Integrative GWAS and landscape genomics prioritize candidate loci associated with high-altitude adaptation in Tibetan sheep

Jianxiang Tang[1,2], Chao Yuan[1,2], Bowen Chen[1,2], Fan Wang[3], Tingting Guo[1,2], Zengkui Lu[1,2,*], Jianbin Liu[1,2,*]

[1] Key Laboratory of Animal Genetics and Breeding on Tibetan Plateau, Ministry of Agriculture and Rural Affairs, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou 730050, China

[2] Sheep Breeding Engineering Technology Research Center of Chinese Academy of Agricultural Sciences, Lanzhou 730050, China

[3] School of Life Sciences and Engineering, Northwest Minzu University, Lanzhou 730100, China

[*]Corresponding authors: luzengkui@caas.cn; liujianbin@caas.cn

ABSTRACT

High-altitude environments expose Tibetan sheep (*Ovis aries*) to hypoxia, low temperatures, strong ultraviolet radiation, and seasonal nutritional stress. Blood indicators related to immunoglobulins, oxidative stress, and inflammatory cytokines can reflect physiological responses to these environmental pressures, but the genetic basis of variation in these indicators remains incompletely understood. We integrated whole-genome resequencing data from 220 adult rams across 11 Tibetan sheep populations with 24 blood physiological and biochemical indicators and environmental variables from the sampling sites. A genome-wide association study (GWAS) prioritized 324 candidate single-nucleotide polymorphisms (SNPs) annotated to 310 genes. Genotype-environment association (GEA) analysis identified 121 environment-associated SNPs at a false discovery rate (FDR)<0.05. Of 67 non-overlapping 50-kb windows containing environment-associated SNPs, 27 (40.30%) overlapped with blood trait-associated windows. Regions corresponding to representative

Uncorrected proof

candidate genes, including *MITF*, *ANTXR1*, *FAS*, and *TTC23L*, were associated with blood indicators such as glutathione peroxidase (GSH-PX) and immunoglobulin D (IgD). Allele-frequency gradients, Bayesian regional prioritization, motif scanning, and public tissue-expression data provided complementary clues for the functional interpretation of the candidate regions. These findings identify candidate loci for independent validation and functional studies of high-altitude adaptation in Tibetan sheep.

Key words: Blood physiological and biochemical indicators; Genotype-environment association; Oxidative stress; Immune regulation; Bayesian fine-mapping

INTRODUCTION

Tibetan sheep are an important indigenous livestock resource on the Qinghai-Xizang Plateau and adjacent high-altitude regions. Their long-term residence in this region makes them a useful livestock model for studying physiological and genetic adaptation to plateau environments. Understanding the genetic basis of these adaptations is also relevant to the conservation and improvement of Tibetan sheep germplasm.

Previous studies suggest that Tibetan sheep entered the Qinghai-Xizang Plateau approximately 3 000 years ago and adapted to high-altitude environments over a relatively short evolutionary period (Liang et al., 2024). High-altitude exposure affects oxygen uptake, blood circulation, energy metabolism, redox homeostasis, and immune function. Adaptation to these conditions usually involves coordinated changes in hypoxia response, oxygen transport, cardiovascular regulation, energy metabolism, and tissue homeostasis (Bigham & Lee, 2014; Storz & Cheviron, 2021). Recent studies of sheep across altitude gradients also implicate oxygen uptake, angiogenesis, erythropoiesis, mitochondrial function, and ion homeostasis (Zhang et al., 2024b). Because hypoxia can regulate immune function, inflammatory signaling, and oxidative stress, blood indicators such as reactive oxygen species (ROS), malondialdehyde (MDA), superoxide dismutase (SOD), and glutathione

Uncorrected proof

peroxidase (GSH-PX) can provide intermediate phenotypes linking environmental stress, genomic variation, and adaptive physiology (Eltzschig & Carmeliet, 2011; Taylor & Colgan, 2017; McGettrick & O'Neill, 2020; Siques et al., 2018; Pena et al., 2022).

Population genomic, structural-variation, and multi-omics studies have substantially advanced the study of high-altitude adaptation in domestic animals. In yaks, Tibetan chickens, Tibetan pigs, and Tibetan sheep, previous studies have identified candidate pathways related to hypoxia response, energy metabolism, oxygen transport, cardiovascular function, and regulatory variation (Qiu et al., 2012; Wang et al., 2015; Ma et al., 2019; Hu et al., 2019; Pham et al., 2021; Zhao et al., 2026). In Tibetan sheep, selection-scan and multi-omics studies have highlighted genes and pathways such as hypoxia-inducible factor 1 (HIF-1) signaling, HBB, *MITF*, beta-globin haplotypes, EGLN1 expression, and structural variants with potential cis-regulatory effects (Hu et al., 2019; Li et al., 2024; Liang et al., 2024; Yan et al., 2024). However, most work has focused on candidate regions or expression changes. Whether these variants are associated with blood physiological and biochemical indicators remains less well characterized. Integrating gene annotation, regulatory prediction, and expression resources may therefore improve interpretation of candidate loci for complex adaptive traits (Zhang et al., 2024a).

A genome-wide association study (GWAS) can identify variants associated with blood physiological indicators. Genotype-environment association (GEA) analysis can identify genomic regions associated with ecological factors such as altitude, temperature, and precipitation. In landscape genomics, environmental association analysis provides a framework for detecting statistical associations between allele frequencies and ecological gradients (Rellstab et al., 2015). Studies in livestock also show that integrating environmental variables with genomic data can help characterize local adaptation in indigenous breeds (Gheyas et al., 2021; Niu et al., 2024).

Uncorrected proof

However, GEA signals alone do not establish that associated variants contribute causally to adaptation. Integrating GWAS with GEA can prioritize candidate regions supported by associations with both blood physiological traits and environmental gradients (Dauphin et al., 2023; Lasky et al., 2023).

Here, we integrated whole-genome resequencing, GWAS of blood physiological and biochemical indicators, and GEA to prioritize candidate regions related to high-altitude adaptation in Tibetan sheep. We then combined functional annotation, genotype-phenotype correlations, spatial allele-frequency patterns, Bayesian fine-mapping, motif prediction, and public transcriptome evidence to refine and interpret the prioritized regions.

MATERIALS AND METHODS

Sample collection and DNA extraction

We collected blood samples from 220 adult rams representing 11 Tibetan sheep populations in Gansu, Qinghai, and the Xizang Autonomous Region (Supplementary Table S1). Twenty rams approximately 3.5 years of age that appeared to be in good physical condition and showed no obvious signs of disease or abnormality upon visual inspection were randomly selected from each population. All samples were collected in late July, and the animals were managed under grazing with supplementary feeding. Because this was a population-level field study rather than a controlled feeding trial, local forage composition, supplementary feed, and management practices were not quantitatively standardized across populations. To reduce major biological heterogeneity, sampling was restricted to adult rams of a similar age class and health status; population structure, geographic distance, and environmental gradients were considered in downstream analyses. For each animal, 5 mL of jugular venous blood was collected into an EDTA anticoagulant tube and stored at −20°C. A second 5 mL blood sample was allowed to stand at room temperature for approximately 40 min and was then centrifuged at 3,000 r/min for 15 min. Separated serum was snap-frozen in liquid

Uncorrected proof

nitrogen, stored at −80°C, and analyzed within one month. Each serum sample was thawed only once before analysis to avoid repeated freeze–thaw cycles. Genomic DNA was extracted using the TIANamp Genomic DNA Kit (Tiangen, China). DNA purity was assessed using a NanoDrop 2000 spectrophotometer, integrity using 1.0% agarose gel electrophoresis, and concentration using Qubit fluorometry. Under the sequencing provider's category-A criteria, DNA concentration was required to be at least 10 ng/μL and total DNA amount at least 0.5 μg, with no or only slight degradation and at most slight RNA contamination. Qualified DNA samples were used for whole-genome resequencing library construction. The animal study protocol was approved by the Institutional Ethics Committee of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences (No. 2026-11).

Measurement of 24 blood physiological and biochemical indicators

The 24 blood physiological and biochemical indicators were selected based on their reported relevance to humoral immunity, redox homeostasis, or inflammatory regulation under high-altitude hypoxia; their suitability for standardized serum assays; and their ability to capture complementary aspects of physiological status. The panel comprised five immunoglobulins, seven oxidative stress indicators, and 12 inflammatory cytokines. These indicators were treated as continuous phenotypes for association analysis.

Commercial kits from Nanjing Jiancheng Bioengineering Institute were used to measure immunoglobulin, oxidative stress, and inflammatory cytokine indicators. The immunoglobulin (Ig) indicators included IgA, IgG, IgM, IgE, and IgD. The oxidative stress indicators included GSH-PX, glutathione (GSH), SOD, total antioxidant capacity (T-AOC), nitric oxide (NO), MDA, and ROS. The inflammatory cytokine indicators included tumor necrosis factors (TNF-α and TNF-β), interferons (IFN-γ, IFN-α, and IFN-β), interleukins (IL-1β, IL-1α, IL-2, IL-4, IL-6, and IL-10), and interleukin-1 receptor antagonist (IL-1Ra). All

Uncorrected proof

assays followed the manufacturer's instructions for reagent preparation, standard-curve construction, sample incubation, signal detection, and calculation of final concentrations or activities. Serum samples from all 11 populations were assayed in a single laboratory batch using the same workflow to minimize technical variation. All 24 indicators were subjected to rank-based inverse normal transformation (RINT) before GWAS. The transformed values were used in GWAS and in the correlations between candidate-allele dosage and blood indicators, whereas the original measured values were used for the descriptive statistics in Supplementary Table S2 and the genotype-group comparisons in Figure 3B.

Whole-genome resequencing and variant detection

Sequencing libraries were constructed according to the TruSeq Nano DNA HT Kit protocol (Illumina, USA). Genomic DNA was enzymatically fragmented, end-repaired, A-tailed, and ligated to adapters. After library quality control, 150-bp paired-end sequencing was performed on the Illumina HiSeq X10 platform. The average sequencing depth was approximately 6.2×. After base calling, raw sequencing data were converted into FASTQ files and quality filtered using Quality Trim v.1.6.0. Reads containing adapter sequences, reads with an N-base proportion >10%, and reads in which >50% of bases had quality scores $Q \leq 20$ were removed. High-quality reads were aligned to the in-house Tibetan sheep reference genome HB, assembled by our team, using Burrows–Wheeler Aligner (BWA) v.0.7.15 with the maximal exact match (MEM) algorithm (parameters: −k 32 −M) (Li & Durbin, 2009). The resulting alignment files were converted to binary alignment/map (BAM) format using SAMtools v.1.17, and polymerase chain reaction (PCR) duplicate reads were marked using Picard v.2.18.7 (https://broadinstitute.github.io/picard/). Local realignment around insertions and deletions (indels) and base quality score recalibration were performed using the Genome Analysis Toolkit (GATK) v.3.4-46. Multi-sample variant calling was subsequently conducted using the UnifiedGenotyper module in GATK v.3.4-46 (DePristo et al., 2011). Filtering flags

Uncorrected proof

were applied using VariantFiltration with the parameters −Window 4, −filter "*QD*<4.0 || *FS*>60.0 || *MQ*<40.0", and −G_filter "*GQ*<20". Here, QD denotes quality by depth, FS the Phred-scaled Fisher strand-bias statistic, MQ root-mean-square mapping quality, and GQ genotype quality. BEDTools v.2.25.0 was used to calculate sequencing depth and genome coverage (Quinlan & Hall, 2010). Functional annotation of detected variants was performed using ANNOVAR (Wang et al., 2010). PLINK v.1.90 was then used to remove low-quality variants, including sites with an average sequencing depth <3× or >30× and non-biallelic sites (Chang et al., 2015).

