The effect of altitude on the expression of immune-related genes in Peruvian rural indigenous

Authors: Luis Jaramillo-Valverde[1,2+], Gilderlanio Santana de Araújo[1+], Julio A. Poterico[1,2], Catalina Martinez-Jaramillo[1], Vicky Roa-Linares[1], Sandra Alvites-Arrieta[1], Nelis Pablo-Ramirez[1], Milward Ubillus[1], Diana Palma-Lozano[1], Carolina Silva-Carvalho[3], Luca Vasconcelos-da-Gama[3], Lucas F Costa[3], Eduardo Tarazona-Santos[3], Soumya Raychaudhuri[4], Heinner Guio[1,2*]

[1]Universidad de Huánuco, Huánuco, Peru
[2]INBIOMEDIC, Research and Technological Center, Lima, Peru
[3]Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Belo Horizonte, Universidade Federal de Minas Gerais, MG, Brazil
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

[+]Equal contribution
*Corresponding Author:
E-mail: heinnerguio@gmail.com

AUTHOR CONTRIBUTIONS

Study design: LJ-V, SR and HG. Performed the experiments: LJ-V, GSA,JAP,CMJ,VRL,SAA,NPR,MU,DPL. Analyzed the data: LJ-V, GSA, CSC, LVG, LFC,ETS,SR and HG All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Lluis Quintana-Murci and David Comas for their constructive comments and perspectives. We thank all those who facilitated the recruitment of participants, including the Direccion Regional de Salud from Huanuco; and all participants in this study.

FUNDING INFORMATION

The Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (FONDECYT) and BANCO MUNDIAL with the contract number 19-2019-BM-INC.INV supported this study.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Luis Jaramillo-Valverde https://orcid.org/0000-0001-5093-6067
Gilderlanio Santana de Araújo https://orcid.org/0000-0001-9199-9419
Julio A. Poterico https://orcid.org/0000-0001-7838-3505
Catalina Martinez-Jaramillo https://orcid.org/0000-0001-6846-0140
Vicky Roa-Linares https://orcid.org/0000-0001-9253-615X
Sandra Alvites-Arrieta -
Nelis Pablo-Ramirez -
Milward Ubillus https://orcid.org/0000-0002-3684-9394
Diana Palma-Lozano https://orcid.org/0000-0003-4520-7374
Carolina Silva-Carvalho https://orcid.org/0000-0002-6356-6779
Luca Vasconcelos-da-Gama -
Lucas F Costa https://orcid.org/0000-0001-9363-0121
Eduardo Tarazona-Santos https://orcid.org/0000-0003-3508-3160
Soumya Raychaudhuri https://orcid.org/0000-0002-1901-8265
Heinner Guio https://orcid.org/0000-0003-0078-1188

ABSTRACT

Background: Some genes associated with immune response have been elucidated in some Andean populations, which may explain part of the immune system adaptation and pathogen response in high-altitude residents. We performed differential expression analysis and mainly focused on genes with high fold-change between groups and those genes with differential expression related to the immune system. Methods: We performed a transcriptome analysis of Peruvian individuals, primarily from rural areas, with high genetic backgrounds from ancient indigenous people, exposed to different living altitudes (high vs low). We collected peripheral blood

mononuclear cells (PBMC) from 62 volunteers and exposed them to bacterial lipopolysaccharide (LPS), Pam3CSK4 (a synthetic triacylated lipopeptide responsible for bacterial components), and R848 (an imidazoquinoline compound related to viral nucleic acids). Results: Differential expression analysis results and data integration analysis with immune system pathway data exhibit a set of 17 genes associated with the immune system, nine down-regulated genes (*ITGAX, CCL22, CSF1, CXCL8, IL12A, MMP9, CSF2, PTGS2, ENSG00000138685.17*), and eight up-regulated genes (*HLA-DPB1, FN1, CD36, MMP2, HLA-DRB1, FCGR1A, CCL17, HLA-DRB5*). These genes are notably enriched in antigen-presenting cells, such as dendritic cells and macrophages. Conclusion: The differential expression of immune system genes in native inhabitants with a prominent Andean genetic background highlights the active involvement of both the innate and adaptive immune systems. Remarkably, our study revealed a distinct gene expression signature in our population, setting it apart from other transcriptomic studies conducted thus far.

Keywords: altitude, expression, immune-related genes, rural indigenous

AUTHOR SUMMARY

Living in high-altitude regions poses several environmental challenges for humans, including reduced oxygen levels, increased exposure to ultraviolet radiation, cold temperatures, and altered responses to pathogens, among other factors. Here we show genomic and transcriptome analyses of blood cells in Peruvian individuals, primarily from rural areas, from indigenous populations. We previously showed a robust adaptation signals related to altitude in Peruvian and Bolivian populations, identifying a significant association with the DUOX2 gene, which plays a crucial role in maintaining homeostasis in thyroid hormone production, as well as influencing the innate immune system and inflammatory response. We now find that differential expression analysis exhibits a set of 17 genes associated with the immune system, nine down-regulated genes (ITGAX, CCL22, CSF1, CXCL8, IL12A, MMP9, CSF2, PTGS2, ENSG00000138685.17), and eight up-regulated genes (HLA-DPB1, FN1, CD36, MMP2, HLA-DRB1, FCGR1A, CCL17, HLA-DRB5). These results suggest a distinct gene expression signature in our population, setting it apart from other transcriptomic studies conducted thus far.

INTRODUCTION

Approximately 1% of the world's population resides in high-altitude areas, specifically at elevations over 2500 meters above sea level (m.a.s.l.). Countries such as Peru and Bolivia have a significant portion (30 percent) of their population living at high altitudes in the Andean mountains (1). Living in high-altitude regions poses several environmental challenges for humans, including reduced oxygen levels, increased exposure to ultraviolet radiation, cold temperatures, and altered responses to pathogens, among other factors (2). Pioneering researchers in the field of population genomics at high altitude have suggested that genes of the Hypoxia Inducible Factor (HIF) pathway have evolved under positive natural selection (3,4), while the genes *ENDRA*, *PRKAA1*, and *NOS2A* have been suggested as being associated with blood pressure regulation and reproductive success in Andeans (5). Additionally, genetic studies have demonstrated that native Andean highlanders depict positive selection of cardiovascular genes, such as *BRINP3*, *NOS2*, *TBX5*(6), and the non-coding region *HAND2-AS1*(7).

A recent study confirms these findings and provides further insights. Analyses of a genome-wide study involving 286 individuals across Peru and over-representation functional analysis identified ten positive natural selection signals among highlander populations related to cardiovascular function (*TLL1*, *DUSP27*, *TBX5*, *PLXNA4*, *SGCD*), the hypoxia-inducible factor pathway (*TGFA*, *APIP*), skin pigmentation (*MITF*), glucose (*GLIS3*) and glycogen metabolism (*PPP1R3C*, *GANC*) (8), as well as genes involved in the *TP53* pathway, such as *SP100*, *DUOX2,* and *CLC* (9).

These studies indicate that Peruvian Andean populations have exhibited distinct adaptations compared to other groups like Tibetans, Sherpas, or Ethiopians in response to high-altitude living. Moreover, certain genes related to immune response have been identified in Andean populations as being positively selected. These genes might provide insights into the adaptation of the immune system and responses to pathogens among those residing at high altitudes. Examples of robust adaptation signals related to altitude in Peruvian and Bolivian populations can be found in the studies conducted by Borda et al. (7) and Jacovas et al. (9). Specifically, these studies identified a significant association with the DUOX2 gene, which plays a crucial role in

maintaining homeostasis in thyroid hormone production, as well as influencing the innate immune system and inflammatory response.

Likewise, Caro-Consuegra et al. found positive selection in Peruvian Andean populations in *CD40*, *TNIP1*, and *CHIA* genes, which are involved in the innate immune system (8). Furthermore, genes such *ANXA1*, *CD44*, *ADGRE1*, *KCNA3,* and *FCRL4* among others, seem to have been positively selected due to *Mycobacterium tuberculosis* in Andean Ecuadorian populations, and this event could have emerged before the European contact with aboriginal communities (10). However, none of the aforementioned genomic studies have included functional analysis to correlate genomic findings with other biomarkers and phenotypes.

To our knowledge, there is currently no investigation assessing the expression of immune genes in native participants (*e.g.*, Andean Peruvians) with long-term exposure to high altitude–*i.e.* residence since birth or childhood. For that reason, we aim in our study, after characterizing the genetic structure and ancestry of the studied population, we aimed to conduct genomic and transcriptome analyses of blood cells in Peruvian individuals, primarily from rural areas, from indigenous populations. These individuals are exposed to varying living altitudes, comparing high versus low altitudes individuals. We collected peripheral blood mononuclear cells (PBMC) from 62 volunteers, which were exposed to bacterial lipopolysaccharide (LPS), Pam3CSK4 (a synthetic triacylated lipopeptide responsible for bacterial components), and R848 (an imidazoquinoline compound related to viral nucleic acids) (11). Subsequently, we discerned variations in gene expression profiles among populations residing at low and high altitudes. Our analysis focused predominantly on genes that displayed differential expression, with a emphasis on those significantly enriched in immunological responses.

