RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Integrated analysis of miRNA and mRNA expression profiles in development of porcine testes

Maoliang Ran[1,3], Bin Chen[1,3], Maisheng Wu[2], Xiaochun Liu[1,3], Changqing He[1,3], Anqi Yang[1,3], Zhi Li[1,3], Yongjun Xiang[2], Zhaohui Li[2], Shanwen Zhang[2]

[1]College of Animal Science and Technology, Hunan Agriculture University, Changsha, 410128, China.

[2]Xiangtan Bureau of Animal Husbandry and Veterinary Medicine and Aquatic Product, Xiangtan, 411104, China.

[3]Hunan Provincial Key Laboratory for Genetic Improvement of Domestic Animal, Changsha, 410128, China.

Correspondence information: Bin Chen, College of Animal Science and Technology, Hunan Agricultural University, Hunan Changsha 410128, P.R. China. Tel: +86 0731 84618176, Fax: +86 0731 84610280, E-mail: chenbin7586@126.com.

Abstract

To understand the complex physiological process underlying pig testes development and spermatogenesis, this study aims to characterize the change in miRNA and mRNA profiles at four developmental stages of embryonic and postnatal testes, including 60 dpc (days post coitus, E60), 90 dpc (E90), 30-day-old (D30) and 180-day-old (D180). A total of 304 mature, 50 novel miRNAs, and 8343 differentially-expressed genes were identified. 93 (48 up and 45 down), 104 (49 up and 55 down), 122(49 up and 73 down) differentially-expressed miRNAs, as well as 1007 (646 up and 361 down), 1929 (911 up and 1018 down), 7420 (3998 up and 3422

down) differentially-expressed genes were identified in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively. Integrating analysis of miRNA and mRNA expression profiles predicted more than 50,000 miRNA-mRNA interaction sites. GO and KEGG pathway analysis of the predicted target genes illustrated the likely roles of differentially expressed miRNAs in testes development and spermatogenesis. Such as, PI3K-Akt signaling pathway and Hippo signaling pathway related developmental, and Carbon metabolism, Fatty acid metabolism, Protein digestion and absorption were involved in metabolite synthesis**.** These integrated high-throughput expression data show that miRNA is a critical factor in porcine testes development, providing a useful resource to understand global genome expression change in porcine testes development and spermatogenesis.

Keywords: RNA-seq, miRNA, mRNA, Porcine, Testes

Introduction

Increasing compelling investigations have attempted to investigate physiological process of porcine testes development and spermatogenesis through miRNA and mRNA profile. After over expression of the testis development-related protein-1 gene (TDRP1) in pig testes, 29 common genes, including 17 up and 12 down-regulated genes, are identified using microarray screening[1]. Using high-throughput RNA-seq, 147 and 109 genes are identified over expressed in the gonad of Large White and Iberian pig, respectively [2]. These differentially expressed genes are enriched in reputation, developmental process and fatty acid metabolic process. 9,061 novel

transcripts are detected in pig testes, including 252 pigs testes specific transcripts, enriching in quantitative trait loci regions for reproduction traits in pigs[3]. In addition, microRNA (miRNA) is a class of non-coding RNA (about 22 nt), which plays important roles in regulating cell growth, development, and apoptosis[4-7]. miRNA expression profile study showed that a large number of miRNAs are differentially expressed in specific developmental stage of the human or mouse testes[8, 9], however, limited study focus on pig testes. Microarry-based and deep sequencing approaches have identified mature miRNAs preferentially expressed in porcine immature and mature testes[10, 11]. Further functional study showed that miRNAs regulates testes development and spermatogenesis[12-14].

Although these miRNA and miRNA profile studies are contributed to investigate the intricate physiological process in pig testes development and spermatogenesis. However, this physiological process is starting in embryonic early stage and maintaining to adult[15, 16], while previous study focus on postnatal developmental stage. A miRNA and mRNA profiles which include embryonic and postnatal developmental stages is indispensably to study the physiological process of pig testes development and spermatogenesis. In addition, the first step is to predict the target genes in understanding the function roles of miRNAs. Some new key miRNA-mRNA interaction pairs are predicted by integrating analysis of miRNA and mRNA profiles[17-20], which point out that it's preferred to establish more miRNA and mRNA transcriptome data. In the present study, we investigated miRNA and mRNA in the testes of Shaziling pig (a Chinese indigenous breed) at four representative

stages: 60 dpc (days post coitus, E60), 90 dpc (E90), 30-day-old (D30) and 180-day-old (D180) using IlluminaHiseq technology. A total of 304 mature and 50 novel miRNAs were identified in these samples. For the mRNA, 646 up- and 361 down-regulated, 911 up- and 1018 down-regulated, 3998 up- and 3422 down-regulated genes were identified in E90vs. E60, D30vs. E90 and D180vs. D30, respectively, and 367 genes were shared. A combined analysis was conducted to identity target genes of miRNAs and to characterize their functional roles in porcine testes development, and 55,339 miRNA-mRNA interaction sites were predicted. Which provide a useful resource to understand global genome expression change and the regulation roles of miRNAs in porcine testes development and spermatogenesis.

Results

Raw data used in this study were submitted in the National Center for Biotechnology Information Short Reads Archive. The accession number is SRP056644.

Overview of small RNA libraries

In order to identify differentially expressed miRNA during development of porcine testes tissues, four small RNA libraries were constructed and sequenced by the IlluminaHiseq 2500 platform. A total of 6,695,453, 6,474,804, 6,457,999 and 6,645,391 raw reads were generated in E60, E90, D30 and D180 libraries, respectively. Firstly, removing the low-quality sequences and adaptors, and then discarding the sequences shorter than 18 nt, 6,347,414, 6,287,768, 6,148,364 and 6,279,272 clean reads were obtained and used for further analysis (Table 1). We then aligned clean reads against the pig genome using Bowtie[21] with perfect matches,

5,853,113(92.21%), 5,911,514(94.02%), 5,678,650(92.36%) and 5,615,968(88.67%) reads from each library, respectively. Approximately 1.3% (81,128) of clean reads could be mapped to rRNA, snRNA, sonRNA and tRNA deposited in NCBI. The un-annotated reads accounted for an average of 8,192,733 which comprised most of the total reads. Reads length distributions of four libraries are shown in Fig. 1. 21 to 23 nt small RNAs were the main size and accounted for at least 70% of the population in E60, E90 and D30, while the D180 had the main size from 29 to 31 nt and followed by 28 nt and 32 nt. These reads of longer than 25 nt that mostly may represent Piwi-interacting RNA (piRNA), a newly identified class of small regulatory RNAs, which has been reported that abundantly generated in the mature testes of animals[10, 22, 23].

microRNA profiling of pig testes in different development stages

In order to identify the known porcine miRNAs, we compared the clean reads of each library with known porcine miRNA precursors in miRBase 20.0. The statistical results showed that a total of 304 mature miRNAs were detected in four libraries (Table S2). In order to know more miRNAs which express in the testes of pig, miRvo[24] and mirdeep2[25] were used to identify porcine novel miRNAs and 50 novel miRNAs were predicted in four libraries (Table S3). The results of further analysis indicated that E60, E90, D30 and D180 testes shared 237 known and 15 novel miRNAs. A total of 285, 291, 287 and 243 known mature miRNAs and 43, 34, 40 and 19 novel miRNAs were identified from E60, E90, D30 and D180, respectively.

