animals

MDPI

Article

Comparative Transcriptomic Analysis of Alpaca (*Vicugna pacos*) Skin Identifies Candidate Genes and Biological Processes Associated with Fiber Color

Alejandra Jiménez-Espinoza [1,2], Lariza Pahuara [1,2], Felipe Avila [3], Flor Corredor [4], Alan Cruz [5], Gustavo Gutiérrez [6], Daisy Obispo [1], Solange R. Paredes-Moscosso [1,†], Alex Yucra [5], Ricardo Fujita [1], Claudio Villegas-Llerena [1,*,†] and Oscar Acosta [1,*]

1 Centro de Genética y Biología Molecular, Facultad de Medicina Humana, Universidad de San Martin de Porres, Lima 15024, Peru; dobispoa@usmp.pe (D.O.); pcblspar@upc.edu.pe (S.R.P.-M.)
2 Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima 15081, Peru
3 Veterinary Genetics Laboratory, Weill School of Veterinary Medicine, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; ffavila@ucdavis.edu
4 Estación Experimental La Molina, Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), La Molina, Lima l2791, Peru
5 Estación Científica Pacomarca, Inca Tops S.A, Av. Miguel Forga 348, Arequipa 040101, Peru; alancruzc@outlook.com (A.C.)
6 Departament of Animal Science, Iowa State University, Ames, IA 50011, USA
* Correspondence: pcblcvil@upc.edu.pe (C.V.-L.); oacostac@usmp.pe (O.A.)
† Current address: Biomolecules Laboratory, School of Biology, Faculty of Health Sciences, Universidad Peruana de Ciencias Aplicadas, Lima 15067, Peru.

Simple Summary

The natural color of alpaca fiber is economically important and represents a valuable component of the species' diversity. We examined gene activity in skin samples from 65 Huacaya alpacas with different fiber colors and related it to measurements of lightness and color. Pigmented alpacas showed numerous differences in gene expression compared with white alpacas. Major candidates included genes involved in melanin production and transport, such as *TYR* (tyrosinase), *TYRP1* (tyrosinase-related protein 1), *MITF* (melanocyte-inducing transcription factor) and *EDNRB* (endothelin receptor type B). Three other candidate genomic regions designated as unassigned loci (LOC) in VicPac4 also showed association with premelanosome protein (*PMEL*) and the OCA2 melanosomal transmembrane protein (*OCA2/P*). Genes with higher expression in pigmented skin were mainly related to pigmentation and melanosomes, whereas genes with lower expression were associated with the epidermis and keratins. Agreement across complementary analyses supports these genes as candidates for further investigation. However, because the number of animals differed across color categories and we did not perform experimental validation, the findings should be interpreted as transcriptomic associations rather than evidence of causal relationships.

Abstract

The transcriptomic basis of natural fiber-color diversity in alpacas remains incompletely understood. We analyzed skin biopsies from 65 Huacaya alpacas by RNA sequencing (RNA-seq) and integrated phenotypic color categories with instrumental Commission Internationale de l'Éclairage L*a*b* CIELAB measurements, a standardized color space system used to objectively measure and quantify color. Differential expression was assessed using DESeq2 differential-expression analysis package with adjustment for collection batch, sex, and fiber-color category, while L* (lightness), a* (green–red axis), and b* (blue–yellow axis) were examined in separate continuous models. The categorical analysis included

Check for updates

Academic Editor: Weimin Wang

Received: 14 September 2026
Revised: 15 October 2026
Accepted: 27 October 2026
Published: 11 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

64 samples and 25,405 genes. We identified 4968, 3993, 1469, and 7607 differentially expressed genes in light brown (LB), dark brown (DB), light fawn intensity Y (LFY), and light fawn intensity Z (LFZ), respectively, relative to white (W), with higher expression predominating in pigmented categories. Global transcriptomic structure and gene-level expression patterns were associated with CIELAB parameters, supporting the use of these measurements as complementary quantitative descriptors of fiber color. However, these associations should be interpreted cautiously because of the imbalance between collection batches and the heterogeneous cellular composition of the skin biopsies. Integration of categorical, continuous, and functional evidence prioritized *TYR* (tyrosinase), *TYRP1* (tyrosinase-related protein 1), *MITF* (melanocyte-inducing transcription factor), *EDNRB* (endothelin receptor type B), *MFSD12* (major facilitator superfamily domain containing 12), and *GPR143* (G protein-coupled receptor 143), among other genes. Two regions with unassigned locus (LOC) identifiers in VicPac4 showed evidence consistent with premelanosome protein (*PMEL*), and a third with *OCA2* melanosomal transmembrane protein (*OCA2/P*), although these identities remained interpretive proposal. Genes with higher expression in pigmented categories were overrepresented in pigmentation and melanosome processes, whereas genes with lower expression were associated with epidermal and keratin-related processes. These findings identify transcriptomic changes associated with natural fiber-color variation and candidates for functional validation.

Keywords: alpaca; CIELAB; differential expression; fiber color; melanogenesis; melanosome; RNA-seq; VicPac4

1. Introduction

The alpaca (*Vicugna pacos*) produces fiber in a wide range of natural colors and is an economically, culturally, and agriculturally important resource for Andean communities and livestock systems [1–3]. The textile industry has historically favored white fiber because it is readily dyed. Nevertheless, naturally pigmented fibers have attracted renewed interest because they can be used in their original color, thereby reducing the water, energy, and chemicals required for dyeing [3,4]. This renewed value contrasts with the historical decline of some colored-alpaca populations and underscores the need to conserve their genetic diversity and understand the biological mechanisms associated with fiber pigmentation [3,5].

In mammals, skin and hair color depend mainly on the amount, type, distribution, and transfer of melanin produced by melanocytes. Eumelanin primarily generates black and brown shades, whereas pheomelanin contributes yellow and reddish tones [6–8]. Both pigments are synthesized in melanosomes, specialized organelles that subsequently transfer their contents to keratinocytes in the hair follicle. This process requires coordinated activity among genes involved in melanocyte development and survival, melanin synthesis, melanosome formation and transport, and regulation of the balance between pigment types [9–11]. Experimental work in melanocyte models has further shown that melanosome maturation depends on coordinated PMEL processing and regulation of organelle pH [12].

Genetic studies in alpacas have described variants and expression differences in classical pigmentation genes, including *MC1R*, *ASIP*, and *SLC7A11* [13–16]. Recent alpaca studies have also reported coat-color associations with TYRP1 and DCT polymorphisms [17] and genome-wide signals involving ASIP and KIT [18]. Preliminary transcriptomic analyses in alpacas, sheep, and goats have also identified differential expression of *TYR*, *TYRP1*, *DCT*, *PMEL*, *MLANA*, and other genes related to melanocyte and melanosome function [19–22]. However, much of the available evidence comes from two-color comparisons or candidate-

gene studies. Consequently, the joint characterization of multiple color categories using transcriptomic designs that account for other relevant sources of variation among animals remains limited [3,22]. More recent integrated transcriptomic and metabolomic work in sheep has linked pigmentation differences to TYR, TYRP1, DCT, PMEL, MLANA, and SLC45A2 [23].

A further challenge is that visual categories simplify a continuous, multidimensional phenotype. The CIELAB color space quantitatively describes L* (lightness), a* (green–red axis), and b* (blue–yellow axis) [24]. This system has proved useful for objectively characterizing alpaca fiber-color variation and its relationship with other fiber traits [4]. Integrating these measurements with RNA sequencing (RNA-seq) profiles makes it possible to determine whether gene expression differs not only among color categories but also along individual colorimetric gradients. This approach is particularly useful when visual categories overlap or have unequal sample sizes [4,22].

Functional interpretation is also constrained by the genomic annotation available for the species. VicPac4 is a recent chromosome-level assembly of *Vicugna pacos* and represents an advance over previous alpaca reference genomes [25]. Nevertheless, in the annotation used here, many gene models retain LOC identifiers or uninformative descriptions. Candidate characterization therefore requires careful integration of native annotation with orthology, functional evidence, and protein similarity. This strategy preserves source traceability and avoids treating a match to a gene in another species as an automatic change to the locus's official identity.

We analyzed skin-biopsy transcriptomes from 65 Huacaya alpacas representing different fiber-color categories. We hypothesized that phenotypic and instrumental color variation would be associated with consistent changes in genes and processes related to melanogenesis, melanosomes, and the structural organization of skin. The objectives were to: (i) assess data quality and characterize global transcriptomic structure; (ii) identify differentially expressed genes among color categories, emphasizing comparisons with white fiber; (iii) examine associations between gene expression and L*, a*, and b*; (iv) identify overrepresented biological processes; and (v) integrate transcriptomic, functional, and protein evidence to prioritize known genes and LOC. By combining categorical analysis, quantitative colorimetry, and traceable annotation, we sought to provide a transcriptomic basis for prioritizing mechanisms and gene candidates associated with natural alpaca fiber-color diversity.

2. Materials and Methods

2.1. Study Design, Animals, and Sample Collection

We conducted an observational comparative study of skin gene expression in Huacaya alpacas (*Vicugna pacos*) grouped into ten fleece-color categories following the classification framework of Cruz [26] and Norma Tecnica Peruana (NTP) 231.301 [27]: white (W; $n = 5$), light fawn intensities X, Y, and Z (LFX, $n = 1$; LFY, $n = 9$; LFZ, $n = 14$), a dark variant of light fawn intensity X (LFXO; $n = 2$), light brown (LB; $n = 13$), dark brown (DB; $n = 12$), reddish black (RB; $n = 4$), black (B; $n = 3$), and gray (G; $n = 2$). Samples were obtained from the Pacomarca Scientific Station of the Inca Group, located in Llalli District, Melgar Province, Puno Region, Peru. We included 65 apparently healthy alpacas (39 females and 26 males), aged 2.47–6.65 years, without visible skin lesions or signs of disease. Each animal was individually identified by a numbered ear tag and had an individual record containing fiber phenotype information. Owner authorization was obtained for all participating animals. Detailed characteristics of the animals included in the categorical differential-expression model are provided in Supplementary Table S1.

