



# Transcriptome expression profiling of fur color formation in domestic rabbits using Solexa sequencing

L.Z. Qin<sup>1</sup>, W.Z. Wang<sup>1</sup>, L.J. Shi<sup>1</sup>, X.Y. Wan<sup>1</sup>, X.R. Yan<sup>1</sup>, Q.Q. Weng<sup>2</sup> and X.S. Wu<sup>1</sup>

<sup>1</sup>College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu, China

<sup>2</sup>Yuyao Xinnong Rabbit Co., Ltd., Zhejiang, Yuyao, China

Corresponding author: X.S. Wu

E-mail: xswu@yzu.edu.cn

Genet. Mol. Res. 15 (2): gmr.15027413

Received August 11, 2015

Accepted November 17, 2015

Published April 25, 2016

DOI <http://dx.doi.org/10.4238/gmr.15027413>

**ABSTRACT.** Fur color is an important, genetically determined characteristic of domestic rabbits, and rabbit furs are of great economic value. To investigate the molecular genetics associated with fur color determination in domestic rabbits, we used Solexa-sequencing technology to probe gene expression in dorsal skin tissues sampled from full-sibling Rex rabbits of different colors. The number of expressed genes in each sample was approximately 14,700. Among the top 30 genes and transcription factors with the highest reads per kilobase per million values, the elongation factor- $\alpha$  1 gene was highly expressed in all samples, as were genes of the ribosomal protein and keratin gene families. Compared with the chinchilla (C) Rex rabbit control sample, the numbers of genes in the black (B) and white (W) rabbit samples were 1809 and 460, respectively, and the number of common differentially expressed genes was 257. Clustering analysis of these 257 genes revealed that 32 were up-regulated in sample B and down-

regulated in sample W. Of these 32 genes, we identified some that are related to fur formation, including Tyrosinase-related protein 1 (*TYRPI*) and Tyrosinase (*TYR*), as well as genes with unknown functions. Quantitative real-time polymerase chain reaction was used to verify the expression patterns of those genes. The findings are expected to provide reference for the further study of fur color formation in rabbits.

**Key words:** Fur color; Rex rabbit skin; Solexa-sequencing technology

## INTRODUCTION

Rex rabbit furs are favored by consumers for their inherent aesthetic appeal, lightness, softness, and heat-retaining properties. Color, style, and material are three essential elements involved in the design of garments, with color attracting the attention of consumers first. Fur color is a genetic marker, which has utility and research value in mammals (Ginsburg, 1944; Bennett and Lamoreux, 2003; Sturm, 2009). The process of color formation in mammals is complex, and is regulated and controlled by diverse, yet related, genes, and can also be affected by the environment. Bennett and Lamoreux (2003) identified approximately 127 genes that are involved in different pathways controlling fur color formation in mice. Diverse fur colors in mammals, ranging from white to black, are mainly determined by the relative quantities of eumelanin and pheomelanin produced by melanocytes and the distribution of these pigments in the body (Cone et al., 1996; Hartmeyer et al., 1997). Chang (2009) summarized findings of previous studies on this mechanism (Cooksey et al., 1997; Schallreuter et al., 2008), and generated a diagram of the melanin synthesis process.

Genetic data from rabbits indicate that the genes affecting fur color are involved in at least 10 systems or locuses, and that specific colors are expressed through interactions between allelic genes and genes at different sites. Evidence suggests that many genes, for example Melanocortin receptor 1 (*Mclr*), Microphthalmia-associated transcription factor (*Mitf*), *Kit*, Agouti signaling protein (*ASIP*), Melanophilin (*MLPH*), and Premelanosome protein 17 (*Pmel-17*), are involved in the production of melanin and therefore, are likely to be involved in the determination of fur color in mammals (Kijas et al., 1998; Marklund et al., 1999; Philipp et al., 2005; Brunberg et al., 2006; Mundy and Kelly, 2006; Bauer et al., 2009). However, many unknown genes remain to be identified.

In our study, Solexa transcriptome sequencing was employed using tissue samples obtained from the back skin of Rex rabbit siblings with different coat colors to screen for genes involved in fur color determination. Genes related to melanin metabolism and color formation were identified through GO functional annotation and KEGG pathway analysis, and qRT-PCR was then used for verification to determine the main genes that control fur color in rabbits.

## MATERIAL AND METHODS

### Sample collection and RNA extraction

The Rex rabbits used for this study were purchased from Yuyao Xinnong Rabbit Co., Ltd. (Zhejiang, China). All animals were healthy, were born in the same nest by the same mother, and were raised under the same conditions for 42 days (to the first molt) to minimize

the impact of genetic background and environment on fur color. The fur colors of the rabbits were characterized as black (B), white (W), and chinchilla (C) (Figure 1). For anesthesia, 5 mL 0.5% sodium pentobarbital solution was injected into the ear vein of the rabbits. After anesthesia, a 1-cm<sup>2</sup> skin tissue sample was obtained from the backs of the animals, placed immediately in liquid nitrogen, and preserved at -80°C until use. An iodine solution was smeared on the resultant lesion to prevent bacterial infection. Total RNA was extracted from the back skin of the Rex rabbits using conventional TRIzol extraction.

### **Transcriptome sequencing and information analysis**

After total RNA was extracted and treated with DNase I, oligo(dT)-magnetic beads were used to isolate mRNA, which was resuspended in fragmentation buffer. The mRNA was then fragmented into short fragments, and cDNA was synthesized using the mRNA fragments as templates. Short fragments were purified and resolved with Elution buffer for end reparation and single nucleotide A (adenine) addition. Next, the short fragments were connected with adapters. During the quality control (QC) steps, the Agilent 2100 Bioanalyzer (Agilent, USA) and ABI StepOnePlus Real-Time PCR System (Applied Biosystems, USA) were used for the quantification and qualification of the sample library. Then, the library was sequenced using IlluminaHiSeq™ 2000 (Illumina, USA).

Primary sequencing data produced by IlluminaHiSeq™ 2000, termed “raw reads”, were subjected to QC analysis to determine whether a resequencing step was needed. After QC, raw reads were filtered into clean reads, which were then aligned to the reference sequences with SOAPaligner/SOAP2, part of the Short Oligonucleotide Analysis Package (SOAP) (Li et al., 2009).

### **Gene expression levels**

The number of gene reads in the unique BLAST analysis and the total number of reads of the reference sequence in the BLAST analysis were used to calculate the gene expression levels, which were calculated using the reads per kilobase per million reads (RPKM) method as follows:  $RPKM = 10^9 * C / (N * L)$ .

### **Screening of differentially expressed genes (DEGs)**

This analysis identified genes that have different expression levels among samples. Regarding the significance of digital gene expression profiles (Audic and Claverie, 1997), the absolute value of the  $\text{Log}_2\text{Ratio} \geq 1$  and false-discovery rate (FDR)  $\leq 0.001$  (Benjamini and Yekutieli, 2001) was used as the threshold to judge the significance of differences in gene expression. More stringent criteria with smaller FDRs and larger fold-change values can also be used to identify DEGs.

