

RESEARCH ARTICLE

De Novo Assembly of the Donkey White Blood Cell Transcriptome and a Comparative Analysis of Phenotype-Associated Genes between Donkeys and Horses

Feng-Yun Xie^{1,2,3}, Yu-Long Feng^{3,4}, Hong-Hui Wang^{1,2,3}, Yun-Feng Ma^{3,4}, Yang Yang³, Yin-Chao Wang⁴, Wei Shen^{1,2,3}, Qing-Jie Pan^{2,3}, Shen Yin^{1,2,3}, Yu-Jiang Sun³, Jun-Yu Ma^{1,2,3*}

1 Institute of Reproductive Science, Qingdao Agricultural University, Qingdao, Shandong, 266109, China, **2** Key Laboratory of Animal Reproduction and Germplasm Enhancement in Universities of Shandong, Qingdao Agricultural University, Qingdao, Shandong, 266109, China, **3** College of Animal Science and Technology, Qingdao Agricultural University, Qingdao, Shandong, 266109, China, **4** Black Donkey Research Institute, Shandong Donggejiao Company Limited, Liaocheng, Shandong, 252000, China

* jun_yu_ma@aliyun.com



OPEN ACCESS

Citation: Xie F-Y, Feng Y-L, Wang H-H, Ma Y-F, Yang Y, Wang Y-C, et al. (2015) *De Novo* Assembly of the Donkey White Blood Cell Transcriptome and a Comparative Analysis of Phenotype-Associated Genes between Donkeys and Horses. PLoS ONE 10 (7): e0133258. doi:10.1371/journal.pone.0133258

Editor: Peng Xu, Chinese Academy of Fishery Sciences, CHINA

Received: April 28, 2015

Accepted: June 24, 2015

Published: July 24, 2015

Copyright: © 2015 Xie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All raw data files are available from the Sequence Read Archive database (accession number SRX973470).

Funding: Funding was provided to Yu-Jiang Sun from Breeding Projects in Shandong Province. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Prior to the mechanization of agriculture and labor-intensive tasks, humans used donkeys (*Equus africanus asinus*) for farm work and packing. However, as mechanization increased, donkeys have been increasingly raised for meat, milk, and fur in China. To maintain the development of the donkey industry, breeding programs should focus on traits related to these new uses. Compared to conventional marker-assisted breeding plans, genome- and transcriptome-based selection methods are more efficient and effective. To analyze the coding genes of the donkey genome, we assembled the transcriptome of donkey white blood cells *de novo*. Using transcriptomic deep-sequencing data, we identified 264,714 distinct donkey unigenes and predicted 38,949 protein fragments. We annotated the donkey unigenes by BLAST searches against the non-redundant (NR) protein database. We also compared the donkey protein sequences with those of the horse (*E. caballus*) and wild horse (*E. przewalskii*), and linked the donkey protein fragments with mammalian phenotypes. As the outer ear size of donkeys and horses are obviously different, we compared the outer ear size-associated proteins in donkeys and horses. We identified three ear size-associated proteins, HIC1, PRKRA, and KMT2A, with sequence differences among the donkey, horse, and wild horse loci. Since the donkey genome sequence has not been released, the *de novo* assembled donkey transcriptome is helpful for preliminary investigations of donkey cultivars and for genetic improvement.

Introduction

For thousands of years of human history, donkeys were mainly bred for farm labor or packing goods. In 2000, the world donkey population was estimated at approximately 43.5 million, but only 41 million donkeys existed in 2006 (a 5.7% reduction) [1]. China has the most donkeys, and its domestic donkey population increased from 7.4 million in 1966 to 10.923 million in 1996 [2]; however, based on the latest population survey, there were only 6.891 million donkeys in 2007 (a 37% reduction) [3]. The decreasing of donkey population in China is mainly attributed to agricultural mechanization, and partly to the slow speed of donkey breeding (single birth and long pregnancy). The accumulation of artificially selected characters important for farm labor over thousands of years are not suited to the modern demands in China for donkey meat, milk, and fur. Therefore, cultivating donkey breeds with new traits based on the demands of modern society is essential for the growth of the donkey industry.

In modern livestock breeding, genomes are important for cultivar characterization and genetic improvement. Conventional marker-assisted selection requires onerous phenotype data collection and its applications in commercial breeding have not generated desired results [4]. The cost of next-generation sequencing technology is sharply decreasing; accordingly, more livestock genomes have been released, and genome selection methods for livestock breeding have been developed [5]. In 2009, the first horse genome was released [6], and a horse-specific whole-genome single nucleotide polymorphism (SNP) chip was developed and successfully used for screening genes associated with Lavender Foal Syndrome [7]. Recently, it has been reported that a high-density horse genome SNP chip can be used to identify the genotypes of extant Perissodactyla, including donkeys [8]. However, as the horse SNP chip does not include all donkey SNPs, it cannot be directly applied to donkey breeding. In addition to a genome SNP chip, cDNA SNPs (cSNPs) or protein polymorphisms are more convenient for phenotype-genotype association analyses and for livestock breeding selection [9]. Livestock transcriptome sequencing can provide information not only regarding cSNP/protein polymorphism, but also about the expression levels of corresponding genes. As obtaining livestock cSNP data by transcriptome sequencing is faster and cheaper than obtaining genome SNP data, particularly for species without released genome sequences like the donkey, transcriptome analysis may be helpful for the initial design of breeding plans.

