* Nucleic Acids Res, 2010. 38(13): p. 4325-36.**

182. **Adam, J., et al., *Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling.* Cancer Cell, 2011. 20(4): p. 524-37.**

183. **Tian, Y.M., et al., *Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors.* J Biol Chem, 2011. 286(15): p. 13041-51.**

184. **Hirsila, M., et al., *Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor.* J Biol Chem, 2003. 278(33): p. 30772-80.**

185. **Koivunen, P., et al., *Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases.* J Biol Chem, 2004. 279(11): p. 9899-904.**

186. **Fandrey, J., T.A. Gorr, and M. Gassmann, *Regulating cellular oxygen sensing by hydroxylation.* Cardiovasc Res, 2006. 71(4): p. 642-51.**

187. **Flashman, E., et al., *Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen.* FEBS J, 2010. 277(19): p. 4089-99.**

188. **McDonough, M.A., et al., *Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2).* Proc Natl Acad Sci U S A, 2006. 103(26): p. 9814-9.**

189. **Mikhaylova, O., et al., *The von Hippel-Lindau tumor suppressor protein and Egl-9-Type proline hydroxylases regulate the large subunit of RNA polymerase II in response to oxidative stress.* Mol Cell Biol, 2008. 28(8): p. 2701-17.**

190. **Kuznetsova, A.V., et al., *von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination.* Proc Natl Acad Sci U S A, 2003. 100(5): p. 2706-11.**

191. **Aldenderfer, M.S., *Moving up in the world.* American Scientist, 2003. 91(6): p. 542-549.**

192. **Zhao, M., et al., *Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau.* Proceedings of the National Academy of Sciences of the United States of America, 2009. 106(50): p. 21230-21235.**

193. **Kwiatkowski, D.P., *How malaria has affected the human genome and what human genetics can teach us about malaria.* Am J Hum Genet, 2005. 77(2): p. 171-92.**

194. **Flint, J., et al., *The population genetics of the haemoglobinopathies.* Baillieres Clin Haematol, 1993. 6(1): p. 215-62.**
195. **Tishkoff, S.A., et al., *Convergent adaptation of human lactase persistence in Africa and Europe.* Nat Genet, 2007. 39(1): p. 31-40.**
196. **Leon-Velarde, F., et al., *Consensus statement on chronic and subacute high altitude diseases.* High Alt Med Biol, 2005. 6(2): p. 147-57.**
197. **McMullin, M.F., *The classification and diagnosis of erythrocytosis.* Int J Lab Hematol, 2008. 30(6): p. 447-59.**
198. **McMullin, M.F., *Idiopathic erythrocytosis: a disappearing entity.* Hematology Am Soc Hematol Educ Program, 2009: p. 629-35.**
199. **Lemarchandel, V., et al., *Compound heterozygosity in a complete erythrocyte bisphosphoglycerate mutase deficiency.* Blood, 1992. 80(10): p. 2643-9.**
200. **Pritlove, D.C., et al., *Novel placental expression of 2,3-bisphosphoglycerate mutase.* Placenta, 2006. 27(8): p. 924-7.**
201. **Schroter, W., *[Congenital non-spherocytic hemolytic anemia by 2,3-diphosphoglycerate mutase deficiency of the erythrocytes in early infancy].* Klin Wochenschr, 1965. 43(21): p. 1147-53.**
202. **Labie, D., et al., *Familial diphosphoglyceratemutase deficiency. Influence on the oxygen affinity curves of hemoglobin.* FEBS Lett, 1970. 9(1): p. 37-40.**
203. **Rosa, R., et al., *The first case of a complete deficiency of diphosphoglycerate mutase in human erythrocytes.* J Clin Invest, 1978. 62(5): p. 907-15.**
204. **Rosa, R., et al., *Isolation, characterization, and structure of a mutant 89 Arg----Cys bisphosphoglycerate mutase. Implication of the active site in the mutation.* J Biol Chem, 1989. 264(14): p. 7837-43.**
205. **Garel, M.C., et al., *Natural and artificial mutants of the human 2,3-bisphosphoglycerate as a tool for the evaluation of structure-function relationships.* Biomed Biochim Acta, 1990. 49(2-3): p. S166-71.**
206. **Hoyer, J.D., et al., *Erythrocytosis due to bisphosphoglycerate mutase deficiency with concurrent glucose-6-phosphate dehydrogenase (G-6-PD) deficiency.* Am J Hematol, 2004. 75(4): p. 205-8.**
207. **Galson, D.L., et al., *Comparison of the human and mouse erythropoietin genes shows extensive homology in the flanking regions.* Blood, 1993. 82(11): p. 3321-6.**
208. **Semenza, G.L., et al., *Polycythemia in transgenic mice expressing the human erythropoietin gene.* Proc Natl Acad Sci U S A, 1989. 86(7): p. 2301-5.**
209. **Semenza, G.L., et al., *Human erythropoietin gene expression in transgenic mice: multiple transcription initiation sites and cis-acting regulatory elements.* Mol Cell Biol, 1990. 10(3): p. 930-8.**
210. **Semenza, G.L., et al., *Cell-type-specific and hypoxia-inducible expression of the human erythropoietin gene in transgenic mice.* Proc Natl Acad Sci U S A, 1991. 88(19): p. 8725-9.**
211. **Yin, H. and K.L. Blanchard, *DNA methylation represses the expression of the human erythropoietin gene by two different mechanisms.* Blood, 2000. 95(1): p. 111-9.**
212. **Rondon, I.J., et al., *Hypoxia up-regulates the activity of a novel erythropoietin mRNA binding protein.* J Biol Chem, 1991. 266(25): p. 16594-8.**
213. **Park, D., et al., *[Removal of the 3'- and 5'-untranslated region amplifies expression of the erythropoietin gene in mammalian cells].* Mol Biol (Mosk), 2001. 35(3): p. 413-6.**
214. **Theil, E.C., *Iron regulatory elements (IREs): a family of mRNA non-coding sequences.* Biochem J, 1994. 304 (Pt 1): p. 1-11.**
215. **Klausner, R.D., T.A. Rouault, and J.B. Harford, *Regulating the fate of mRNA: the control of cellular iron metabolism.* Cell, 1993. 72(1): p. 19-28.**

