The Thai Journal of Veterinary Medicine

Volume 49
Issue 1 *March, 2019*

Article 4

3-1-2019

Transcriptome analysis of Zi geese (Anser cygnoides) prehierarchical follicles and verification of related genes in arginine metabolic pathway

Yongfeng Sun

Lulu Liu

Chang Liu

Cornelius Tlotliso Sello

Huiyan Wu

See next page for additional authors

Transcriptome analysis of Zi geese (Anser cygnoides) prehierarchical follicles and verification of related genes in arginine metabolic pathway

Authors

Yongfeng Sun, Lulu Liu, Chang Liu, Cornelius Tlotliso Sello, Huiyan Wu, Hongtao Lu, Chenguang Xu, and Yujian Sui

Original Article

Transcriptome analysis of Zi geese (*Anser cygnoides*) prehierarchical follicles and verification of related genes in arginine metabolic pathway

Yongfeng Sun[1,2,†*] Lulu Liu[1,†] Chang Liu[1,†] Cornelius Tlotliso Sello[1]

Huiyan Wu[1] Hongtao Lu[1] Chenguang Xu[1] Yujian Sui[1]

Abstract

Arginine metabolism plays a pivotal role in the mammalian reproductive system by controlling various physiological processes of the reproductive organs including follicular development, ovulatory mechanisms and egg production. However little is known about the molecular mechanisms regulating arginine metabolism at early stages of folliculogenesis. The objective of the current study was to determine the arginine metabolic pathways in prehierarchical follicles of Zi geese through the Illumina Hiseq 2500 platform and their transcripts through the *de novo* assembly. In this study, five different stages of the geese prehierarchical follicles transcriptome were sequenced. A total of 31722729162 bp from 224929214 high quality clean sequences was obtained. There were 74.75% transcripts annotated pathways, and 106 unigenes were mapped in the Arginine and proline metabolism pathway. Combined with Gene Ontology (GO) functions, *ArgE*, *NOS* and *ODC* were selected as the candidate genes to explore the arginine metabolic pathway and determine their relative expression in the prehierarchical follicles of geese. The results showed that the three genes were expressed in all five prehierarchical stages of geese follicles, with the predominant expression in primary follicles ($p < 0.05$), indicating promotion of follicular development. This study provides useful inclusive gene expression profiling information that would certainly provide basic theory for future research studies on follicular development in Zi geese.

Keywords: Annotation, *de novo* assembly, follicular development, gene expression

[1]*Department of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, Xincheng Street, No: 2888. Changchun City, Jilin Agricultural University, Changchun, People's Republic of China*

[2]*Key laboratory for animal production, product quality and safety of the Ministry of education, China*

[†]*These authors contributed equally to this article and share senior coauthor ship*

***Correspondence**: sunyongfeng1977@126.com*

Introduction

The geese (*Anas cygnoides*) are commercially important grass-feeding poultry species largely cultivated in China (Xu et al. 2015). Zi geese are the native breed in Jilin and Heilongjiang provinces of the northeast of China with a high feed conversion ratio, and better adaptability (Zhang et al. 2013). However, the low egg production performance (on average 80-100 eggs per annual reproduction cycle) hinders the goose's large-scale production (Ding et al. 2015). In recent years studies on improving geese production performance (egg laying performance) at the molecular level over the conversional hybridization breeding method has gained more attention (Xu et al. 2013). Poultry egg laying performance depends largely on the continuous growth and development of the ovarian follicles from non-hierarchical to hierarchical follicles (Liu et al. 2015). It is therefore important to understand the molecular ovarian follicular development mechanisms to conduct targeted regulation of the laying cycle recruitment to improve geese laying performance. Transcriptome focuses mainly on understanding the biogenetical functions through differentially expressed genes (DEGs) analysis (Yang et al. 2016). RNA-Seq is one of the most competent and profound next generation sequencing (NGS) technology tool used to disclose the transcriptional profiles for gene expression analysis in biological molecular mechanisms (Yang et al. 2013). In recent years, RNA-Seq has been useful in the computable analysis of gene expression patterns in various tissues in geese including nonhierarchical follicle ovaries (Yu et al. 2016), spleen (Wang et al. 2015), hypothalamus (Gao et al. 2015), and the liver (Yen et al. 2009). Identification of genes regulating geese follicular development and the difference in their expression patterns between the different follicular stages would lay a foundation for further understanding of molecular and cellular mechanisms of folliculogenesis in geese.

The initial stages of follicles development are critical reproductive central units of the ovary that lead to healthy dominant follicles for ovulation (Choi and Rajkovic 2006). Therefore, the transcription of the various genes necessary for folliculogenesis needs more attention at early stages of development. Arginine provides the most nitrogen for organisms, approximately 14% and does not only serves as substrates to synthesize proteins, but can also regulate vital cellular functions through its biological activities and its metabolites (Wu 2013). Arginine affects the expression of genes and the signal transduction of nerves (Wiesinger 2001), the endocrine system (Tong and Barbul 2004) and plays a role in the female reproductive system (Che et al. 2013). Mateo et al. (2007) have demonstrated that dietary L-arginine supplementation to gilts improves fetal survival and growth and the birth weights of piglets. Zeng et al. (2013) previous study has shown that arginine supplementation in early pregnancy in rats enhanced embryo implantation. Contrary to other animal species, poultry lack the key enzymes carbamoyl phosphatase, which synthesize arginine, and can only obtain this amino acid through their diet (Oso et al. 2017). Polyamine can regulate cell development and growth, gene expression, cell proliferation and reproductive capacity (Agostinelli et al. 2010). Hence, it is necessary to understand the expression patterns of genes in the follicles of geese that relate to arginine metabolic pathways. The aim of the study was to determine the arginine metabolic pathways in prehierarchical follicles of Zi geese through the Illumina Hiseq 2500 platform and their transcripts through the *de novo* assembly. Through the functional annotation and analysis of differential gene expression, we obtained biological information on the genes and proteins of Zi geese.