To investigate protein sequence differences between donkeys and horses and to link donkey genotypes to phenotypes, we assembled the donkey white blood cell transcriptome *de novo*. By BLAST searching against the public non-redundant (NR) protein database, we annotated the donkey transcriptome. These results will be helpful for preliminary investigations of donkey genotype-phenotype associations. We also linked predicted donkey protein fragments with mammalian phenotypes [10]. As the outer ears of donkey are notably bigger than those of horses or wild horses, we analyzed the outer ear morphology-associated genes in the donkey, horse, and wild horse using the predicted donkey protein fragments. This association analysis improves our understanding of the phenotypic differences between donkeys and horses.

Results

De novo assembly of the donkey white blood cell transcriptome

Using high-throughput deep sequencing of the transcriptome of white blood cells sampled from the Dezhou donkey, we obtained approximately 129 million clean reads. We used these clean reads to assemble 358,946 unigenes (average length, 1,412 nt), of which 264,714 were distinct singletons. The average sequencing depth of these distinct singletons was 59 and the average GC content of the singletons was 44.77% (S1 Dataset). By aligning these singleton unigenes

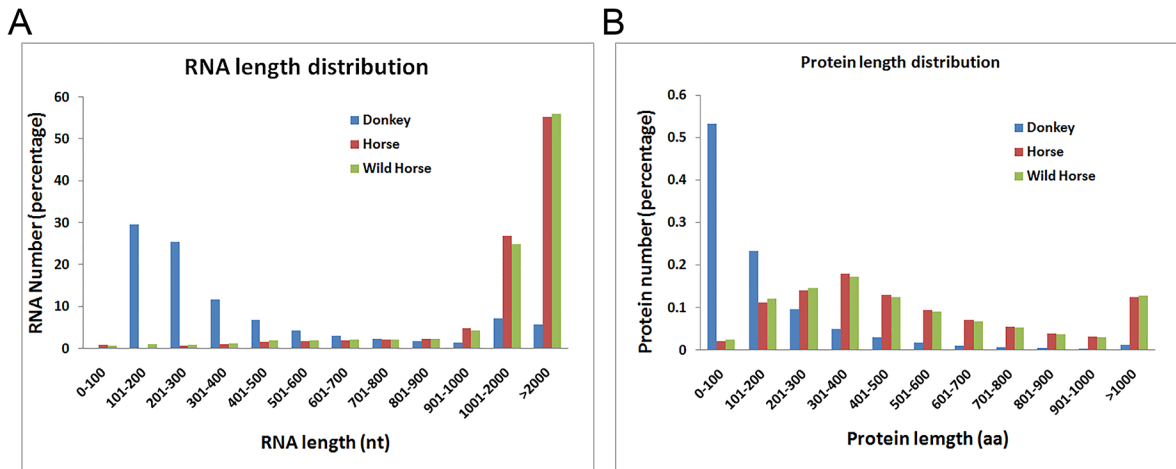


Fig 1. Length distribution of the *de novo* assembled donkey RNAs and predicted donkey protein fragments. The RNA and protein data for the horse and wild horse were downloaded from the NCBI genome database (<ftp://ftp.ncbi.nih.gov/genomes>).

doi:10.1371/journal.pone.0133258.g001

to the NR protein database using the blastx program, we predicted 38,949 unique donkey protein fragments (average length, 168 aa). As the donkey genome has not been fully sequenced and annotated, we compared the donkey unigenes and predicted proteins with data from the horse and wild horse. The length distributions of singleton unigenes and their corresponding predicted proteins are shown in Fig 1. The number and length of RNA fragments and proteins of the white blood cells of the donkey, horse, and wild horse are shown in Table 1. The total RNA length of *de novo* assembled donkey unigenes was 161.07% of the total RNA length of the horse and 122.20% of the total RNA length of the wild horse. However, the total protein length of predicted donkey protein fragments was only 35.27% of the total protein length of the horse and 29.55% of the total protein length of the wild horse, indicating that the *de novo* assembled donkey transcriptome includes a great many non-coding regions, such as untranslated regions, non-coding RNAs, or mis-assembled sequences. The assembled donkey unigenes and predicted protein fragments are listed in Supporting Information (S2 Dataset).

Analysis of predicted donkey proteins

To further analyze the predicted donkey proteins, we aligned them with horse and wild horse proteins by using the blastp program. As a result, 28,098 donkey protein fragments (72.14%) were aligned to 10,693 horse proteins (33.05%), and 28,180 donkey protein fragments (72.35%) were aligned to 11,213 wild horse proteins (29.20%). Based on the blast results, we found that the average identity between donkey proteins and horse and wild horse proteins

Table 1. Summary of the RNA/protein numbers and lengths of donkey white blood cells, horse, and wild horse.

