Transcriptome-based Selection and Validation of Optimal House-keeping Genes for Skin Research in Goats (Capra hircus)

Jipan Zhang
Southwest University https://orcid.org/0000-0002-3562-4118
Chengchen Deng
Southwest University
Jialu Li
Southwest University
Yong-Ju Zhao (zyongju@163.com)
https://orcid.org/0000-0001-9256-8856

Research article

Keywords: House-keeping genes; Reference genes; Goat; Skin; ComprFinder method

Posted Date: March 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17424/v1

Version of Record: A version of this preprint was published on July 18th, 2020. See the published version at https://doi.org/10.1186/s12864-020-06912-4.

Abstract

Background: In Quantitative real-time polymerase chain reaction (qRT-PCR) experiments, accurate and reliable target gene expression data is dependent on optimal amplification of house-keeping genes (HKGs). The RNA-seq technology offers a novel approach to detect new HKGs with improved stability. Goat (Capra hircus) is an economically important livestock species, and plays an indispensable role in the world animal fiber and meat industry. Unfortunately, uniform and reliable HKGs be used in skin research of goats have not been identified. Therefore, this study seeks to identify a set of stable HKGs for the skin tissue of C. hircus using the new high-throughput sequencing technology.

Results: Based on the transcriptome dataset of 39 goat skin tissues, 8 genes (SRP68, NCBP3, RRAGA, EIF4H, CTBP2, PTPRA, CNBP, and EEF2) with relatively stable expression levels were identified and selected as new candidate HKGs. The classical HKGs including SDHA and YWHAZ from a previous study, and 2 conventional genes (ACTB and GAPDH) were also considered. Four different experimental materials: (1) different development stages, (2) hair follicle cycle stages, (3) breeds and (4) sampling sites were provided for determination and validation. Four algorithms (geNorm, NormFinder, BestKeeper, and ΔCt method) and a comprehensive algorithm (ComprFinder, developed in-house) were used to assess the stability of each HKG. It was observed that NCBP3+SDHA+PTPRA was more stably expressed than previously used genes, in all conditions analyzed. This combination was effective at normalizing target gene expression. Moreover, a new algorithm, ComprFinder, was developed and released for comprehensive analysis.

Conclusion: This study presents the first data of candidate HKGs selection for skin tissues of C. hircus based on an RNA-seq dataset. We propose the use of the NCBP3+SDHA+PTPRA combination as the triplet HKGs in skin molecular biology in C. hircus and other closely related species in order to standardize analyses across studies. In addition, we also encourage researchers who are performing candidate HKG evaluations and have the needs of a comprehensive analysis to adopt our new algorithm, ComprFinder.

Background

In much molecular biology research, determining the relative changes of target gene expression at the transcriptional level requires precise quantitative analysis. The emergence and development of quantitative real-time polymerase chain reaction (qRT-PCR) has enabled the comprehensive application of mRNA quantification. Furthermore, qRT-PCR is a widely used technique due to its accuracy, sensitivity, reproducibility, and cost-effectiveness in analyzing gene expression [1, 2]. The qRT-PCR technology detects real-time fluorescence generated by a dye bound to, and proportional to, the total amount of nucleic acid. This value is typically reported as a cycle threshold value (Ct) in the comparative Ct method [3]. The qRT-PCR assay is critically dependent on house-keeping genes (HKGs) to obtain the relative expression level [4, 5]. Therefore, the choice of HKGs has become a major source of error and bottlenecks in qRT-PCR experiments.

In qRT-PCR experiments, the selection of inadequate HKGs may lead to an inappropriate interpretation of target gene expression [6]. There are two commonly held misconceptions with regard to the selection of HKGs: (I) HKGs are selected based on experience without verification or without reviewing HKG research papers, and (II) using a single HKG with poor stability. In recent years, it has been reported with increasing frequency [7, 8] that the applicability of commonly used HKGs, such as ACTB, GAPDH, and 18sRNA, had crucial limitations. Ideal endogenous HKGs should exhibit consistent expression levels across all experimental conditions (e.g. cell types,

physiological states, and growth conditions [9, 10]. Unfortunately, the ideal HKGs does not exist, which means that each experimental system may need to use unique HKG(s) to accurately represent the specific research question being probed.

Goat (Capra hircus) is an economically important livestock species as a source of meat, hair, and dairy products [11]. The skin tissue, as the largest biological organ, serves many functions such as physical protection from injury and infection, thermal insulation, and a substrate for growing hair. In order to reveal the molecular regulatory mechanism of hair follicle activity, it is necessary to clarify the pattern of target gene expression under different conditions, such as in different stages of the hair follicle cycle or in different breeds. Unfortunately, most molecular studies examining goat skin have blindly used a single HKG such as ACTB [12–14] or GAPDH [15, 16]. In 2014, Bai et al. [17] selected 10 commonly used HKGs based on a literature review to explore their stability in different hair follicle cycles of Liaoning cashmere goats. However, due to the lack of materials involved in and restriction of traditional HKGs, that study resulted in a limited impact. The development of high-throughput RNA-seq technology provides a means of determining spatio-temporal expression at the transcriptome level, and provides a novel approach for the identification of novel HKGs [18, 19]. This strategy was successfully used to identify candidate HKGs for Artemisia sphaerocephala [7], fish [20], tomato leaves [21], Pyropia yezoensis [22], Euscaphis [23], and Arabidopsis pumila [24], all from transcriptome datasets. Therefore, it is hypothesized that novel and credible HKGs can be predicted and validated via transcriptome sequencing data from goat skin tissue.

In this study, the transcriptome dataset of 39 skin tissues of goats was analyzed. Potential HKGs were predicted, of which 8 genes (SRP68, NCBP3, RRAGA, EIF4H, CTBP2, PTPRA, CNBP, and EEF2) were selected based on their relatively stable expression levels. An additional 4 widely used HKGs (SDHA, YWHAZ, ACTB, and GAPDH) were selected for comparison and validation. These 12 genes were amplified using qRT-PCR in 4 groups with different experimental treatments. Four different algorithms (geNorm, ΔCt method, NormFinder, and BestKeeper) and a comprehensive method (developed the ComprFinder method) were used to evaluate the stability of each HKG. Finally, the reliability of the recommended optimal HKGs were validated and confirmed.

Results

Selection of novel candidate HKGs based on RNA-seq data

From a complete transcriptome dataset, the FPKM values of all transcripts from each individual animal were obtained. We first removed some transcripts which did not have a credible function annotation, or exhibited abnormal expression levels (FPKM = 0). This resulted in 15853 unigenes for further selection. Next, expression level and stability were evaluated, as determined by the average of the FPKM value, coefficients of variation (CV, %), maximum fold change (MFC), and dispersion measure (DPM). As shown in Fig. 1, the probability density curve of all 15853 unigenes was evaluated based on the following 4 indicators:

(1) FPKM. Potential HKGs were relatively highly expressed genes (expression levels ≥ the 80th percentile or FPKM ≥ 10) [8]. In this study, including 5623 genes with FPKM values > 10 (35.5% of 15853 genes, the green area in Fig. 1A).

(2) CV (%). The coefficient of variation is defined as the ratio of the standard deviation (SD_{FPKM}) to the mean of the FPKM of all samples for one gene. The CV value represents the degree of variation. The most promising HKGs

would have the lowest CV values. A total of 2266 genes with a CV ≤ 20% (14.3% of 15853 genes, the red area in Fig. 1B) were retained in this step; in fact, CVs ranged from 7.7–20.0%.

(3) DPM. The Dispersion measure parameter was introduced for the identification of the HKGs on pattern gene finder (PaGeFinder) in 2012 [25]. The jar package was downloaded and run from the PaGeFinder website. Most stable genes exhibited lower DPM values. The default parameters of DPM < 0.3 returned and excessive 7025 unigenes, a more stringent DPM < 0.2 was used. This produced 2026 genes (12.8% of 15853 genes, the yellow area in Fig. 1C) were retained in this step, For the selected genes, DPM values ranged from 0.09 to 0.2.

(4) MFC. The maximum fold change, which is defined as the fold change between the largest and smallest FPKM values within 39 RNA-seq were calculated. This parameter reflects the range of extreme value, and the lowest values are preferable. In this study, MFC < 2.5 were selected, producing 2508 genes (15.8% of 15853 genes, the blue area in Fig. 1D), all within the range of 1.35 to 2.5.

Furthermore, a Venn diagram was constructed for the 4-color blocks (green, red, yellow, and blue corresponding to those used in Fig. 1A-D, respectively). This indicated 1325 genes (Fig. 1E) that meet all 4 of the above requirements. These were considered candidate HKGs. Of these, 8 genes including RRAGA, PTPRA, SRP68, EIF4H, NCBP3, CTBP2, CNBP, and EEF2 were selected for further qualification as novel candidate HKGs. Each gene was ranked based on its CV value. Those with a lower CV received a higher ranking order (Table 1). In addition, 4 genes outside of the 1325 were considered, including SDHA and YWHAZ, as they had previously been proposed by other researchers [17], and ACTB and GAPDH genes were included as the most commonly used endogenous HKGs for exploring target gene expression in goats. Therefore, 12 candidate HKGs were analyzed in subsequent steps.

