

## Article

# Microarray-Based Prediction of Polycythemia after Exposure to High Altitudes

Haijing Wang<sup>1,2,3,4,†</sup>, Daoxin Liu<sup>1,2,4,5,†</sup>, Pengfei Song<sup>1,2,4</sup> , Feng Jiang<sup>1,2,4</sup> and Tongzuo Zhang<sup>1,2,4,\*</sup>

<sup>1</sup> Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China; wanghj@nwipb.cas.cn (H.W.); liudx@qhu.edu.cn (D.L.); pfsong@nwipb.cas.cn (P.S.); jiangfeng@nwipb.cas.cn (F.J.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Medical School, Qinghai University, Xining 810016, China

<sup>4</sup> Qinghai Provincial Key Laboratory of Animal Ecological Genomics, Xining 810008, China

<sup>5</sup> Kunlun College, Qinghai University, Xining 810016, China

\* Correspondence: zhangtz@nwipb.cas.cn

† These authors contributed equally to this paper.

**Abstract:** In high-altitude environments, the prevalence of high-altitude polycythemia (HAPC) ranges between 5 and 18 percent. However, there is currently no effective treatment for this condition. Therefore, disease prevention has emerged as a critical strategy against this disease. Here, we looked into the microarray profiles of GSE135109 and GSE29977, linked to either short- or long-term exposure to the Qinghai Tibet Plateau (QTP). The results revealed inhibition in the adaptive immune response during 30 days of exposure to QTP. Following a gene set enrichment analysis (GSEA) discovered that genes associated with HAPC were enriched in Cluster1, which showed a dramatic upregulation on the third day after arriving at the QTP. We then used GeneLogit to construct a logistic prediction model, which allowed us to identify 50 genes that classify HAPC patients. In these genes, *LRRC18* and *HCAR3* were also significantly altered following early QTP exposure, suggesting that they may serve as hub genes for HAPC development. The in-depth study of a combination of the datasets of transcriptomic changes during exposure to a high altitude and whether diseases occur after long-term exposure in Hans can give us some inspiration about genes associated with HAPC development during adaption to high altitudes.

**Keywords:** high altitude polycythemia; logistic model; prediction; microarray



**Citation:** Wang, H.; Liu, D.; Song, P.; Jiang, F.; Zhang, T. Microarray-Based Prediction of Polycythemia after Exposure to High Altitudes. *Genes* **2022**, *13*, 1193. <https://doi.org/10.3390/genes13071193>

Academic Editor: Giuseppe Alberto Palumbo

Received: 17 May 2022

Accepted: 30 June 2022

Published: 2 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With their hypoxic circumstances, high altitudes have a profound impact on the physical health of locals and visitors [1]. In order to adapt to hypobaric hypoxia, dwellers from plain to plateau must allow the lungs to acquire more atmospheric oxygen, and the number of red blood cells (RBCs) will rise compensatively [2,3]. Mild polycythemia is considered a beneficial response to high-altitude hypoxia [4]. However, in certain people, the number of RBCs continues to grow, resulting in high altitude polycythemia (HAPC), which is characterized by headache, disorientation, sleeplessness, and bone discomfort. It was first described by Viault in 1980 [5] and is the most typical characteristic sign of chronic mountain sickness [6].

Over 140 million people reside at altitudes greater than 2500 meters above sea level [7]. The Qinghai-Tibet Plateau (QTP), the highest and largest plateau in the world, is characterized by low oxygen levels, and millions of people live and work in this region. According to reports, the prevalence of HAPC in the QTP increases with altitude and ranges between 5% and 18% [8,9]. As the Qinghai-Tibet Railway was finished, an increasing number of Han people migrated and traveled to Tibet. HAPC is much more prevalent among Han than among natives [2]. Historically, studies focused more on the pathophysiologic processes of

HAPC [10]. Furthermore, numerous investigations have verified the associations between gene polymorphisms and HAPC by analyzing the genetic backgrounds of Han people and Tibetans [9,11,12]. Nonetheless, disease-related genes require further investigation, particularly in Han people. Here, we explored the microarray profile of HAPC and discovered that genes involved in the development of this disease even changed at the early time of exposure to hypoxia.

This study investigated the microarray profiles of GSE135109 [13] and GSE29977 [6], respectively. We classified the differential expression genes (DEGs) in GSE135109 into 6 clusters with different dynamic expression patterns. Furthermore, 415 DEGs in GSE29977 were used to construct a logistic model to predict HAPC. At last, we overlapped the genes identified in two datasets and acquired *LRRC18* and *HCAR3* as hub genes related to HAPC development after exposure to QTP. This research will give us some inspiration about key genes associated with HAPC development during adaption to high altitudes and significantly enhance the early warning for HAPC.

## 2. Materials and Methods

### 2.1. Data Collection and Preliminary Processing

We searched in NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo/>, accessed on 17 March 2022), using the keywords “high altitude” and “Homo sapiens” and downloaded the microarray datasets regarding people adjusting to high altitudes, including GSE135109 [13] and GSE29977 [6]. The GSE135109 microarray dataset comprised 8 samples, which were sampled at a low-altitude starting point (500 m) and at 3, 7, and 30 days after a quick ascent to a high altitude (5200 m). The volunteers in this dataset were never previously at high altitudes and had not developed HAPC during 30 days of exposure to high altitudes. Using the GPL24539 platform, white blood cells from the periphery were collected and sequenced. The oligo package was used to read CEL files from the GSE135109 dataset by annotating gene symbols. The GSE29977 microarray data included 10 subjects who were Hans and had migrated to the high altitude (4550 m) for more than 8 months, 5 of which were high altitude polycythemia (HAPC) patients, while the remaining 5 were matched controls. The peripheral blood leukocytes of 10 individuals were collected, and the gene expression profiling was analyzed by GPL570 platform. GEOquery and AnnoProbe packages were conducted to download and annotate this dataset. The participants in both two datasets were young men. After the collected datasets were annotated with official gene symbols and normalized ( $\log_2$ ), we acquired 18,190 genes and 18 samples for the following analysis.

### 2.2. Identification of Differential Expression Genes

A differential expression gene (DEG) analysis was performed on the GSE135109 and GSE29977 datasets, respectively. The limma (v 3.50.0) package [14] and the normalizeBetweenArrays function were used to normalize the quantiles. The significant DEG threshold was defined as  $\log_2(\text{FC})$  with an absolute value more than 1 and a  $p$ -value less than 0.05.

### 2.3. Microarray Time-Course Analysis

The union DEG sets (1135 genes) between three groups (D03vsD0, D07vsD0, D30vsD0) were acquired for a further time-course analysis. The *avereps* function in the limma package was used to compute the average expression values at various times. The *Mfuzz* package [15] was implemented to detect clusters with consistent expression trends during exposure to QTP.

### 2.4. Logistic Prediction Model Using Procedure Logistic

To develop a logistic prediction model, the GeneLogit software, a logistic regression method applied to microarray data in small samples, was utilized [16]. Firstly, the localFDR function was used to estimate the false discovery rate (FDR). Secondly, the theta estimate provided by the model estimation function was 0.012. Thirdly, the bootstrap prediction function was used to find the optimal  $\tau$  and the corresponding prediction error for different

values of  $q$ . The results are displayed in Table 1. Finally, we selected the procedure logistic ( $q = 50$ ,  $\tau = 0.379$ ) to construct our prediction model.

