iScience, Volume *23*

Supplemental Information

A Unique Energy-Saving Strategy during Hibernation Revealed by Multi-Omics Analysis in the Chinese Alligator

Jian-Qing Lin, Yun-Yi Huang, Meng-Yao Bian, Qiu-Hong Wan, and Sheng-Guo Fang

SUPPLEMENTARY FIGURES

A. Female

KEGG pathway	Gene number	Down-regulated DEGs Number	Down-regulated DEGs q value	Up-regulated DEGs Number	Up-regulated DEGs q value	Key genes
Liver						
Primary bile acid biosynthesis	14	7	0.011	0	/	*CYP39A1, HSD3B7, AKR1D1, HSD17B4, BAAT*
Bile secretion	79	14	0.139	2	1.000	*SCARB1, NCEH1, EPHX1, BAAT, AQP9, ABCB1, FXR*
Small intestine						
Fat digestion and absorption	26	16	4.024E-05	1	1.000	*CD36, SCARB1, NPC1L1, FABP1, FABP2, MGAT2, DGAT1, DGAT2, MTTP, APOB, APOA4, APOA1*
Vitamin digestion and absorption	16	8	0.018	0	/	*THTR, RFC1, SCARB1, APOB, APOA4, APOA1, IF, CUBN*
Carbohydrate digestion and absorption	35	8	0.211	2	1.000	*LCT, MGAM, SGLT1,G6PC,GLUT5*
Mineral absorption	32	10	0.046	0	/	*DAR,SLC26A9, ZIP4, CTR1, SGLT1, FPN1, ATP1A, ATP1B, HEPH, NRAMP*
Protein digestion and absorption	87	15	0.210	1	1.000	*PEPT1, XPNPEP2, LAT2, TAT1, ATP1A, ATP1B*

B. Male

KEGG pathway	Gene number	Down-regulated DEGs Number	Down-regulated DEGs Enrichment	Up-regulated DEGs Number	Up-regulated DEGs Enrichment	Key genes
Liver						
Primary bile acid biosynthesis	14	5	0.183	1	1.000	*CYP39A1, CYP27A, AKR1D1, HSD17B4, BAAT*
Bile secretion	79	13	0.337	6	1.000	*SCARB1, NCEH1, EPHX1, BAAT, AQP1, ABCB1*
Small intestine						
Fat digestion and absorption	26	17	1.172E-04	1	1.000	*CD36, SCARB1, NPC1L1, FABP1, FABP2, MGAT2, DGAT1, DGAT2, MTTP, APOB, APOA4, APOA1*
Vitamin digestion and absorption	16	7	0.117	0	/	*RBP2, SCARB1, APOB, APOA4, APOA1, IF, CUBN*
Carbohydrate digestion and absorption	35	9	0.286	2	1.000	*LCT, MGAM, SGLT1,G6PC,GLUT5*
Mineral absorption	32	12	0.085	0	/	*DAR,SLC26A9, VDR, CTR1, SGLT1, FPN1, ATP1A, ATP1B, HEPH, NRAMP, FTH1*
Protein digestion and absorption	87	20	0.117	2	1.000	*MME, PEPT1, RBAT, LAT2, TAT1, ATP1A, ATP1B*

C. Heatmap

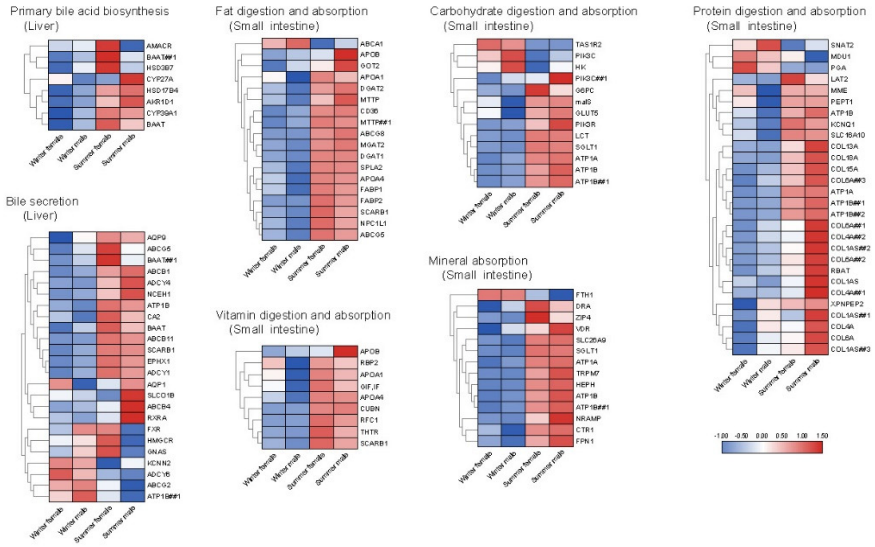
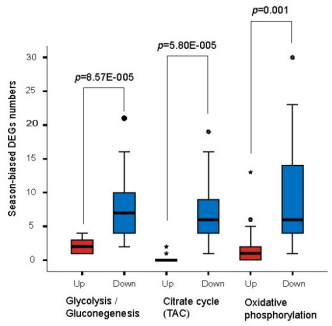
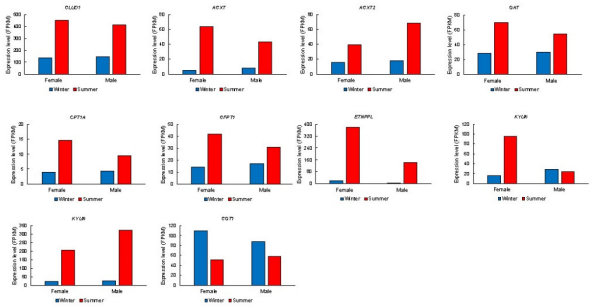