MATERIAL AND METHODS

Population and samples

We studied the population from Huanuco, a department in Central Peru, focusing on a region in the eastern slope of the Andes that includes the Andean Mountains and

167 the Amazonian Tropical Forest (in particular the so-called Amazon Yunga or Higher
168 Amazonia). In addition to biological samples, we collected demographic data such as
169 sex, age, body mass index (BMI), and altitude of residence. In the first part of the
170 study, 35 individuals from the studied population were analyzed for genomic diversity.
171 In the second part of the study, we studied 62 individuals for transcriptome analyses,
172 with an overlap of 19 individuals between the two parts of the study. The study was
173 approved by the Inmensa Ethics Committee of Peru (number 0011-2021). Written
174 informed consent was obtained from all participants involved in this study.
175
176 **Genetic structure and ancestry of the studied Huanuco population**
177 Genotyping and Data Quality Control (QC). Thirty-five individuals were genotyped with
178 the Illumina GSA MG v3 array by Macrogen (Seul, Korea). After Illumina Genome
179 Studio processing, we obtained 737 534 SNVs. After further QC with MosaiQC
180 software (https://github.com/ldgh/Smart-cleaning-public), a tool that automates the
181 Laboratory of Human Genetic Diversity (LDGH) quality control pipeline using PLINK,
182 703 515 SNVs remained.
183
184 Individual ancestry analyses. For the genetic clustering analysis, we ran
185 ADMIXTURE(12) 14 times in the unsupervised mode for different values of K (2 to 14),
186 and present the results with the lowest log-likelihood, and for these results, we used
187 cross-validation (CV) errors to assess the K values that better fit the data and the
188 results. ADMIXTURE analysis assumes independence among SNPs, for this reason,
189 we pruned all datasets for linkage disequilibrium (LD) (r^2<0.4) using PLINK 1.7. Firstly,
190 we run ADMIXTURE with K=3 (189 837 SNVs) to infer African, European and Native
191 American ancestry using the entire dataset in the Table S1, and used these results for
192 relatedness analyses (see below). Then, we pruned our dataset for the few related
193 individuals (see below) and ran again 14 replicates of ADMIXTURE for K from 4 to 12,
194 considering only unrelated individuals and 162 957 unlinked SNVs, presenting the
195 results of the most likely run, assessing their CV error for different K. Because the
196 ADMIXTURE results with K=3 informed us that African ancestry was negligible in our
197 dataset (see below), we retired African individuals for the ADMIXTURE analyses for
198 K>3. Because ADMIXTURE analyses require unrelated individuals, we estimated
199 Kinship coefficients between Huánuco individuals (Φij) REAP (Relatedness Estimation
200 in Admixed Populations), which is more appropriate for admixed individuals (13). The

201 estimates of Φij by REAP are conditioned on the estimated individual ancestry
202 proportions (IAP, estimated with ADMIXTURE) for K=3 parental populations and the
203 3 parental population's allele frequencies per SNV. To identify related individuals to
204 be pruned for ancestry analyses we used NAToRA, a relatedness-pruning method to
205 minimize the loss of dataset in genetic and omics analyses (14), using the relatedness
206 cutoff of Φij>0.06.
207

208 **Transcriptome analyses in the Huanuco population**

209 Preparation of biological samples. In 62 individuals, PBMCs were isolated from whole
210 blood by Ficoll-Paque centrifugation within 4 hours from collection and cryopreserved.
211 PBMCs were cultured in RPMI-1640 (Fisher) supplemented with 10% heat-inactivated
212 FBS (FBS premium, US origin, Wisent) and 1% L-glutamine (Fisher). All samples were
213 stimulated by pathogens molecules (virus and bacteria) in independent processes: for
214 each of the tested individuals, and one process without stimulation. PBMCs (2 million
215 per condition) were stimulated for 4 hours at 37° C with 5% CO2 with bacterial
216 lipopolysaccharide (LPS), Pam3CSK4 (Invitrogen - Cat#tlrl-pms) a synthetic
217 triacylated lipopeptide responsible for bacterial components, and R848 ((Invitrogen -
218 Cat#tlrl-r848-5) an imidazoquinoline compound, related to viral nucleic acids,
219 according to the manufacturer's instructions.

220 Library Construction, Quality Control, and Sequencing. The process of library
221 construction involved the purification of messenger RNA from total RNA using
222 magnetic beads attached to poly-T oligos (TruSeq RNA Sample Prep Kit v2, Illumina
223 Inc.). Subsequently, the mRNA was fragmented, and the first strand of cDNA was
224 synthesized using random hexamer primers. This was followed by the synthesis of the
225 second strand cDNA. The library was considered complete after undergoing end
226 repair, A-tailing, adapter ligation, size selection, amplification, and purification.
227 Quantification and size distribution analysis of the library was performed using Qubit
228 and real-time PCR, as well as a Bioanalyzer (Agilent, CA, USA). The libraries were
229 quantified to determine their concentration and assessed for size distribution using the
230 Bioanalyzer (Agilent, CA, USA). Once quantified, the libraries with appropriate
231 concentrations and data amounts were pooled together. Finally, the pooled libraries
232 were sequenced using Illumina platforms. The total data output sequencing product
233 was 514.1 Gb. The Data Quality Summary is available in Supplementary Material 1.

234 Transcriptome data pre-processing and differential gene analysis. Total RNA
235 sequencing was performed by a pair-end strategy for all samples using Novaseq
236 PE150 (Illumina Inc). Raw data is stored in .fastq format (.fq.gz compressed) for
237 analysis. The NOVOGENE provides raw RNA sequencing data by removing adapters
238 and low-quality reads regarding bases that cannot be determined (N > 10%) and those
239 with QScore <= 5. For our quality control analysis, each sample in .fastq was
240 resubmitted for the quality control pipeline using fastQC and multiQC to generate
241 sequencing quality reports to verify quality per bases, GC content, the proportion of
242 undefined nucleotide bases, and the presence of adapter sequences. The Phred score
243 Q30 was the base threshold for filtering high-quality samples. Read alignment was
244 performed with STAR with the Release 41 (GRCh38.p13) genome reference. The
245 software STAR was set to consider that a read may overlap on the same or opposite
246 strand and to follow The Encyclopedia of DNA Elements (ENCODE) best practices.
247 The STAR outputs alignments in a sorted BAM format. After genome alignment, read
248 quantification was performed with HTSeq with the GRCh38 human genome annotation
249 file (.gff). The read-counting process considered the large set of transcripts, in total
250 61801 coding and non-coding elements cataloged in ENCODE.

251 Differential gene expression analysis. The 17 control samples (i.e. not stimulated) and
252 62 treated PBMC samples LPS (n = 16), Pam3CSK4 (n = 24), R (n = 22)) were
253 organized in two groups: low altitude (less than 1000 m.a.s.l.) and high altitude (>3000
254 m.a.s.l.) to perform differential expression analysis (DEA) with the edgeR package
255 (23). The dispersion of gene expression was estimated to fit a generalized linear model
256 and covariate correction for individual (categorical variable), sex (categorical variable),
257 BMI (numerical variable), age (numerical variable), and stimulus by pathogens (one
258 virus and two bacteria, categorical variable). Genomic ancestry was not included as
259 covariates because only 19 individuals have both array genotyping and transcriptome
260 data. Those genes with a false-discovery rate (FDR < 0,05) and fold-change (|FC| >
261 2) were considered differentially expressed genes (DEG). The distribution of DEG
262 genes (down- and up-regulated) regarding statistical significance and genes from the
263 immune system, before and after stimulation, was represented with volcano plots with
264 *ggplot2* implemented in the R package.
265
266 Functional Enrichment Analysis. We conducted a thorough functional enrichment

analysis of gene sets, encompassing both upregulated and downregulated genes, employing Gene Ontology (GO) and utilizing the ClusterProfile (15) package in the R language. To enhance statistical rigor, multiple tests were corrected using the False Discovery Rate (FDR). The outcomes, comprising the top biological processes, were visually represented through bar plots and networks depicting the interrelation of genes and processes.

Immune system genes dataset. DEA results were merged with biological pathway data with a focus only on Immune System genes. In the Reactome database, the adaptive *immune system* catalogs the interaction of 868 genes, the *innate immune system* showed 724 transcripts, and 1160 genes comprise the *cytokine signaling pathway*. The Reactome data is available at reacomte.org, and we used version 81 released on June 16, 2022.

RESULTS

Genomic diversity and ancestry of Huanuco population

ADMIXTURE analyses (Figure 1, K=3 and 4) of 35 individuals reveals that Huanuco individuals are predominantly Native American (87.3% on average) with a low European ancestry (average: 11.82%) and negligible African ancestry (<1%) (Table S2). Given the negligible African ancestry, which is consistent with the history of the Huanuco region, we present ADMIXTURE results excluding African individuals (results in Tables S3).