**Table 1.** The optimal  $\tau$  and the prediction error for different  $q$ .

$q$	1	10	20	50	100
Optimal $\tau$	4.502	2.316	0.559	0.379	0.214
prediction error	0.059	0.058	0.056	0.051	0.051

### 2.5. Analysis of Weighted Gene Co-Expression Network Analysis (WGCNA)

We conducted the weighted gene co-expression network analysis (WGCNA) to discover the relationships between gene expression patterns and phenotypes in GSE29977 using the WGCNA package [17]. For the subsequent analysis, we kept 14043 genes with a median absolute deviation (MAD) greater than 0.01. The soft power was estimated at 14, and we constructed an unsigned scale-free co-expression network for the genes with a minModuleSize of 100 and mergeCutHeight of 0.25. Ultimately, we identified 35 modules. The moduleEigengenes function calculated the dissimilarity of the module eigengenes (MEs), and then the Spearman correlation was used to assess the associations between the MEs and phenotypes. The exportNetworkToCytoscape function was used to export the network of genes selected in the logistic prediction model. At last, Cytoscape software [18] was acquired to visualize the network.

### 2.6. Evaluation of Immune Cell Types

The CIBERSORT algorithm was applied to the log transformed normalized microarray data of GSE135109, and the leukocyte signature matrix (Cibersort: LM22) was used to infer immune cell types in each sample [19]. We utilized 1000 permutations from the default signature matrix and calculated the  $p$ -value and root mean square error for each sample file to improve the accuracy of the deconvolution algorithm. The scores for each signature were summarized for each group and median centered to permit comparisons between groups. The wilcox.test was used to discover the significance between comparisons.

### 2.7. Functional and Pathway Enrichment Analysis

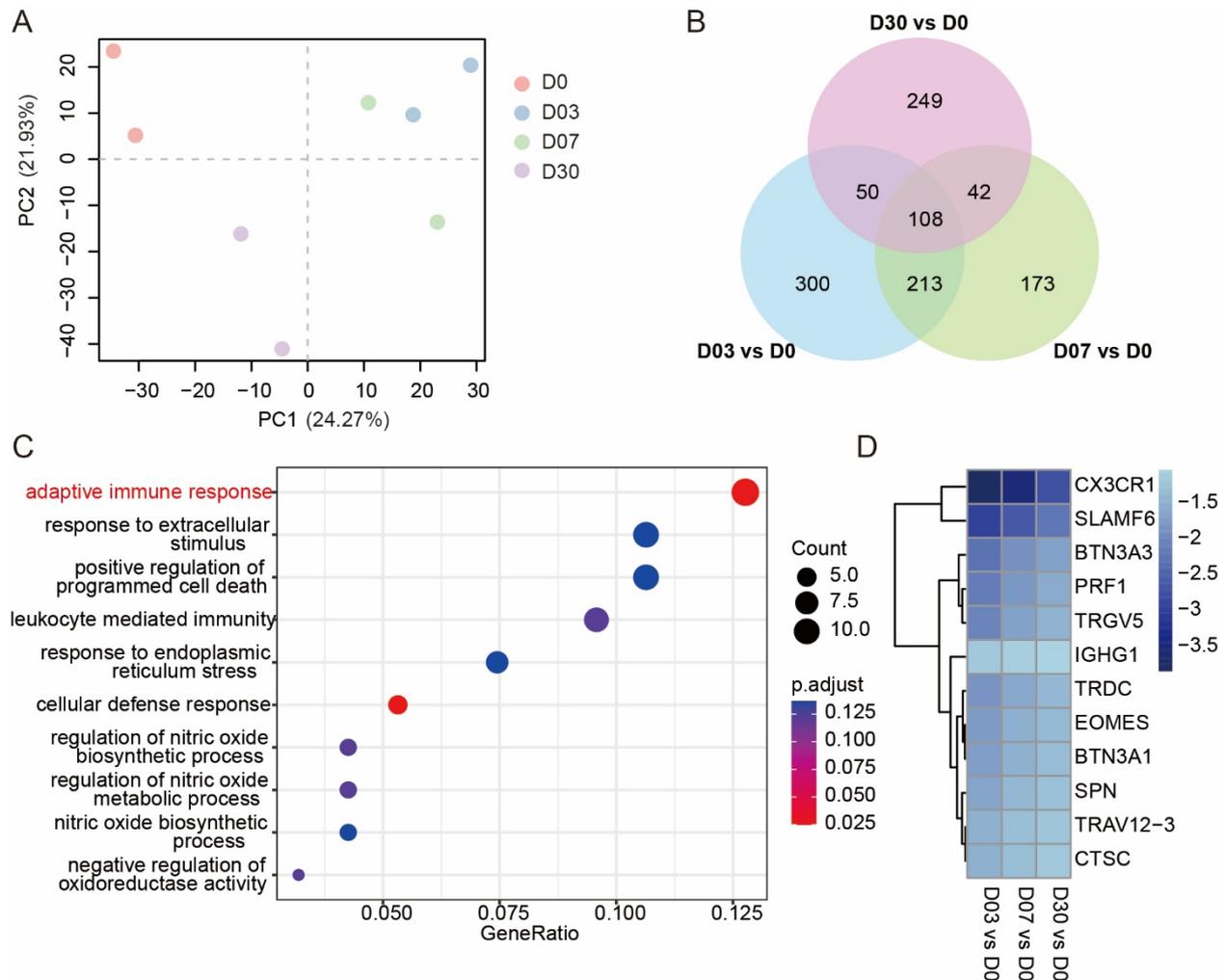
A Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and visualization were performed using the package clusterProfiler [20] for genes with different expression patterns. The GO comparison across clusters was performed using the compareCluster function. The gene set enrichment analysis (GSEA) [21] was performed to determine whether a series of pre-defined geneset were enriched in the gene rank. The Molecular Signatures Database (MSigDB) [22] was also used here.  $p$ -values of  $<0.05$  denoted statistical significance.

## 3. Results

### 3.1. Exposure to the QTP Weakened the Adaptive Immune Response in Leukocytes

To determine the differences in the gene expressions of the four groups, the principal component analysis (PCA) was conducted (Figure 1A). The apparent distinction was that the gene expression changes after exposure to the QTP for 3 days were similar to those for 7 days (increasing along the PC1 axis). However, this trend reduced along the PC2 axis after 30 days, indicating that the continuous exposure altered the expression of genes in white blood cells. Then, the differential expression gene analysis was conducted to identify the differential expression genes (DEGs) between different time points (D03, D07, D30) and the control group (D0). The Venn diagram demonstrated that 108 genes were consistently significant along with the exposure duration (Figure 1B). We performed GO enrichment to determine the function of these 108 genes, and the dot plot revealed that these genes were enriched in the adaptive immune response category (Figure 1C). We extracted 12 genes related to the adaptive immune response, and the log<sub>2</sub> fold change of each treatment group

declined with exposure time (Figure 1D). These findings suggest that hypoxia in the QTP might inhibit the adaptive immune response of white blood cells but that it would recover as the exposure duration increases.



