



Dynamic Transcriptome Analysis Reveals Potential Long Non-coding RNAs Governing Postnatal Pineal Development in Pig

Yalan Yang^{1†}, Rong Zhou^{2†}, Wentong Li^{1,2}, Ying Liu², Yanmin Zhang², Hong Ao², Hua Li¹ and Kui Li^{1,2*}

¹ Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, School of Life Sciences and Engineering, Foshan University, Foshan, China, ² Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:

Yun Xiao,
Harbin Medical University, China
Tao Zhou,
Auburn University, United States

*Correspondence:

Kui Li
likui@caas.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 28 January 2019

Accepted: 15 April 2019

Published: 03 May 2019

Citation:

Yang Y, Zhou R, Li W, Liu Y,
Zhang Y, Ao H, Li H and Li K (2019)
Dynamic Transcriptome Analysis
Reveals Potential Long Non-coding
RNAs Governing Postnatal Pineal
Development in Pig.
Front. Genet. 10:409.
doi: 10.3389/fgene.2019.00409

Postnatal development and maturation of pineal gland is a highly dynamic period of tissue remodeling and phenotype maintenance, which is genetically controlled by programmed gene expression regulations. However, limited molecular characterization, particularly regarding long noncoding RNAs (lncRNA), is available for postnatal pineal at a whole transcriptome level. The present study first characterized the comprehensive pineal transcriptome profiles using strand-specific RNA-seq to illustrate the dynamic mRNA/lncRNA expression at three developmental stages (infancy, puberty, and adulthood). The results showed that 21,448 mRNAs and 8,166 novel lncRNAs were expressed in pig postnatal pineal gland. Among these genes, 3,573 mRNAs and 851 lncRNAs, including the 5-hydroxytryptamine receptors, exhibited significant dynamic regulation along maturation process, while the expression of homeobox genes didn't show significant differences. Gene Ontology analysis revealed that the differentially expressed genes (DEGs) were significantly enriched in ion transport and synaptic transmission, highlighting the critical role of calcium signaling in postnatal pineal development. Additionally, co-expression analysis revealed the DEGs could be grouped into 12 clusters with distinct expression patterns. Many differential lncRNAs were functionally enriched in co-expressed clusters of genes related to ion transport, transcription regulation, DNA binding, and visual perception. Our study first provided an overview of postnatal pineal transcriptome dynamics in pig and demonstrated that dynamic lncRNA regulation of developmental transitions impact pineal physiology.

Keywords: pineal gland, pig, long noncoding RNA, postnatal development, transcriptome

Abbreviations: DAVID, database for annotation, visualization, and integrated discovery; DGE, differentially expressed gene; GO, gene ontology; lincRNA, long intergenic noncoding RNAs; lncRNA, long noncoding RNAs; MCL, Markov clustering; PCA, principal component analysis; RPKM, reads per kilobase per million reads; RT-qPCR, real-time quantitative PCR; TF, transcription factor; Y, Yorkshire.

INTRODUCTION

The mammalian pineal gland is a neuroendocrine transducer whose main and most conserved function is converting photoperiodic information into the nocturnal hormonal signal of melatonin synthesis and secretion (Maronde and Stehle, 2007). Melatonin regulates a variety of circadian and circannual physiological processes, such as the sleep-wake cycle, feeding, and cognition rhythms (Leon et al., 2004; Acuna-Castroviejo et al., 2007). Recent studies have revealed that melatonin also regulates many general physiological functions, including lipid and glucose metabolism, immune function, and carcinogenesis (Carrillo-Vico et al., 2005; Jha et al., 2015; Trivedi et al., 2016). Exploring pineal development will contribute to an improved understanding of its functions and mechanisms of regulation. The pineal gland develops as a tubular evagination from the dorsal diencephalon between the habenular and posterior commissures in the embryonic brain. The pineal gland displays a phase of rapid cell proliferation during the prenatal periods. However, cell proliferation activity terminates rapidly (Sapède and Cau, 2013), and pinealoblasts differentiate into pinealocytes during the two first postnatal weeks in rats (Calvo and Boya, 1983). After postnatal maturation, the parenchyma of the pineal gland is composed primarily of pinealocytes and interstitial cells (Moller and Baeres, 2002).

Pineal development is a complicated and dynamic process that is precisely genetically controlled by the programmed expression of gene cascades and TFs. Several TFs responsible for the establishment and maintenance of the pineal phenotype have been identified, such as the homeobox TFs *PAX6*, *LHX9*, and *OTX2* (Rath et al., 2013). However, gene abundance represents only part of the complexity of the transcriptome, as it has emerged that lncRNAs, which are a subgroup of transcripts that are longer than 200 nucleotides (nt) yet have limited protein-coding potential, have recently emerged as pivotal regulators in governing various developmental processes (Batista and Chang, 2013). For example, lncRNAs could regulate skeletal muscle differentiation during myogenesis, such as *MyoD* and *H19* (Dey et al., 2014; Gong et al., 2015a). lncRNAs show precisely spatiotemporal expression patterns and regulate specific neuronal functions in brain (Briggs et al., 2015). Until now, the expression dynamics of mRNAs and lncRNAs involved in pineal gland have not been extensively explored. Describing the transcriptome profiles of the pineal gland through development may improve our understanding of the molecular pathways and regulatory mechanisms that are responsible for postnatal pineal development in mammals.

The pig (*Sus scrofa*) not only is an important agricultural animal but also serves as an attractive model organism for biomedical research, due to the similarity of its organ size, anatomy and physiology, and developmental processes with those of humans (Groenen et al., 2012; Prather, 2013; Niu et al., 2017; Yan et al., 2018). Hence, we can understand the developmental patterns of the pineal gland in mammals by using the pig as a model. In this study, we characterized high-resolution pineal transcriptome profiles in Y pigs using strand-specific total RNA sequencing, which allowed us to comprehensively illustrate

the dynamic characteristics and functions of mRNAs/lncRNAs across three postnatal developmental stages: infancy (30 days, Y30), puberty (180 days, Y180), and adulthood (300 days, Y300). These results establish a general overview of the pineal transcriptome dynamics and pave the road for further investigations of the underlying functions and regulatory mechanisms of lncRNAs governing postnatal development of the mammalian pineal gland.

MATERIALS AND METHODS

Sample Collection

Nine Y pigs with the same genetic background at postnatal days 30, 180, and 300 (three replicates per stage) were obtained from the Tianjin Ningheyuan Swine Breeding Farm (Tianjin, China) and slaughtered during daytime, between 10:00 and 14:00 Beijing time. The pineal sample of each pig was collected and immediately frozen in liquid nitrogen until RNA isolation. All animal procedures were performed according to the protocols of the Chinese Academy of Agricultural Sciences and the Institutional Animal Care and Use Committee.