Analyses assuming K=5 ancestry clusters (K=4 and 5) reveal that Native American ancestry of Huánuco individuals may be dissected in a predominant Andean-associated cluster (84.5% [SD:5.6%]of the total ancestry) and a minor Amazonian-associated cluster (3.7% [SD: 1.1%] of the total ancestry). Interestingly, the Andean-associated ancestry cluster is higher in highlanders (average 86%) than in lowlanders (average 82%, <600 m vs. >3000m, Kruskal-Wallis $H=4.057$, $p<0.05$, $N=35$).

A higher-resolution clustering (K=6 and 7) reveals that Huanuco individuals are similar to Tallanes and Moche (indigenous groups from the northern Peruvian Coast) and Chacapoyas (an indigenous group from the northern Andes/Amazonian geographic interface). Altogether, Huanuco individuals are similar to Tallanes, Moche and Chachapoyas from Northern Peru, and all of them present different levels of admixture between Andean and Amazonian ancestries.

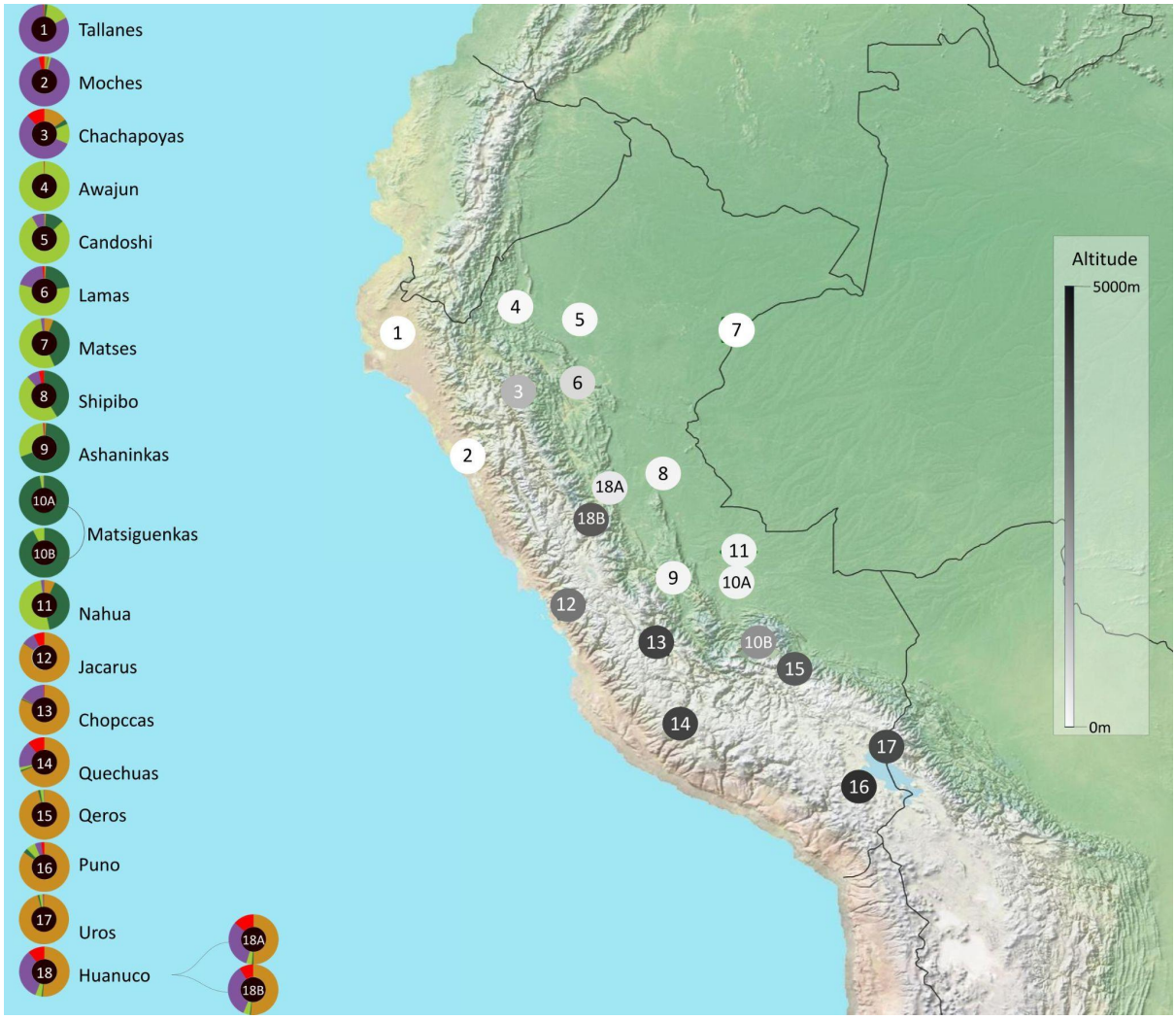


Figure 1. Genetic and geographic landscapes for Peruvian populations. (A): Genetic structure of seventeen native populations from coast (n. 1 and 2), Amazon (n. 3-11), Andes (n. 12-17) and Huanuco samples (18A and 18B), inferred by ADMIXTURE (K=6). Donuts plots show the average percentages of individuals for six ancestry clusters in each population. Four clusters were associated with distinct Native American groups: one Andean (brown), one North Amazonian (light green), one South Amazonian (dark green) and one coast (purple). Two clusters were associated with non-Native continental ancestries, both European (red), caused by admixture with these populations in the last 500 years. **(B)**: Information for geographic location and altitude of samples. The color of the population markers goes from white (at sea level) to black (5000 meters above the sea). **(C)**: Individual ancestry inferred by ADMIXTURE (K=3) for Huanuco samples, where we can see the clusters Native American (green), African (blue) and European (red).

318 **Demographic data overview of Huanuco transcriptome study**

319

320 Forty-three (69.4% or total) participants were highland sojourners (3226 – 3539
321 m.a.s.l.) whereas 19 (30.6%) inhabited lowlands (596 – 634 m.a.s.l.). The samples
322 were obtained at the residences of participants and comprise 11 males (11.7%) and
323 51 females (82.3%) with an age mean of 48 years (SD±15). The samples showed a
324 BMI median of 26.4 Kg/m^2. It is worth noting that almost all participants had ancestors
325 (e.g. parents, grandparents) from high Andean areas of the Huánuco department
326 (Table 1).

327

328 **Table 1.** Demographic data of Peruvian individuals, from rural areas, with high genetic
329 backgrounds from ancient indigenous people, exposed to different living altitudes (high vs
330 low).

331

Variables	Total		Living altitudes				p-Value
			Low		High		
	N	%	N	%	N	%	
Sex							
Female	51	82.26	15	78.95	36	83.72	
Male	11	17.74	4	21.05	7	16.28	0.650
Age (years)	-	-	42	(34 - 58) †	48	(37 - 64) †	0.392
Cholesterol (mg/dl)	-	-	169	(136 – 185)†	166	(145 – 204)†	0.262
Hemoglobin (g/dl)	-	-	14.4	(13.4 - 15.1)†	14.1	(13.4 - 14.4)†	0.267
Glucose (mg/dl)	-	-	87	(78 - 96)†	83	(76 – 89)†	0.184
BMI (kg/m^2)	-	-	28.3	(24.3 - 30.2)†	25.1	(22.3 - 28.5)†	0.031

332 **Median. †(Q1 – Q3)*
333 *BMI (Body mass index)*
334 *Statistically significant (p < 0.05, Chi squared).*

335

336 **High altitude vs. low altitude differential expression analysis**

337 We obtained raw sequencing data for a total of 61,806 transcripts that were
338 subsequently analyzed to compare gene expression between individuals living at high
339 and low altitudes. Specifically, we explored how altitude affects the transcriptional
340 response to pathogen stimulations.

341

342 In the analysis comparing individuals from low and high altitudes, we identified 843
343 genes that exhibited differential expression in our model considering covariates
344 individual, sex, BMI, age, and stimulus by pathogens (virus and bacteria). Out of these,

397 genes were down-regulated, while 446 genes were up-regulated in high-altitude individuals compared to those at low altitude (FDR < 0,05 and |FC| > 2, Supplementary Table 4). We visualized the most statistically significant differentially expressed genes in Figure 2.

Differential expression analysis was conducted on all samples, both with and without stimulation, while also accounting for covariates. The two genes with the highest differential gene expression between highlanders and lowlanders were: (i) the *APELA* (Apelin Receptor Early Endogenous Ligand) gene, down-regulated in high-altitude, with a FC=-11.87 and FDR<= 2.56E-06, associated with processes such as angiogenesis, vasculogenesis, heart development, hormone activity, and endoderm development (16,17); and (ii) *HS3ST2* (heparan sulfate-glucosamine 3-sulfotransferase 2), one of the commonest isoenzymes involved in heparan sulfate (HS) structural modifications (18), showed significant up-regulation in high-altitude individuals (FC=9.2, FDR <= 6.78E-06).