Species (sample)	Donkey (<i>Equus africanus asinus</i>) white blood cells	Horse (<i>Equus caballus</i>)	Wild horse (<i>Equus przewalskii</i>)
RNA number	264,714	35,999	45,076
Average RNA length	594 nt	2,711 nt	2,854 nt
Total RNA length	157,214,590 nt	97,609,315 nt	128,651,487 nt
Protein number	38,949	32,352	38,416
Average protein length	168 aa	573 aa	575 aa
Total protein length	6,538,837 aa	18,540,986 aa	22,124,456 aa

doi:10.1371/journal.pone.0133258.t001

were 94.32% and 94.12%, respectively, and the average positive sequence rates were 95.79% and 95.68%.

To determine associations between donkey protein fragments and mammalian phenotypes, we annotated donkey and horse proteins using the mammalian phenotype dataset from the Mouse Genome Informatics database (www.informatics.jax.org). As a result, 10,480 distinct donkey protein fragments were classified according to their associations with mammalian phenotypes (Table 2 and S3 Dataset). Although white blood cell mRNAs cannot be used to analyze associations between all coding regions and mammalian phenotypes, Table 2 shows that white blood cell mRNAs contain more than a half of the corresponding genes underlying mammalian phenotypes such as tumorigenesis (53.99%), liver/biliary system phenotypes (52.95%), hematopoietic system phenotypes (50.96%), embryogenesis (50.31%), and cellular phenotypes (50.37%).

Comparison of outer ear morphology-associated proteins between donkeys and horses

As is commonly known, the outer ear of the donkey is longer than those of horses and wild horses. Using a mammalian phenotype database, we obtained 125 mouse genes that are associated with outer ear morphology. Based on BLAST searches of donkey, horse, and mouse proteins, we found 138 predicted donkey protein fragments (corresponding to 39 mouse proteins) associated with outer ear morphology (Fig 2). Three of these proteins (KMT2A, lysine-specific methyltransferase 2A; HIC1, hypermethylated in cancer 1; and PRKRA, protein kinase, interferon inducible double-stranded RNA dependent activator) were not conserved between horses and donkeys. We obtained multiple sequence alignments of the homologous proteins of the donkey, horse, wild horse, mouse, human, and pig. We found that the N-terminal region of the horse HIC1 protein (1–368 aa of GI: 545180679) was not conserved with those of the other species (partially displayed in Fig 3). We also found that horse and wild horse PRKRA proteins have the same 10 N-terminal amino acids, which cannot be aligned to the PRKRA locus of all other species including the donkey (Fig 4). Additionally, the horse KMT2A lacked the regions spanning amino acids 1304–1569 (region A) and 2038–2092 (region B) of donkey KMT2A. The lack of region A was horse-specific and the extra region B of donkey KMT2A was donkey-specific (S1 Fig).

Discussion

As the donkey genome has not been published, the *de novo* assembled donkey white blood cell transcriptome is useful for preliminary investigations of associations between donkey genotypes and phenotypes. In our study, we predicted 6,538,837 amino acids of donkey protein sequences (accounting for 35% of the total protein length observed in horses). Using these predicted donkey proteins, we investigated the proteins that control donkey phenotypes, such as the outer ear size. We identified 3 outer ear size-associated proteins, HIC1, KMT2A, and PRKRA, and examined sequence differences between donkeys and horses/wild horses. HIC1 is a tumor suppressor protein that suppresses the overexpression of sirtuin 1 and maintains the activity of p53 to induce apoptosis of DNA-damaged cells [11]. HIC1-deficient mice show general developmental delay and underdeveloped outer ears [12]. The mutated N-terminal region of horse HIC1 includes the GLDLSKK motif, which mediates the transcriptional repression activity of HIC1 [13]. PRKRA is involved in cell apoptosis induced by its translation inhibition activity. It is also required for DICER1-mediated small interfering RNA production [14, 15]. In mouse embryos, *PRKRA* mRNA can be detected in the developing ear at embryonic day 12 [16]. In adult mice, *PRKRA* is expressed in all the regions of the pinna, middle ear, and cochlea

Table 2. Statistical summary of predicted donkey protein fragments associated with mammalian phenotypes.