These above mentioned mature and novel miRNAs were used differentially

expression analysis by calculating the $\log_2$Ratio with q-value＜0.01 and |$\log_2$(Fold change)|＞1 as the cut-off. A total of 48 up- and 45 down-regulated, 49 up- and 55 down-regulated, and 49 up- and 73 down-regulated miRNAs were detected in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively (Fig. 2A; Table S4). Venn diagrams showed that 42 miRNAs were shared among these three groups (Fig. 2B). In addition, the hierarchial clustering results of differentially expressed miRNAs between libraries depicted that there are much more differences in miRNAs expressed in porcine testes development (Fig. 2C). All differentially expressed miRNAs between libraries were clustered together after 6 rounds of clustering.

mRNA expression profiles in porcine testes development

For knowing the mRNA expression profiles in porcine testes development, a total of four libraries, E60, E90, D30 and D180, were constructed and sequenced by IlluminaHiseq 2500 platform. The major characteristics of four libraries were summarized in Table 2. The E60, E90, D30 and D180 libraries were found contain 58,184,970, 51,031,524, 50,603,772 and 43,386,314 raw reads, respectively. After removing low-quality reads, adaptors and all possible contaminants, 56,296,294, 49,177,458, 48,837,300 and 41,545,444 clean reads were retained, respectively. The error rate and GC content of each library were calculated for controlling the quality of libraries (Table 2). The quality of the process of library construction was also assessed using transcript homogeneity (Fig. S1). These results showed that the quality of our four libraries were excellent.

The clean reads were aligned to the reference genome of *Susscrofa* by using

Tophat v2.0.9[26]. The results showed that 77.89%, 78.37%, 77.44% and 78.35% of clean reads were matched the *Susscrofa* genome from E60, E90, D30 and D180 libraries, respectively. In total, 72.47%, 72.65%, 72.37% and 72.97% of clean reads from these above mentioned libraries were uniquely mapped to the reference genome. In addition, more than 40% of clean reads were non-splice reads in each sample. However, 24.29%, 27.19%, 28.97% and 31.13% of clean reads were mapped to the border of exon (also called junction reads) from the E60, E90, D30 and D180 libraries, respectively (Table 2). For realizing the gene expression, the reads numbers mapped to each gene were calculated and normalized to RPKM. The statistic result of RPKM interval and total genes of each library was shown in Table S5. Quantitative saturation analysis was performed to reflect the requirement of the number of sequencing data for quantitative the level of gene expression. As shown in Fig. S2, the RPKM with more than 1 almost reach plateau when the proportion of mapped reads approached 70%, while the RPKM less than 1 was detected a significant shower at approximately 80%. These results showed that the level of gene expression of each library was accurate.

The differentially expressed genes between libraries were screened based on DEGSeq[27] analysis with taking $|\log_2(\text{Fold change})|>1$ and $P<0.05$ as cut-off. Taking a viewpoint on adjoining libraries, 646 up- and 361 down-regulated, 911 up- and 1018 down-regulated, 3998 up- and 3422 down-regulated genes were identified in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively (Fig. 3A). Venn diagram were generated by using their data and depicted that these three groups shared 367

genes (Fig. 3B). As shown in Fig. 3C of the hierarchial clustering results, differentially expressed genes were divided into four groups with eight clusters while gene expression altogether was similar form E60, E90 to D30, but they were differentially expressed compared to D180 library.

Gene ontology (GO) and pathway enrichment analysis were used to explore the functions of differentially expressed genes in testes development. In GO enrichment analysis, a total 213 GO terms involved in biological processes, molecular functions and cellular components were significantly ($P<0.05$) enriched across the above mentioned three different DEGs' groups, while 54 GO terms were found when taken $P<0.001$ as cut-off (Table S6). 71(11($P<0.001$)), 31(10) and 111(33) GO terms were detected in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively. In E90 vs. E60, the most enriched GO terms were mainly involved in binding, protein binding and extracellular region etc. In D30 vs. E90, catalytic activity, oxidoreuctase activity and single-organism metabolic process were the mainly GO terms, while the D180 vs. D30 were mainly represented by binding, catalytic activity, cell part, metabolic process and organic substance metabolic process etc.

The KEGG pathway enrichment analysis were performed on these differently expressed genes and also taken a cut-off criterion of $P<0.05$. A total of 62 enriched pathway were detected, of these, 15, 27 and 20 enriched pathway were found from E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively (Table S7). The classifications indicated that focal adhesion, PI3K-Akt signaling pathway, metabolic pathways, ribosome, endocytosis and carbon metabolism etc. were highly enriched.

Furthermore, there was three overlapping pathway in E90 vs. E60, D30 vs. E90 and D180 vs. D30, focal adhesion, ECM (extracellular matrix)-receptor interaction and phagosome, which indicated that these three pathways have a critical role in testes development.

Confirmation of differential miRNAs and mRNAs by qRT-PCR

To validate the RNA-seq results, we used real-time qRT-PCR to investigate the relative expression levels of randomly selecting 6 miRNAs and 6 mRNAs (miR451, miR10a-5p, *INSL3*, *COL1A1* from E90 vs. E60, novel_46, miR191, *DHRS4*, *CLU* from D30 vs. E90 and miR9-1, miR450a, *HBA*, *PRM1* from D180 vs. D30). As shown in Fig. 4, the results of RNA-seq data and qRT-PCR data were identical. Three genes (novel_46, *INSL3* and *PRM1*) did not show consistent expression between RNA-seq data and qRT-PCR data, which probably caused by biological differences between samples and the sensitivity and capability of the different methods. In general, the results of qRT-PCR validated the RNA-seq results and added more weight to credibility of the differentially expressed miRNAs and genes.

Combined expression analysis of microRNAs and their target mRNAs

In order to minimize false positive rates to predict the miRNA-mRNA interaction pairs, the data of miRNA-seq and mRNA-seq were correlated analysis followed the criterion of anti-regulation of a miRNA and a corresponding mRNA which was used in previous studies [18, 28]. miRanda[29] was used to predict the target genes of miRNAs. As shown in Figure 2A and Figure 3B, the differentially expressed miRNAs and mRNAs of E90 vs. E60, D30 vs. E90 and D180 vs. D30 were performed

correlated expression analysis for predicting miRNAs targets. As a result, 55,339 miRNA-mRNA interaction sites were predicted in our study. In addition, one target (DUSP10) of miR-450a, three targets (COL1A2, COL14A1 and PLXND1) of let-7g and one target (PAK2) of miR-26a were validated by performing qRT-PCR in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively. These results showed that it's a reliable method to predict target genes of miRNAs by integrating analysis the data of RNA-seq and mRNA-seq.

In our further analysis, we interested in the functional roles of target genes in porcine testes development according GO and KEGG pathway enrichment analysis. In GO enrichment analysis, a total of 65 GO terms were significantly enriched in these three differentially expressed groups. More than half of GO terms were enriched at the E90 vs. E60 with 38 terms, while 17 and 10 GO terms were enriched at D30 vs. E90 and D180 vs. D30, respectively (Fig. 6). In the E90 vs. E60, the 38 GO terms mainly referred to the molecular function, such as binding, protein binding, cation binding and metal ion binding, while more than half of GO terms were enriched biological process, such as cell adhesion, biological adhesion, cell migration, cell growth, regulation of embryonic development and regulation of development process(Fig. 6A). In the D30 vs. E90, catalytic activity, oxidoreductase activity and single-organism metabolic process were the most significantly enriched terms(Fig. 6B). In the D180 vs. D30, almost all of GO terms were enriched in metabolic process, such as organic substance metabolic process, primary metabolic process, cellular metabolic process, followed by protein and single-organism metabolic process (Fig.

6C).

KEGG analysis identified a total of 27 pathways (Table S8). These include the categories “Focal adhesion”, “proteoglycans in cancer”, “PI3K-Akt signaling pathway” and “Hippo signaling pathway” related developmental, while “Carbon metabolism”, “Fatty acid metabolism”, “Fatty acid degradation”, “Glyoxylate and dicarboxylate metabolism”, “Protein digestion and absorption” and “Citrate cycle” were involved in metabolite synthesis. Otherwise, 7 and 6 pathways were co-represented in E90 vs. E60 and D30 vs. E90, D30 vs. E90 and D180 vs. D30, respectively (Table S9), showing that these pathways are significantly regulated in the four stages of testes development investigated in this study.