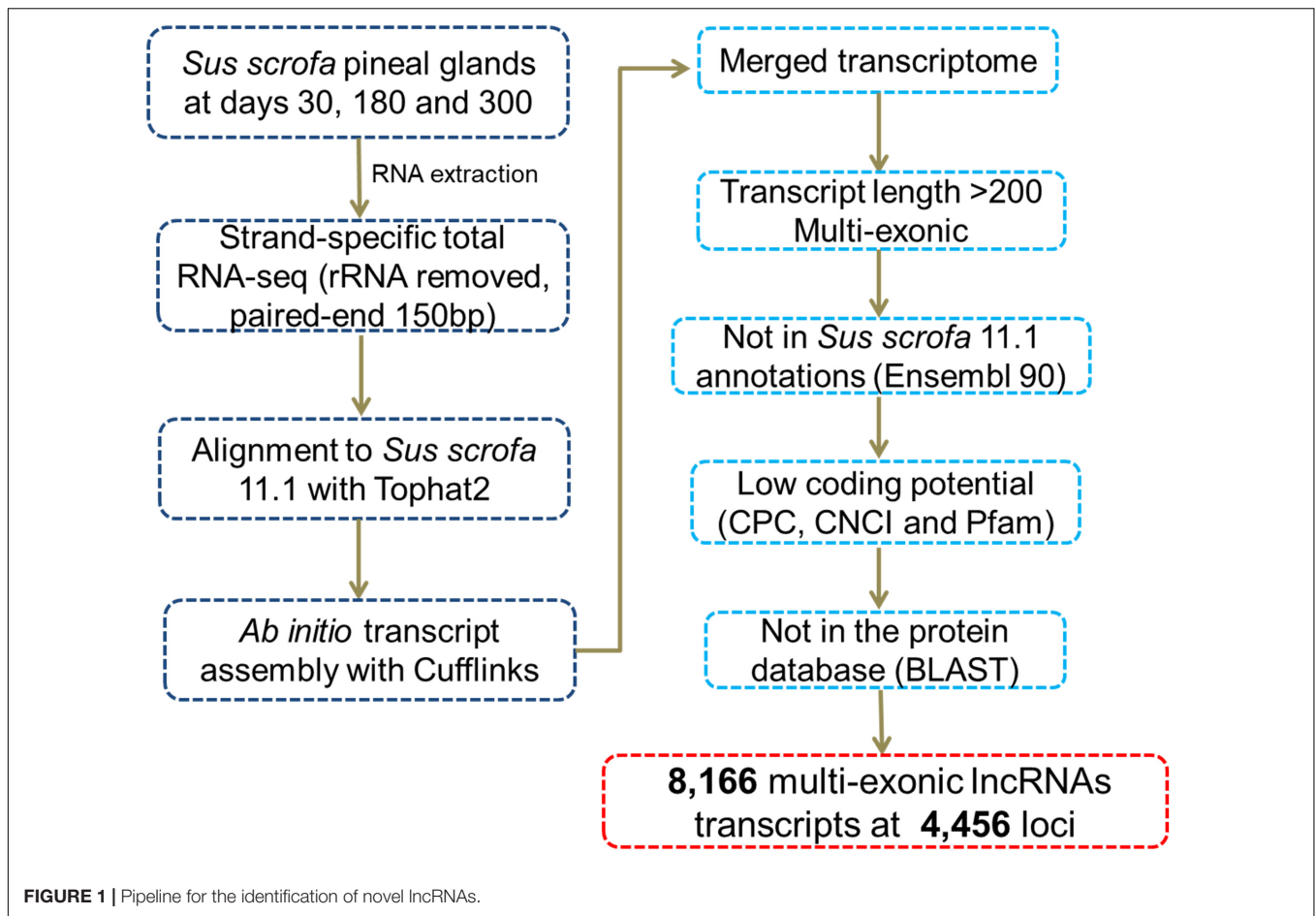
Transcriptome Library Preparation and Sequencing

Total RNA from pineal glands was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's directions. The purified RNA was treated with DNase I (Qiagen, Beijing, China). The quantity and purity of the RNA samples were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, United States). Ribosomal RNA was depleted using the Epicentre Ribo-zeroTM rRNA Removal Kit (Epicentre, Madison, WI, United States). Next, strand-specific RNA-seq libraries for paired-end sequencing were prepared using the NEBNext[®] UltraTM Directional RNA Library Prep Kit for Illumina[®] (NEB, United States) according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 4000 platform to generate 150 bp paired-end reads (Novogene Bioinformatics Technology Co. Ltd., Tianjin, China).

Transcriptome Assembly

The raw reads were firstly subjected to remove adaptor sequences and low-quality reads using custom scripts. The processed clean reads from each sample were then mapped to the reference genome of *Sus scrofa* (v11.1) using TopHat2 (v2.1.0) (Trapnell et al., 2009) with known gene annotation, parameters were set for strand-specific mapping (library-type "fr-secondstrand"). The reference genome sequence and gene annotation files were downloaded from the Ensembl database (release 90)¹. After mapping, duplicate reads were removed using the rmdup tool in the samtools package (Li et al., 2009) to limit the influence of PCR artifacts. The remaining unique mapped reads of each sample were assembled into transcripts independently using Cufflinks (v1.3.0) (Trapnell et al., 2012) with the assistance of known annotations. Finally, assembled transcripts from each sample

¹<http://asia.ensembl.org/index.html>



were merged into a consensus transcriptome using Cuffmerge (v1.0.0) (Trapnell et al., 2012).

Identification of lncRNAs

We identified novel lncRNAs in the pig pineal transcriptome using similar methods to those reported in our previous studies (Tang et al., 2017; Yang et al., 2017). A series of stringent filtering steps were utilized (Figure 1) as follows: (i) Single-exon transcripts and the transcripts less than 200 bp were removed to avoid unreliable transcripts; (ii) We filtered transcripts overlapping (>1 bp) with known gene models deposited in the Ensembl database; (iii) Coding Potential Calculator (CPC, v0.9-r2) (Kong et al., 2007) and Coding-Non-Coding Index (CNCI, v2) (Sun et al., 2013) programs were used to evaluate the coding potential of each transcript. Transcripts predicted to have coding potential (score > 0) by any of these two programs were filtered out; (iv) The transcripts whose corresponding translated protein sequences had a known protein-coding domain in the Pfam database (version 30.0) were removed by PfamScan (v1.3) (Finn et al., 2014); and (v) BLASTs (BLAST 2.2.26+) was used to remove transcripts with similarity to known proteins in the UniRef90 database (UniProt Consortium, 2015) with an *E*-value cutoff of 10^{-5} . Transcripts remaining after the stringent filtering described above were considered putative lncRNAs.

Expression Analysis

The raw read counts for each gene (mRNA/lncRNA) were calculated using HTSeq-count (Anders et al., 2015). For genes with multiple transcripts of different lengths, the longest transcript was selected to compute the gene expression level, measured as RPKM. Genes with RPKM ≥ 0.1 in at least one sample were defined as expressed genes. Highly expressed genes were defined as genes with a maximum RPKM ≥ 50 across the samples (Wang et al., 2014). The edgeR (exact test for negative binomial distribution) Bioconductor package (Robinson et al., 2009) in R software was used to identify DEGs between developmental stages. Gene expression normalization among samples to adjust for different sequencing depths across samples was performed using edgeR (Robinson et al., 2009). After estimating the dispersion of each gene, significantly DEGs were identified using cutoffs of false discovery rate (FDR) ≤ 0.05 and $|\log_2 FC| \geq 1$ according to the edgeR's recommendation (Robinson et al., 2009) and previous studies (Xue et al., 2013; Vanlandewijck et al., 2018).

Co-expression Network Construction

Normalized, non-log transformed gene expression data (RPKM values) of all the differentially expressed mRNAs/lncRNAs were imported into Biolayout Express (3D) (Theocharidis et al., 2009).