Mammalian Phenotype Term	Mouse symbol number	Mouse symbol number 2(percentage)	Horse protein number	Donkey protein fragment number	Donkey protein length (aa)	Donkey RNA length (nt)
adipose tissue phenotype	663	289 * (43.59%)	732	885	196,803	1,704,605
behavior/neurological phenotype	2,701	954 (35.32%)	2,765	3,012	624,409	5,874,119
cardiovascular system phenotype	2,053	893 (43.50%)	2,257	2,858	609,864	6,025,705
cellular phenotype	2,702	1361 (50.37%)	3,213	4,024	901,466	8,036,005
craniofacial phenotype	1,067	386 (36.18%)	1,024	1,274	285,326	2,325,180
digestive/alimentary phenotype	1,021	393 (38.49%)	1,077	1,138	215,507	1,830,532
embryogenesis phenotype	1,451	730 (50.31%)	1,694	2,099	512,414	4,358,464
endocrine/exocrine gland phenotype	1,636	705 (43.09%)	1,846	2,108	455,282	4,297,417
growth/size/body phenotype	3,649	1494 (40.94%)	3,787	4,644	1,076,374	9,211,980
hearing/vestibular/ear phenotype	595	186 (31.26%)	562	575	143,938	1,201,663
hematopoietic system phenotype	2,549	1299 (50.96%)	2,784	4,198	928,483	8,584,817
homeostasis/metabolism phenotype	3,649	1586 (43.46%)	3,977	5,096	1,046,174	9,943,604
immune system phenotype	2,604	1287 (49.42%)	2,873	4,269	927,245	8,861,768
integument phenotype	1,699	588 (34.61%)	1,541	1,911	413,608	3,736,364
limbs/digits/tail phenotype	986	298 (33.15%)	790	891	227,258	1,696,015
liver/biliary system phenotype	899	476 (52.95%)	1,055	1,318	298,347	2,662,206
mortality/aging	4,173	1903 (45.60%)	4,576	6,156	1,400,869	12,975,182
muscle phenotype	1,070	466 (43.55%)	1,305	1,456	309,052	3,014,111
nervous system phenotype	2,689	1068 (39.72%)	3,108	3,267	702,258	6,064,300
other phenotype	160	73 (45.63%)	173	248	43,176	551,780
pigmentation phenotype	582	150 (25.77%)	371	544	127,213	1,149,165
renal/urinary system phenotype	913	372 (40.74%)	972	1,195	265,898	2,162,417
reproductive system phenotype	1,649	658 (39.90%)	1,760	2,020	462,698	3,716,326
respiratory system phenotype	1,024	426 (41.60%)	1,127	1,251	256,160	2,294,374
skeleton phenotype	1,579	639 (40.47%)	1,620	1,996	428,884	3,722,388
taste/olfaction phenotype	111	37 (33.33%)	123	122	19,594	172,984
tumorigenesis	552	298 (53.99%)	627	1,177	247,409	1,907,737
vision/eye phenotype	1,274	497 (39.01%)	1,263	1,528	289,137	2,620,264

*Mouse symbol number 2 represents the number of mouse symbols whose donkey homologous protein fragments are detected in donkey white blood cells.