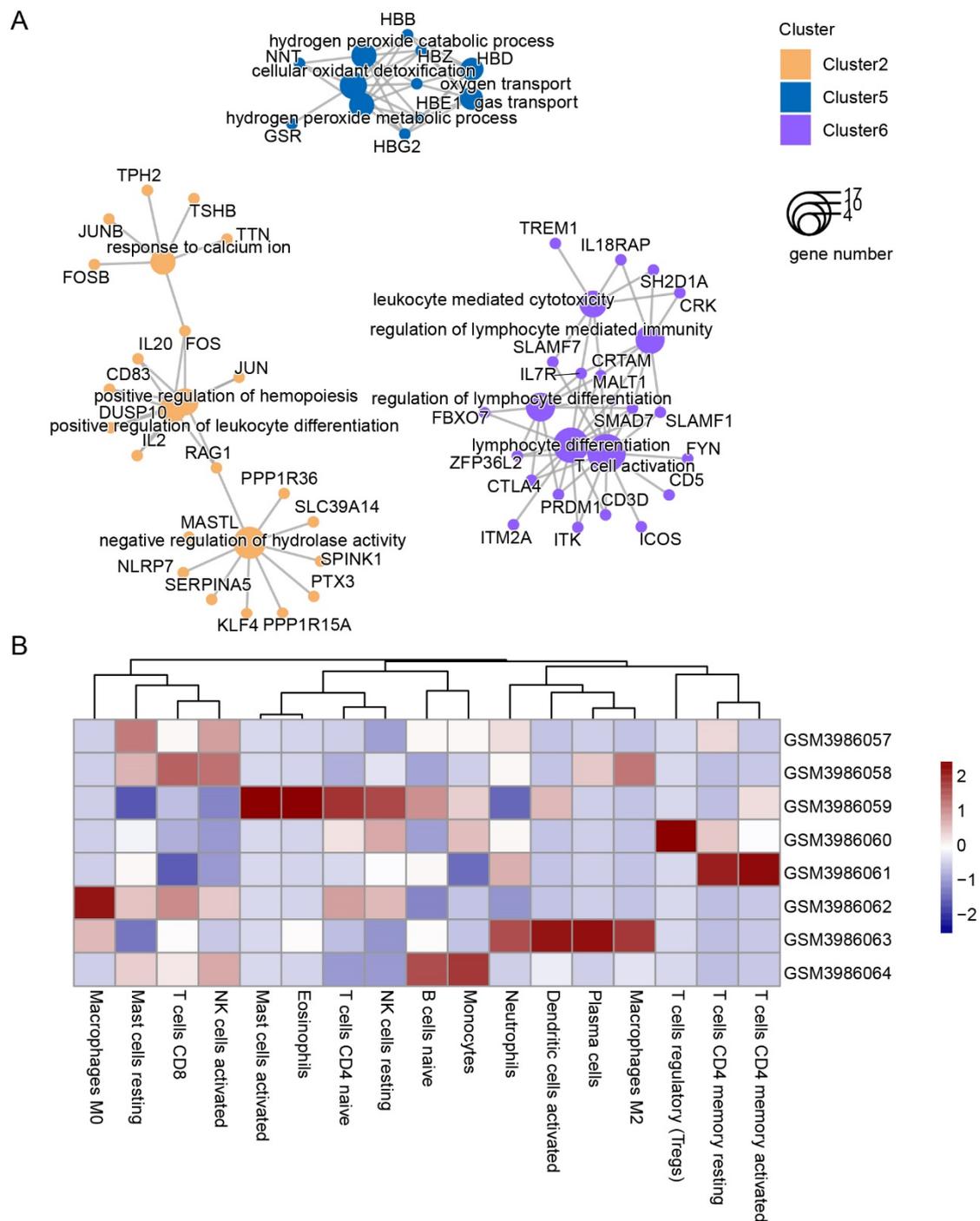
**Figure 1.** Altitude exposure weakened the adaptive immune response: (A) PCA plot of GSE135109; (B) Venn diagram showed the number of shared DEGs in three contrasts; (C) the dot plot was used for visualization of GO enrichment; (D) the heatmap showed log(FC) values in three contrasts of genes related to adaptive immune response.

### 3.2. Dynamic Expression and Functional Characteristics of Leukocytes upon Arrival in the QTP

We investigated the gene expression dynamics in human leukocytes at four time points using 1135 genes that were significantly different from the control, and six clusters were obtained (Figure 2A). Cluster 1 (206 genes) and cluster 3 (244 genes) were dramatically changed with an increase in cluster1 and a decrease in cluster 3 on the third day after arrival in the QTP. The dynamic expression of these 450 genes exhibited the greatest altitude sensitivity. Then, we performed an enrichment analysis to determine their function (Figure 2B). Both overlaps and variances existed between the two clusters. Cluster 1 was specifically enriched in terms of the intrinsic apoptotic signaling pathway and the cellular response to interleukin-1, and *HIF1 $\alpha$*  was also found in this cluster. Cluster 3 was enriched in terms of receptor-mediated endocytosis and adaptive thermogenesis. Both of them were enriched in the positive regulation of cytokine production.



(Figure 3B and Figure S1A). This result reminded us that the expression patterns in leukocytes were in entirely different forms.