Materials and Methods

Ethics statement: A healthy maternal line of laying Zi geese (*Anser cygnoides*) was provided by Jilin Agricultural University. The geese were reared under natural light and the feed was provided during daytime with free access to water. The Animal welfare ethics committee of Jilin University (20150903), Changchun, People's Republic of China approved this experiment.

Tissue preparation: Six Zi geese aged 35 - 37 weeks laying in regular sequences from a minimum of 2 - 3 eggs were used in this study. The geese were anaesthetized by cervical dislocation approximately 7 h to 9 h after mid-sequence ovulation. The follicles were divided into prehierarchical follicles and hierarchical follicles. The prehierarchical follicles were classified as primary follicles (< 2 mm, PE), small white follicles (2 - 4 mm, SWF), middle white follicles (4 - 6 mm, MWF), large white follicles (6 - 10 mm, LWF) and small yellow follicles (10 - 15 mm, SYF). The vascular layer at the exterior and the yolk at the innermost layer were removed, and the membrane was washed in phosphate buffered saline (PBS) and placed in a -80 °C refrigerator for future use.

RNA extraction for RNA-sequence and quality controls: Total RNA was extracted from five stages of the prehierarchical follicles of Zi geese using TRIzol ® Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. The mixed solution was transferred to an RNA adsorption column and centrifuged at 13,000 g for 1 min at 4 ° C, the filtrate was discarded, and the prepared DNAse was added and allowed to stand at room temperature for 10 min, which requires 5μl total RNA for library construction. The Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) was used to determine the concentration and the purity of the RNA. The quality of the RNA was detected by 1% agarose gel electrophoresis, and the integrity of the RNA was measured by using an Agilent 2100 Bio analyzer (Agilent, Santa Clara, CA, USA).

Construction of cDNA library: Using Oligo-(dT) magnetic beads the mRNA was isolated from total RNA, and the mRNA was then broken up into about 100 bp fragments using metalions. Under the action of reverse transcriptase, using random primers, the mRNA was used as a template to synthesize the first-

strand of cDNA. The RNA template was then removed and a replacement strand synthesized to generate ds cDNA. The 3′ to 5′ exonuclease activity of this mix removed the 3′ overhangs, and the polymerase activity fills in the 5′ overhangs. A single 'A' nucleotide was added to the 3'ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the '3′ end of the adapter provided a complementary overhang for ligating the adapter to the fragment. After 15 cycles of PCR amplification, DNA fragments that had adapter molecules on both ends amplified the amount of DNA in the library.

De novo assembly: The data were sequenced by the Illumina Hiseq 2500 platform, and the sequenced runs were used to produce 100 bp paired-end sequenced reads. Raw reads were obtained from this step. We disposed the adapter sequences, low-quality bases at the end of sequence 3'-ends of sequences, the reads with N ratios of more than 10%, and the sequences of lengths of less than 20 base pairs. The filtered reads obtained by filtering the raw data allowed the subsequent work to proceed. Using the assembly Trinity (Haas et al., 2013), we performed *de novo* assembly of the filtered reads to obtain contigs and singletons.

Open Reading Frame prediction: An Open Reading Frame (ORF) is a base sequence that can code for protein. ORF can be predicted from transcripts (forecasting process: http: //trinityrnaseq.sourceforge.net/analysis/extract_proteins_from_trinity_transcripts. html). The prediction results were corrected using the Pfam (version 27.0) database, and the protein sequences that matched in the database were preserved. BlastX (version 2.2.25, E-value threshold ≤ 1*10 - 5) tested all the nucleotide sequences obtained from splice against the NCBI Non-redundant protein database. Non-redundant protein database is a synthesis database, and it includes Swissprot, Protein Information Resource (PIR), Protein Research Foundation (PRF), and Protein Data Bank (PDB) databases (These databases were downloaded in December 2014).

Annotation of transcripts: The KEGG functional annotation assisted in comprehending biological functions at the system level, such as metabolic pathways, the transmission of genetic information and cellular processes. The arginine metabolic pathway was annotated from the KEGG pathway database and obtained KEGG Orthology (KO) serial numbers corresponding to the transcripts through comparison with the KEGG databases were obtained (KEGG is obtained from KOBAS software, http://kobas.cbi.pku.edu.cn/home.do, Version, KOBAS). Based on the KO identifiers (KO00330), the functional information and metabolic pathways of genes related to the arginine metabolism pathway were determined. All genes related to the arginine metabolism pathway annotated using the Blast2Go (Version: 2.5.0) software (Conesa et al., 2005). The samples were compared with the clusters of orthologous groups (COG, http://www.ncbi.nlm.nih.gov/COG) databases to perform functional annotation, functional classification, and phylogenetic analysis of the proteins. We obtained the biological information of genes related to arginine metabolism and consequently identified the gene involved in arginine catabolism.

Comparison of assembly: We obtained trimmed and filtered reads using SeqPrep software and the Sickle software. Trimmed and filtered reads were compared to assemble transcriptome sequences (allowing two base mismatches). With no contamination and good assembly results, the mapping rate would be greater than 60%, as the reads sequenced by the Illumina Hiseq 2500 platform are usually short with few insertion or deletion errors. To complete this part of the analysis, we selected the currently authoritative short sequence alignment software Bowtie2. The software was used for analysis of the genome sequence, followed by a backtracking algorithm and search to locate reads. Base substitution can account for allowed mismatches. The mapping ratios of the five samples were higher than 85%, further indicating that the assembly results were of good quality (Langmead et al. 2009; Langmead and Salzberg 2012).