216. **Gray, N.K. and M.W. Hentze, *Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs.* EMBO J, 1994. 13(16): p. 3882-91.**

217. **Allerson, C.R., M. Cazzola, and T.A. Rouault, *Clinical severity and thermodynamic effects of iron-responsive element mutations in hereditary hyperferritinemia-cataract syndrome.* J Biol Chem, 1999. 274(37): p. 26439-47.**

218. **Sanchez, M., et al., *Iron-regulatory proteins limit hypoxia-inducible factor-2alpha expression in iron deficiency.* Nat Struct Mol Biol, 2007. 14(5): p. 420-6.**

219. **Fleming, R.E., et al., *Mechanism of increased iron absorption in murine model of hereditary hemochromatosis: increased duodenal expression of the iron transporter DMT1.* Proc Natl Acad Sci U S A, 1999. 96(6): p. 3143-8.**

220. **Wiestner, A., et al., *An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia.* Nat Genet, 1998. 18(1): p. 49-52.**

221. **Kondo, T., et al., *Familial essential thrombocythemia associated with one-base deletion in the 5'-untranslated region of the thrombopoietin gene.* Blood, 1998. 92(4): p. 1091-6.**

222. **Ghilardi, N. and R.C. Skoda, *A single-base deletion in the thrombopoietin (TPO) gene causes familial essential thrombocythemia through a mechanism of more efficient translation of TPO mRNA.* Blood, 1999. 94(4): p. 1480-2.**

223. **Jacobs, K., et al., *Isolation and characterization of genomic and cDNA clones of human erythropoietin.* Nature, 1985. 313(6005): p. 806-10.**

224. **Shoemaker, C.B. and L.D. Mitsock, *Murine erythropoietin gene: cloning, expression, and human gene homology.* Mol Cell Biol, 1986. 6(3): p. 849-58.**

225. **Saleque, S., et al., *Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1.* Mol Cell, 2007. 27(4): p. 562-72.**

226. **Saleque, S., S. Cameron, and S.H. Orkin, *The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages.* Genes Dev, 2002. 16(3): p. 301-6.**

227. **van Haaften, G., et al., *Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer.* Nat Genet, 2009. 41(5): p. 521-3.**

228. **Dalgliesh, G.L., et al., *Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes.* Nature, 2010. 463(7279): p. 360-3.**

229. **Miyake, N., et al., *KDM6A Point Mutations Cause Kabuki Syndrome.* Hum Mutat, 2012.**

230. **Lederer, D., et al., *Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome.* Am J Hum Genet, 2012. 90(1): p. 119-24.**

231. **Montagner, M., et al., *SHARP1 suppresses breast cancer metastasis by promoting degradation of hypoxia-inducible factors.* Nature, 2012. 487(7407): p. 380-4.**