Figure S1. Gene expression patterns in key pathways involved in digestion and absorption in the liver and small intestine, Related to Figure 2. A, B. Season-biased differentially expressed genes (DEGs) in key pathways involved in nutrient digestion and absorption in the liver and small intestine in female (A) and male (B) Chinese alligators. **C.** Expression heatmap of seasonally DEGs in key pathways involved in digestion and absorption in the liver and small intestine.

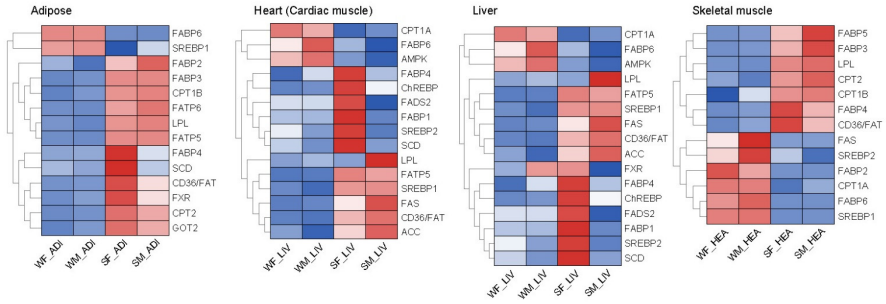
A.



D.



B.



C.

	Female					Male				All			
	Total genes	All DEGs	Down DEGs	Up DEGs	q	All DEGs	Down DEGs	Up DEGs	q	All DEGs	Down DEGs	UP DEGs	q
Alanine, aspartate and glutamate metabolism	44	11	8	3		8	7	1		6	5	1	
Glycine, serine and threonine metabolism	33	14	14	0		11	11	0		6	6	0	
Cysteine and methionine metabolism	39	13	9	4		12	9	3		5	5	0	
Valine, leucine and isoleucine degradation	70	10	9	1	Wilcoxon signedrank test, Z=-2.814, q=0.005	13	11	2	Wilcoxon signedrank test, Z=-2.943, q=0.003	7	7	0	Wilcoxon signedrank test, Z=-2.608, q=0.009
Lysine degradation	46	10	8	2		6	6	0		3	3	0	
Arginine biosynthesis	16	6	3	3		3	2	1		1	0	1	
Arginine and proline metabolism	41	8	5	3		7	5	2		3	2	1	
Histidine metabolism	20	4	3	1		5	4	1		2	1	1	
Tyrosine metabolism	37	11	8	3		9	7	2		8	8	0	
Tryptophan metabolism	36	10	8	2		9	7	2		6	6	0	
Phenylalanine metabolism	18	5	3	2		2	2	0		1	1	0	

Figure S2. Gene expression patterns in key pathways involved in nutrient metabolism, Related to Figure 2. A. Box plots of season-biased differentially expressed gene (DEG) numbers in pathways involved in carbohydrate metabolism. **B.** Expression heatmap of season-biased DEGs involved in lipid metabolism in the liver, adipose tissues, cardiac muscle, and skeletal muscle. **C.** Numbers of DEGs in amino acid metabolic pathways in the liver. **D.** Expression patterns of genes crucial for amino acid metabolism.