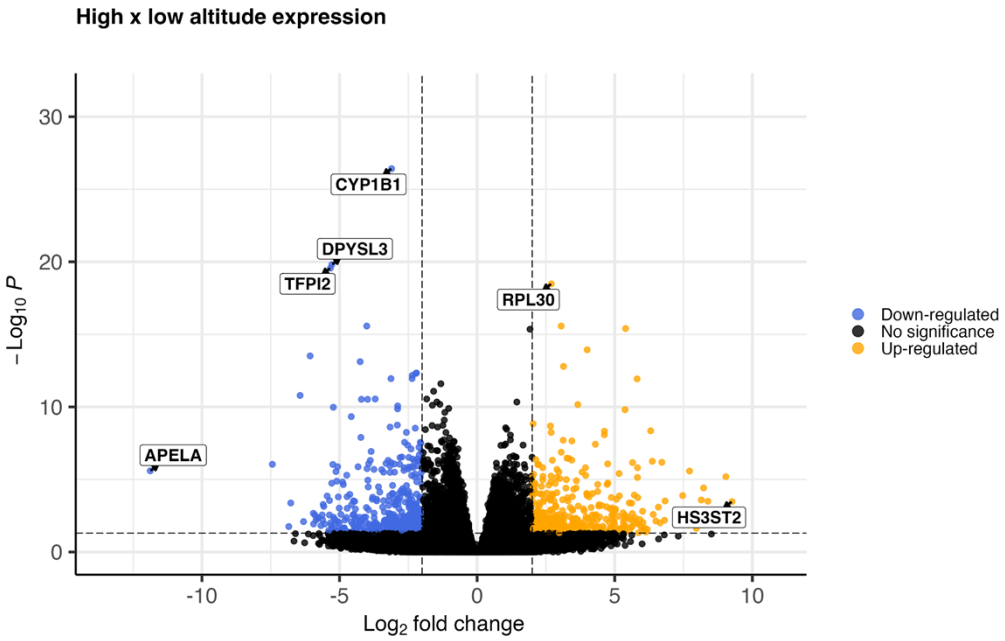


Figure 2. The volcano plot highlights the highest significant genes with differential gene expression regarding high and low-altitude samples. When comparing gene expression between individuals living at low and high altitudes, we identified 843 genes with differential expression. Out of these, 397 genes were down-regulated (depicted as blue points), while 446 genes were up-regulated (represented as orange points) in high-altitude individuals as compared to those at low altitudes.

Differential expression of immune genes possibly to adaptation to pathogens

As expected, following individual pathogen stimulation, a group of genes associated with the immune system exhibited differential expression at different altitudes. Our differential expression analysis results, coupled with data integration involving immune system pathway information, revealed a total of 17 genes linked to both the adaptive and innate immune systems. Among these genes, *ITGAX*, *CCL22, CSF1, CXCL8, IL12A, MMP9*, *CSF2, PTGS2*, and ENSG00000138685.17 (*FGF2*) demonstrated down-regulation, whereas *HLA-DPB1, FN1, CD36, MMP2, HLA-DRB1, FCGR1A, CCL17, and HLA-DRB5* exhibited up-regulation (as depicted in Figure 3) in the Andes. Importantly, these genes are notably enriched in antigen-presenting cells, such as dendritic cells and macrophages, as illustrated in Figure S1.

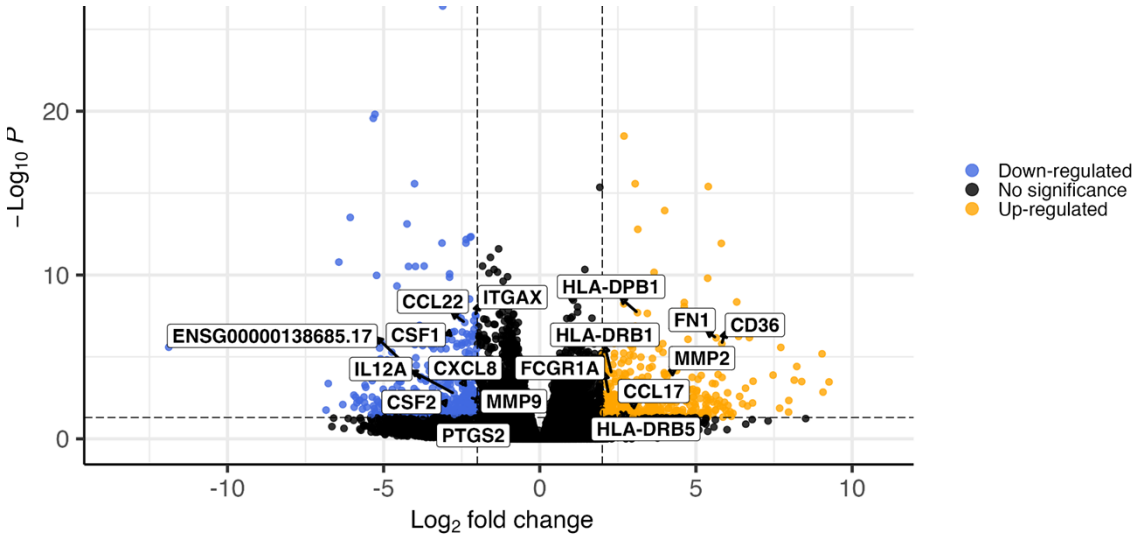


Figure 3. The volcano plot highlights the 17 significant immune system genes with differential gene expression regarding high and low-altitude samples. Nine down-regulated genes (*ITGAX, CCL22, CSF1, CXCL8, IL12A, MMP9, CSF2, PTGS2, ENSG00000138685.17*), and eight up-regulated genes (*HLA-DPB1, FN1, CD36, MMP2, HLA-DRB1, FCGR1A, CCL17, HLA-DRB5*).

Interplay between upregulated and downregulated genes

Functional enrichment by gene ontology analysis sheds light on the intricate molecular landscape underlying cellular responses, particularly in the context of immune and inflammatory processes (see Figure 4). Downregulated genes exhibit a notable impact on immune cell dynamics, as evidenced by their involvement in critical processes such as leukocyte chemotaxis and migration. These genes also contribute to the orchestration of angiogenesis, vasculature development, and responses to external stimuli like lipopolysaccharide, indicating their pivotal role in shaping the cellular microenvironment during immune challenges.

Conversely, the upregulated genes demonstrate a distinct functional profile, emphasizing their significance in reinforcing the host defense mechanisms. Processes like phagocytosis showcase the cell's ability to engulf and eliminate pathogens, while complement activation highlights a sophisticated arm of the immune system. Membrane invagination suggests an intricate regulatory mechanism, potentially influencing cellular internalization processes. Moreover, the involvement of upregulated genes in the humoral immune response, coupled with their role in defending against bacterial threats, underscores their crucial contribution to the overall immunological landscape.

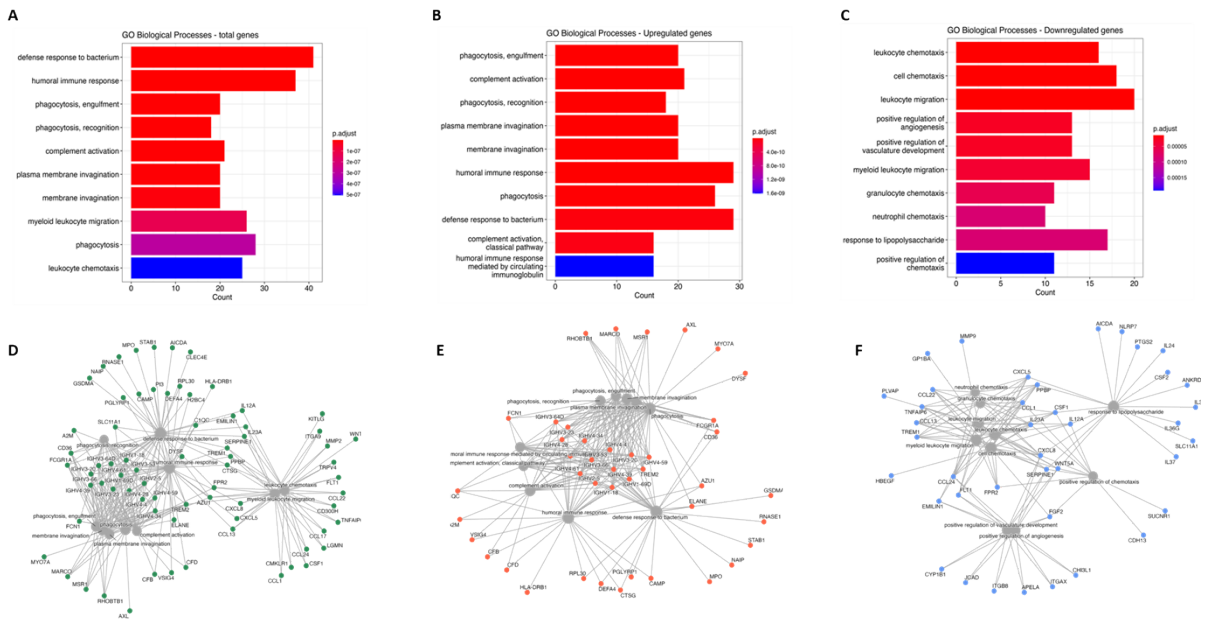


Figure 4. Gene Ontology enrichment analysis was conducted using ClusterProfile and the enrichGO function, focusing on the biological processes of the entire set of differentially expressed genes (A, D), upregulated genes (B, E), and downregulated genes (C, F). The x-axis denotes the number of genes associated with each biological process. The false discovery rate (p.adjust) is depicted by the red gradient, providing a visual representation of the statistical significance across different biological processes.