doi:10.1371/journal.pone.0133258.t002

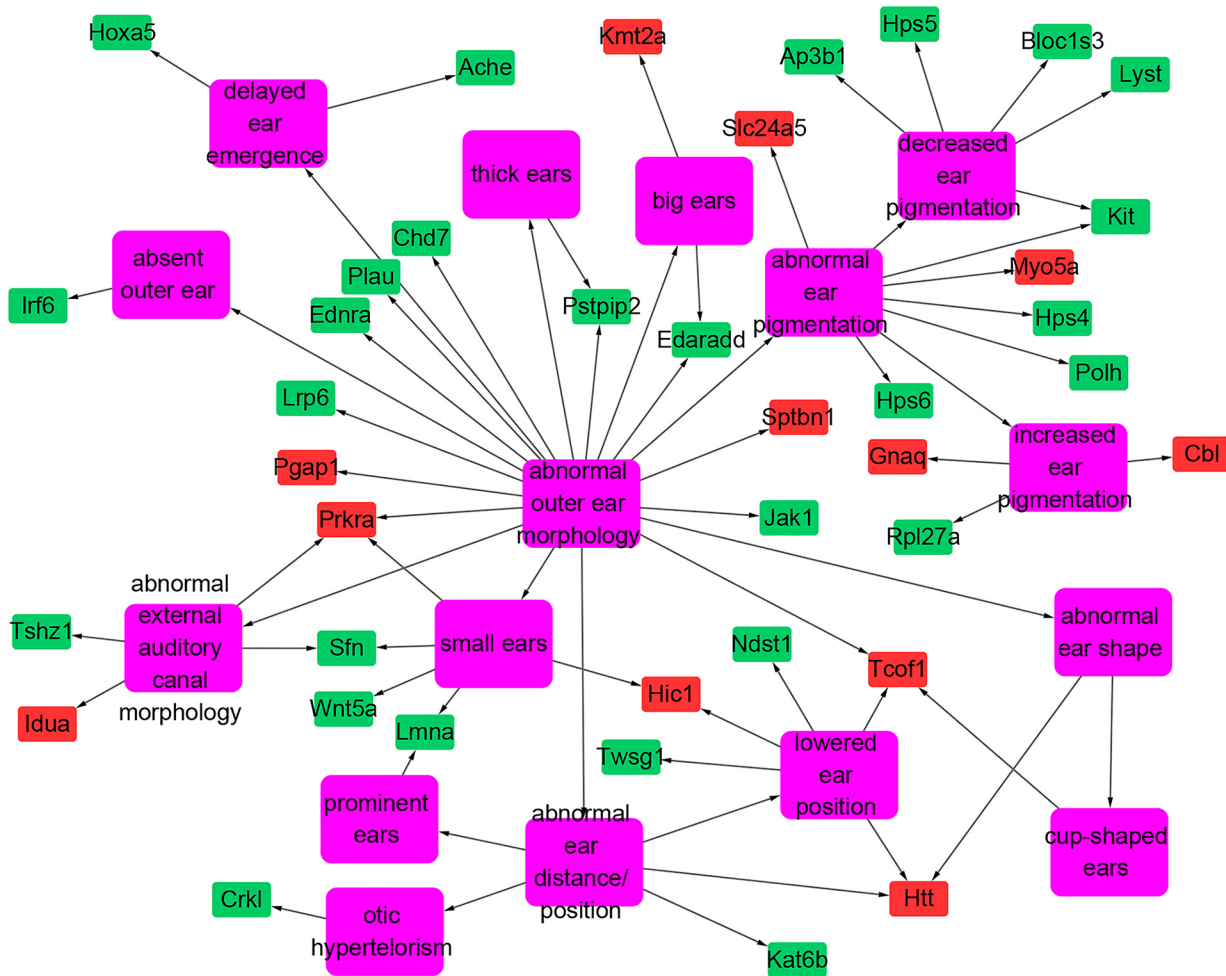


Fig 2. Outer ear morphology-associated genes. Only those genes whose homologous sequences were identified in the donkey white blood cell transcriptome are shown. Pink, mammalian phenotype term; green and red, mouse protein symbols associated with outer ear morphology. Red symbols represent the homologous donkey proteins that are not conserved between donkeys and horses.

doi:10.1371/journal.pone.0133258.g002

[16]. When PRKRA is deleted, both ear development and auditory senses are impaired [16]. The N-terminal sequence variation between horse and wild horse PRKRA and other species indicates that transcription or translation differs from that of other species. KMT2A, a myeloid/mixed lymphoid leukemia gene, is a histone H3 lysine 4-specific methyltransferase. When fused with AF4/FMR2 family member proteins like AF4, KMT2A induces lymphoid and myeloid deregulation, and even induces hematologic malignancy [17]. In addition, KMT2A-AF4 induces body growth retardation and an abnormally large outer ear [17]. We found that horse KMT2A (gi|545221883) lacks 266-aa-long fragments corresponding to region A of donkey KMT2A. However, other versions of horse KMT2A (Uniprot: F6U6A9_HORSE) and wild horse KMT2A do not lack the 266-aa fragment, indicating a sequencing error or differences among breeds.

In this study, we developed a workflow to link donkey protein sequences with mammalian phenotypes, and compared phenotype-associated proteins between donkey and horse accessions. This workflow is helpful for analyzing functional genes in donkeys, and it will improve donkey breeding. Although protein differences cannot fully explain the phenotype differences among donkeys, horses, and wild horses, our data can be used to preliminarily explore

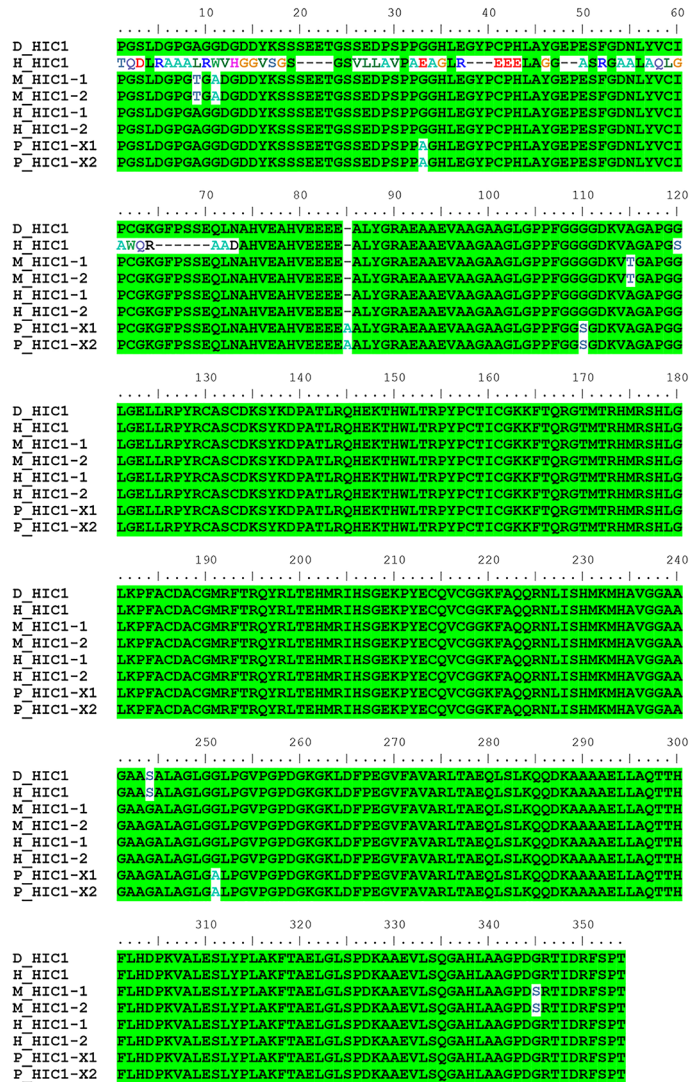


Fig 3. Multiple alignment of the HIC1 protein. The protein IDs and corresponding GenBank accession numbers are listed in [Table 3](#).

doi:10.1371/journal.pone.0133258.g003

differences in body size and other traits important for human breeding programs. In addition, there are some limitations in our study. Firstly, a gene mutation which induces a phenotype in mice may not necessarily induce the same phenotype in donkeys; secondly, as the *de novo* assembled genes cannot contain all of the associated genes of a phenotype, our data can only be used as a reference resource for donkey phenotype analysis.

In summary, we assembled the donkey white blood cell transcriptome *de novo* and linked donkey unigenes to mammalian phenotypes. We will further investigate cSNP selection methods in donkeys and identify gene markers associated with body size, milk production, and dermal thickness.