Samples were collected during two field campaigns, in April 2024 and August 2025, corresponding to collection batches 1 and 2, respectively. During both campaigns, the same operators obtained skin biopsies from the right mid-rib region under local anaesthesia with 2% lidocaine, using sterile, disposable 3 mm skin biopsy punches (Integra Miltex, York, PA, USA). The same sampling, RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) stabilization, and cold-chain transport protocol was applied to all samples. Two independent samples were obtained from each animal: one intended for RNA sequencing and one retained as a backup. Immediately after collection, biopsies were immersed in RNAlater, samples were then transferred to dry ice and transported under frozen conditions to the Centro de Investigación de Genética y Biología Molecular, Universidad de San Martín de Porres, Lima, Peru, where they were stored at −80 °C until further processing. The protocol was approved through expedited review by the Institutional Ethics Committee for the Use of Animals of Universidad Peruana Cayetano Heredia (SIDISI 24002; certificate CIEA-008-03-24; 23 February 2024).

2.2. CIELAB Colorimetric Characterization

Alpacas were classified by phenotypic fleece-color category, recorded in the source metadata as ColorCode. All sampled animals had solid-colored fleeces; no spotted individuals were included. Fiber samples were obtained from the mid-rib region of the same animals used for skin-biopsy transcriptomic analysis. Fiber color was measured instrumentally using a Chroma Meter CR-210 (Konica Minolta, Tokyo, Japan), following the alpaca-fleece colorimetry procedure described by Yucra Mendoza [28]. The resulting CIELAB parameters were L* (lightness), a* (green–red axis), and b* (blue–yellow axis). Agreement between phenotypic categories and instrumental CIELAB measurements was evaluated using group centroids and colorimetric distances. Each sample was also assigned to its nearest centroid in a leave-one-out procedure, in which the evaluated sample did not contribute to its own group's centroid. Category separation was further explored using silhouette coefficients and principal component analysis of L*, a*, and b* (Table S2). This colorimetric characterization was descriptive only and was not used to change the original categories or reassign samples.

2.3. RNA Extraction, Library Preparation, and Sequencing

Total RNA extraction, quality control, library preparation and sequencing were performed by Novogene (Sacramento, CA, USA). One set of skin biopsies from each animal was shipped under frozen conditions. RNA was extracted using RNeasy minikit (QIAGEN, Hilden, Germany) and sample quality was assessed using Tapestation 4150 (Agilent Technologies, Santa Clara, CA, USA). Libraries were prepared using Illumina RNA library preparation with polyadenylated [poly(A)] RNA enrichment and sequenced on an Illumina NovaSeq X Plus platform (Illumina, San Diego, CA, USA), generating 150 bp paired-end reads. A minimum of 12 Gb of raw sequencing data was obtained for each sample. Raw sequencing data were then provided to our group for downstream bioinformatics analysis.

2.4. Quality Control, Alignment, and Quantification

The quality of the raw sequencing reads was initially assessed using FastQC v0.12.1 [29]. Reads were subsequently processed using fastp v1.3.4 [30], with automatic adapter detection, quality trimming at both ends using four-nucleotide sliding windows with a minimum mean Phred score of 20, and removal of reads shorter than 50 nucleotides. The JavaScript Object Notation (JSON) reports generated by fastp were used to summarize read retention, Phred quality score ≥ 20 (Q20), Phred quality score ≥ 30 (Q30), and guanine–cytosine (GC) content across the 65 libraries. Per-library processing and read-quality metrics are provided in Supplementary Table S3.

Filtered reads were aligned with Spliced Transcripts Alignment to a Reference (STAR) v2.7.11b [31] to the *Vicugna pacos* VicPac4 reference assembly (RefSeq GCF_048564905.1) and its corresponding Gene Transfer Format (GTF) annotation. The STAR index was built using sjdbOverhang = 149. The resulting coordinate-sorted Binary Alignment/Map (BAM) files were indexed using SAMtools v1.21 [32]. Gene-level expression was quantified using featureCounts v2.0.8 [33], with paired-end fragments assigned to exons and subsequently summarized by gene_id. The featureCounts options -p, –countReadPairs, -t exon, -g gene_id, and -s 0 were used, with -s 0 specifying unstranded read counting. Per-library alignment and quantification metrics are reported in Supplementary Table S4.

2.5. Differential Expression Analysis Among Fiber-Color Categories

The count matrix and metadata were integrated in R v4.4.3 after verifying one-to-one correspondence and identical ordering of all 65 sample identifiers. Differential expression was analyzed with DESeq2 v1.46.0 [34]. Before model fitting, genes with very low expression were filtered out. A gene was retained when it had at least 10 counts in at least two samples. Two samples were chosen as the minimum because some fleece-color groups contained only two animals, allowing genes expressed specifically in these small groups to remain in the analysis. This filtering criterion does not imply that comparisons based on two animals provide strong statistical evidence; therefore, contrasts involving small groups were interpreted as exploratory.

The main design included collection batch and sex as adjustment covariates and fiber-color category as the variable of interest (~ Batch + Sex + ColorCode). Sex and batch were included to account for potential variation in skin gene expression. Batch represented the two collection campaigns; seasonal effects could not be estimated independently of other campaign-associated differences. Age was not included as an additional covariate. Exploratory analyses detected no significant correlations between age and the first ten transcriptomic principal components after Benjamini–Hochberg correction across the covariate-association tests. Age distributions also did not differ significantly among color categories or between collection batches. These findings supported retaining the model without age but do not exclude age-related effects on individual genes. A sensitivity analysis using a reduced model without collection batch (~ Sex + ColorCode) assessed whether the color-associated results depended on adjustment for collection period. In the primary model, W, females, and collection batch 2 were set as the reference levels. Automatic replacement of outlier counts was disabled by setting minReplicatesForReplace = Inf, thereby retaining the original observed counts. Pedigree records and genomic relatedness estimates were not available for this analysis; consequently, relatedness among animals within or across fleece-color categories could not be evaluated or included in the models. Adjustment for sex and collection batch does not account for relatedness.

Although automatic replacement of outlier counts was disabled, Cook's distance was still used with the default DESeq2 cutoff to identify gene-level results strongly influenced by individual observations. DESeq2's independent filtering procedure was also retained to increase statistical power by excluding genes with insufficient information from the multiple-testing correction. Categorical differential-expression analysis included only fleece-color groups represented by at least two animals. Therefore, the single LFX sample was excluded from categorical comparisons but retained in descriptive analyses and in the continuous models of L*, a*, and b*. This exclusion was based on the absence of replication within the LFX category, rather than on sample quality. The nine eligible color groups yielded 36 pairwise comparisons. The eight comparisons between each non-white color category and white fleece were defined a priori as primary contrasts, whereas comparisons among

non-white categories were considered secondary. The complete categorical contrast plan is provided in Supplementary Table S5.

Any contrast involving a group represented by fewer than five samples was classified as exploratory. Four primary contrasts with adequate sample representation in both analytical datasets were selected for the main analysis: LB versus W, DB versus W, LFY versus W, and LFZ versus W. Differentially expressed genes (DEGs) were selected using a dual criterion of a Benjamini–Hochberg-adjusted p value < 0.05 and an absolute log2 fold change ≥ 1, corresponding to a minimum twofold difference in expression. The absolute log2 fold-change (log2FC) cutoff was applied to the original estimated coefficients as an effect-size filter after Wald testing against zero, rather than as a threshold-based hypothesis test. Sensitivity to this filter was assessed at stricter cutoffs of 1.5 and 2, retaining the original adjusted p values and gene-evaluability criteria (Supplementary Table S6). Sensitivity of the highlighted candidates is reported in Supplementary Table S7. Multiple-testing correction was performed separately within each contrast. A nested subset of 53 samples comprising W, LB, DB, LFY, and LFZ was used to assess the internal stability of these four shared contrasts.

2.6. Exploratory Analysis and Transcriptomic Associations with CIELAB Parameters

Variance-stabilizing transformation (VST) was applied for principal component analysis (PCA) and heatmaps. These displays were used only for exploratory and descriptive purposes and not for differential-expression inference. L*, a*, and b* were standardized and analyzed separately using continuous DESeq2 models. The main model included collection batch, sex, and the standardized CIELAB parameter (~ Batch + Sex + CIELAB_z). Sensitivity analyses used models without collection batch and models excluding the sample from the singleton color category. The estimated coefficient was interpreted as the log2 change in gene expression per one-standard-deviation increase in the evaluated colorimetric parameter. p values were adjusted by the Benjamini–Hochberg procedure separately for each parameter; a joint correction across all three CIELAB parameters was also calculated. An association was considered robust when statistical significance and effect direction were retained in the main model and both sensitivity analyses. The transcriptomic PCA used the 500 genes with the highest variance in blind VST expression values, with centering and without scaling. PC1-PC3 and PC2-PC3 projections were added for the same 65-sample PCA (Supplementary Figure S3), and eigenvalues, explained variance, and cumulative explained variance are reported in Supplementary Table S8.

2.7. Functional Annotation, Enrichment Analysis, and Candidate Prioritization

A traceable gene dictionary was constructed for VicPac4 using gene_id as the primary identifier and retaining links to National Center for Biotechnology Information (NCBI) GeneID, transcript, and protein accessions. NCBI orthology relationships were recorded according to relationship type and target species; orthologs were not assigned solely from matching gene symbols. For each Gene Ontology (GO) association, the annotation source, evidence code, qualifier, and hierarchical relationships among terms were retained. To balance annotation coverage and evidence strength, positive, non-obsolete GO associations were classified according to their level of functional support. Level 1 comprised associations supported by both a direct *Vicugna pacos* annotation and at least one formally assigned ortholog; Level 2 comprised associations supported only by a direct *Vicugna pacos* annotation; and Level 3 comprised associations without a direct *Vicugna pacos* annotation but supported by orthologs from at least two species. Together, these three levels defined the primary functional annotation profile ('PrimarySupport_L1_L3') used for the main enrichment analysis. Associations supported by a single non-human ortholog (Level 4) or exclusively by a human ortholog (Level 5) were evaluated only in sensitivity and exploratory analyses,

respectively. Detailed definitions of the functional-evidence levels and annotation profiles are provided in Supplementary Table S9.

Functional enrichment was evaluated using two complementary Gene Ontology (GO) approaches. First, one-sided hypergeometric tests were used to identify GO terms overrepresented among all differentially expressed genes (DEGs) and, separately, among genes showing increased or decreased expression. Second, a threshold-independent gene set enrichment analysis was performed using the multilevel algorithm implemented in the fgsea R package (v1.32.4; fgseaMultilevel function). For this analysis, all tested genes were ranked according to the signed Wald statistic obtained from DESeq2 for each of the eight primary color-versus-white contrasts.