### **Gene ontology (GO) and pathway-enrichment analysis of DEGs**

GO is an international standard gene functional classification system that describes three ontologies: molecular function, cellular component, and biological process. GO-enrichment analysis provided all GO terms that were significantly enriched in the list of DEGs in comparison to a particular genome background, and the DEGs that correspond to

specific biological functions were filtered. This method firstly mapped all DEGs to GO terms in the database (<http://www.geneontology.org/>) and calculated gene numbers for every term. A hypergeometric test was then used to identify any significantly enriched GO terms in the input DEGs. The calculated P value was then corrected using the Bonferroni correction, and a corrected P value of  $\leq 0.05$  was used as a threshold. GO terms that fulfilled this condition were defined as significantly enriched DEGs. This analysis allowed recognition of the main biological functions that comprise the DEGs. Thus, the main biological functions of the DEGs identified in our study were determined by GO-enrichment analysis.

Genes usually interact with each other to function in biological processes. Therefore, pathway-based analyses further described the biological functions of the DEGs. KEGG (Kanehisa et al., 2008), which is the major public pathway-related database, was used to perform pathway-enrichment analysis of DEGs.

### **Gene structure refinement and novel transcript prediction**

For model species such as humans, gene annotation of the genome is relatively complete; however, for other species that have not been fully studied, we can refine the gene structure with additional reads to further complete the gene annotation. We assembled transcripts from the sequencing reads using Cufflink (Roberts et al., 2011). Through the comparison of assembled transcripts and gene annotation from reference information, it is possible to build assembled transcripts that extend 5' and/or 3' past the end of the annotated gene, permitting refinement of the gene structure. To discover novel transcribed regions, we compared our assembled transcripts and annotated genomic transcripts from reference sequences. To be reported as a novel transcript, an assembled transcript must meet three requirements: the transcript must be at least 200 bp away from the annotated gene, the transcript must be more than 180 bp in length, and the sequencing depth must be no less than 2. To study the function of a novel transcript, we used the Support Vector Machine-based classifier, Coding Potential Calculator, to assess the protein-coding potential of the RNAs.

### **Alternative splicing**

Although alternative splicing is universal in eukaryotes, the number of genes that undergo alternative splicing is likely underestimated. Seven types of alternative splicing have been described: exon skipping, intron retention, alternative 5'-splice site, alternative 3'-splice site, alternative first exon, alternative last exon, and mutually exclusive exon. We used the splice junctions tested by SOAP splice (Huang et al., 2011) to identify alternative splicing events occurring in the transcript in order to determine the method of splicing used.

This program calculates the likelihood of each genotype at each site based on the alignment of short reads to a reference sequence together with the corresponding sequencing quality scores and then infers the genotype with the highest posterior probability at each site based on Bayes' theorem (the reverse probability model). Thus, we have taken into account the intrinsic bias or errors that are common in Illumina GA sequencing data and recalibrated the quality values for use in inferring consensus sequence.

### **Real-time quantitative reverse transcription (RT)-PCR**

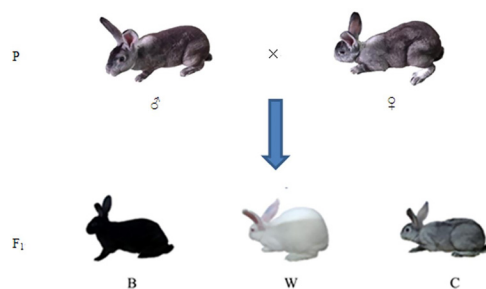
Six genes were selected from the DEGs for validation by quantitative real-time PCR

analysis. Quantification was performed with a two-step reaction process involving RT and PCR. Each RT reaction consisted of 0.5  $\mu$ g RNA, 2  $\mu$ L PrimerScript Buffer, 0.5  $\mu$ L oligo dT, 0.5  $\mu$ L random 6-mers, and 0.5  $\mu$ L PrimerScript RT Enzyme Mix I (TaKaRa, Japan) in a total volume of 10  $\mu$ L. Reactions were performed in a GeneAmp<sup>®</sup> PCR System 9700 (Applied Biosystems) for 15 min at 37°C, followed by heat inactivation of the reverse transcriptase for 5 s at 85°C. The 10- $\mu$ L RT reaction mix was then diluted 1:10 in nuclease-free water and maintained at -20°C. Real-time PCR was performed using the LightCycler<sup>®</sup> 480 II Real-time PCR Instrument (Roche, Switzerland) with a 10- $\mu$ L PCR mixture that included 1  $\mu$ L cDNA, 5  $\mu$ L 2X LightCycler<sup>®</sup> 480 SYBR Green I Master (Roche), 0.2  $\mu$ L forward primer, 0.2  $\mu$ L reverse primer, and 3.6  $\mu$ L nuclease-free water. Reactions were incubated on a 384-well optical plate (Roche) at 95°C for 10 min, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. Each sample was run in triplicate for analysis. The primer sequences were designed in the laboratory and synthesized by Generay Biotech (Generay, PRC) based on the RNA sequences obtained from the NCBI database ([S1 Table](#)). At the end of the PCR, melting curve analysis was performed to validate the specific generation of the expected PCR product. mRNA expression levels were normalized to that of *GAPDH* and were calculated using the  $2^{-\Delta\Delta C_t}$  method (Livak and Schmittgen, 2001).

## RESULTS

### Basic analysis of raw data

To minimize the impact of an individual's genetic background and environment on fur color, we collected samples from full-sibling Rex rabbits of different colors. The fur colors of the rabbits were characterized as B, W, and C (Figure 1). To obtain an overview of the rabbit gene expression profile at the first molt, cDNA from the back-skin of Rex rabbits was sequenced using the Solexa-sequencing platform. A summary of the number of raw reads and their mapping to the reference database is presented in Table 1. In total, 47,831,184 (sample B), 46,142,230 (sample W), and 48,743,588 (sample C) clean reads were obtained by sequencing. In our study, clean reads were mapped to the rabbit genome in the Ensemble database. Gene coverage was calculated as the percentage of a gene covered by reads. This value is equal to the ratio of the number of bases in a gene covered by unique mapping reads to the total number of bases in the coding region in that gene. Gene coverage statistics of the samples were similar (Figure 2). Approximately 73% of the genes in the samples exhibited gene coverage of at least 80%.