Genotype imputation and dataset construction

Genotypes from low-depth sequencing were phased and missing genotypes were imputed using BEAGLE v.5.4 (beagle.22Jul22.46e.jar) (Browning et al., 2018; Browning et al., 2021), without an external reference panel. The imputed data were converted to binary PLINK format and further filtered for site missingness and minor allele frequency (MAF). Missingness before and after BEAGLE imputation and quality control is summarized in Supplementary Table S3.

To evaluate imputation performance, genotypes with read depth (DP)<3 were set to missing, and biallelic single-nucleotide polymorphisms (SNPs) on Chr1–Chr26 were retained if site missingness≤0.50, MAF≥0.05, and the Hardy-Weinberg equilibrium test $P \geq 1\times10^{-6}$. We then randomly masked 5% of observed genotypes with DP≥8, GQ≥20, and allele depths supporting the original genotype call, and reran the same BEAGLE workflow. Imputed genotypes were compared with their corresponding pre-masking genotypes. Evaluation metrics included genotype concordance and the squared Pearson correlation coefficient (R^2) between alternate (ALT) allele counts (0, 1, or 2) derived from the pre-masking observed genotypes and the post-imputation best-guess genotypes. These metrics were calculated

Uncorrected proof

within four MAF bins and across all masked genotype calls. This within-dataset masking analysis assessed internal imputation concordance (Supplementary Table S4).

Because some raw genotypes were missing in the low-depth data, the final GWAS, GEA, and FINEMAP analyses were based mainly on the imputed genotype dataset. To assess reliability at core candidate SNPs, these sites were traced back to the PLINK dataset converted from the raw variant call format (VCF) file. For these sites, we calculated the genotype missing rate, allele frequency, and number of individuals with non-missing raw genotypes. We also compared non-missing raw genotypes with the corresponding imputed genotypes (Supplementary Table S5).

Genome-wide association study

All 24 blood physiological and biochemical indicators were subjected to rank-based inverse normal transformation before GWAS to reduce the influence of distributional skewness on model fitting and association testing. GWAS was performed in R using rMVP v.1.4.5 (Yin et al., 2021). Association analysis used the Fixed and Random Model Circulating Probability Unification (FarmCPU) model (Liu et al., 2016). The primary GWAS jointly analyzed all 220 animals from the 11 sampling populations. Separate within-population GWAS and meta-analysis were not performed because each population contained only 20 animals, which was insufficient for stable estimation of population-specific SNP effects and standard errors. To control confounding from population structure and relatedness, principal components (PCs) and a kinship matrix were constructed from genome-wide SNP data. Population structure correction used PC1–PC5 derived from 5 367 autosomal SNPs pruned for linkage disequilibrium (LD). Principal components were included as fixed effects, and the kinship matrix was included as a random-effect covariate. Sex and age were not included as additional covariates because all samples were adult rams of a similar age class. As a sensitivity analysis for potential site-level confounding, each of the 24 traits was centered

Uncorrected proof

within sampling populations after outlier handling and then subjected to rank-based inverse normal transformation. GWAS was repeated using the same genotype data, covariates, and FarmCPU parameters as in the primary analysis (Supplementary Table S6). In addition, each of the 11 populations was excluded in turn, leaving 200 animals in each leave-one-population-out analysis. All 24 traits were reanalyzed using the same phenotype-processing procedures, covariates, and FarmCPU parameters as in the primary analysis, and the resulting sensitivity statistics are summarized in Supplementary Table S7. The within-population phenotype-centered analysis assessed sensitivity to population-level confounding, whereas the leave-one-population-out analysis assessed the influence of each population on the association results. Both were sensitivity analyses within the study cohort, not independent replication in separate populations. Because the sample size was limited and the 24 blood indicators may be controlled by multiple small-effect loci, we used a two-step candidate-screening strategy. First, $P<1\times10^{-6}$ ($-\log_{10}(P)>6$) was used as a display threshold for strong candidate association signals (Supplementary Table S8). Second, the top 0.05% of SNPs ranked by P value for each GWAS trait were retained as the GWAS-prioritized candidate set for window-level integration with GEA results. This threshold was used for candidate-region screening and cross-analysis and should not be interpreted as a strict genome-wide significance threshold for causal loci. FarmCPU GWAS diagnostics are provided in Supplementary Figure S1. Statistical power to detect SNP associations was evaluated across different genetic effect sizes and minor allele frequencies, assuming a sample size of 220 and an association threshold of $P<1\times10^{-6}$ (Supplementary Figure S2).

Environmental data collection and dimensionality reduction

Based on the longitude and latitude of the 11 Tibetan sheep sampling sites, absolute altitude and 19 bioclimatic variables were extracted from the WorldClim 2 global high-resolution climate database (https://www.worldclim.org/) (Fick & Hijmans, 2017). The climate data

Uncorrected proof

225 covered 1970–2000 and had a spatial resolution of 30 arc-seconds. To reduce
226 multicollinearity among the 20 environmental variables, Pearson correlation coefficients were
227 calculated, and highly correlated variables were removed using $|r|>0.85$ as the threshold.
228 Eight relatively independent variables were retained: altitude, mean annual temperature,
229 temperature seasonality, maximum temperature of the warmest month, mean temperature of
230 the driest quarter, precipitation of the wettest month, precipitation of the driest month, and
231 precipitation seasonality. Principal component analysis (PCA) was performed using R v.4.3.1
232 with FactoMineR v.2.13 (Lê et al., 2008) and factoextra v.2.0.0. Environmental biplots were
233 generated to evaluate environmental differentiation and the main contributing factors among
234 populations.

235 **Analysis of environmental heterogeneity and geographical isolation effects using Mantel**
236 **tests**

237 This analysis was performed to evaluate the relationships of isolation by distance (IBD) and
238 isolation by environment (IBE) with genetic differentiation among Tibetan sheep
239 populations. PLINK v.1.90 and the R package geosphere v.1.5-20 were used to calculate
240 genetic, geographic, and environmental distance matrices among populations. When
241 constructing the environmental distance matrix, squared Pearson correlation coefficients (R^2)
242 were recalculated among the 20 environmental variables. Highly collinear variables were
243 removed using $R^2>0.70$ as the threshold. This filter retained ten environmental variables,
244 which were standardized using Z-score transformation before Euclidean distances were
245 calculated. The R package vegan v.2.7-1 was used to assess correlations between genetic and
246 geographic distances and between genetic and environmental distances. Mantel tests were
247 performed with 9 999 permutations. Multiple matrix regression with randomization (MMRR)
248 was further performed at the population level, with population genetic distance as the
249 response matrix and geographic and environmental PC1 distances as explanatory matrices.

Uncorrected proof

The 11 populations were treated as independent units, and significance was assessed using 9 999 permutations. The ten retained environmental variables were also tested separately, followed by Benjamini–Hochberg correction. Complete model results are provided in Supplementary Table S9.

Genotype-environment association analysis

We used latent factor mixed models (LFMM) to test genotype-environment associations while accounting for population structure (Frichot et al., 2013; Caye et al., 2019). Based on the PCA of environmental variables, PC1 of the integrated environmental gradient was used as the environmental explanatory variable. Each population-level PC1 score was assigned to all individuals from that population, and the resulting environmental vector was matched to the genotype matrix by sample ID and genotype-file order before analysis. The genomic inflation factor was used to correct LFMM association *P* values. The Benjamini–Hochberg method was then used to control the false discovery rate (FDR), with FDR<0.05 as the significance threshold (Benjamini & Hochberg, 1995). Significant environment-associated SNPs were retained for GWAS–GEA integration (Supplementary Table S10). As a sensitivity analysis, population-level allele counts were analyzed using the standard covariate model in BayPass v.2.41, with the 11 populations treated as independent environmental units. Environmental PC1 was tested using 47 075 background SNPs systematically selected from an LD-pruned set together with 121 LFMM candidate SNPs, while accounting for covariance among population allele frequencies. Three independent runs were performed using different random seeds, and the median Bayes factor on the decibel scale [BF(dB)] across runs was used to rank the tested SNPs (Gautier, 2015; Supplementary Table S11).

Integration of GWAS and GEA results and screening and annotation of candidate SNPs

Overlap between GWAS and GEA signals was evaluated using an UpSet matrix based on non-overlapping 50-kb genomic windows (Lex et al., 2014). For each GWAS trait, SNPs in

Uncorrected proof

the top 0.05% by *P* value were assigned to 50-kb windows according to chromosomal position. These SNPs were grouped into three phenotype modules: immunoglobulins, oxidative stress, and inflammatory cytokines. Significant GEA SNPs (FDR<0.05) were assigned to the same window framework. The resulting matrix was used to evaluate overlap between environment-associated and phenotype-associated regions (Supplementary Table S12). Genes within 50 kb upstream or downstream of phenotype-environment overlapping regions were annotated using the Tibetan sheep reference genome. Enrichment analysis was performed using the 38 genes annotated from the 27 GWAS–GEA overlapping windows, with a 50-kb flank on each side, as the foreground gene set. Using the same annotation rule, the background set was constructed from 50 957 analyzable non-overlapping 50-kb windows containing at least one tested SNP and comprised 19 711 non-redundant genes. Gene Ontology (GO) analysis used annotations for the longest protein-coding transcript of each gene. Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were generated using KOBAS v.3.0.3_3 (Bu et al., 2021) and the *Ovis aries* database. Enrichment was evaluated using a one-sided hypergeometric test. *P* values were adjusted using the Benjamini–Hochberg method separately across all 53 GO terms and all 80 KEGG pathways. Complete results are provided in Supplementary Tables S13 and S14. In parallel with the window-based GWAS–GEA overlap analysis, we prioritized the 121 significant LFMM candidate SNPs. For each candidate SNP, we counted the number of the 24 FarmCPU GWAS traits for which it reached nominal significance (P<0.05) and recorded its minimum GWAS *P* value. Candidate SNPs were first ranked in descending order of the number of supporting traits and then in ascending order of the minimum *P* value. To reduce redundant representation of neighboring signals, physical-distance-based regional deduplication was performed, retaining only the highest-ranked SNP among candidates on the same chromosome separated by less than 50 kb. We then retained candidate SNPs supported by at least three traits, or by at least two traits with a

Uncorrected proof

minimum $P<0.005$, yielding 17 region-representative candidate SNPs. Gene-structure annotation showed that eight of these SNPs were intragenic and corresponded to *GDPD4*, *TTC23L*, *MITF*, *OSBPL10*, *GSTA1-1*, *FAS*, *CFAP58*, and *ANTXR1*. These loci were therefore prioritized for downstream analyses (Supplementary Table S15). These lead SNPs were used for exploratory follow-up genotype-phenotype analyses and population-level allele-frequency analyses. For each biallelic SNP, genotypes were additively coded as 0, 1, or 2 according to the dosage of the effect allele specified in Supplementary Table S15. At the individual level, Spearman rank correlation was used to assess associations between effect-allele dosage and the RINT-transformed values of the 24 blood indicators. At the population level, Spearman rank correlation was used to assess associations between effect-allele frequencies and non-redundant environmental variables across the 11 populations. For the four locus-trait combinations shown in Figure 3B, overall differences in phenotype distributions among the three genotype groups were assessed using the Kruskal–Wallis test, followed by two-sided pairwise Wilcoxon rank-sum tests. Benjamini–Hochberg adjustment was applied separately to the *P* values from the four Kruskal–Wallis tests and the 12 pairwise Wilcoxon tests. For environmental correlation analysis, Spearman correlation was used to test associations between population-level candidate allele frequencies and non-redundant environmental variables after removing highly correlated variables ($|r|>0.70$).