Discussion

Most of clean reads (77.44-78.37%) identified by this study could match the released *S. Scrofa* genome, which was similar to the described in the pig muscle and ovary transcriptome (78.7%)[30], however, it was lower than that of pig adipose (80-87%)[31]. We found that the remained clean reads (21.63-22.56%) were mismatched, which could be due to the low sequence coverage of the swine reference genome (0.66×)[32]. In addition, a large proportions of splice reads (24.29-31.13%) were found in our study, which is much higher than that in other species[33, 34], reinforces the need to improve the current pig annotation.

A total of 304 mature and 50 novel miRNAs were identified, while almost all novel miRNAs had a small reads number, which suggest that their roles in testes development or spermatogenesis need to be further confirmed. 93, 104 and 122

differentially-expressed miRNAs were identified in E90 vs. E60, D30 vs. E90 and D180 vs. D30, respectively, with 42 miRNAs were shared in these three groups. 122 miRNAs are identified from the porcine developing testes at D30 and D180 developmental stages using Solexa deep sequencing, while 96 up and 26 down-regulated in D180[10]. In addition, 51 up and 78 down-regulated in D180 vs. D60 are detected from the developing testes of Large White pig using microarray technology [11]. A previous study showed that 156 miRNAs/miRNA*s were identified in the testes of adult Tibetan pigs (210-days-old, a black, Chinese indigenous breed)[35]. All of these differences pointed that sequencing data from different studies are usually hard to be repeated, which can be explained mainly by experimental technology, breeds, developmental stages, the threshold and small animal number. But, our study found a large of miRNA express profiles variability compared with these previous studies through constructing and sequencing four libraries involved embryo to adult, and thus provided a more comprehensive result.

Approximately ten thousand differentially-expressed genes were detected in our four mRNA-seq libraries, while most of them were identified between D30 and D180 libraries. All of these differentially-expressed genes may explain the physiological process of the development of porcine testes since lots of important differentially-expressed genes at the four different physiological stages may participate in porcine testes development and spermatogenesis. There are several lines of evidence, SRY HMG box related gene 9 (*Sox9*), one of the differentially-expressed genes in all libraries, is activated by *Sry* (sex determining region of the Y, also

identified in our study) and maintained by SOX9 protein both of which directly activate the core 1.3 kb testes-specific enhancer of *Sox9* [36], and it functions as a critical Sertoli cell differentiation factor, perhaps in all vertebrates [37]. GATA binding protein 4 (*GATA4*), a differentially-expressed gene, serves as a key transcriptional regulator for proper development of the murine fetal testes, Sertoli cell function in adult mice [38], and human testicular development [39]. In addition, the physical interaction between the transcription factor GATA4 and its co-factor FOG2 are required for normal gonadal development through regulating the normal expression of *Sox9* and *Sry* [40, 41]. The signals which come from the transforming growth factor β (TGF-β) superfamily of proteins play critical roles in governance of the testes development and spermatogenesis [42]. In our study, several members of TGF-β superfamily genes were detected in the differentially-expressed gene analysis between samples, such as *TGF-β1*, *TGF-β2*, *TGF-β3*, *TGF-βR3*, *BMP* and *BMP6*. The expressive quantity of these TGF-β superfamily genes in E60 and E90, especially in E90, were higher than D30 and D180, which indicate that TGF-β signaling may play more important regulation roles in porcine testes development in embryonic.

Previous studies have indicated that many pathways participate in regulating testes development and spermatogenesis, such as MAPK [43, 44], Hedegehog, Wnt/β-catenin [45, 46] and PI3K-Akt signaling pathway [47, 48]. Many protein kinases, such as MAPK, MAP kinase kinase (MAP2K) and MAP kinase kinasekinase (MAP3K) , play vital roles in activating MAPK pathway [44]. In our analysis, *MAPK*, *MAP2K* and *MAP3K* were up-regulated in young testes (D30) and activated the

MAPK pathway, which might promote testes development and spermatogenesis during testes early development stage, while these genes were down-regulated in adult testes (D180). But, MAPK pathway still plays crucial roles in spermatogenesis through other regulation styles. For example, Nek2, a serine-threonine kinase, activated by the MAPK pathway in mouse pachytene spermatocytes [49] and *HMGA2* (High-Mobility Group Protein A2) were up-regulated in adult testes (D180) in our study. Previous investigations reveal that the functional interaction between Nek2 and HMGA2 plays a crucial role in the correct process of chromatin condensation in meiosis in mouse spermatocytes [50]. Wnt/β-catenin pathways also was revealed play an important suppression role in mouse and human spermatogonia [45], and the suppression of Wnt/β-catenin signaling is a prerequisite for the normal development of Primordialgermcells [46]. *Wt1*, a up-regulated gene in young testes (D30), was certified that it is a negative regulator of β-catenin signaling during testes development [51]. The activation of PI3K-Akt signaling pathway play s a central role in embryonic testes cord formation, mesonephros cell migration and the self-renewal division of spermatogonial stem cells [47, 48]. In our study, a total of 32 and 8 up-regulated genes, detected in E60 and E90 testes were significantly enriched in the PI3K-Akt signaling pathway, which indicated that the PI3K-Akt signaling pathway were activated in embryonic testes. In addition, *Akt3* gene was up-regulated in adult testes (D180) in our analysis, which might suggest the PI3K-Akt signaling pathway participate in regulating spermatogenesis. More importantly, a large number of differentially-expressed genes in D30 vs. E90 were significantly enriched in

metobolic pathways, such as Glyoxylate and dicarboxylate, Fatty acid, Propanoate, Pyruvate, and Glutathione metabolisms, which provided important evidence that the metabolism of porcine testes after born is much more complexity than that in embryonic.

In general, miRNAs play important gene-regulatory roles in animals by targeting the mRNAs of protein coding genes and repressing their posttranscriptional [52]. In this way, an up-regulation of a miRNA indicates a decrease activity of the target genes. It is an important step to identify the miRNA target genes for understanding their roles in gene regulatory networks. Recently, various studies taken the computational methods to predict the target genes of miRNAs, and integrating analysis of miRNA and mRNA expression profiles is one way to minimize false positive rates and identify the real target genes [18, 28]. In the current study, 55,339 miRNA-mRNA interaction sites were predicted, which is far above our expectation. Lower fold change used as criterion to identify differentially expression miRNAs and genes may be a crucial reason to explain this result, and we will identify less interaction pairs through lowering the threshold. In addition, the online database of predicting targets can't be overlooked, which may be not represent the actual existing interactions. In previous study, limited interactions predicted in our study, were certified using different methods. For example, miR34c be pivotal in murine embryonic stem cells differentiation into male germ cells through targeting retinoic acid receptor gamma gene [53]. In mouse, miR18, a member of Oncomir-1, was revealed that it regulated heat shock factor 2 (HSF2) activity in spermatogenesis and link miR18 and

HSF2-mediated male germ cell maturation [54]. However, a large number of interactions failed to be certified, and some false positive predictions may still involve in our study. In the further studies, the interactions predicted in our study will be validated through different experimental methods, and some key interaction sites may be found in testes development and spermatogenesis.

Materials and methods

Ethics statement

This study was conducted according to the guidelines of the Declaration of Helsinki and all procedures involving animal subjects were approved by the animal welfare committee of College of Animal Science and Technology, Hunan Agriculture University (No.2013-06). Sows and piglets were had a general anesthesia (Zoletil 50, Virbac Co., France) before sampling. Animals did not suffer unnecessarily at any stage of this experiment.

Animals and Sample collection

Four healthy Shaziling sows with the same parent, were fertilized via artificial insemination method and the semen adopted from a healthy adult *Shaziling* boar. Sows and piglets were raised in the pig farm at Xiangtan bureau of animal husbandry and aquatic product (Hunan, China). Two sows were slaughtered at E60 and E90, respectively, and the testes of embryos were collected. The piglets of another two sows were castrate at D30 and D180 for obtaining the testes samples. Testes samples were immediately snap frozen in liquid nitrogen and stored at -80℃ until used.