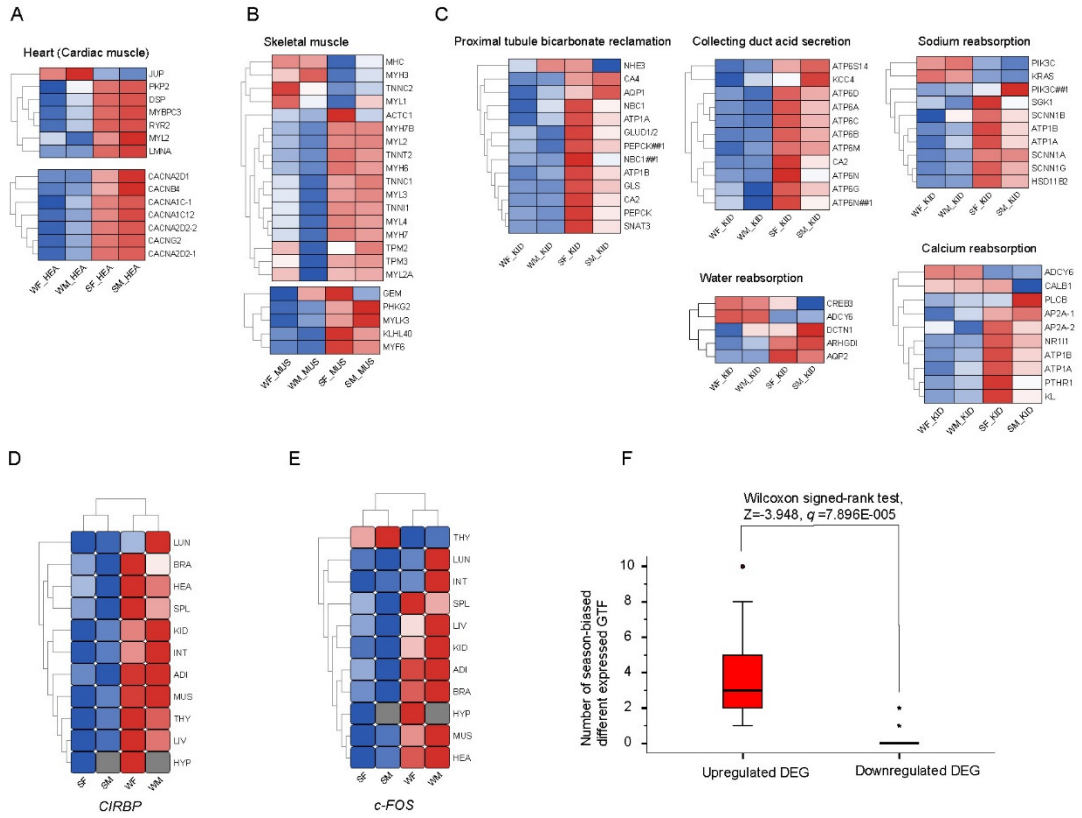


Figure S3. Gene expression patterns in key pathways involved in physiological function, Related to Figure 2. A, B, C. Heatmaps of differentially expressed genes (DEGs) involved in cardiac contraction and heart disease (A), skeletal muscle contraction and development (B), reabsorption and excretion in the kidneys (C). **D, E.** Expression patterns of *CIRBP* (D) and *c-FOS* (E). **F.** Comparison of the numbers of upregulated and downregulated general transcription factor genes (*GTF*s) during hibernation.

	mCG density			Methylation level of mCG			Methylation level of all CG		
	Summer	Winter	Wilcoxon signed-rank test	Summer	Winter	Wilcoxon signed-rank test	Summer	Winter	Wilcoxon signed-rank test
Adipose	97.19	95.35		77.84	74.86		79.32	77.42	
Brain	98.12	96.71		76.88	79.16		78.84	78.67	
Heart	97.58	95.21		74.74	71.55		77.15	75.16	
Hypothalamus	97.95	98.84		66.05	49.27		80.08	78.23	
Small intestine	97.50	96.28		74.70	72.59		75.89	73.58	
Kidney	98.43	98.29	Z=-1.778 q=0.083	80.15	77.73	Z=-1.423 q=0.155	81.97	79.45	Z=-1.867 q=0.067
Skeletal muscle	97.51	97.14		77.96	78.30		74.92	74.67	
Liver	98.05	97.00		74.98	78.56		77.58	78.67	
Lung	98.31	98.26		79.07	78.82		80.17	81.19	
Spleen	98.45	98.51		81.62	79.20		82.91	80.63	
Thyroid	97.72	98.74		80.68	79.43		80.87	81.32	

