Whole blood and peripheral mononuclear cell transcriptional response to prolonged altitude exposure in well-trained runners

Shaun Sutehall[1], Fernanda Malinsky[2], Jonathan Shurlock[3], Guan Wang[4], Andrew Bosch[1] and Yannis Pitsiladis[2, 5-7]

[1]Division of Exercise Science and Sport Medicine, University of Cape Town, Cape Town, South Africa; [2]Centre for Stress and Age-related Disease, University of Brighton, Brighton, United Kingdom; [3]Somerset NHS Foundation Trust, Taunton, United Kingdom [4]Sport and Exercise Science and Sports Medicine Research and Enterprise Group, University of Brighton, Brighton, United Kingdom; [5]Centre for Exercise Sciences and Sports Medicine, FIMS Collaborating Centre of Sports Medicine, Rome, Italy; [6]European Federation of Sports Medicine Associations (EFSMA), Lausanne, Switzerland; [7]International Federation of Sports Medicine (FIMS), Lausanne, Switzerland.

Corresponding Author:
Professor Yannis P Pitsiladis
Centre for Stress and Age-related Disease
University of Brighton,
Brighton
United Kingdom
Email: Y.Pitsiladis@brighton.ac.uk

Abstract

Recombinant human erythropoietin (rHuEpo) abuse by athletes threatens the integrity of sport. Due to the overlap in physiological response to rHuEpo and altitude exposure, it remains difficult to differentiate changes in haematological variables caused by rHuEpo or altitude and therefore, other molecular methods to enhance anti-doping should be explored.

Objective: To identify the haematological and transcriptomic response to prolonged altitude exposure typical of practices used by elite athletes.

Design: Longitudinal study

Setting: University of Cape Town and Altitude Training Centre in Ethiopia

Participants and Intervention: Fourteen well-trained athletes sojourned to an altitude training camp in Sulutla, Ethiopia (~2400 – 2500 m above sea level) for 27 days. Blood samples were taken prior, 24 hr, 9, 16 and 24 days after arrival at altitude in addition to 24 hr, 6, 13 and 27 days upon return to sea level.

Main Outcome Measures: Blood samples were analysed for haemoglobin concentration, haematocrit and reticulocyte percentage. The transcriptomic response in whole blood and peripheral blood mononuclear cells (PBMC) were analysed using gene expression microarrays.

Results: A unique set of 29 and 10 genes were identified to be commonly expressed at every altitude timepoint in whole blood and PBMC, respectively. There were no genes identified upon return to sea level in whole blood and only one gene within PBMC.

Conclusions: The current study has identified a series of unique genes that can now be integrated with genes previously validated for rHuEpo abuse, thereby enabling the differentiation of rHuEpo from the altitude confounder.

Introduction

Following the dominance of East African athletes in endurance sports in the 1968 Mexico Olympic Games (~2300 m above sea level), the effects of altitude on endurance performance became an area of intense research. Sojourning to altitude has become common practice with the expectation of gaining the same advantage as the East African athletes. When travelling to altitude, decreased pressure of inspired oxygen (PiO_2) reduces arterial blood saturation (SaO_2) thereby challenging oxidative metabolism (1). A number of physiological adaptations take place in response to altitude exposure to minimise the disruption caused by the decreased PiO_2. Short term adaptations to altitude include a decreased plasma volume, increased heart rate and an increased ventilation rate (2). Longer-term adaptations follow that include an increase in endogenous erythropoietin (Epo) production that stimulates red blood cell production (3). It is this primary adaptation that is believed to enhance exercise performance at sea level (4). However, the ergogenic effects of altitude training in elite athletes has recently been challenged (5).

Initial research into altitude training aimed to replicate the training habits of the East African athletes who predominately live at an altitude in excess of ~2000 m above sea level and often train at, or higher altitudes (6). Early investigations into the performance benefits of living and training at high altitude (LHTH) generated conflicting results. Some studies demonstrated an 8-10% increase in sea level maximal oxygen uptake ($\dot{V}O_2max$) and improved time trial performance (7), while others reported no difference (8). In response, Levine and Stray-Gundersen introduced the novel concept of living at moderate altitude (2500 m above sea level) while training at low altitude (<1500 m above sea level, LHTL) in order to attenuate the decrease in training intensity seen in previous LHTH studies (9). It was found that both training paradigms (LHTL and LHTH) led to an improvement in $\dot{V}O_2max$ but only LHTL led to a significant increase in sea level 5000 m time trial performance. The LHTL paradigm has remained the most popular method of altitude training, with minor modifications on the design such as performing all but high intensity training sessions at "high" altitude to maximise physiological adaptation. Despite a plethora of studies on altitude training, the scientific literature has been deemed not strong enough to make firm recommendations regarding altitude training's effectiveness (10). Nonetheless, altitude training has continued to be an extremely popular training method used by a large number of elite athletes.