RNA isolation and qualification

Total RNA was isolated using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA). RNA degradation and contamination were monitored on 1% agarose gels. RNA concentration and integrity were measured using Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA) and the RNA Nano 6000Assy Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively. The experimental protocols were strictly performed according to the manufacturers' technical instructions.

miRNA sequencing and analysis

For constructing each small RNA library, equality total RNA of three pig testes tissues in same developmental stage were pooled. Four small RNA libraries were constructed and named E60, E90, D30 and D180, respectively. Sequencing libraries were generated using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. The small RNAs ligated with 5′ and 3′ adaptors were reverse transcribed and amplified. The libraries were quantified by the Agilent Bioanalyzer 2100 system using DNA high sensitivity chips. The library preparations were sequenced at the NovogeneBioninformatics Institute (Beijing, China) on an IlluminaHiseq 2500 platform and 50bp single-end reads were generated after the clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After removing adapter sequences, reads containing poly-N and low quality reads, all of the clean reads were mapped to Repeatmasker

and Rfam database (ftp://selab.janelia.org/pub/Rfam) to remove tags originating from repeat sequences, rRNA, tRNA, snRNA and snoRNA. To identify known porcine miRNA, the remaining clean reads were further compared with the mature miRNAs in miRBase 20.0 (http://www.mirbase.org/) and their reads were counted. For predicting porcine novel miRNA, we used the available software miRvo[24] and mirdeep2[25]. The miRNA expression levels were estimated by TPM (transcript per million) values (TPM=(miRNA total reads/total clean reads)$\times 10^6$)[55]. Differential expression analysis between groups was performed using the DEGseq[27] R package. P-value was adjusted using q-value[56], and q-value$<$0.01 and |log2(Fold change) |$>$1 was set as the threshold for significantly differential expression by default. GO enrichment analysis was implemented using GOseq[57] R package as well as GO terms with q-values less than 0.05 were considered significantly enriched by differentially expressed miRNAs. KEGG enrichment analysis on differential expressed miRNAs was performed by KOBAS[58] software using hypergeometric test.

mRNA sequencing and statistical analysis

The total RNA from the testes of three boars of different developmental stage (E30, E90, D30 and D180) were pooled with equal quantity to construct cDNA libraries following the manufacturer's technical instruction. Sequencing libraries were generated using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA) in accordance with the manufacturer's recommendations. After cluster generation, four libraries were sequenced at the NovogeneBioninformatics Institute

(Beijing, China) on an IlluminaHiseq 2500 platform and 125bp paired-end reads were generated. Clean reads were obtained by filtering out adaptor sequences and removing low quality reads from raw data. Then the clean reads were aligned to the reference genome of *Susscrofa* 10.2 at ftp://ftp.ensembl.org/pub/release-74/fasta/sus_scrofa/dna/ by using Tophat[26] v2.0.9. Gene expression level was calculated by Reads Per Kilo base per Million reads (RPKM)[59] after the reads numbers mapped to each gene was counted by HTSeq[60] v0.6.1. Differential genes were performed by DEGSeq[27] R package (1.12.0), corrected P-value of 0.05 and absolute value of log2(Fold change) of 1 were set as the threshold for significantly differential expression. The methods of GO and KEGG enrichment analysis on differential expressed genes were similar with the analysis of differential expressed miRNAs.

Quantitative real-time PCR validation of miRNAs and mRNA

After acquiring high quality total RNA, miRNAs and mRNAs were reverse transcribed by using miRNA first-strand cDNA synthesis kit (CWBIO, China, CW2141) and RevertAid™ First Strand cDNA Synthesis Kit(Fermentas, Republic of Lithuania, 11917-020) respectively. Quantitative real-time PCR (qRT-PCR) analyses on the miRNAs and the mRNAs were performed by using the SYBRgreen PCR Master Mix (ABI, USA, 4304437) in a Thermo PIKO REAL 96 system. 5s RNA and β-actin were used as internal controls of miRNA and mRNA, respectively. Primer sequences of all miRNA and mRNA were designed by using Primer 5 (Additional file1: Table S1). For miRNA quantification, the thermal cycling program was set as follows: 50℃ for 2 min, 95℃ for 10 min, followed by 40 cycles of 95℃ for 5 s and

60℃ for 30 s. For mRNA, the reaction conditions were: 95℃ for 10 min, followed by 40 cycles of 95℃ for 10 s and 59℃ for 50 s. Relative miRNA and mRNA expression were evaluated using the $2^{-\Delta\Delta Ct}$ method. At least three independent biological replicates were used for each miRNAs and mRNA.

Conclusion

In this study, four miRNA and mRNA libraries were constructed and sequenced with porcine testis from 60 dpc (days post coitus), 90 dpc, 30-day-old and 180-day-old. A total of 304 mature and 50 novel miRNAs were identified in these four testis developmental stages. For the mRNA, 646 up- and 361 down-regulated, 911 up- and 1018 down-regulated, 3998 up- and 3422 down-regulated genes were identified in E90 vs. E60, D30 vs. E90 and DM vs. D30, respectively, while 367 differentially-expressed genes were co-expressed. These findings provide new insights into the embryonic and postnatal developing process of the pig testis. Our on-going effort will focus on some interaction sites through experimental approaches for expecting to provide more fundamental information in understanding these regulatory mechanisms of porcine testis development or spermatogenesis at molecular level.

Acknowledgements

M. L. Ran, B. Chen and X. C. Liu conceived and designed the experiments. M. L. Ran, M. S. Wu, C. Q. He, Y. J. Xiang, Z. H. Li and S.W. Zhang performed the experiments. M. L. Ran, A. Q. Yang, Z. Li analyzed the data. M. L. Ran wrote the manuscript. All authors read and approved the final manuscript and declare that they

have no competing interests. This work was financially supported by China Agriculture Research System (CARS-36).

Dedications

The authors would like to appreciate Jian Zhao to perform the experiments and Wenkai Ren and Guan Yang for critical reading of the manuscript.

References

1. M. Liu, W. Zhang, W. Luo and Q. Nie, *DNA Cell Biol*, 2013, 32(9):511-6.
2. A. Esteve-Codina, R. Kofler, N. Palmieri, G. Bussotti, C. Notredame and M. Perez-Enciso, *BMC Genomics*, 2011, 12:552.
3. Z. Q. Du, C. J. Eisley, S. K. Onteru, O. Madsen, M. A. Groenen, J. W. Ross and M. F. Rothschild, *Anim Genet*, 2014, 45(2):198-204.
4. C. Clarke, M. Henry, P. Doolan, S. Kelly, S. Aherne, N. Sanchez, P. Kelly, P. Kinsella, L. Breen, S. F. Madden, L. Zhang, M. Leonard, M. Clynes, P. Meleady and N. Barron, *BMC Genomics*, 2012, 13:656.
5. D. Baltimore, M. P. Boldin, R. M. O'Connell, D. S. Rao and K. D. Taganov, *Nat Immunol*, 2008, 9(8):839-45.
6. I. Alvarez-Garcia and E. A. Miska, *Development*, 2005, 132(21):4653-62.
7. A. M. Cheng, M. W. Byrom, J. Shelton and L. P. Ford, *Nucleic Acids Res*, 2005, 33(4):1290-7.
8. N. Yan, Y. Lu, H. Sun, D. Tao, S. Zhang, W. Liu and Y. Ma, *Reproduction*, 2007, 134(1):73-9.
9. J. Zhang, Q. Liu, W. Zhang, J. Li, Z. Li, Z. Tang, Y. Li, C. Han, S. H. Hall and Y. Zhang, *Acta Biochim Biophys Sin (Shanghai)*, 2010, 42(2):145-53.
10. C. Lian, B. Sun, S. Niu, R. Yang, B. Liu, C. Lu, J. Meng, Z. Qiu, L. Zhang and Z. Zhao, *FEBS J*, 2012, 279(6):964-75.
11. L. Luo, L. Ye, G. Liu, G. Shao, R. Zheng, Z. Ren, B. Zuo, D. Xu, M. Lei, S. Jiang, C. Deng, Y. Xiong and F. Li, *PLoS One*, 2010, 5(8):e11744.
12. J. Bao, D. Li, L. Wang, J. Wu, Y. Hu, Z. Wang, Y. Chen, X. Cao, C. Jiang, W. Yan and C. Xu, *J Biol Chem*, 2012, 287(26):21686-98.
13. Q. E. Yang, K. E. Racicot, A. V. Kaucher, M. J. Oatley and J. M. Oatley, *Development*, 2013, 140(2):280-90.
14. J. Rakoczy, S. L. Fernandez-Valverde, E. A. Glazov, E. N. Wainwright, T. Sato, S. Takada, A. N. Combes, D. J. Korbie, D. Miller, S. M. Grimmond, M. H. Little, H. Asahara, J. S. Mattick, R. J. Taft and D. Wilhelm, *Biol Reprod*, 2013, 88(6):143.
15. Y. Matsui, *J Androl*, 2010, 31(1):61-5.
16. D. G. de Rooij and L. D. Russell, *J Androl*, 2000, 21(6):776-98.
17. S. Yun, S. U. Lee, J. M. Kim, H. J. Lee, H. Y. Song, Y. K. Kim, H. Jung, Y. J. Park, S. R. Yoon, S. R. Oh, T. D. Kim and I. Choi, *PLoS One*, 2014, 9(10):e108913.
18. M. Tang, D. Mao, L. Xu, D. Li, S. Song and C. Chen, *BMC Genomics*, 2014, 15:835.
19. S. Sree, K. Radhakrishnan, S. Indu and P. G. Kumar, *Biol Reprod*, 2014, 91(3):69.