A pairwise gene-to-gene Pearson correlation matrix was calculated as a measure of similarity between genes. Based on a Pearson correlation coefficient cut-off threshold of $r \geq 0.90$, a weighted, undirected co-expression network of mRNA-lncRNA interactions was generated. In this network, each node represents one gene (mRNA or lncRNA) and the edge between two nodes represents the Pearson correlation coefficients above the selected threshold. The network was clustered into groups of mRNAs/lncRNAs sharing similar expression patterns using the MCL algorithm (Enright et al., 2002), which has been demonstrated to be one of the most effective graph-based clustering algorithms available. To control the size of the clusters, the inflation coefficient was set to 2.4 and each cluster must contain at least 30 genes. This network was checked manually and clusters with no particular expression pattern were removed. Clusters were named according to their relative size, the largest cluster being designated Cluster 1.

Functional Enrichment Analysis

Gene ontology enrichment analysis were performed by the DAVID website (v6.7²) (Huang et al., 2008) with a background set of human orthologues included in this study.

Real-Time Quantitative PCR (RT-qPCR)

The total RNA of each sample was reverse transcribed into cDNA using a RevertAid First Strand cDNA Synthesis Kit (Thermo, Waltham, MA, United States) according to the manufacturer's instructions. The RT-qPCR reaction solution was comprised of 10 μ l of 2 \times SYBR Premix Ex Taq (Takara, Dalian, China), 0.4 μ l of each primer, 1 μ l of cDNA, 0.4 μ l of Dye II, and sterile water to a volume 20 μ l. The RT-qPCR cycling parameters were as follows: 95°C for 5 min, followed by 40 cycles at 95°C for 5 s and 60°C for 1 min. Next, a dissociation program was carried out at 95°C for 15 s, 60°C for 1 min, and 95°C for 15 s. Each reaction was performed in triplicate. The 2- $\Delta\Delta$ Ct method was used to determine the gene expression level. The porcine *GAPDH* gene was selected as an internal control. All primer sequences are listed in **Supplementary Table S1**.

²<http://david.abcc.ncifcrf.gov/>

RESULTS

Overview of the *Sus scrofa* Pineal Transcriptome Data

To identify changes in mRNA/lncRNA expression during postnatal pineal gland development, we generated RNA-seq libraries from the pineal glands of female Y pigs at infancy (Y30), puberty (Y180), and adulthood (Y300). Three biological replicates were evaluated per stage. Utilizing strand-specific RNA-seq of total RNA, a total of 1.05 billion clean sequencing reads (150 bp paired-end) were obtained after discarding low-quality and adaptor reads, corresponding to an average of 116.4 million sequence reads per sample. Of the clean reads, 79.9–90.0% could be mapped to the pig reference genome (version 11.1) by the Tophat2 pipeline (Trapnell et al., 2009) (**Table 1**). After removing duplicate reads, the remaining uniquely mapped reads were used for further lncRNA identification and gene expression analyses.

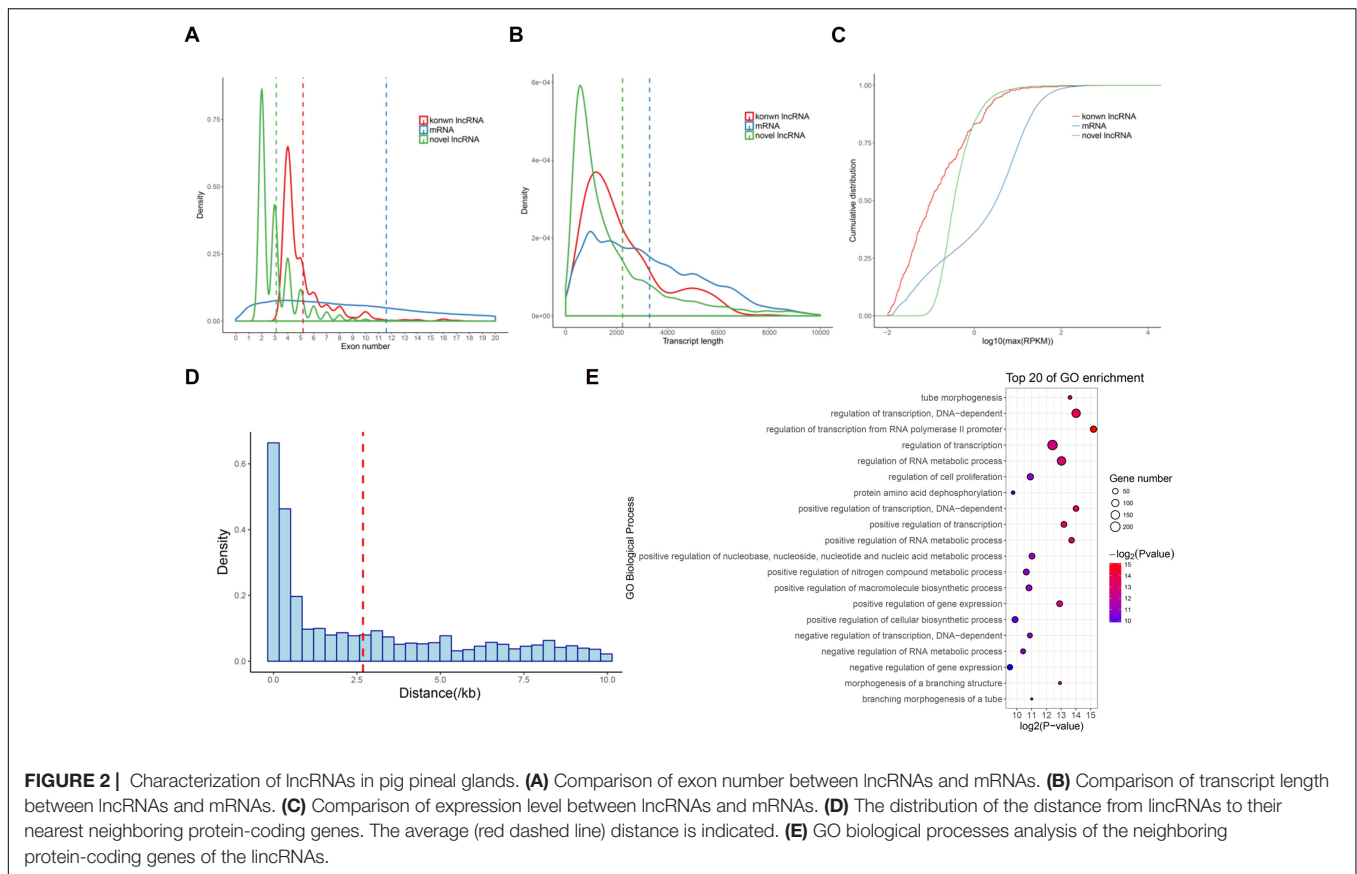
Identification and Characterization of lncRNAs in the Pineal Transcriptome

After reconstructing the transcriptome using Cufflinks and Cuffmerge (Trapnell et al., 2009), we identified putative lncRNAs in the pineal transcriptome using a pipeline (**Figure 1**) similar to those reported in our previous studies (Tang et al., 2017; Yang et al., 2017). Eventually, a total of 8,166 multi-exonic lncRNA transcripts corresponding to 4,456 genomic loci were obtained (**Supplementary Table S2**). According to their genomic location, most (6,505) were lincRNAs located in intergenic regions, while 1,129 were lncRNAs transcribed from the antisense strand of the reference coding transcript, and the remaining 532 lncRNAs overlapped with middle coding exon regions.