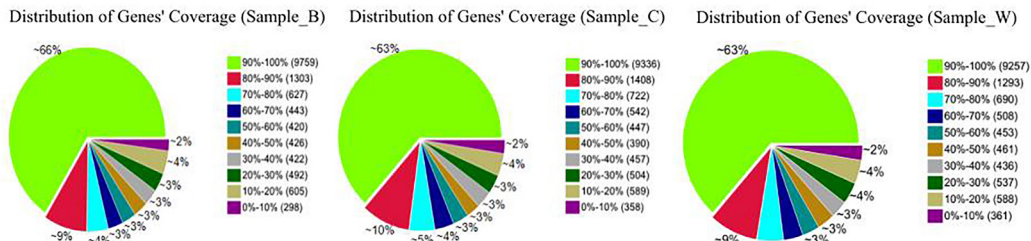


**Figure 1.** Fur colors of the Rex rabbits and their parents used in this study. B = black; W = white; C = chinchilla.

**Table 1.** Number of raw reads and their mapping to the reference database.

| Statistics of data production |                                  | B                   | W                   | C                   |
|-------------------------------|----------------------------------|---------------------|---------------------|---------------------|
| Total reads                   |                                  | 57,672,967          | 53,767,964          | 58,360,826          |
| Total base pairs              |                                  | 5,767,297,600       | 5,376,796,400       | 5,836,082,600       |
| Q20 bases count (bp)          |                                  | 4,928,221,650       | 4,633,767,111       | 4,960,249,349       |
| Q20 bases ratio (%)           |                                  | 85.45               | 86.18               | 84.99               |
| Clean reads                   |                                  | 47,831,184          | 46,142,230          | 48,743,588          |
| Total mapped reads to gene    | Total mapped reads (Num./Pct.)   | 16,482,200 (35.72%) | 17,726,278 (37.06%) | 16,705,914 (34.27%) |
|                               | Perfect match (Num./Pct.)        | 11,017,416 (23.88%) | 11,692,021 (24.44%) | 10,897,143 (22.36%) |
|                               | ≤5 bp mismatch (Num./Pct.)       | 5,464,784 (11.84%)  | 6,034,257 (12.62%)  | 5,808,771 (11.92%)  |
|                               | Unique match (Num./Pct.)         | 14,827,845 (32.14%) | 16,241,027 (33.95%) | 14,999,378 (30.77%) |
|                               | Multi-position match (Num./Pct.) | 1,654,355 (3.59%)   | 1,485,251 (3.11%)   | 1,706,536 (3.50%)   |
|                               | Total unmapped reads (Num./Pct.) | 29,660,030 (64.28%) | 30,144,906 (62.94%) | 32,037,674 (65.73%) |
| Total mapped reads to genome  | Total mapped reads (Num./Pct.)   | 33,248,388 (72.06%) | 34,251,090 (71.61%) | 35,085,424 (71.98%) |
|                               | Perfect match (Num./Pct.)        | 20,360,205 (44.12%) | 20,718,190 (43.32%) | 21,100,746 (43.29%) |
|                               | ≤5 bp mismatch (Num./Pct.)       | 12,888,183 (27.93%) | 13,532,900 (28.29%) | 13,984,678 (28.69%) |
|                               | Unique match (Num./Pct.)         | 29,226,201 (63.34%) | 30,573,140 (63.92%) | 30,851,491 (63.29%) |
|                               | Multi-position match (Num./Pct.) | 4,022,187 (8.72%)   | 3,677,950 (7.69%)   | 4,233,933 (8.69%)   |
|                               | Total unmapped reads (Num./Pct.) | 12,893,842 (27.94%) | 13,580,094 (28.39%) | 13,658,164 (28.02%) |

B = black; W = white; C = chinchilla.



**Figure 2.** Gene coverage statistics of each fur samples [B (black), W (white), and C (chinchilla)].

## Quality evaluation of sequencing

Clean reads were obtained by processing data on the adaptor sequences and low-quality sequences from the raw reads, and the clean reads were used as the basis for subsequent analyses. The proportion of clean reads in the samples accounted for 85% of the raw reads, indicating that the sequencing quality of the samples was good and that the number of clean reads obtained met the needs of the study. The curve describing the number of sequencing reads tended to reach a plateau after 20 M ([S1 Figure](#)), suggesting that the number of sequences reached saturation. Therefore, the number of sequence reads in this study were sufficient for DEG analysis. The analysis provided the distribution of reads in the reference genome and the location of the reads, including exons, introns, and intergenic regions. We show the distribution

of reads in the reference genome on Chr13 of sample B as an example ([S2 Figure](#)). In the gene number diagram, the whole chromosome was divided into 500 windows. Each window represented the number of genes in the 286721-nt region, and the longitudinal coordinate represented the gene number corresponding to each window. Reads were well distributed along the 13th chromosome.

## Analysis of gene differential expression

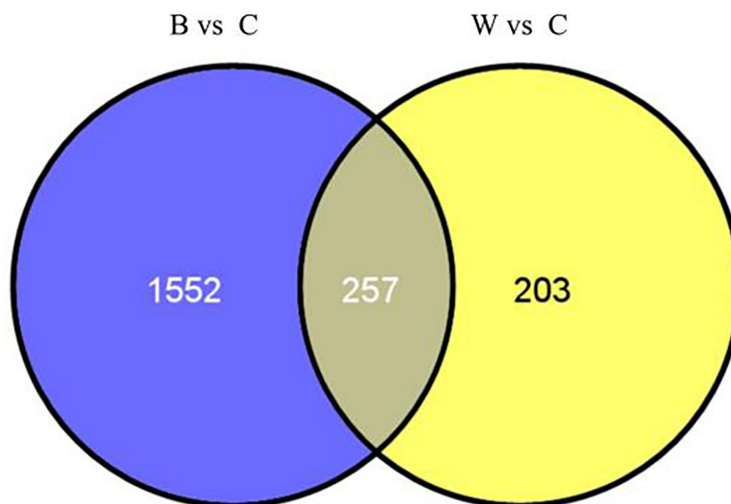
During bioinformatic analysis, DEGs were defined as those with  $FDR \leq 0.001$  and at least a 2-fold-change in expression. After screening with these criteria, the number of up- and down-regulated genes in the B and W samples compared to the C sample was determined. The number of expressed genes in B, C and W sample were 14792, 14752 and 14854, respectively. Of the 30 genes and/or transcription factors with the highest RPKM values in the samples, expression of the elongation factor 1-alpha 1 gene (*EEF-1*) in the samples was very high. Ribosomal protein family genes and keratin family genes were also highly expressed in these samples (Table 2).