Both approaches were conducted separately for the three GO domains—biological process (BP), molecular function (MF), and cellular component (CC)—and were restricted to GO terms containing between 10 and 500 genes. The overrepresentation background comprised genes eligible for DEG classification after independent filtering and annotated in the corresponding functional profile, whereas fgsea used propagated GO associations from the PrimarySupport_L1_L3 profile. p values were adjusted using the Benjamini–Hochberg procedure within each analysis stratum. For fgsea, 10,000 simple permutations and $eps = 0$ were used, and terms without calculable p values were excluded from significance summaries.

Candidate genes were required to meet the differential-expression criteria (adjusted $^*p^* < 0.05$ and $|\log2FC| \geq 1$) in at least one categorical contrast and to belong to a predefined functional module within the PrimarySupport_L1_L3 profile. Candidates were then ranked using an 11-point scoring system that combined functional relevance (0–2 points), direct *Vicugna pacos* annotation (0–1), ortholog support (0–1), DEG detection across the 53- and 65-sample analyses (1–2), consistency of expression direction across comparable contrasts (0–1), categorical–L* concordance and robustness across sensitivity models (0–3), and robust association with a* or b* (0–1). CIELAB associations were considered significant at adjusted $p < 0.05$ and robust when they remained significant with the same effect direction in the primary model and both sensitivity models. Candidates were grouped into five priority tiers (P1–P5) and ranked within each tier according to their total score and extent of DEG support. The complete scoring criteria and definitions of the five priority tiers are provided in Supplementary Table S10. Because the 53- and 65-sample datasets overlapped, agreement between them was interpreted as cross-analysis robustness rather than independent replication. For the compact candidate table, calls versus white were counted separately for each analysis set and contrast across the four focal comparisons shared by the 53- and 65-sample datasets. The maximum absolute categorical coefficient was summarized over retained DEG results from these four comparisons and the secondary LFZ-versus-LFY comparison in both datasets.

For priority loci bearing provisional LOC designations, protein-sequence similarity was assessed using Basic Local Alignment Search Tool for proteins (BLASTP v2.17.0+) against Swiss-Prot, and conserved domains were examined using Reverse Position-Specific Basic Local Alignment Search Tool (RPS-BLAST v2.17.0+) against the Conserved Domain Database (CDD). This evidence was used to propose interpretive functional labels and confidence levels without replacing or modifying the official VicPac4 identifiers. The functional-evidence levels and annotation profiles used to construct the Gene Ontology analysis sets are described in Supplementary Table S9, whereas the protein-similarity, conserved-domain, and other supporting evidence for the priority LOC is reported in Supplementary Table S11.

3. Results

3.1. Sample Characteristics and RNA-seq Performance

The study included 65 Huacaya alpacas (39 females and 26 males) distributed across ten fiber-color categories. The animals included in the categorical model ranged from 2.47 to 6.65 years of age, with a mean age of 4.37 years. Age distributions were comparable across fiber-color categories and collection batches, with no evidence of substantial imbalance. Due to its single-individual sample size, the LFX sample was excluded from categorical differential expression but retained for descriptive and continuous CIELAB analyses. The resulting categorical model comprised 64 samples from nine color categories and 25,405 genes after filtering. The design matrix was full rank, and no invalid dispersion estimates were detected. Among the included genes, only LOC140700246 failed to converge numerically and was automatically excluded from the differential-expression results (Table 1).

Table 1. Experimental design and primary inferential model.

Characteristic	Description
Animals	65 Huacaya alpacas; 39 females and 26 males
Nominal categories	10 categories; nine eligible for categorical inference
Categorical samples	64; LFX excluded because $n = 1$
Modeled genes	25,405
Main model	~ Batch + Sex + ColorCode
Sensitivity model	~ Sex + ColorCode
DEG criterion	Adjusted $p < 0.05$ and $\|\log2FC\| \geq 1$
Convergence	One non-convergent gene; no invalid dispersions

DEG, differentially expressed gene; LFX, light fawn intensity X; ColorCode, source metadata variable for fleece-color category; log2FC, log2 fold change.

Cook's distance filtering resulted in p values not being assigned to 140 genes (0.55% of the 25,405 modeled genes) in each of the eight primary contrasts against white fleece. Depending on the contrast, independent filtering excluded between 0 and 2369 additional low-information genes from multiple-testing adjustment. No samples were excluded on the basis of Cook's distance. Although 295 genes exhibited at least one observation above the Cook's distance diagnostic threshold, the highest sample-specific count was 60 genes, representing only 0.24% of the genes evaluated. Contrast- and sample-level Cook's distance diagnostics and independent-filtering summaries are provided in Supplementary Table S12.

Across all 65 libraries, 4.214 billion read pairs were generated and 4.153 billion were retained after fastp processing. Median retention per library was 98.52%. The median number of read pairs per library decreased slightly from 66.35 million before filtering to 65.34 million afterward, while median Q30 increased from 95.73% to 96.62%. STAR alignment yielded median uniquely mapped and total mapped read rates of 89.91% and 95.91%, respectively. FeatureCounts assigned 3.125 billion fragments to annotated genes, with a median assignment rate of 69.21% per library. Given the high mapping rate, the lower assignment rate likely reflects exon-based counting against the incomplete VicPac4 annotation, with many aligned fragments falling outside annotated exons. Other unassigned categories, including multimapping and ambiguous overlaps, are reported in Table S4.

3.2. CIELAB Revealed a Colorimetric Gradient with Overlap Among Categories

Complete L*, a*, and b* measurements were available for all 65 samples. Mean L* values described a lightness gradient from the lightest categories, LFX (88.40) and W

(84.29), to the darkest, B (14.66) and RB (13.32). Between-sample differences were also represented by a* and b*. However, phenotypic categories did not form completely distinct colorimetric groups. In leave-one-out nearest-centroid classification, 39 of 64 evaluable samples (60.9%) were assigned to their original category. LFX could not be evaluated because it contained only one individual. Thirty samples had negative silhouette coefficients, indicating colorimetric proximity to other categories, and three were identified as potential within-group outliers. Thus, nominal categories represented a general color gradient with overlap among groups. Original categories were retained without reassignment, and CIELAB measurements were used as complementary quantitative evidence (Figure 1A; Supplementary Figure S1).

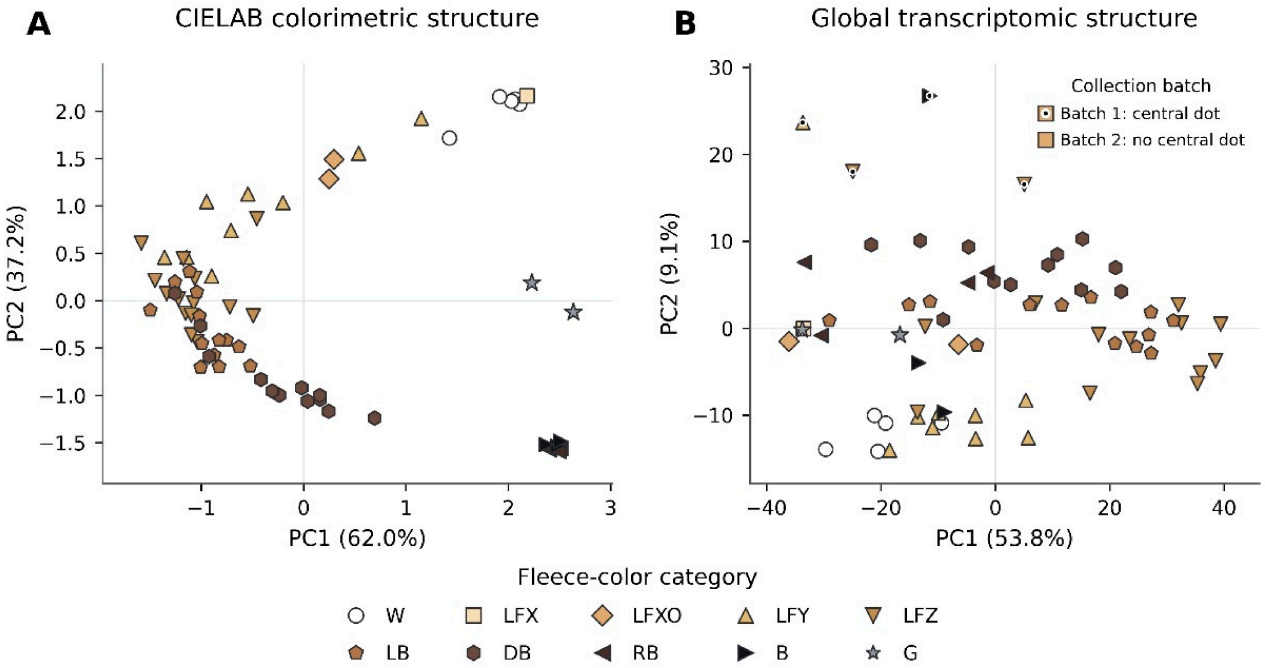


Figure 1. Colorimetric characterization and global transcriptomic structure of 65 alpaca skin samples. (**A**) Principal component analysis (PCA) based on standardized Commission Internationale de l'Éclairage Lab* (CIELAB) color-space parameters: L* (lightness), a* (green–red axis), and b* (blue–yellow). (**B**) PCA of blind variance-stabilized gene expression values using the 500 most variable genes. Each point represents one animal. Colors and marker shapes identify fleece-color categories consistently across both panels. In panel B, a contrasting central dot identifies collection batch 1, whereas markers without this dot indicate collection batch 2. Axis labels show the percentage of variance explained by each principal component (PC1, first principal component; PC2, second principal component). W, white; LFX, LFY, and LFZ, light fawn intensities X, Y, and Z, respectively; LFXO, dark variant of light fawn intensity X; LB, light brown; DB, dark brown; RB, reddish black; B, black; G, gray.