The drug recombinant human erythropoietin (rHuEpo) increases red blood cell production and therefore oxygen carrying capacity of blood is increased, a process similar to the response to altitude exposure. The use of rHuEpo has been banned by the World Anti-Doping Agency due to its well-known performance enhancing effects (11–13). The efficacy of rHuEpo to increase red blood cell production has been shown to be superior compared with simulated altitude exposure (14) and also expected to be greater compared to real altitude exposure. Current blood doping anti-doping detection methods are inadequate in detecting small, or "microdoses" of rHuEpo (15, 16). In recent years, there has been a rapid development of Omics technologies which involve the use of the transcriptional, translational or epigenomic response to a given stimulus or environment. We have previously demonstrated the potential for a transcriptomic approach to enhance the detection of blood doping by identifying 34 robust transcripts in response to high dose rHuEpo administration (17). We have also validated this transcriptomic signature using microdoses of rHuEpo (18). In this study (18), an initial investigation was also performed to identify the effects of altitude exposure on transcriptomic markers. Twenty-one elite runners sojourned to Sierra Nevada, Spain (2320 m above sea level) for 2-3 weeks. After 10 days at 2320 m above sea level, 13 genes tended to be significantly up-regulated, and a trend towards down-regulation of 20 genes one week after return from altitude was observed in the 21 elite runners. In the same study, 4 elite rowers also sojourned to Santa Caterina, Italy (1850 m above sea level) for 2 weeks but there were no differentially expressed transcripts following altitude exposure. Insufficient altitude dose (i.e., height and duration) in addition to the low number of athletes may account for the lack of significant changes seen in these two groups of athletes. Further investigations of the transcriptomic response to altitude in modes commonly used by athletes is required, which may aid in the development of biomarkers of altitude exposure and allow a quantitative method of differentiating adaptations gained through altitude exposure and those gained through rHuEpo. Therefore, the aim of this study was to identify the haematological and transcriptomic response to prolonged altitude exposure and in particular, altitude dose typical of practices used by elite athletes. At this early discovery stage, no specific hypothesis can be formulated as the use of microarray analysis is untargeted and considered a "hypothesis free" or an "agnostic" methodological approach. Rather than focusing on biological candidates, the transcriptome is screened without any prior selection for specific genes or variants. This interrogation of the entire transcriptome has numerous advantages such as overcoming the challenges imposed by the incomplete understanding of physiology, in this case altitude physiology.

Methods

Participants: Fifteen endurance trained athletes, regularly competing at local and national level and training at least six times per week were recruited from Cape Town, South Africa (13 males, 2 females, age: 25±4 yrs; height: 169±8 cm; weight: 56±8 kg; $\dot{V}O_2max$: 72.8±5.9 $mL \cdot kg \cdot min^{-1}$). None of the participants had been to altitude above 1700 m in the previous 6 months prior the commencement of this study. All participants provided written informed consent prior to testing in accordance with the Declaration of Helsinki and approved protocol by the University of Cape Town Ethics Committee (Cape Town, South Africa).

Study Design

Preliminary testing: Prior to data collection, serum ferritin levels were assessed 3 weeks before the planned departure. If blood ferritin levels were below 40 $\mu g \cdot L^{-1}$, 200 mg of elemental iron (ferrous sulphate) was recommended to be consumed twice daily for 3 weeks, if after the administration blood ferritin remained below 40 $\mu \cdot L^{-1}$, the participant was excluded from the study due to the importance of adequate blood ferritin levels to facilitate red blood cell production and subsequent adaptation to altitude. Three participants required iron supplementation, one of which did not meet the cut off after iron supplementation and was excluded from the study (n=14 participants were included in the study, 12 males, 2 females).

Participants provided blood samples four days and one day before departure, blood was taken from an antecubital vein and collected into 6 mL K_3EDTA (four days prior to departure only, Greiner Bio-One Ltd, Stonehouse, UK) and 3 mL Tempus™ Blood RNA tubes (Life Technologies, Carlsbad, CA, USA) after remaining in the supine position for a minimum of 10 min. Blood samples were then analysed for haemoglobin concentration, (HGB), haematocrit (HCT) and reticulocyte percentage (RET%) using Beckman Coulter DXH800 (Beckam Coulter, Inc, Cape Town, South Africa). The Tempus tubes were mixed vigorously for 15 s immediately after collection, incubated at room temperature for 3 hours and then stored at -80°C before RNA extraction. PBMCs were collected in CPT Vacutainer, inverted 10 times and spun at room temperature (RT-18-25°C) for 20 minutes at 1600 RCF within 1 hr of collection. Tubes were spun for an additional 5 min, if needed, for adequate cell layer separation. The cell layer containing PBMCs was collected using a sterile Pasteur pipette, washed twice in a 1 X Phosphate Buffer Solution and spun at 300 RCF for 15 min at RT. The cells were resuspended in 1 mL of pre-aliquoted RNALater solution (Thermo Fisher

Scientific, Wilmington, DE, USA) according to manufacturers' instructions and placed on ice for 10 minutes before being frozen and stored at -80°C, until further analysis.

Altitude exposure: Once preliminary testing was completed, participants flew to Addis Ababa, Ethiopia (~2400 m above sea level). Participants remained here for 24 hrs and then sojourned to a training camp in Sulutla, Ethiopia (~2500 m above sea level). At 24 hrs (D1), 9 (D2), 16 (D3) and 24 days (D4) after arrival at altitude, participants were transported to a laboratory in Addis Ababa to provide blood samples. Blood was collected into 4 mL K_2EDTA (Greiner Bio-One Ltd, Stonehouse, UK) and 3 mL Tempus Blood RNA tubes which were stored at -20°C for subsequent analysis. Blood samples were taken after remaining in the supine position for at least 10 min prior to blood sampling which was performed while seated. Blood samples were analysed for HGB, HCT and RET% using ABBOT Cell DYE Ruby (ABBOT, Illinois, USA), due to practical difficulties, RET% could not be assessed 24 hrs after arrival and blood samples were analysed using Sysmex XS-500i (Sysmex cooperation, Kobe, Japan). After blood analysis, participants were transported back to the training camp where they would remain until the next laboratory visit. On the final laboratory visit, participants remained in Addis Ababa for three days and subsequently flew to Cape Town, South Africa after spending 27 days at altitude (~2400-2500 m above sea level).