We next analyzed the features of these newly identified lncRNAs, namely novel lncRNAs. As expected, the novel lncRNAs contained fewer exons (3.1 exons on average) than mRNAs (11.6 exons on average; $P < 2.2e - 16$) (**Figure 2A**). The average transcript length of these lncRNAs (2235.2 nt) was significantly shorter than that of mRNAs (3296.1 nt; $P < 2.2e - 16$) (**Figure 2B**). Moreover, the expression levels of the lncRNAs (average RPKM = 2.7) were also significantly lower than those of the mRNAs (average RPKM = 10.9; $P < 2.2e - 16$).

TABLE 1 | Summary of sequencing metrics and read mapping for the RNA-seq of pig pineal glands.

Sample	Stage	Length	Reads	Mapped reads		Mapped ratio	
				Read1	Read2	Read1	Read2
Y30_1	30 days	150 bp	54,049,538 \times 2	47,598,866	43,352,218	88.1%	80.2%
Y30_2	30 days	150 bp	50,796,645 \times 2	44,735,289	41,004,165	88.1%	80.7%
Y30_3	30 days	150 bp	58,589,821 \times 2	51,633,312	46,824,908	88.1%	79.9%
Y180_1	180 days	150 bp	79,812,929 \times 2	70,250,218	67,703,427	88.0%	84.8%
Y180_2	180 days	150 bp	55,857,541 \times 2	49,638,178	45,750,849	88.9%	81.9%
Y180_3	180 days	150 bp	58,459,709 \times 2	51,560,101	46,995,972	88.2%	80.4%
Y300_1	300 days	150 bp	58,236,109 \times 2	51,208,023	46,802,188	87.9%	80.4%
Y300_2	300 days	150 bp	53,832,028 \times 2	48,429,903	44,651,018	90.0%	82.9%
Y300_3	300 days	150 bp	54,051,461 \times 2	48,432,224	43,497,472	89.6%	80.5%



(Figure 2C). These results were consistent with the previous lncRNAs reports in pigs and other mammals (Iyer et al., 2015; Tang et al., 2017; Yang et al., 2017). Additionally, we found that 2,481 lincRNA were transcribed near (<10 kb) their protein-coding neighbors. The average distance from lincRNAs to their neighboring genes was 2.68 kb (Figure 2D). GO analysis revealed that these neighboring genes were significantly enriched in regulation of transcription and tube morphogenesis functions (Figure 2E), indicating that these lincRNAs are preferentially located in the vicinity of genes with specific functions that are closely associated with postnatal pineal development.

Dynamic Expression of mRNAs and lncRNAs in Pineal Gland

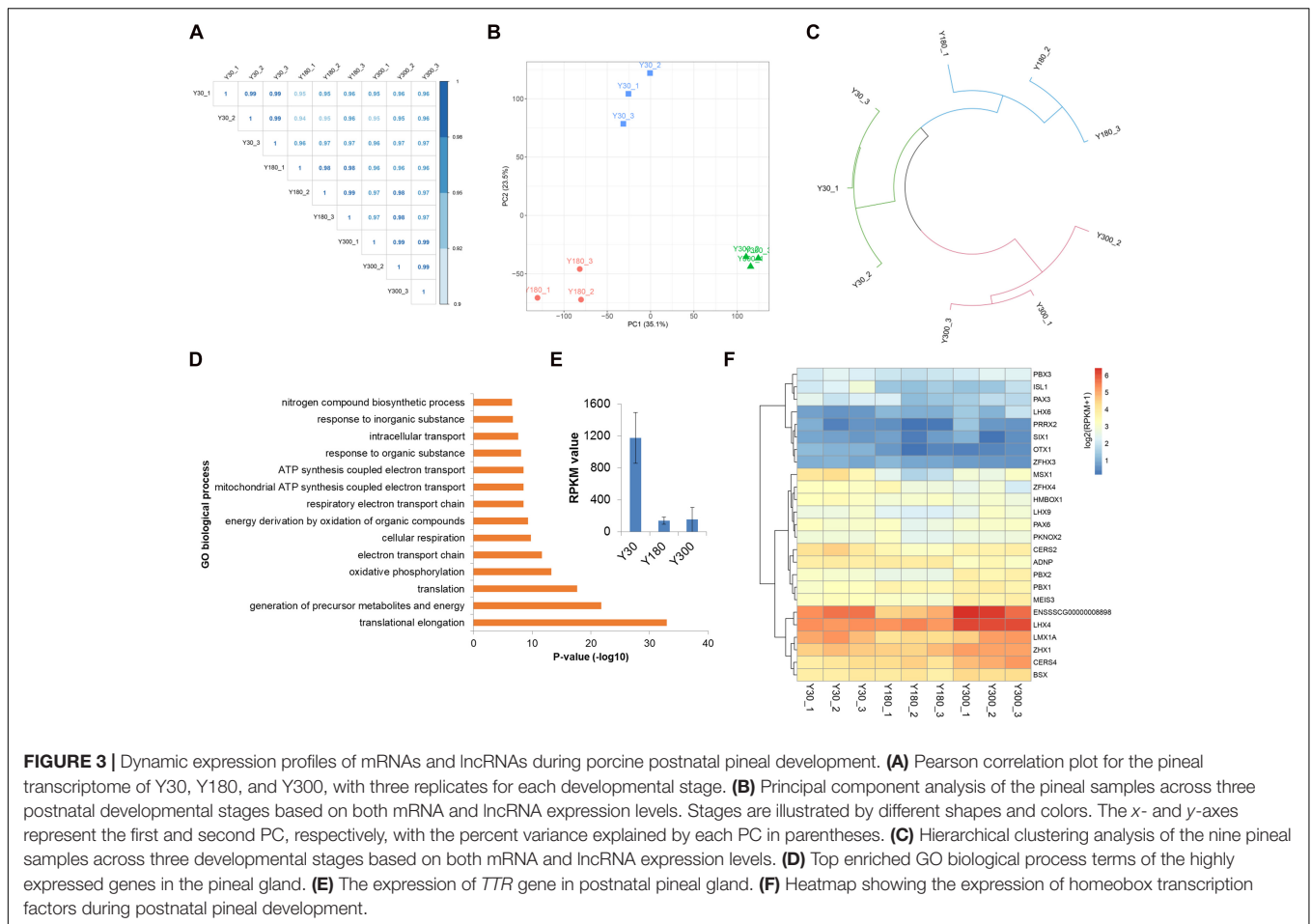
We next evaluated the expression of novel lncRNAs, known lincRNAs, and mRNAs across postnatal pineal development and found a high Pearson correlation within and across stages ($R > 0.95$) (Figure 3A), indicating a high level of measurement consistency among biological replicates. A PCA was performed in order to understand the expression patterns of all mRNAs and lncRNAs during postnatal pineal development. We found that the PCA could clearly separate the three developmental stages from each other; the first two principal components (PC1 and PC2) could explain 35.1 and 23.5% of the transcriptional variation, respectively (Figure 3B). Clustering analysis revealed that samples within stages were clustered together first, and then

Y30 and Y180 were grouped to form a larger cluster, and finally, clustered with Y300 (Figure 3C). These findings demonstrated a very high reproducibility within stages and distinct expression patterns across postnatal pineal development.