3.3. Global Transcriptomic Variation Was Associated with Fiber Color

PCA of VST-transformed expression showed that PC1 and PC2 explained 53.83% and 9.09% of transcriptomic variation among the 65 samples, respectively. Samples displayed structure associated with fiber-color category, together with differences related to collection batch, supporting inclusion of batch as a covariate in the main differential-expression model (Supplementary Figure S2). After adjustment for collection batch and sex, PC1 was significantly associated with L* (adjusted $p = 0.0037$), a* (adjusted $p = 4.47 \times 10^{-7}$), and b* (adjusted $p = 0.0012$). PC2 was also significantly associated with L* (adjusted $p = 1.16 \times 10^{-7}$) (Figure 1B). PC3 explained 3.84% of the variance. The LFX sample fell within the range of PC3 scores observed among the five white samples (Supplementary

Figure S3; Supplementary Table S8). This comparison is descriptive because LFX was represented by a single animal.

3.4. Pigmented-Fiber Categories Showed Extensive Transcriptomic Differences from White Fiber

The four contrasts shared by the 53- and 65-sample analyses differed markedly in the number of DEGs relative to white fleece (W). In the full dataset, with 64 animals included in the categorical model, we identified 4968 DEGs for LB versus W, 3993 for DB versus W, 1469 for LFY versus W, and 7607 for LFZ versus W. Across these contrasts, 81.1–87.7% of DEGs showed higher expression in the pigmented category than in white alpacas (Figure 2; Table 2).

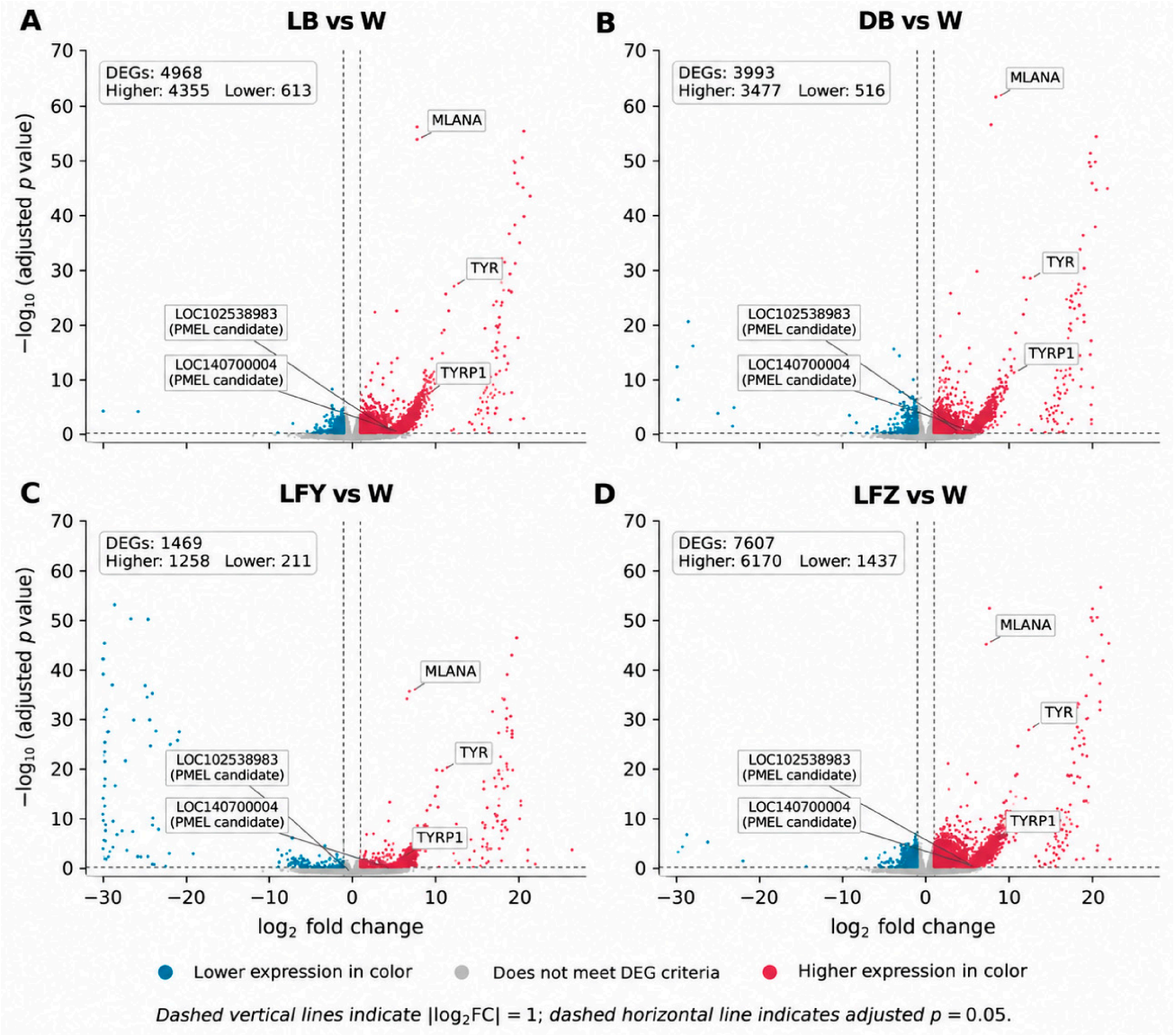


Figure 2. Differential expression in the four core contrasts against white fleece. Volcano plots showing (**A**) LB versus W, (**B**) DB versus W, (**C**) LFY versus W, and (**D**) LFZ versus W. Differentially expressed genes (DEGs) were defined by adjusted $p < 0.05$ and an absolute log2 fold change $|\log2FC| \geq 1$ in the model adjusted for collection batch and sex. Blue and red points denote lower and higher expression, respectively, in the pigmented category. Selected pigmentation genes and PMEL-compatible loci are connected to their corresponding points. W, white; LB, light brown; DB, dark brown; LFY and LFZ, light fawn intensities Y and Z, respectively.

Table 2. Differential expression in the four core contrasts against white fleece (W).

Contrast	n Color	n W	DEGs	Higher in Color	Lower in Color	Jaccard Without Batch
LB vs. W	13	5	4968	4355	613	0.920
DB vs. W	12	5	3993	3477	516	0.920
LFY vs. W	9	5	1469	1258	211	0.840
LFZ vs. W	14	5	7607	6170	1437	0.802

DEG, differentially expressed gene; W, white; LB, light brown; DB, dark brown; LFY and LFZ, light fawn intensities Y and Z, respectively. Positive and negative log2FC values indicate higher and lower expression, respectively, in the pigmented category. Jaccard values compare the main model with the model without collection-batch adjustment.

The black-versus-white comparison (B versus W) identified 1962 DEGs using the same selection criteria (adjusted $p < 0.05$ and $|\log2FC| \geq 1$). This comparison was considered exploratory because only three black alpacas were included. Complete gene-level results for all eight color-versus-white contrasts are provided in Supplementary Table S13. Results for the four main contrasts showed substantial agreement between models with and without adjustment for collection batch. Jaccard indices measuring overlap between DEG sets ranged from 0.802 to 0.920. Among genes with finite log2FC estimates in both models, 93.7–99.6% retained the same direction of expression change (Supplementary Table S14).

The 53-sample analysis also preserved the ranking of the four contrasts by DEG count, supporting the internal stability of the findings (Supplementary Table S15).

Among the secondary comparisons, LFZ versus LFY identified 3240 DEGs, of which 2996 showed higher expression and 244 showed lower expression in LFZ (Supplementary Figure S4).

Applying stricter absolute log2 fold-change cutoffs reduced the number of selected genes. At cutoffs of 1.5 and 2, respectively, DEG counts were 3508 and 2903 for LB versus W, 2951 and 2487 for DB versus W, 1368 and 1300 for LFY versus W, and 5141 and 3894 for LFZ versus W. The cutoff of 2 retained 51.2–88.5% of the genes selected at the original cutoff of 1 across the four main contrasts (Supplementary Table S6). Of the nine highlighted named genes, nine met the selection criteria in at least one main contrast at a cutoff of 1, seven at 1.5, and four at 2 (Supplementary Table S7.

3.5. Gene Expression Showed Consistent Associations with CIELAB Parameters

The colorimetric characteristics of the fleece-color categories are summarized in Table 3A. Continuous models separately evaluated associations between 25,405 genes and each CIELAB parameter. At a false discovery rate below 0.05, 5839 associations were identified for L*, 10,829 for a*, and 6697 for b*. Of these, 1125, 1388, and 537 associations, respectively, also had an absolute log2 expression change of at least 1 per standard-deviation increase in the evaluated parameter (Table 3B). Complete gene-level results are provided in Supplementary Table S16.

Negative coefficients predominated for L*, occurring in 4348 genes and indicating increased expression as fiber lightness decreased. Estimated effects were highly consistent across alternative model specifications. For L*, correlations between coefficients from the primary and sensitivity models were ≥0.962, calculated across genes with evaluable coefficients in both models. Across all three CIELAB parameters, more than 99.6% of genes that were significant in at least one of the compared models and had evaluable coefficients in both retained the same direction of association (Supplementary Figure S5).

Table 3. CIELAB characterization and continuous transcriptomic associations. (**A**) Colorimetric characterization. (**B**) Continuous transcriptomic associations.

(**A**)

Group	n	L* Centroid	L* Range	a* Centroid	b* Centroid	Agreement
W	5	84.29	75.80–87.41	0.63	9.39	0.400
LB	13	32.35	26.01–42.96	11.88	18.55	0.462
DB	12	24.70	18.69–38.78	10.30	13.28	0.750
RB	4	13.31	12.98–13.68	0.64	0.10	1.000
G	2	46.27	44.61–47.93	0.32	2.52	1.000
LFX	1	88.40	88.40–88.40	0.13	7.57	NE
LFXO	2	68.27	66.52–70.01	6.61	16.07	1.000
LFY	9	56.90	42.93–78.93	8.71	19.73	0.444
LFZ	14	38.99	29.91–55.54	11.60	20.57	0.571
B	3	14.66	14.05–15.66	0.67	0.11	0.667

(**B**)

Channel	n	Genes	FDR < 0.05	Positive	Negative	FDR and |effect| ≥ 1
a*	65	25,405	10,829	6883	3946	1388
b*	65	25,405	6697	3798	2899	537
L*	65	25,405	5839	1491	4348	1125

Agreement is the proportion assigned to the original category by leave-one-out nearest-centroid classification; NE, not evaluable because n = 1. W, white; LFX, LFY, and LFZ, light fawn intensities X, Y, and Z, respectively; LFXO, dark variant of light fawn intensity X; LB, light brown; DB, dark brown; RB, reddish black; B, black; G, gray. FDR, false discovery rate; L*, lightness; a*, green–red axis; b*, blue–yellow axis; effect, log2 fold change per one-standard-deviation increase in the corresponding CIELAB channel.