**Table 2.** Prior 30 genes or transcription factors with the highest RPKM value in each fur sample.

| Gene name  | RPKM (B) | RPKM (W) | RPKM (C) |
|--|----------|----------|----------|
| Elongation factor 1-alpha  | 10355.9  | 12003.0  | 13772.4  |
| PREDICTED: keratin 25-like   | 8475.1   | 10880.3  | 5744.8   |
| Keratin 10   | 5650.2   | 8905.3   | 15600.0  |
| 40S ribosomal protein S12  | 5601.5   | 7274.1   | 8458.9   |
| PREDICTED: 60S acidic ribosomal protein P1 isoform 1                                   | 4564.2   | 6359.3   | 6523.9   |
| PREDICTED: Uncharacterized protein LOC466117 isoform 7                                 | 4326.9   | 5110.0   | 5665.2   |
| 40S ribosomal protein S20 isoform 2  | 3907.7   | 5296.1   | 5926.1   |
| PREDICTED: Keratin 6A-like   | 3805.6   | 6019.9   | 9858.1   |
| 60S ribosomal protein L13a, partial  | 3648.3   | 5258.4   | 6047.1   |
| Ribosomal protein L26  | 3645.3   | 4793.4   | 5364.2   |
| Ribosomal protein L4 variant   | 3410.5   | 4432.9   | 4725.2   |
| PREDICTED: Ubiquitin-40S ribosomal protein S27a  | 3365.0   | 4723.4   | 4757.4   |
| 60S ribosomal protein L11, partial   | 3254.1   | 4395.8   | 5417.1   |
| Keratinocyte differentiation-associated protein precursor                              | 3120.2   | 5274.5   | 10423.3  |
| 40S ribosomal protein S11  | 3120.1   | 4613.3   | 5108.1   |
| 40S ribosomal protein S3   | 3069.7   | 4133.8   | 4466.6   |
| PREDICTED: 60S ribosomal protein L12-like  | 3033.2   | 4172.8   | 4899.4   |
| PREDICTED: Ribosomal protein L7-like   | 2964.2   | 3839.1   | 4134.4   |
| mCG19129   | 2942.7   | 3738.4   | 5028.8   |
| 40S ribosomal protein S5   | 2822.9   | 3879.8   | 4924.7   |
| PREDICTED: Trichohyalin-like   | 7865.0   | 6699.4   | -        |
| PREDICTED: Fatty acid binding protein 9, testis  | 4475.9   | 4989.7   | -        |
| PREDICTED: Keratin-associated protein 7-1-like   | 4417.8   | 4179.7   | -        |
| PREDICTED: Cystatin M  | 3562.8   | 3725.7   | -        |
| PREDICTED: Hypothetical protein  | 3562.3   | -        | -        |
| PREDICTED: Keratin-associated protein 3.1-like   | 3541.7   | -        | -        |
| PREDICTED: Keratin 27-like   | 3374.5   | -        | -        |
| PREDICTED: Hypothetical protein  | 3165.3   | -        | -        |
| PREDICTED: Keratin-associated protein 11-1-like  | 2906.8   | -        | -        |
| hypothetical protein A306_01592, partial   | 2852.5   | -        | -        |
| PREDICTED: ADP-ribosylation factor-like protein 6-interacting protein 1-like isoform 2 | -        | 3994.4   | 4471.4   |
| PREDICTED: Calmodulin-like 3-like  | -        | 3848.4   | 4691.0   |
| PREDICTED: 40S ribosomal protein S16   | -        | 3772.6   | 3759.5   |
| 60S ribosomal protein L10  | -        | 3618.6   | 4105.5   |
| 40S ribosomal protein S25  | -        | 3610.0   | 3915.8   |
| PREDICTED: 60S ribosomal protein L23   | -        | 3539.6   | 3556.0   |
| Peptidyl-prolylcis-trans isomerase A   | -        | 3525.3   | -        |
| PREDICTED: Keratin 14-like   | -        | -        | 4595.5   |
| PREDICTED: Ribosomal protein L14-like  | -        | -        | 3655.2   |
| PREDICTED: Ribosomal protein S19-like  | -        | -        | 3489.6   |

These analyses revealed that 1809 genes were differentially expressed in sample B compared to sample C rabbit skin. Of these, 1150 genes (1008 known and 142 unknown) were up-regulated ( $\geq 2$ -fold) and 659 genes (567 known and 92 unknown) were down-regulated ( $\leq 2$ -fold) in skin from sample B compared with skin from sample C (**S2 Table**). Comparison of sample W and sample C revealed 460 DEGs, of which 157 (99 known) were up-regulated and 303 (256 known) were down-regulated in skin from sample W compared with skin from sample C (**S3 Table**). The number of DEGs identified between the B and C rabbits was 1809, while W and C rabbits was 460. This is an indirect evidence that there is a closer genetic relationship between the W and C rabbits than between B and C in terms of pigment formation.

The Venn diagram showing the 1809 and 460 DEGs included 257 common genes (Figure 3). Hierarchical clustering analysis was also performed (**S3 Figure**). Thirty-two genes were up-regulated in sample B and down-regulated in sample W, and these genes may contribute to the formation of melanin (Table 3). Only the fatty acid binding protein 7 (*FABP7*) gene was down-regulated in sample B and up-regulated in sample W, suggesting that this gene product may inhibit the formation of melanin. Hierarchical clustering analysis of the 33 genes (Figure 4) showed the relationship between up- and down-regulated genes and changes in gene expression between samples B and W compared to sample C. According to the clustering diagram, the genes were divided into two categories, with 12 sub-categories. Many sarcoplasmic proteins, including actins and myomesins, were observed among these 33 differential genes; however, no new genes with known function were found. This analysis also identified *TYRP1* and *TYR*, which represented the only genes known to be associated with fur color formation in rabbits.



**Figure 3.** Venn analysis of the differentially expressed genes. B = black; C = chinchilla; W = white.

### Functional annotation of DEGs

To understand the functions of DEGs, we carried out GO functional enrichment and KEGG pathway analysis. GO analysis comparing sample B with sample C rabbit skin indicated that 836, 786, and 758 DEGs were classified in the cellular component, molecular function,