Geographic analysis and altitude gradient regression of GWAS–GEA lead variants

Candidate-allele frequencies of the lead variants were calculated across the 11 Tibetan sheep populations to characterize their spatial distributions. High-resolution digital elevation model data were extracted based on the sampling-site coordinates, and geographic distribution maps were generated using the R packages elevatr v.0.99.1, sf v.1.0-21 (Pebesma, 2018), ggplot2 v.4.0.0, and scatterpie v.0.2.6. Linear regression models were fitted with candidate-allele frequency as the dependent variable and sampling-site altitude as the independent variable.

Uncorrected proof

Regression slopes were expressed per 1 000 m increase in altitude and reported with 95% confidence intervals. Spearman rank correlations were also calculated. *P* values from the linear regression and Spearman analyses were adjusted separately across the eight candidate loci using the Benjamini–Hochberg method. Sensitivity to individual sampling sites was assessed by refitting each model after excluding one population at a time.

Bayesian fine-mapping of regions defined by GWAS–GEA lead SNPs

Bayesian fine-mapping was used to prioritize variants within candidate regions. Each GWAS–GEA lead SNP in Supplementary Table S15 was used as a regional anchor (Benner et al., 2016). Genotypes of all variants within ±500 kb of each lead SNP were extracted from the same BEAGLE-imputed dataset used for the GWAS and GEA analyses, and an LD matrix was calculated for each region. FINEMAP integrated GWAS marginal effects, standard errors, and the LD matrix to estimate posterior probabilities of causal configurations using the Shotgun Stochastic Search algorithm. The marginal posterior inclusion probability (PIP) was calculated for each variant. Variants were ranked by PIP within each ±500-kb region, and the highest-PIP variant in each region was retained as the FINEMAP-prioritized variant. Because these variants were selected from the local region rather than restricted to the original lead SNP, their coordinates could differ from those of the GWAS–GEA lead SNPs. The prioritized variants were then used for raw-VCF concordance assessment, functional annotation, and transcription factor binding motif prediction.

PCR amplification and Sanger sequencing verification

For each of the eight candidate variants, DNA samples were selected from each of the three genotype classes assigned using the whole-genome sequencing (WGS) data after genotype imputation. The target regions were amplified by PCR and subjected to Sanger sequencing. Sequencing chromatograms were inspected to confirm the corresponding alleles and genotype calls. Primer sequences are provided in Supplementary Table S16.

Uncorrected Proof

FIMO-based transcription factor motif analysis

For each of the eight FINEMAP-prioritized variants, 51-bp reference (REF) and alternate (ALT) sequences were constructed, with the variant at the center and 25 bp on each side. Find Individual Motif Occurrences (FIMO) v.5.5.9 (Grant et al., 2011) was used to scan 841 matrices from the JASPAR 2022 (https://jaspar.elixir.no/) CORE vertebrates non-redundant motif set (Castro-Mondragon et al., 2022) on both the forward and reverse-complement strands. Only motif matches spanning the central variant were evaluated, and REF and ALT matches were paired according to the same motif matrix, position, and strand. No fixed percentage-similarity threshold was applied. Motif matches were retained when FIMO $q \leq 0.05$ in either allele. The q value controlled multiple testing of motif matches but did not test the statistical significance of the REF-ALT score difference.

Public RNA-seq expression and correlation analysis

Public RNA sequencing (RNA-seq) data from GSE261409 were used to examine candidate-gene expression (Yan et al., 2024). The subset analyzed in this study comprised 49 adult female Hu sheep approximately 1.5 years old, sampled at the low-altitude baseline (day 0) and at 7 days, 14 days, 21 days, and approximately 8 months after translocation to high altitude. The five sampling stages included 10, 10, 10, 10, and 9 animals, respectively. Each animal was sampled at only one stage, so there were no longitudinal repeated measurements. Multiple tissues were collected from each animal. The cross-tissue analysis included 243 samples: 48 lung, 48 heart, and 49 each from spleen, muscle, and liver. Raw counts were normalized using the trimmed mean of M-values method. The lung analysis included one sample from each of 48 animals. Differential-expression statistics were obtained from the published DESeq2 analysis.

Statistical analysis

Uncorrected proof

Unless otherwise stated, statistical analyses and data visualization were performed in R. Phenotypic data are presented as mean ± standard error. Spearman rank correlations were used to assess relationships among blood indicators and between individual-level effect-allele dosage and blood indicators. At the population level, correlations were assessed between effect-allele frequencies and environmental variables. Pearson correlation analysis was used to assess expression relationships among candidate genes in the public lung RNA-seq dataset after adjustment for sampling stage and RNA integrity number (RIN). The three pairwise *P* values were adjusted using the Benjamini–Hochberg method. Overall differences among genotype groups were assessed using the Kruskal–Wallis test, followed by two-sided pairwise Wilcoxon rank-sum tests. The Benjamini–Hochberg method was used to control the false discovery rate. Statistical significance was defined as **P*<0.05, ***P*<0.01, and ****P*<0.001.

RESULTS

Genotype imputation performance

The final analysis dataset comprised 3 352 196 SNPs in 220 animals, with 0% genotype missingness after imputation and quality control (Supplementary Table S3). At the eight FINEMAP-prioritized variants, 121–192 animals had non-missing raw genotypes, with 100% concordance between these calls and the corresponding imputed genotypes (Supplementary Table S5). The validation dataset comprised 3 809 791 SNPs, and 2 962 754 masked sample-by-variant genotype calls were evaluated; all masked calls were imputed. Overall genotype concordance was 81.88%, and the squared Pearson correlation coefficient between pre-masking and post-imputation ALT-allele counts was 0.629. For the MAF bins 0.05–0.10, 0.10–0.20, 0.20–0.30, and 0.30–0.50, genotype concordance was 86.77%, 80.92%, 76.35%, and 74.92%, respectively, and the corresponding R^2 values were 0.654, 0.616, 0.531, and

Uncorrected proof

0.443. Concordance for homozygous-reference, heterozygous, and homozygous-alternate genotypes was 97.94%, 62.68%, and 92.15%, respectively (Supplementary Table S4).

GWAS analysis of blood physiological and biochemical indicators and detection of related candidate variants

Single-trait GWAS was performed for 24 blood physiological and biochemical indicators. A total of 368 phenotype-associated signals were detected. After removing duplicates, 324 candidate SNPs covering all 24 indicators were retained (Figure 1). These signals were distributed non-uniformly across the genome and formed concentrated peaks in several chromosomal regions. The 324 candidate SNPs were annotated to 310 genes. After further restricting candidate sites to intragenic regions or regions within 50 kb upstream or downstream of the nearest gene, 216 phenotype-associated signals remained. These signals corresponded to 193 SNPs and 188 non-redundant candidate genes. In addition, 20 SNPs and 26 genes were associated with two or more blood indicators. Figure 1A shows the candidate SNP screening workflow. Figure 1B–D shows the genome-wide distribution of association signals across phenotype modules and highlights candidate genes including *ANTXR1*, *CFAP58*, *FAS*, *GDPD4*, *GSTA1-1*, *MITF*, *OSBPL10*, and *TTC23L*. Genomic inflation factors for the 24 FarmCPU analyses are summarized in Supplementary Table S17.

For a sample size of 220 and a significance threshold of $P<1\times10^{-6}$, the approximate power analysis indicated that a single variant would need to explain approximately 13.99% of the phenotypic variance to achieve 80% power (Supplementary Figure S2). In the within-population phenotype-centered sensitivity analysis, all eight prespecified prioritized variants retained effect directions consistent with those in the primary GWAS. Although the exact-SNP associations were attenuated, the strongest signal within each prespecified ±50-kb candidate region ranked within the top 5% genome-wide in the corresponding centered GWAS (Supplementary Table S6). To further assess whether these regional signals were

Uncorrected proof

disproportionately influenced by any single population, we conducted a leave-one-population-out analysis. Across 88 prespecified locus-trait evaluations, the exact-SNP effect direction remained consistent in 72 evaluations, whereas a regional best signal within the genome-wide top 5% was retained in 84 evaluations (Supplementary Table S7). Together, these analyses indicated that regional rankings were generally more stable than exact-SNP estimates.

Environmental heterogeneity and geographic isolation are associated with genetic differentiation among Tibetan sheep populations

Mantel tests were used to evaluate relationships between genetic distance, geographic distance, and environmental distance among Tibetan sheep populations (Figure 2A). Genetic distance was positively correlated with geographic distance, supporting isolation by distance (IBD; Mantel's r=0.173, P=1×10^{-4}). Genetic distance was also positively correlated with environmental distance, supporting isolation by environment (IBE; Mantel's r=0.377, P=1×10^{-4}). The Mantel correlation was higher for environmental distance than for geographic distance. In the population-level MMRR analysis, environmental PC1 distance remained positively associated with genetic distance after geographic distance was included in the model (standardized β=0.641, P=0.0033). The overall model had an R^2 of 0.311 and a permutation P value of 0.0185 (Supplementary Table S9).

Environmental differences among the 11 sampling sites were further analyzed. After redundancy reduction of altitude and 19 bioclimatic variables, eight relatively independent environmental factors were retained. PCA separated several Tibetan sheep populations in environmental space (Figure 2B). Altitude and temperature- and precipitation-related variables contributed strongly to the first two principal components.

GEA was performed after the Mantel tests indicated a significant relationship between environmental distance and genetic distance. After genomic-inflation-factor correction, λ was

Uncorrected proof

1. Using Benjamini–Hochberg correction, 121 environment-associated SNPs were identified at FDR<0.05. The maximum raw *P* value corresponding to this FDR threshold was 1.77×10^{-6}. These signals were distributed across multiple genomic regions, with association peaks near *GDPD4*, *TTC23L*, *MITF*, *GSTA1-1*, *FAS*, *CFAP58*, *ANTXR1*, and *OSBPL10* (Figure 2D). The quantile–quantile (QQ) plot showed a reasonable overall distribution of association tests (Figure 2E). The population-level BayPass sensitivity analysis showed that 107 of the 121 LFMM candidate SNPs ranked within the top 1% of all 47 194 tested SNPs, whereas all eight core SNPs ranked within the top 1.1% (Supplementary Table S11). Correlation analysis linked several candidate sites or genes with environmental variables related to altitude, temperature, or precipitation (Figure 2C). Together, the Mantel tests, environmental PCA, and GEA results showed that genetic differentiation was associated with both geographic distance and environmental differences.