Sea level testing: Participants reported to the laboratories 24 hrs (Post1), 6 (Post), 13 (Post3) and 27 days (Post4) after return from altitude. On these days, participants provided blood samples. The sampling protocol and analysis equipment used was identical to the preliminary testing. An incremental running test was also performed 3, 8, 15 and 26 days after return from altitude. One participant withdrew from the study after Post1 for reasons unrelated to the study.

RNA extraction and gene expression analysis: Whole blood RNA from the Tempus tubes was isolated from Tempus tubes following the manufacturers' instructions (Tempus^{TM} Spin RNA Isolation Kit, Life Technologies, Carlsbad, CA, USA). PBMC RNA was isolated using RNeasy Mini kit, following the manufacturers' instructions (RNeasy Mini Kit, Qiagen, Hilden, Germany). Purified RNA was eluted in 90 μL elution buffer and stored in three aliquots at -80°C until subsequent analysis. RNA quality was assessed using the Nanodrop® ND-2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). RNA

integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). One hundred nanogram of total RNA was used to process the Affymetrix GeneChip® Human Transcriptome 2.0 Array using the GeneChip® WT Plus Reagent Kit according to the manufacturer's instructions (Affymetrix, Thermo Fisher Scientific, Santa Clara, CA, USA). Single-stranded cDNA (ss-cDNA) was synthesized by the reverse transcription of complimentary RNA (cRNA). 200µL hybridization cocktail (containing approximately 5.2µg fragmented and labelled single-stranded complimentary DNA targets) was loaded to the GeneChip® Human Transcriptome Array 2.0 (Affymetrix, Thermo Fisher Scientific, Santa Clara, CA, USA). The GeneChip arrays were incubated in the GeneChip Hybridization Oven 645 for 16 hrs, washed and stained on the GeneChip Fluidics Station 450. Subsequently, arrays were scanned using the GeneChip® Scanner 3000 7G. Expression Console (Affymetrix, Thermo Fisher Scientific, Santa Clara, CA, USA) was used to perform initial data quality control and visualisation of the data generated following array scanning. Due to technical issues, PBMCs could not be collected at D1 (24 hrs after arrival at altitude). Following the analysis of the whole blood transcriptomic response to altitude, the sample collected four days prior to the altitude sojourn differed significantly from the sample collected 24 hrs prior to the sojourn. Upon further investigation, this trend was not observed within the PBMC samples; for consistent comparisons across tissue types and to maximise baseline data reliability, this timepoint was excluded from the analysis and "baseline" will refer to samples collected 24 hrs prior to departure for all transcriptomic analysis and four days prior to departure for haematological parameters.

Statistical analysis:

Following Shapiro-Wilk test for normality, $\dot{V}O_2max$ and haematological parameters were analysed via a one-way ANOVA. If significant, a Bonferroni multiple comparison test was performed to determine which timepoints significantly differed from baseline with significance set at $p \leq 0.05$. Statistical analysis of these variables was performed using GraphPad (Version 9.0.0; GraphPad Software Inc, La Jolla, CA, USA). The Bioconductor package "oligo" within the program R (RStudio, Version 1.2.5042, R Foundation for Statistical Computing, Vienna, Austria) was employed to read in the intensity CEL files, and the RMA function was used for background correction, normalisation, calculating expression and creating the expression dataset for further analysis. Following background correction and normalisation, using R, the Limma package was employed to perform differential expression

analysis (19, 20). Those differentially expressed transcripts exceeding a fold change of 1.2 were reported at 5% FDR.

Gene set enrichment analysis was performed using GSEA (version 4.1.0, Broad Institute Inc, CA, USA, (21)), following recommended procedures where appropriate (22). The gene set enrichment analysis was performed by examining the Molecular Signatures Database's Hallmark (50 gene sets, (23)) and Gene Ontology Biological Process (7573 gene sets, (24, 25)) collections of gene sets. Significant gene pathways of interested were determined using a nominal $p \leq 0.05$ and an FDR of 10%. If multiple significant pathways were discovered, a biological network map was created using pathways exceeding $p \leq 0.05$, FDR of 10%, a Jaccard overlap coefficient of >0.375 and a combined constant k=0.5 using EnrichmentMap (26) and AutoAnnotate (27) within the Cytoscape software (28).

Results

Physiological markers

HGB was significantly increased 9 days after ascent to altitude compared with baseline (15.4±0.8 vs 14.0±0.7 vs $g \cdot dL^{-1}$, $p<0.01$, respectively). HGB remained significantly elevated throughout the altitude exposure and 24 hrs after returning from altitude (15.1±0.5 $g \cdot dL^{-1}$, $p<0.01$) but did not differ from baseline thereafter (Figure 1A). After ascent to altitude, HCT increased significantly 24 hrs after arrival and increased continually until reaching peak values 24 days after arrival at altitude (41.4±2.1% vs 47.0±2.8%, $p<0.01$, respectively) (Figure 1B). Upon return to sea level, HCT declined but remained significantly higher than baseline throughout the duration of the study (Figure 1B). While a one-way ANOVA indicated RET% significantly differed between time points ($p=0.02$), multiple comparisons failed to reveal any significant differences compared with baseline ($p>0.05$, Figure 1C). $\dot{V}O_2max$ did not change throughout the duration of altitude exposure or return to sea level.