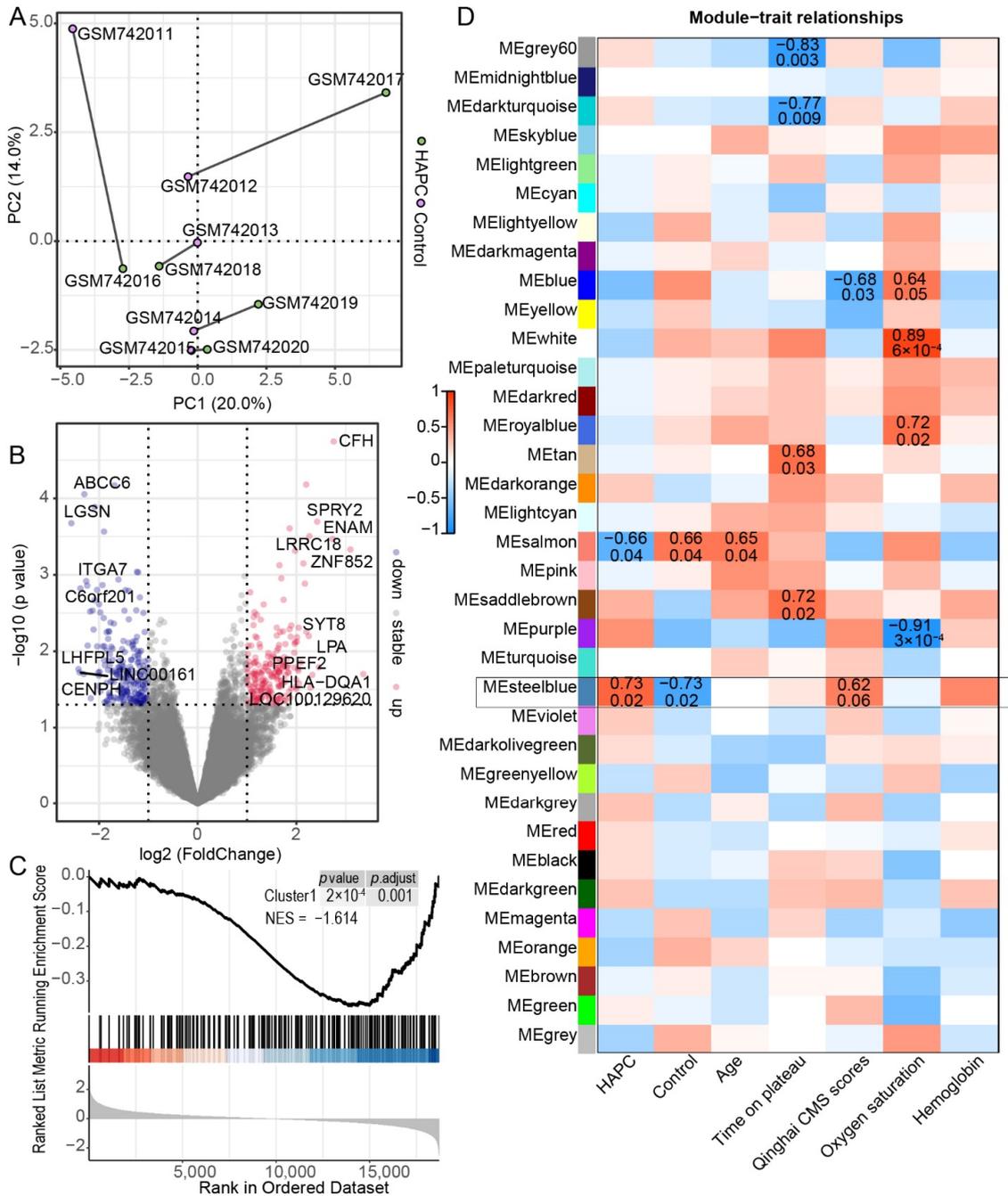


**Figure 3.** The biological functions of genes insensitive to hypoxia and the changes in prediction in immune cell types: (A) genes in cluster 2, cluster 5, and cluster 6 were compared to identify enriched biological processes, with the sizes of the dots representing the numbers of genes involved and with colors identifying different clusters; (B) the heatmap of the scores for each immune cell type.

### 3.3. Gene Expression in Leukocytes of HAPC Patients Exhibited Substantial Individual Variation

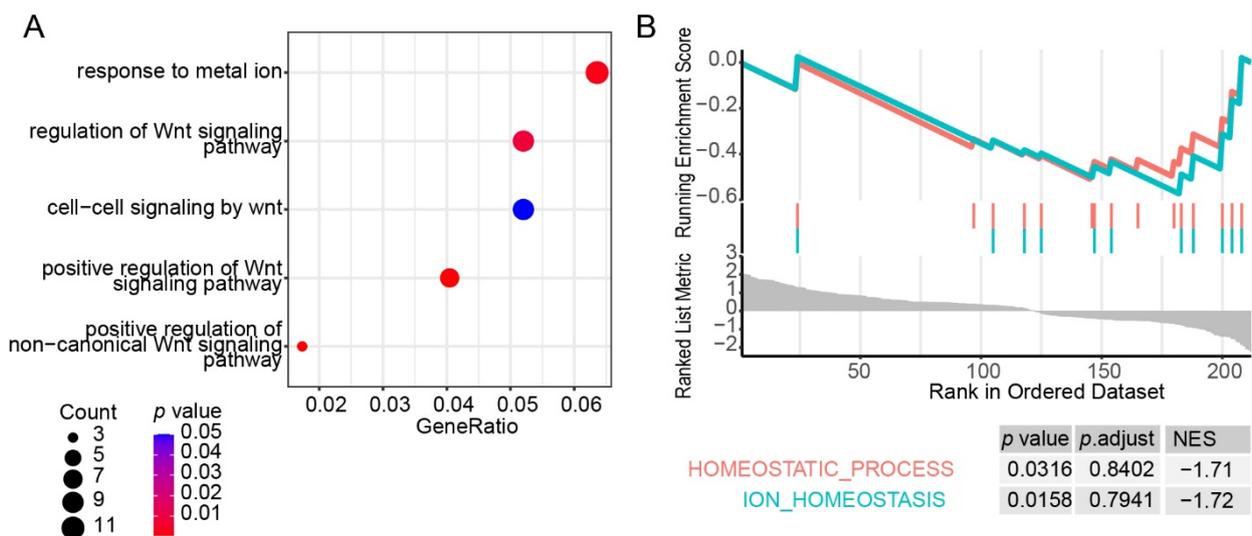
The GSE29977 dataset contains the microarray expression levels of 5 patients and 5 healthy human blood leukocytes that are paired with additional personal information. However, the PCA revealed significant individual differences, indicating a large dispersion

in the group (Figure 4A). Next, we performed DEG analysis and acquired 212 downregulated and 203 upregulated genes (Figure 4B). Then, the gene set enrichment analysis (GSEA) was conducted to figure out the potential function, and the result indicated that genes involved in the metabolism of heme were downregulated and HBZ was included in this geneset (Figure S1B).



**Figure 4.** Analysis of genes related to HAPC in GSE135109: (A) PCA plot of GSE135109; (B) a volcanic map showing DEGs, where the red dots indicates DEGs with upregulation and the blue dots indicate downregulation; the genes with absolute log(FC) values higher than 2.25 were labeled; (C) the GSEA plot of geneset of cluster 1 in Figure 2A; (D) The heatmap of module trait relationships showing the correlations between each gene module and the phenotypes.

To figure out the relationship with DEGs in GSE135109 and GSE29977, we used the GSEA. The result was only enriched in cluster1 (Figure 4C), which increased quickly upon QTP exposure and also reminded us that genes related to HAPC were affected at an early stage of exposure to high altitudes. However, when we examined genes that were overlapped in two datasets, a few of them were identified (Figure S1C). In order to figure out the molecular mechanism underlying HAPC, we obtained the pivotal modules from the WGCNA analysis (Figures 4D and S2A). The steelblue and salmon modules were significant in the HAPC patients. The blue module was also negatively correlated with Qinghai CMS scores but positive with oxygen saturation. This meant that these three modules were directly related to the disease. The oxygen saturation was also relevant with the white, royalblue, and purple modules. Additionally, the length of time on the plateau showed different signs and was significant with modules of grey60, datkturquoise, tan, and saddlebrown. This result demonstrated that the age, time on plateau, and oxygen saturation also impact the disease outcome. The steelblue module drew our attention. It was not only related to the disease phenotype, but also to CMS which is its diagnostic criterion (Figure 4D). Therefore, we further analyzed the genes in this module. The results showed that the biological process of this gene set was enriched in the Wnt signaling pathway (Figure 5A), and the homeostasis in leukocytes was inhibited by genes in this module (Figure 5B).