3.6. Integrated Evidence Prioritized Genes Related to Pigmentation, Melanosome Biology, and Skin Structure

Integration of categorical expression, CIELAB associations, and functional evidence identified an expanded panel of 20 candidate genes. It included 12 genes related to pigmentation and melanosome biology (*TYR*, *TYRP1*, *MITF*, *MLANA*, *DCT*, *GPR143*, *RAB27A*, *MFSD12*, *HPS4*, *MREG*, *EDNRB*, and *EDNRA*), 5 genes related to epidermal or fiber structure (*EPPK1*, *OVOL1*, *KRT86*, *KRTAP11-1*, and *KRT72*), and 3 prioritized loci (*LOC102538983*, *LOC140700004* and *LOC102543550*). Expression patterns for all 20 genes are presented in Figure 3, with supporting evidence provided in Table S17.

Among known candidates, *TYR* and *TYRP1* showed the most consistent evidence. Both were differentially expressed in all four main contrasts shared by the 53- and 65-sample analyses and had strong negative associations with L*, with standardized coefficients of $\beta = -2.02$ for *TYR* and $\beta = -3.00$ for *TYRP1*. Their expression therefore increased as fiber lightness decreased. *GPR143*, *EDNRB*, *MITF*, *MFSD12*, *HPS4*, and *MREG* also showed patterns consistent with higher expression in darker fibers (Figures 3 and 4; Table 4).

The complete evidence matrix for the expanded 20-gene panel, including the eight candidates not discussed individually, is provided in Table S17.

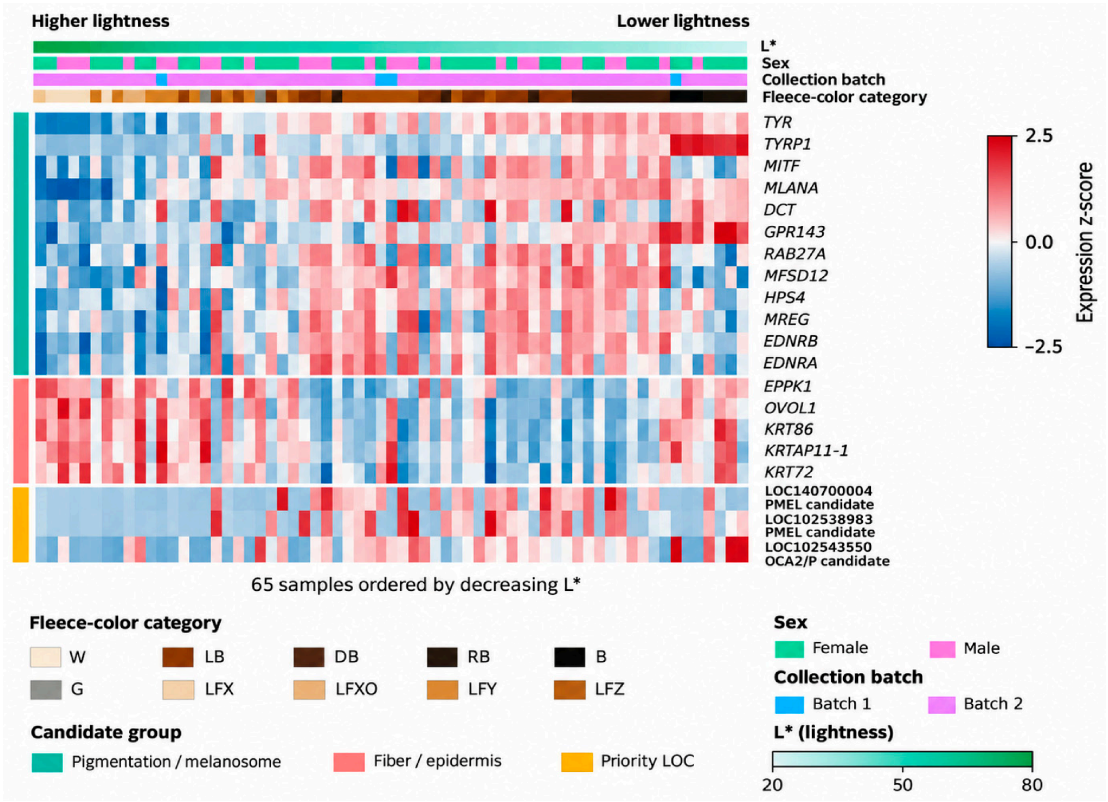


Figure 3. Expression pattern of the expanded candidate-gene panel. Heatmap of variance-stabilizing transformation (VST)-transformed expression standardized by gene across 65 samples. Columns represent samples ordered from higher to lower lightness (L*), and rows represent candidate genes. Top annotations indicate fiber-color category, sex, collection batch, and lightness. Category colors approximate the phenotypic fiber shades and do not replace instrumental CIELAB measurements. The blue-white-red scale represents lower, intermediate, and higher relative expression, respectively.

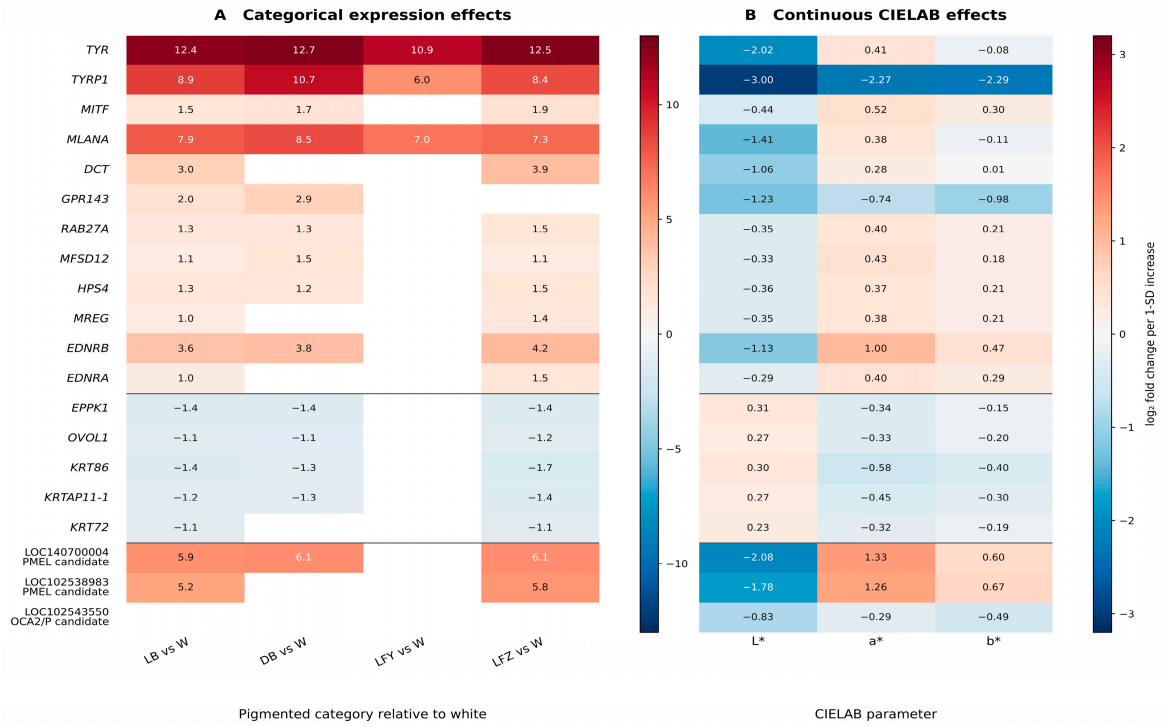


Figure 4. Integrated categorical and continuous evidence for candidate genes. (**A**) Log2 fold changes (log2FC) in the four core contrasts against white fleece. Blank cells indicate that no retained cat-

egorical differential-expression result was available in the candidate source table. (**B**) Log2 fold changes per one-standard-deviation increase in L* (lightness), a* (green–red axis), or b* (blue–yellow axis). Proposed premelanosome protein (PMEL) and OCA2 melanosomal transmembrane protein (OCA2/P) identities are interpretive labels and do not replace official VicPac4 locus (LOC) identifiers. W, white; LB, light brown; DB, dark brown; LFY and LFZ, light fawn intensities Y and Z, respectively.

Table 4. Prioritized candidate genes supported by categorical, CIELAB, and functional evidence.

Gene/Label	Identity	Calls vs. W	Max |log2FC|	β L*	Adjusted *p* L*	Robust	Tier
TYRP1	Known gene	8	10.75	−3.00	5.75×10^{-14}	Yes	P1
EDNRB	Known gene	6	4.26	−1.13	2.17×10^{-6}	Yes	P1
MITF	Known gene	7	1.97	−0.44	4.67×10^{-4}	Yes	P1
EDNRA	Known gene	3	1.54	−0.29	0.012	Yes	P1
MFSD12	Known gene	6	1.46	−0.33	0.003	Yes	P1
HPS4	Known gene	6	1.54	−0.36	9.07×10^{-4}	Yes	P1
MREG	Known gene	4	1.45	−0.35	0.007	Yes	P1
GPR143	Known gene	4	2.91	−1.23	9.87×10^{-14}	Yes	P1
TYR	Known gene	8	12.69	−2.02	3.96×10^{-27}	Yes	P1
PMEL candidate [LOC102538983]	LOC supported, moderate	5	5.79	−1.78	0.002	Yes	P1
PMEL candidate [LOC140700004]	LOC supported, high	6	6.11	−2.08	2.90×10^{-4}	Yes	P1
OCA2/P candidate [LOC102543550]	LOC supported, high	0	2.55	−0.83	0.029	No	P2

Calls vs. W: number of significant gene-by-analysis-set-by-contrast results across LB–W, DB–W, LFY–W, and LFZ–W in the 53-sample subset and the 65-sample analysis (64 animals in its categorical model), giving a maximum of eight calls. These overlapping datasets do not constitute independent replication. Max |log2FC|: largest absolute original, unshrunken coefficient among retained DEG results across both datasets and five contrasts, including the secondary LFZ–LFY comparison. Selection required adjusted $p < 0.05$ and $|\log2FC| \geq 1$. A gene may therefore have zero calls vs. W and a nonzero maximum derived from LFZ–LFY. LOC, provisional locus identifier; PMEL, premelanosome protein; OCA2/P, OCA2 melanosomal transmembrane protein; β L*, log2 fold change per one-standard-deviation increase in L*; adjusted P, Benjamini–Hochberg-adjusted P value; P1 and P2, integrated priority tiers; Robust, retention of significance and effect direction in the prespecified CIELAB sensitivity analyses.