DISCUSSION

Our study is the first attempt to investigate the gene expression of indigenous populations from Peru (Figure 1) related to the immune response concerning their altitude of residence, and therefore, generally, the time of life exposed to low oxygen concentration, cold, dryness, ultraviolet rays, and differentiated immune response. In this sense, Quach et al. recently demonstrated that studying the effect of immunity leads us to consider what type of ancestry an individual or population belongs to, as Africans tend to have a stronger immune response than Europeans (19).

We have found that among the 843 genes showing differential expression between high and low-altitude inhabitants in the presence of pathogen-associated molecule stimulation, approximately half are up-regulated in high altitude. A striking discovery was the significant downregulation of the *APELA* gene in high altitude residents. Consistent with prior studies, the downregulation of *APELA* has been associated with vascular constriction and the development of micro- and macrovasculature disorders, including conditions such as preeclampsia (20) and pulmonary artery disease in individuals with chronic lung diseases (21). This observation aligns with epidemiological reports indicating an elevated risk of preeclampsia in pregnancies at higher altitudes (22), and arterial remodeling in dwellers with chronic exposure to these settings (23).

Furthermore, we have pinpointed genes exhibiting substantial differential expression, implicating both the innate and adaptive immune systems in proportionate measures. Notably, the up-regulation of *FCGR1A* and *HLA-DRB5* genes, alongside the down-regulation of the *HLA-DQB1* gene, highlights their significance in this context. Another

443 noteworthy gene possibly linked to immune functions is *HS3ST2*, whose expression
444 was upregulated in high-altitude inhabitants. This gene has previously been
445 associated with heightened expression in the polarization of M2 macrophages,
446 indicative of an anti-inflammatory response (24).
447
448 Our results are congruent with what has been observed in previous research
449 conducted on Andean populations, which have identified genes under positive
450 selection involved in the immune system, such as *DUOX2*, *CD40*, *TNIP1*, and *CHIA*
451 (7,8). Nevertheless, our understanding is currently limited due to the absence of
452 information regarding functional studies that explore the responses of these genes,
453 transcripts, proteins, or metabolites to altitude in these individuals.
454
455 It is worth noting that genetic positive selection does not necessarily determine what
456 may occur at the transcriptomic or proteomic level. Sharma *et al.* found that of the 162
457 genes and 251 proteins that have shown a signature of positive selection in high-
458 altitude research, only 7 (1.7%) genes underwent protein translation. These genes
459 and their proteins, which converge in positive selection with the effect of altitude on
460 inhabitants from different geographical areas, denote cardiovascular and lipid, and
461 glucose metabolism adaptation (25). Our findings do not coincide with those 7 filtered
462 genes since it was done for a global analysis of all residents exposed to high altitude.
463 Additionally, Sharma *et al.* did not consider transcriptomic studies because they are
464 scarce in the group of people exposed to high altitudes.
465
466 A recent study published by Li *et al.* presents a meta-analysis and bioinformatics
467 approach to available databases to identify which immune genes and how they varied
468 in expression when exposed to altitude conditions. The databases used by these
469 authors were from studies of cells exposed to normoxia or hypoxia, and in the
470 bioinformatics analysis, the cells came from European descent subjects exposed to
471 altitude but not native residents of a high-altitude area (26). Li *et al.* demonstrated that
472 the genes *ERH*, *VBP1*, *PSMA4*, *TOMM5*, *BNIP3L*, and *POLR2K* had a downregulation
473 when associated with hypoxia, except for *BNIP3L*. Additionally, these authors show
474 that altitude promotes the increase of some immune cells (T cells CD4 memory
475 resting, T cells gamma delta, Dendritic cells activated, Mast cells resting) at the
476 expense of others relevant to the innate or adaptive immune system (*e.g.* T cells CD8,

NK cells resting, Neutrophils).

In contrast, Manella *et al.* discovered that genes associated with immune response can exhibit differential expression patterns based on both the duration of exposure to high altitude and the specific altitude level (e.g., 3,800 meters in Puno compared to 5,100 meters in La Rinconada). Additionally, the timing of the measurements during the day also plays a role in these gene expression variations (27). The investigation revealed that time-associated gene alterations were linked to innate immunity, specifically pertaining to the activation of neutrophils. Additionally, distinct transcriptional patterns were identified in relation to the factors *RELA* and *IRF8*, recognized as mediators of inflammation and immune responses. Furthermore, the study conducted by Manella et al. unveiled an elevation in M2 macrophages and monocytes, influenced by both altitude and time. Conversely, a decrease contingent on altitude was observed in NK cells and regulatory T cells. Moreover, the researchers demonstrated the temporal variation in the expression of *NR1D1*, with its overall expression being notably reduced in response to altitude (28).

The Manella team's investigation is particularly intriguing due to its unique focus on individuals of European descent in one of the world's highest inhabited cities, located in Peru. However, it focused on European descent men subjected to a relatively short-term exposure, spanning several weeks, which cannot be directly compared to the chronic exposure experienced by native inhabitants at such altitudes. Moreover, our study population in Huánuco regularly experiences altitude fluctuations due to the necessity for individuals residing in the high Andean regions to descend to the city of Huánuco for various activities such as shopping, making payments, receiving government assistance, engaging in recreational pursuits, among others. As a result, a prospective study conducted within the Huánuco population, using a sample resembling ours, could investigate how genes are regulated, either upregulated or downregulated, in relation to the duration of absence from high-altitude environments followed by return or conversely, the effects of prolonged high-altitude exposure and subsequent return to lower altitudes. Additionally, it could investigate how the circadian rhythm of immune expression (29) might be altered in response to altitude changes in native inhabitants chronically exposed to high-altitude conditions.

Thus, we can assert that our study has successfully demonstrated a distinct genetic expression signature in high-altitude dwellers with a native Andean genetic background, as compared to their counterparts living at lower altitudes. To the best of our knowledge, this is the first study conducting a transcriptomic analysis in Peruvian inhabitants residing in the Andean region. Furthermore, we have identified several genes exhibiting both upregulation and downregulation, with a distribution observed within genes related to the innate and adaptive immune systems. However, we acknowledge that our sample size is relatively small, particularly among our group of individuals living at lower altitudes.

It's worth emphasizing that most of our study participants were female, and we did not implement age matching procedures during the sample selection process. This absence of age matching could potentially introduce selection bias, as gene expression can vary based on both the participant's sex and age, as has been published elsewhere (30). To mitigate potential sex-related bias, we incorporated sex as a covariate variable in our differential expression analysis. However, we acknowledge that our study was limited in its ability to consider all environmental conditions and travel habits, factors that may have exerted an influence on our research findings..

Epigenetic findings also appear to hold relevance in populations such as ours, as demonstrated by Childebayeva *et al.* (31), who revealed distinctive DNA methylation patterns among Andean Quechua individuals. The study indicated that individuals born and raised at high altitudes exhibited increased DNA methylation levels in the Long Interspersed Nuclear Element-1 (*LINE-1*) repetitive DNA element. This specific element has been shown to influence the expression of nearby genes through its methylation status. Interestingly, the study further unveiled a significant positive correlation between the duration of residence at high altitudes and the elevation of *LINE-1* DNA methylation levels. Notably, average *LINE-1* methylation was found to be associated with the genes *MTHFD1*, *TYMS*, *FOLH1*, and *GLDC*, suggesting potential regulatory effects on these genes. Furthermore, the study identified a negative association between the methylation of *EPAS1* and the duration of high-altitude exposure, implying a potential role of EPAS1 methylation in altitude-related adaptations.

Childebayeva *et al.*, recently assessed 850,000 CpG sites in 113 Andean Quechua individuals, demonstrating differentially methylated positions associated with the hypoxia-inducible factor pathway, red blood cell production, blood pressure, sugar metabolism, and other biological processes (32). These robust epigenetic changes could provide an explanation for epidemiological data supporting that the younger Andean individuals migrate to low-altitude coastal regions, the greater the risk of developing cardiometabolic risk factors (33–35), posing questions related to what could happen inversely.

This comprehensive analysis unveils a nuanced interplay between upregulated and downregulated genes, collectively shaping the intricate mosaic of immune and inflammatory responses. The differential regulation of these genes provides valuable insights into the molecular mechanisms governing cellular behaviors, offering a foundation for further exploration and targeted therapeutic interventions in the context of immune-related disorders.

Despite these limitations, our study stands as the pioneering effort demonstrating how immune response gene expression varies in individuals chronically exposed to high altitude (natives). Further investigations could validate and expand upon our findings through genomic, transcriptomic, proteomic, or even epigenetic analyses.