Integration of GWAS and GEA reveals candidate regions with both phenotype and environment associations

To identify regions supported by both GWAS and GEA, the 121 significant environment-associated SNPs and the top 0.05% of SNPs from GWAS of each of the 24 blood indicators were assigned to non-overlapping 50-kb windows. An UpSet matrix summarized overlap among the resulting candidate-window sets. The union of the GWAS and GEA candidate-window sets comprised 22 022 non-redundant 50-kb windows. The environmental association set contained 67 windows. The immunoglobulin, oxidative stress, and inflammatory cytokine sets contained 5 935, 8 044, and 12 907 windows, respectively (Figure 2F). Sensitivity results for 25-, 50-, and 100-kb windows and four GWAS thresholds are provided in Supplementary Table S18.

Among the 67 environment-associated windows, 27 overlapped with at least one GWAS candidate window, accounting for 40.30%. The numbers of overlapping windows involving

Uncorrected proof

473 immunoglobulin, oxidative-stress, and inflammatory-cytokine traits were 8, 9, and 17,
474 respectively. However, a chromosome-wise circular-shift permutation test with matching for
475 window-level MAF and SNP density showed that the observed overlap count did not exceed
476 the permutation-based expectation (observed=27, permutation mean=31.75, enrichment
477 ratio=0.850, empirical P=0.9057, and Benjamini–Hochberg (BH)-adjusted P=1.000;
478 Supplementary Table S18). These window overlaps were therefore used for exploratory
479 prioritization of candidate regions. Of the 38 foreground and 19 711 background genes, 28
480 (73.68%) and 13 350 (67.73%) received at least one GO annotation, respectively. KOBAS
481 mapped 37 foreground genes (97.37%) and 19 028 background genes (96.53%). GO
482 enrichment analysis identified 11 terms with BH-adjusted $P \leq 0.05$, of which six were
483 supported by at least two foreground genes. No KEGG pathway remained statistically
484 significant after BH correction. Figure 2G therefore presents eight nominally enriched
485 pathways with $P<0.05$ that were supported by at least two foreground genes. Complete GO
486 and KEGG results are provided in Supplementary Tables S13 and S14.

487 **Association patterns of GWAS–GEA lead SNPs with blood indicators and**
488 **environmental gradients**

489 Physical-distance-based regional deduplication and cross-trait support filtering of the 121
490 LFMM candidate SNPs yielded 17 region-representative candidate SNPs. Eight of these
491 SNPs were intragenic and were retained as GWAS–GEA lead SNPs for downstream
492 analyses, corresponding to *ANTXR1*, *GDPD4*, *GSTA1-1*, *MITF*, *FAS*, *CFAP58*, *OSBPL10*,
493 and *TTC23L* (Supplementary Table S15). A phenotype correlation network was constructed
494 to evaluate relationships between these lead SNPs and blood indicators. The 24 indicators
495 formed modules related to immunoglobulins, oxidative stress, and inflammatory cytokines.
496 Several GWAS–GEA lead SNPs were associated with blood indicators, with the direction of
497 association varying among loci and traits (Figure 3A).

Uncorrected proof

Overall differences among genotype groups were significant after Benjamini–Hochberg adjustment for all four locus-trait combinations shown in Figure 3B: *MITF*-GSH-PX, H(2)=34.17, adjusted P=1.52×10^{-7}; *ANTXR1*-IgD, H(2)=23.81, adjusted P=1.35×10^{-5}; *FAS*-ROS, H(2)=20.17, adjusted P=5.55×10^{-5}; and *TTC23L*-SOD, H(2)=15.19, adjusted P=5.04×10^{-4}. Of the 12 pairwise Wilcoxon comparisons, 10 remained significant after Benjamini–Hochberg adjustment (Figure 3B). A correlation heatmap also showed significant correlations between GWAS–GEA lead SNP dosage and multiple blood indicators (Figure 3C). Additional genotype–phenotype comparisons at the eight candidate loci are shown in Supplementary Figures S3–S10.

Spatial distribution of GWAS–GEA lead variants and relationships with altitude

Candidate allele frequencies at the *MITF*, *ANTXR1*, *TTC23L*, and *FAS* lead loci were positively correlated with altitude in Spearman analyses across the 11 populations (Figure 4). The regression slopes per 1 000 m were 0.138 (95% confidence interval (CI), 0.020–0.256), 0.140 (0.030–0.250), 0.109 (0.018–0.201), and 0.095 (−0.001–0.192), respectively; the corresponding BH-adjusted regression P values were 0.053, 0.053, 0.053, and 0.083. Spearman's ρ values were 0.659, 0.827, 0.648, and 0.653, with BH-adjusted P values of 0.041, 0.0068, 0.041, and 0.041, respectively. In leave-one-population-out analyses, the regression slope remained positive in all 11 iterations for each locus, indicating that the direction of association was not attributable to a single sampling population. Population genotype distributions and altitude associations at four additional loci are shown in Supplementary Figures S11 and S12, respectively.

Fine-mapping and predicted regulatory effects within candidate regions

Bayesian fine-mapping was performed in representative regions centered on GWAS–GEA lead SNPs related to *MITF*, *ANTXR1*, *FAS*, and *TTC23L*. Within each ±500-kb region, variants were ranked by PIP and the highest-PIP variant was retained as the FINEMAP-

uncorrected proof

prioritized variant (Figure 5A; Supplementary Figure S13). These variants were not necessarily identical to the original GWAS–GEA lead SNPs; therefore, the coordinates reported for the two analysis stages differ in some regions. PCR-Sanger sequencing confirmed the expected alleles and genotype calls at eight selected candidate variants. One representative chromatogram for each of the eight loci is shown in Supplementary Figure S14. Functional validation of the candidate variants was not performed. Across the eight candidate loci, 32 locus-matrix pairs contained at least one variant-overlapping FIMO match with $q \leq 0.05$ in either allele. Representative matches and REF-ALT scores for *MITF-SPIC*, *ANTXR1-GATA1*, *FAS-PRDM9*, and *TTC23L-ZNF76* are shown in Figure 5B, C. The complete results are provided in Supplementary Table S19 and Supplementary Figure S15.

Tissue expression patterns and adjusted expression correlations of candidate genes

Public RNA-seq data showed that *ANTXR1*, *FAS*, and *MITF* were expressed in lung, heart, spleen, muscle, and liver tissues (Figure 5D). After adjustment for sampling stage and RIN, the lung expression correlations between *ANTXR1* and *MITF*, *ANTXR1* and *FAS*, and *MITF* and *FAS* were 0.672, 0.546, and 0.821, respectively. The corresponding BH-adjusted *P* values were 1.21×10^{-6}, 1.52×10^{-4}, and 4.67×10^{-11}, respectively (Figure 5E). Published DESeq2 results showed that *MITF* was significantly upregulated in the lung at 21 days after high-altitude translocation and that *FAS* was significantly upregulated in the heart after approximately 8 months, whereas *ANTXR1* did not meet the combined differential-expression criteria used in the original study (Supplementary Figure S16).

DISCUSSION

By integrating GWAS of blood physiological and biochemical indicators with GEA, we prioritized candidate regions supported by both blood phenotype associations and environmental gradients. GWAS links variants to blood indicators, whereas GEA links genotypes to environmental gradients. These complementary associations support candidate

Uncorrected proof

prioritization but do not establish causal effects of the variants (Forester et al., 2018). Within the regions defined by the GWAS–GEA lead SNPs, FINEMAP provided a further step of variant prioritization. The FINEMAP-prioritized variants may therefore differ from the original lead SNPs.

Recent genomic and multi-omic studies indicate that high-altitude livestock share certain functional features involving hypoxia responses, vascular regulation, and tissue homeostasis, although different genomic mechanisms have been highlighted across species. A multi-assembly graph-genome study including wild and domestic yaks and related bovine species systematically characterized structural variation and its evolutionary origins in domestic yaks (Liu et al., 2023). An integrative genomic, transcriptomic, and epigenomic study in Tibetan pigs suggested that cis-regulatory variants affecting *EPAS1*, *KLF13*, *SPRED1*, and *CFD* may contribute to angiogenesis and circulatory system development (Lai et al., 2025). A graph-based goat pangenome study identified a structural variant in the *PAPSS2* region under selection in Tibetan goats (Bian et al., 2024). Complementing these studies, we integrated GWAS of 24 blood physiological and biochemical indicators with GEA to prioritize Tibetan sheep candidate regions related to redox homeostasis, immune regulation, apoptosis, and tissue homeostasis.

Functional annotation and previous studies provide biological context for the candidate genes prioritized here. *ANTXR1*, *GDPD4*, *GSTA1-1*, *MITF*, *FAS*, *CFAP58*, *OSBPL10*, and *TTC23L* may be involved in redox homeostasis, immune regulation, tissue homeostasis, lipid metabolism, apoptosis, or cellular-structure maintenance. For example, *GSTA1-1*, *MITF*, *FAS*, *ANTXR1*, and *OSBPL10* have been linked to glutathione-mediated antioxidant defense, cell differentiation and oxidative stress response, immune homeostasis, extracellular matrix homeostasis, and lipid metabolism (Liang et al., 2005; Perttila et al., 2009; Reeves et al.,

Uncorrected proof

2013; Hu et al., 2017). These annotations provide functional clues for blood physiological and biochemical variation under high-altitude conditions.

The *MITF*-related candidate site was associated with GSH-PX levels. Its candidate-allele frequency was also positively correlated with altitude in the population-level Spearman analysis. *MITF* is a transcriptional regulator involved in melanocyte development, differentiation, pigmentation, and stress responses (Levy et al., 2006; Cheli et al., 2010). Recent studies also suggest that *MITF* can contribute to cellular antioxidant defense through the nuclear factor erythroid 2-related factor 2 (NRF2) pathway or ROS-related physiological factors (Han et al., 2020; Roider et al., 2024).

The *ANTXR1*-related SNP was associated with IgD levels, and its candidate-allele frequency was positively correlated with altitude in the population-level Spearman analysis. *ANTXR1*, also known as tumor endothelial marker 8 (TEM8), is involved in cell-matrix interactions, extracellular matrix homeostasis, and maintenance of tissue structure. It may also regulate vascular and connective tissue homeostasis (Hotchkiss et al., 2005; Reeves et al., 2013; Besschetnova et al., 2015; Andersen et al., 2016; Hu et al., 2017). Because hypoxia can influence tissue homeostasis, vascular regulation, and extracellular matrix remodeling, *ANTXR1*-related variants provide candidate clues for vascular and tissue-remodeling processes in Tibetan sheep.

FAS is a classical apoptosis- and immune-regulation gene. The *FAS*-related SNP was associated with ROS levels, and its candidate-allele frequency was positively correlated with altitude in the Spearman analysis; however, the linear regression was not statistically significant after BH correction (adjusted $P = 0.083$). Thus, the evidence supports a monotonic association between candidate-allele frequency and altitude, while support for a linear relationship remains limited. FAS–FAS ligand (FASL) signaling contributes to immune-cell clearance, peripheral immune tolerance, and apoptosis regulation (Strasser et al., 2009; Yolcu

Uncorrected proof

et al., 2017). *FAS* has also been reported to stabilize adult erythropoiesis and support stress erythropoiesis responses through negative feedback regulation (Koulnis et al., 2011). Apoptosis-related signals such as FASL may change under high-altitude environments (Kosanovic et al., 2019). The *FAS*-related region therefore remains a candidate for testing links among apoptosis, oxidative stress, blood phenotypes, and plateau environmental responses.