Whole blood

A total of 135 samples were collected over the duration of the study (96% of planned sample collection). Compared with baseline, a unique series of genes was discovered throughout the altitude exposure (Supplementary table 1). Specifically, compared with baseline there were 138 transcripts, representing 117 genes (24 up, 93 down) differentially expressed at D1, 263 transcripts representing 229 genes (179 up, 50 down) at D2, 372 transcripts representing 316

genes (222 up, 94 down) at D3 and 554 transcripts representing 478 genes (129 up, 349 down) at D4. There were 29 genes commonly expressed across each timepoint during altitude exposure (Table 1). Upon return to sea level, there were no genes that were differentially expressed when compared with baseline at any time.

To explore the biological processes related to the identified gene expression in response to altitude exposure, gene set enrichment analysis was performed, firstly using the Hallmark database (Supplementary data 1). Within this database, it was found that at D2, Haem metabolism was significantly upregulated compared with baseline (p=0.016, FDR=1.3%). Haem metabolism continued to be significantly upregulated at D3 (p=0.002, FDR=0.8%), D4 (p=0.034, FDR=2.4%) and 24 hrs (Post 1) after return to sea level (p=0.026, FDR=4%) but did not differ from baseline thereafter. Compared with D1, 20 pathways were significantly upregulated in baseline (Supplementary data 1). Following this, when compared with D2 and D3, only pathways related to beta catenin signalling and glycolysis remained significantly upregulated in baseline. Both these pathways in addition to pathways related to androgen response and fatty acid metabolism were significantly upregulated in baseline when compared with D4 (Supplementary data 1). There were no pathways that were significantly upregulated in baseline when compared to any timepoint after return to sea level. There were no pathways identified using the Gene ontology database that were significantly up regulated during the altitude sojourn, however, at Post1, regulation of hippo signalling was significantly upregulated compared with baseline (Supplementary data 2). Thereafter, there were no other pathways significantly upregulated following the return to sea level. There were 884 pathways that were upregulated at baseline when compared with D1 (Supplementary data 2, Figure 2) but no other pathways were significantly upregulated with the exception of Post 2, where there were six pathways upregulated at baseline.

PBMC

Compared with baseline, there were 200 transcripts representing 163 genes (129 up, 34 down) differentially expressed at D2, 51 transcripts representing 46 genes (18 up, 28 down) at D3, 21 transcripts representing 17 genes (9 up, 8 down) at D4, 2 transcripts representing 1 gene (downregulated) at Post 1 and 2 transcripts representing 1 gene (downregulated) at Post 2 (Supplementary table 2). Of these expressed genes, there were 10 genes commonly expressed throughout the duration of the altitude exposure (Table 2).

When investigating the biological processes related to the identified gene expression within PBMC, 18 pathways from the Hallmark database were significantly upregulated, and exceeded both the desired *p*-value and FDR at D2 when compared with baseline (Supplementary data 3). No pathways were identified in D3, however, in D4 a pathway related to hypoxia was significantly upregulated when compared with baseline ($p<0.01$, FDR=3%) however no pathways were significantly upregulated upon return to sea level. There were no time points through the altitude sojourn that resulted in a significantly upregulated pathways in baseline, however, three pathways were significantly upregulated in baseline when compared with Post3.

Notably, when using the gene ontology data base, 1250 pathways were identified to be significantly upregulated at D2 when compared with baseline. When visualised within cystoscope, the most closely linked pathways are involved with fatty acid/lipid processes (Figure 3, supplementary data 4). Subsequent to D2, there were no pathways identified during altitude exposure or upon return to sea level when compared with baseline. Similarly, there were no pathways that were significantly upregulated in baseline when compared with the altitude or upon return to sea level.

Discussion

This is the first study to investigate and quantify the transcriptomic response of well-trained athletes to a prolonged (27 day) sojourn to moderate altitude (~2400-2500 m above sea level). The main finding of this study was the identification of a 29-gene signature of altitude exposure in whole blood and a 10-gene signature of altitude in PBMC. A surprising result of this study was the lack of transcriptomic response in whole blood and minimal response in PBMC (one gene) in the days following altitude exposure, suggesting that the return to homeostatic conditions is achieved rapidly. This overall response mirrors many of the haematological findings, in that despite significant increases in the haematological markers during altitude exposure, there were no significant changes 24 hrs after return to sea level, aside from HCT, which may in part, explain the lack of a significant transcriptomic response on return to sea level.

The most important finding from this study is that a series of genes could be identified in whole blood and may have a superior diagnostic ability when creating a test to distinguish

athletes who have visited altitude and those using rHuEpo than the current athlete biological passport. For example, genes TCF7 (transcription factor 7) and ABCG2 (ATP binding cassette subfamily G member 2) are both protein coding genes involved with cellular process of haemopoietic stem cells (29, 30). Specifically, TF7 has been shown to control the rate at which haemopoietic stem cells either self-renew or differentiate primarily by binding to genes within stem cells containing CD34+, promoting self-renewal while repressing genes related differentiation (29). A downregulation in this gene, such as was identified in this study, suggests a promotion of haemopoietic stem cell differentiation into mature blood cells, a physiological process expected during altitude exposure (3). ABCG2 is highly conserved in all species of vertebrates sequenced (30). ABCG2 has been identified to play a key role within the protection of haemopoietic stem cells under hypoxic stress (31). The changes in the expression of these genes observed (i.e., downregulation of TCF7 and upregulation of ABCG2) may relate to the unique physiological response to hypoxia experienced by the participants within the present study.