232. **Ingram, D.A., et al., *Identification of a novel hierarchy of endothelial progenitor cells using human peripheral and umbilical cord blood.* Blood, 2004. 104(9): p. 2752-60.**

233. **Toshner, M., et al., *Evidence of dysfunction of endothelial progenitors in pulmonary arterial hypertension.* Am J Respir Crit Care Med, 2009. 180(8): p. 780-7.**

234. **Purdue, M.P., et al., *Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3.* Nat Genet, 2011. 43(1): p. 60-5.**

235. **Han, S.S., et al., *The chromosome 2p21 region harbors a complex genetic architecture for association with risk for renal cell carcinoma.* Hum Mol Genet, 2012. 21(5): p. 1190-200.**

APPENDIX

Reagents:

1. Water

Water was purified using an Elix^R 10 System (*Millipore*) and subsequently purified using a MilliQ System (*Millipore*) with a 0.22 μm filter at the outlet. "MilliQ-grade" water was used for all buffers and solutions. Nuclease-free water was purchased from *Sigma Aldrich*.

2. Buffers

- **Dilution Buffer:** 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 15.7 mM Tris-HCl/pH 8.1, 167 mM NaCl
- **Elution Buffer:** 1% SDS, 0.1 M $NaHCO_3$
- **Hypotonic Extraction Buffer (HEB):** 20 mM Tris/pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$
- **High Salt (wash) Buffer:** 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl/pH 8.1, 500 mM NaCl
- **IgePal Lysis Buffer:** 10 mM Tris/pH 7.6, 0.25 M NaCl, 0.5% Igepal
- **LiCl (wash) Buffer:** 0.25 M LiCl, 1% NP-40, 1% sodium deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl/pH 8.1
- **Low Salt (wash) Buffer:** 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl/pH 8.1, 150 mM NaCl
- **Phosphate-buffered saline (PBS):** 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2PO_4.7H_2O$, 2 mM KH_2PO_4
- **PBST (10×):** 1% Tween-20, 13.7 mM NaCl, 27 mM KCl, 100 mM $Na_2PO_4.7H_2O$, 20 mM KH_2PO_4
- **SDS Lysis Buffer:** 1% SDS, 10 mM EDTA, 50 mM Tris-HCl /pH 8.1
- **SDS/PAGE Running Buffer (10×):** 250 mM Tris-base, 1.92 M glycine, 1% SDS
- **TE Buffer:** 10 mM Tris-HCl/pH 8, 1 mM EDTA
- **Transfer Buffer (20×):** 0.2 mM Tris-base, 2M Glycine

- **Transfer Buffer (1×):** 10% methanol (*Sigma*), 5% Transfer Buffer (20×)
- **Tris-Borate EDTA (TBE) Buffer (1×):** 10 × TBE (*Sigma)* was diluted to 1× with water
- **Tris-HCl/SDS/pH 8.8 Buffer (4×):** 1.5 M Tris/HCl, 0.4% SDS
- **Tris-HCl/SDS/pH 6.8 Buffer (4×):** 0.5 M Tris/Hcl, 0.4% SDS

3. Gels

- **SDS-PAGE Gel:**

  Separating Gel (9% acrylamide): 1.5 ml 30% acrylamide/0.8% bisacrylamide (*Bio-rad Laboratories*), 1.25 ml 4×Tris-HCl/SDS pH 8.8, 2.25 ml water, 17 µl 10% APS (ammonium persulfate solution) , 3.4 µl TEMED (tetramethylethylenediamine) (*Sigma*)

  Stacking Gel (3.9 % acrylamide): 0.65 ml 30% acrylamide/0.8% bisacrylamide (*Bio-rad Laboratories*), 1.25 ml 4×Tris-HCl/SDS pH 6.8, 3.05 ml water, 33 µl 10% APS, 6.7 µl TEMED.

4. TaqMan Gene expression assays:

- Vascular Endothelial Growth Factor A (VEGFA): Hs00173626_m1
- Aldolase C (ALDC): Hs00193059_m1
- Adrenomedullin (ADM): Hs00181605_m1
- Prolyl-4-Hydroxylase-Alpha1 (P4HA1): Hs00168575_m1
- N-myc downstream regulated 1 (NDRG1): Hs00608387_m1
- EPAS1/HIF2A: Hs01026149_m1
- EGLN1: Hs00254392_m1
- HIF1A: Hs00153153_m1
- B2 microglobulin (B2M) endogenous control: Hs99999907_m1

5. TaqMan SNP genotyping assays (for allele-specific expression):

- rs59901247: C_89911839_10
- SNP rs1447563: C_2163022_10