20. J. G. Joung, D. Kim, S. Y. Lee, H. J. Kang and J. H. Kim, *BMC Med Genomics*, 2014, 7 Suppl 1:S10.
21. B. Langmead, C. Trapnell and M. Pop, S. L. Salzberg, *Genome Biol*, 2009, 10(3):R25.
22. A. Girard, R. Sachidanandam, G. J. Hannon and M. A. Carmell, *Nature*, 2006, 442(7099):199-202.
23. N. C. Lau, A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel and R. E. Kingston, *Science*, 2006, 313(5785):363-7.
24. M. Wen, Y. Shen, S. Shi and T. Tang, *BMC Bioinformatics*, 2012, 13:140.
25. M. R. Friedlander, S. D. Mackowiak, N. Li, W. Chen and N. Rajewsky, *Nucleic Acids Res*, 2012, 40(1):37-52.
26. C. Trapnell, L. Pachter and S. L. Salzberg, *Bioinformatics*, 2009, 25(9):1105-11.
27. L. Wang, Z. Feng, X. Wang, X. Wang and X. Zhang, *Bioinformatics*, 2010, 26(1):136-8.
28. X. Xu, Y. Shen, J. Fu, L. Lu and J. Li, *PLoS One*, 2014, 9(11):e112722.
29. A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander and D. S. Marks, *Genome Biol*, 2003, 5(1):R1.
30. Q. Nie, M. Fang, X. Jia, W. Zhang, X. Zhou, X. He and X. Zhang, *DNA Res*, 2011, 18(5):343-51.
31. J. Corominas, Y. Ramayo-Caldas, A. Puig-Oliveras, J. Estelle, A. Castello, E. Alves, R. N. Pena, M. Ballester and J. M. Folch, *BMC Genomics*, 2013, 14:843.
32. A. L. Ferraz, A. Ojeda, M. Lopez-Bejar, L. T. Fernandes, A. Castello, J. M. Folch and M. Perez-Enciso, *BMC Genomics*, 2008, 9:173.
33. M. Mank, A. F. Santos, S. Direnberger, T. D. Mrsic-Flogel, S. B. Hofer, V. Stein, T. Hendel, D. F. Reiff, C. Levelt, A. Borst, T. Bonhoeffer, M. Hubener and O. Griesbeck, *Nat Methods*, 2008, 5(9):805-11.
34. T. Lu, G. Lu, D. Fan, C. Zhu, W. Li, Q. Zhao, Q. Feng, Y. Zhao, Y. Guo, W. Li, X. Huang and B. Han, *Genome Res*, 2010, 20(9):1238-49.
35. M. Li, Y. Liu, T. Wang, J. Guan, Z. Luo, H. Chen, X. Wang, L. Chen, J. Ma, Z. Mu, A. A. Jiang, L. Zhu, Q. Lang, X. Zhou, J. Wang, W. Zeng, N. Li, K. Li, X. Gao and X. Li, *Int J Biol Sci*, 2011, 7(7):1045-55.
36. S. Bagheri-Fam, A. H. Sinclair, P. Koopman and V. R. Harley, *Int J Biochem Cell Biol*, 2010, 42(3):472-7.
37. S. Morais da Silva, A. Hacker, V. Harley, P. Goodfellow, A. Swain and R. Lovell-Badge, *Nat Genet*, 1996, 14(1):62-8.
38. A. Kyronlahti, R. Euler, M. Bielinska, E. L. Schoeller, K. H. Moley, J. Toppari, M. Heikinheimo and D. B. Wilson, *Mol Cell Endocrinol*, 2011, 333(1):85-95.
39. D. Lourenco, R. Brauner, M. Rybczynska, C. Nihoul-Fekete, K. McElreavey and A. Bashamboo, *Proc Natl Acad Sci U S A*, 2011, 108(4):1597-602.
40. N. L. Manuylov, Y. Fujiwara, Adameyko, II, F. Poulat and S. G. Tevosian, *Dev Biol*, 2007, 307(2):356-67.
41. S. G. Tevosian, K. H. Albrecht, J. D. Crispino, Y. Fujiwara, E. M. Eicher and S. H. Orkin, *Development*, 2002, 129(19):4627-34.
42. C. Itman, S. Mendis, B. Barakat and K. L. Loveland, *Reproduction*, 2006, 132(2):233-46.
43. W. F. Marei, M. Salavati and A. A. Fouladi-Nashta, *Mol Hum Reprod*, 2013, 19(9):590-9.
44. T. Almog and Z. Naor, *Mol Cell Endocrinol*, 2008, 282(1-2):39-44.

45. N. Golestaneh, E. Beauchamp, S. Fallen, M. Kokkinaki, A. Uren and M. Dym, *Reproduction*, 2009, 138(1):151-62.
46. T. Kimura, T. Nakamura, K. Murayama, H. Umehara, N. Yamano, S. Watanabe, M. M. Taketo and T. Nakano, *Dev Biol*, 2006, 300(2):545-53.
47. M. Uzumcu, S. D. Westfall, K. A. Dirks and M. K. Skinner, *Biol Reprod*, 2002, 67(6):1927-35.
48. J. Lee, M. Kanatsu-Shinohara, K. Inoue, N. Ogonuki, H. Miki, S. Toyokuni, T. Kimura, T. Nakano, A. Ogura and T. Shinohara, *Development*, 2007, 134(10):1853-9.
49. S. Di Agostino, P. Rossi, R. Geremia and C. Sette, *Development*, 2002, 129(7):1715-27.
50. S. Di Agostino, M. Fedele, P. Chieffi, A. Fusco, P. Rossi, R. Geremia and C. Sette, *Mol Biol Cell*, 2004, 15(3):1224-32.
51. H. Chang, F. Gao, F. Guillou, M. M. Taketo, V. Huff and R. R. Behringer, *Development*, 2008, 135(10):1875-85.
52. D. P. Bartel, *Cell*, 2009, 136(2):215-33.
53. S. Zhang, M. Yu, C. Liu, L. Wang, Y. Hu, Y. Bai and J. Hua, *Cell Biochem Funct*, 2012, 30(8):623-32.
54. J. K. Bjork, A. Sandqvist, A. N. Elsing, N. Kotaja and L. Sistonen, *Development*, 2010, 137(19):3177-84.
55. L. Zhou, J. Chen, Z. Li, X. Li, X. Hu, Y. Huang, X. Zhao, C. Liang, Y. Wang, L. Sun, M. Shi, X. Xu, F. Shen, M. Chen, Z. Han, Z. Peng, Q. Zhai, J. Chen, Z. Zhang, R. Yang, J. Ye, Z. Guan, H. Yang, Y. Gui, J. Wang, Z. Cai and X. Zhang, *PLoS One*, 2010, 5(12):e15224.
56. J. D. Storey and R. Tibshirani, *Methods Mol Biol*, 2003, 224:149-57.
57. M. D. Young, M. J. Wakefield, G. K. Smyth and A. Oshlack, *Genome Biol*, 2010, 11(2):R14.
58. X. Mao, T. Cai, J. G. Olyarchuk and L. Wei, *Bioinformatics*, 2005, 21(19):3787-93.
59. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer and B. Wold, *Nat Methods*, 2008, 5(7):621-8.
60. S. Anders, P. T. Pyl and W. Huber, *Bioinformatics*, 2015, 31(2):166-9.