There are a variety of other physiological mechanisms related to the 29 genes significantly altered throughout altitude exposure compared with baseline in whole blood (Table 1). For example, the gene HACD3 (3-hydroxyacyl-CoA dehydratase 3) is protein coding (PTPLAD1), involved in the elongation process of fatty acids (FA) (32). PTPLAD1 is one of four 3-OH acyl-CoA dehydratases and catalyses the final reaction of FA elongation, although expression has only a have small effect on the elongation of saturated and monosaturated FA and a few effects on polyunsaturated FAs (32). Other genes have a clearer function in response to altitude exposure, such as YME1L1 (YME1 like 1 ATPase) which is an mitochondrial protein, ensuring cell prefiltration, maintaining cristae morphology among other important roles (35). It has been shown that YME1L1 is stress-sensitive and is rapidly degraded in response to oxidative stress (36). YME1L1 has been identified as a marker of high-altitude pulmonary oedema (HAPE, (37)). None of the participants of the present study were diagnosed with HAPE and were in good health through the altitude sojourn and upon return to sea level, suggesting that YME1L1 may not reflect HAPE specifically, but a "normal" response to altitude exposure.

Not all genes identified within the current investigation have a clear link with altitude exposure *per se*, such as MT1F (metallothionein 1F). MT1F codes for the protein metallothionine, which has been implicated in a range of physiological processes such as

toxic metal detoxification, metal ion homeostasis and oxidative stress (38). MT1F has been shown to effectively control the homeostasis of zinc (39) and copper (40) and other toxic metals (38). MT1F has been shown to scavenge reactive oxygen species and bind metals in response to particulate matter (42). Both zinc and copper have been well established as pollutants delivered from motor vehicles (43), which are likely to be present in the air masses of Cape Town (44). In contrast, the levels of air pollution in the rural village used as the training camp within this study are likely to be significantly lower (no data available). Therefore, downregulation of MT1F may reflect a change in the constitution of inhaled air, not in response to altitude exposure. Findings such as this illustrate the careful analysis required of each identified gene prior to adoption within an anti-doping test.

In this study, not only was whole blood collected but also PBMC, which has not previously investigated as a potential tissue of interest to detect blood doping (45). While PBMC are relatively easy to extract and provide a unique insight into the immune response to a stimulus, they only contain lymphocytes, monocytes, natural killer cells and dendritic cells (46). The majority of the body's immune cells are therefore excluded from analysis (i.e., neutrophils, basophils and eosinophils). Nevertheless, PBMCs have shown to provide a valuable insight into exposure to altitude. Several genes related to hypoxia have been identified within the PBMC, such as two members of CXC chemokine gene family. For example, CXCL8, a major mediator of the inflammatory response, secreted by both the leucocytes and non-leucocyte cell population and is associated with several cancers and their response to hypoxia (47, 48) and CXCR4 (C-X-C motif chemokine receptor 4), a receptor specific for stromal cell-derived factor-1, located on the cell surface and can be induced through the activation of the HIF1-a pathway (49). The diagnostic potential of PBMC as a sample type to be used for anti-doping is growing. For example, a study has investigated the effect of rHuEpo on cytokine gene expression in tumour necrosis factor (TNF)-treated human brain microvascular endothelial cells and demonstrated a unique 96 gene signature (51). It was found that interleukin-6, interleukin-1 and CXCR4 genes were all downregulated when treated with rHuEPO. This is notable considering that in response to altitude, CXCR4 was upregulated but downregulated in the presence of rHuEpo, suggesting it may be a suitable "candidate gene" to detect rHuEpo use, considering the different responses.

A unique aspect of the present investigation is the use of both whole blood and PBMC to identify unique signatures of altitude exposure. As anticipated, the gene expression varied

388 greatly between both sample types. Specifically, whole blood demonstrated an increase in the
389 number of differentially expressed transcripts throughout the altitude exposure whereas the
390 number of differentially expressed transcripts decreased throughout the altitude sojourn in
391 PBMC (Supplementary table 1 and 2). There was only one gene that was common across
392 sample types at D2 (TENT5C), seven at D3 (H3C10, CAVIN2, GPR183, ITGB3,
393 SH3BGRL2, PPBP and PRKAR2B) and none at D4 (Supplementary table 1 and 2). The
394 initial high number of transcripts in PBMC followed by a decline could be indicative of an
395 immune response, as has been suggested by others (52). This is further highlighted by the
396 large number of pathways identified at D2 within the PBMC samples using gene set
397 enrichment analysis (Figure 2). After D2, there was no longer any pathways significantly
398 altered during the altitude exposure or upon return to sea level. Due to this probable immune
399 response to hypoxia, it has been suggested that athletes acclimatise for at least one week prior
400 to the commencement of intense physical activity (52). Considering a strong PBMC response
401 at D2 (9 days at altitude), but not at D3 (16 days at altitude) was identified within the present
402 study, a longer period of acclimatisation should be recommended of 9-16 days in length. The
403 whole blood response to altitude was markedly different to that observed with the PBMC and
404 may reflect the adaptation to altitude exposure and an increase in training intensity as the
405 athletes become more familiar with training at altitude (none of the athletes had trained at
406 altitude prior to this study). Similarly, the increased number of transcripts identified within
407 whole blood matches the continual increase in the haematological variables (Figure 1), likely
408 demonstrating adaptation to altitude.
409
410 Differentiating rHuEpo abuse and altitude is a difficult considering the limitations of the ABP
411 (53), potentially enabling athletes to “dope” with impunity. The present study has identified a
412 unique series of gene “markers” in both blood and PBMC of altitude exposure which may aid
413 in the differentiation of altitude exposure and rHuEpo abuse. A surprising finding of this
414 study was the lack of genes found upon return to sea level, a finding which will render the
415 “altitude genes” identified in this study somewhat redundant in an anti-doping situation (i.e.,
416 anti-doping agencies will know if a sample is taken at altitude). When taken in context of the
417 other available literature however, the genes identified within this study can be used to reduce
418 a list of genes identified with rHuEpo abuse (i.e., (17, 18)), to a subset of genes that have not
419 be observed during altitude exposure and thus enable the creation of a robust test, able to
420 differentiate the two. Similarly, since no genes were identified upon return to sea level, those
421 genes identified following rHuEpo administration (i.e., (17)) can be used as evidence for