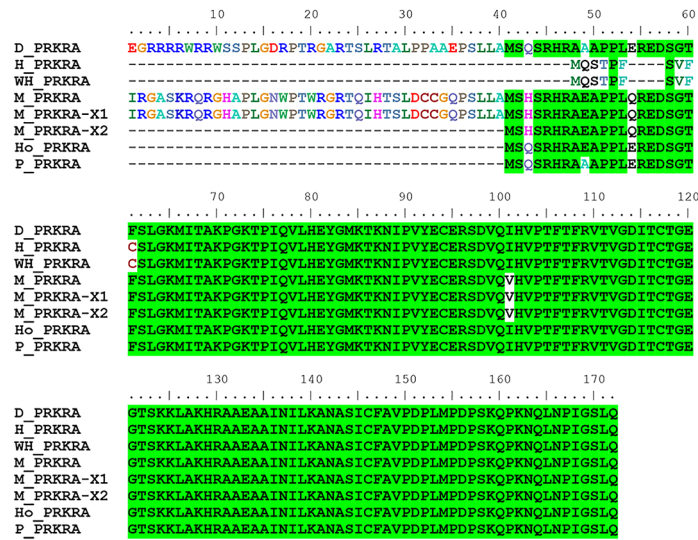


Fig 4. Multiple alignment of the PRKRA protein. The protein IDs and corresponding GenBank accession numbers are listed in [Table 3](#).

doi:10.1371/journal.pone.0133258.g004

Materials and Methods

Ethics Statement

The donkey blood isolation manipulations were performed according to the suggestions in the Guideline for Animal Experimentation in Qingdao Agricultural University. Donkey jugular vein blood (5–8 ml) was collected using disposable vacuum blood collection tubes and blood-taking needles. All animal manipulations were performed under the approval of the Animal Care and Ethics Committee of Qingdao Agricultural University.

Preparation donkey white blood cell samples

The white blood cells used for deep transcriptome sequencing were isolated from a female Dezhou donkey from the Black Donkey Research Institute (Liaocheng, Shandong, China) using a peripheral blood white cell separation kit (WBC1094, TBD Science, Tianjin, China). The isolated white cells were stored in RNAlater Stabilization Solution (AM7020, Life Technologies) and sent to the Beijing Genome Institute (BGI-Shenzhen) for transcriptome library construction and RNA-seq.

Deep transcriptome sequencing of donkey white blood cells

All deep transcriptome sequencing procedures were performed at the Beijing Genome Institute (Shenzhen, China). Briefly, total RNAs were isolated and mRNAs were extracted by magnetic beads with Oligo(dT). These mRNAs were fragmented and used as templates to synthesize cDNAs. After purification and single nucleotide adenine addition, short fragments were linked to adapters for amplification. The amplified library was sequenced using the Illumina HiSeq 2000 platform. The raw sequencing data in fastq format was uploaded to the Sequence Read Archive (SRA) under the accession number SRX973470.

Table 3. Protein sequences used for multiple sequence alignment.

Symbol / Full name	Abbreviation	Species	ID
HIC1 / hypermethylated in cancer 1	D_HIC1	Donkey	Unigene110778_NormalA
	H_HIC1	Horse	gj 545180679
	M_HIC1-1	Mouse	gj 148226885
	M_HIC1-2	Mouse	gj 148228529
	H_HIC1-1	Human	gj 61676186
	H_HIC1-2	Human	gj 148237270
	P_HIC1-X1	Pig	gj 311268073
	P_HIC1-X2	Pig	gj 545859615
	PRKRA / protein kinase, interferon inducible double stranded RNA dependent activator	D_PRKRA	Donkey
H_PRKRA		Horse	gj 545191414
WH_PRKRA		Wild Horse	gj 664713499
M_PRKRA		Mouse	gj 755499568
M_PRKRA-X1		Mouse	gj 755499566
M_PRKRA-X2		Mouse	gj 6755162
Ho_PRKRA		Human	gj 49065476
P_PRKRA		Pig	gj 194043952
KMT2A / lysine-specific methyltransferase 2A		D_KMT2A	Donkey
	H_KMT2A	Horse	gj 545221883
	WH_KMT2A	Wild Horse	gj 664709525
	M_KMT2A	Mouse	gj 124486682
	H_KMT2A-1	Human	gj 308199413
	H_KMT2A-2	Human	gj 56550039
	P_KMT2A	Pig	gj 350588548

doi:10.1371/journal.pone.0133258.t003

RNA-seq data processing and donkey transcriptome annotation

The read adaptors and low-quality reads from raw RNA-seq reads were removed, and the *de novo* assembly was carried out with the RNA-seq *de novo* assembly software Trinity [18]. The new assembled donkey unigenes were annotated and the corresponding protein fragments were predicted using BAST searches of the unigene sequence to the non-redundant (NR) protein database.

Comparative analysis of RNA and protein sequences among the donkey, horse, and wild horse

The protein data for the mouse, horse [6], and wild horse [19] were downloaded from the NCBI genome database (<ftp://ftp.ncbi.nih.gov/genomes>). The number and length of RNA and protein fragments were summarized using a python script. The predicted donkey unigene proteins were aligned to the horse or wild horse protein dataset using the blastp program. Homologous sequences among the donkey, horse, and wild horse were identified by applying an identity threshold of 80% [20].