We detected an average of 15,388 mRNAs (a total of 21,448 mRNAs, with a range of 15,227–15,611 mRNAs per sample) and 2,740 lncRNAs (2,511–3,015 lncRNAs per sample) expressed ($RPKM \geq 0.1$) in pineal glands, which accounted for 68.9 and 57.0% of the total mRNAs and lncRNAs, respectively. The RPKM values of most of the mRNAs were greater than 1, while the majority of the lncRNAs were lowly expressed ($RPKM \leq 0.1$). Of these RNAs, 853 genes (842 mRNAs and 11 lncRNAs) were highly expressed in pineal glands ($RPKM \geq 50$ in at least one sample). As expected, these genes were significantly enriched in translation, oxidative phosphorylation, and ATP synthesis-coupled electron transport functions (Figure 3D), all of which are essential for protein synthesis and other basic requirements for postnatal pineal development. Additionally, *TTR*, a pineal-specific gene, was highly expressed in our samples, especially at the Y30 stage (Figure 3E). Most of the homeobox TFs were lowly expressed in postnatal pineal gland (Figure 3F).

Differentially Expressed mRNAs and lncRNAs

We found a total of 4,424 genes (including 3,573 mRNAs and 851 lncRNAs) with a significant difference in expression ($|\log_2$

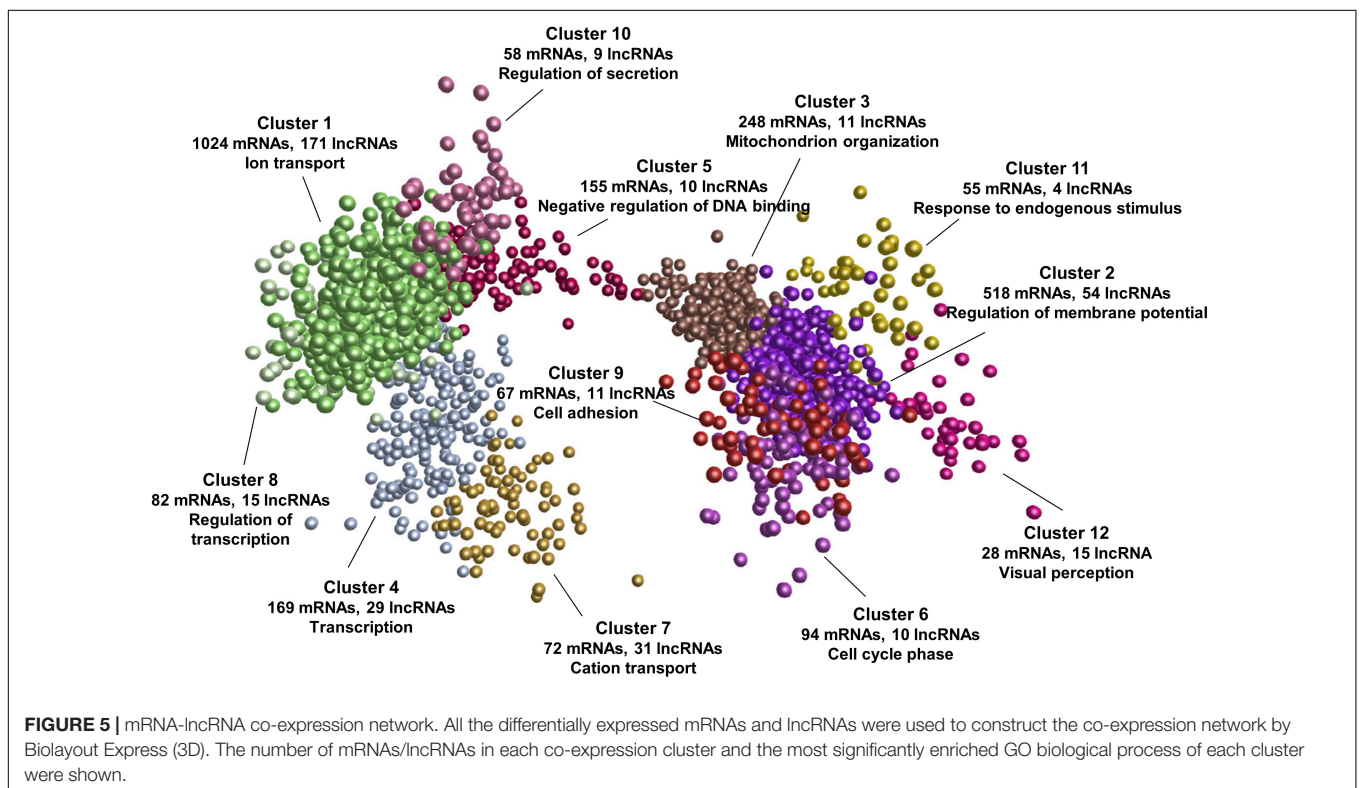
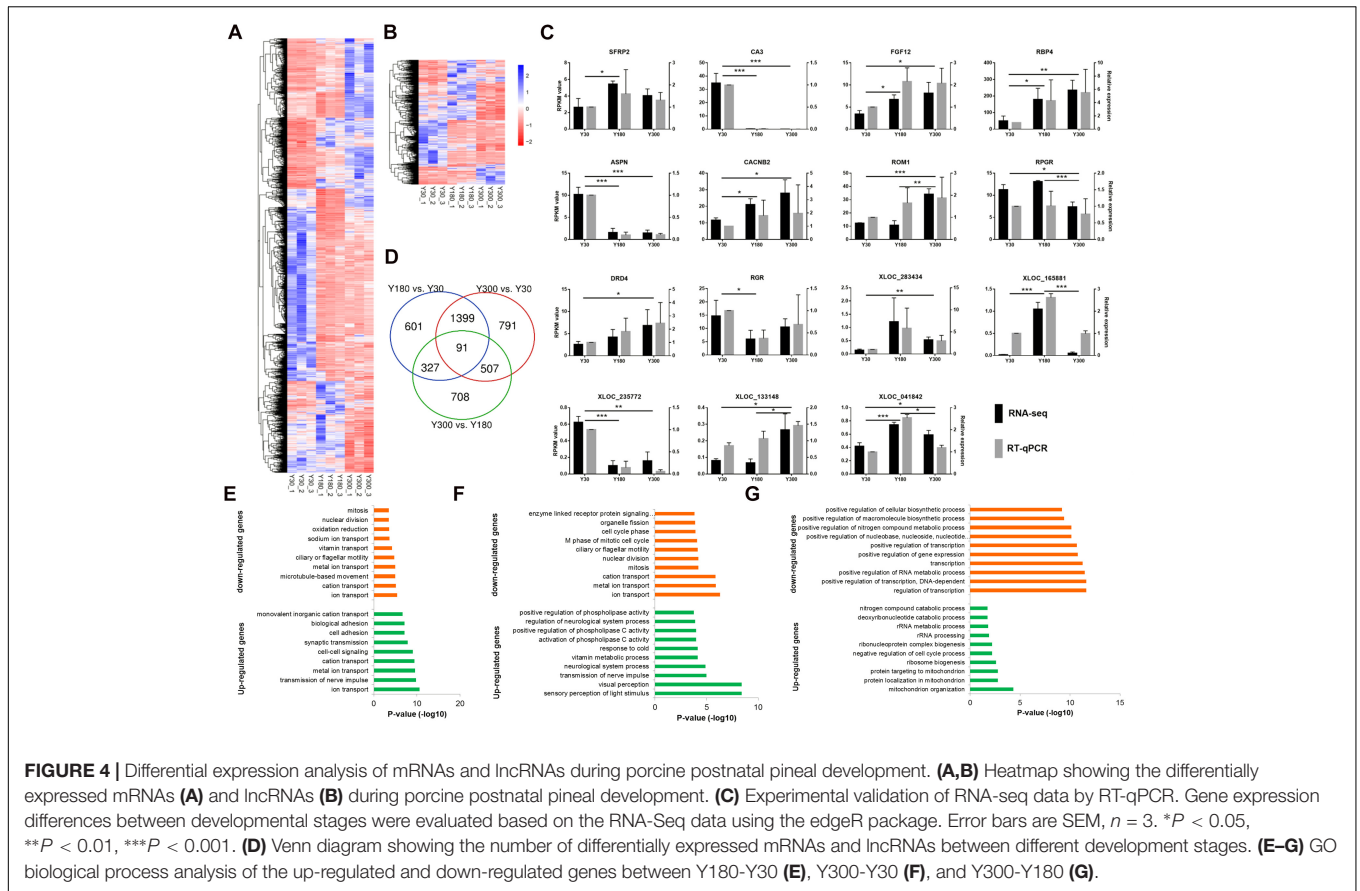