Table 1

The summarised information of 12 potential HKGs based on transcriptome data

Type	Gene symbol	Mean_FPKM	CV (%)	Ranking order [a]	MFC [b]	DPM [c]
New predicted candidate HKGs	RRAGA	51.4	8.4%	6	1.416	0.083
	PTPRA	23.8	9.1%	8	1.474	0.090
	SRP68	27.2	9.2%	9	1.510	0.091
	EIF4H	133.0	9.5%	16	1.479	0.094
	NCBP3	10.0	9.5%	17	1.542	0.094
	CTBP2	22.5	9.9%	25	1.566	0.098
	CNBP	226.5	14.3%	458	1.880	0.141
	EEF2	499.7	15.1%	619	1.923	0.149
Suggested by previous study	SDHA	44.0	18.5%	1679	2.710	0.182
	YWHAZ	137.5	19.2%	1946	2.320	0.189
Conventional HKGs	ACTB	556.1	24.6%	4456	2.962	0.239
	GAPDH	391.6	29.9%	6855	2.945	0.286

[a] Ranking order in all genes based on CV value within all 15853 unigenes

[b] MFC, maximum fold change, highest/lowest FPKM value of one gene within 39 transcriptome profiles

[c] DPM, dispersion measure, were determined by PaGeFinder method and an acceptable value should be ≤ 0.3

Amplification specificity and efficiency of the candidate HKGs and target genes

A total of 15 primer pairs targeting 12 candidate HKGs and 3 target genes were designed for qRT-PCR. The gene, primer and amplicon specifications are presented in Additional file 1: Table S1. All primers were designed to anneal at around 60℃. Amplification efficiency for all 15 genes ranged from 96.4% for DKK1 to 103.9% for PTPRA, and the coefficient of determination varied from 0.9986 to 0.9999. The specificity for each paired primer was validated by the melting curve analysis, which showed a single amplification peak (Additional file 1: Figure S1).

Expression profiles of the candidate HKGs

The mean Ct (the average of 3 technical replicates in the same sample) values were used to calculate gene expression levels among samples with distinct experimental factors. As shown in Fig. 2 and Additional file 1: Table S2, the expression level of the 12 candidate HKGs varied widely from 20.74 to 31.60. The most highly expressed gene was ACTB (mean Ct value: 23.25 cycles), and the lowest was SPR68 (mean Ct value: 29.07 cycles). The top 3 genes with low standard deviations were SRP68 (0.875), NCBP3 (0.970), and PTPRA (0.972). The 3 most deviantly expressed genes were ACTB (1.483), CNBP (1.277), and GAPDH (1.258). The narrower standard deviation range of a gene means it has a higher expression stability in different samples. Although

some genes had a lower standard deviation than others, experimental errors are always possible. Therefore, in order to obtain reliable HKGs to normalize target gene expression levels, further analysis with more scientific algorithms and evaluation for those candidate HKGs is necessary.

Analysis of HKG expression stability

In this study, 4 publically available algorithms were used to evaluate HKGs for higher-accuracy stability rankings: geNorm, NormFinder, BestKeeper, and ΔCt method.

geNorm analysis

Gene expression stability was determined by M-value in geNorm analysis; the lower M value suggests a higher gene expression stability. For group 1, the two most stable genes were EIF4H and EEF2 with the lowest M value, and GAPDH was the most unstable gene (Fig. 3A). For group 2, the two most stable genes were EIF4H and PTPRA, and ACTB was the most unstable gene (Fig. 3B). For group 3, the two most stable genes were EIF4H and PTPRA, whereas ACTB was the most unstable gene (Fig. 3C). For group 4, the two most stable genes were NCBP3 and PTPRA, and GAPDH was the most unstable gene (Fig. 3D). For all samples, geNorm analysis was conducted on 39 samples and 12 HKGs. It was determined that the 3 most stable genes were PTPRA, EIF4H, and NCBP3. Conversely, ACTB, CNBP, and GAPDH were the most unstable genes (Fig. 3E).

geNorm can be used to determine the minimum optimal number of HKGs needed for accurate normalization under different experimental treatments by analyzing pairwise variation (V_n/V_{n+1}). This method recognizes V_n/V_{n+1} <0.15 as a threshold value, and “n” as an appropriate number of the HKG needed. The V2/3 values for all of the experimental variables were below the cut-off value of 0.15 (0.067, 0.078, 0.099, 0.091 and 0.081 for group 1, group 2, group 3, group 4 and all samples, respectively), which indicate that using the double HKGs (first two genes in each group) are sufficiently accurate for use in normalizing qRT-PCR derived gene expression data (Additional file 1: Figure S2). The triplet or more gene combinations can also be used as V_n/V_{n+1} <0.15.

NormFinder analysis

Expression stability values, as determined by NormFinder, are shown in Table 2. For group 1, SDHA and EIF4H were the most stable HKGs, and ACTB was the least stable gene, which was the same as was determined by geNorm. In group 2, SDHA and NCBP3 were the most stable HKGs while ACTB was the least stable gene. In group 3, SDHA and YWHAZ got the top rank, while ACTB ranked at the lowest. In group 4, PTPRA and NCBP3 were the most stable, while GAPDH ranked at the lowest. In all samples, SDHA and NCBP3 were the most stable, while ACTB was the least.

Table 2
Gene expression stability calculated by NormFinder

Gene name	Group 1	Group 2	Group 3	Group 4	All samples
SDHA	0.007(1)	0.006(1)	0.006(1)	0.008(8)	0.009(1)
NCBP3	0.011(5)	0.007(2)	0.017(8)	0.006(2)	0.011(2)
PTPRA	0.008(3)	0.011(5)	0.014(4)	0.005(1)	0.012(3)
EEF2	0.012(7)	0.008(3)	0.014(6)	0.006(3)	0.012(4)
CTBP2	0.013(8)	0.012(7)	0.013(3)	0.007(6)	0.013(5)
EIF4H	0.008(2)	0.012(6)	0.017(7)	0.007(4)	0.014(6)
YWHAZ	0.010(4)	0.017(10)	0.011(2)	0.007(5)	0.015(7)
RRAGA	0.015(9)	0.010(4)	0.019(9)	0.008(7)	0.015(8)
SRP68	0.011(6)	0.015(9)	0.019(10)	0.009(9)	0.016(9)
GAPDH	0.019(11)	0.015(8)	0.014(5)	0.014(12)	0.018(10)
CNBP	0.016(10)	0.019(11)	0.026(11)	0.012(10)	0.021(11)
ACTB	0.021(12)	0.021(12)	0.028(12)	0.014(11)	0.026(12)

BestKeeper analysis

The BestKeeper algorithm used std-values to assess the stability of the HKGs. The lower the std-value, the more stable the HKG expression was. As is shown in Table 3, in group 1, SDHA and PTPRA were the most stable HKGs, whereas RRAGA was the least stable gene. The same was observed with the geNorm analysis. In group 2, SDHA and EEF2 were the most stable HKGs, while ACTB was the least stable gene. In group 3, SDHA and YWHAZ got the top rank, while SPR68 ranked at the lowest. In group 4, EEF2 and NCBP3 were most stable, while GAPDH was the least. In all samples, SDHA and EEF2 were most stable, while ACTB was the least.

Table 3
Expression stability std-values calculated using BestKeeper

Gene name	Group 1	Group 2	Group 3	Group 4	All samples
SRP68	0.468(1)	0.506(3)	0.631(1)	0.767(3)	0.663(1)
SDHA	0.516(2)	0.464(1)	0.737(3)	0.760(2)	0.733(2)
NCBP3	0.611(8)	0.510(4)	0.743(4)	0.826(6)	0.753(3)
CTBP2	0.536(3)	0.538(6)	0.776(6)	0.746(1)	0.764(4)
EIF4H	0.546(4)	0.573(7)	0.709(2)	0.912(9)	0.768(5)
EEF2	0.557(5)	0.473(2)	0.813(8)	0.875(7)	0.775(6)
PTPRA	0.562(6)	0.615(9)	0.758(5)	0.822(5)	0.779(7)
RRAGA	0.693(9)	0.530(5)	0.789(7)	0.820(4)	0.811(8)
YWHAZ	0.586(7)	0.742(12)	0.856(10)	0.901(8)	0.871(9)
CNBP	0.778(11)	0.583(8)	0.987(11)	0.986(12)	0.962(10)
GAPDH	0.948(12)	0.674(10)	0.837(9)	0.981(11)	0.984(11)
ACTB	0.719(10)	0.679(11)	1.203(12)	0.978(10)	1.114(12)

ΔCt analysis

The 12 candidate HKGs were analyzed using the Delta Ct method, the data of which is presented in Table 4. The stability of the gene is inversely related to the std-value, thus a lower value indicates greater stability. In group 1, the two most stably expressed genes were PTPRA and SDHA, and the lowest were GAPDH and ACTB. In group 2, the two most stable genes were EEF2 and SDHA, and the least were ACTB and CNBP. In group 3, SDHA and PTPRA were the most stably expressed, whereas ACTB and CNBP were the least. In group 4, the top two stably expressed genes were NCBP3 and EEF2, whereas CNBP and GAPDH were the least. In all samples, the 3 most stable genes were PTPRA, EEF2, and SDHA, while GAPDH, CNBP, and ACTB were the least stable genes.