Analysis of the differential gene expression: In the previous step, we identified genes related to the arginine metabolic pathway. Then we compared the expression of genes in five stages of prehierarchical follicles. We calculated the reads per kilo base of the exon model per million mapped reads (FPKM) of each gene sample as the expression quantity. We identified the significant difference of gene expression levels based on the condition that False Discovery Rate (FDR) < 0.05 & log | Fold Change (FC) | >= 2 through the EdgeR (version 2.12) software and used Goatools (https://github.com/tanghaibao/GOatools) to analyze the GO enrichment significance. The Fisher exact test was used to test the results. To control the false positive rate, we adopted four methods (Bonferroni, Holm, Sidak and false discovery rate) to test the P-value. The correction p-value (P-FDR) ≤ 0.05 reveals the significance of Gene Ontology functional enrichment.

Identification of RNA-seq by RT-qPCR: Total RNA was extracted from five stages of prehierarchical follicles. Every 10μl reaction system containing 1μl RNA, and cDNA was synthesized by reverse transcription (TOYOBO, FSQ-101, Osaka, Japan). Reaction conditions were as follows: 37°C, 15min; 98°C, 5min. The RT-qPCR technique was used to confirm expression levels. Using the $2^{-\Delta\Delta CT}$ algorithm, the gene expression levels in different stages of follicles were obtained. Primers were designed by Primer Premier 5.0 software (Primer-E Ltd., Plymouth, UK) according to the nucleotide sequences (ORF prediction). The primer sequences used for RT-qPCR are listed in Table 1. *β-actin* was used as an internal control. Following the SYBR Green Kit (Toyobo, Osaka, Japan) manufacturer's protocol with a 20 μl reaction system, the reaction conditions were 95°C, 1 min; 95°C, 15 s; 60°C, 45 s; 40 cycles; a melt curve was drawn at the end of the reaction.

Table 1 Primers used for RT-PCR.

Primers	Sequence (5′-3′)	Annealing temperature (°C)	Amplicon size (bp)
ODC	F: GCATAAAGCCCTTCCTCG R: TTCCGTCTTACTGGCACAAT	59	118
ArgE	F: CAACACCCCTCTTACAACTCA R: TCCACATCCCTCAAACCAA	59	162
NOS	F: GTCCTGGTGTTTGGGTG R: GTATTTCTTTGGTTTGTCCG	57	135
β-actin	F: GACCACCTTCAACTCCATC R: GGCTGTGATCTCCTTCTG	59	130

Results

Reads assembly and analysis of the transcriptome: Total RNA was extracted from five stages of prehierarchical follicles of Zi geese. The RNA integrity number (RIN) values of all samples were greater than 8.5. The OD 260/280 of the RNA samples ranged from 1.9 to 2.0. Based on the above results, the quality of RNA was satisfactory to be used in subsequent experiments. Raw data produced by the Illumina Hiseq 2500 platform reached 231,060,052 reads. The raw data was submitted to SRA (SRP071765) and filtered to obtain high-quality sequencing data which was used for subsequent analysis (Erlich et al. 2008). The filtered reads were *de novo* assembled using the Trinity software, because the Zi geese belong to the species with no reference genome. After the low quality reads and adaptors had been removed, 224929214 clean reads remained, among which 37618718 reads were from PE, 39715266 reads were from SWF, 48168070 reads were from MWF, 55265868 reads were from LWF and 44161292 reads were from SYF. The details of filtered reads are shown in Table 2. The percentages of GC content were 45.44% and 44.82% in the transcripts and unigenes, respectively. Most of the transcripts were 1 – 600 bp in length (32.16%), followed by 601 – 800 bp (15.4%), and 801 – 1000 bp (9.42%). The complete long sequences of good quality allowed us to acquire more information on the genes, and these transcriptome results provide useful data for future experiments.

Table 2 The *de novo* assembly results.

	Unigenes	Transcripts
Total sequence num	114,486	145,020
Total base sequences	114,645,812	228,399,949
Percentage of GC	44.82%	45.44%
Largest transcript	28,471	28,471
Smallest transcript	401	401
Average length	1263.44	1574.95
N50	1,817	2,689

Open Reading Frame results: Functional annotation of these transcripts for the better understanding of the biological functions. Based on the results, we predicted 74,405 sequences. The annotation information was obtained by comparison with the Non-redundant protein database, String, and Swissprot databases. Meanwhile, primers were synthesized according to the sequence predicted and verified by PCR. The product was compared with those of homologous species to verify its reliability.

Non-redundant database annotation results statistics: The Non-redundant protein database is a synthesized database (approximately 91.25% of the sequences are annotated) that contains SwissProt, Protein Information Resource (PIR), Protein Research Foundation (PRF), and Protein Data Bank (PDB) databases. By comparison with the Non-redundant protein database, we examined the similarity between the geese and similar species, as well as the functional information on homologous sequences. 21,771 unigene sequences were annotated in the Non-redundant protein database, with Swissprot providing 16,156 unigene sequences. The unannotated unigene sequences are coding sequences for which functions have not yet been identified or non-coding sequences and some of them might be novel genes. Based on the sequence alignment results, the species with a high homology to the geese include *Anas platyrhynchos* (20,208), *Gallus gallus* (7,295), and *Columba livia* (1,735). We used an E-value threshold ≤1*10-5 as a condition to filter out unqualified sequences (Camacho et al. 2009; Grabherr et al. 2011). The smaller the E-value, the more the reliability of the results. The E-values of 32393 unigene sequences were 0; 3082 unigene sequences ranged from 1e - 20 to 1e - 10; and 2745 unigene sequences ranged from 1e - 10 to 1e - 5, demonstrating that the matching results are reliable. Also, 37,647 unigenes were defined in the database with 80% – 100% of the total amino acids and 3,739 unigenes were defined in the database with 60% – 80% of the total amino acids.