CONCLUSIONS

The differential expression of immune system genes in native inhabitants with a prominent Andean genetic background highlights the active involvement of both the innate and adaptive immune systems. Remarkably, our study revealed a distinct gene expression signature in our population, setting it apart from other transcriptomic studies conducted thus far. To fully comprehend the impact of altitude on immune system gene expression and genes related to high-altitude adaptation, a comprehensive multi-omic approach is indispensable. By collectively exploring these factors, we can gain profound insights into the intricate molecular mechanisms governing immune responses and the process of adaptation to high-altitude environments.

REFERENCES

1. Tremblay JC, Ainslie PN. Global and country-level estimates of human population at high altitude. Proc Natl Acad Sci. 2021 May 4;118(18):e2102463118.
2. Moore LG. Measuring high-altitude adaptation. J Appl Physiol. 2017 Nov 1;123(5):1371–85.
3. Bigham AW. Genetics of human origin and evolution: high-altitude adaptations. Curr Opin Genet Dev. 2016 Dec;41:8–13.
4. Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, et al. Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. Begun DJ, editor. PLoS Genet. 2010 Sep 9;6(9):e1001116.
5. Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, et al. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Hum Genomics. 2009 Dec;4(2):79.
6. Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, et al. Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genet. 2017 Nov;101(5):752–67.
7. Borda V, Alvim I, Mendes M, Silva-Carvalho C, Soares-Souza GB, Leal TP, et al. The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. Proc Natl Acad Sci. 2020 Dec 22;117(51):32557–65.
8. Caro-Consuegra R, Nieves-Colón MA, Rawls E, Rubin-de-Celis V, Lizárraga B, Vidaurre T, et al. Uncovering Signals of Positive Selection in Peruvian Populations from Three Ecological Regions. Mol Biol Evol. 2022 Aug 3;39(8):msac158.
9. Jacovas VC, Couto-Silva CM, Nunes K, Lemes RB, De Oliveira MZ, Salzano FM, et al. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep. 2018 Aug 24;8(1):12733.
10. Joseph SK, Migliore NR, Olivieri A, Torroni A, Owings AC, DeGiorgio M, et al. Genomic evidence for adaptation to tuberculosis in the Andes before European contact. iScience. 2023 Feb;26(2):106034.
11. Quach H, Rotival M, Pothlichet J, Loh YE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, Deschamps M, Naffakh N, Duffy D, Coen A, Leroux-Roels G, Clément F, Boland A, Deleuze JF, Kelso J, Albert ML, Quintana-Murci L. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell. 2016 Oct 20;167(3):643-656.e17. doi: 10.1016/j.cell.2016.09.024.
12. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-1664. doi:10.1101/gr.094052.109
13. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. Am J Hum Genet. 2012;91(1):122-138. doi:10.1016/j.ajhg.2012.05.024
14. Leal TP, Furlan VC, Gouveia MH, et al. NAToRA, a relatedness-pruning method to minimize the loss of dataset size in genetic and omics analyses. *Comput Struct Biotechnol J*. 2022;20:1821-1828. Published 2022 Apr 9. doi:10.1016/j.csbj.2022.04.009
15. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu x, Liu S, Bo X, Yu G (2021). “clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.” The Innovation, 2(3), 100141. doi:10.1016/j.xinn.2021.100141.

16. Christian SM HelkerJean EberleinKerstin WilhelmToshiya SuginoJulian

626 MalchowAnnika SchuermannStefan BaumeisterHyouk-Bum KwonHans-Martin
627 MaischeinMichael PotenteWiebke HerzogDidier YR Stainier (2020) Apelin signaling
628 drives vascular endothelial cells toward a pro-angiogenic state eLife 9:e55589.

629 17.Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, Mitchell A, Ma
630 J, Dubrulle J, Reyon D, Tsai SQ, Joung JK, Saghatelian A, Schier AF. Toddler: an
631 embryonic signal that promotes cell movement via Apelin receptors. Science. 2014
632 Feb 14;343(6172):1248636. doi: 10.1126/science.1248636. Epub 2014 Jan 9. PMID:
633 24407481; PMCID: PMC4107353.

634 18. Denys A, Allain F. The Emerging Roles of Heparan Sulfate 3-O-Sulfotransferases
635 in Cancer. Front Oncol. 2019 Jun 12;9:507. doi: 10.3389/fonc.2019.00507. PMID:
636 31249810; PMCID: PMC6582251.

637 19. Quach H, Rotival M, Pothlichet J, Loh YHE, Dannemann M, Zidane N, et al.
638 Genetic Adaptation and Neandertal Admixture Shaped the Immune System of
639 Human Populations. Cell. 2016 Oct;167(3):643-656.e17.

640 20. Shimada N, Nakayama T, Umemura H, Kawana K, Yamamoto T, Uchigasaki S. A
641 Case-Control Study of the *APELA* Gene and Hypertensive Disorders of Pregnancy.
642 Medicina (Kaunas). 2022 Apr 26;58(5):591. doi: 10.3390/medicina58050591. PMID:
643 35630008; PMCID: PMC9146499.

644 21. Yang X, Bian Y, Wan J, Li L, Yang P, Zhao S, Zhao H. Variants in the 5'-UTR of
645 APELA gene in women with preeclampsia. Prenat Diagn. 2019 Mar;39(4):308-313.
646 doi: 10.1002/pd.5431. Epub 2019 Feb 21. PMID: 30719741.

647 22. Dimitriadis E, Rolnik DL, Zhou W, Estrada-Gutierrez G, Koga K, Francisco RPV,
648 Whitehead C, Hyett J, da Silva Costa F, Nicolaides K, Menkhorst E. Pre-eclampsia.
649 Nat Rev Dis Primers. 2023 Feb 16;9(1):8. doi: 10.1038/s41572-023-00417-6. Erratum
650 in: Nat Rev Dis Primers. 2023 Jul 3;9(1):35. PMID: 36797292

651 23. Sydykov A, Mamazhakypov A, Maripov A, Kosanovic D, Weissmann N, Ghofrani
652 HA, Sarybaev AS, Schermuly RT. Pulmonary Hypertension in Acute and Chronic High
653 Altitude Maladaptation Disorders. Int J Environ Res Public Health. 2021 Feb
654 10;18(4):1692. doi: 10.3390/ijerph18041692. PMID: 33578749; PMCID:
655 PMC7916528.

656 24. Martinez P, Denys A, Delos M, Sikora AS, Carpentier M, Julien S, Pestel J, Allain
657 F. Macrophage polarization alters the expression and sulfation pattern of
658 glycosaminoglycans. Glycobiology. 2015 May;25(5):502-13. doi:
659 10.1093/glycob/cwu137. Epub 2014 Dec 11. PMID: 25504800.

660 25. Sharma V, Varshney R, Sethy NK. Human adaptation to high altitude: a review
661 of convergence between genomic and proteomic signatures. Hum Genomics. 2022
662 Jul 15;16(1):21.
663 26. Li Q, Xu Z, Fang F, Shen Y, Lei H, Shen X. Identification of key pathways,
664 genes and immune cell infiltration in hypoxia of high-altitude acclimatization via meta-
665 analysis and integrated bioinformatics analysis. Front Genet. 2023 Mar
666 24;14:1055372.
667 27. Manella G, Ezagouri S, Champigneulle B, Gaucher J, Mendelson M, Lemarie

668 E, et al. The human blood transcriptome exhibits time-of-day-dependent response to
669 hypoxia: Lessons from the highest city in the world. Cell Rep. 2022 Aug;40(7):111213.
670 28. Scheiermann C, Kunisaki Y, Frenette PS. Circadian control of the immune
671 system. Nat Rev Immunol. 2013 Mar;13(3):190–8.
672 29. Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, et al. Distinctive
673 roles of age, sex, and genetics in shaping transcriptional variation of human immune
674 responses to microbial challenges. Proc Natl Acad Sci [Internet]. 2018 Jan 16 [cited
675 2023 May 15];115(3). Available from:
676 https://pnas.org/doi/full/10.1073/pnas.1714765115
677 30. Childebayeva A, Jones TR, Goodrich JM, Leon-Velarde F, Rivera-Chira M,
678 Kiyamu M, et al. LINE-1 and *EPAS1* DNA methylation associations with high-altitude
679 exposure. Epigenetics. 2019 Jan 2;14(1):1–15.
680 31. Childebayeva A, Goodrich JM, Leon-Velarde F, Rivera-Chira M, Kiyamu M,
681 Brutsaert TD, et al. Genome-Wide Epigenetic Signatures of Adaptive Developmental
682 Plasticity in the Andes. Betancourt A, editor. Genome Biol Evol. 2021 Feb
683 3;13(2):evaa239.
684 32. Miranda JJ, Gilman RH, Smeeth L. Differences in cardiovascular risk factors in
685 rural, urban and rural-to-urban migrants in Peru. Heart. 2011 May 15;97(10):787–96.
686 33. Bernabe-Ortiz A, Sanchez JF, Carrillo-Larco RM, Gilman RH, Poterico JA,
687 Quispe R, et al. Rural-to-urban migration and risk of hypertension: longitudinal results
688 of the PERU MIGRANT study. J Hum Hypertens. 2017 Jan;31(1):22–8.
689 34. Carrillo-Larco RM, Guzman-Vilca WC, Leon-Velarde F, Bernabe-Ortiz A,
690 Jimenez MM, Penny ME, et al. Peru – Progress in health and sciences in 200 years
691 of independence. Lancet Reg Health - Am. 2022 Mar;7:100148.
692 35. Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. "edgeR: a
693 Bioconductor package for differential expression analysis of digital gene expression
694 data." bioinformatics 26.1 (2010): 139-140.