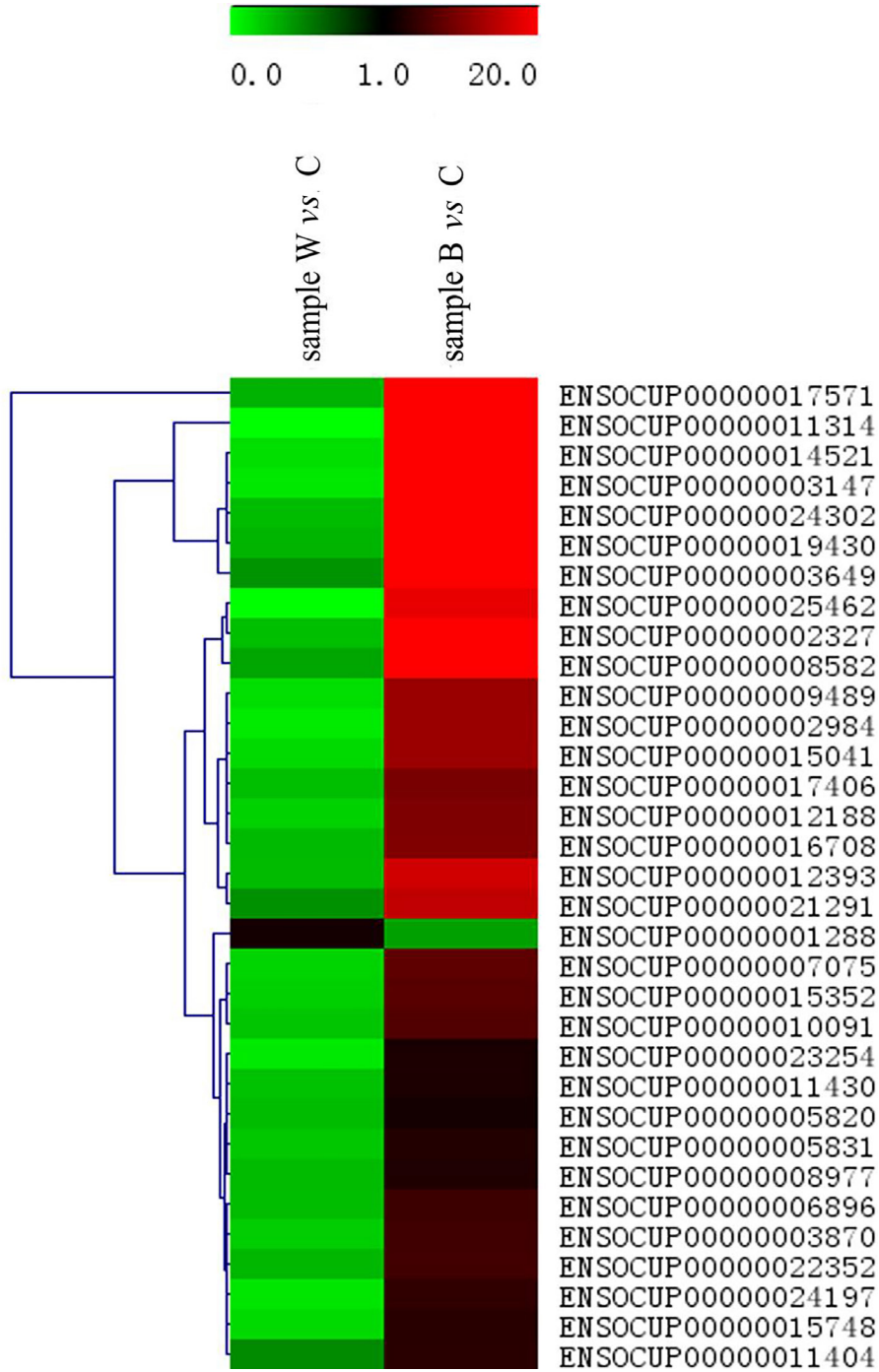


and biological process categories, respectively. In a comparison of sample W with sample C rabbit skin, 189, 171, and 176 DEGs were grouped in the cellular component, molecular function, and biological process categories, respectively. Of these GO terms, the majority of the DEGs were classified into cellular processes, cell and cell part, and binding categories ([S4 Figure](#)).

**Table 3.** Genes up-regulated in sample B (black) and down-regulated in sample W (white) compared with sample C (chinchilla).

| Gene ID             | Name  |
|---------------------|---|
| ENSOCUP00000017571  | Myosin-4  |
| ENSOCUP00000011314  | Myomesin family member 2  |
| ENSOCUP00000003649  | TRP-1   |
| ENSOCUP00000024302  | PREDICTED: Actinin, alpha 3   |
| ENSOCUP00000003147  | Voltage-dependent L-type calcium channel subunit alpha-1S   |
| ENSOCUP00000014521  | PREDICTED: titin  |
| ENSOCUP00000019430  | Myozenin-1  |
| ENSOCUP00000008582  | Tyrosinase precursor  |
| ENSOCUP00000002327  | Chain A, Crystal Structure of the recombinant Serca1a (calcium Pump of Fast Twitch Skeletal Muscle) in the E1.mg2+State |
| ENSOCUP00000025462  | PREDICTED: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2-like                                      |
| ENSOCUP00000012393  | PREDICTED: Myotilin   |
| ENSOCUP000000021291 | PREDICTED: Myosin, heavy chain 7, cardiac muscle, beta  |
| ENSOCUP00000015041  | Nebulin   |
| ENSOCUP00000002984  | PREDICTED: Myosin heavy chain IIa isoform 2   |
| ENSOCUP00000016708  | Myosin regulatory light chain 2   |
| ENSOCUP00000012188  | Myosin light chain 1/3, skeletal muscle isoform   |
| ENSOCUP00000017406  | PREDICTED: Myosin binding protein C, slow type  |
| ENSOCUP000000007075 | PREDICTED: Neuronal tropomodulin-like   |
| ENSOCUP00000015352  | PREDICTED: Actinin, alpha 2-like  |
| ENSOCUP00000010091  | PREDICTED: Carbonic anhydrase 3-like  |
| ENSOCUP00000022352  | PREDICTED: Kelch repeat and BTB (POZ) domain containing 10  |
| ENSOCUP000000003870 | PREDICTED: Cysteine and glycine-rich protein 3 isoform 1  |
| ENSOCUP000000006896 | Troponin C, slow skeletal and cardiac muscles   |
| ENSOCUP000000024197 | PREDICTED: Leucine rich repeat containing 39  |
| ENSOCUP00000011404  | PREDICTED: SH3 and cysteine rich domain 3   |
| ENSOCUP00000015748  | PREDICTED: Myosin regulatory light chain 2, ventricular/cardiac muscle isoform-like                                     |
| ENSOCUP000000005831 | PREDICTED: Myosin XVIIIIB-like  |
| ENSOCUP000000008977 | troponin I, slow skeletal muscle  |
| ENSOCUP00000011430  | PREDICTED: SCM-like with four MBT domains 2   |
| ENSOCUP00000023254  | RecName: Full=Myoglobin   |
| ENSOCUP000000005820 | PREDICTED: Myomesin family, member 3  |
| ENSOCUP000000009489 | PREDICTED: Nebulin-related anchoring protein isoform 2  |

The majority of the GO terms, including pigmentation, did not appear to be significantly enriched in the DEGs. The DEGs and the associated GO terms related to pigmentation and melanogenesis. A comparison of sample C with sample W is shown in Table 4. In contrast to the GO analysis, results of the KEGG pathway analysis are more indirect as these describe the interactions between proteins, and thus, a change in the pathway can be caused by changes in the expression and activity of the proteins involved. Of the 1809 DEGs identified between samples B and C rabbit skin, 1502 were annotated into 237 specific KEGG pathways. Of the 460 DEGs identified between samples W and C rabbit skin, 373 were annotated into 190 specific KEGG pathways. These pathways were found to be primarily related to metabolism and diseases; however, q-values for the association with pathways related to pigmentation were all greater than 0.05, indicating a lack of any significant difference. The highest q-value was for DEGs in the pathway associated with pigmentation (Figure 5).

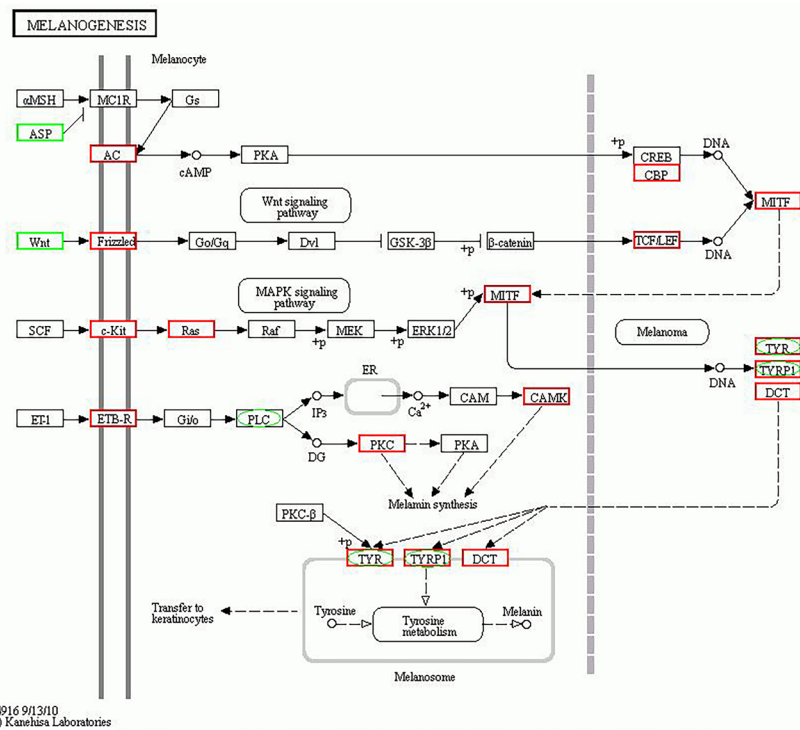


**Figure 4.** Hierarchical clustering of differentially expressed genes in B (black) and W (white) samples compared with the C (chinchilla) sample. Red indicates up-regulation, and green indicates down-regulation.