fold change (FC)|≥1 and FDR≤0.05) between developmental stages, including 2,417 Y180-Y30 (including 1,982 mRNAs and 436 lncRNAs), 2,788 Y300-Y30 (including 2,264 mRNAs and 524 lncRNAs), and 1,633 Y300-Y180 (including 1,187 mRNAs and 446 lncRNAs) DEGs (Figures 4A,B). Several 5-hydroxytryptamine (serotonin) receptors were included in this list, including *HTR2A*, *HTR2B*, *HTR2C*, and *HTR7*. We randomly verified 15 of the DEGs (10 mRNAs and 5 lncRNAs) by RT-qPCR and found a high concordance between the RT-qPCR and the RNA-seq data (Figure 4C), suggesting that the differential expression analysis based on the RNA-seq data was reliable. The highest number of DEGs was observed in the Y300-Y30 comparison, which was correlated with the difference in development time among the three stages. Most DEGs were observed in at least two of the three comparisons, and 91 of them (56 mRNAs and 35 lncRNAs) were found in all three comparisons (Figure 4D), including genes related to phosphate metabolic (*PPM1J*, *ND4*, *PRLR*, and *ND5*) and cell motility (*CCK*, *FOXJ1*, *DCDC2*, and *DNAH2*). We further examined the enriched functions of the DEGs through GO enrichment analysis. Compared with Y30, the up-regulated genes in Y180 were significantly enriched in ion transport, transmission of nerve impulse, cell-cell signaling, and synaptic transmission functions, while the down-regulated genes were

associated with ion transport, oxidation-reduction, and cell cycle categories (Figure 4E and Supplementary Table S3). The up-regulated genes in Y300 when compared with Y30 were significantly enriched in sensory perception of light stimulus, visual perception, transmission of nerve impulse, and neurological system process functions, while the down-regulated genes were enriched in ion transport and cell cycle functions (Figure 4F and Supplementary Table S3). Compared with Y180, the up-regulated genes in Y300 were significantly enriched in mitochondrion organization, ribosome biogenesis, and negative regulation of cell cycle process functions, while the down-regulated genes were associated with transcription and RNA metabolic and phosphate metabolic processes (Figure 4G and Supplementary Table S3).

Inference of Pineal lncRNA Function Using Co-expressed Network

To explore the potential functions and regulatory mechanisms of lncRNAs during postnatal pineal development, we constructed a co-expression interaction network of differentially expressed mRNAs and lncRNAs. The network consisted of 605,831 interaction pairs. These genes were grouped into 12 co-expression clusters by MCL algorithm (Enright et al., 2002).



The expression pattern of each cluster during postnatal pineal development was shown in **Supplementary Figure S1**. Some of these clusters contained mRNAs that are closely associated with postnatal pineal development (**Figure 5** and **Supplementary Table S4**). Cluster 1 was the biggest one, which contained 1024 mRNAs and 171 lncRNAs, genes in this cluster were highly expressed at Y30 stage, such as members of the solute carrier family genes (*SLC5A5*, *SLC13A5*, and *SLC39A12*). Cluster 7 was highly expressed at Y180 stage and contained 72 mRNAs and 31 lncRNAs. Interestingly, GO enrichment analyses suggested that ion transport was the most significantly enriched term of genes in these two clusters. The genes in both cluster 2 (518 mRNAs and 54 lncRNAs) and cluster 3 (248 mRNAs and 11 lncRNAs) were abundantly expressed at Y300 stage. The genes in cluster 3 were higher expressed at Y30 than at Y180 stage, while the genes in cluster 2 were stably expressed at these two stages. Cluster 2 and cluster 3 mainly functioned in regulation of membrane potential and mitochondrion organization, respectively. Additionally, cluster 4 (169 mRNAs and 29 lncRNAs) was enriched with transcription and oxidative phosphorylation genes, including the core subunits of mitochondrial membrane respiratory chain NADH dehydrogenase (*ND1*, *ND2*, *ND4*, and *ND5*). The genes in cluster 4 were higher expressed at Y180 stage than at Y30 and Y300 stages. Whereas the genes in cluster 5 exhibited an inversed expression patterns with the genes in cluster 4. Negative regulation of DNA binding was the most enriched biological process for cluster 5, which contained 155 mRNAs and 10 lncRNAs, such as *PTHLH*, *SMO*, *ID1*, and *XLOC_050558*. Genes in cluster 6 (94 mRNAs and 10 lncRNAs) were specifically expressed at Y300 stage, which were closely associated with cell cycle phase and mitosis, such as *CCNB1*, *CDC20*, and *DLGAP5*. Remarkably, the expressions of genes in cluster 9 (67 mRNAs and 11 lncRNAs) and cluster 10 (58 mRNAs and 9 lncRNAs) were continuously increased during postnatal pineal development, cell adhesion and regulation of secretion was the most significantly enriched biological processes in these two clusters, respectively. The continuously decreased genes were grouped into cluster 8 (82 mRNAs, 15 lncRNAs) and mostly enriched in regulation of transcription, such as *PLAG1*, *ATF7IP*, and *CRTC3*. Genes in cluster 11 and cluster 12 were associated with response to endogenous stimulus and visual perception, respectively. These results suggested putative regulatory functions for a subset of lncRNAs in postnatal pineal development.

DISCUSSION

In this study, we provided deep strand-specific RNA-seq of total RNA from three representative postnatal developing stages (infancy, puberty, and adulthood) of the porcine pineal gland, which we used to study the expression profiles of mRNAs/lncRNAs. We expect that this new resource will contribute to the understanding of the importance of transcription regulation in mammalian postnatal pineal gland development and maturation.