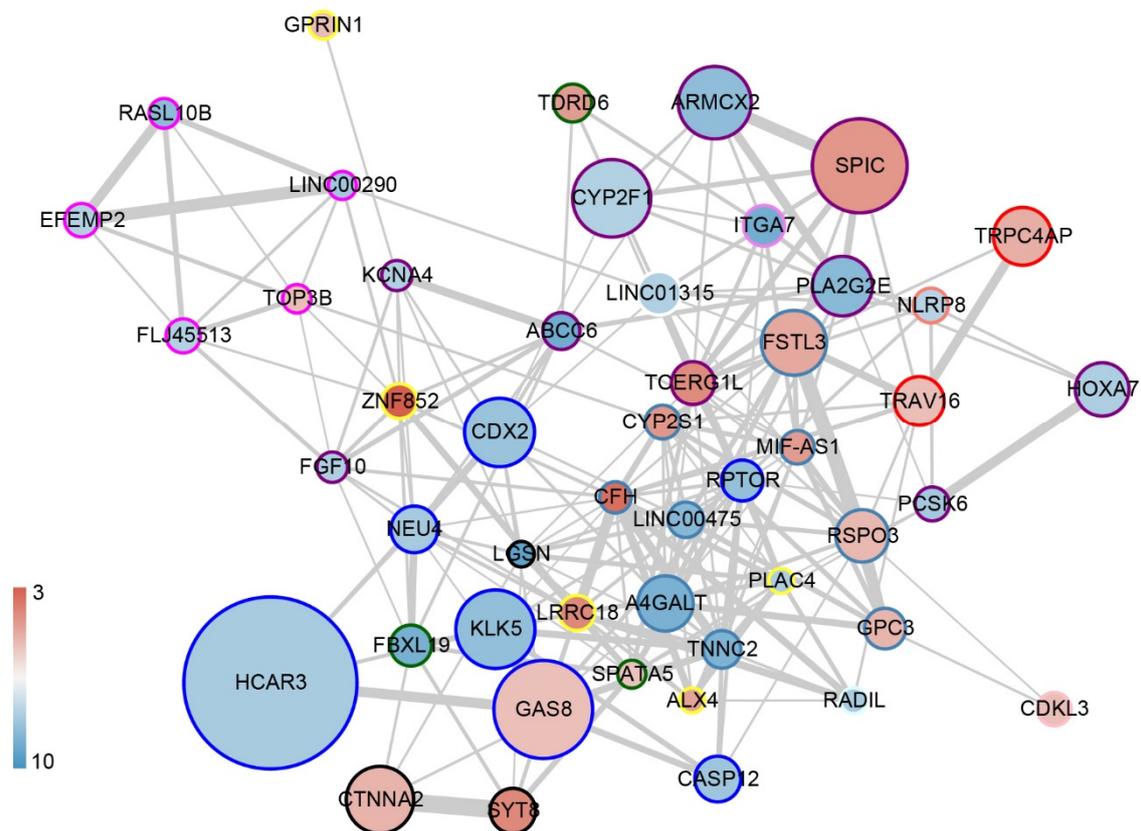
**Figure 5.** Enrichment of genes in the steelblue module: (A) the dot plot of functional annotations of genes in the steelblue module using the Gene Ontology biological process; (B) the GSEA plot of the “Homeostatic process” and “ion homeostasis” genesets in the steelblue module.

### 3.4. Construction of a Prediction Model for HAPC

To identify the marker genes that predict HAPC, we constructed the logistic regression model using the 415 DEGs. The optimal  $\tau$  and the corresponding prediction errors for  $q = 1, 10, 20, 50,$  and  $100$  are given in Table 1. The optimal  $\tau$  decreases as  $q$  increases and choosing a  $q$  value beyond 50 does not help much in further reducing the prediction error. Next, we chose the procedure logistic ( $q = 50, \tau = 0.379$ ) to build our prediction model, with the results given in Table 2, and obtained 50 marker genes to construct the prediction model. However, there was no separate test dataset to assess the prediction error for our model. Then, we used this dataset itself to cross-validate the model and illustrated a higher efficiency (Figure S2B). *LRRC18* and *HCAR3* were also overlapped in the DEGs of GSE135109 (Table S1). To visualize the contribution of these genes, we built a network of these 50 marker genes using a weight of edge generated threshold of 0.02 in WGCNA (Figure 6). *HCAR3* showed higher total and inner connectivity (kTotal) in this network. Here, we constructed a logistic model to screen the marker genes to predict HAPC.

**Table 2.** Logistic prediction model using logistic regression ( $q = 50$ ,  $\tau = 0.379$ ).

Gene	Intercept	<i>PLA2G2E</i>	<i>TNNC2</i>	<i>GAS8</i>	<i>RPTOR</i>	<i>GPC3</i>
coefficient	3.405	−0.078	−0.081	0.041	−0.067	0.056
Gene	<i>CFH</i>	<i>ZNF852</i>	<i>NLRP8</i>	<i>TCERG1L</i>	<i>FLJ45513</i>	<i>ARMCX2</i>
coefficient	0.107	0.122	−0.047	0.086	−0.042	−0.068
Gene	<i>MIF-AS1</i>	<i>A4GALT</i>	<i>ABCC6</i>	<i>LGSN</i>	<i>RADIL</i>	<i>NEU4</i>
coefficient	0.073	−0.081	−0.096	−0.105	−0.055	−0.052
Gene	<i>LINC00290</i>	<i>PCSK6</i>	<i>FGF10</i>	<i>SPIC</i>	<i>LRRC18</i>	<i>EFEMP2</i>
coefficient	−0.057	−0.064	−0.048	0.067	0.086	−0.048
Gene	<i>LINC01315</i>	<i>CYP2F1</i>	<i>CDX2</i>	<i>RASL10B</i>	<i>KLK5</i>	<i>RSPO3</i>
coefficient	−0.045	−0.044	−0.059	−0.070	−0.057	0.050
Gene	<i>PLAC4</i>	<i>TDRD6</i>	<i>ITGA7</i>	<i>KCNA4</i>	<i>SPATA5</i>	<i>HCAR3</i>
coefficient	−0.054	0.073	−0.082	−0.053	0.042	−0.048
Gene	<i>RBMS3-AS3</i>	<i>HOXA7</i>	<i>CTNNA2</i>	<i>FBXL19</i>	<i>LINC00475</i>	<i>SYT8</i>
coefficient	−0.065	−0.056	0.047	−0.082	−0.072	0.080
Gene	<i>CYP2S1</i>	<i>CDKL3</i>	<i>ALX4</i>	<i>GPRIN1</i>	<i>TRAV16</i>	<i>CASP12</i>
coefficient	0.084	0.053	0.069	0.051	0.044	−0.057
Gene	<i>TRPC4AP</i>	<i>FSTL3</i>	<i>TOP3B</i>			
coefficient	0.050	0.056	0.050			



**Figure 6.** The co-expression network of marker genes predicting HAPC. The co-expression network visualization for marker genes was obtained from the WGCNA with a weight cutoff of 0.2. The colors of the dots indicate the log(FC) values, where red means upregulated genes and blue means downregulated. The colored circles of the dots indicate the modules to which they belong.