3.7. Protein Evidence Supported Functional Identities for Three Priority Loci

Integration of native annotation, orthology relationships, similarity to curated proteins, and conserved domains supported proposed functional identities for three priority loci. *LOC140700004* and *LOC102538983* showed evidence consistent with the premelanosome protein *PMEL*, with high and moderate confidence, respectively. Both loci also showed differential expression among color categories and negative associations with L*, consistent with higher expression in darker fibers.

LOC102543550 showed high-confidence protein evidence consistent with the P protein encoded by *OCA2*. However, this locus was not differentially expressed in the four main contrasts against white fiber, and its association with L* was no longer robust after exclusion of the single LFX sample. It was therefore retained as a secondary candidate. These proposed identities are interpretive functional labels and do not constitute formal renaming of VicPac4 genes (Figure 4; Table 4; Table S11).

3.8. Functional Enrichment Revealed Pigmentation and Skin-Structure Patterns

Overrepresentation analysis showed that genes with higher expression in pigmented-fiber categories were enriched for terms related to pigmentation and melanosomes. In contrast, genes with lower expression were enriched for intermediate filament organization, epidermis development, keratin filament, and structural constituent of skin epidermis.

Selected terms were significant under the primary GO annotation profile and showed concordant patterns across the 53- and 65-sample analyses (Figure 5; Table S18).

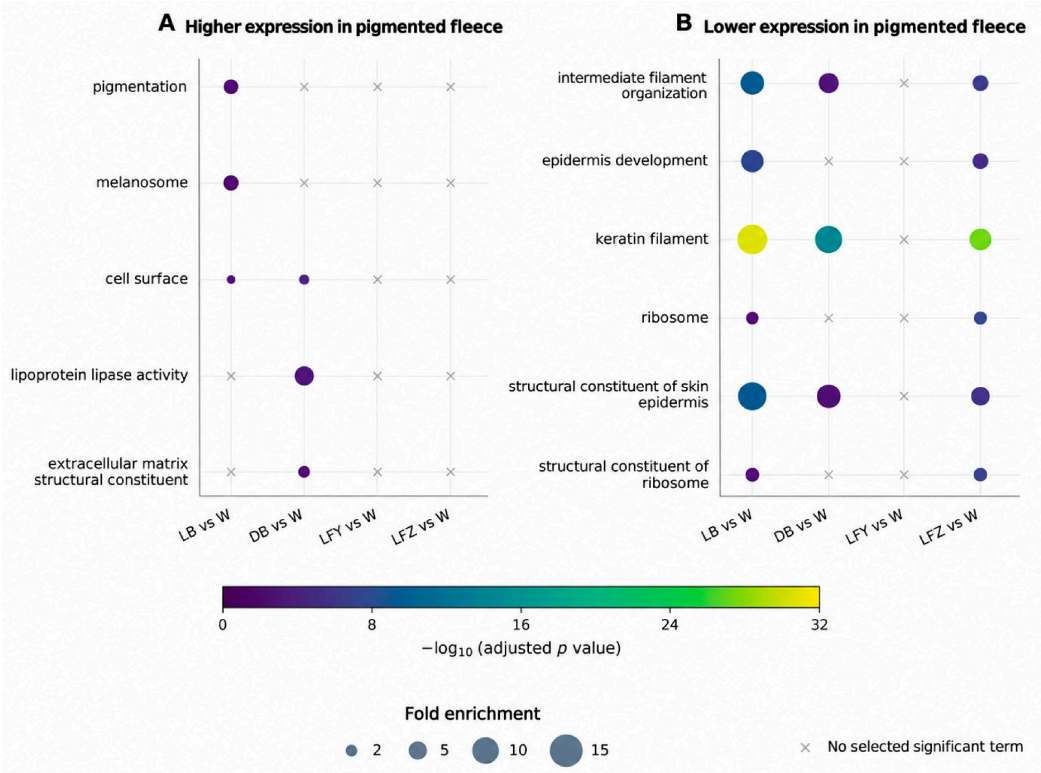


Figure 5. Representative Gene Ontology (GO) terms associated with fleece color, shown separately for genes with higher and lower expression in pigmented categories. Dot size represents fold enrichment, color represents adjusted statistical significance, and gray crosses indicate that no selected significant term was present for that contrast. Terms were selected from the primary annotation profile while considering stability between the 65- and 53-sample analyses. W, white; LB, light brown; DB, dark brown; LFY and LFZ, light fawn intensities Y and Z, respectively.

The complementary rank-based analysis identified 772 significant GO term–contrast associations across the eight primary comparisons. Within the prespecified pigmentation, melanosome, keratin, and epidermis categories, 44 significant associations representing 14 unique GO terms were detected. Pigmentation- and melanosome-related gene sets showed positive enrichment, whereas keratin-, epidermis-, and skin-development-related gene sets showed negative enrichment. These patterns were most consistent in the LB-versus-W, DB-versus-W, and LFZ-versus-W contrasts. Melanosome, pigment granule, and keratin filament were each significantly enriched in six contrasts, whereas no significant focal enrichment was detected for LFXO versus W (Supplementary Figure S6; Supplementary Table S19).

4. Discussion

4.1. Main Transcriptomic Findings Associated with Fiber Color

This study integrated phenotypic color categories, instrumental CIELAB measurements, and skin gene expression to examine variation associated with alpaca fiber color. Global transcriptomic variation was related to quantitative color dimensions, and the four main pigmented-fiber categories showed extensive expression differences from white fiber. Higher expression predominated in these contrasts, and genes known to participate in melanogenesis and melanosome biology were prioritized. Functional analysis also distin-

guished pigmentation-related processes from structural functions of the epidermis and keratin filaments.

Effect directions remained largely stable in sensitivity analyses, and the 53-sample subset preserved the relative ordering of the four shared contrasts by DEG count. These observations support the internal consistency of the results but do not constitute independent validation or establish causal relationships.

PCA showed that a substantial proportion of transcriptomic variation was associated with L*, a*, and b* (Figure 1; Table 3). This relationship is compatible with differences in the abundance or activity of pigmentation-related cells but cannot distinguish between these possibilities. Skin biopsies contain melanocytes, keratinocytes, fibroblasts, immune cells, blood vessels, and hair follicles at different stages [35,36]; consequently, expression profiles represent a mixture of cell types and biological processes.

Analysis of the colorimetric data also revealed overlap among several phenotypic fleece-color categories in CIELAB space. Fiber-color category and instrumental measurements should therefore be viewed as complementary sources of information: the former preserves the classification used in production settings, whereas L*, a*, and b* quantify the color of each sample [4].

4.2. Magnitude and Stability of Differences Among Color Categories

The magnitude of expression differences was not uniform among categories. LFZ versus W produced the largest number of DEGs, followed by LB, DB, and LFY versus W. This ordering should not be interpreted as a direct darkness scale. DEG counts depend not only on phenotypic distance between groups but also on sample size, within-category variability, biopsy cell composition, effect magnitude, and statistical power [37].

Despite these quantitative differences, positive log2FC values predominated in all four contrasts, indicating that most DEGs had higher expression in pigmented-fiber categories than in white fiber [19–22]. This pattern was shared by LB, DB, LFY, and LFZ, although each category had a distinct magnitude and gene composition.

Although eumelanin and pheomelanin provide a biological framework for interpreting fleece-color variation, their amounts and relative proportions were not measured in this study. Therefore, the observed gene-expression patterns cannot establish the pigment composition of individual fleece-color categories.

The model without collection batch remained highly concordant with the main model, particularly for LB and DB versus W. Concordance was somewhat lower for LFY and LFZ, indicating that batch adjustment affected the statistical classification of a larger proportion of genes in these groups, although the overall direction of effects remained stable.

Including collection batch in the main model was justified by systematic differences observed in the PCA. However, the marked imbalance in sample numbers between batches prevents complete separation of collection-time variation from related biological differences. Main conclusions were therefore based on the adjusted model, while the model without batch was used only to evaluate stability rather than to select the configuration yielding more favorable findings.

4.3. Candidate Genes Related to Melanogenesis and Melanosome Biology

Among prioritized genes, *TYR* and *TYRP1* stood out for the recurrence and magnitude of their expression differences and their negative associations with L*. Thus, expression of both genes tended to increase as fiber lightness decreased. *TYR* encodes an essential enzyme in the early stages of melanin synthesis [6,10], whereas *TYRP1* contributes to melanosome function and pigment production and stability [10,19]. Their identification is consistent with transcriptomic studies of pigmented skin in alpacas and other domestic

species, in which several melanogenesis components show higher expression in pigmented tissues [19–22]. Recent alpaca genetic evidence also supports TYRP1 as a coat-color candidate [17], while sheep transcriptomic and metabolomic findings implicate both TYR and TYRP1 in pigmentation-associated networks [23]. These cross-study observations provide biological context but do not independently validate expression differences in the animals analyzed here.

The integrated panel also included *MITF* [10], *EDNRB* [32], *EDNRA*, *MFSD12* [38], *GPR143*, *HPS4*, and *MREG* [11]. These genes contribute to different stages of melanocyte differentiation, pigment synthesis, and melanosome formation, maturation, or transport. *MITF* regulates melanocyte differentiation and function and controls the expression of several melanogenic genes. *EDNRB* contributes to melanocyte development and maintenance [39], whereas *GPR143*, *HPS4*, *and MREG* are associated with the organization and transport of pigment organelles.

Agreement between categorical contrasts and continuous CIELAB models strengthened the prioritization of these genes. Nevertheless, the results do not show that each gene independently determines a specific color category. The observed patterns more likely reflect coordinated processes involving melanocyte abundance, melanogenic activity, melanosome transport, and pigment transfer to keratinocytes.

The association with L* was especially informative because it linked gene expression to an individual measure of lightness. However, a* and b* also showed numerous associations, indicating that transcriptomic variation was not limited to a light-versus-dark distinction. These parameters represent chromatic dimensions distinct from lightness and should be interpreted in relation to one another and to the color classification used at Pacomarca.