Gene Ontology (GO) and Cluster of Orthologous Group of proteins (COGs) annotations: Blast2Go software was used to analyze the unigenes and compare them with the GO database. There are three categories in GO: biological process, molecular function and cellular component. These three major categories include many levels, and the larger the level number, the more detailed the function will be obtained. The top three categories are considered level

1, followed by level 2, level 3 and level 4. The classification of all the gene products of geese is briefly described. According to the alignment results, 1,744 unigene sequences were assigned to the molecular functions category, 3,811 to the cellular component category, and 6,126 to the biological process category. The top three groups in the molecular function category are binding (68.42%), catalytic activity (47.85%) and transporter activity (9.76%). The top three in the cellular component category are cell (68.52%), cell part (68.52%) and organelle (56.75%). The top three terms in the biological process category are cellular process (74.74%), metabolic process (67.37%) and single-organism process (66.89%). These total values were above 100% because the same unigene can be annotated with different functions. The biological process category which was the focus in this research was related to signaling (311 unigenes), metabolic process (704 unigenes), cellular metabolic process (583 unigenes), nitrogen compound metabolic process (827 unigenes), regulation of metabolic process (311 unigenes), cellular macromolecule metabolic process (457 unigenes), cellular nitrogen compound metabolic process (383 unigenes) and organic substance biosynthetic process (318 unigenes) in the fourth-level classifier cartogram.

COG is a protein orthologous cluster database, allowing unknown sequences to be annotated by known proteins. The functional annotation, classification, and analysis of protein evolution were performed by the COG database. The COG number is obtained from the String (http://string-db.org/) database, to allow functional classification. The results showed that 5,308 unigenes could be categorized into 24 COG categories. Amongst the most numerous unigenes only (967 unigenes) had general function predictions followed by "signal transduction mechanisms" (604 unigenes), "posttranslational modification, protein turnover, chaperones" (479 unigenes) and "translation, ribosomal structure and biogenesis" (282 unigenes). The smaller groups are cell motility (6 unigenes) and nuclear structure (3 unigenes). In addition, the COG categories of amino acid transport and metabolism included 218 unigenes.

KEGG pathway analysis: We assigned 26,532 unigenes to 342 KEGG pathways. The pathways containing the most genes were as follows: pathways in cancer (721 unigenes), the PI3K-Akt signaling pathway (688 unigenes), and the MAPK signaling pathway (660 unigenes). A signaling pathway that contains more unigenes is generally active. For various research purposes, subsequent experiments were conducted. The remaining active pathways are shown in Fig 1. The metabolic pathway relevant to this research through KEGG pathway analysis were arginine and proline metabolism (106 unigenes) as shown in Fig 2. Genes related to arginine metabolism in geese prehierarchical follicles, such as arginine (*Arg*), nitric oxide synthase (*NOS*), ornithine decarboxylase (*ODC*), argininosuccinate synthase (*ASS*), arginine decarboxylase (*ADC*) and argininosuccinate lyase (*ASL*) were found as shown in Table 3.

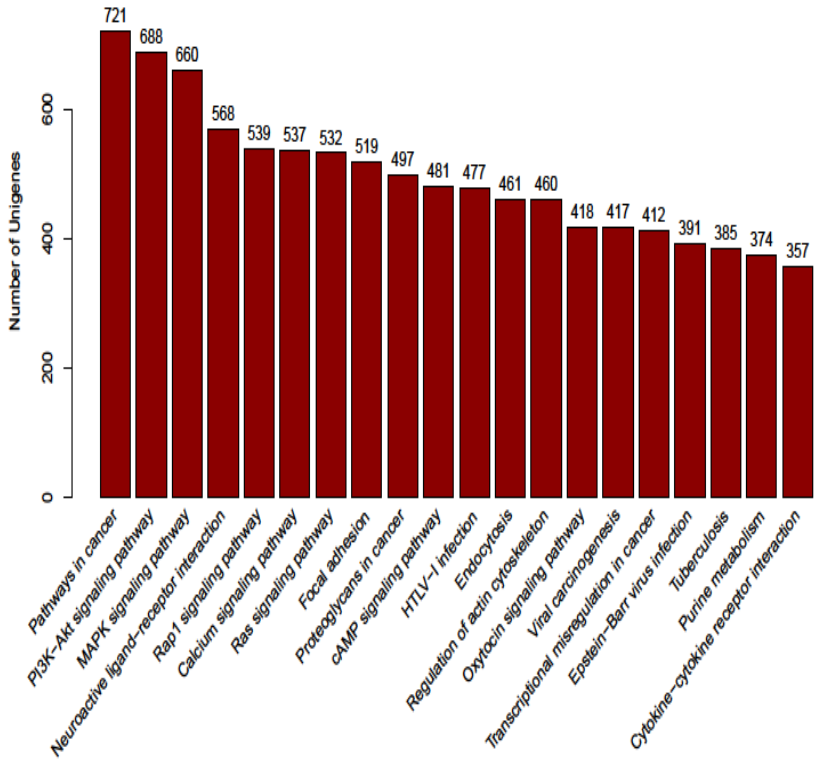

Figure 1 The top 20 KEGG pathways**.** The column height indicates that the activity of the biological pathway in the sample.