#### 4. Discussion

After exposure to the plateau environment, the changes in hematopoiesis are more complicated in humans than in experimental animals. This study merged two datasets regarding the expression levels of leukocytes in peripheral blood in humans exposed to QTP

and identified genes that could predict HAPC. The results indicated that the acclimation to high altitude affects the immune system and that immune suppression would be restored as time progressed. HAPC patients and the adapted people had distinct expression patterns, and the steelblue module was related to both the disease phenotype and CMS. Then the prediction model for HAPC was also developed, and significant alterations in *HCAR3* and *LRRC18* were also observed during the beginning of the exposure to a high-altitude environment, suggesting that HAPC patients can be identified early.

At short timescales, the impact on the adaptive immune response cut a striking figure. The adaptive immune system relies on B cells and T cells, and in vitro experiments demonstrated that the human adaptive immune response is dependent on the oxygen level [23]. *CX3CR1* is stably expressed on CD8+ T cells [24], and its expression was highly suppressed on the third day following a quick climb to a high altitude. Physiological hypoxia impacts the immune cell function, ultimately controlling innate and adaptive immune responses, mainly through transcriptional regulation via hypoxia-inducible factors (HIFs) [25]. Here, we showed that HIF1 $\alpha$  was rapidly upregulated and associated with multiple biological processes, including intrinsic apoptotic signaling. Severe acute hypoxia promotes oxidative damage, which could trigger apoptotic mechanisms [26]. Nonetheless, a study conducted on the Kyrgyz population revealed that circulating levels of the apoptotic signal were diminished following acute and chronic exposure to high altitude conditions [27].

Inflammatory factors are reduced with acclimations to high altitude and are closely related to the pathogenesis of HAPC and high-altitude acclimation [28]. During the first month of exposure, immune-related genes were mainly affected. Genes associated with lymphocytes were decreased in leukocytes and increased in lymphocytes. The CIBERSORT predictions also suggested that the expression profile of white blood cells exhibited inconsistencies. This result committed to the study at 3700 m demonstrated an increase in neutrophils and a slight decrease in lymphocytes [29] and exposure to acute hypoxia marginally raised the neutrophils [30]. However, several investigations revealed contradictory results. Hypoxic preconditioning combined with altitude training increased T lymphocyte CD4/CD8 expression [31]. The latest cytological evidence showed that metabolic stress under hypoxia rapidly drives T cell exhaustion [32]. Different immune cell types might be related to distinct metabolic features.

HAPC is an adverse consequence of the adaption to high altitude, whereby the RBC count continues to rise. As a result, significantly higher concentrations of hemoglobin were found in HAPC patients than in matched controls [6]. However, genes involved in the metabolism of heme were decreased. Heme is essential in detecting and utilizing molecular oxygen and must be synthesized and degraded within an individual nucleated cell [33]. A chronic hypoxia increased erythrocyte lifespan was discovered in a recent study [34], which might explain why the heme metabolism was inhibited in HAPC. Furthermore, the DEGs between HAPC and control subjects were also enriched in cluster 1, rapidly upregulated after high altitude exposure, and inhibited in HAPC patients. These results indicate that people who developed HAPC were sensitive to hypoxia. In this study, we discovered that the critical module related to the disease was enriched in Wnt signaling, and it was also found to be related to adaptation to high altitude in Ethiopia [35]. In addition, the Wnt pathway impacts the ability of *Drosophila melanogaster* to complete its life cycle under hypoxia [36], reminding us that Wnt signaling might serve an essential role in hypoxia adaptation and may disturb the steady-state in HAPC patients.

Prediction models are designed to assist individuals in making decisions on the use of diagnostic testing and starting or stopping treatments [37]. Here, we constructed a logistic prediction model to discover 50 genes to predict HAPC. Serum inflammatory factor profiles revealed several proinflammatory factors that were higher expressed, including IL-1  $\beta$ , IL-2, IL-3, TNF- $\alpha$ , MCP-1, and IL-16 between the HAPC and plateau control groups in Han people [28]. However, the ages between the two compared groups showed a large difference ( $37.4 \pm 9.7$  in HAPC and  $27.2 \pm 3.5$  in control), and aging may be associated with an upregulated inflammatory response [38]. In another study, plasma biomarkers of

HAPC in older Tibetans (about 50 years old) were identified, including C4A, C6, CALR, MASP1, and CNDP1 [39]. Among these proteins, *MASP1* showed higher transcriptional expression ( $\log_2(\text{FC}) = 1.04$ ,  $p$  value = 0.07) than others in this dataset. This inconsistency might remind us that the mechanisms between Tibetans and Hans in developing HAPC was differ.

*HCAR3* and *LRRRC18* were also sensitive to exposure to high altitudes and might be able to predict whether a person would develop HAPC. *HCAR3* activates Gi signaling in immune cells and is also a therapeutic target for breast cancer and inflammatory bowel diseases [40–42], and these diseases cause hypoxia response. *HCAR3* activates ERK1/2, which in turn activates NF $\kappa$ B-mediated hypoxia responses. Its transcriptional expression also is more downregulated in smokers than nonsmokers at altitude, which might be related to the inhibition of the hypoxic response [43]. *LRRRC18* was also identified as a critical gene in coronary artery disease diagnosis [44], and its single-nucleotide polymorphisms are associated with systemic lupus erythematosus [45]. However, the relationship between these genes and the HAPC remains unclear. To investigate the underlying mechanism, additional research is required.

This study had several limitations. It was necessary to consider the restricted sample size while evaluating the results. Although split-sample validation was not performed, the goodness of fit of our logistic models was assessed. A cohort will be needed to evaluate the genes we identified here, along with a depth study of the mechanism of HAPC development.

## 5. Conclusions

To discover the molecular mechanism during human exposure to high altitudes, we explored two datasets related to short and long-term exposure to the QTP. The short-term microarray expression revealed an impaired adaptive immune response, and several immune cell types revealed dynamic alterations. In HAPC patients, the gene set for the heme metabolism was downregulated and also enriched in cluster1, which was immediately upregulated after arriving at high altitude. *HCAR3* and *LRRRC18* expression might relate to the development of HAPC. Here, we established a model and preliminarily explored people who might develop HAPC.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13071193/s1>, Figure S1: (A) The relative proportion of immune cell types in each group; (B) GSEA plot of geneset of “HALLMARK HEME METABOLISM”; (C) VENN plot of DEGs in two datasets; Figure S2: (A) Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA) (B) The result of cross-validate prediction.  $y$  indicates real disease state, fitted indicates the predict result of Logistic model; Table S1: The  $\log_2(\text{FC})$  of *LRRRC18* and *HCAR3* in different compared group.