To link the donkey predicted protein fragments to mammalian phenotypes, mammalian phenotype data was extracted from the Mammalian Phenotype Browser (www.informatics.jax.org) [10]. As the predicted donkey protein fragments were not integrated, horse proteins were used to query homologous proteins between the mouse and *Equus* species. The mouse protein sequences were compared with those in horses, and the most similar sequences with identities of greater than 80% were considered homologous.

Analysis of *Equus* outer ear morphology-associated proteins

The information regarding genes associated with outer ear morphology (MP: 0002177) was extracted from the Mammalian Phenotype Browser [10]. The outer ear morphology-associated proteins that had no homology to horse genes or donkey sequences identified in the *de novo* assembled transcriptome were ignored. The donkey protein fragments that were not fully aligned to homologous horse proteins and whose protein identities were less than 98%, were analyzed to determine the genotype controlling *Equus* outer ear morphology. To analyze the outer ear size associated-proteins, outer ear size associated-protein sequences of the mouse, human, and pig were obtained, and a multiple sequence alignment was generated and analyzed (Table 3).

Cytoscape was used to visualize the gene–phenotype associations [21]. Python scripts were written to extract and analyze data and these scripts can be obtained on requested.

Supporting Information

S1 Fig. Multiple alignment of the KMT2A protein. The protein IDs and corresponding GenBank accession numbers are listed in Table 3.

(PDF)

S1 Dataset. Lengths and GC percents of assembled donkey genes.

(ZIP)

S2 Dataset. Assembled donkey genes and predicted proteins.

(ZIP)

S3 Dataset. Phenotype annotation of donkey assembled genes.

(XLSX)

Acknowledgments

We thank the staff at the Black Donkey Institute for their help during blood sample collection. We also thank the office staff at the Shandong Province Horse Association for their help with webpage construction.

Author Contributions

Conceived and designed the experiments: JYM WS YJS. Performed the experiments: FYX YLF HHW YFM YY YCW JYM. Analyzed the data: JYM SY. Contributed reagents/materials/analysis tools: YCW QJP YJS WS JYM. Wrote the paper: JYM SY.

References

1. Waltraud Kugler H-PG, Elli Broxham. Donkey Breeds in Europe: Inventory, Description, Need for Action, Conservation Report 2007/2008. Monitoring Institute, 2008.
2. Denis Fielding PS. Donkeys, People and Development: A Resource Book of the Animal Traction Network for Eastern and Southern Africa (ATNESA): Technical Centre for Agricultural and Rural Cooperation; 1997.
3. Resources CNCoAG. Animal Genetic Resources Blog: horses and donkeys. camel Chi. Beijing: China Agriculture Press; 2011.
4. Dekkers JC. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. Journal of animal science. 2004; 82 E-Suppl:E313–28. Epub 2004/10/09. PMID: [15471812](#).