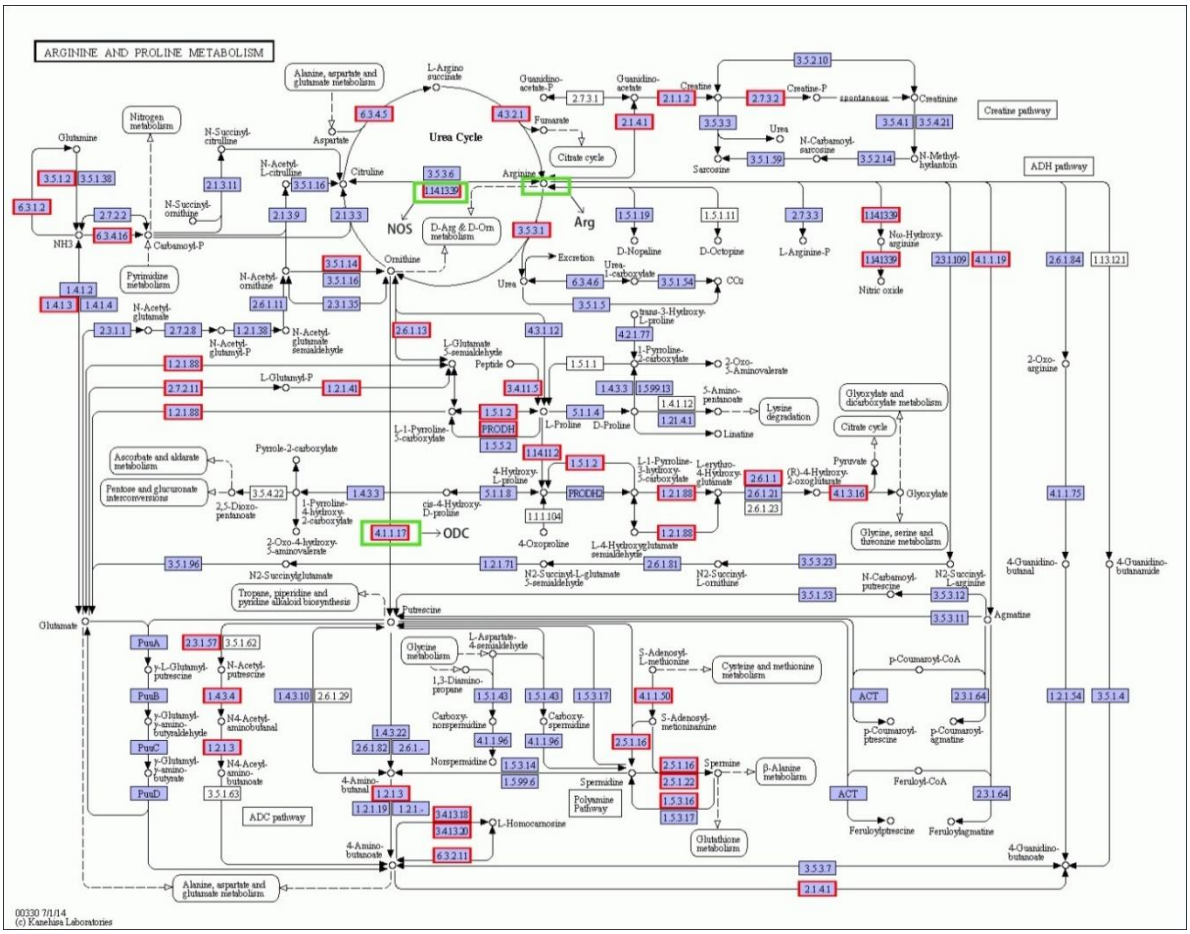

Figure 2 Arginine and Proline Metabolism**.** We annotated the arginine metabolic pathway from the KEGG pathway database and obtained KEGG Orthology (KO) serial numbers corresponding to the transcripts through comparison with the KEGG databases.
Blue Background Rectangular Box: Gene Products belong to KEGG ORTHOLOGY
Red Border Rectangular Box: Gene Products Detected by This Sequencing
–>: Activation –|: Inhibition —: Binding/Association
+p: Phosphorylation -p: Dephosphorylation : Direction of Reaction
NOS: nitric oxide synthase *ArgE*: arginase *ODC*: ornithine decarboxylase

Table 3 Some genes related to the arginine metabolism pathway.

KO	ID	Definition
K13240	c83328_g1	PREDICTED: nitric oxide synthase, brain [*Anas platyrhynchos*]
K01476	c91586_g1	Arginase, type II [*Gallus gallus*]
K00802	c92150_g1	Spermine synthase [*Gallus gallus*]
K01581	c187690_g1	PREDICTED: ornithine decarboxylase [*Falco peregrinus*]
K12259	c90045_g1	PREDICTED: spermine oxidase isoform X1 [*Anas platyrhynchos*]
K01583	c213995_g1	PREDICTED: arginine decarboxylase isoform X3 [*Gallus gallus*]

Differential gene expression analysis: To understand the expression of arginase (*ArgE*), nitric oxide synthase (*NOS*), and ornithine decarboxylase (*ODC*) in geese prehierachical follicles, RNA-Seq by Expectation Maximization (RSEM, version 1.2.27) was used to calculate their relative expression quantities. The standard measure of gene expression levels is the FPKM value, read as a unit. All the arginine metabolic related genes were expressed in all the prehierarchical follicles with the highest expression in PE ($p < 0.05$). However, the least detectable amounts were observed in arginine in middle white follicles ($p < 0.05$). The RT-qPCR results are consistent with transcriptome in Fig 3.

Discussion

The statistics of the unigenes annotated using different databases, such as Pfam, KEGG, Swissprot and Non-redundant protein database suggest that

these unigenes have more detailed biological function information and are of the great reference value for our area of interest. The Non-redundant protein database is not only a comprehensive database but also a non-redundant protein database. The result is ideal because approximately 91.25% of the sequences are annotated in the Non-redundant protein database.