The *TTC23L*-related SNP was associated with SOD levels, and its candidate-allele frequency was positively correlated with altitude in the population-level Spearman analysis. However, direct functional evidence linking *TTC23L* to high-altitude adaptation, oxidative stress, or blood physiological regulation remains limited. The *TTC23L* region is therefore a candidate supported by association evidence, rather than an established functional locus for high-altitude adaptation. *GDPD4* and *CFAP58* are additional candidates of interest. *GDPD4* encodes glycerophosphodiester phosphodiesterase domain-containing protein 4. Genetic variation within *GDPD4* has been reported to be significantly associated with fatty acid metabolism-related traits in livestock, and the gene has also been identified as a candidate associated with local adaptation in environmental genomic studies (Gheyas et al., 2021; Liu et al., 2024). *CFAP58* encodes cilia and flagella associated protein 58, but its potential role in high-altitude adaptation in livestock remains poorly understood. The contributions of *GDPD4* and *CFAP58* to high-altitude adaptation require further investigation.

Recent studies of high-altitude adaptation increasingly emphasize non-coding and cis-regulatory variants. For example, chromatin accessibility maps and regulatory networks have revealed open chromatin regions and target-gene relationships involved in human hypoxia adaptation (Xin et al., 2020). FIMO identified variant-overlapping motif matches involving *SPIC*, *GATA1*, *PRDM9*, and *ZNF76* at representative FINEMAP-prioritized variants. Public RNA-seq data showed that *ANTXR1*, *FAS*, and *MITF* were expressed across multiple tissues

Uncorrected proof

and retained positive expression correlations in lung after adjustment for sampling stage and RNA integrity number. These results provide complementary functional clues for the candidate regions.

Several limitations should be considered when interpreting these findings. Although local nutritional and husbandry conditions could not be standardized, within-population phenotype centering and leave-one-population-out analyses generally retained the regional signals. In the leave-one-population-out analysis, the best regional signal remained within the genome-wide top 5% in 84 of 88 evaluations. These analyses assessed robustness within the study cohort and did not constitute independent replication. BayPass and MMRR analyses supported the population-level GEA patterns, but the limited numbers of sampling sites and animals constrained separation of environmental effects from population-history effects and reduced power to detect small-effect or low-frequency variants. Imputation performance was assessed by within-dataset genotype masking rather than comparison with independent high-depth genotypes. PCR-Sanger sequencing supported genotype calls at eight candidate variants but did not establish their regulatory functions. These association-based findings therefore require validation in independent populations and functional experiments.

CONCLUSION

By integrating GWAS of 24 blood physiological and biochemical indicators with GEA, we prioritized eight candidate loci associated with high-altitude adaptation in Tibetan sheep, involving *ANTXR1*, *GDPD4*, *GSTA1-1*, *MITF*, *FAS*, *CFAP58*, *OSBPL10*, and *TTC23L*. Genotype–phenotype associations, population-level allele-frequency gradients, and regional fine-mapping supported the prioritization of these candidate regions. Motif predictions and tissue-expression analyses provided complementary clues for their functional interpretation. Together, these findings link blood physiological and biochemical variation to environment-

Uncorrected proof

associated genetic variation, providing candidate loci for functional studies, marker development, and breeding for environmental resilience.

DATA AVAILABILITY

The raw whole-genome sequencing datasets supporting this study are available from the NCBI Sequence Read Archive under BioProject accession number PRJNA1138910: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1138910. The public multi-tissue RNA-seq data used in this study are available from the Gene Expression Omnibus under accession GSE261409. All other data generated or analyzed during this study are included in the article and its supplementary information files. Scripts for the GWAS and GEA analyses are available at https://github.com/TJX-gwas/bioinformatics-scripts. The in-house Tibetan sheep reference genome HB and its annotation files have not yet been deposited in a public repository but are available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Figures:

Supplementary Figure S1. FarmCPU GWAS diagnostics for 24 blood traits.

Supplementary Figure S2. Analytical GWAS power under effect-size and MAF scenarios.

Supplementary Figure S3. Exploratory genotype–phenotype comparisons at the *ANTXR1* candidate locus.

Supplementary Figure S4. Exploratory genotype–phenotype comparisons at the *CFAP58* candidate locus.

Supplementary Figure S5. Exploratory genotype–phenotype comparisons at the *FAS* candidate locus.

Supplementary Figure S6. Exploratory genotype–phenotype comparisons at the *GDPD4* candidate locus.

Uncorrected proof

Supplementary Figure S7. Exploratory genotype–phenotype comparisons at the *GSTA1-1* candidate locus.

Supplementary Figure S8. Exploratory genotype–phenotype comparisons at the *MITF* candidate locus.

Supplementary Figure S9. Exploratory genotype–phenotype comparisons at the *OSBPL10* candidate locus.

Supplementary Figure S10. Exploratory genotype–phenotype comparisons at the *TTC23L* candidate locus.

Supplementary Figure S11. Genotype frequency distributions of eight candidate loci across 11 sheep populations.

Supplementary Figure S12. Relationships between candidate-allele frequencies and altitude at four additional GWAS–GEA lead SNPs.

Supplementary Figure S13. Regional GWAS association signals and FINEMAP statistical prioritization across eight candidate regions.

Supplementary Figure S14. Technical validation of eight candidate variants by PCR-Sanger sequencing.

Supplementary Figure S15. FIMO motif predictions spanning four additional candidate variants.

Supplementary Figure S16. Candidate-gene expression during high-altitude acclimatization in Hu sheep.

Supplementary Tables:

Supplementary Table S1. Sampling information for 11 Tibetan sheep populations.

Supplementary Table S2. Descriptive statistics of 24 blood physiological and biochemical indicators in Tibetan sheep.

Uncorrected proof

694 Supplementary Table S3. Summary of genotype missingness before and after BEAGLE
695 imputation and quality control.
696 Supplementary Table S4. MAF-stratified BEAGLE imputation performance assessed by
697 random genotype masking.
698 Supplementary Table S5. Summary of raw VCF support and imputation concordance for
699 eight FINEMAP-prioritized candidate variants.
700 Supplementary Table S6. Within-population phenotype-centered GWAS sensitivity analysis
701 of eight prioritized candidate regions.
702 Supplementary Table S7. Leave-one-population-out GWAS sensitivity analysis of eight
703 prespecified candidate regions.
704 Supplementary Table S8. Candidate SNPs identified by GWAS for 24 blood physiological
705 and biochemical indicators.
706 Supplementary Table S9. Population-level multiple matrix regression with randomization
707 analysis.
708 Supplementary Table S10. Significant genotype–environment association SNPs identified by
709 LFMM.
710 Supplementary Table S11. Population-level BayPass sensitivity analysis of the 121 LFMM
711 candidate SNPs.
712 Supplementary Table S12. GWAS–GEA overlapping 50-kb windows and annotated genes.
713 Supplementary Table S13. Gene Ontology enrichment analysis of genes near GWAS–GEA
714 overlapping candidate regions using an analyzable-gene background.
715 Supplementary Table S14. KEGG pathway enrichment analysis of genes near GWAS–GEA
716 overlapping candidate regions using an analyzable-gene background.
717 Supplementary Table S15. Prioritization criteria and annotation of eight GWAS–GEA lead
718 SNPs.

Uncorrected proof

Supplementary Table S16. Primer information for PCR amplification and Sanger sequencing of eight selected candidate variants.

Supplementary Table S17. Genomic inflation factors for FarmCPU analyses of 24 blood physiological and biochemical indicators in Tibetan sheep.

Supplementary Table S18. Circular-shift assessment of overlap between GWAS and GEA candidate windows.

Supplementary Table S19. FIMO predictions for variant-overlapping transcription factor motif matches at eight candidate loci.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

J.X.T., Z.K.L., and J.B.L. conceived and designed the study and jointly supervised the work. J.X.T. and Z.K.L. performed the experiments, conducted the bioinformatics and statistical analyses, designed the figures and tables, and wrote the manuscript. J.X.T., C.Y., B.W.C., F.W., T.T.G., Z.K.L., and J.B.L. participated in animal management and sample collection. C.Y., Z.K.L., and J.B.L. critically revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank the staff involved in animal management and sample collection.

FUNDING

This work was supported by the Science and Technology Projects of Xizang Autonomous Region (XZ202502ZY0046), the Innovation Project of Chinese Academy of Agricultural Sciences (25-LZIHPS-07), the National Technical System for Wool Sheep Industry (CARS-39-02), and the Kunlun Talents High-End Innovative and Entrepreneurial Talents Program of Qinghai Province (QHKLYC-GDCXCY-2024-300).

Uncorrected proof

744 **REFERENCES**

745 Andersen NJ, Boguslawski EA, Naidu AS, et al. 2016. Anthrax toxin receptor 1 is essential
746 for arteriogenesis in a mouse model of hindlimb ischemia. *PLoS ONE*, **11**(1): e0146586.
747 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and
748 powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
749 *(Methodological)*, **57**(1): 289–300.
750 Benner C, Spencer CC, Havulinna AS, et al. 2016. FINEMAP: efficient variable selection
751 using summary data from genome-wide association studies. *Bioinformatics*, **32**(10):
752 1493–1501.
753 Besschetnova TY, Ichimura T, Katebi N, et al. 2015. Regulatory mechanisms of anthrax
754 toxin receptor 1-dependent vascular and connective tissue homeostasis. *Matrix Biology*,
755 **42**: 56–73.
756 Bian P, Li J, Zhou S, et al. 2024. A graph-based goat pangenome reveals structural variations
757 involved in domestication and adaptation. *Molecular Biology and Evolution*, **41**(12):
758 msae251.
759 Bigham AW, Lee FS. 2014. Human high-altitude adaptation: forward genetics meets the HIF
760 pathway. *Genes & Development*, **28**(20): 2189–2204.
761 Browning BL, Tian X, Zhou Y, et al. 2021. Fast two-stage phasing of large-scale sequence
762 data. *The American Journal of Human Genetics*, **108**(10): 1880–1890.
763 Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-
764 generation reference panels. *The American Journal of Human Genetics*, **103**(3): 338–
765 348.
766 Bu D, Luo H, Huo P, et al. 2021. KOBAS-i: intelligent prioritization and exploratory
767 visualization of biological functions for gene enrichment analysis. *Nucleic Acids*
768 *Research*, **49**(W1): W317–W325.