Supplementary material

Table S1: Primers sequence used in this study.

Table S2: Summary of known porcine miRNAs.

Table S3: Summary of novel porcine miRNAs.

Table S4: Differentially expressed miRNAs.

Table S5: RPKM distribution of different genes in each library.

Table S6: 213 GO terms of differentially expressed genes.

Table S7: 62 KEGG pathways of differentially expressed genes.

559 Table S8: 27 KEGG pathways of predicting target genes.

560 Table S9: 13 shared KEGG pathway between differentially analysis groups.

561 Figure S1: Transcript homogeneity of mRNA read libraries.

562 Figure S2: Saturation cure analysis of the gene expression read libraries generated
563 from four samples.

564 **Tables**

565 Table 1: categorization of reads of small RNAs in porcine at different developmental
566 stages.

567 Table 2: basic characteristics of mRNA.

568 **Figure legends**

569 Figure1: Length distribution and abundance of the small RNA libraries.

570 Figure 2: Differentially expressed profiles of miRNA. (A) the number of
571 differentially-expressed miRNA in each analyze group; (B) Venn diagram of
572 differential expression of miRNAs between E90 vs. E60, D30 vs. E90 and D180 vs.
573 D30; (C) Hierarchical clustering of miRNA expression. miRNA profiles from four
574 development stages of pig testes were clustered.

575 Figure 3: Differentially expressed profiles of mRNA. (A) the number of
576 differentially-expressed mRNA in each analyze group; (B) Venn diagram of
577 differential expression of mRNAs between E90 vs. E60, D30 vs. E90 and D180 vs.
578 D30; (C) Hierarchical clustering of mRNA expression. mRNA profiles from four
579 development stages of pig testes were clustered.

580 Figure 4: Validation of the RNA-seq data by qRT-PCR. Two miRNAs (A) and

581 mRNAs (B) were randomly selected from each analyze group were shown. 5s RNA
582 and β-actin were used as internal controls of miRNA and mRNA, respectively. Three
583 biological replicates were used.

584 Figure 5: Validation of the predicted target genes of miRNAs by qRT-PCR. (A) The
585 expression pattern of miR-450a and its target gene (DUSP10) were validated in E90
586 vs. E60. (B) The expression pattern of miR-26a and its target genes (PAK2) were
587 validated in D180 vs. D30. (C) The expression pattern of let-7g and its target genes
588 (COL1A2, COL14A1 and PLXND1) were validated in D30 vs. E90. 5s RNA and
589 β-actin were used as internal controls of miRNA and mRNA, respectively. Three
590 biological replicates were used.

591 Figure 6: GO analysis of target genes of miRNAs. Biological process (BP), molecular
592 function (MF), and cellular component (CC) of target genes in E90 vs. E60, D30 vs.
593 E90 and D180 vs. D30 in combined expression analysis. The left and right y-axis
594 represent the percent and the number of genes, respectively. The x-axis indicate the
595 names of the cluster.

Tables

Table 1 categorization of reads of small RNAs in porcine at different developmental stages

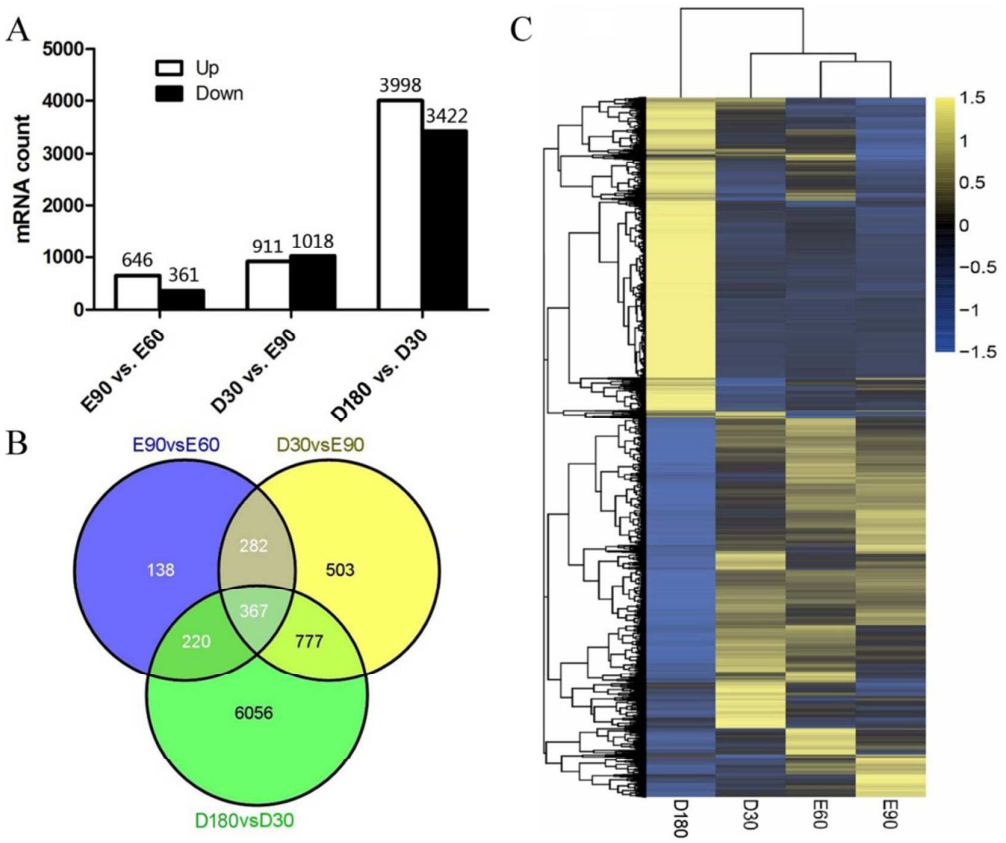
Category	Development stages				Average
	E60	E90	D30	D180	
Raw reads	6,695,453	6,474,804	6,457,999	6,645,391	6,568,421
Low qulity	1,908	1,836	1,954	1,579	1,819
5 adapter contamin	407	137	415	326	321
3 adapter	136,872	116,544	111,208	132,193	124,204
With ployA/T/G/C	13,117	2,371	11,647	28,139	13,818
Smaller than 18nt	195,734	66,148	184,411	149,613	148,976
Clean reads	6,347,414	6,287,768	6,148,364	6,333,541	6,279,272
Exon:+	87,329	64,502	161,842	93,317	101,747
Exon:-	8,722	3,597	5,951	8,153	6,605
Intron:+	48,849	19,511	46,949	82,944	49,563
Intron:-	19,191	7,683	13,404	63,138	25,854
Know miRNA	2,900,573	3,608,302	2,651,173	377,434	2,384,371
Novel miRNA	8,422	1,209	14,834	4,037	7,125
rRNA etc.*	105,483	50,846	121,533	46,653	81,128
Repeat	465,955	107,074	338,955	750,750	415,683
Un-annotated	2,208,589	24,048,790	2,324,009	4,189,542	8,192,733