**Author Contributions:** Conceptualization, H.W. and T.Z.; data curation, H.W.; formal analysis, H.W.; funding acquisition, T.Z.; methodology, H.W. and T.Z.; project administration, T.Z.; resources, H.W.; software, H.W. and P.S.; supervision, T.Z.; validation, D.L.; visualization, H.W.; writing—original draft, H.W. and D.L.; writing—review and editing, H.W., D.L., P.S., F.J. and T.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (2019QZKK0501); the Joint Grant from Chinese Academy of Sciences-People’s Government of Qinghai Province on Sanjiangyuan National Park (LHZX-2020-01) and Science and Technology Department of Qinghai Province Major Project “Sanjiangyuan National Park Animal Genome Program”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We acknowledge GEO database for providing their platforms and Xiao et al., Jiang et al. for uploading their meaningful datasets. We also want to acknowledge BioRender for generating the graphic abstract.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Jia, Z.; Zhao, X.; Liu, X.; Zhao, L.; Jia, Q.; Shi, J.; Xu, X.; Hao, L.; Xu, Z.; Zhong, Q.; et al. Impacts of the Plateau Environment on the Gut Microbiota and Blood Clinical Indexes in Han and Tibetan Individuals. *mSystems* **2020**, *5*, e00660-19. [[CrossRef](#)] [[PubMed](#)]
- Reeves, J.T.; Leon-Velarde, F. Chronic Mountain Sickness: Recent Studies of the Relationship between Hemoglobin Concentration and Oxygen Transport. *High Alt. Med. Biol.* **2004**, *5*, 147. [[CrossRef](#)] [[PubMed](#)]
- Martin, D.; Windsor, J. From Mountain to Bedside: Understanding the Clinical Relevance of Human Acclimatisation to High-Altitude Hypoxia. *Postgrad. Med. J.* **2008**, *84*, 622. [[CrossRef](#)] [[PubMed](#)]
- Jefferson, J.A.; Escudero, E.; Hurtado, M.-E.; Kelly, J.P.; Swenson, E.R.; Wener, M.H.; Burnier, M.; Maillard, M.; Schreiner, G.F.; Schoene, R.B.; et al. Hyperuricemia, Hypertension, and Proteinuria Associated with High-Altitude Polycythemia. *Am. J. Kidney Dis.* **2002**, *39*, 1135–1142. [[CrossRef](#)] [[PubMed](#)]
- Viault, F. On the Large Increase in the Number of Red Cells in the Blood of the Inhabitants of the High Plateaus of South America. In *High Altitude Physiology*; West, J.B., Ed.; Hutchinson Ross: Stroudsburg, PA, USA, 1981; pp. 333–334.
- Jiang, C.; Liu, F.; Luo, Y.; Li, P.; Chen, J.; Xu, G.; Wang, Y.; Li, X.; Huang, J.; Gao, Y. Gene Expression Profiling of High Altitude Polycythemia in Han Chinese Migrating to the Qinghai-Tibetan Plateau. *Mol. Med. Rep.* **2012**, *5*, 287–293. [[CrossRef](#)]
- Chen, K.; Li, N.; Fan, F.; Geng, Z.; Zhao, K.; Wang, J.; Zhang, Y.; Tang, C.; Wang, X.; Meng, X. Tibetan Medicine Duoxuekang Capsule Ameliorates High-Altitude Polycythemia Accompanied by Brain Injury. *Front. Pharmacol.* **2021**, *12*, 680636. [[CrossRef](#)]
- Windsor, J.S.; Rodway, G.W. Heights and Haematology: The Story of Haemoglobin at Altitude. *Postgrad. Med. J.* **2007**, *83*, 148–151. [[CrossRef](#)]
- Fan, X.; Ma, L.; Zhang, Z.; Li, Y.; Hao, M.; Zhao, Z.; Zhao, Y.; Liu, F.; Liu, L.; Luo, X.; et al. Associations of High-Altitude Polycythemia with Polymorphisms in PIK3CD and COL4A3 in Tibetan Populations. *Hum. Genomics* **2018**, *12*, 37. [[CrossRef](#)]
- Liu, Y.-S.; Huang, H.; Zhou, S.-M.; Tian, H.; Li, P. Excessive Iron Availability Caused by Disorders of Interleukin-10 and Interleukin-22 Contributes to High Altitude Polycythemia. *Front. Physiol.* **2018**, *9*, 548. [[CrossRef](#)]
- Xu, J.; Yang, Y.-Z.; Tang, F.; Ga, Q.; Tana, W.; Ge, R.-L. EPAS1 Gene Polymorphisms Are Associated With High Altitude Polycythemia in Tibetans at the Qinghai-Tibetan Plateau. *Wilderness Environ. Med.* **2015**, *26*, 288–294. [[CrossRef](#)]
- Liu, L.; Zhang, Y.; Zhang, Z.; Zhao, Y.; Fan, X.; Ma, L.; Zhang, Y.; He, H.; Kang, L. Associations of High Altitude Polycythemia with Polymorphisms in EPHA2 and AGT in Chinese Han and Tibetan Populations. *Oncotarget* **2017**, *8*, 53234. [[CrossRef](#)] [[PubMed](#)]
- Xiao, J.; Li, X.; Fan, X.; Fan, F.; Lei, H.; Li, C. Gene Expression Profile Reveals Hematopoietic-Related Molecule Changes in Response to Hypoxic Exposure. *DNA Cell Biol.* **2020**, *39*, 548–554. [[CrossRef](#)] [[PubMed](#)]
- Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)] [[PubMed](#)]
- Kumar, L.; E Futschik, M. Mfuzz: A Software Package for Soft Clustering of Microarray Data. *Bioinformatics* **2007**, *2*, 5–7. [[CrossRef](#)] [[PubMed](#)]
- Liao, J.G.; Chin, K.-V. Logistic Regression for Disease Classification Using Microarray Data: Model Selection in a Large p and Small n Case. *Bioinformatics* **2007**, *23*, 1945–1951. [[CrossRef](#)] [[PubMed](#)]
- Langfelder, P.; Horvath, S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
- Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
- Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)]
- Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
- Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
- Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)] [[PubMed](#)]
- Hochgerner, M.; Sturm, E.M.; Schnoegl, D.; Kwapiszewska, G.; Olschewski, H.; Marsh, L.M. Low Oxygen Levels Decrease Adaptive Immune Responses and Ameliorate Experimental Asthma in Mice. *Allergy* **2022**, *77*, 870–882. [[CrossRef](#)] [[PubMed](#)]
- Yamauchi, T.; Hoki, T.; Oba, T.; Jain, V.; Chen, H.; Attwood, K.; Battaglia, S.; George, S.; Chatta, G.; Puzanov, I.; et al. T-Cell CX3CR1 Expression as a Dynamic Blood-Based Biomarker of Response to Immune Checkpoint Inhibitors. *Nat. Commun.* **2021**, *12*, 1402. [[CrossRef](#)]
- Chen, Y.; Gaber, T. Hypoxia/HIF Modulates Immune Responses. *Biomedicines* **2021**, *9*, 260. [[CrossRef](#)]