rHuEpo abuse and are likely not in response to a recent sojourn to altitude. Furthermore, the present study has identified several genes related to hemopoietic stem cell production, some of which require a hypoxic environment to be stimulated. Genes such as these should be considered as candidates as "negative controls" of rHuEpo. Since rHuEpo increases erythropoiesis *without* a hypoxic environment, the lack of these stem cells indicates erythropoiesis without the hypoxic stimulus and indicate rHuEpo abuse. Considering the popularity of altitude training, athletes could, in theory use rHuEpo during an altitude training camp, merging both the response of altitude and rHuEpo. Such cases will require more analysis and further study to identify candidate genes representing these unique situations.

There are several limitations of this study which must be considered when interpreting the presented result. Firstly, the height and duration of altitude training camps varies significantly between athletes and therefore, further studies should investigate the transcriptomic response to altitude exposure, with the effect of changing height, duration and training structure (i.e., LHTL) quantified. No measure of total haemoglobin mass (tHbmass) was made within the present study and therefore an accurate quantification of the participant's adaptation to altitude cannot be made. Future studies should consider measuring tHbmass and gene expression to explore potential correlations. The genes expressed within this study were determined via microarray and not subsequently validated using quantitative polymerase chain reaction (q-PCR). Validation of the genes will be required prior to the creation of an anti-doping test which will have to be defended in a court of law, if successfully used to catch and prosecute doping athletes. During the analysis of the whole blood gene expression, it became apparent that the baselines (four days and one day prior to departure) differed significantly. This difference was not observed within the PBMC baselines, suggesting an issue relates to whole blood samples specifically. No data is available one day prior to departure and thus we are unaware of any potential haematological changes that may have occurred over the four-day baseline period, as has been observed when athletes are unfamiliar with needles/blood drawing as has been demonstrated elsewhere (i.e., Figure 1A and 1B (54)). However, considering the lack of differences between both baselines in the PBMC, the authors believe this issue is related to the whole blood samples taken four days prior to departure only and have therefore excluded these data from the analysis. Only two female athletes were included within the current study and therefore sex-specific differences in gene expression in response to altitude exposure cannot be determined.

In conclusion, the present study has demonstrated a unique molecular signature of altitude exposure in both whole blood and PBMC. Surprisingly, there was only a minimal transcriptomic response to altitude in both tissue types upon return to sea level. The identified genes can be used to eliminate markers of "natural" stimulation of erythropoiesis from the established markers of rHuEpo abuse. Prior to the creation of an "OMICS" anti-doping test for rHuEpo, further studies are needed to quantify the effect of other confounding factors such as exercise, sex and repeated exposure to altitude. Similarly, further efforts should be made to assess the variability in gene expression in a much larger sample size than that used within the present study.

Acknowledgements

This work was funded by both the World Anti-Doping Agency (WADA) grant no. 13C28YP and the South African Institute for drug-free sport (SAIDS).

Conflict of interest

None declared

Figure legends

Figure 1. Haematological variables across 27-day sojourn to altitude (~2400 – 2500 m above sea level), specifically haemoglobin concentration (A), haematocrit (B) and Reticulocyte percentage (C). Significant differences ($p<0.05$) from baseline are indicated by *.

Figure 2. Biological network of the whole blood dataset at D1 using the Gene Ontology Biological process gene set enrichment analysis and Cystoscope visualisation. Each circle represents an individual gene set, and each interconnected line represents a shared gene between gene sets. The size and width of the circles and lines is proportional to the number of genes within a set and the number of shared genes, respectively. Pathways related to similar biological themes are grouped together. The enrichment map was created with the following cut offs; <0.1% FDR, $p<0.05$, Jaccard Overlap coefficient >0.375 with a combined constant K=0.5 and a normalised enrichment score $\geq$1.70.

Figure 3. Biological network of the PBMC dataset at D2 using the Gene Ontology Biological process gene set enrichment analysis and Cystoscope visualisation. Each circle represents an

individual gene set, and each interconnected line represents a shared gene between gene sets. The size and width of the circles and lines is proportional to the number of genes within a set and the number of shared genes, respectively. Pathways related to similar biological themes are grouped together. The enrichment map was created with the following cut offs; <0.1% FDR, $p<0.05$, Jaccard Overlap coefficient >0.375 with a combined constant K=0.5 and a normalised enrichment score $\geq$1.90.

Table 1. 29 genes, commonly expressed in whole blood during altitude exposure over 24 hr, 9, 16 and 24 days after arriving at altitude (~2400-2500 m above sea level). All genes exceed the predetermined 1.2-fold change, 5% FDR and $p<0.05$. Average FC indicates the mean FC expressed across each time point during altitude exposure.