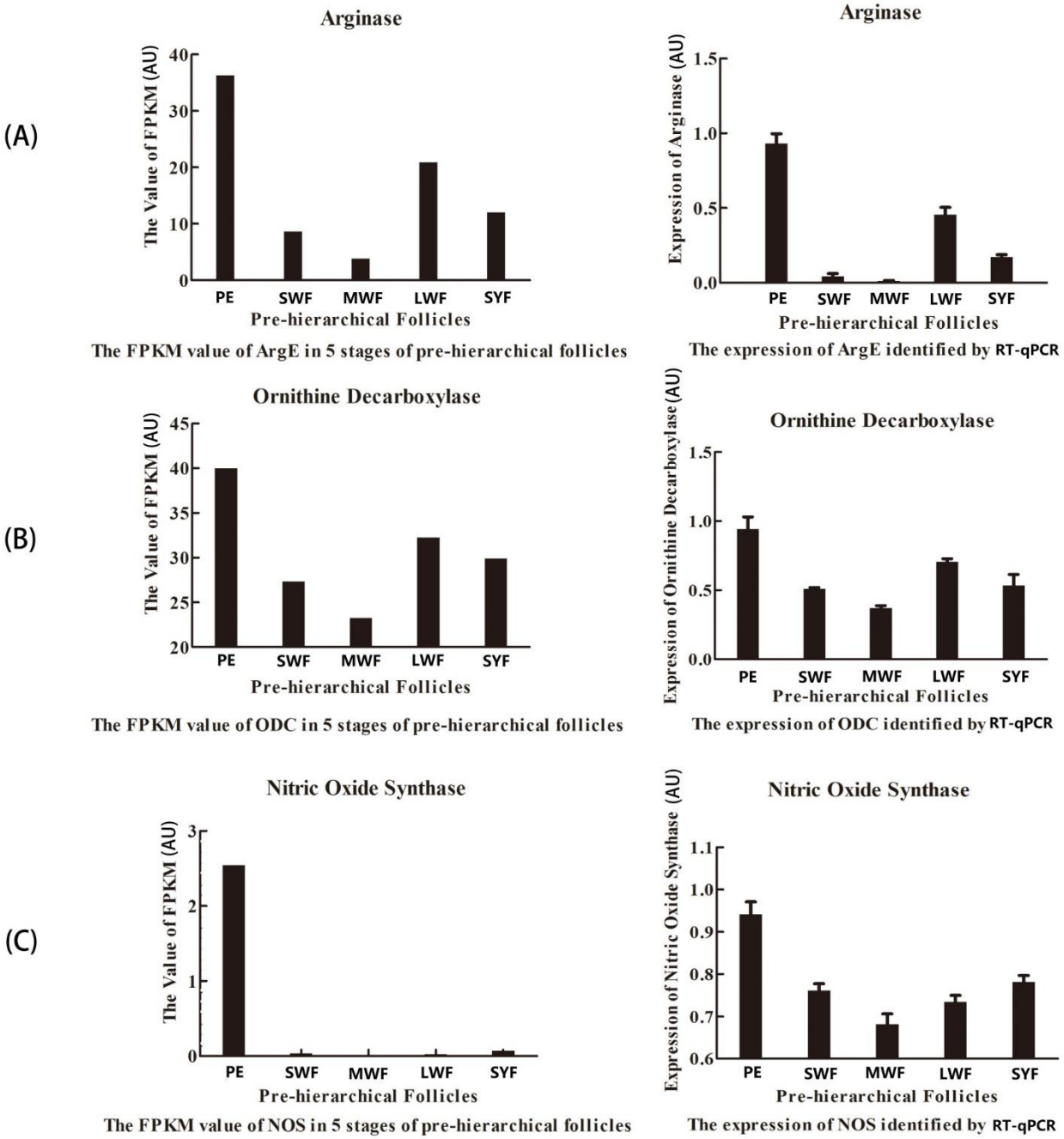

Figure 3 The expression of genes in five stages of prehierarchical follicles (PE, SWF, MWF, LWF, SYF)**.** The FPKM (Fragments Per Kilobase of exon per Million fragments mapped) value of *ArgE* (Arginase) in five stages of prehierarchical follicles (A) and the expression of *ArgE* identified by RT-qPCR (B). The FPKM value of *ODC* (ornithine decarboxylase) in five stages of prehierarchical follicles (C) and the expression of *ODC* identified by RT-qPCR (D). The FPKM value of *NOS* (nitric oxide synthase) in five stages of prehierarchical follicles (E) and the expression of *NOS* identified by RT-qPCR (F). The prehierarchical follicles were graded as PE (primary follicles; < 2 mm), SWF (small white follicles; 2 - 4 mm), MWF (middle white follicles; 4 - 6 mm), LWF (large white follicles; 6 - 10 mm) and SYF (small yellow follicles; 10 - 15 mm).The data is represented as means ± SEM; and different letters indicate the significant amount of mRNA expression between different sizes of prehierarchical follicles ($p < 0.05$) normalised with *β-actin*.

GO enrichment significant analysis was performed for the differentially expressed genes. The number of background genes was 114,486 unigenes. These results did not only suggest which differential genes were concentrated on which functions, but also clarified the differences between different levels of prehierarchical follicles at the genetic level. Four different multiple-testing procedures were used to correct the p value to control the false positive rate. Notably, there was significant KEGG pathway when the correction p-value (p-fdr) ≤ 0.05. The differential expression of genes was more enriched in the protein

metabolism of prehierarchical follicles, such as the arginine and proline metabolism pathway related to this research. The GO database is a comprehensive database and unifies the definitive descriptions of gene and protein functions. Combined with the GO annotation analysis, *ArgE, NOS,* and *ODC* were selected.

Ovarian folliculogenesis begins with the development of primordial follicles, which contain an oocyte detained at the first mitotic division diplotene stage, encircled by a few flattened undifferentiated granulosa cells layers. During the follicles further development, the primordial structures transit from the resting pool to primary follicles, with a single layer of granulosa cells to form a rounded, cuboidal shape as they increase in size and proliferate (Johnson 2015). Ovarian granulosa cells play an important role in the development of follicles since they secrete essential factors (estradiol and insulin-like growth factor) promoting follicular growth and development (Matsuda et al. 2012). The proliferation and apoptosis of follicular granulosa cells play a key role in regulating and determining the fate of follicles and ovulation (Shimizu et al. 2012). Studies have shown that granulosa cells apoptosis leads to follicles atresia when more than 10% of granulosa cells degenerate and hence fail to develop to maturity stage (Lin et al. 2010).