Uncorrected proof

769 Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, et al. 2022. JASPAR 2022: the
770 9th release of the open-access database of transcription factor binding profiles. *Nucleic*
771 *Acids Research*, **50**(D1): D165–D173.
772 Caye K, Jumentier B, Lepeule J, et al. 2019. LFMM 2: fast and accurate inference of gene-
773 environment associations in genome-wide studies. *Molecular Biology and Evolution*,
774 **36**(4): 852–860.
775 Chang CC, Chow CC, Tellier LC, et al. 2015. Second-generation PLINK: rising to the
776 challenge of larger and richer datasets. *GigaScience*, **4**(1): 7.
777 Cheli Y, Ohanna M, Ballotti R, et al. 2010. Fifteen-year quest for microphthalmia-associated
778 transcription factor target genes. *Pigment Cell & Melanoma Research*, **23**(1): 27–40.
779 Dauphin B, Rellstab C, Wüest RO, et al. 2023. Re-thinking the environment in landscape
780 genomics. *Trends in Ecology & Evolution*, **38**(3): 261–274.
781 DePristo MA, Banks E, Poplin R, et al. 2011. A framework for variation discovery and
782 genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**(5): 491–
783 498.
784 Eltzschig HK, Carmeliet P. 2011. Hypoxia and inflammation. *New England Journal of*
785 *Medicine*, **364**(7): 656–665.
786 Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for
787 global land areas. *International Journal of Climatology*, **37**(12): 4302–4315.
788 Forester BR, Lasky JR, Wagner HH, et al. 2018. Comparing methods for detecting multilocus
789 adaptation with multivariate genotype-environment associations. *Molecular Ecology*,
790 **27**(9): 2215–2233.
791 Frichot E, Schoville SD, Bouchard G, et al. 2013. Testing for associations between loci and
792 environmental gradients using latent factor mixed models. *Molecular Biology and*
793 *Evolution*, **30**(7): 1687–1699.

Uncorrected proof

794 Gautier M. 2015. Genome-wide scan for adaptive divergence and association with
795 population-specific covariates. *Genetics*, **201**(4): 1555–1579.
796 Gheyas AA, Vallejo-Trujillo A, Kebede A, et al. 2021. Integrated environmental and
797 genomic analysis reveals the drivers of local adaptation in African indigenous chickens.
798 *Molecular Biology and Evolution*, **38**(10): 4268–4285.
799 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
800 *Bioinformatics*, **27**(7): 1017–1018.
801 Han S, Chen J, Hua J, et al. 2020. MITF protects against oxidative damage-induced retinal
802 degeneration by regulating the NRF2 pathway in the retinal pigment epithelium. *Redox*
803 *Biology*, **34**: 101537.
804 Hotchkiss K, Basile C, Spring S, et al. 2005. TEM8 expression stimulates endothelial cell
805 adhesion and migration by regulating cell-matrix interactions on collagen. *Experimental*
806 *Cell Research*, **305**(1): 133–144.
807 Hu K, Olsen BR, Besschetnova TY. 2017. Cell autonomous ANTXR1-mediated regulation of
808 extracellular matrix components in primary fibroblasts. *Matrix Biology*, **62**: 105–114.
809 Hu XJ, Yang J, Xie XL, et al. 2019. The genome landscape of Tibetan sheep reveals adaptive
810 introgression from argali and the history of early human settlements on the Qinghai-
811 Tibetan Plateau. *Molecular Biology and Evolution*, **36**(2): 283–303.
812 Kosanovic D, Platzek SM, Petrovic A, et al. 2019. Circulating apoptotic signals during acute
813 and chronic exposure to high altitude in Kyrgyz population. *Frontiers in Physiology*, **10**:
814 54.
815 Koulnis M, Liu Y, Hallstrom K, et al. 2011. Negative autoregulation by Fas stabilizes adult
816 erythropoiesis and accelerates its stress response. *PLoS ONE*, **6**(7): e21192.

Uncorrected proof

817 Lai D, Qu X, Wang Z, et al. 2025. Integrative genomic, transcriptomic and epigenomic
818 analysis reveals cis-regulatory contributions to high-altitude adaptation in Tibetan pigs.
819 *Molecular Biology and Evolution*, **42**(7): msaf169.
820 Lasky JR, Josephs EB, Morris GP. 2023. Genotype-environment associations to reveal the
821 molecular basis of environmental adaptation. *The Plant Cell*, **35**(1): 125–138.
822 Lê S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. *Journal*
823 *of Statistical Software*, **25**(1): 1–18.
824 Levy C, Khaled M, Fisher DE. 2006. MITF: master regulator of melanocyte development and
825 melanoma oncogene. *Trends in Molecular Medicine*, **12**(9): 406–414.
826 Lex A, Gehlenborg N, Strobelt H, et al. 2014. UpSet: visualization of intersecting sets. *IEEE*
827 *Transactions on Visualization and Computer Graphics*, **20**(12): 1983–1992.
828 Li C, Chen B, Langda S, et al. 2024. Multi-omic analyses shed light on the genetic control of
829 high-altitude adaptation in sheep. *Genomics, Proteomics & Bioinformatics*, **22**(2):
830 qzae030.
831 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
832 transform. *Bioinformatics*, **25**(14): 1754–1760.
833 Liang FQ, Alssadi R, Morehead P, et al. 2005. Enhanced expression of glutathione-S-
834 transferase A1-1 protects against oxidative stress in human retinal pigment epithelial
835 cells. *Experimental Eye Research*, **80**(1): 113–119.
836 Liang X, Duan Q, Li B, et al. 2024. Genomic structural variation contributes to evolved
837 changes in gene expression in high-altitude Tibetan sheep. *Proceedings of the National*
838 *Academy of Sciences*, **121**(27): e2322291121.
839 Liu J, Sebastià C, Jové-Juncà T, et al. 2024. Identification of genomic regions associated with
840 fatty acid metabolism across blood, liver, backfat and muscle in pigs. *Genetics Selection*
841 *Evolution*, **56**(1): 66.

Uncorrected proof

842 Liu X, Huang M, Fan B, et al. 2016. Iterative usage of fixed and random effect models for
843 powerful and efficient genome-wide association studies. *PLoS Genetics*, **12**(2):
844 e1005767.
845 Liu X, Liu W, Lenstra JA, et al. 2023. Evolutionary origin of genomic structural variations in
846 domestic yaks. *Nature Communications*, **14**(1): 5617.
847 Ma YF, Han XM, Huang CP, et al. 2019. Population genomics analysis revealed origin and
848 high-altitude adaptation of Tibetan pigs. *Scientific Reports*, **9**(1): 11463.
849 McGettrick AF, O'Neill LA. 2020. The role of HIF in immunity and inflammation. *Cell*
850 *Metabolism*, **32**(4): 524–536.
851 Niu Y, Li Y, Zhao Y, et al. 2024. Whole-genome sequencing identifies functional genes for
852 environmental adaptation in Chinese sheep. *Journal of Genetics and Genomics*, **51**(11):
853 1278–1285.
854 Pebesma E. 2018. Simple features for R: standardized support for spatial vector data. *The R*
855 *Journal*, **10**(1): 439–446.
856 Pena E, El Alam S, Siques P, et al. 2022. Oxidative stress and diseases associated with high-
857 altitude exposure. *Antioxidants*, **11**(2): 267.
858 Perttilä J, Merikanto K, Naukkarinen J, et al. 2009. OSBPL10, a novel candidate gene for
859 high triglyceride trait in dyslipidemic Finnish subjects, regulates cellular lipid
860 metabolism. *Journal of Molecular Medicine*, **87**(8): 825–835.
861 Pham K, Parikh K, Heinrich EC. 2021. Hypoxia and inflammation: insights from high-
862 altitude physiology. *Frontiers in Physiology*, **12**: 676782.
863 Qiu Q, Zhang G, Ma T, et al. 2012. The yak genome and adaptation to life at high altitude.
864 *Nature Genetics*, **44**(8): 946–949.
865 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
866 features. *Bioinformatics*, **26**(6): 841–842.

Uncorrected proof

867 Reeves C, Charles-Horvath P, Kitajewski J. 2013. Studies in mice reveal a role for anthrax
868 toxin receptors in matrix metalloproteinase function and extracellular matrix
869 homeostasis. *Toxins*, **5**(2): 315–326.
870 Rellstab C, Gugerli F, Eckert AJ, et al. 2015. A practical guide to environmental association
871 analysis in landscape genomics. *Molecular Ecology*, **24**(17): 4348–4370.
872 Roider E, Lakatos AIT, McConnell AM, et al. 2024. MITF regulates IDH1, NNT, and a
873 transcriptional program protecting melanoma from reactive oxygen species. *Scientific*
874 *Reports*, **14**(1): 21527.
875 Siques P, Brito J, Pena E. 2018. Reactive oxygen species and pulmonary vasculature during
876 hypobaric hypoxia. *Frontiers in Physiology*, **9**: 865.
877 Storz JF, Cheviron ZA. 2021. Physiological genomics of adaptation to high-altitude hypoxia.
878 *Annual Review of Animal Biosciences*, **9**: 149–171.
879 Strasser A, Jost PJ, Nagata S. 2009. The many roles of FAS receptor signaling in the immune
880 system. *Immunity*, **30**(2): 180–192.
881 Taylor CT, Colgan SP. 2017. Regulation of immunity and inflammation by hypoxia in
882 immunological niches. *Nature Reviews Immunology*, **17**(12): 774–785.
883 Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants
884 from high-throughput sequencing data. *Nucleic Acids Research*, **38**(16): e164.
885 Wang MS, Li Y, Peng MS, et al. 2015. Genomic analyses reveal potential independent
886 adaptation to high altitude in Tibetan chickens. *Molecular Biology and Evolution*, **32**(7):
887 1880–1889.
888 Xin J, Zhang H, He Y, et al. 2020. Chromatin accessibility landscape and regulatory network
889 of high-altitude hypoxia adaptation. *Nature Communications*, **11**(1): 4928.

Uncorrected proof

Yan Z, Yang J, Wei WT, et al. 2024. A time-resolved multi-omics atlas of transcriptional regulation in response to high-altitude hypoxia across whole-body tissues. *Nature Communications*, **15**(1): 3970.

Yin L, Zhang H, Tang Z, et al. 2021. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, **19**(4): 619–628.

Yolcu ES, Shirwan H, Askenasy N. 2017. Fas/Fas-ligand interaction as a mechanism of immune homeostasis and β-cell cytotoxicity: enforcement rather than neutralization for treatment of type 1 diabetes. *Frontiers in Immunology*, **8**: 342.

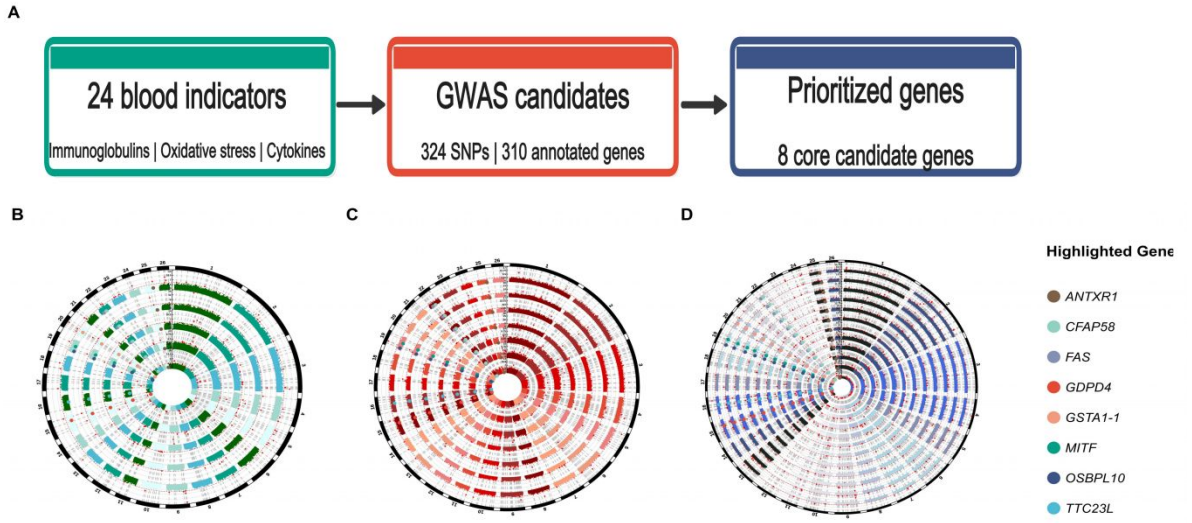
Zhang D, Cheng J, Li X, et al. 2024a. Comprehensive multi-tissue epigenome atlas in sheep: a resource for complex traits, domestication, and breeding. *iMeta*, **3**(6): e254.