Pineal gland is an neuroendocrine organ for the regulation of the circadian clock system in all vertebrate species

(Macchi and Bruce, 2004). It's well known that homeobox genes are essential for normal pineal development and are key regulators in the maintenance of the postnatal pineal phenotype (Rath et al., 2013). We observed that most of the homeobox genes were lowly expressed (RPKM < 1) in our study, *HOPX* and *LHX4* were the most abundantly expressed ones in pig postnatal pineal gland, implying these two genes might play important roles in pineal development, though their function in pineal has not been reported. *LHX9* and *PAX6* are essential for early development of the mammalian pineal gland (Rath et al., 2013; Yamazaki et al., 2015), our study confirmed that these two genes were lowly expressed in the postnatal pineal gland after 30 days. *OTX2* displayed decreased expression in the postnatal pineal gland of rat (Rath et al., 2006), and was barely expressed in our samples. Additionally, the neurogenic differentiation factor 1 (*NEUROD1*) gene, a member of the bHLH TF family, is known to influence the fate of specific neuronal and endocrine cells (Munoz et al., 2007), and we confirmed that it was highly expressed in the postnatal pineal gland.

The expression of most DEGs changed constantly across postnatal pineal development, reflecting a dynamic regulation of gene expression. For example, the expression level of the transthyretin (*TTR*) gene, the major thyroid hormone transporter in the CNS, was much higher at Y30 than at Y180 or Y300. *TTR* has also been reported to be differentially expressed between midnight and midday in the pineal gland (Acuna-Castroviejo et al., 2007). We observed that DEGs were significantly associated with the ion transport, cell-cell signaling, synaptic transmission, and developmental maturation. Especially, 48 genes in calcium signaling pathway were differentially expressed, such as *CALB1* and *CACNB2*, which are involved in a variety of calcium-dependent processes, including cell motility, cell division, and hormone or neurotransmitter release (Dolphin, 2007). It's reported that melatonin could modulate neural development through the regulation of calcium signaling (Poloni et al., 2011). Additionally, 88 DEGs were involved in transmission of nerve impulse, such as *HTR2A*, *HTR2C*, and *HTR7*, which are 5-hydroxytryptamine (serotonin) receptors. 5-hydroxytryptamine is a precursor for melatonin production and is produced abundantly in the pineal gland of all vertebrate animals (Sapède and Cau, 2013). These results provide evidences that the critical roles of ion transport, especially calcium signaling, in postnatal pineal development, which might contribute to deeply understand the complexity of the pineal architectures and functions.

With the rapid adoption of RNA-seq technologies, thousands of lncRNA in the genome have been discovered in various species, their functions in various biological processes have been demonstrated (Geng et al., 2013; Gong et al., 2015b; Volders et al., 2015; Liang et al., 2018). However, compared with those of human and mouse, the lncRNA resources in pig are relatively limited (Quek et al., 2015; Liang et al., 2018). In this study, we identified a total of 8,166 novel lncRNAs, greatly expanding the genomic information of non-coding RNAs in pigs. These lncRNAs exhibited similar genomic characteristics with those of lncRNAs described in previous studies of pigs and other mammals (Iyer et al., 2015; Tang et al., 2017;

Yang et al., 2017). 851 lncRNAs, including 35 known and 816 novel lncRNAs, were differentially expressed across postnatal pineal development. Remarkably, 282 of them were transcribed near their protein-coding neighbors. For instance, XLOC_199747 located upstream of neurotrophin 3 (*NTF3*). There was a significantly positive correlation between the expressions of these two genes ($r = 0.81$). These results suggested that these differentially expressed lncRNAs might act on mRNAs involving in postnatal development by *cis* regulation. Co-expression analysis identified coordinated gene clusters that were shared in a developmental-specific expression fashion, which is an effective approach to uncover the function of lncRNAs (Pauli et al., 2012; Anamaria et al., 2014). We found most clusters containing genes with interesting functions. For example, GO enrichment analyses suggested that the cluster 1 was mostly associated with ion-transport, including many solute carrier family genes, which play important roles in the adrenergic regulation of cAMP and cGMP in pinealocytes (Sugden et al., 1986). Genes in cluster 2 were highly expressed at Y300 stages, which were closely related to regulation of membrane potential and transmission of nerve impulse, implying the critical roles of these mRNAs (such as *SCN1B* and *SYT4*) and lncRNAs (such as *XLOC_018250* and *XLOC_179558*) in mature pineal gland. Cell cycle genes (such as *CCNB1*, *CCNB2*, *CCNB3*, and *CCNF*), and 10 lncRNAs (such as *XLOC_280714* and *XLOC_156756*) were grouped into Cluster 6. The expression of these genes was decreased dramatically during postnatal pineal development, which was in consistence with the termination of pinealoblasts proliferation after birth (Sapède and Cau, 2013). Another intriguing example is cluster 12, which includes 15 lncRNAs (such as *XLOC_046348* and *XLOC_196944*). The mRNAs in this cluster were enriched in functional terms related to visual and sensory perception, which have been shown to play essential roles in circadian melatonin rhythm (Reiter, 1993). The dynamic changes observed in the co-expression networks offer insights regarding to the functions and regulation of lncRNAs during postnatal pineal development.

CONCLUSION

Overall, our data cataloged the pineal transcriptional profiles and basic gene expression features during postnatal development and maturation in pig. Novel lncRNAs were identified, which provide rich resources for understanding the molecular mechanisms and regulatory network of postnatal pineal development in mammals. The lncRNAs in the co-expression network may be considered as promising targets for postnatal pineal development, maturation,

REFERENCES

- Acuna-Castroviejo, D., Escames, G., Rodriguez, M. I., and Lopez, L. C. (2007). Melatonin role in the mitochondrial function. *Front. Biosci.* 12, 947–963.
- Anamaria, N., Magali, S., Maria, W., Angélica, L., Tasman, D., Ulrich, Z., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Batista, P. J., and Chang, H. Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152, 1298–1307. doi: 10.1016/j.cell.2013.02.012
- Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L., and Guy, B. (2015). Mechanisms of long non-coding rnas in mammalian nervous system and phenotype maintenance, but their function still needs to be further explored at the molecular, cellular, and individual levels.

ETHICS STATEMENT

All animal procedures were performed according to the protocols of the Chinese Academy of Agricultural Sciences and the Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

KL designed and managed the project. YY administered the computational analysis. YY and RZ analyzed the data and wrote the manuscript. RZ, WL, YL, YZ, and YY performed animal work and collected biological samples. WL, YL, and YZ performed molecular experiments. HL, HA, and KL revised the manuscript. All the authors approved the final manuscript.

FUNDING

This work was supported by National Key Basic Research Program of China (2015CB943100), National Science and Technology Major Project (2018ZX08009-26B, 2016ZX080011-006), the Key Project of National Natural Science Foundation of China (31330074), National Nonprofit Institute Research Grant (Y2016JC07 and 2018-YWF-YB-7), Foshan University Initiative Scientific Research Program, and Open Subject of Key Laboratory of Animal Molecular Design and Precise Breeding (2018A09).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00409/full#supplementary-material>

FIGURE S1 | The dynamic expression pattern of each co-expression cluster during postnatal pineal development.

TABLE S1 | Primer sequences of mRNAs and lncRNAs selected for validation by RT-qPCR.

TABLE S2 | Summary of the predicted lncRNAs in the *Sus scrofa* pineal gland.

TABLE S3 | GO biological process enrichment analysis of the differentially expressed genes between different developmental stages.

TABLE S4 | GO biological process enrichment analysis of the mRNAs in each co-expression cluster.