Table 2. 10 genes, commonly expressed in peripheral monocyte blood cells (PBMC) during altitude exposure over 24 hr, 9, 16 and 24 days after arriving at altitude (~2400-2500 m above sea level). All genes exceed the predetermined 1.2-fold change, 5% FDR and $p<0.05$. Average FC indicates the mean FC expressed across each time point during altitude exposure.

Supplementary table 1. Complete list of all differentially expressed genes in whole blood when compared with baseline over the 27-day altitude sojourn and upon return to sea level.

Supplementary table 2. Complete list of all differentially expressed genes in PBMCs when compared with baseline over the 27-day altitude sojourn and upon return to sea level.

Supplementary data 1. Gene set enrichment analysis of whole blood when compared with baseline, using the Hallmark database. Data highlighted in green indicates that the pathway has exceeded the predetermined cut offs of $p<0.05$ and 10% FDR.

Supplementary data 2. Gene set enrichment analysis of whole blood when compared with baseline, using the Gene Ontology biological process database. Data highlighted in green indicates that the pathway has exceeded the predetermined cut offs of $p<0.05$ and 10% FDR.

Supplementary data 3. Gene set enrichment analysis of PBMC when compared with baseline, using the Hallmark database. Data highlighted in green indicates that the pathway has exceeded the predetermined cut offs of $p<0.05$ and 10% FDR.

Supplementary data 4. Gene set enrichment analysis of PBMC when compared with baseline, using the Gene Ontology biological process database. Data highlighted in green indicates that the pathway has exceeded the predetermined cut offs of $p<0.05$ and 10% FDR.

References:

1. Porter R, Knight J. *High Altitude Physiology: cardiac and respiratory aspects*. John Wiley & Sons; 2009.
2. Brown JPR, Grocott MPW. Humans at altitude: physiology and pathophysiology. *Contin Educ Anaesth Crit Care Pain*. 2013;13(1):17–22.
3. Ge R-L, Witkowski S, Zhang Y, et al. Determinants of erythropoietin release in response to short-term hypobaric hypoxia. *J Appl Physiol*. 2002;92(6):2361–7.
4. Mazzeo RS. Physiological Responses to Exercise at Altitude: An Update. *Sport Med*. 2008;38(1):1–8.
5. Lundby C, Robach P. Does 'altitude training' increase exercise performance in elite athletes?: Should altitude training be recommended to elite athletes? *Exp Physiol*. 2016;101(7):783–8.
6. Wilber RL, Pitsiladis YP. Kenyan and Ethiopian distance runners: what makes them so good? *Int J Sports Physiol Perform*. 2012;7(2):92–102.
7. Faulkner JA, Daniels JT, Balke B. Effects of training at moderate altitude on physical performance capacity. *J Appl Physiol*. 1967;23(1):85–9.
8. Buskirk ER, Kollias J, Akers RF, Prokop EK, Reategui EP. Maximal performance at altitude and on return from altitude in conditioned runners. *J Appl Physiol*. 1967;23(2):259–66.
9. Levine BD, Stray-Gundersen J. "Living high-training low": effect of moderate-altitude acclimatization with low-altitude training on performance. *J Appl Physiol*. 1997;83(1):102–12.
10. Lundby C, Millet GP, Calbet JA, Bärtsch P, Subudhi AW. Does "altitude training" increase exercise performance in elite athletes? *Br J Sports Med*. 2012;46(11):792–5.
11. Durussel J, Daskalaki E, Anderson M, et al. Haemoglobin mass and running time trial performance after recombinant human erythropoietin administration in trained men. *PLoS One*. 2013;8(2):e56151.
12. Thomsen JJ, Rentsch RL, Robach P, et al. Prolonged administration of recombinant human erythropoietin increases submaximal performance more than maximal aerobic capacity. *Eur J Appl Physiol*. 2007;101(4):481–6.
13. Annaheim S, Jacob M, Krafft A, Breymann C, Rehm M, Boutellier U. RhEPO improves time to exhaustion by non-hematopoietic factors in humans. *Eur J Appl Physiol*. 2016;116(3):623–33.
14. Ashenden MJ, Hahn AG, Martin DT, Logan P, Parisotto R, Gore CJ. A comparison of

the physiological response to simulated altitude exposure and r-HuEpo administration. *J Sports Sci*. 2001;19(11):831–7.
15. Ashenden M, Gough CE, Garnham A, Gore CJ, Sharpe K. Current markers of the Athlete Blood Passport do not flag microdose EPO doping. *Eur J Appl Physiol*. 2011;111(9):2307–14.
16. Ashenden M. Contemporary issues in the fight against blood doping in sport. *Haematologica*. 2004;89(8):901–3.
17. Durussel J, Haile DW, Mooses K, et al. Blood transcriptional signature of recombinant human erythropoietin administration and implications for antidoping strategies. *Physiol Genomics*. 2016;48(3):202–9.
18. Wang G, Durussel J, Shurlock J, et al. Validation of whole-blood transcriptome signature during microdose recombinant human erythropoietin (rHuEpo) administration. *BMC Genomics*. 2017;18(8 LB-Wang2017):817.
19. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47.
20. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 2005;21(9):2067–75.
21. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545–50.
22. Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482–517.
23. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
24. Consortium GO. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–8.
25. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
26. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5(11):e13984.