Zhang W, Yuan C, An X, et al. 2024b. Genomic insights into Tibetan sheep adaptation to different altitude environments. *International Journal of Molecular Sciences*, **25**(22): 12394.

Zhao Z, Nie R, Fan H, et al. 2026. Large-scale genomics reveals three-source ancestry and layered adaptation to high altitude in Tibetan chickens. *Advanced Science*, **13**(42): e22994.

uncorrected proof

915 **Figure legends**

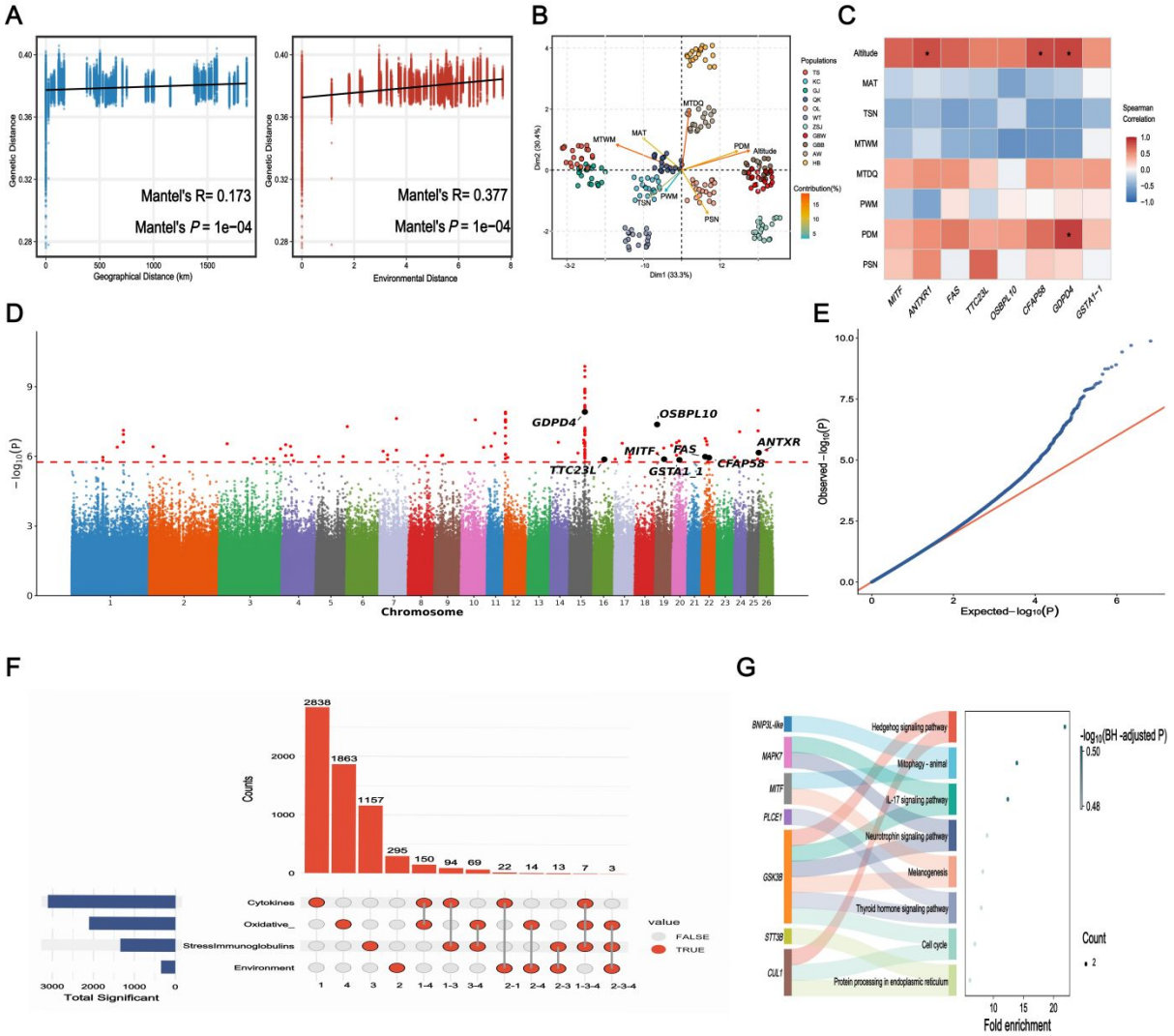


916

917 **Figure 1. Genome-wide association candidate signals for 24 blood physiological and**
918 **biochemical indicators**

919 A: Workflow for screening genome-wide association study (GWAS) candidate single-
920 nucleotide polymorphisms (SNPs) across 24 blood indicators. B–D: Circular Manhattan plots
921 showing GWAS signals for different phenotype modules. Colored markers indicate key
922 candidate genes, including *ANTXR1*, *CFAP58*, *FAS*, *GDPD4*, *GSTA1-1*, *MITF*, *OSBPL10*,
923 and *TTC23L*.

uncorrected p



924

925 **Figure 2. Environmental factors, genetic structure, and integration of genomic and**
926 **environmental associations in Tibetan sheep populations**

927 A: Mantel-test relationships between genetic distance and geographic or environmental
928 distance. B: Principal component analysis (PCA) of environmental variables from 11
929 sampling sites. C: Spearman correlations between effect-allele frequencies at lead single-
930 nucleotide polymorphisms (SNPs) from integrated genome-wide association study (GWAS)
931 and genotype-environment association (GEA) analyses and environmental variables across
932 populations. D: Manhattan plot of GEA results. The dashed line indicates the false discovery
933 rate (FDR) significance threshold, and candidate genes are labeled. E: Quantile–quantile

Zoological Research

(QQ) plot of GEA results. F: UpSet analysis of environmental association windows and GWAS candidate windows related to immunoglobulins, oxidative stress, and inflammatory cytokines. G: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with nominal $P<0.05$ supported by at least two foreground genes. None of the displayed pathways remained statistically significant after Benjamini–Hochberg correction. MAT, mean annual temperature; TSN, temperature seasonality; MTWM, maximum temperature of the warmest month; MTDQ, mean temperature of the driest quarter; PWM, precipitation of the wettest month; PDM, precipitation of the driest month; PSN, precipitation seasonality.

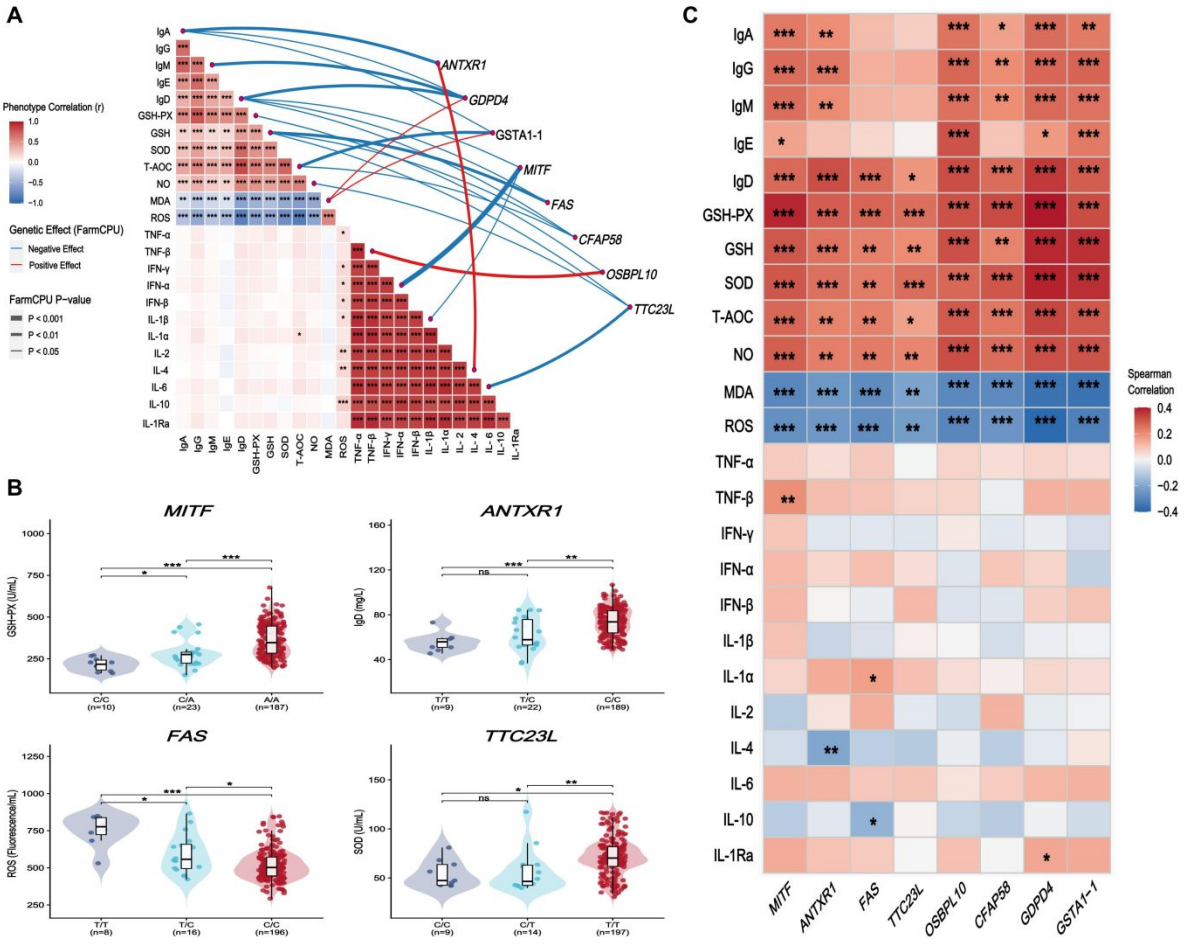


Figure 3. Exploratory follow-up associations between representative lead variants and blood physiological and biochemical indicators

A: Correlation network among 24 blood indicators and associations between lead single-nucleotide polymorphisms (SNPs) from integrated genome-wide association study (GWAS)

and genotype-environment association (GEA) analyses and blood indicators. B: Genotype-group comparisons for key blood indicators at lead loci, including *MITF*–GSH-PX, *ANTXR1*–IgD, *FAS*–ROS, and *TTC23L*–SOD (GSH-PX, glutathione peroxidase; IgD, immunoglobulin D; ROS, reactive oxygen species; SOD, superoxide dismutase). C: Spearman correlations between lead-SNP effect-allele dosage and 24 blood indicators.