5. Stock KF, Reents R. Genomic selection: Status in different species and challenges for breeding. *Reproduction in domestic animals = Zuchthygiene*. 2013; 48 Suppl 1:2–10. Epub 2013/08/24. doi: [10.1111/rda.12201](https://doi.org/10.1111/rda.12201) PMID: [23962210](https://pubmed.ncbi.nlm.nih.gov/23962210/).
6. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*. 2009; 326(5954):865–7. Epub 2009/11/07. doi: [10.1126/science.1178158](https://doi.org/10.1126/science.1178158) PMID: [19892987](https://pubmed.ncbi.nlm.nih.gov/19892987/); PubMed Central PMCID: [PMC3785132](https://pubmed.ncbi.nlm.nih.gov/PMC3785132/).
7. Brooks SA, Gabreski N, Miller D, Brisbin A, Brown HE, Streeter C, et al. Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS genetics*. 2010; 6(4):e1000909. Epub 2010/04/27. doi: [10.1371/journal.pgen.1000909](https://doi.org/10.1371/journal.pgen.1000909) PMID: [20419149](https://pubmed.ncbi.nlm.nih.gov/20419149/); PubMed Central PMCID: [PMC2855325](https://pubmed.ncbi.nlm.nih.gov/PMC2855325/).
8. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS genetics*. 2012; 8(1):e1002451. Epub 2012/01/19. doi: [10.1371/journal.pgen.1002451](https://doi.org/10.1371/journal.pgen.1002451) PMID: [22253606](https://pubmed.ncbi.nlm.nih.gov/22253606/); PubMed Central PMCID: [PMC3257288](https://pubmed.ncbi.nlm.nih.gov/PMC3257288/).
9. Hoehenwarter W, van Dongen JT, Wienkoop S, Steinfath M, Hummel J, Erban A, et al. A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. *Proteomics*. 2008; 8(20):4214–25. Epub 2008/10/17. doi: [10.1002/pmic.200701047](https://doi.org/10.1002/pmic.200701047) PMID: [18924179](https://pubmed.ncbi.nlm.nih.gov/18924179/).
10. Hancock JM, Adams NC, Aidinis V, Blake A, Bogue M, Brown SD, et al. Mouse Phenotype Database Integration Consortium: integration [corrected] of mouse phenome data resources. *Mammalian genome: official journal of the International Mammalian Genome Society*. 2007; 18(3):157–63. Epub 2007/04/17. doi: [10.1007/s00335-007-9004-x](https://doi.org/10.1007/s00335-007-9004-x) PMID: [17436037](https://pubmed.ncbi.nlm.nih.gov/17436037/); PubMed Central PMCID: [PMC4230762](https://pubmed.ncbi.nlm.nih.gov/PMC4230762/).
11. Chen WY, Wang DH, Yen RC, Luo J, Gu W, Baylin SB. Tumor suppressor HIC1 directly regulates SIRT1 to modulate p53-dependent DNA-damage responses. *Cell*. 2005; 123(3):437–48. Epub 2005/11/05. doi: [10.1016/j.cell.2005.08.011](https://doi.org/10.1016/j.cell.2005.08.011) PMID: [16269335](https://pubmed.ncbi.nlm.nih.gov/16269335/).
12. Carter MG, Johns MA, Zeng X, Zhou L, Zink MC, Mankowski JL, et al. Mice deficient in the candidate tumor suppressor gene Hic1 exhibit developmental defects of structures affected in the Miller-Dieker syndrome. *Human molecular genetics*. 2000; 9(3):413–9. Epub 2000/02/03. PMID: [10655551](https://pubmed.ncbi.nlm.nih.gov/10655551/).
13. Deltour S, Pinte S, Guerardel C, Wasylyk B, Leprince D. The human candidate tumor suppressor gene HIC1 recruits CtBP through a degenerate GLDLSKK motif. *Molecular and cellular biology*. 2002; 22(13):4890–901. Epub 2002/06/08. PMID: [12052894](https://pubmed.ncbi.nlm.nih.gov/12052894/); PubMed Central PMCID: [PMC133903](https://pubmed.ncbi.nlm.nih.gov/PMC133903/).
14. Peters GA, Li S, Sen GC. Phosphorylation of specific serine residues in the PKR activation domain of PACT is essential for its ability to mediate apoptosis. *The Journal of biological chemistry*. 2006; 281(46):35129–36. Epub 2006/09/20. doi: [10.1074/jbc.M607714200](https://doi.org/10.1074/jbc.M607714200) PMID: [16982605](https://pubmed.ncbi.nlm.nih.gov/16982605/).
15. Kok KH, Ng MH, Ching YP, Jin DY. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *The Journal of biological chemistry*. 2007; 282(24):17649–57. Epub 2007/04/25. doi: [10.1074/jbc.M611768200](https://doi.org/10.1074/jbc.M611768200) PMID: [17452327](https://pubmed.ncbi.nlm.nih.gov/17452327/).
16. Rowe TM, Rizzi M, Hirose K, Peters GA, Sen GC. A role of the double-stranded RNA-binding protein PACT in mouse ear development and hearing. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(15):5823–8. Epub 2006/03/31. doi: [10.1073/pnas.0601287103](https://doi.org/10.1073/pnas.0601287103) PMID: [16571658](https://pubmed.ncbi.nlm.nih.gov/16571658/); PubMed Central PMCID: [PMC1458657](https://pubmed.ncbi.nlm.nih.gov/PMC1458657/).
17. Chen W, Li Q, Hudson WA, Kumar A, Kirchhof N, Kersey JH. A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy. *Blood*. 2006; 108(2):669–77. Epub 2006/03/23. doi: [10.1182/blood-2005-08-3498](https://doi.org/10.1182/blood-2005-08-3498) PMID: [16551973](https://pubmed.ncbi.nlm.nih.gov/16551973/); PubMed Central PMCID: [PMC1895483](https://pubmed.ncbi.nlm.nih.gov/PMC1895483/).
18. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 2011; 29(7):644–52. Epub 2011/05/17. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/); PubMed Central PMCID: [PMC3571712](https://pubmed.ncbi.nlm.nih.gov/PMC3571712/).
19. Huang J, Zhao Y, Shiraigol W, Li B, Bai D, Ye W, et al. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Scientific reports*. 2014; 4:4958. Epub 2014/05/16. doi: [10.1038/srep04958](https://doi.org/10.1038/srep04958) PMID: [24828444](https://pubmed.ncbi.nlm.nih.gov/24828444/); PubMed Central PMCID: [PMC4021364](https://pubmed.ncbi.nlm.nih.gov/PMC4021364/).
20. Pearson WR. An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics / editorial board, Andreas D Baxeavanis [et al]*. 2013; Chapter 3:Unit3 1. Epub 2013/06/12. doi: [10.1002/0471250953.bi0301s42](https://doi.org/10.1002/0471250953.bi0301s42) PMID: [23749753](https://pubmed.ncbi.nlm.nih.gov/23749753/); PubMed Central PMCID: [PMC3820096](https://pubmed.ncbi.nlm.nih.gov/PMC3820096/).
21. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011; 27(3):431–2. Epub 2010/12/15. doi: [10.1093/bioinformatics/btq675](https://doi.org/10.1093/bioinformatics/btq675) PMID: [21149340](https://pubmed.ncbi.nlm.nih.gov/21149340/); PubMed Central PMCID: [PMC3031041](https://pubmed.ncbi.nlm.nih.gov/PMC3031041/).