**Table 4.** Differentially expressed genes and their GO terms related to pigmentation and melanogenesis.

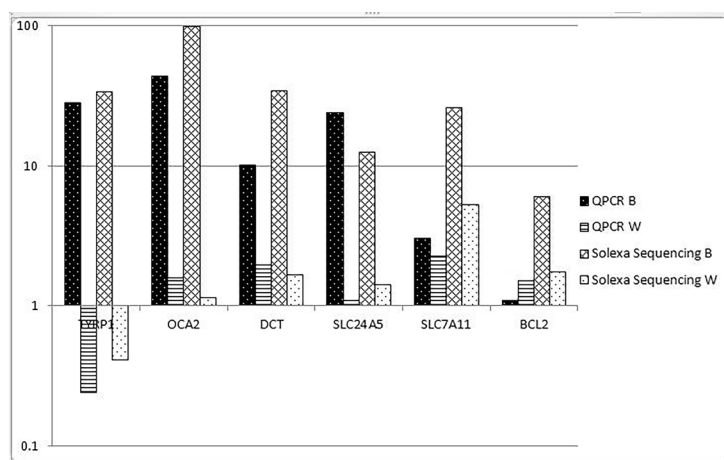
| Comparison   | GO terms  | Genes           | Relative expression | Fold-change | Corrected P value of GO term |
|--------------|---|-----------------|---------------------|-------------|------------------------------|
| B versus C   | Pigment metabolic process                               | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
|              |   | <i>ASIP</i>     | Down-regulation     | 0.44        |                              |
|              |   | <i>TYR</i>      | Up-regulation       | 22.69       |                              |
|              | Pigment cell differentiation                            | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
|              |   | <i>MITF</i>     | Up-regulation       | 2.32        |                              |
|              | Pigmentation and developmental pigmentation             | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
|              |   | <i>ASIP</i>     | Down-regulation     | 0.44        |                              |
|              |   | <i>TYR</i>      | Up-regulation       | 22.69       |                              |
|              |   | <i>MITF</i>     | Up-regulation       | 2.32        |                              |
|              | Pigment granule organization and cellular pigmentation  | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
|              |   | <i>ASIP</i>     | Down-regulation     | 0.44        |                              |
|              | Melanin metabolic process                               | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
| <i>TYR</i>   |   | Up-regulation   | 22.69               |             |                              |
| <i>ASIP</i>  |   | Down-regulation | 0.44                |             |                              |
| <i>TYRP1</i> |   | Up-regulation   | 33.92               |             |                              |
| W versus C   | Pigment metabolic process                               | <i>TYRP1</i>    | Up-regulation       | 33.92       | 1                            |
|              | Pigmentation  | <i>TYR</i>      | Up-regulation       | 22.69       | 1                            |
|              | Developmental pigmentation<br>Melanin metabolic process |                 |                     |             |                              |

B = black; W = white; C = chinchilla.



**Figure 5.** Differentially expressed coat color genes in rabbit skin and their involvement in the melanogenesis pathway. Red boxes represent up-regulated genes in sample B compared with sample C, whereas green boxes represent down-regulated genes in this sample. Green circles represent the down-regulated genes in sample W compared with sample C. *TYRP1* and *TYR* were the only genes up-regulated in sample B but were down-regulated in sample W. In this pathway, many genes were up-regulated in sample B, whereas only three were down-regulated in sample W and no up-regulated genes were found in this sample.

The results of the RT-PCR and sequencing analyses demonstrated consistent up- and down-regulation of similar magnitude for each of the six selected genes, suggesting that the transcriptome sequencing results were reliable (Figure 6).



**Figure 6.** Quantitative PCR validation of the transcriptome data. The gene RPKM value was set as 1 for sample C, and therefore, a value of  $>1$  indicates up-regulation, while a value of  $<1$  indicates down-regulation.

## Gene structure refinement, novel transcript prediction, and annotation

By comparing the SOAP results with the existing gene annotations, we optimized the gene structure of the samples and predicted further transcribed sequences from intergenic regions. These expressed genes may perform important functions and may act as regulatory sequences in animals. We also optimized the structure of 6747, 6430, and 6693 genes in samples B, W, and C, respectively. However, the identity of the products of these transcripts and the functions of these proteins remain unknown.

## Alternative splicing

Alternative splicing is an important mechanism involved in gene regulation, which commonly affects gene expression. Alternative splicing allows various proteins with different structures and different functions to be produced from a single gene. Intron retention, a type of alternative splicing, was found in all samples. In this analysis, we detected 8736 alternative splice variants in the 3891 genes in sample B, 6612 alternative splice variants in the 3202 genes in sample W, and 7597 alternative splice variants in the 3559 genes in sample C.

## DISCUSSION

In this study, we employed transcriptome sequencing and functional analysis to investigate genes associated with color determination in the fur of Rex rabbits using samples of the back skin and high-throughput Solexa technology. Sequence analysis revealed DEGs between samples B and C. Products of the following genes were found to be associated with

fur color formation: *TYRP1*, *TYR*, *c-KIT*, *Ras*, *Mitf*, Dopachrome tautomerase (*DCT*), Solute carrier family 7, member 11 (*Slc7a11*), and Oculocutaneous albinism II (*OCA2*). Only *TYRP1*, *TYR*, and *Slc7a11* were expressed differentially between samples W and C. However, compared with sample C, *Slc7a11* was up-regulated in samples B and W with RPKM ratios of 25.99 and 5.44, respectively. Li et al. (2009) found that the *Slc7a11* gene was most highly expressed in brown Kazak sheep, was moderately expressed in black sheep, and was expressed at the lowest levels in white sheep (Li et al., 2012). This indicates that the expression of *Slc7a11* does not increase as the fur color deepens. Therefore, *Slc7a11* may serve as a candidate gene for fur color determination, and it should be explored further. Although only two treatment samples and one control sample were investigated in our study, the DEGs identified by comparing sample B with sample C, and sample W with sample C may serve as candidate genes for the further in-depth study of fur color.