Arginine is a major nitrogen-supplying amino acid for poultry. However, birds lack all obligatory enzymes to synthesize arginine from the urea cycle and hence depend greatly on the exogenous dietary supply of this amino acid (Murakami et al. 2012). The arginine metabolism proceeds by two main routes. In the first route, arginase catalyzes the conversion of L-arginine to ornithine and urea, and ornithine decarboxylase catalyzes (*ODC*) the conversion of ornithine to polyamines. L-arginine serves as a substrate for many molecules comprised of creatine, polyamines, glutamine, glutamate, ornithine, proline, protein, agmatine, dimethylargininesa and nitric oxide hence it implicates numerous important bio physiological functions in avian (Tapiero et al. 2002).The study reported by Weiger and Hermann (2014) indicated that polyamines are involved in cell growth, differentiation and proliferation. The second route involves nitric oxide synthase (*NOS*), which catalyzes the conversion of arginine to nitric oxide and citrulline (Bautista-Ortega J et al. 2014). Many studies have shown that *NO* is an important regulatory factor for ovarian development and can effect ovarian function. Kim et al., (2005) provided evidence that *NO* can prevent premature follicular atresia through the autocrine and paracrine inhibition of rat granulosa cells. The study on bovine follicular granulosa cells showed that the physiological level of *NO* might contribute to cell survival (Zamberlam et al. 2011). Fouad et al. (2012) showed that nitric oxide regulates reproductive performance in poultry as well as improving the egg-laying performance traits in Japanese quails. The results of transcriptome sequencing demonstrated that *NOS*, *ArgE*, and *ODC* were expressed in the prehierarchical follicles of geese, with the highest expression in the primary follicles ($p < 0.05$), suggesting that differentially expressed genes were active in arginine metabolism and might play a substantial role in the development of follicles. A possible reason for the higher expression of *NOS*, *ArgE* and *ODC* in the primary follicles compared to the other follicle stages might be the change in granulosa cell morphology from flattened to cuboidal shape suggesting that arginine metabolism plays a critical role in follicular development at the early stage (Sangha et al. 2011; Nóbrega Júnior et al. 2014). These results indicate that the arginine metabolic related genes play an important role during the growth and development of Zi geese prehierarchical follicles.

Conclusion

In this study, *de novo* assembly was performed to investigate five stages of prehierarchical follicles of geese using Illumina Hiseq 2500 platform. Based on the transcriptome results, we analyzed the arginine metabolic pathway and the expression of related genes in prehierarchical follicles of geese. This study explored whether the related genes in arginine metabolism pathway may play an important role in the follicular development of the geese, which would certainly provide theoretic foundation for the follow-up research on goose follicular development.

Acknowledgements

We acknowledge Xiaomei Luo (Shanghai Majorbio Biopharm Biotechnology Co., Ltd.) and Lina Sun (Beijing Ori-gene Technology Co., Ltd) for their kind help with transcriptome analysis. We also thank all the laboratory members, Lijun Wang and Wenyu Li for their help in sample collection and RNA extraction.

References

Agostinelli E, Marques MP, Calheiros R, Gil FP, Tempera G, Viceconte N, Battaglia V, Grancara S and Toninello A 2010. Polyamines: fundamental characters in chemistry and biology. Amino Acids. 38(2): 393-403.

Bautista-Ortega J, Cortes-Cuevas A, Ellis EA and Ruiz-Feria CA 2014. Supplemental L-arginine and vitamins E and C preserve xanthine oxidase activity in the lung of broiler chickens grown under hypobaric hypoxia. Poult Sci. 93(4): 979-988. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL 2009. BLAST+: architecture and applications. Bmc Bioinformatics. 10(1): 421.Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL 2009. BLAST+: architecture and applications. Bmc Bioinformatics. 10(1): 421.

Che L, Yang P, Fang Z, Lin Y and Wu D 2013. Effects of dietary arginine supplementation on the reproductive performance and immunity of sows. Czech J. Anim. Sci. 58: 167-75.

Choi Y and Rajkovic A 2006. Genetics of early mammalian folliculogenesis. Cellular and Molecular Life Sciences. 63(5): 579.

Conesa A, Götz S, Garcíagómez J M, Terol J, Talón M and Robles M 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21(18): 3674-3676.

Ding N, Han Q, Zhao XZ, Li Q, Li J, Zhang HF, Gao GL, Luo Y, Xie YH, Su J and Wang QG 2015. Differential gene expression in pre-laying and laying period ovaries of Sichuan White geese (Anser cygnoides). Genetics & Molecular Research Gmr. 14(2): 6773.

Erlich Y, Mitra PP, Delabastide M, Mccombie WR and Hannon GJ 2008. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. Nature Methods. 5(8): 679-82.

Fouad AM, El-Senousey HK, Yang XJ and Yao JH 2012. Role of dietary L-arginine in poultry production. International Journal of Poultry Science. 11: 718.

Gao G, Li Q, Zhao X, Ding N, Han Q, Su J and Wang Q 2015. Transcriptome profiling of the hypothalamus during prelaying and laying periods in Sichuan white geese (A nser cygnoides). Animal Science Journal. 86(8): 800-05.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29(7): 644.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N and Regev A 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8(8): 1494-1512.

Johnson AL 2015. Ovarian follicle selection and granulosa cell differentiation. Poultry Science. 94(4): 781.

Kim H, Moon C, Ahn M, Lee Y, Kim H, Kim S, Ha T, Jee Y and Shin T 2005. Expression of nitric oxide synthase isoforms in the porcine ovary during follicular development. Journal of Veterinary Science. 6(2): 97-101.

Langmead B and Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9(4): 357-359.