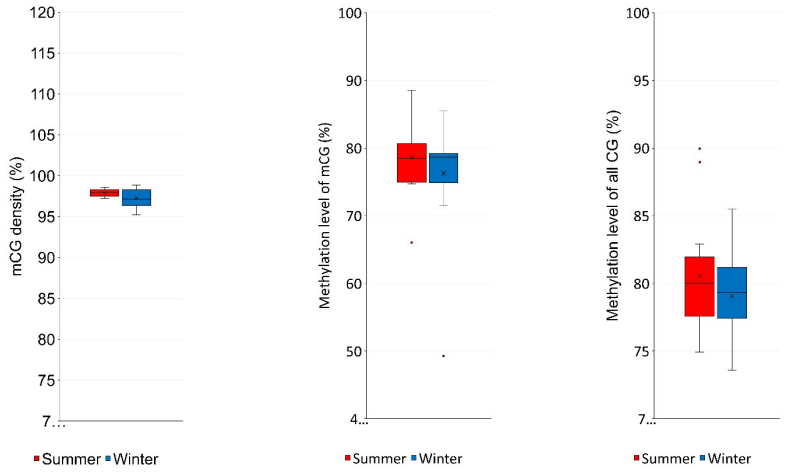


Figure S4. Comparison of DNA methylation in CG context in inactive (winter) and active (summer) periods, Related to Figure 3.

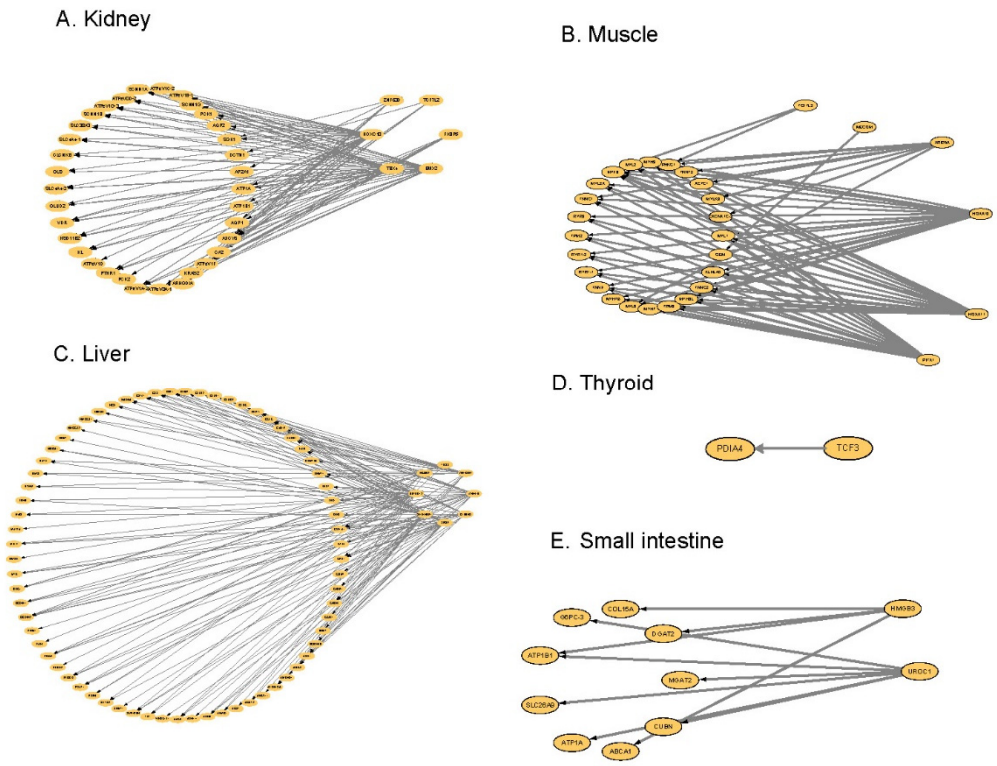


Figure S5. Associations between transcription factors (TFs) and key functional genes in each tissue, Related to Figure 5.