Table 4
Gene expression stability calculated by the ΔCt method

Gene name	Group 1	Group 2	Group 3	Group 4	All samples
PTPRA	0.391(1)	0.439(4)	0.573(2)	0.443(3)	0.499(1)
EEF2	0.423(4)	0.412(1)	0.587(4)	0.428(2)	0.500(2)
SDHA	0.391(2)	0.417(2)	0.535(1)	0.475(4)	0.503(3)
NCBP3	0.457(6)	0.424(3)	0.643(7)	0.422(1)	0.512(4)
EIF4H	0.392(3)	0.441(5)	0.604(5)	0.483(6)	0.520(5)
CTBP2	0.486(8)	0.516(7)	0.619(6)	0.493(8)	0.553(6)
YWHAZ	0.425(5)	0.549(8)	0.583(3)	0.486(7)	0.557(7)
RRAGA	0.573(10)	0.442(6)	0.672(8)	0.479(5)	0.568(8)
SRP68	0.460(7)	0.583(9)	0.702(10)	0.499(9)	0.590(9)
GAPDH	0.740(12)	0.594(10)	0.699(9)	0.731(12)	0.741(10)
CNBP	0.555(9)	0.690(11)	0.957(11)	0.644(11)	0.757(11)
ACTB	0.582(11)	0.826(12)	1.198(12)	0.597(10)	0.973(12)

A comprehensive ranking of the four methods examined

Next, the ComprFinder algorithm was employed to obtain a comprehensive score by which to rank the potential HKGs. The results are presented in Table 5. In group 1, the 3 most stable HKGs were EIF4H, PTPRA, and SDHA. In group 2, SDHA, NCBP3, and EEF2 were the most stable HKGs analyzed. In group 3, SDHA, PTPRA, and EIF4H were the three most stable HKGs analyzed, whereas NCBP3, PTPRA, and EEF2 were the most stable genes from group 4. The overall rankings, from the highest to the lowest stability, were NCBP3 > SDHA > PTPRA > EEF2 > EIF4H > SRP68 > CTBP2 > YWHAZ > RRAGA > GAPDH > CNBP > ACTB. It is interesting to note that the top 3 of different group rankings have at least 2 of NCBP3, SDHA and PTPRA. In contrast, the commonly used HKGs, ACTB and GAPDH, were relegated to the bottom 2 and 4 positions, respectively.

Table 5
Comprehensive rankings calculated using the ComprFinder method

Ranking No	Group 1		Group 2		Group 3		Group 4		All samples	
	Gene	Score	Gene	Score	Gene	Score	Gene	Score	Gene	Score
1	EIF4H	0.063	SDHA	0.059	SDHA	0.129	NCBP3	0.105	NCBP3	0.096
2	PTPRA	0.090	NCBP3	0.082	PTPRA	0.170	PTPRA	0.105	SDHA	0.099
3	SDHA	0.093	EEF2	0.090	EIF4H	0.180	EEF2	0.193	PTPRA	0.108
4	EEF2	0.171	EIF4H	0.210	SRP68	0.230	SDHA	0.211	EEF2	0.129
5	SRP68	0.174	RRAGA	0.227	EEF2	0.247	SRP68	0.245	EIF4H	0.143
6	YWHAZ	0.256	PTPRA	0.236	NCBP3	0.252	CTBP2	0.263	SRP68	0.192
7	NCBP3	0.282	CTBP2	0.322	YWHAZ	0.277	EIF4H	0.309	CTBP2	0.248
8	CTBP2	0.358	SRP68	0.430	CTBP2	0.293	RRAGA	0.327	YWHAZ	0.311
9	RRAGA	0.605	GAPDH	0.609	GAPDH	0.399	YWHAZ	0.361	RRAGA	0.320
10	CNBP	0.637	YWHAZ	0.630	RRAGA	0.404	ACTB	0.795	GAPDH	0.603
11	ACTB	0.677	CNBP	0.697	CNBP	0.730	CNBP	0.820	CNBP	0.680
12	GAPDH	0.971	ACTB	0.943	ACTB	1.000	GAPDH	0.994	ACTB	1.000

NCBP3, SDHA, PTPRA were the 3 most stable HKGs across all samples. Furthermore, their scores were within a tight range, calculated at 0.096, 0.099, and 0.108, respectively. They were also preferably ranked in groups 1–4 relative to other genes. Therefore, they were considered to be the 3 most promising candidate HKGs, and were advanced for further validation.

Validation of HKGs by DKK1, SHH, and FGF5 genes

Based on the above analyses, 3 target genes, including DKK1, SHH, and FGF5 were further characterized based on their changes in expression levels during the secondary hair follicle cycle (T1, T2, T3) with normalizations of different single HKG and multi-gene combinations. It was observed that NCBP3, SDHA, and EEF2 were the top 3 HKGs in group 2 (factor: hair follicle cycle) based on expression stability. Therefore, it can be concluded that the combination of NCBP3 + SDHA + EEF2 was the best-normalized gene set for Group 2. Since these 3 genes (NCBP3, SDHA, and PTPRA) are possibly the most important candidate HKGs, they were further characterized to determine optimal combinations for normalization of gene expression studies. Four multi-gene combinations, including NCBP3 + SDHA + PTPRA, NCBP3 + SDHA, NCBP3 + PTPRA, and SDHA + PTPRA, in addition to 3 single-genes (NCBP3, SDHA, and PTPRA) were used for the analysis. Conversely, ACTB and GAPDH were used for comparison, and were also examined as the multi-gene combination ACTB + GAPDH. In total, 11 multi-gene combinations or single genes were used as the normalization factors. For multiple gene combinations, the geometric average of their Ct value was calculated. The relative gene expression level was calculated as $2^{-\Delta Ct}$, $\Delta Ct = \Delta\ (Ct_{target\ gene}\text{-}Ct_{HKGs})$.

As is shown in Fig. 4A, the expression profiles of DKK1 were similarly obtained using the 7 stable single-gene and multi-gene combinations. Furthermore, it was observed that DKK1 was more highly expressed in the T2 period compared to T1, and it was the most highly expressed during the T3 period. Among the unstable single- and multi-gene combinations, only the ACTB and ACTB + GAPDH performed similarly to the stable genes. However, the gene expression profile as normalized by GAPDH different from the other conditions, and no significant difference has been identified among T1, T2 and T3 periods. Expression of the SHH gene was even during the T1 and T2 periods, but a significant decrease in the T3 period compared to T1 and T2 was observed (Fig. 4B). Similarly, the seven stable internal HKG combinations identified this trend, but ACTB did not. The combination of ACTB + GAPDH identified this expression change as a trend, but was not able to detect significant changes in expression. The expression profile of the FGF5 gene, when normalized by the most stable candidate HKGs used individually or in combination here, were very similar. High levels of expression was observed in the T2 period, but no statistical significance expression was identified relative to T1 and T3 (Fig. 4C). The combination of ACTB + GAPDH showed a similar pattern to the stable HKGs, but when ACTB and GAPDH were used individually, the expression patterns were completely different. Furthermore, ACTB also identified significant differences in the T2 period relative to T1.

The above-mentioned results derived from Fig. 4 reflect the differences of expression profiles of a single target gene normalized by 11 types of single or multiple-gene combinations. In order to further understand the relationship of those single or multi-HKG combinations, a correlation analysis on these relative expression data ($2^{-\Delta Ct}$) of 3 target genes was performed. As is shown in Fig. 5, the normalized results by NCBP3 + SDHA + EEF2 and NCBP3 + SDHA + PTPRA had a high correlation coefficient ($R = 0.990$, $P < 0.001$), suggesting that they have extremely similar normalization capabilities. Other double-gene combinations including NCBP3 + SDHA, NCBP3 + PTPRA, and SDHA + PTPRA had high correlation coefficients, ranging from 0.969–0.997 with NCBP3 + SDHA + EEF2. Also, these double-gene combinations had high correlation coefficients of 0.989–0.994 with NCBP3 + SDHA + PTPRA. This indicated that these 3 types of double-gene combinations exhibited similar normalization capabilities like NCBP3 + SDHA + EEF2 and NCBP3 + SDHA + PTPRA. For single stable HKGs NCBP3, SDHA, and PTPRA also exhibited high correlation coefficients with NCBP3 + SDHA + EEF2 (0.942–0.973) and with NCBP3 + SDHA + PTPRA (0.952–0.977). The ACTB, GAPDH, and ACTB + GAPDH combinations had relatively low correlation coefficients with anyone of stable single- (0.513–0.780) and multi-gene combinations (0.548–0.738).