- development, plasticity, disease, and evolution. *Neuron* 88, 861–877. doi: 10.1016/j.neuron.2015.09.045
- Calvo, J., and Boya, J. (1983). Postnatal development of cell types in the rat pineal gland. *J. Anat.* 137(Pt 1), 185–195.
- Carrillo-Vico, A., Guerrero, J. M., Lardone, P. J., and Reiter, R. J. (2005). A review of the multiple actions of melatonin on the immune system. *Endocrine* 27, 189–200. doi: 10.1385/endo:27:2:189
- Dey, B. K., Pfeifer, K., and Dutta, A. (2014). The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev.* 28, 491–501. doi: 10.1101/gad.234419.113
- Dolphin, A. (2007). Voltage-dependent calcium channels. *Encycl. Pain* 24(Suppl. 1):2654. doi: 10.1007/978-3-540-29805-2_4827
- Enright, A. J., Dongen, S., and Van Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Geng, C., Wang, Z., Wang, D., Qiu, C., Liu, M., Xing, C., et al. (2013). LncRNADisease: a database for long non-coding RNA-associated diseases. *Nucleic Acids Res.* 41, D983–D986. doi: 10.1093/nar/gks1099
- Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S., et al. (2015a). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* 34, 181–191. doi: 10.1016/j.devcel.2015.05.009
- Gong, J., Liu, W., Zhang, J., Miao, X., and Guo, A. Y. (2015b). lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.* 43, D181–D186. doi: 10.1093/nar/gku1000
- Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi: 10.1038/ng.3192
- Jha, P. K., Challet, E., and Kalsbeek, A. (2015). Circadian rhythms in glucose and lipid metabolism in nocturnal and diurnal mammals. *Mol. Cell. Endocrinol.* 418(Pt 1), 74–88. doi: 10.1016/j.mce.2015.01.024
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35(Suppl. 2), W345–W349.
- Leon, J., Acuna-Castroviejo, D., Sainz, R. M., Mayo, J. C., Tan, D. X., and Reiter, R. J. (2004). Melatonin and mitochondrial function. *Life Sci.* 75, 765–790. doi: 10.1016/j.lfs.2004.03.003
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(1 Pt 2), 1653–1654.
- Liang, G., Yang, Y., Li, H., Yu, H., Li, X., Tang, Z., et al. (2018). LncRNAnet: a comprehensive *Sus scrofa* lncRNA database. *Anim. Genet.* 49, 632–635. doi: 10.1111/age.12720
- Macchi, M. M., and Bruce, J. N. (2004). Human pineal physiology and functional significance of melatonin. *Front. Neuroendocrinol.* 25:177–195. doi: 10.1016/j.yfrne.2004.08.001
- Maronde, E., and Stehle, J. H. (2007). The mammalian pineal gland: known facts, unknown facets. *Trends Endocrinol. Metab.* 18, 142–149. doi: 10.1016/j.tem.2007.03.001
- Moller, M., and Baeres, F. M. M. (2002). The anatomy and innervation of the mammalian pineal gland. *Cell Tissue Res.* 309, 139–150. doi: 10.1007/s00441-002-0580-5
- Munoz, E. M., Bailey, M. J., Rath, M. F., Shi, Q., Morin, F., Coon, S. L., et al. (2007). NeuroD1: developmental expression and regulated genes in the rodent pineal gland. *J. Neurochem.* 102, 887–899. doi: 10.1111/j.1471-4159.2007.04605.x
- Niu, D., Wei, H. J., Lin, L., George, H., Wang, T., Lee, I. H., et al. (2017). Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* 357, 1303–1307. doi: 10.1126/science.aan4187
- Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. doi: 10.1101/gr.133009.111
- Poloni, J. D., Feltes, B. C., and Bonatto, D. (2011). Melatonin as a central molecule connecting neural development and calcium signaling. *Funct. Integrat. Genomics* 11, 383–388. doi: 10.1007/s10142-011-0221-8
- Prather, R. S. (2013). Pig genomics for biomedicine. *Nat. Biotechnol.* 31, 122–124.
- Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., et al. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43, D168–D173. doi: 10.1093/nar/gku988
- Rath, M. F., Munoz, E., Ganguly, S., Morin, F., Shi, Q., Klein, D. C., et al. (2006). Expression of the Otx2 homeobox gene in the developing mammalian brain: embryonic and adult expression in the pineal gland. *J. Neurochem.* 97, 556–566. doi: 10.1111/j.1471-4159.2006.03773.x
- Rath, M. F., Rohde, K., Klein, D. C., and Moller, M. (2013). Homeobox genes in the rodent pineal gland: roles in development and phenotype maintenance. *Neurochem. Res.* 38, 1100–1112. doi: 10.1007/s11064-012-0906-y
- Reiter, R. J. (1993). The melatonin rhythm: both a clock and a calendar. *Experientia* 49, 654–664. doi: 10.1007/bf01923947
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sapède, D., and Cau, E. (2013). The pineal gland from development to function. *Curr. Top. Dev. Biol.* 106, 171–215. doi: 10.1016/B978-0-12-416021-7.00005-5
- Sugden, A. L., Sugden, D., and Klein, D. C. (1986). Essential role of calcium influx in the adrenergic regulation of cAMP and cGMP in rat pinealocytes. *J. Biol. Chem.* 261, 11608–11612.
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646
- Tang, Z. L., Wu, Y., Yang, Y. L., Yang, Y. C. T., Wang, Z. S., Yuan, J. P., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutionary conservation in *Sus scrofa*. *Sci. Rep.* 7:43166. doi: 10.1038/srep43166
- Theocharidis, A., Van, D. S., Enright, A. J., and Freeman, T. C. (2009). Network visualization and analysis of gene expression data using biolayout express(3D). *Nat. Protoc.* 4, 1535–1550. doi: 10.1038/nprot.2009.177
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Trivedi, P. P., Jena, G. B., Tikoo, K. B., and Kumar, V. (2016). Melatonin modulated autophagy and Nrf2 signaling pathways in mice with colitis-associated colon carcinogenesis. *Mol. Carcinog.* 55, 255–267. doi: 10.1002/mc.22274
- UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, 204–212. doi: 10.1093/nar/gku989
- Vanlandewijck, M., He, L., Mae, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. *Nature* 554, 475–480. doi: 10.1038/nature25739
- Volders, P. J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., et al. (2015). An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.* 43, 174–180.
- Wang, J., Garrey, J., and Davis, R. (2014). Transcription in pronuclei and one-to-four-cell embryos drives early development in a nematode. *Curr. Biol.* 24, 124–133. doi: 10.1016/j.cub.2013.11.045
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597. doi: 10.1038/nature12364

- Yamazaki, F., Möller, M., Fu, C., Clokie, S. J., Zykovich, A., Coon, S. L., et al. (2015). The Lhx9 homeobox gene controls pineal gland development and prevents postnatal hydrocephalus. *Brain Struct. Funct.* 220, 1497–1509. doi: 10.1007/s00429-014-0740-x
- Yan, S., Tu, Z., Liu, Z., Fan, N., Yang, H., Yang, S., et al. (2018). A Huntingtin knockin pig model recapitulates features of selective neurodegeneration in huntington's disease. *Cell* 173, 989–1002. doi: 10.1016/j.cell.2018.03.005
- Yang, Y., Zhou, R., Zhu, S., Li, X., Li, H., Yu, H., et al. (2017). Systematic identification and molecular characteristics of long noncoding rnas in pig tissues. *Biomed. Res. Int.* 2017:6152582. doi: 10.1155/2017/6152582

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yang, Zhou, Li, Liu, Zhang, Ao, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.