SUPPLEMENTARY TABLES

Table S1. Sample names for bisulfite sequencing, mRNA sequencing, and small RNA sequencing, Related to Figure 1.

	mRNA sequencing				Bisulfte sequencing		Small RNA sequencing	
	Winter		Summer		Winter	Summer	Winter	Summer
	Male	Female	Male	Female	Female	Female	Female	Female
Liver	WM_LIV	WF_LIV	SM_LIV	SF_LIV	WF_LIV	SF_LIV		
Adipose	WM_ADI	WF_ADI	SM_ADI	SF_ADI	WF_ADI	SF_ADI	WF_ADI	SF_ADI
Heart	WM_HEA	WF_HEA	SM_HEA	SF_HEA	WF_HEA	SF_HEA	WF_HEA	SF_HEA
Brain	WM_BRA	WF_BRA	SM_BRA	SF_BRA	WF_BRA	SF_BRA	WF_BRA	SF_BRA
Hypothalamus	WM_SPL	WF_SPL	SM_SPL	SF_SPL	WF_HYP	SF_HYP		
Spleen		WF_HYP		SF_HYP	WF_SPL	SF_SPL		
Kidney	WM_KID	WF_KID	SM_KID	SF_KID	WF_KID	SF_KID		
Lung	WM_LUN	WF_LUN	SM_LUN	SF_LUN	WF_LUN	SF_LUN		
Skeletal muscle	WM_MUS	WF_MUS	WM_MUS	SF_MUS	WF_MUS	SF_MUS	WF_MUS	SF_MUS
Small intestine	WM_INT	WF_INT	SM_INT	SF_INT	WF_INT	SF_INT	WF_INT	SF_INT
Thyroid		WF_THY1 / WF_THY2		SF_THY1/ SF_THY2	WF_THY1	SF_THY1		

Table S3. Serum thyroid hormone concentrations, related Figure 2.

	Active (summer) period	Inactive (winter) period	Unit
T3	>10.00	0.95	nmol/L
T4	88.09	6.42	nmol/L
FT3	18.83	1.5	pmol/L
FT4	56.48	<0.30	pmol/L
TG-Ab	271.6	20.04	KIU/L
TPO-Ab	>600.00	30.96	KIU/L
TSH	<0.005	<0.005	mIU/L

Table S4. Comparison of DNA methylation in CG context in inactive (winter) and active (summer) periods, related Figure 2.

Protein ID	Protein name	Protein description	Ratio# (iTRAQ)	Ratio# (TMT)
Alsi17612	LV001	Ig lambda chain V region 4A	1.38	1.42
Alsi14324	AF1L1	Actin filament-associated protein 1-like 1	1.24	1.3
Alsi15971	FBLN3	EGF-containing fibulin-like extracellular matrix protein 1	1.35	1.43
Alsi19475	KACB	Ig kappa chain C region, B allele	1.42	1.28
Alsi05485	SRCA	Sarcalumenin	2.93	2.21
Alsi01963	FCGBP	IgGFc-binding protein	1.76	1.89
Alsi16621	CAH6	Carbonic anhydrase 6	1.72	1.48
Alsi17112	A1AT	Alpha-1-antitrypsin	0.67	0.56
Alsi10652	RABL6	Rab-like protein 6	0.65	0.75
Alsi14288	VWF	von Willebrand factor	0.73	0.56
Alsi13698	VWA7	von Willebrand factor A domain-containing protein 7	0.79	0.65
Alsi01866	A2ML1	Alpha-2-macroglobulin-like protein 1	0.69	0.66
Alsi23845	ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	0.64	0.6
Alsi23375	XPO2	Exportin-2	0.53	0.78
Alsi12372	MCF2L	Guanine nucleotide exchange factor DBS	0.53	0.56
Alsi26905	KPYM	Pyruvate kinase PKM	0.28	0.34
Alsi02255	KPYR	Pyruvate kinase PKLR	0.56	0.64
Alsi20997	CLC11	C-type lectin domain family 11 member A	0.73	0.74
Alsi17022	RET4	Retinol-binding protein 4	0.51	0.63
Alsi04866	COMP	Cartilage oligomeric matrix protein	0.65	0.51
Alsi09977	LEG4	Beta-galactoside-binding lectin	0.45	0.44
Alsi21045	APOH	Beta-2-glycoprotein 1	0.47	0.39
Alsi01416	ACT5	Actin, cytoplasmic type 5	0.64	0.79
Alsi12373	FA10	Coagulation factor X	0.51	0.56

Winter/Summer

SUPPLEMENTARY TEXT

Data summary

We collected tissues from adult Chinese alligators in winter and summer and used multiple omics technologies (mRNA-Seq, BS-Seq, sRNA-Seq, iTRAQ, and TMT) (Figure 1, Table S1) to comprehensively explore the molecular mechanisms underlying reptile hibernation.

We produced 42 strand-specific mRNA-seq libraries, and generated 2,439 million paired-end reads, 88.36% of which were uniquely mapped to the Chinese alligator reference genome. Using these new transcriptome data as well as transcriptome data of Chinese alligator embryo (Lin et al., 2018), we annotated 27,500 protein-coding genes in the Chinese alligator genome.