Discussion

Standard criteria for HKG screening for skin tissue research in goats

Which candidate HKGs should we choose

Four original algorithms were used to determine the expression stability values of 12 candidate HKGs. Other final values were determined using a comprehensive algorithm. However, even for the final ComprFinder value, the results varied by different groups. If the top 3 genes were used, groups 1–4 should theoretically be EIF4H + PTPRA + SDHA, SDHA + NCBP3 + EEF2, SDHA + PTPRA + EIF4H, and NCBP3 + PTPRA + EEF2, a total of 5 HKGs (EIF4H, PTPRA, SDHA, NCBP3, EEF2) would be needed. In theory, it is preferable to use multiple high-performing HKGs as a normalization factor. However, in practice, the additional cost, and an excessive number of HKGs limits

the number of samples that can be tested. Therefore, the fewest number of HKGs possible should be used in order to meet the relevant statistical needs in addition to reducing experimental costs [10, 26]. In the present study, NCBP3, SDHA, and PTPRA were the top 3 most stable HKGs for all samples, and 2 of those ranked in the top 3 of any group. Therefore, they were considered to be the 3 most powerful candidate HKGs, and were considered for further validation.

How many candidate HKGs should be used?

There is still no specific theory prescribing a number of HKGs to be used. In the above discussion, NCBP3, SDHA, and PTPRA were proposed for their excellent stability. However, which single or multiple gene combinations (NCBP3 + SDHA + PTPRA, NCBP3 + SDHA, NCBP3 + PTPRA, SDHA + PTPRA, NCBP3, SDHA, PTPRA) should be used? In other words, is there enough efficacy to use a single or double HKG? Compared with NCBP3 + SDHA + EEF2 or NCBP3 + SDHA + PTPRA, the detection efficacy of NCBP3 (Fig. 4A), SDHA (Fig. 4B), and PTPRA (Fig. 4B, Fig. 4C) were not consistent with them. Considering that the performance of the single gene was not good, it is recommended that single HKGs should be avoided, even if they were top-ranked HKGs.

For the double gene combinations NCBP3 + SDHA, NCBP3 + PTPRA and SDHA + PTPRA (Fig. 4A-C), similar expression patterns and detection efficacy were consistent with NCBP3 + SDHA + EEF2 or NCBP3 + SDHA + PTPRA combination. It was observed that NCBP3 + SDHA behaved similarly with NCBP3 + SDHA + EEF2 or NCBP3 + SDHA + PTPRA combination than NCBP3 + PTPRA and SDHA + PTPRA, possibly due to the fact that NCBP3 and SDHA were among the top in the final stability ranking in group 2. This also implies that SDHA + PTPRA, SDHA + PTPRA, and NCBP3 + PTPRA may be the optimal double gene combinations for group 1, group 3, and group 4, respectively. There are still 3 genes, NCBP3, SDHA, and PTPRA, and there is no any type double gene combination able to cope with multiple factors (groups 1–4).

Considering that the 3-gene combination of NCBP3 + SDHA + PTPRA exhibited better stability performance, it can be applied to various factors in goat dermatologic research, and the number of 3 HKGs is still an acceptable range in qRT-PCR experiments. Therefore, it is recommended that NCBP3 + SDHA + PTPRA be adopted as the HKG combination for dermatologic research in goats.

The HKGs of skin tissue in goats and other species

As was said before, a common HKG in traditional skin research of goats was either ACTB [12–14], or GAPDH [15, 16]. Almost all of these studies used a single HKG, so it can be speculated that some of these previous studies may have obtained erroneous data. Therefore, it is crucial that scientists appreciate the advantage of a set of appropriate HKGs. For the target genes that have undergone significant changes, they can also be identified by less stable HKGs. But for the target genes that show slight changes, they can only be identified by optimal HKGs [26]. As far as we know, only one previous HKGs study reported on skin tissue of goat, Bai et al. [17] selected ten commonly used HKGs by consulting the literature to select optimal HKGs to span 3 stages of hair follicle cycle in Liaoning cashmere goats (referred to here as T1, T2, and T3 of IMCG), the authors recommended SDHA + YWHAZ + UBC as the HKGs be used in combination, but their geNorm values (V_2/V_3 = 0.159 and V_3/V_4 = 0.144) imply that the combination of 3 genes was not ideal. In the present study, an RNA-seq dataset was used for prediction and selection, and a large number of biological samples were provided for determination and validation, and multiple algorithms were introduced for evaluation. Therefore, in terms of both the number and quality of HKGs, this study is a significant step forward from previous studies.

When studying the expression of target genes in skin tissue from other species, such as Angora rabbits [27], mink [28], mice [29], and even humans [30], ACTB or GAPDH are generally used as the HKGs. The selection of HKGs from neighboring species has been widely recognized and accepted. Therefore, the data presented here can describe HKGs in skin tissue research, not only for goats, but likely also for other relative species.

The selection and validation of HKGs based on RNA-seq data

Selection and validation of HKGs by the approach of RNA-seq produced results that were more reproducible, had greater sensitivity, and yielded better correlation with protein expression levels, in addition to more accurate detection and higher coverage [31]. To our knowledge, this study is the first report on the selection and validation of suitable HKGs for qRT-PCR analysis in goats. Two novel HKGs (NCBP3 and PTPRA) and an known HKG (SDHA) belonged to the NCBP3 + SDHA + PTPRA combination that was finally recommended. Consistent with other studies [19, 22, 23], new and improved HKGs were identified through the approach of analyzing an RNA-seq dataset. In total, 1325 candidate genes were identified that met the criteria of HKGs (Fig. 1E). Eight of them were selected for additional determination and validation analyses. While this study demonstrated the advantages of using RNA-seq datasets in the discovery of new HKGs, it is also possible that the prediction of HKGs by RNA-seq datasets may be lacking in some respects. For example, the ranking order of these candidate HKGs (Table 1) and the determined final score (Table 5) did not match (compared in Additional file 1: Table S3). Specifically, the CV values of RRAGA and SRP68 were in the top 3, although in the final ranking they did not appear in the top 3 positions of any group (groups 1–4 and all samples). This may be the reason the RNA-seq samples in the selection stage did and the qRT-PCR samples used in the determination stage did not completely overlap. This phenomenon is consistent with those reported by Gao et al. [22]. In any case, this phenomenon implies that the HKGs predicted by RNA-seq screening were not completely reliable, and needs further confirmation and validation by qRT-PCR experiments.

As said before, the ACTB and GAPDH genes are currently the most popular HKG in the literature, but their limited normalizing capacity was verified here. This suggests that scientists must be cautious when selecting traditional HKGs, especially to identify the target genes that have slight changes in expression. Therefore, it is recommended that common HKGs be included as a comparison, to provide direct evidence. Of course, it must also be acknowledged that mining reliable HKGs requires scientific experimental design, complete experimental materials, more algorithmic tools, and a certain amount of scientific research time and funding [26], though these are not available to every research laboratory. Therefore, it is recommended that those experimental systems that do not meet the above conditions search for HKGs in close species on the ICG platform [32], then use as many HKGs as possible and geometric them as a normalization factor to increase the experimental stability, instead of blindly using a single HKG such as ACTB or GAPDH.

ComprFinder algorithm

The requirement of comprehensive analysis, and the shortcomings of the previous algorithms

After evaluating candidate HKGs with the above-mentioned algorithms (geNorm [33], NormFinder [34], BestKeeper [35], and the ΔCt method [36]), it is not surprising that the rankings of candidate genes may vary depending on the algorithm used [26]. Thus, another algorithm is needed for comprehensive ranking. After reviewing the literature on HKGs, it was determined that 2 types of comprehensive algorithms were predominantly used: (1)

according to the primary ranking order, then calculates the arithmetic average to get the final ranking [21, 37]; (2) according to the primary ranking order, then calculates the geometric average to get the final ranking. RefFinder [38] is a typical representative (https://www.heartcure.com.au/reffinder /? type = reference). Many studies [7, 19, 22, 24, 39–41] used this RefFinder algorithm (Times Cited: 352, on Web of Science, 2020/2/3), which illustrates the great impact of RefFinder. For both of the above types, they must depend on the ranking number of the original algorithms. Because of this, it is possible that the use of ranking numbers may cause some errors.

These ranking numbers reflect the true size of stability values of these candidate HKGs, but they should not be used as the input numerical value for the next calculation. Doing so would excessively reduce or enlarge the real differences among them. Taking our experimental data as an example, these candidate HKGs were evaluated and unevenly distributed (Fig. 6A) on the axis of the four algorithms. The RefFinder algorithm provides them with the uniform rank of 1–12 (Fig. 6B), and then calculates the geometric average for each candidate HKG. The RefFinder algorithm will increase the gap between PTPRA and EIF4H, and reduce the gap between CNBP and ACTB (the axis of the geNorm algorithm in Fig. 6A and Fig. 6B). We consider this unreasonable for the comprehensive evaluation. Therefore, in this study, a new algorithm for comprehensive analysis was developed.