26. Coimbra-Costa, D.; Alva, N.; Duran, M.; Carbonell, T.; Rama, R. Oxidative Stress and Apoptosis after Acute Respiratory Hypoxia and Reoxygenation in Rat Brain. *Redox Biol.* **2017**, *12*, 216–225. [[CrossRef](#)]
27. Kosanovic, D.; Platzeck, S.M.; Petrovic, A.; Sydykov, A.; Maripov, A.; Mamazhakypov, A.; Sartmyrzaeva, M.; Muratali Uulu, K.; Cholponbaeva, M.; Toktosunova, A.; et al. Circulating Apoptotic Signals during Acute and Chronic Exposure to High Altitude in Kyrgyz Population. *Front. Physiol.* **2019**, *10*, 54. [[CrossRef](#)] [[PubMed](#)]
28. Yi, H.; Yu, Q.; Zeng, D.; Shen, Z.; Li, J.; Zhu, L.; Zhang, X.; Xu, Q.; Song, H.; Kong, P. Serum Inflammatory Factor Profiles in the Pathogenesis of High-Altitude Polycythemia and Mechanisms of Acclimation to High Altitudes. *Mediat. Inflamm.* **2021**, *2021*, 8844438. [[CrossRef](#)]
29. Huang, H.; Liu, B.; Wu, G.; Xu, G.; Sun, B.-D.; Gao, Y.-Q. Hematological Risk Factors for High-Altitude Headache in Chinese Men Following Acute Exposure at 3700 m. *Front. Physiol.* **2017**, *8*, 801. [[CrossRef](#)]
30. Pedersen, B.K.; Steensberg, A. Exercise and Hypoxia: Effects on Leukocytes and Interleukin-6-Shared Mechanisms? *Med. Sci. Sports Exerc.* **2002**, *34*, 2004–2013. [[CrossRef](#)]
31. Wang, X.; Huang, L.; Gao, H. Effects of Hypoxic Preconditioning Combined with Altitude Training on CD55, CD59 and the Immune Function of Swimmers. *Ann. Palliat. Med.* **2021**, *10*, 509–517. [[CrossRef](#)]
32. Scharping, N.E.; Rivadeneira, D.B.; Menk, A.V.; Vignali, P.D.A.; Ford, B.R.; Rittenhouse, N.L.; Peralta, R.; Wang, Y.; Wang, Y.; DePeaux, K.; et al. Mitochondrial Stress Induced by Continuous Stimulation under Hypoxia Rapidly Drives T Cell Exhaustion. *Nat. Immunol.* **2021**, *22*, 205–215. [[CrossRef](#)] [[PubMed](#)]
33. Shibahara, S.; Han, F.; Li, B.; Takeda, K. Hypoxia and Heme Oxygenases: Oxygen Sensing and Regulation of Expression. *Antioxid Redox Signal* **2007**, *9*, 2209–2225. [[CrossRef](#)]
34. Tang, F.; Feng, L.; Li, R.; Wang, W.; Liu, H.; Yang, Q.; Ge, R.-L. Inhibition of Suicidal Erythrocyte Death by Chronic Hypoxia. *High Alt. Med. Biol.* **2019**, *20*, 112–119. [[CrossRef](#)]
35. Scheinfeldt, L.B.; Soi, S.; Thompson, S.; Ranciaro, A.; Woldemeskel, D.; Beggs, W.; Lambert, C.; Jarvis, J.P.; Abate, D.; Belay, G.; et al. Genetic Adaptation to High Altitude in the Ethiopian Highlands. *Genome Biol.* **2012**, *13*, R1. [[CrossRef](#)] [[PubMed](#)]
36. Gersten, M.; Zhou, D.; Azad, P.; Haddad, G.G.; Subramaniam, S. Wnt Pathway Activation Increases Hypoxia Tolerance during Development. *PLoS ONE* **2014**, *9*, e103292. [[CrossRef](#)]
37. Shipe, M.E.; Deppen, S.A.; Farjah, F.; Grogan, E.L. Developing Prediction Models for Clinical Use Using Logistic Regression: An Overview. *J. Thorac. Dis.* **2019**, *11*, S574–S584. [[CrossRef](#)] [[PubMed](#)]
38. Wu, D.; Meydani, S.N. Age-Associated Changes in Immune and Inflammatory Responses: Impact of Vitamin E Intervention. *J. Leukoc. Biol.* **2008**, *84*, 900–914. [[CrossRef](#)] [[PubMed](#)]
39. Wang, Z.; Liu, F.; Ye, S.; Jiang, P.; Yu, X.; Xu, J.; Du, X.; Ma, L.; Cao, H.; Yuan, C.; et al. Plasma Proteome Profiling of High-Altitude Polycythemia Using TMT-Based Quantitative Proteomics Approach. *J. Proteom.* **2019**, *194*, 60–69. [[CrossRef](#)]
40. Kapolka, N.J.; Isom, D.G. HCAR3: An Underexplored Metabolite Sensor. *Nat. Rev. Drug Discov.* **2020**, *19*, 745. [[CrossRef](#)]
41. McGuire Sams, C.; Shepp, K.; Pugh, J.; Bishop, M.R.; Merner, N.D. Rare and Potentially Pathogenic Variants in Hydroxycarboxylic Acid Receptor Genes Identified in Breast Cancer Cases. *BMC Med. Genomics* **2021**, *14*, 284. [[CrossRef](#)]
42. Wnorowski, A.; Wnorowska, S.; Kurzepa, J.; Parada-Turska, J. Alterations in Kynurenine and NAD<sup>+</sup> Salvage Pathways during the Successful Treatment of Inflammatory Bowel Disease Suggest HCAR3 and NNMT as Potential Drug Targets. *Int. J. Mol. Sci.* **2021**, *22*, 13497. [[CrossRef](#)] [[PubMed](#)]
43. Nicholson, S.J.; Munster, S.K.; Uyhelji, H.A.; Hutchings, D.C.; White, V.L.; Booth, J.L.; Burian, D.M.; Metcalf, J.P. Transcriptional Responses to Altitude-Induced Hypoxia in Bronchial Epithelium, Broncho-Alveolar Cells, and Blood Differ between Smokers and Nonsmokers. United States, Department of Transportation, Federal Aviation Administration, Office of Aviation, Civil Aerospace Medical Institute: Washington, DC, USA; Venesco LLC: Chantilly, VA, USA, 2021. Available online: [https://www.faa.gov/data\\_research/research/med\\_humanfacs/oamtechreports/2020s/media/202125.pdf](https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2020s/media/202125.pdf) (accessed on 17 March 2022).
44. Yang, Y.; Xu, X. Identification of Key Genes in Coronary Artery Disease: An Integrative Approach Based on Weighted Gene Co-Expression Network Analysis and Their Correlation with Immune Infiltration. *Aging* **2021**, *13*, 8306–8319. [[CrossRef](#)] [[PubMed](#)]
45. Wang, C.; Ahlford, A.; Järvinen, T.M.; Nordmark, G.; Eloranta, M.-L.; Gunnarsson, I.; Svenungsson, E.; Padyukov, L.; Sturfelt, G.; Jönsen, A.; et al. Genes Identified in Asian SLE GWASs Are Also Associated with SLE in Caucasian Populations. *Eur. J. Hum. Genet.* **2013**, *21*, 994–999. [[CrossRef](#)] [[PubMed](#)]