Using tissues from female alligators, we produced 22 BS-seq libraries and generated 6,975 million clean read pairs (1,609 Gb of clean data), with an average depth of 16.11 per strand for each sample, and an average of 85.97% of genomic cytosines (Cs) being covered by at least five unique reads in each sample.

For sRNAs, we generated 129 million clean single-end reads from 10 sRNA-Seq libraries. 89.34% of the reads were 18–35 nt in length and were successfully mapped to the genome. We excluded other RNA species (rRNAs, tRNAs, snRNAs, snoRNAs), repetitive sequences, and transcript sequences, and with the aid of data from eight other tissues (testes and ovaries), we annotated 950 mature miRNAs. Among them, 132 miRNAs from 77 families were found in at least one other species and were thus identified as conserved miRNAs.

EXPERIMENTAL PROCEDURES

Sample sources and DNA and RNA extraction

Chinese alligator tissues (including liver, adipose, cardiac muscle, skeletal muscle, brain, hypothalamus, spleen, kidney, lung, small intestine, and thyroid tissues) were provided by the Changxing Yinjiabian Chinese Alligator Nature Reserve (CCANR) (Figure 1, Table S1). Winter samples were collected from two adult hibernating individuals (one male and one female) in January 2015, and summer samples were collected from two different adult individuals (one male and one female) during the breeding season, in May 2015. Serum samples for iTRAQ and TMT were collected from three female individuals in winter and summer, respectively. All samples were immediately stored in liquid nitrogen until use. Sample collection was performed with permission from the State Forestry Administration of China [Forest Conservation Permission Document (2014) 1545] and the Animal Ethics Committee of Zhejiang University (ZJU2015-154-13).

The gDNA used for BS-Seq and RNA used for mRNA-Seq was isolated from tissue samples using an AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Total RNA for sRNA-Seq was isolated using a TRIzol RNA isolation kit (Invitrogen, Waltham, MA, USA) according to the manufacturer's instructions.

Strand-specific cDNA library construction and sequencing

Three micrograms of RNA was used for strand-specific cDNA library construction using a NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA), according to the manufacturer's instructions. Library quality was evaluated on a Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). The index-coded samples were clustered using a TruSeq PE Cluster Kit v3-cBot-HS on a cBot Cluster Generation System (Illumina Inc., San Diego, CA, USA). The library was sequenced on the Illumina HiSeq 2500 platform, generating 125-bp paired-end reads. In-house Perl scripts were used to preprocess the raw reads in fastq format. Reads containing adapter or poly-N sequences, and low-quality reads were filtered out. The Q20, Q30, and GC content of the remaining reads were calculated, and all subsequent analyses were based on these clean reads.

BS-Seq library construction and sequencing

An unmethylated λ DNA fragment was added to the gDNA to evaluate the bisulfite conversion efficiency for quality control of the bisulfite treatment. Six micrograms of gDNA and 30 ng λ DNA were mixed and fragmented into 200–300 bp by sonication. After end repair and acetylation, barcodes with methylated cytosines were added to the fragmented DNA. DNA bisulfite conversion was carried out twice using an EZ DNA methylation-GoldTM Kit (Zymo Research, Irvine, CA, USA). The DNA fragments were then amplified by PCR with KAPA Hifi HotStart Uracil + ReadyMix (Kapa Biosystems, Wilmington, MA, USA). DNA concentrations in the BS libraries were determined with a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and insert sizes were evaluated using the Bioanalyzer 2100 system. The index-coded samples were clustered using the TruSeq PE Cluster Kit v3-cBot-HS on the cBot Cluster Generation System, according to the manufacturer's instructions. The BS library was sequenced on the Illumina HiSeq 2500 platform, generating 125-bp paired-end reads. In-house Perl scripts were used to preprocess the raw reads in fastq format. Low-quality reads, reads containing adapter and poly-N sequences, and reads shorter than 36 nt following adapter removal were filtered out. The Q20, Q30, and GC content of the remaining reads were calculated, and all subsequent analyses were based on these clean reads.

Small RNA library construction and sequencing

Three micrograms of total RNA was used to construct sRNA libraries using a NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (New England Biolabs), according to the manufacturer's instructions. The Agilent Bioanalyzer 2100 system was used for library quality assessment. The index-coded samples were clustered using the TruSeq SR Cluster Kit v3-cBot-HS on the cBot Cluster Generation System. After cluster generation, the sRNA libraries were sequenced on an Illumina HiSeq 2500 platform, generating 50-bp single-end reads.