The ComprFinder algorithm

ComprFinder performs the next step of the integration analysis on the results obtained through the algorithms including, but not limited to, the above mentioned geNorm, BestKeeper, NormFinder, and ΔCt method. The schematic diagram of the ComprFinder algorithm is presented in Fig. 7. Specifically, STEP 1, according to the original algorithm results ordered the values from small to large to find the minimum and the maximum. From this, the range was calculated. In STEP 2, the values from the original algorithm were standardized in the interval [0, 1], where the minimum value = 0 and the maximum value = 1. All of the other data were assigned normalized values between 0 and 1. This step makes these stability values that belong to the same HKG but from different original algorithms abide by the additive property. In STEP 3, arithmetic averages of the standardized values for each candidate HKG were calculated, and then these final scores were sorted to obtain their final rankings (Fig. 7).

ComprFinder was intentionally developed to replace the previously used RefFinder. The comparison of ComprFinder and RefFinder results were shown in Additional file 1: Table S4. It can be found that the relative final scores and ranking order of the 12 candidate HKGs calculated using the two algorithms are different. For example, NCBP3(0.096), SDHA(0.099), and PTPRA(0.108) were the top 3 genes for all samples calculated by ComprFinder, whereas PTPRA(2.14), SDHA(2.45), and EEF2(3.13) calculated by RefFinder. The ComprFinder algorithm directly standardizes the results of the original algorithms, unlike the RefFinder, which uses the ranking numbers of original algorithms. Using the ranking numbers, the real differences among the candidate HKGs will excessively reduce or increase. Therefore, the new comprehensive algorithm can overcome the intrinsic errors caused by artificial assignment. Standardized processing allows the results of different algorithms to have the same dimension, and make them essentially comparable. Finally, the standardized results can be integrated to get a series of scores, and the final evaluation. Therefore, ComprFinder is a more reasonable algorithm than RefFinder for the comprehensive evaluation of HKGs.

We provide a ComprFinder algorithm tool (Additional file 2) for researchers who have comprehensive evaluation needs for candidate HKGs. This tool is based in Microsoft Excel, which can be downloaded from the supplementary materials of this article. Briefly, after inputting the original algorithm results into the input area, the ComprFinder algorithm automatically processes the data. Then, all candidate HKGs will be scored and presented

in the output area. Although the use of ComprFinder in this study was based on geNorm, BestKeeper, NormFinder, and the ΔCt method, the analysis is not limited to these 4 algorithms.

Conclusion

In this study, we presented the first data of candidate HKG selection for skin tissues of goats based on transcriptome data. The NCBP3 + SDHA + PTPRA combination was identified and finally recommended as the triplet HKGs for skin molecular biology studies in goats and other closely related species. In addition, a comprehensive algorithm tool, ComprFinder, was developed for comprehensive evaluation of candidate HKGs.

Methods

Animals and skin tissue samples

All animals and sampling procedures in this experiment were supervised and approved by the Institutional Animal Care and Use Committee of Southwest University. The Inner Mongolia cashmere goats (IMCG) were kept in Breeding Farm, Southwest University (Chongqing, China) and two private farms in Inner Mongolia Autonomous Region, China. Dazu black goat (DBG), Hechuan white goat (HCWG), F1 of IMCG and DBG were kept in Breeding Farm, Southwest University. Each 1 cm2 skin sample was collected after commercial slaughter of animal and flash frozen in liquid nitrogen, or stored in an RNA/DNA sample protection reagent (Takara, Dalian, China). Some skin samples were collected from the experimental animals after anesthesia(intravenous injection, the anesthetic from Baite Company, Changsha, China). All samples were stored at − 80 °C until the time of analysis.

A total of 48 skin tissue samples were collected for the determination and validation of potential HKGs. In the determination stage of this study (Fig. 8C), all samples were collected from does. Four groups (factors) were sampled, including age (4 development stages, group 1), sampling time (3-stage of hair follicle cycle, group 2), breed (4 different breeds, group 3), and sampling site (5 different sampling sites on the body of the goat, group 4). As shown in Fig. 8C, group 1 including F_1_P0, F_1_P60, F_1_P240 and F_1_Adult, which sampled 0-day, 2 months, 8 months and 2 years after birth from F_1. Group 2 included IMCG_T1, IMCG_T2, and IMCG_T3, which were sampled during the anagen (September), catagen (December) and telogen (March) from IMCG. Group 3 consisted of animals of 4 breeds, including IMCG, DBG, F_1, and HCWG. Skin samples from each of the breeds were sampled in the anagen phase of the hair follicle cycle. Skin biopsies collected from group 4 (#4, #5, #6, #12, and #14 from IMCG), were sampled from the forearm, dorsal chest, lateral chest, thigh, and the inner side of the forearm, respectively. Except for samples belonging to #4, #5, #12 or #14, all other samples analyzed here were collected from the lateral chest of the goat body. Except for samples belonging to F1_P0, F1_P60 or F1_P240, all other samples were collected from adult goats. In the validation stage of this study (Fig. 8E), 3 bucks of IMCG were added to group 2, to enhance the accuracy of the validation.

RNA isolation and cDNA synthesis

Total RNA was extracted using the RNAiso Plus kit (#9109, TaKaRa, China) according to the manufacturer's instructions. The concentration and purity were determined using the Nanodrop2000 (Thermo, USA) with the 260/280 ratios being between 1.8 and 2.0, and the 260/230 ratios were greater than 1.6 in all analyzed RNA samples. First-strand cDNA was synthesized using the 5X All-In-One RT MasterMix (with AccuRT Genomic DNA Removal Kit) (#G492, ABM, Canada) and 1:4 volumes of DEPC water was added to dilute the samples.

Selection of candidate HKGs

Transcriptome sequencing data of 39 goat skin tissues (unpublished data, Fig. 8A) was performed using the Illumina paired-end sequencing technology on an Illumina Hi-Seq™ 2000 platform. After assembly and annotation, the gene expression profiles and read counts of unigenes were converted into fragments per kilobase of exon model per million mapped reads (FPKM values). Based on the FPKM value of every gene in each transcriptome, the coefficient of variation (CV), the maximum fold change (MFC) and the dispersion measure (DPM) [25] were calculated. Genes with RPKM, CV, MFC, and DPM fulfilling the criteria of HKGs were retained for further analysis (Fig. 8B). Moreover, two HKGs (SDHA and YWHAZ) from a previous study by Bai [17], and two commonly used HKGs (ACTB and GAPDH) were also considered as HKGs. All candidate HKGs were amplified using qRT-PCR for subsequent determination and validation. The probability density curves were drawn by self-written scripts (Additional file 3) run on MATLAB software (https://ww2.mathworks.cn/products/matlab.html). Venn diagram analysis was performed using the OmicShare online platform tools (http://www.omicshare.com/tools).

Primer design and amplification efficiency analysis

Specific primers were designed using the NCBI web tool based on the sequences of the unigenes. The criteria for primer design were as follows: primer lengths of 17–24 bp, GC content of 50%– 66%, melting temperature of 58–61 °C, and amplicon lengths of 100–200 bp. All primers were synthesized by the Beijing Genomics Institute (Beijing, China).

qRT-PCR analysis

Sample reactions were performed in a 10 μL reaction volume with 5 μL of 2 × qPCR MasterMix (#MasterMix-S, ABM, Canada), 1 μL cDNA template, and 0.3 μL each primer, and 3.4 μL DNase/RNase-free water. The reaction was run on the Bio-Rad CFX96 Real-Time PCR Detection System. The thermal cycling conditions were conducted according to the reagent kit instructions as follows: enzyme activation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing/extension at 60 °C for 60 s. The specificity of the SYBR green PCR signal was confirmed by melting curve analysis. The samples consisted of 4 experimental groups with 3 biological replicates in every level of each group, and 3 technical replicates for each sample (Fig. 8C). All qRT-PCR experiments and data analyses in the present study were performed in accordance with the MIQE guidelines [42, 43].

Determination of expression stability of HKGs

The cycle threshold (Ct) data of all candidate HKGs obtained from the qRT-PCR experiments were determined by 4 algorithms which including geNorm [33], NormFinder [34], BestKeeper [35], and the ΔCt method [36](Fig. 8D). After using the above-mentioned traditional evaluation algorithms, another algorithm was developed for comprehensive ranking. The new algorithm, ComprFinder, standardizes the output values from the above 4 algorithms, and then arithmetically averages them to get the final score and final ranking order.

Experimental validation of the HKGs

In order to verify the results, the 3 best candidate HKGs were selected, including NCBP3, SDHA, and PTPRA. As well as the most unstable and most commonly used HKGs including ACTB and GAPDH. Next, the HKGs were verified and evaluated using 3 target genes, which are the most important genes in hair follicle research of goats,

including DKK1, SHH, and FGF5 (Fig. 8E). Considering the requirement of accuracy evaluate in their expression, 3 new bucks were included to the original sample size of 3 does, (6 adult IMCG, 3♀ and 3♂) were sampled over 3 time-points (T1, T2, and T3). The qRT-PCR was conducted as described above. The paired sample t-test was performed using Microsoft Excel, and the graph was plotted using GraphPad Prism 6. The results are presented as Mean ± SEM, * $P < 0.05$, ** $P < 0.01$. For multiple gene combinations, the geometric average of their Ct value was calculated [33]. The relative gene expression level was calculated as $2^{-\Delta Ct}$, $\Delta Ct = \Delta (Ct_{target\ gene}\text{-}Ct_{HKGs})$.