Langmead B, Trapnell C, Pop M and Salzberg SL 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 10(3): 25.

Lin PF, Hao Y B, Guo HL, Liu HL and Rui R 2010. Role of Fas/FasL on apoptosis of porcine follicular granulosa cells derived from isolated follicles during culture in vitro. Zoological Research. 31(3): 268.

Liu H, Zhang W, Li Q, Liu J, Zhang T, Zhou T, Li L, Wang J, Xu H and He H 2015. The comprehensive mechanisms underlying nonhierarchical follicular development in geese (Anser cygnoides). Animal Reproduction Science. 159: 131-40.

Mateo RD, Wu G, Bazer FW, Park JC, Shinzato I and Kim SW 2007. Dietary L-arginine supplementation enhances the reproductive performance of gilts. Journal of Nutrition. 137(3): 652.

Matsuda F, Inoue N, Manabe N and Ohkura S 2012. Follicular growth and atresia in mammalian ovaries: regulation by survival and death of granulosa cells. The Journal of reproduction and development. 58: 44-50.

Murakami AE, Fernandes JIM, Hernandes L and Santos TC 2012. Effects of starter diet supplementation with arginine on broiler production performance and on small intestine morphometry. Pesquisa Veterinária Brasileira. 32: 259-66.

Nóbrega Júnior JE, Rossetto R, Matos MHT, Chaves RN, Magalhães DM, Lima-Verde IB, Báo SN, Campello CC, Figueiredo JR, Gonçalves PBD and Oliveira JFC 2014. Sphingosine 1-phosphate promotes activation of caprine preantral follicle in vitro. Arq Bras Med Vet Zoo. 66(4): 977-985.

Oso AO, Williams GA, Oluwatosin OO, Bamgbose AM, Adebayo AO, Olowofeso O, Pirgozliev V, Adegbenjo AA, Osho SO and Alabi JO 2017. Growth performance, nutrient digestibility, metabolizable energy, and intestinal morphology of growing turkeys fed diet supplemented with arginine. Livestock Science. 198: 24-30.

Sangha GK, Khera KS and Kapil K 2011. Changes in number of granulosa cells during folliculogenesis in prepubertal goat ovary. Indian J Anim Sci. 78(4): 348-352.

Shimizu T, Kayamori T, Murayama C and Miyamoto A 2012. Bone morphogenetic protein (BMP)-4 and BMP-7 suppress granulosa cell apoptosis via different pathways: BMP-4 via PI3K/PDK-1/Akt and BMP-7 via PI3K/PDK-1/PKC. Biochemical & Biophysical Research Communications. 417(2): 869-873.

Tapiero H, Mathé G, Couvreur P and Tew KD 2002. I. Arginine. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 56: 439-45.

Tong BC and Barbul A 2004. Cellular and physiological effects of arginine. Mini reviews in medicinal chemistry. 4: 823-32.

Wang A, Liu F, Chen S, Wang M, Jia R, Zhu D, Liu M, Sun K, Wu Y and Chen X 2015. Transcriptome analysis and identification of differentially expressed transcripts of immune-related genes in spleen of gosling and adult goose. International journal of molecular sciences. 16(9): 22904-26.

Weiger TM and Hermann A 2014. Cell proliferation, potassium channels, polyamines and their interactions: a mini review. Amino Acids. 46(3): 681-8.

Wiesinger H 2001. Arginine metabolism and the synthesis of nitric oxide in the nervous system. Progress in neurobiology . 64: 365-91.

Wu G. Amino acids : biochemistry and nutrition. CRC Press. 2013.

Xu Q, Song YD, Chen Y, Liu R, Zhang Y, Li Y, Huang ZY, Zhao WM, Chang GB and Chen GH 2015. Molecular cloning and expression patterns of the cholesterol side chain cleavage enzyme (CYP11A1) gene during the reproductive cycle in goose (Anas cygnoides). Journal of animal science and biotechnology. 6(2):54.

Xu Q, Zhao WM, Chen Y, Tong YY, Rong GH, Huang ZY, Zhang Y, Chang GB, Wu XS and Chen GH 2013. Transcriptome profiling of the goose (Anser cygnoides) ovaries identify laying and broodiness phenotypes. PloS one. 8(2): e55496.

Yang EW, Girke T and Jiang T 2013. Differential gene expression analysis using coexpression and RNA-Seq data. Bioinformatics. 29 (17): 2153-2161.

Yang W, Rosenstiel PC and Schulenburg H 2016. ABSSeq: a new RNA-Seq analysis method based on modelling absolute expression differences. BMC Genomics. 17(1): 541.

Yen CF, Lin EC, Wang YH, Wang PH, Lin HW, Hsu JC, Wu LS, Jiang YN and Ding ST 2009. Abundantly expressed hepatic genes and their differential expression in the liver of prelaying and laying geese. Poultry science. 88 (9) :1955-1962.

Yu J, Lou YP and Zhao A 2016. Transcriptome analysis of follicles reveals the importance of autophagy and hormones in regulating broodiness of Zhedong white goose. Scientific Reports. 6: 36877.

Zamberlam G, Portela V, Oliveira JFCD, Gonçalves PBD and Price CA 2011. Regulation of inducible nitric oxide synthase expression in bovine ovarian granulosa cells. Molecular & Cellular Endocrinology. 335(2): 189-194.

Zeng X, Mao X, Huang Z, Wang F, Wu G and Qiao S 2013. Arginine enhances embryo implantation in rats through PI3K/PKB/mTOR/NO signaling pathway during early pregnancy. Reproduction. 145(1): 1-7.

Zhang X, Kang B, Zhang LN, Guo JR, Jiang DM, Ji H, Zhen L and Yang HM 2013. Gene expression profiles of LH, prolactin and their receptors in female Zi geese (Anser cygnoides) during development. Folia Biologica. 61(1-2): 59-64.