27. Kucera M, Isserlin R, Arkhangorodsky A, Bader GD. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations [Internet]. *F1000Research*. 2016;5.
28. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
29. Wu JQ, Seay M, Schulz VP, et al. Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. *PLoS Genet*. 2012;8(3):e1002565.
30. Robey RW, To KKK, Polgar O, et al. ABCG2: a perspective. *Adv Drug Deliv Rev*. 2009;61(1):3–13.
31. Krishnamurthy P, Ross DD, Nakanishi T, et al. The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme. *J Biol Chem*. 2004;279(23):24218–25.
32. Sawai M, Uchida Y, Ohno Y, et al. The 3-hydroxyacyl-CoA dehydratases HACD1 and HACD2 exhibit functional redundancy and are active in a wide range of fatty acid elongation pathways. *J Biol Chem*. 2017;292(37):15538–51.
33. Boutchueng-Djidjou M, Belleau P, Bilodeau N, et al. A type 2 diabetes disease module with a high collective influence for Cdk2 and PTPLAD1 is localized in endosomes. *PLoS One*. 2018;13(10):e0205180.
34. Katayama K, Goto K, Ishida K, Ogita F. Substrate utilization during exercise and recovery at moderate altitude. *Metabolism*. 2010;59(7):959–66.
35. Stiburek L, Cesnekova J, Kostkova O, et al. YME1L controls the accumulation of respiratory chain subunits and is required for apoptotic resistance, cristae morphogenesis, and cell proliferation. *Mol Biol Cell*. 2012;23(6):1010–23.
36. Rainbolt TK, Saunders JM, Wiseman RL. YME 1L degradation reduces mitochondrial proteolytic capacity during oxidative stress. *EMBO Rep*. 2015;16(1):97–106.
37. Tang C, Yu C, Liu X, Ran Z, Luo Y. Genome-wide association study of mitochondrial DNA in Chinese men identifies seven new susceptibility loci for high-altitude pulmonary oedema. *Vojnosanit Pregl*. 2020;(00):63.
38. Sutherland DEK, Stillman MJ. The “magic numbers” of metallothionein. *Metallomics*. 2011;3(5):444–63.
39. Kelly EJ, Quaife CJ, Froelick GJ, Palmiter RD. Metallothionein I and II protect against zinc deficiency and zinc toxicity in mice. *J Nutr*. 1996;126(7):1782–90.

631 40. Kelly EJ, Palmiter RD. A murine model of Menkes disease reveals a physiological
632 function of metallothionein. *Nat Genet*. 1996;13(2):219–22.
633 41. Courtade M, Carrera G, Paternain JL, et al. Metallothionein expression in human lung
634 and its varying levels after lung transplantation. *Chest*. 1998;113(2):371–8.
635 42. Huang Y-CT, Li Z, Carter JD, Soukup JM, Schwartz DA, Yang I V. Fine ambient
636 particles induce oxidative stress and metal binding genes in human alveolar
637 macrophages. *Am J Respir Cell Mol Biol*. 2009;41(5):544–52.
638 43. Fang G-C, Wu Y-S, Wen C-C, Huang S-H, Rau J-Y. Ambient air particulate
639 concentrations and metallic elements principal component analysis at Taichung Harbor
640 (TH) and WuChi Traffic (WT) near Taiwan Strait during 2004–2005. *J Hazard Mater*.
641 2006;137(1):314–23.
642 44. Williams J, Petrik L, Wichmann J. PM 2.5 chemical composition and geographical
643 origin of air masses in Cape Town, South Africa. *Air Qual Atmos Heal*. 2020;1–12.
644 45. Reichel C. OMICS-strategies and methods in the fight against doping. *Forensic Sci*
645 *Int*. 2011;213(1):20–34.
646 46. Kleiveland CR. Peripheral blood mononuclear cells. *impact food Bioact Heal*.
647 2015;161–7.
648 47. Russo RC, Garcia CC, Teixeira MM, Amaral FA. The CXCL8/IL-8 chemokine family
649 and its receptors in inflammatory diseases. *Expert Rev Clin Immunol*. 2014;10(5):593–
650 619.
651 48. Korbecki J, Kojder K, Kapczuk P, et al. The effect of hypoxia on the expression of
652 CXC chemokines and CXC chemokine receptors—a review of literature. *Int J Mol Sci*.
653 2021;22(2):843.
654 49. Krzywinska E, Stockmann C. Hypoxia, metabolism and immune cell function.
655 *Biomedicines*. 2018;6(2):56.
656 50. Piovan E, Tosello V, Indraccolo S, et al. Differential regulation of hypoxia-induced
657 CXCR4 triggering during B-cell development and lymphomagenesis. *Cancer Res*.
658 2007;67(18):8605–14.
659 51. Avasarala JR, Konduru SS. Recombinant erythropoietin down-regulates IL-6 and
660 CXCR4 genes in TNF-α-treated primary cultures of human microvascular endothelial
661 cells. *J Mol Neurosci*. 2005;25(2):183–9.
662 52. Mazzeo RS. Altitude, exercise and immune function. *Exerc Immunol Rev*.
663 2005;11(6):16.
664 53. Sutehall S, Muniz-Pardos B, Lima G, et al. Altitude Training and Recombinant Human

Erythropoietin: Considerations for Doping Detection. *Curr Sports Med Rep*. 2019;18(4):97–104.

54. Haile DW, Durussel J, Mekonen W, et al. Effects of EPO on Blood Parameters and Running Performance in Kenyan Athletes. *Med Sci Sports Exerc*. 2018;51(2):299–307.