Supplementary

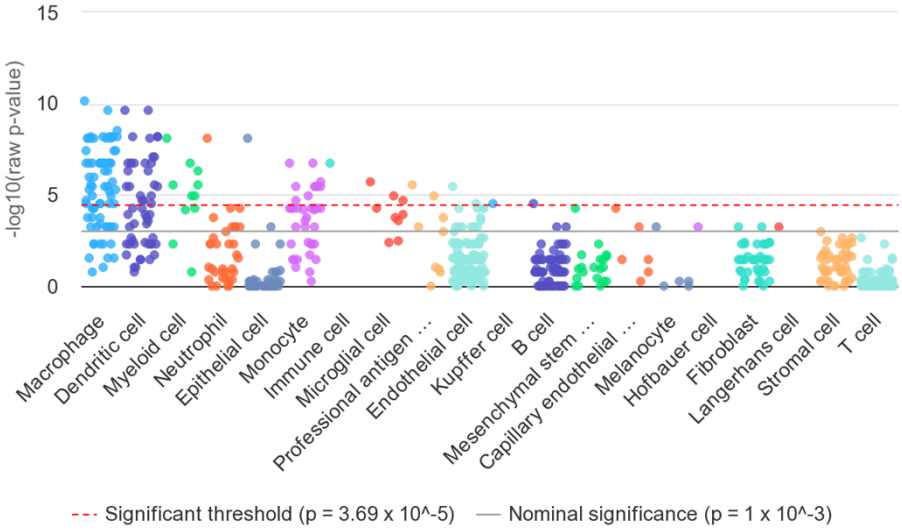
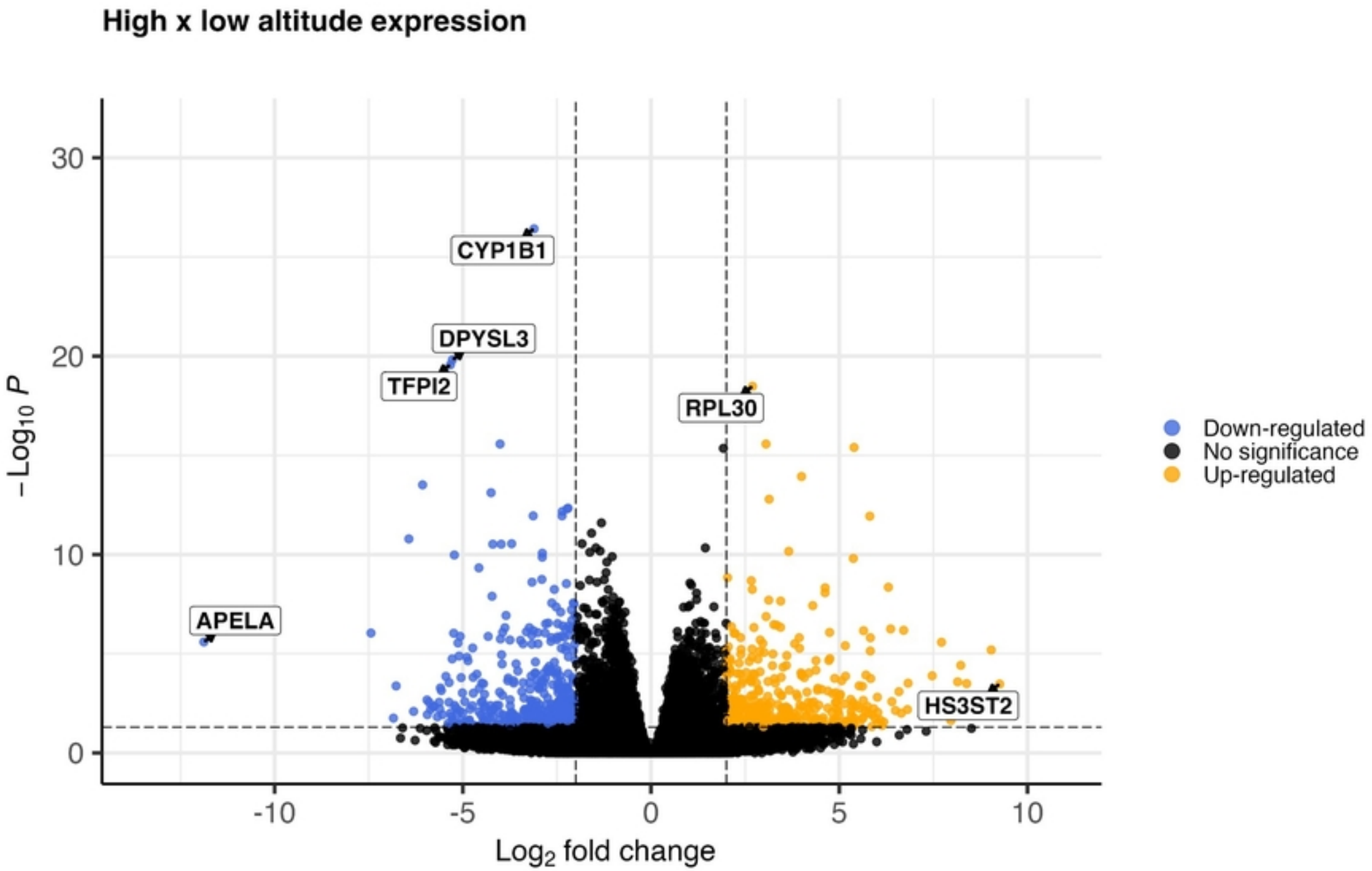
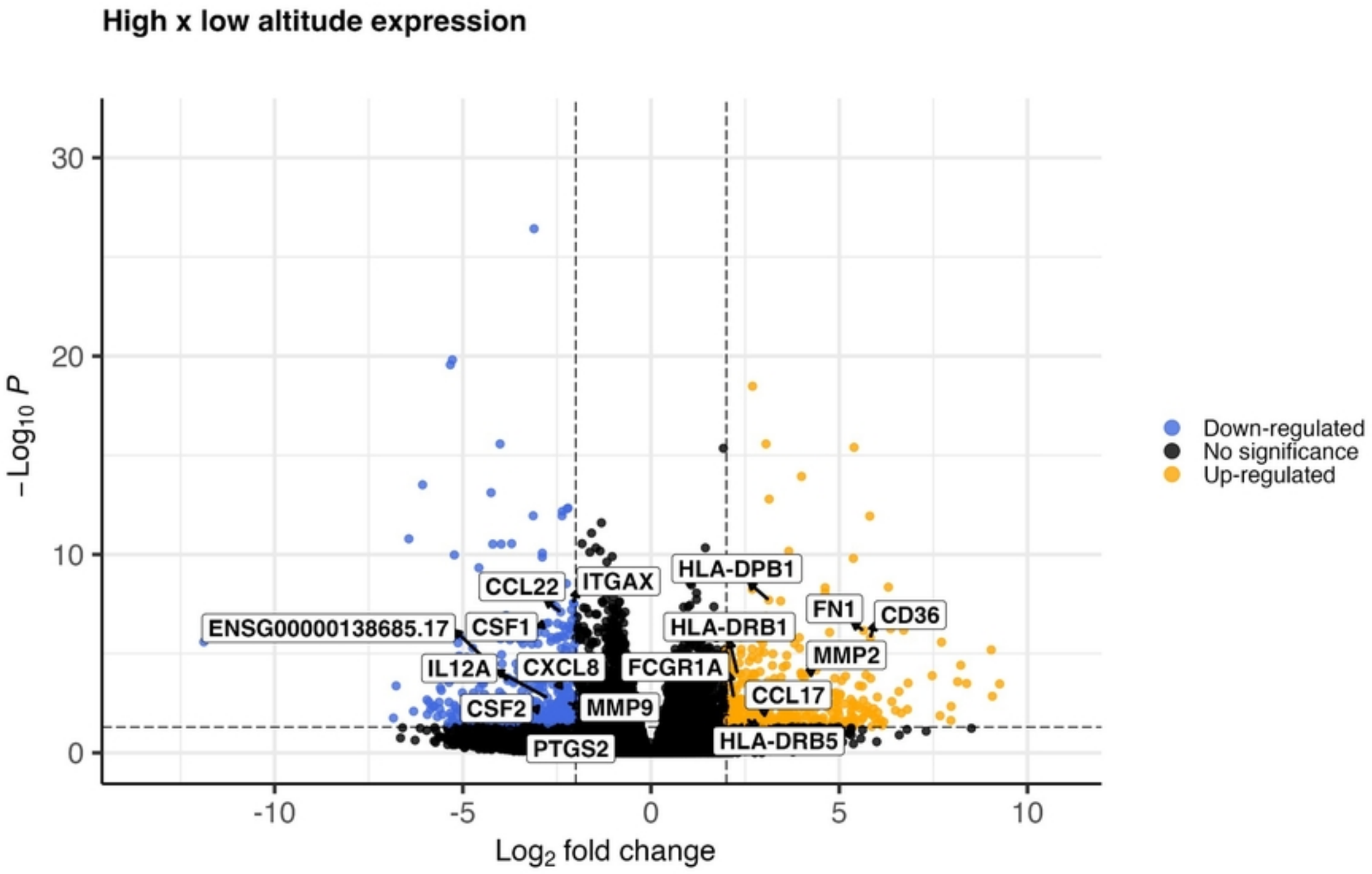


Figure S1. Cell-type-specific Enrichment Analysis (CSEA) for differentially expressed genes was obtained using WebCSEA (Web-based Cell-type Specific Enrichment Analysis of Genes).