RNA-Seq data analysis

The new transcriptome data produced in this study as well as transcriptome data of Chinese alligator embryo (Lin et al., 2018) were used to identify genes in Chinese alligator genome. An index of the Chinese alligator reference genome was built using Bowtie2 (Langmead and Salzberg, 2012),

and the cleaned reads were mapped to the reference genome using TopHat v. 2.0.12 (Kim et al., 2013). All aligned reads were assembled, and Cufflinks v. 2.1.1 was used to identify genes (Trapnell et al., 2010). TransDecoder and CPC were used to identify the coding region and protein-coding potential of each novel transcript (Haas et al., 2013; Kong et al., 2007). Genes with open reading frames larger than 150 bp and high protein-coding potential (score > 0) were subjected to further analysis. Transcription factor genes were identified by aligning all 27,500 genes to the animal transcription factor database TFDB2.0 (Zhang et al., 2015), using hmmsearch (Eddy, 2011).

Reads mapped to each gene were counted using HTSeq v. 0.6.1 (Anders et al., 2015), and differential gene expression between each pair of samples was analyzed using the DEGseq R package v. 1.12.0 (Mortazavi et al., 2008). Genes with an FDR < 0.005 and |log2(fold change)| > 1 were considered differentially expressed genes (DEGs). To correspond this with BS-Seq and sRNA-Seq data, we focused on the different season-biased transcriptomes of the female alligator for most of our analyses. However, we noted that gene expression patterns were largely similar in male and female non-gonadal tissues. The number of fragments per kilobase of exon per million mapped fragments (FPKM) was calculated to estimate gene expression levels. The WGCNA R package (Langfelder and Horvath, 2008) was used for weighted gene co-expression network analysis based the RNA-Seq data to analyze correlations in expression patterns among Chinese alligator genes. Gene expression level heatmaps were constructed using TBtools (Chen et al., 2018).

BS-Seq data analysis

BS-Seq reads were aligned to the Chinese alligator reference genome using Bismark (v. 0.12.5) with default parameters (Krueger and Andrews, 2011). The Chinese alligator reference genome was transformed into fully BS-converted versions termed “T genome” (C-to-T converted) and “A genome” (G-to-A converted) and then indexed using Bowtie2 (Langmead and Salzberg, 2012). All cytosines of the BS-converted reads were transformed to thymines, and the reads were aligned to the “T genome.” All guanines of the BS-converted reads were transformed to adenosines and the reads were aligned to the “A genome.” The sequence reads that produced a unique best alignment from the two alignments (original Watson and Crick strand) were re-aligned to the original reference genome to infer the methylation state of all cytosines in sequence reads. Multiple reads mapped to

the same regions of genome were defined as clonal duplicates and were filtered out to avoid inaccuracy that might be caused by bias during PCR amplification. The BS library conversion rate was estimated as the percentage of cytosines sequenced at cytosine positions in the λ reference genome.

To identify methylation sites, we modeled the sum of methylated cytosines (mCs) as a binomial (Bin) random variable with methylation rate (r), as mC ~ Bin (mC + umC*r).

The methylation level of each cytosine site was determined by the number of reads containing a methylated cytosine at the site of interest divided by the total number of reads covering the cytosine site. The methylation level of a specific region was calculated as the average methylation level of all cytosine sites in this region.

Differentially methylated regions (DMRs) between two samples were identified using swDMR (Wang et al., 2015). Since most of the mCs were in the CG context and the methylation levels in CHG and CHH contexts were low, we focused solely on CG sites for subsequent analyses. The sliding window was set to 1000 bp with a step length of 100 bp. To ensure statistical power, only windows with at least 10 CG sites and a coverage of 5 in each of the two compared samples were considered. Fisher's exact test was employed and only windows with an FDR-adjusted p (q) < 0.05 and greater than two-fold methylation level change were considered DMRs. Genes containing DMRs in their putative promoter or/and gene body regions were defined as differentially methylated genes (DMGs).