To further evaluate the internal relationship of these candidate HKGs, a correlation analysis was performed. First, the target genes were normalized by different HKGs or HKG combinations. Their copy numbers were converted to relative expression levels with $2^{-\Delta Ct}$, and then their normalized-based expression levels were examined by the correlation analysis.

Abbreviations

ACTB:Actin Beta; CNBP:CCHC-Type Zinc Finger Nucleic Acid Binding Protein; Ct:Cycle threshold value; CTBP2:C-Terminal Binding Protein 2; CV:Coefficient of variation; DBG:Dazu black goat; DKK1:Dickkopf WNT Signaling Pathway Inhibitor 1; DPM:Dispersion measure; EEF2:Eukaryotic Translation Elongation Factor 2; EIF4H:Eukaryotic Translation Initiation Factor 4H; F1:First filial generation; FGF5:Fibroblast Growth Factor 5; FPKM:Fragments per Kilobase per Million; GAPDH:Glyceraldehyde-3-Phosphate Dehydrogenase; HCWG:Hechuan white goat; HKGs:House-keeping genes; IMCG:Inner Mongolia cashmere goat; MFC:Maximum fold change; NCBP3:Nuclear Cap Binding Subunit 3; PTPRA:Protein Tyrosine Phosphatase Receptor Type A; qRT-PCR:Quantitative real-time polymerase chain reaction; RNA-seq:RNA Sequencing; RRAGA:Ras Related GTP Binding A; SDHA:Succinate Dehydrogenase Complex Flavoprotein Subunit A; SHH:Sonic Hedgehog Signaling Molecule; SRP68:Signal Recognition Particle 68; YWHAZ:Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta.

Declarations

Ethics approval and consent to participate

All animals and sampling procedures in this experiment were supervised and approved by the Institutional Animal Care and Use Committee of Southwest University.

Consent for publication

Not applicable.

Availability of data and materials

The RNA-seq dataset analyzed during the current study is not publicly available as it belongs to another unpublished study, but is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (No.31772564), National Key R&D Program of China (2018YFD0502003), and the Chongqing Technology Innovation and Application Development Special Project (No. cstc2019jscx-lyjsAX0018). The funders had no role in the design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

ZJP and DCC designed the experiments. ZJP and LJL fed animals, collected the experimental tissues, and extracted the RNA. ZJP performed all qRT-PCR experiments, analyzed and interpreted the data, drafted the manuscript. ZYJ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgment

The authors would like to thank Tao Li (Harbin Institute of Technology at Weihai, School of Information Science and Engineering) for revising the Matlab script.

References

1. Erickson HS, Albert PS, Gillespie JW, Rodriguez-Canales J, Linehan WM, Pinto PA, Chuaqui RF, Emmert-Buck MR: **Quantitative RT-PCR gene expression analysis of laser microdissected tissue samples**. *Nature Protocols* 2009, **4**(6):902-922.
2. Bustin SA, Benes V, Nolan T, Pfaffl MW: **Quantitative real-time RT-PCR - a perspective**. *J Mol Endocrinol* 2005, **34**(3):597-601.
3. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative C-T method**. *Nature Protocols* 2008, **3**(6):1101-1108.
4. Huggett J, Dheda K, Bustin S, Zumla A: **Real-time RT-PCR normalisation; strategies and considerations**. *Genes Immun* 2005, **6**(4):279-284.
5. Zhang L, Dai Z, Yu J, Xiao M: **CpG-island-based annotation and analysis of human housekeeping genes**. *Briefings in bioinformatics* 2020.
6. Kloubert V, Rink L: **Selection of an inadequate housekeeping gene leads to misinterpretation of target gene expression in zinc deficiency and zinc supplementation models**. *J Trace Elem Med Bio* 2019, **56**:192-197.
7. Hu X, Zhang L, Nan S, Miao X, Yang P, Duan G, Fu H: **Selection and validation of reference genes for quantitative real-time PCR in Artemisia sphaerocephala based on transcriptome sequence data**. *Gene* 2018, **657**:39-49.
8. Brown AJ, Gibson S, Hatton D, James DC: **Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data**. *Biotechnology journal* 2018, **13**(1).

9. Bustin SA: **Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems**. *J Mol Endocrinol* 2002, **29**(1).
10. Eisenberg E, Levanon EY: **Human housekeeping genes, revisited**. *Trends in Genetics* 2013, **29**(10):569-574.
11. Waldron S, Brown C, Komarek AM: **The Chinese Cashmere Industry: A Global Value Chain Analysis**. *Dev Policy Rev* 2014, **32**(5):589-610.
12. Zhang Y, Wu K, Wang L, Wang Z, Han W, Chen D, Wei Y, Su R, Wang R, Liu Z *et al*: **Comparative study on seasonal hair follicle cycling by analysis of the transcriptomes from cashmere and milk goats**. *Genomics* 2019.
13. Rile N, Liu Z, Gao L, Qi J, Zhao M, Xie Y, Su R, Zhang Y, Wang R, Li J *et al*: **Expression of Vimentin in hair follicle growth cycle of inner Mongolian Cashmere goats**. *Bmc Genomics* 2018, **19**.
14. Su R, Fan Y, Qiao X, Li X, Zhang L, Li C, Li J: **Transcriptomic analysis reveals critical genes for the hair follicle of Inner Mongolia cashmere goat from catagen to telogen**. *Plos One* 2018, **13**(10).
15. He N, Dong Z, Zhu B, Nuo M, Bou S, Liu D: **Expression of pluripotency markers in Arbas Cashmere goat hair follicle stem cells**. *In Vitro Cellular & Developmental Biology-Animal* 2016, **52**(7):782-788.
16. Guo HY, Cheng GH, Li YJ, Zhang H, Qin KL: **A Screen for Key Genes and Pathways Involved in High-Quality Brush Hair in the Yangtze River Delta White Goat**. *Plos One* 2017, **12**(1).
17. Bai WL, Yin RH, Yin RL, Jiang WQ, Wang JJ, Wang ZY, Zhu YB, Zhao ZH, Yang RJ, Luo GB *et al*: **Selection and validation of suitable reference genes in skin tissue of Liaoning cashmere goat during hair follicle cycle**. *Livestock Science* 2014, **161**:28-35.
18. Dos Santos KCG, Desgagne-Penix I, Germain H: **Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis**. *BMC genomics* 2020, **21**(1):35-35.
19. Zhao Z, Zhang Z, Ding Z, Meng H, Shen R, Tang H, Liu Y-G, Chen L: **Public-transcriptome-database-assisted selection and validation of reliable reference genes for qRT-PCR in rice**. *Science China-Life Sciences* 2019.
20. Li Y, Han J, Wu J, Li D, Yang X, Huang A, Bu G, Meng F, Kong F, Cao X *et al*: **Transcriptome-based evaluation and validation of suitable housekeeping gene for quantification real-time PCR under specific experiment condition in teleost fishes**. *Fish & shellfish immunology* 2020, **98**:218-223.
21. Pombo MA, Ramos RN, Zheng Y, Fei Z, Martin GB, Rosli HG: **Transcriptome-based identification and validation of reference genes for plant-bacteria interaction studies using Nicotiana benthamiana**. *Scientific Reports* 2019, **9**.
22. Gao D, Kong F, Sun P, Bi G, Mao Y: **Transcriptome-wide identification of optimal reference genes for expression analysis of Pyropia yezoensis responses to abiotic stress**. *Bmc Genomics* 2018, **19**.
23. Liang W, Zou X, Carballar-Lejarazu R, Wu L, Sun W, Yuan X, Wu S, Li P, Ding H, Ni L *et al*: **Selection and evaluation of reference genes for qRT-PCR analysis in Euscaphis konishii Hayata based on transcriptome data**. *Plant Methods* 2018, **14**.
24. Jin Y, Liu F, Huang W, Sun Q, Huang X: **Identification of reliable reference genes for qRT-PCR in the ephemeral plant Arabidopsis pumila based on full-length transcriptome data**. *Scientific Reports* 2019, **9**.
25. Pan JB, Hu SC, Wang H, Zou Q, Ji ZL: **PaGeFinder: quantitative identification of spatiotemporal pattern genes**. *Bioinformatics* 2012, **28**(11):1544-1545.
26. Kozera B, Rapacz M: **Reference genes in real-time PCR**. *Journal of Applied Genetics* 2013, **54**(4):391-406.