*rRNA/snRNA/snoRNA/tRNAconsidered

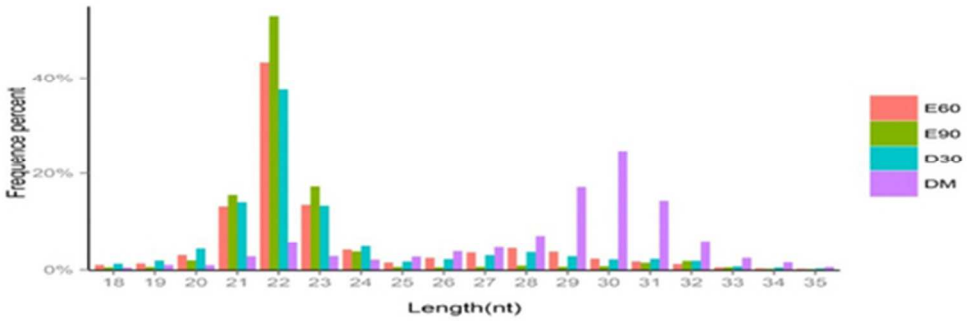
Table 2 basic characteristics of mRNA

Sample name	E60	E90	D30	D180
Raw reads	58,184,970	51,031,524	50,603,772	43,386,314
Clean reads	56,296,294	49,177,458	48,837,300	41,545,444
Q30 (%)	87.84	89.14	89.37	87.87
GC content(%)	47.67	49.74	50.65	49.31
Total mapped	43,851,191 (77.89%)	38,539,918 (78.37%)	37,820,316 (77.44%)	32,552,478 (78.35%)
Uniquely mapped	40,797,558 (72.47%)	35,727,629 (72.65%)	35,344,938 (72.37%)	30,316,956 (72.97%)
Reads map to '+'	20,379,021 (36.2%)	17,852,253 (36.3%)	17,650,519 (36.14%)	15,150,872 (36.47%)
Reads map to '-'	20,418,537 (36.27%)	17,875,376 (36.35%)	17,694,419 (36.23%)	15,166,084 (36.5%)
Non-splice reads	27,123,343 (48.18%)	22,354,684 (45.46%)	21,195,391 (43.4%)	17,383,593 (41.84%)
Splice reads	13,674,215 (24.29%)	13,372,945 (27.19%)	14,149,547 (28.97%)	12,933,363 (31.13%)

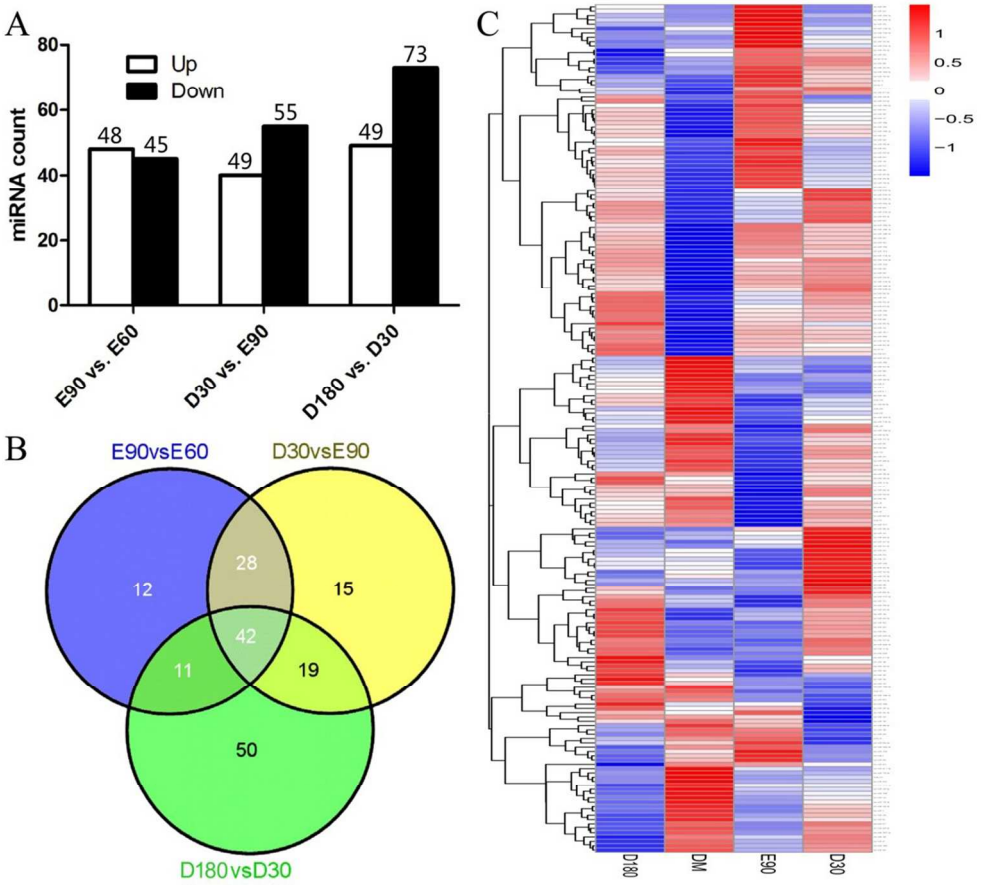
Note:1. “+”and“-” mean refers to sense strands and anti-sense strand, respectively.
2. “Non-splice reads” means reads for the entire sequence is mapped to one exon; “Splice reads” means reads mapped to the border of exon, also called junction reads.



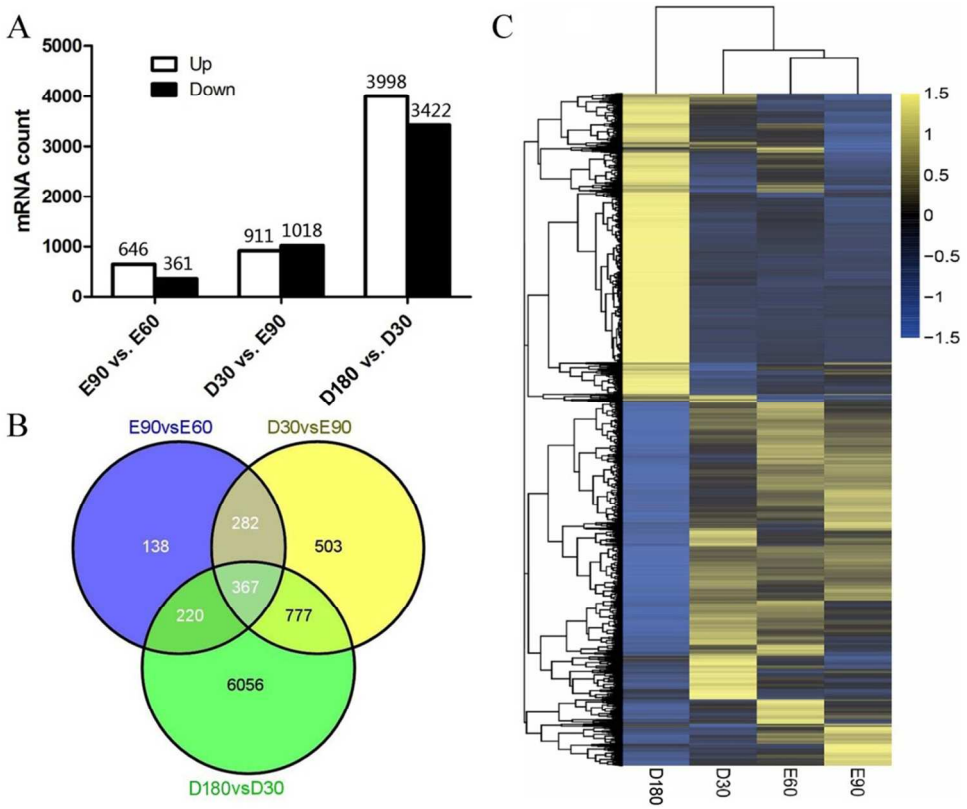
Gene expression profile in development of porcine testes investigates the intricate physiological process in pig testes development and spermatogenesis.