4.4. Functional Interpretation of Priority Loci

The annotation strategy retained official VicPac4 identifiers and added an interpretive layer based on transcript and protein information, orthology relationships, similarity to curated Swiss-Prot proteins, and conserved domains. Under this framework, *LOC140700004* and *LOC102538983* showed evidence consistent with the premelanosome protein PMEL.

PMEL contributes to the formation of the melanosomal fibrillar matrix on which melanin is deposited [40]. The expression differences among color categories and negative associations of these loci with L* are therefore consistent with a possible pigmentation-related function. *LOC140700004* and *LOC102538983* nevertheless had different confidence levels based on the quality and concordance of available evidence. This difference supports retaining them as independent loci rather than merging them or replacing their official identifiers.

LOC102543550 showed evidence consistent with the *OCA2*-encoded P protein, which is involved in melanosome function [11,41]. However, it was not differentially expressed in the four main contrasts against white fiber, and its association with L* lost robustness when the single LFX sample was excluded. It therefore remains a secondary candidate requiring additional confirmation.

More generally, even strong protein similarity does not by itself establish definitive gene identity. Proposed labels facilitate biological interpretation but do not replace the official VicPac4 annotation. Confirmation will require additional evidence from conserved genomic position (synteny), gene structure, transcribed isoforms, and, where possible, experimental function.

4.5. Pigmentation- and Skin-Structure-Related Processes

Genes with higher expression in pigmented-fiber categories showed overrepresentation of GO terms related to pigmentation and melanosomes. This provided a functional

perspective complementary to individual genes such as *TYR*, *TYRP1*, *MITF*, and *GPR143*. The presence of these terms in both the 53- and 65-sample analyses showed that the overall interpretation remained stable when the analytical set was changed. Because the two analyses share samples and procedures, however, this concordance represents internal stability rather than independent functional validation.

Conversely, genes with lower expression in the pigmented-fiber categories were enriched for functions related to intermediate filament organization, epidermis development, keratin filaments, and structural components of the skin epidermis. Keratins are components of the epithelial cell cytoskeleton and, in addition to providing mechanical support, contribute to cellular organization, stress responses, and other regulatory processes in epithelia [42]. Therefore, the observed pattern suggests that transcriptomic differences among fiber-color categories were not limited to melanin synthesis but were also accompanied by variation in processes related to skin and hair follicle organization.

The absence of histological information and cell-type-specific expression profiles prevents us from determining whether these signals arose from expression changes within the same cells, differences in the proportions of cell types represented in the biopsies, or variation in hair follicle cycle stage. This caution is particularly important because melanocytes, keratinocytes, fibroblasts, and whole-skin samples exhibit clearly distinct expression profiles [35]. Consequently, these findings should be interpreted as transcriptomic associations observed in whole-skin tissue rather than as evidence of direct effects of pigmentation on skin structure.

Taken together, the agreement between the threshold-based overrepresentation analysis and the complementary rank-based analysis supports two main biological components underlying the observed transcriptomic differences: increased representation of pigmentation- and melanosome-related processes in pigmented-fiber categories and coordinated changes in pathways related to epidermal structure, keratinization, and keratin filaments. These findings suggest that fiber-color variation involves not only pigment production and melanosome biology but also broader differences in the structural and cellular context of the skin.

4.6. Strengths, Limitations, and Perspectives

Key strengths include analysis of 65 individual biological samples, use of a recent genome assembly, inclusion of sex and collection batch in the statistical model, and explicit evaluation of result stability. Individual CIELAB measurements complemented phenotypic categories with quantitative color descriptions. Contrast-specific statistical universes strengthened functional analysis, while traceable annotation enabled examination of uncharacterized loci without replacing their official identifiers. Sequencing, alignment, and fragment-assignment metrics also supported the overall technical quality of the dataset.

The study also has limitations. First, animal numbers differed among categories: RB, G, LFXO, and B had fewer than five individuals, while LFX had only one. LFX was therefore excluded from the categorical model, and contrasts involving small groups were treated as exploratory. Second, collection batches were markedly imbalanced, with four samples in batch 1 and 61 in batch 2. This distribution limits the ability to separate batch-associated variation from biological differences, even after statistical adjustment. Small and unequal category sizes can reduce sensitivity to detect expression differences and increase uncertainty, but its contribution cannot be isolated from other biological and technical sources of variation in this dataset. Third, skin biopsies contained mixtures of cell types, and neither matched histological information nor single-cell expression profiles were available. Fourth, the VicPac4 annotation remains incomplete; therefore, the proposed functional identities of LOC should be regarded as interpretive rather than official annotations. Fifth,

although age distributions were comparable across fiber-color categories, age was not included as a covariate in the differential-expression models.

Pedigree records and genomic relatedness estimates were not available, so relatedness could not be evaluated or modeled within or across fleece-color categories, including the five categories emphasized in the main comparisons. Residual age-associated variation also cannot be excluded.

These limitations affect statistical inference as well as causal interpretation. Agreement across models sharing the same animals is an internal stability check and cannot rule out confounding by relatedness. Future studies should increase and balance animal numbers across fiber-color categories, sex, age, and collection batch and should incorporate pedigree- or genomic-based relationship estimates, histological analyses, and methods capable of distinguishing skin cell types. Independent RT-qPCR validation was not performed in this study. TYR and TYRP1 are priorities for confirmation using gene-specific assays and reference genes whose expression stability has been verified in alpaca skin.

5. Conclusions

In this study, we found that natural variation in alpaca fiber color was associated with broad and consistent differences in skin gene expression. Both color-category comparisons and continuous CIELAB models highlighted genes related to melanogenesis and melanosome biology, including *TYR*, *TYRP1*, *MITF*, *EDNRB*, *MFSD12*, and *GPR143*. *LOC140700004* and *LOC102538983* showed evidence consistent with *PMEL*, whereas *LOC102543550* showed similarity to *OCA2/P* but weaker transcriptomic support. These identities are interpretive proposals rather than changes to the official VicPac4 annotation.

Genes with higher expression in pigmented-fiber categories were primarily associated with pigmentation and melanosome processes, whereas genes with lower expression were enriched for epidermal, keratin, and intermediate-filament processes. The prioritized genes and processes provide a basis for further research, but their interpretation must account for unequal category sizes, strong batch imbalance, heterogeneous skin cell composition, unmodeled relatedness, and the absence of independent experimental validation. Confirmation by RT-qPCR, functional studies, and additional balanced cohorts with documented relatedness is therefore required.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ani16223567/s1: The supporting information includes a supplementary table workbook, six supplementary figures, and three separate CSV files containing the complete results for Tables S13, S16, and S19. Table S1: Characteristics of the 64 alpacas included in the categorical differential-expression model; Table S2: Sample metadata, instrumental CIELAB measurements, and colorimetric classification diagnostics; Table S3: Per-library fastp processing and sequencing-read quality metrics; Table S4: STAR alignment and featureCounts quantification metrics; Table S5: Complete categorical differential-expression contrast plan; Table S6: Sensitivity of gene selection to absolute log2 fold-change cutoffs of 1, 1.5, and 2; Table S7: Sensitivity of highlighted named genes and LOC candidates to the same fold-change cutoffs. Table S8: Eigenvalues and explained variance for the 65-sample transcriptomic PCA. Table S9: Functional-evidence levels and annotation profiles used for Gene Ontology analysis; Table S10: Criteria used for integrated candidate-gene prioritization; Table S11: Evidence supporting the proposed functional identities of priority LOC loci; Table S12: Cook's distance and independent-filtering diagnostics for categorical differential-expression contrasts; Table S13: Complete gene-level differential-expression results for the eight primary color-versus-white contrasts; Table S14: Concordance between the primary model and the model without collection-batch adjustment; Table S15: Internal stability assessment in the 53-sample subset; Table S16: Complete gene-level associations with L*, a*, and b*; Table S17: Complete evidence matrix for the expanded 20-gene candidate panel; Table S18: Gene Ontology overrepresentation results; Table S19: Complete rank-based Gene Ontology enrichment results across the eight primary color-versus-white

contrasts; Figure S1: Agreement between phenotypic fleece-color categories and instrumental CIELAB measurements; Figure S2: Transcriptomic principal component analysis stratified by sex and collection batch; Figure S3: PC1-PC3 and PC2-PC3 projections of the transcriptomic PCA, with the LFX sample labeled; Figure S4: Secondary LFZ-versus-LFY differential-expression contrast; Figure S5: Robustness of gene-expression associations with CIELAB parameters; Figure S6: Rank-based Gene Ontology enrichment across the eight primary color-versus-white contrasts.

Author Contributions: A.J.-E. and L.P.: Methodology, investigation, formal analysis, data curation, writing—original draft; F.A., F.C., A.C., G.G., D.O., S.R.P.-M., A.Y. and R.F.: review, editing, supervision; C.V.-L. and O.A.: Conceptualization, supervision, administration, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CONCYTEC through the PROCIENCIA program under call E041-2023-01, contract/agreement PE501082695-2023 and the Universidad de San Martín de Porres (Project: E10012023030).

Institutional Review Board Statement: The animal study protocol was approved through expedited review by the Institutional Ethics Committee for the Use of Animals of Universidad Peruana Cayetano Heredia (SIDISI 24002; certificate CIEA-008-03-24; approval issued on 23 February 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors thank the Government of Peru, specifically the Instituto Nacional de Innovación Agraria (INIA), for providing access to *Vicugna pacos*. This research was conducted under Access Contract to Genetic Resources No. 007-2025-MIDAGRI-INIA/DGIA. We explicitly recognize Peru as the country of origin of the genetic resources used in this study.

Conflicts of Interest: The authors have read the journal's policy and have the following competing interests: the co-authors (A.C., A.Y.) are employees of Inca Tops S.A company which is participating in the project as an associated entity and provided the animals. The other authors have no competing interests.