27. Ding H, Zhao H, Cheng G, Yang Y, Wang X, Zhao X, Qi Y, Huang D: **Analyses of histological and transcriptome differences in the skin of short-hair and long-hair rabbits**. *BMC Genomics* 2019, **20**(1):140.
28. Song X, Xu C, Liu Z, Yue Z, Liu L, Yang T, Cong B, Yang F: **Comparative Transcriptome Analysis of Mink (Neovison vison) Skin Reveals the Key Genes Involved in the Melanogenesis of Black and White Coat Colour**. *Sci Rep* 2017, **7**(1):12461.
29. Song Y, Boncompagni AC, Kim SS, Gochnauer HR, Zhang Y, Loots GG, Wu D, Li Y, Xu M, Millar SE: **Regional Control of Hairless versus Hair-Bearing Skin by Dkk2**. *Cell reports* 2018, **25**(11):2981-2991 e2983.
30. Jacobs LC, Hamer MA, Gunn DA, Deelen J, Lall JS, van Heemst D, Uh HW, Hofman A, Uitterlinden AG, Griffiths CEM *et al*: **A Genome-Wide Association Study Identifies the Skin Color Genes IRF4, MC1R, ASIP, and BNC2 Influencing Facial Pigmented Spots**. *The Journal of investigative dermatology* 2015, **135**(7):1735-1742.
31. Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li YX, Zeng R *et al*: **Estimating accuracy of RNA-Seq and microarrays with proteomics**. *Bmc Genomics* 2009, **10**.
32. Sang J, Wang ZN, Li M, Cao JB, Niu GY, Xia L, Zou D, Wang F, Xu XJ, Han XJ *et al*: **ICG: a wiki-driven knowledgebase of internal control genes for RT-qPCR normalization**. *Nucleic Acids Res* 2018, **46**(D1):D121-D126.
33. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. *Genome Biology* 2002, **3**(7).
34. Andersen CL, Jensen JL, Orntoft TF: **Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets**. *Cancer Res* 2004, **64**(15):5245-5250.
35. W. P, Tichopad A, Prgomet C, Neuvians TP: **Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations**. *Biotechnol Lett* 2004, **26**(6):509-515.
36. Silver N, Best S, Jiang J, Thein SL: **Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR**. *Bmc Molecular Biology* 2006, **7**.
37. Liang L, He Z, Yu H, Wang E, Zhang X, Zhang B, Zhang C, Liang Z: **Selection and Validation of Reference Genes for Gene Expression Studies in Codonopsis pilosula Based on Transcriptome Sequence Data**. *Scientific reports* 2020, **10**(1):1362-1362.
38. Xie FL, Xiao P, Chen DL, Xu L, Zhang BH: **miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs**. *Plant Molecular Biology* 2012, **80**(1):75-84.
39. Zhou X, Wu X, Chu M, Liang C, Ding X, Pei J, Xiong L, Bao P, Guo X, Yan P: **Validation of Suitable Reference Genes for Gene Expression Studies on Yak Testis Development**. *Animals* 2020, **10**(2).
40. Dos Santos CP, da Cruz Saraiva KD, Batista MC, Germano TA, Costa JH: **Identification and evaluation of reference genes for reliable normalization of real-time quantitative PCR data in acerola fruit, leaf, and flower**. *Molecular biology reports* 2020, **47**(2):953-965.
41. Feng K, Liu J-x, Xing G-M, Sun S, Li S, Duan A-Q, Wang F, Li M-Y, Xu Z-S, Xiong A-S: **Selection of appropriate reference genes for RT-qPCR analysis under abiotic stress and hormone treatment in celery**. *Peerj* 2019, **7**.
42. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL *et al*: **The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments**. *Clin Chem* 2009, **55**(4):611-622.

43. Bustin SA: **Why the need for qPCR publication guidelines?-The case for MIQE**. *Methods* 2010, **50**(4):217-226.

Supplementary Files Legend

Supplementary Figure and Tables
Figure S1 Melting curves for the 12 candidate HKGs and 3 target genes
Figure S2 Optimal number of HKGs in different experimental groups calculated by geNorm
Table S1 Primer sequences and amplicon information of candidate HKGs and target genes for qRT-PCR Table S2 Ct values of the 12 candidate HKGs in all samples
Table S3 The comparison of experimental results and RNA-seq data
Table S4 The comparison of final results using ComprFinder and RefFinder algorithms

Figures

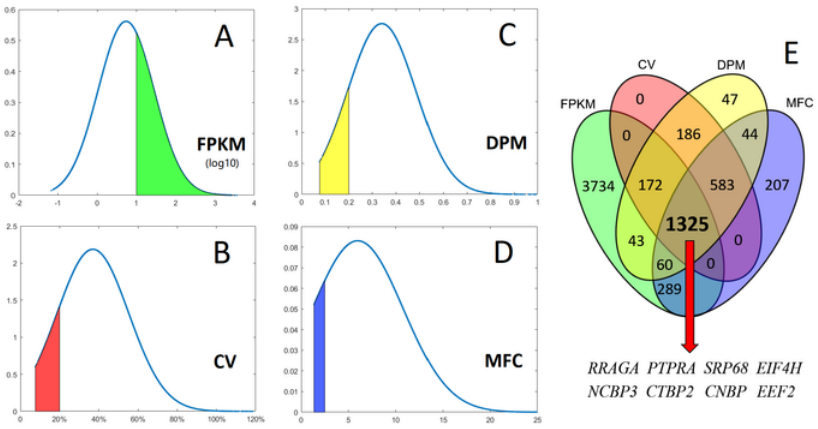

Figure 1 Probability density curve of FPKM, CV, DPM and MFC of 15853 unigenes
The y-axes indicate the probability values, and x-axes indicate FPKM(A, $\log_{10}$FPKM), CV (B), DPM (C) and MFC (D)

Figure 1

Probability density curve of FPKM, CV, DPM and MFC of 15853 unigenes

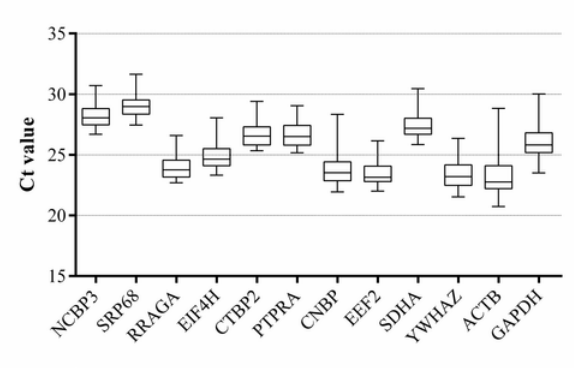

Figure 2 Boxplot of absolute Cq value of the 12 candidate genes in all skin tissue samples
Boxes indicated median (Q2) and quartiles first and third (Q1 and Q3) and whiskers corresponded to the minimum and maximum values.

Figure 2

Boxplot of absolute Cq value of the 12 candidate genes in all skin tissue samples

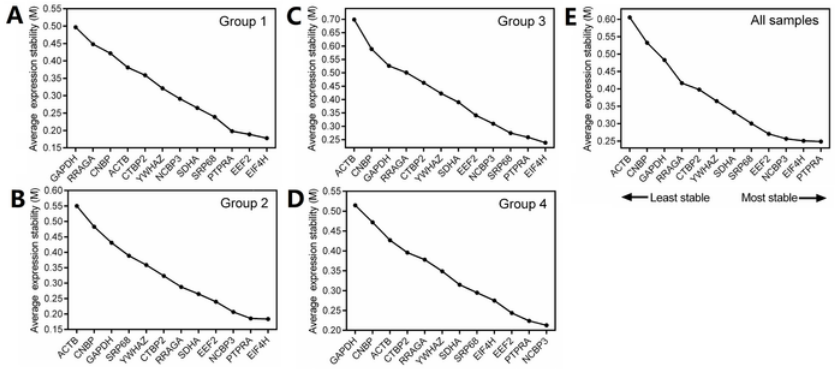

Figure 3 Average expression stability (M-value) calculated by geNorm
(A) Group 1, 4 different development stages; (B) Group 2, 3 time-points in hair follicle cycle; (C) Group 3, 4 goat breeds; (D) Group 4, 5 sampling sites on the body of the goat. (E) All samples including roups 1-4.