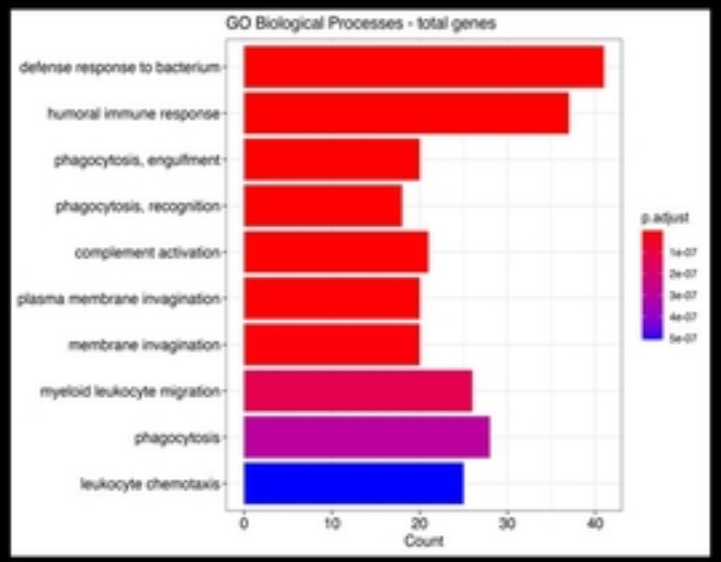
Genetic and geographic landscapes for Peruvian populations

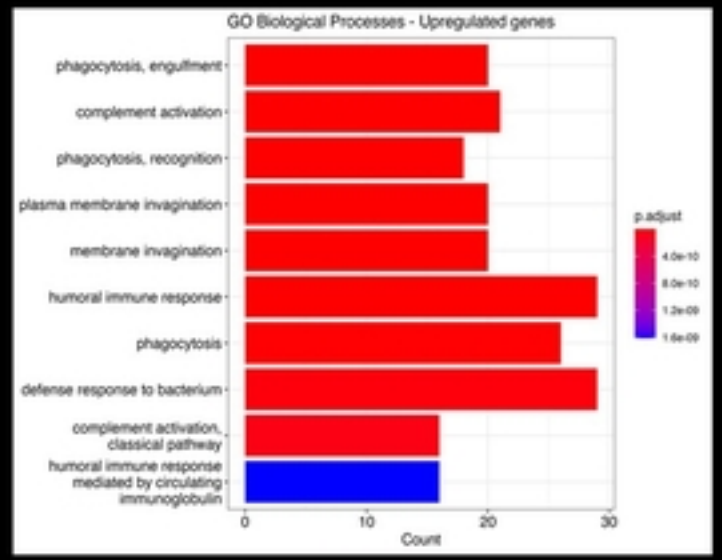


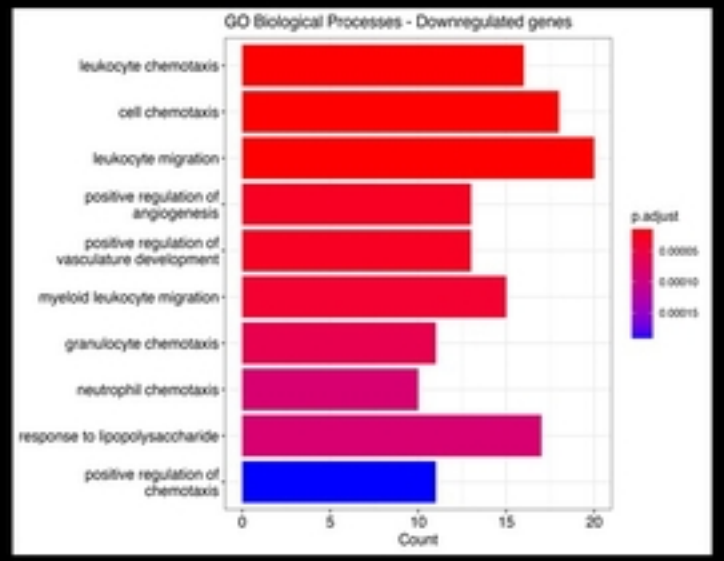
The volcano plot highlights the highest significant genes with dif

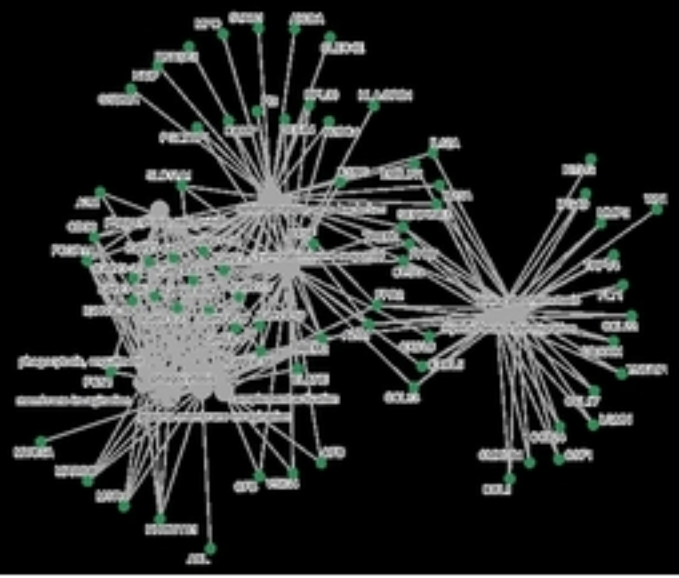


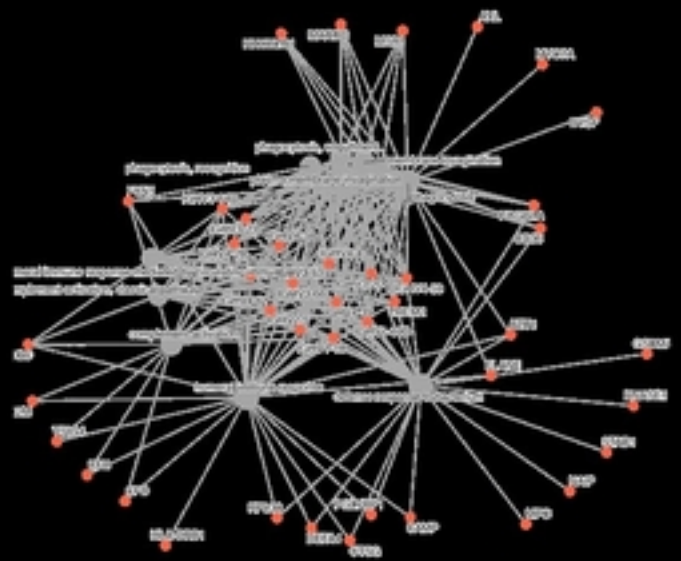
The volcano plot highlights the 17 significant immune system ge

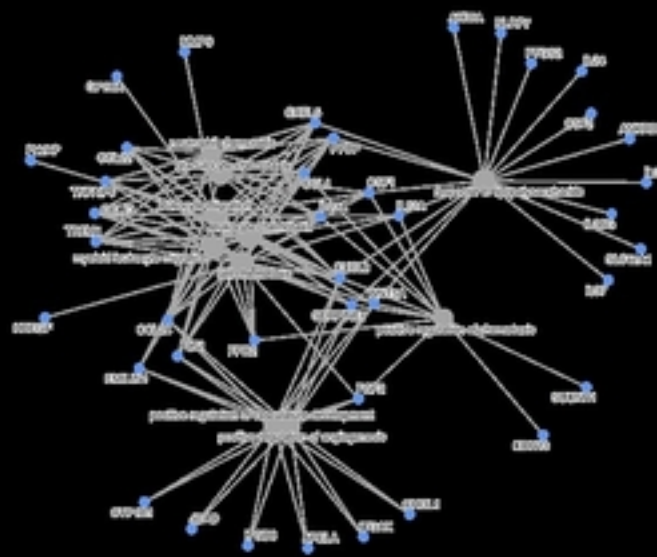












Gene Ontology enrichment analysis was conducted using Cluste