References

1. Quispe, E.C.; Rodríguez, T.C.; Iñiguez, L.R.; Mueller, J.P. Producción de Fibra de Alpaca, Llama, Vicuña y Guanaco En Sudamérica. *Anim. Genet. Resour. Inf.* **2009**, *45*, 1–14. [CrossRef]
2. Mueller, J.P.; Rigalt, F.; Lamas, H.; Sacchero, D.M.; Cancino, A.K.; Wurzinger, M. Fibre Quality of South American Camelids in Argentina: A Review. *Anim. Genet. Resour.* **2015**, *56*, 97–109. [CrossRef]
3. Anello, M.; Daverio, M.S.; Di Rocco, F. Genetics of Coat Color and Fiber Production Traits in Llamas and Alpacas. *Anim. Front.* **2022**, *12*, 78–86. [CrossRef] [PubMed]
4. Cruz, A.; Yucra, A.; Gutiérrez, G.A.; Burgos, A.; Morante, R.; Gutiérrez, J.P.; Cervantes, I.; Wurzinger, M. Colorimetry Analysis of Coat Color and Its Relationship with Fiber Traits in Alpacas. *Animal* **2021**, *15*, 100219. [CrossRef] [PubMed]
5. Morante, R.; Goyache, F.; Burgos, A.; Cervantes, I.; Pérez-Cabal, M.A.; Gutiérrez, J.P. Genetic Improvement for Alpaca Fibre Production in the Peruvian Altiplano: The Pacomarca Experience. *Anim. Genet. Resour. Inf.* **2009**, *45*, 37–43. [CrossRef]
6. Ito, S.; Wakamatsu, K. Chemistry of Mixed Melanogenesis—Pivotal Roles of Dopaquinone†. *Photochem. Photobiol.* **2008**, *84*, 582–592. [CrossRef] [PubMed]
7. Cieslak, M.; Reissmann, M.; Hofreiter, M.; Ludwig, A. Colours of Domestication. *Biol. Rev.* **2011**, *86*, 885–899. [CrossRef] [PubMed]
8. Kaelin, C.B.; Barsh, G.S. Genetics of Pigmentation in Dogs and Cats. *Annu. Rev. Anim. Biosci.* **2013**, *1*, 125–156. [CrossRef] [PubMed]
9. Barsh, G.; Cotsarelis, G. How Hair Gets Its Pigment. *Cell* **2007**, *130*, 779–781. [CrossRef] [PubMed]
10. D'Mello, S.A.N.; Finlay, G.J.; Baguley, B.C.; Askarian-Amiri, M.E. Signaling Pathways in Melanogenesis. *Int. J. Mol. Sci.* **2016**, *17*, 1144. [CrossRef]
11. Le, L.; Sirés-Campos, J.; Raposo, G.; Delevoye, C.; Marks, M.S. Melanosome Biogenesis in the Pigmentation of Mammalian Skin. *Integr. Comp. Biol.* **2021**, *61*, 1517–1545. [CrossRef] [PubMed]

12. Beyers, W.C.; Detry, A.M.; Di Pietro, S.M. OCA7 Is a Melanosome Membrane Protein That Defines Pigmentation by Regulating Early Stages of Melanosome Biogenesis. *J. Biol. Chem.* **2022**, *298*, 102669. [CrossRef] [PubMed]
13. Powell, A.J.; Moss, M.J.; Tree, L.T.; Roeder, B.L.; Carleton, C.L.; Campbell, E.; Kooyman, D.L. Characterization of the Effect of Melanocortin 1 Receptor, a Member of the Hair Color Genetic Locus, in Alpaca (*Lama pacos*) Fleece Color Differentiation. *Small Rumin. Res.* **2008**, *79*, 183–187. [CrossRef]
14. Chandramohan, B.; Renieri, C.; La Manna, V.; La Terza, A. The Alpaca Agouti Gene: Genomic Locus, Transcripts and Causative Mutations of Eumelanic and Pheomelanic Coat Color. *Gene* **2013**, *521*, 303–310. [CrossRef] [PubMed]
15. Chandramohan, B.; Renieri, C.; La Manna, V.; La Terza, A. The Alpaca Melanocortin 1 Receptor: Gene Mutations, Transcripts, and Relative Levels of Expression in Ventral Skin Biopsies. *Sci. World J.* **2015**, *2015*, 265751. [CrossRef] [PubMed]
16. Tian, X.; Meng, X.; Wang, L.; Song, Y.; Zhang, D.; Ji, Y.; Li, X.; Dong, C. Molecular Cloning, mRNA Expression and Tissue Distribution Analysis of *Slc7a11* Gene in Alpaca (*Lama paco*) Skins Associated with Different Coat Colors. *Gene* **2015**, *555*, 88–94. [CrossRef] [PubMed]
17. Melo-Rojas, C.; Bravo-Matheus, P.W.; Araoz, C.A.; Zapata-Coacalla, C. Identification of Polymorphisms in TYRP1, DCT and RAB38 Genes and Their Association with Coat Color in Alpacas. *Front. Anim. Sci.* **2023**, *4*, 1236582. [CrossRef]
18. Tan, K.; Adeniyi, O.O.; Letko, A.; Manz, E.; Wagner, H.W.; Zanolari, P.; More, M.; Gutierrez-Reynoso, G.A.; Drögemüller, C.; Lühken, G. Genome-Wide SNP-Based Genomic Diversity and Population Structure Analysis in Alpaca Populations from Europe and Peru. *G3 Genes Genomes Genet.* **2026**, *16*, jkag157. [CrossRef] [PubMed]
19. Fan, R.; Xie, J.; Bai, J.; Wang, H.; Tian, X.; Bai, R.; Jia, X.; Yang, L.; Song, Y.; Herrid, M.; et al. Skin Transcriptome Profiles Associated with Coat Color in Sheep. *BMC Genom.* **2013**, *14*, 389. [CrossRef] [PubMed]
20. Peng, Y.; Wang, Y.; Wang, R.; Geng, L.; Ma, R.; Zhang, C.; Liu, Z.; Gong, Y.; Li, J.; Li, X. Exploring Differentially Expressed Genes Associated with Coat Color in Goat Skin Using RNA-Seq. *Can. J. Anim. Sci.* **2019**, *99*, 357–366. [CrossRef]
21. Bhat, B.; Yaseen, M.; Singh, A.; Ahmad, S.M.; Ganai, N.A. Identification of Potential Key Genes and Pathways Associated with the Pashmina Fiber Initiation Using RNA-Seq and Integrated Bioinformatics Analysis. *Sci. Rep.* **2021**, *11*, 1766. [CrossRef] [PubMed]
22. Cransberg, R. Insights into the Alpaca Skin Transcriptome in Relation to Fibre Colour. Ph.D. Thesis, Curtin University, Perth, Australia, 2017.
23. Zhang, M.; Xu, X.; Chen, Y.; Wei, C.; Zhan, S.; Cao, J.; Guo, J.; Dai, D.; Wang, L.; Zhong, T.; et al. Transcriptomic and Metabolomic Analyses Reveal Molecular Regulatory Networks for Pigmentation Deposition in Sheep. *Int. J. Mol. Sci.* **2024**, *25*, 8248. [CrossRef] [PubMed]
24. CIE TC 1-85 International Commission on Illumination (CIE). *CIE 015:2018 Colorimetry*, 4th ed.; CIE Central Bureau: Vienna, Austria, 2018. [CrossRef]
25. Cerna, M.N.M.; Stroupe, S.; Kolb, E.; Jevit, M.; Li, S.; Raudsepp, T.; Davis, B.W. Alpaca Chromosome-Level Reference Genome VicPac4 Reveals a Novel NOR-Associated Satellite Specific to South American Camelids. *Genomics* **2026**, *118*, 111212. [CrossRef] [PubMed]
26. Cruz, A. Parámetros Genéticos de Caracteres Funcionales y Secundarios en Alpacas. Doctoral Thesis, Universidad Complutense de Madrid, Madrid, Spain, 2017.
27. *NTP 231.301*; Fibra de Alpaca Clasificada: Definiciones, Clasificación por Grupo de Calidades, Requisitos y Rotulado. Comisión de Normalización y de Fiscalización de Barreras Comerciales No Arancelarias (INDECOPI): Lima, Peru, 2014.
28. Mendoza, A.Y. Parámetros Genéticos de los Valores CIELAB del Vellón en Alpaca de la Estación Científica Pacomarca, Puno. Master's Thesis, Universidad Nacional Agraria La Molina, Lima, Peru, 2025.
29. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*; Babraham Bioinformatics: Cambridge, UK, 2010.
30. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]
31. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef] [PubMed]
32. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [CrossRef] [PubMed]
33. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef] [PubMed]
34. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
35. Reemann, P.; Reimann, E.; Ilmjärv, S.; Porosaar, O.; Silm, H.; Jaks, V.; Vasar, E.; Kingo, K.; Kõks, S. Melanocytes in the Skin—Comparative Whole Transcriptome Analysis of Main Skin Cell Types. *PLoS ONE* **2014**, *9*, e115717. [CrossRef]
36. Gerber, P.A.; Buhren, B.A.; Schrumpf, H.; Homey, B.; Zlotnik, A.; Hevezi, P. The Top Skin-Associated Genes: A Comparative Analysis of Human and Mouse Skin Transcriptomes. *Biol. Chem.* **2014**, *395*, 577–591. [CrossRef] [PubMed]
37. Stark, R.; Grzelak, M.; Hadfield, J. RNA Sequencing: The Teenage Years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [CrossRef] [PubMed]

38. Adelmann, C.H.; Traunbauer, A.K.; Chen, B.; Condon, K.J.; Chan, S.H.; Kunchok, T.; Lewis, C.A.; Sabatini, D.M. MFSD12 Mediates the Import of Cysteine into Melanosomes and Lysosomes. *Nature* **2020**, *588*, 699–704. [CrossRef] [PubMed]
39. Saldana-Caboverde, A.; Kos, L. Roles of Endothelin Signaling in Melanocyte Development and Melanoma: Endothelins in Melanocyte Development and Melanoma. *Pigment Cell Melanoma Res.* **2010**, *23*, 160–170. [CrossRef] [PubMed]
40. Bissig, C.; Rochin, L.; Van Niel, G. PMEL Amyloid Fibril Formation: The Bright Steps of Pigmentation. *Int. J. Mol. Sci.* **2016**, *17*, 1438. [CrossRef] [PubMed]
41. Bellono, N.W.; Escobar, I.E.; Lefkovith, A.J.; Marks, M.S.; Oancea, E. An Intracellular Anion Channel Critical for Pigmentation. *elife* **2014**, *3*, e04543. [CrossRef] [PubMed]
42. Pan, X.; Hobbs, R.P.; Coulombe, P.A. The Expanding Significance of Keratin Intermediate Filaments in Normal and Diseased Epithelia. *Curr. Opin. Cell Biol.* **2013**, *25*, 47–56. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.