As shown in Table 3, the genes that were up-regulated in sample B and down-regulated in sample W relative to sample C mostly encode sarcoplasmic proteins. In terms of the clustering analysis diagram, the *TYRP1* gene belongs to the same class as the Myomesin 2 (*MYOM2*), Actinin alpha 3 (*ACTN3*), calcium channel, voltage-dependent, L type, alpha 1S subunit (*Cacnals*), *titin*, and Myozenin 1 (*MYOZ1*) genes. The *TYP* gene together with the Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2 (*APOBEC2*) and Sarco-Endoplasmic Reticulum Calcium ATPase 1 (*SERCA1*) genes, also belongs to this class. Provance et al. (2002) revealed that melanophilin is required for the targeting of myosin-Va to melanosomes. Melanosomes detach from microtubules and bind to subcortical actin bundles through an interaction with myosin-Va (Goud, 2002). Melanosomes and their melanin pigment content are transferred to surrounding keratinocytes via a process that is essential for skin and hair pigmentation. Many sarcoplasmic protein genes were found in the 32 DEGs identified in our study. According to previous studies, sarcoplasmic protein-Va is involved in melanin transfer (Goud, 2002; Provance et al., 2002). With respect to the results of the clustering analysis, sarcoplasmic proteins have some impact on the formation of melanin in epidermal hair. In the future, knocking out specific genes encoding sarcoplasmic proteins would allow their validation in cells and animals.

*EEF-α1* belongs to the G-protein family, and was found to be highly expressed in all samples. This factor is mainly involved in mRNA translation but also functions as a GTPase and forms a triploid complex by combining with GTP and aminoacyl tRNA. It recruits aminoacyl tRNA to ribosomes and regulates the translocation of polypeptides from the ribosomal A position to the P position. As the ribosome is an indispensable organelle for protein synthesis, high expression of genes encoding ribosomal proteins reinforce their importance in protein translation. Abundant translation activity was reflected in the high transcription levels of the ribosomal protein genes, which was consistent with previous findings (Karsi et al., 2002). Keratin is a fibrous protein that has connective and protective functions, and exists mainly in the epidermal structures of animals. Powell and Rogers (1997) found that the expression of keratin requires the coordination of 50-100 genes for the production of mature wool fiber. Therefore, keratin family genes were predicted to have high RPKM values. García-Gómez et al. (2011) proposed that no single keratin gene is responsible for the coat color trait. Fan et al. (2013) revealed that high expression of keratin only affected the strength of wool. Other researchers, however, have not found a relationship between keratin genes and fur color formation, and therefore, we did not study these genes.

*TYRP1* and *TYR*, which are known to be related to fur color formation, were identified

in the present study. As a member of the tyrosinase-related family and the first successfully cloned pigment gene, *TYRP1* plays an important role in the downstream pathway of melanin synthesis. Winder et al. (1994) showed that the mouse brown locus is influential in determining the structure of the melanosome, and *TYRP1* plays an important role in stabilizing tyrosinase and other melanosomal enzyme activities. Other reports have also described a significant correlation between *TYRP1* expression and fur color formation (Schmidt-Küntzel et al., 2005; Duchesnes et al., 2009). Fan et al. (2013) showed that *TYRP1* is the most differentially expressed color formation gene between black and white Sunit sheep. This result is consistent with the experimental results reported here. Rooryck et al. (2006) reported two *TYRP1* mutations in a Caucasian patient; a missense mutation (1066 g > a/c. p.A rg356Glu), which was inherited from the mother, and a *de novo* single-base deletion (c.106 del/p.L eu36X). Therefore, *TYRP1* is a very important candidate gene.

TYR is a key enzyme in the process of melanin synthesis, and this enzyme catalyzes different reactions in three steps (Hearing and Tsukamoto, 1991). *TYR*, *TYRP1*, and *TYRP2* constitute the tyrosinase gene family and their products catalyze specific reactions during melanin synthesis (Jimbow, 1999). Tyrosinase is encoded by tyrosinase genes at the albino locus, and functional mutation of tyrosinase genes leads to albinism, which has been demonstrated in mice, rabbits, humans, and cows (Aigner et al., 2000; King et al., 2003; Scullin et al., 2003; Schmutz et al., 2004).

Compared with sample C, only *FABP7* was down regulated in sample B and up regulated in sample W. *FABP7* is a very important member of the poly-adenylate binding protein family. Most researchers have focused on the relationship between *FABP7* and cancers, including brain glioma. Slipicevic et al. (2008) showed that the *FABP7* protein regulates the proliferation and invasion of melanoma cells *in vitro*. However, these investigators did not elaborate on whether this protein could weaken melanin in melanoma cells. Furthermore, Colombo et al. (2012) analyzed the transcriptome of mouse embryonic skin cells and revealed previously unreported genes that were expressed in melanoblasts. In addition, *FABP7* was significantly expressed in melanoblasts, melanocytes, and melanomas. Based on the findings of these previous studies, we further explored the activity of *FABP7* and its involvement in melanin synthesis.

In conclusion, using Solexa transcriptome sequencing of full-sib domestic rabbits, we identified genes known to be related to fur formation (including *TYRP1* and *TYR*) and new genes with unknown functions. These findings are expected to provide reference for future studies associated with the mechanisms of fur color formation.

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

Research supported by the Modern Agricultural Industrial System Special Funding (CARS-44-1) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (#2011-137). We are grateful to Ling Pan in Shanghai Oebiotech Co., Ltd. for advice on data analysis.