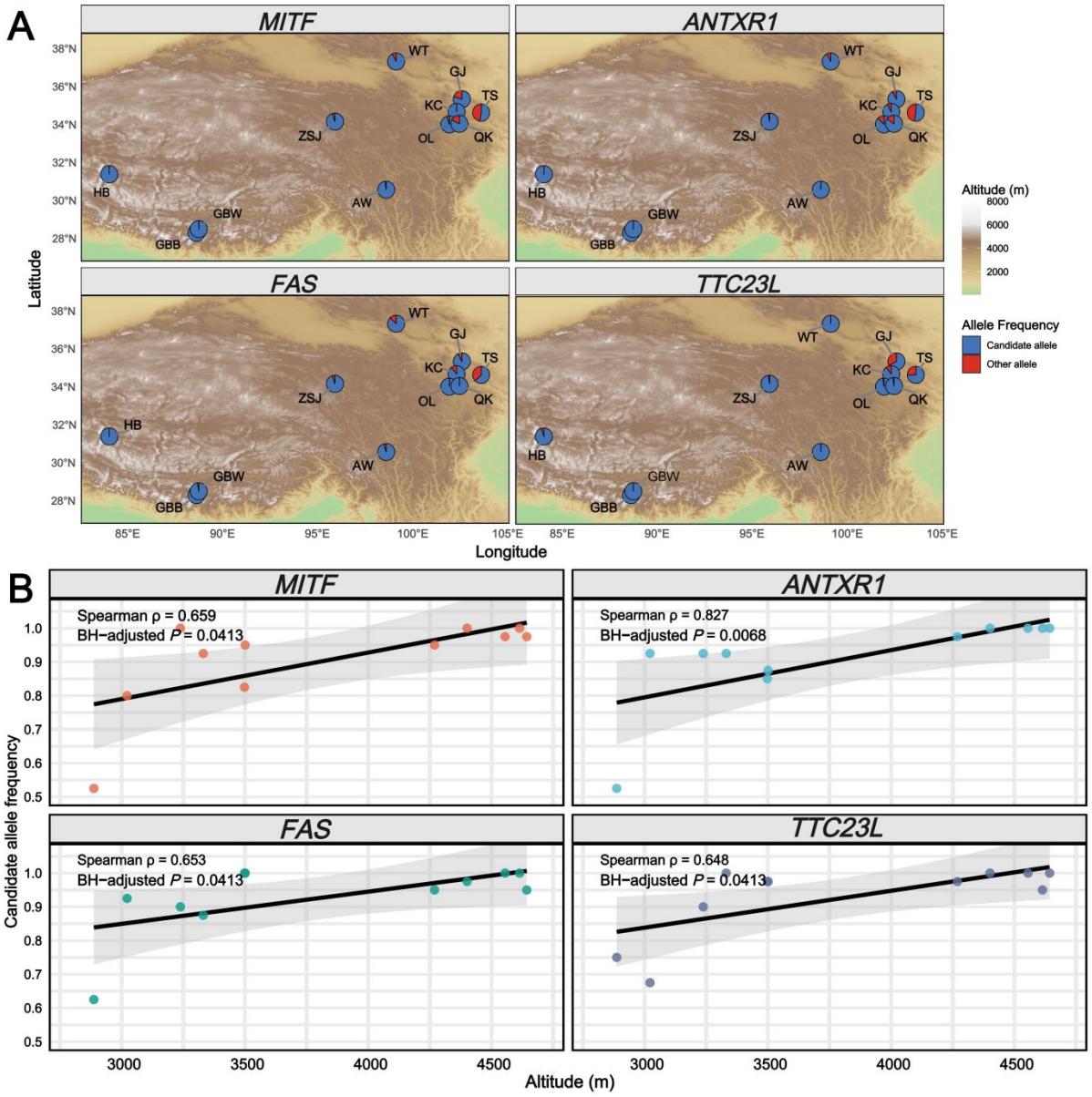


Figure 4. Geographic distributions of candidate alleles at representative lead variants and their associations with altitude

A: Geographic distributions of allele frequencies at lead single-nucleotide polymorphisms (SNPs) related to *MITF*, *ANTXR1*, *FAS*, and *TTC23L* across 11 Tibetan sheep populations. B: Relationships between population-level candidate-allele frequencies and altitude. Each point represents one population; lines and shaded bands indicate linear fits and their 95% confidence intervals, respectively. Spearman's ρ and Benjamini–Hochberg (BH)-adjusted *P* values are shown.

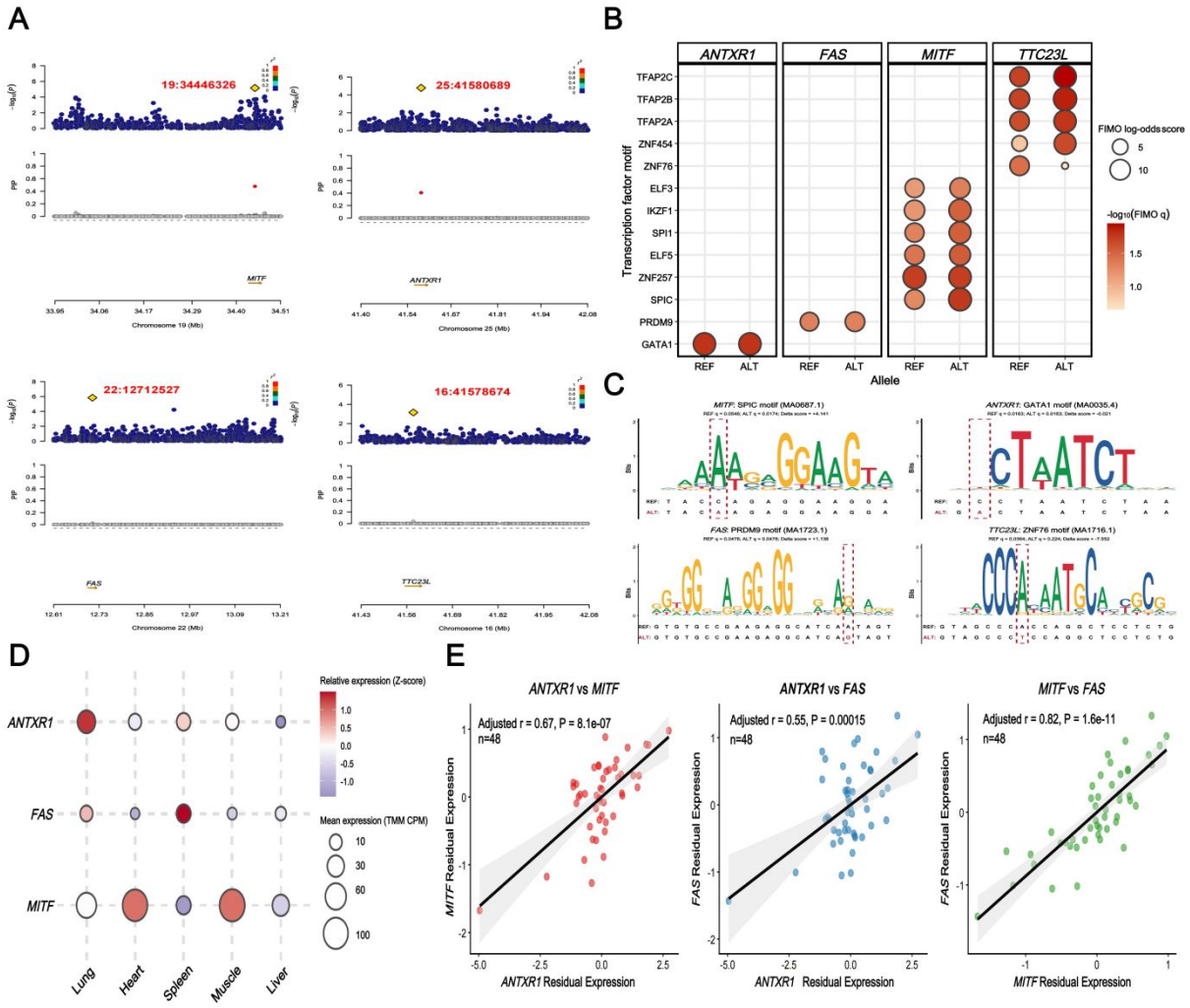


Figure 5. FINEMAP, motif prediction, and expression analyses of representative candidate regions

A: Local genome-wide association study (GWAS) signals and FINEMAP results for candidate regions near *MITF*, *ANTXR1*, *FAS*, and *TTC23L*. B: Motif matches identified by Find Individual Motif Occurrences (FIMO) that span FINEMAP-prioritized variants in

Zoological Research

reference (REF) and alternate (ALT) sequences. Bubble size represents the FIMO log-odds score, and color represents $-\log_{10}(q)$. C: Sequence logos of representative FIMO motifs. Red dashed boxes indicate variant positions. D: Expression patterns of *ANTXR1*, *FAS*, and *MITF* in lung, heart, spleen, muscle, and liver. E: Expression correlations among *ANTXR1*, *FAS*, and *MITF* in 48 lung samples from 48 animals after adjustment for sampling stage and RNA integrity number.

整合全基因组关联分析与景观基因组学筛选藏绵羊高海拔适应相关候选位点

唐健翔[1,2]，袁超[1,2]，陈博雯[1,2]，王凡[3]，郭婷婷[1,2]，卢曾奎[1,2,*]，刘建斌[1,2,*]

[1]中国农业科学院兰州畜牧与兽药研究所，农业农村部青藏高原畜禽遗传育种重点实验室，兰州 730050，中国

[2]中国农业科学院羊育种工程技术研究中心，兰州 730050，中国

[3]西北民族大学生命科学与工程学院，兰州 730100，中国

* 通讯作者：

刘建斌：liujianbin@caas.cn;

卢曾奎：luzengkui@caas.cn

摘要：

高海拔环境使藏绵羊（*Ovis aries*）面临低氧、低温、强紫外线辐射及季节性营养压力。免疫球蛋白、氧化应激及炎症细胞因子指标可用于表征对上述环境压力的生理响应，但这些指标变异的遗传基础尚未完全阐明。本研究整合了不同海拔 11 个种群 220 只成年藏绵羊公羊的全基因组重测序数据、24 项血液生理生化指标及采样地点环境变量。全基因组关联分析（GWAS）筛选出 324 个候选单核苷酸多态性位点（SNP），注释涉及 310 个基因。基因型－环境关联分析（GEA）在错误发现率（FDR）<0.05 时鉴定

uncorrected proof

出 121 个环境关联 SNP。在包含环境关联 SNP 的 67 个互不重叠的 50-kb 窗口中，27 个 SNP（40.30%）与血液性状关联窗口重叠。*MITF*、*ANTXR1*、*FAS* 和 *TTC23L* 等代表性候选基因对应的区域与谷胱甘肽过氧化物酶（GSH-PX）、免疫球蛋白 D（IgD）等血液指标相关。等位基因频率梯度、贝叶斯区域优先排序、基序扫描及公开组织表达数据为候选区域的功能解释提供了互补线索。这些发现为藏绵羊高海拔适应相关位点的独立验证和功能研究提供了候选目标。

关键词：血液生理生化指标；基因型-环境关联；氧化应激；免疫调控；贝叶斯精细定位

uncorrected proof