Figure 3

Average expression stability (M-value) calculated by geNorm

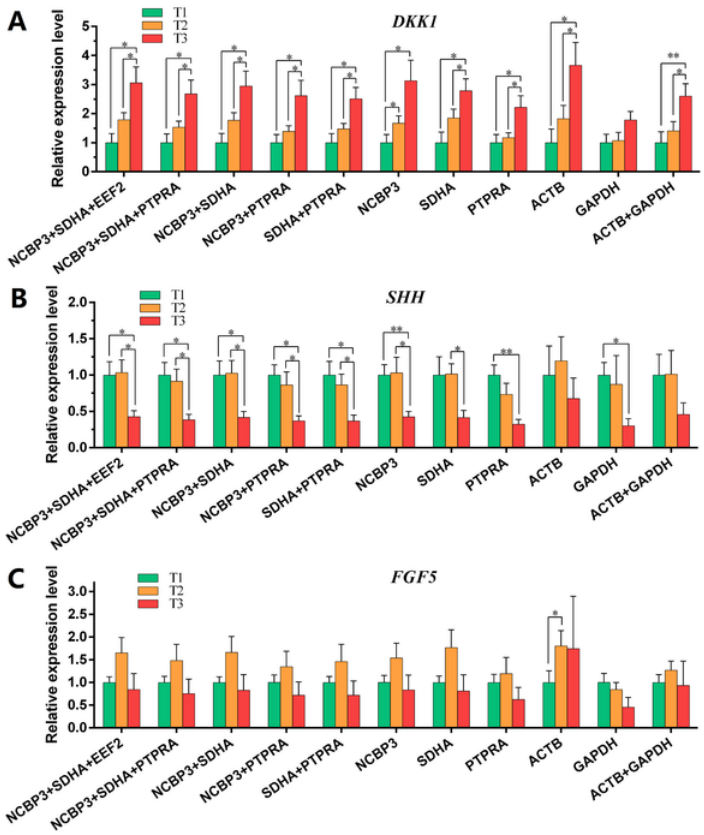

Figure 4 Relative expression levels normalized by 11 types of single or multiple gene combinations of HKGs
Expression of *DKK1*(A), *SHH*(B), and *FGF5*(C) were normalized by the most stable single or multiple gene combinations (*NCBP3+SDHA+EEF2*, *NCBP3+SDHA+PTPRA*, *NCBP3+SDHA*, *NCBP3+PTPRA*, *SDHA+PTPRA*, *NCBP3*, *SDHA*, *PTPRA*,) and the most unstable single or multiple genes combination (*ACTB*, *GAPDH*, *ACTB+GAPDH*). The error bars represent the SEM, and the paired t-test in any two stages, (**P* <0.05, ***P* <0.01, n=6) for each hair follicle cycle time-point of IMCG. T1, T2, and T3 indicate the anagen, catagen, and telogen, respectively.

Figure 4

Relative expression levels normalized by 11 types of single or multiple gene combinations of HKGs

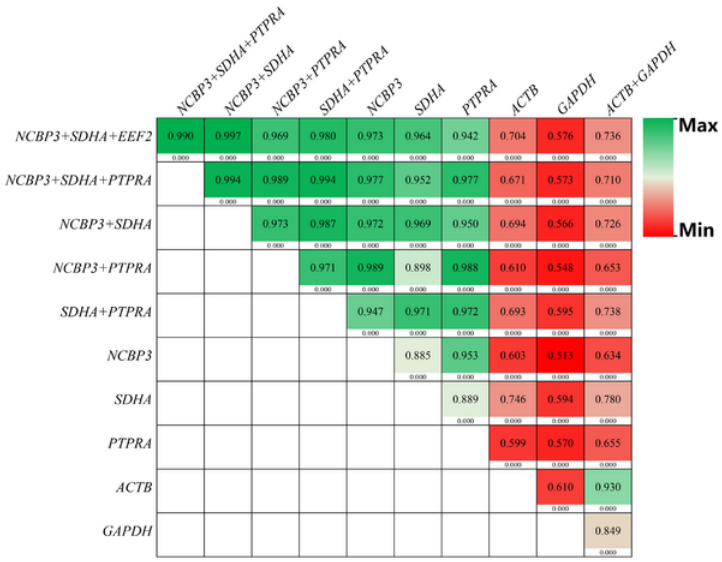

Figure 5 Heat map of correlation coefficients of relative expression levels based on different normalized HKGs
Three target genes were detected in 18 tissue samples and normalized by different types of HKGs. The number in each color block is correlation coefficient (R value), and the number below the color block is the *P* value of the corresponding R value.

Figure 5

Heat map of correlation coefficients of relative expression levels based on different normalized HKGs

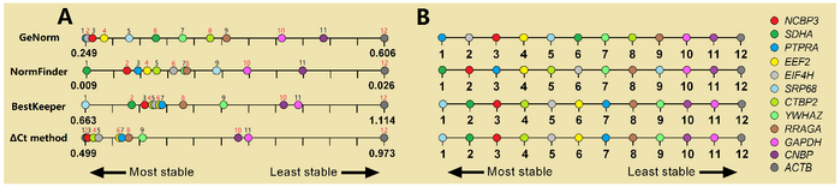

Figure 6 Gene stability values and rank order
The same color presents the same gene. (A) Tweleve candidate HKGs unevenly distributed on the axis, this is their true distribution; (B) Tweleve candidate HKGs evenly distributed on the axis ordered by 1-12.

Figure 6

Gene stability values and rank order

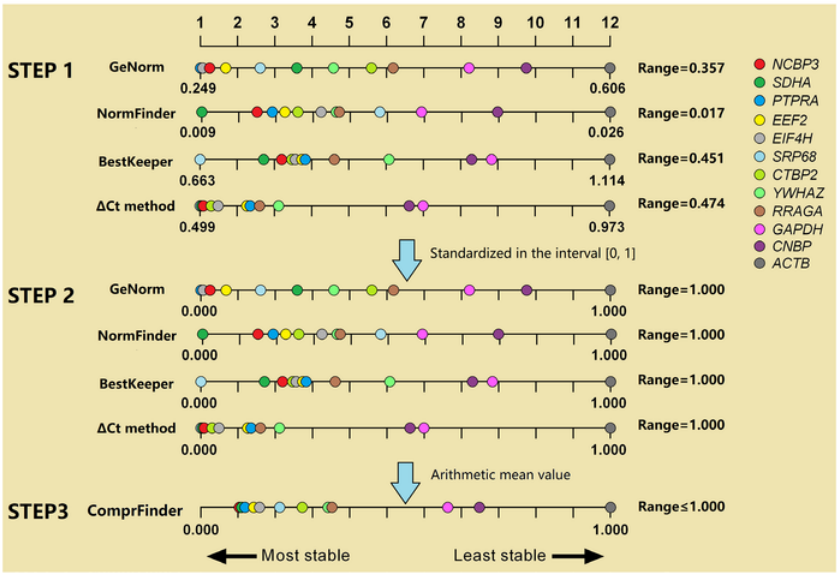

Figure 7 Schematic diagram of the ComprFinder algorithm
The same color presents the same gene. From step 1 to step 2, the uneven distribution of these 12 HKGs will not change, but they are proportionally enlarged or reduced to the range of 0-1. Calculated the arithmetic mean value for each gene to determine the final ranking order.

Figure 7

Schematic diagram of the ComprFinder algorithm

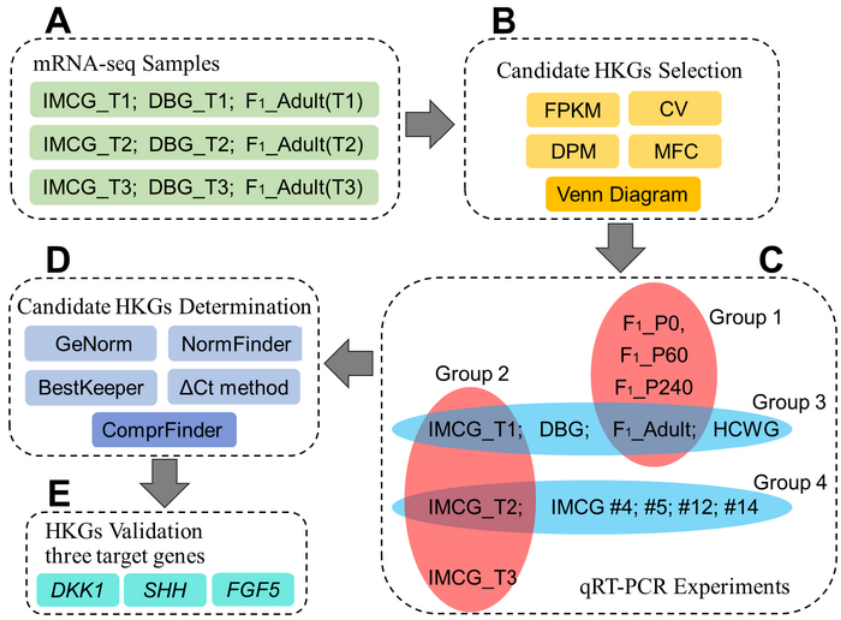

Figure 8 The workflow of this study

(A) The sample information of transcriptome sequence data of these IMCG (n=3*3), DBG (n=3*3) and F1_Adult (n=7*3); **(B)** Candidate housekeeping genes were preliminarily selected by four indicators which including FPKM, CV, DPM, and MFC, and were further selected by Venn diagram analysis; **(C)** The sample information of the qRT-PCR experiments on the 4 experimental groups, with 3 biological replicates in every level of each group. Group 1, different development stages (n=4*3); Group 2, hair follicle cycle stages (n=3*3); Group 3, breeds (n=3*3); Group 4, sampling sites (n=5*3). **(D)** Candidate housekeeping genes were determined using 4 algorithms, including GeNorm, NormFinder, BestKeeper and the ΔCt method. An additional comprehensive analysis was conducted using ComprFinder, a new algorithm developed by the authors. **(E)** The selected HKGs were validated, yielding 3 target genes.

Figure 8

The workflow of this study

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- renamed43c40.pdf
- Matlabscript.m

- DownloadableExcelfilewiththeComprFinderalgorithm.xlsx