## REFERENCES

- Aigner B, Besenfelder U, Müller M and Brem G (2000). Tyrosinase gene variants in different rabbit strains. *Mamm. Genome* 11: 700-702. <http://dx.doi.org/10.1007/s003350010120>
- Audic S and Claverie JM (1997). The significance of digital gene expression profiles. *Genome Res.* 7: 986-995.
- Bauer GL, Praetorius C, Bergsteinsdóttir K, Hallsson JH, et al. (2009). The role of MITF phosphorylation sites during coat color and eye development in mice analyzed by bacterial artificial chromosome transgene rescue. *Genetics* 183: 581-594. <http://dx.doi.org/10.1534/genetics.109.103945>
- Benjamini Y and Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29: 1165-1188.
- Bennett DC and Lamoreux ML (2003). The color loci of mice--a genetic century. *Pigment Cell Res.* 16: 333-344. <http://dx.doi.org/10.1034/j.1600-0749.2003.00067.x>
- Brunberg E, Andersson L, Cothran G, Sandberg K, et al. (2006). A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genet.* 7: 46. <http://dx.doi.org/10.1186/1471-2156-7-46>
- Chang TS (2009). An updated review of tyrosinase inhibitors. *Int. J. Mol. Sci.* 10: 2440-2475. <http://dx.doi.org/10.3390/ijms10062440>
- Colombo S, Champeval D, Rambow F and Larue L (2012). Transcriptomic analysis of mouse embryonic skin cells reveals previously unreported genes expressed in melanoblasts. *J. Invest. Dermatol.* 132: 170-178. <http://dx.doi.org/10.1038/jid.2011.252>
- Cone RD, Lu D, Koppula S, Vage DI, et al. (1996). The melanocortin receptors: agonists, antagonists, and the hormonal control of pigmentation. *Recent Prog. Horm. Res.* 51: 287-317, discussion 318.
- Cooksey CJ, Garratt PJ, Land EJ, Pavel S, et al. (1997). Evidence of the indirect formation of the catecholic intermediate substrate responsible for the autoactivation kinetics of tyrosinase. *J. Biol. Chem.* 272: 26226-26235. <http://dx.doi.org/10.1074/jbc.272.42.26226>
- Duchesnes CE, Naggert JK, Tatnell MA, Beckman N, et al. (2009). New Zealand Ginger mouse: novel model that associates the tyrp1b pigmentation gene locus with regulation of lean body mass. *Physiol. Genomics* 37: 164-174. <http://dx.doi.org/10.1152/physiolgenomics.90336.2008>
- Fan R, Xie J, Bai J, Wang H, et al. (2013). Skin transcriptome profiles associated with coat color in sheep. *BMC Genomics* 14: 389. <http://dx.doi.org/10.1186/1471-2164-14-389>
- García-Gámez E, Reverter A, Whan V, McWilliam SM, et al.; International Sheep Genomics Consortium (2011). Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in merino sheep. *PLoS One* 6: e21158. <http://dx.doi.org/10.1371/journal.pone.0021158>
- Ginsburg B (1944). The Effects of the Major Genes Controlling Coat Color in the Guinea Pig on the Dopa Oxidase Activity of Skin Extracts. *Genetics* 29: 176-198.
- Goud B (2002). How Rab proteins link motors to membranes. *Nat. Cell Biol.* 4: E77-E78. <http://dx.doi.org/10.1038/ncb0402-e77>
- Hartmeyer M, Scholzen T, Becher E, Bhardwaj RS, et al. (1997). Human dermal microvascular endothelial cells express the melanocortin receptor type 1 and produce increased levels of IL-8 upon stimulation with alpha-melanocyte-stimulating hormone. *J. Immunol.* 159: 1930-1937.
- Hearing VJ and Tsukamoto K (1991). Enzymatic control of pigmentation in mammals. *FASEB J.* 5: 2902-2909.
- Huang S, Zhang J, Li R, Zhang W, et al. (2011). SOAPSsplice: Genome-Wide ab initio Detection of Splice Junctions from RNA-Seq Data. *Front. Genet.* 2: 46. <http://dx.doi.org/10.3389/fgene.2011.00046>
- Jimbow K (1999). Biological role of tyrosinase-related protein and its relevance to pigmentary disorders (vitiligo vulgaris). *J. Dermatol.* 26: 734-737. <http://dx.doi.org/10.1111/j.1346-8138.1999.tb02084.x>
- Kanehisa M, Araki M, Goto S, Hattori M, et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36: D480-D484. <http://dx.doi.org/10.1093/nar/gkm882>
- Karsi A, Cao D, Li P, Patterson A, et al. (2002). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene* 285: 157-168. [http://dx.doi.org/10.1016/S0378-1119\(02\)00414-6](http://dx.doi.org/10.1016/S0378-1119(02)00414-6)
- Kijas JM, Wales R, Törnsten A, Chardon P, et al. (1998). Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* 150: 1177-1185.
- King RA, Pietsch J, Fryer JP, Savage S, et al. (2003). Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. *Hum. Genet.* 113: 502-513. <http://dx.doi.org/10.1007/s00439-003-0998-1>
- Li HT, He X, Zhou ZY, Zhao SH, et al. (2012). [Expression levels of Slc7a11 in the skin of Kazakh sheep with different coat colors]. *Yi Chuan* 34: 1314-1319. <http://dx.doi.org/10.3724/SP.J.1005.2012.01314>

- Li R, Yu C, Li Y, Lam TW, et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966-1967. <http://dx.doi.org/10.1093/bioinformatics/btp336>
- Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-408. <http://dx.doi.org/10.1006/meth.2001.1262>
- Marklund S, Moller M, Sandberg K and Andersson L (1999). Close association between sequence polymorphism in the KIT gene and the roan coat color in horses. *Mamm. Genome* 10: 283-288. <http://dx.doi.org/10.1007/s003359900987>
- Mundy NI and Kelly J (2006). Investigation of the role of the agouti signaling protein gene (ASIP) in coat color evolution in primates. *Mamm. Genome* 17: 1205-1213. <http://dx.doi.org/10.1007/s00335-006-0056-0>
- Philipp U, Hamann H, Mecklenburg L, Nishino S, et al. (2005). Polymorphisms within the canine MLPH gene are associated with dilute coat color in dogs. *BMC Genet.* 6: 34. <http://dx.doi.org/10.1186/1471-2156-6-34>
- Powell BC and Rogers GE (1997). The role of keratin proteins and their genes in the growth, structure and properties of hair. *EXS* 78: 59-148.
- Provance DW, James TL and Mercer JA (2002). Melanophilin, the product of the leaden locus, is required for targeting of myosin-Va to melanosomes. *Traffic* 3: 124-132. <http://dx.doi.org/10.1034/j.1600-0854.2002.030205.x>
- Roberts A, Pimentel H, Trapnell C and Pachter L (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27: 2325-2329. <http://dx.doi.org/10.1093/bioinformatics/btr355>
- Rooryck C, Roudaut C, Robine E, Müsebeck J, et al. (2006). Oculocutaneous albinism with TYRP1 gene mutations in a Caucasian patient. *Pigment Cell Res.* 19: 239-242. <http://dx.doi.org/10.1111/j.1600-0749.2006.00298.x>
- Schallreuter KU, Kothari S, Chavan B and Spencer JD (2008). Regulation of melanogenesis - controversies and new concepts. *Exp. Dermatol.* 17: 395-404. <http://dx.doi.org/10.1111/j.1600-0625.2007.00675.x>
- Schmidt-Küntzel A, Eizirik E, O'Brien SJ and Menotti-Raymond M (2005). Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. *J. Hered.* 96: 289-301. <http://dx.doi.org/10.1093/jhered/esi066>
- Schmutz SM, Berryere TG, Ciobanu DC, Mileham AJ, et al. (2004). A form of albinism in cattle is caused by a tyrosinase frameshift mutation. *Mamm. Genome* 15: 62-67. <http://dx.doi.org/10.1007/s00335-002-2249-5>
- Scullin P, Sheahan P and Sheila K (2003). Myoclonic jerks associated with gabapentin. *Palliat. Med.* 17: 717-718.
- Slipicevic A, Jørgensen K, Skrede M, Rosnes AKR, et al. (2008). The fatty acid binding protein 7 (FABP7) is involved in proliferation and invasion of melanoma cells. *BMC Cancer* 8: 276. <http://dx.doi.org/10.1186/1471-2407-8-276>
- Sturm RA (2009). Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* 18 (R1): R9-R17. <http://dx.doi.org/10.1093/hmg/ddp003>
- Winder A, Kobayashi T, Tsukamoto K, Urabe K, et al. (1994). The tyrosinase gene family - interactions of melanogenic proteins to regulate melanogenesis. *Cell. Mol. Biol. Res.* 40: 613-626.

## Supplementary material

**S1 Figure.** Sequencing saturation analysis of samples B, W, and C.

**S2 Figure.** Uniformity of the distribution of reads in the reference genome on chromosome 13 of sample B.

**S3 Figure.** Venn analysis and hierarchical clustering of 257 common differentially expressed genes.

**S4 Figure.** GO functional classification of differentially expressed genes.

**S1 Table.** Primers used for quantitative real-time PCR.

**S2 Table.** Differentially expressed fur color genes in black versus chinchilla rabbit skin.

**S3 Table.** Differentially expressed fur color genes in white versus chinchilla rabbit skin.