sRNA-Seq data analysis

sRNA reads were cleaned to eliminate unqualified reads and then mapped to the Chinese alligator genome reference sequence using Bowtie v. 2.2.3 (Langmead and Salzberg, 2012) without allowing any mismatch. We then annotated the reference genome sequence using the Rfam database (Kalvari et al., 2018) and RepeatMaker (Smit et al., 2013-2015), and reads corresponding to rRNAs, tRNAs, snRNAs, snoRNAs, and repeat sequences, as well as exons and introns were removed. miREvo v. 1.1 (Wen et al., 2012) and mirdeep2 (Friedlander et al., 2012) were used to predict miRNAs through exploration of the secondary structure, the Dicer cleavage site, and the minimum

free energy of the reads. Predicted miRNAs were subjected to Rfam (Kalvari et al., 2018) for miRNA family analysis and identification of conserved miRNAs (found in at least one other species). In-house scripts were applied to obtain miRNA counts. Finally, the expression level of each miRNA was normalized as the number of transcripts mapped to the miRNA per million transcripts (TPM). DEGseq R package was used to analyze differential miRNA expression between paired samples (Wang et al., 2010). miRNAs with $q < 0.01$ and $|\log2(\text{fold change})| > 1$ were assigned as DEmiRs. MiRanda, RNAhybrid, and PITA were used to predict target genes of miRNAs (Betel et al., 2010; Kertesz et al., 2007; Rehmsmeier et al., 2004). Target genes approved by at least one software package were considered targets. DESeq R package (Anders and Huber, 2010) was used to carry out principal component analysis (PCA) of sRNA-Seq data of the gonad samples and construct the plot.

GO and KEGG enrichment analyses

The GOseq 2.12 R package was used for GO enrichment analysis (Young et al., 2010). The KOBAS software package was employed for KEGG pathway enrichment analysis (Mao et al., 2005). GO terms and KEGG pathways with a $q < 0.05$ were regarded as significantly enriched.

TMT/iTRAQ protein quantification and data analysis

To reduce protein complexity and interference from highly abundant proteins, we used a ProteoMinerTM Kit (Bio-Rad Laboratories, Hercules, CA, USA) according to the manufacturer's instructions to deplete high- abundance proteins in the serum samples. Protein concentrations were determined by Bradford protein assay (Bio-Rad Laboratories). Total protein (100 μg) was digested with Trypsin Gold (Promega, Madison, WI, USA) using a protein/trypsin ratio of 30:1 at 37 °C for 16 h. The peptides were dried by vacuum centrifugation and reconstituted in 0.5 M triethylammonium bicarbonate buffer and processed according to the manufacturer's instruction for 8-plex iTRAQ (Applied Biosystems, Waltham, MA, USA) and 6-TMT labeling (Thermo Fisher Scientific). The labeled peptide mixtures were then pooled and dried by vacuum centrifugation. The peptides were separated by SCX chromatography using a LC-20AB HPLC Pump system (Shimadzu, Kyoto, Japan) and analyzed by LC-ESI-MS/MS analysis using a Q Exactive mass spectrometer (Thermo Fisher Scientific) coupled to the HPLC.

Peptides and proteins were identified by searching against the Chinese alligator database containing 27,500 sequences using the Mascot search engine v. 2.3.02 (Matrix Science, London, UK). Quantitative protein ratios were weighted and normalized by the median ratio in Mascot. We compared pairs of winter and summer samples. Only ratios with $p < 0.05$ and a fold change > 1.2 or < 0.83 in a least one pair were considered a significantly season-biased differentially expressed protein, and those with a different change trend in any pair were excluded.

SUPPLEMENTAL REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166-169.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11, R90.

Chen, C., Xia, R., Chen, H., and He, Y. (2018). TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. bioRxiv, 289660.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput Biol 7, e1002195.

Friedlander, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 40, 37-52.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8, 1494-1512.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res 46, D335-D342.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat Genet 39, 1278-1284.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35, W345-349.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571-1572.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.

Lin, J.Q., Zhou, Q., Yang, H.Q., Fang, L.M., Tang, K.Y., Sun, L., Wan, Q.H., Fang, S.G. (2018). Molecular mechanism of temperature-dependent sex determination and differentiation in Chinese alligator revealed by developmental transcriptome profiling. Science Bulletin, 63:209-212.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21, 3787-3793.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 5, 621-628.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507-1517.

Smit, A., Hubley, R., and Green, P. (2013-2015). RepeatMasker Open-4.0.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-515.

Wang, L.K., Feng, Z.X., Wang, X., Wang, X.W., and Zhang, X.G. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136-138.

Wang, Z., Li, X., Jiang, Y., Shao, Q., Liu, Q., Chen, B., and Huang, D. (2015). swDMR: a sliding window approach to identify differentially methylated regions based on whole genome bisulfite sequencing. PLoS One 10, e0132866.

Wen, M., Shen, Y., Shi, S., and Tang, T. (2012). miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. BMC Bioinformatics 13, 140.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11, R14.

Zhang, H.M., Liu, T., Liu, C.J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y., and Guo, A.Y. (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Res 43, D76-81.