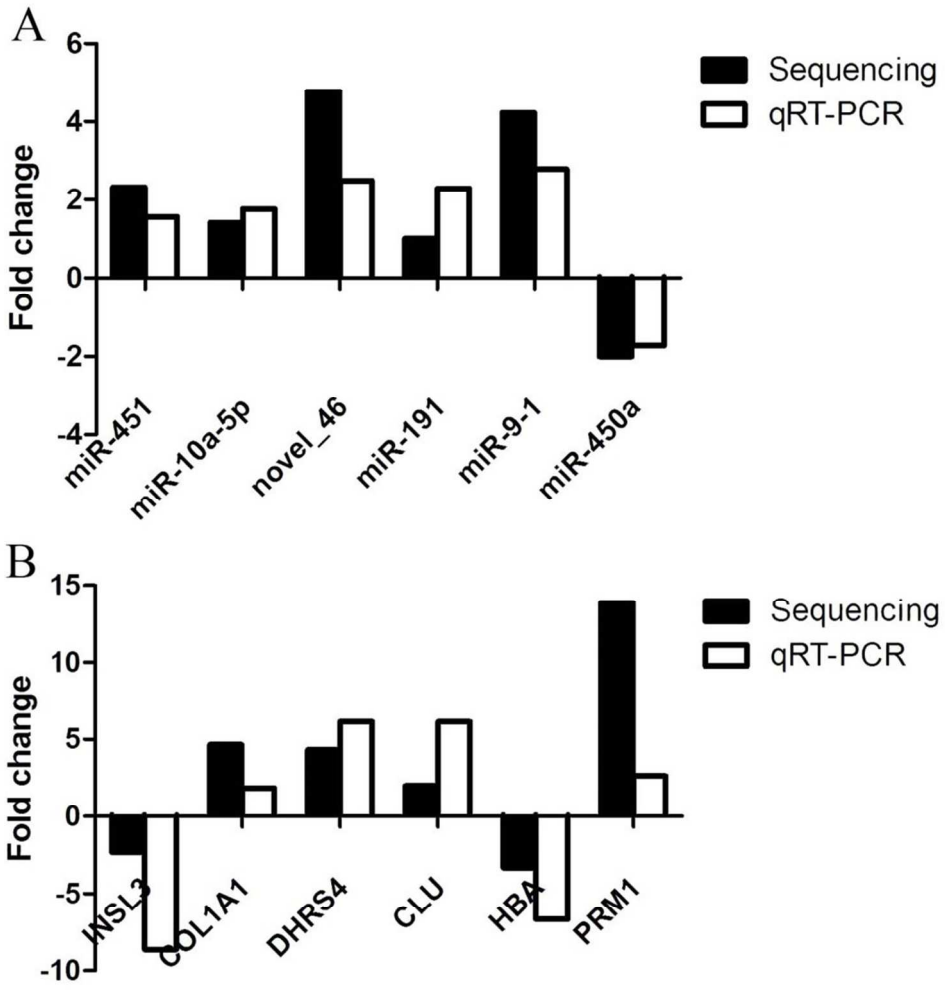
47x15mm (300 x 300 DPI)



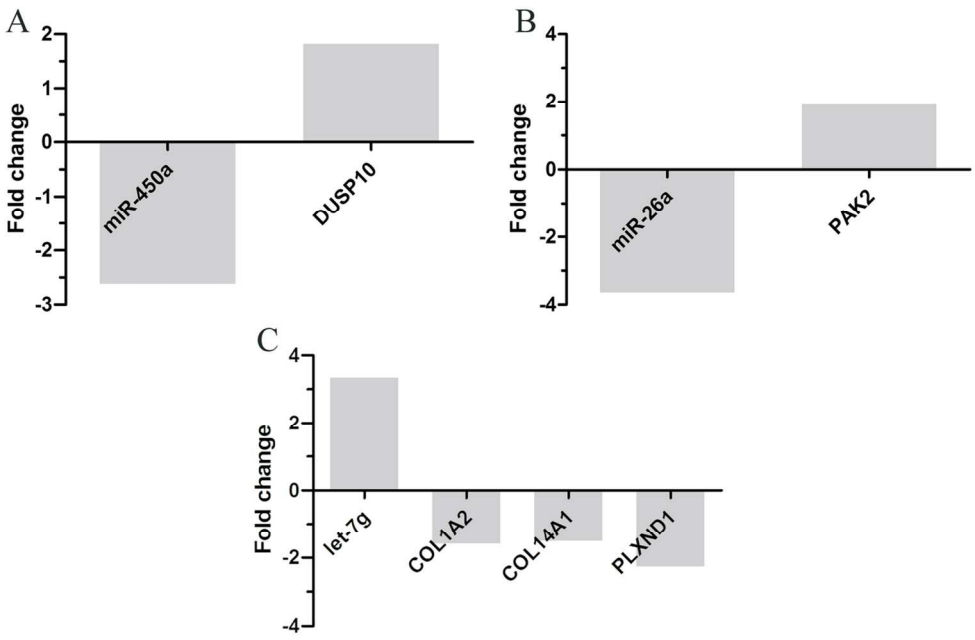
130x117mm (300 x 300 DPI)



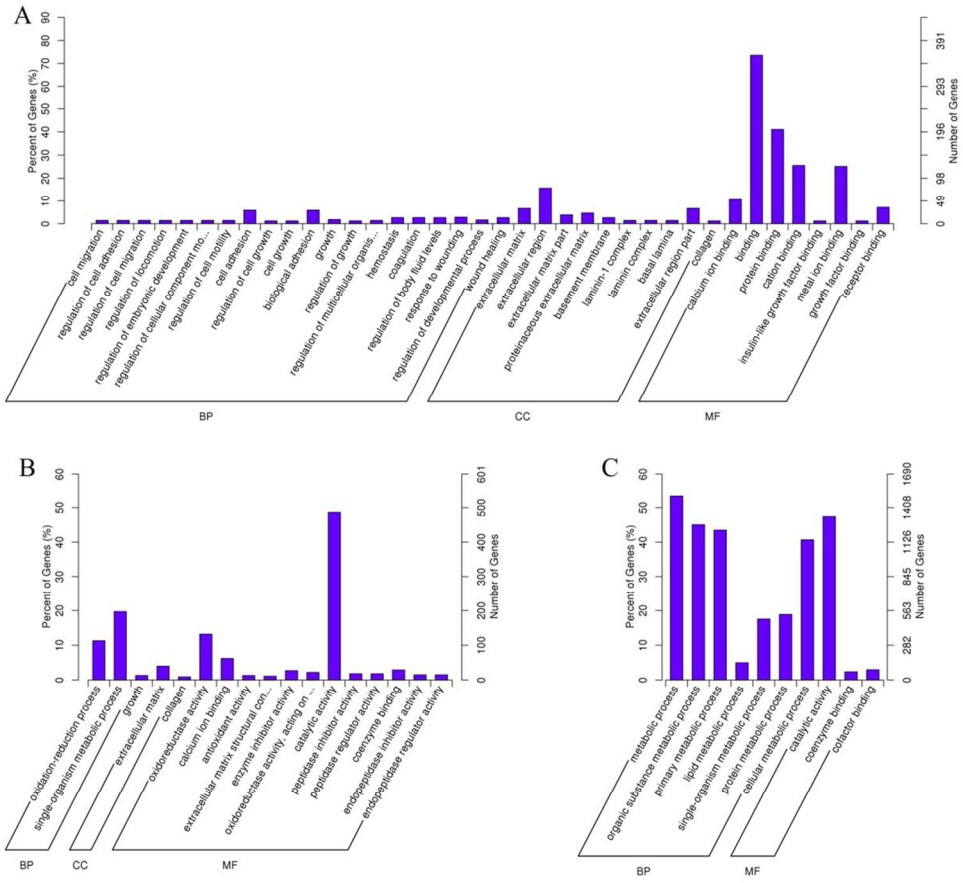
119x97mm (300 x 300 DPI)



132x134mm (300 x 300 DPI)



120x77mm (300 x 300 DPI)



130x116mm (300 x 300 DPI)