Transcriptome and chromatin accessibility in porcine intestinal epithelial cells upon Zearalenone exposure

Haifei Wang¹, Jian Jin¹, Jiayun Wu¹, Huan Qu¹, Shenglong Wu¹,² & Wenbin Bao¹,²*

Zearalenone (ZEA) is one of the main mycotoxins widely spread in contaminated cereal crops, which poses a great threat to food safety as well as human and animal health. Biological control strategies are emerging as important solutions to eliminate mycotoxin contaminations. However, molecular mechanisms underlying ZEA cytotoxic effects are only partly understood. Noncoding RNAs and chromatin accessibilities are important regulators of gene expression and implicate in a variety of biological processes. Here, we established a study model of porcine intestinal epithelial cells upon ZEA exposure and presented a RNA-seq dataset for mRNA, microRNA, and lncRNA profiling in 18 experimental samples. In addition, chromatin accessibilities of four samples were also explored by ATAC-seq. This dataset will shed new light on gene expression profiling and transcriptional regulation of animal cells in the response to ZEA exposure, which further contributes to detecting biomarkers and drug targets for predicting and controlling ZEA contamination.

Background & Summary
Zearalenone (ZEA) is one of the main mycotoxins produced by a variety of Fusarium fungal species and widely spread in contaminated cereal crops including maize, wheat, barley and oats¹. After ingestion and absorption, ZEA is mainly metabolized by intestinal cells and hepatocytes. Because of the structural similarities of ZEA to endogenous estrogen, ZEA can result in serious endocrine disruption and reproductive disorders in animals²,³. In addition, ZEA was also found to cause toxic effects on liver and kidney functions⁴,⁵, and lymphocyte proliferation⁶. ZEA is chemically stable and cannot be removed by the manufacturing process, which poses great risks to food safety as well as human and animal health. Biological control strategies are emerging as promising solutions to eliminate mycotoxin contaminations. Therefore, it is becoming increasingly important to further understand the molecular mechanisms underlying ZEA toxic effects for developing strategies controlling ZEA contamination. Disruption of gene expression programs is an important event through which mycotoxins exert cytotoxic effects. Recent studies have preliminarily investigated the effects of ZEA exposure on genome wide gene expression in porcine epithelial cells⁷,⁸. However, the regulatory networks involved in gene expression alterations in animal cells upon ZEA exposure remain largely unknown.

Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are noncoding RNAs that act as important regulators involved in a variety of physiological, developmental and disease processes at the post-transcriptional level of their target genes⁹. LncRNAs are a class of transcripts with the length of greater than 200 nucleotides, and miRNAs are transcripts with the length of ~22 nucleotides. LncRNAs and miRNAs can either independently regulate target mRNA expression or functionally interact to control the expression of target mRNAs⁹,¹⁰. Therefore, identification of expression patterns of lncRNAs and miRNAs can greatly contribute to revealing the molecular events relevant to the phenotypic changes. Chromatin accessibility represents genomic regions binding with regulatory factors responsible for gene transcription, which can be measured by Tn5 transposase-accessible chromatin sequencing (ATAC-seq). Recently, ATAC-seq has become an effective and powerful tool to capture open chromatin to identify the regulatory elements of gene transcription¹¹.
In this study, we performed genome-wide analyses of the expressions of mRNA, miRNA, and lncRNA in porcine intestinal epithelial cells upon ZEA exposure (Fig. 1a). In total, 18 samples were sequenced on the Illumina Hiseq Platform, generating a total of 1,052,122,031 clean reads after quality control (Tables 1, 2). In addition, changes in chromatin accessibilities upon ZEA exposure were also explored by ATAC-seq (Fig. 1a; Table 1), which yielded a total of 230,639,896 clean reads (Table 3). Integrative bioinformatic analysis workflow of RNA-seq and ATAC-seq data is shown in Fig. 1b. These data will provide comprehensive insight into gene expression profiles and transcriptional regulation of animal cells in the response to ZEA exposure, which may aid the detection of biomarkers and drug targets for predicting and controlling ZEA contamination.

Methods

Sample preparation and collection. Porcine intestinal epithelial cells (IPEC-J2) were inoculated in 6-well plate at a density of $5 \times 10^5$ cells/mL and cultured overnight in a CO$_2$ incubator at 37°C. ZEA was then added to the medium of experimental wells at a final concentration of 10 $\mu$g/mL, which can induce cytotoxicity in porcine intestinal epithelial cells as previously reported$^{12,13}$. An equal volume of phosphate buffer saline was added to the medium of control wells. ZEA-treated and control cells were cultured for 48 h and collected for RNA-seq and ATAC-seq (Table 1). Cell viability was gauged using the Cell Counting Kit-8 following the manufacturer’s protocols (Dojindo, Kumamoto, Japan) on the platform of Tecan Infinite 200 microplate reader (Sunrise, Tecan, Switzerland). Significant reduction of cell viability was observed upon ZEA exposure, indicating the toxic effects elicited by ZEA on IPEC-J2 cells (Fig. 1c). Three ZEA-treated and three control samples were collected for mRNA, microRNA, and IncRNA sequencing, respectively (Table 1). In addition, two ZEA-treated and two control samples were collected for ATAC-seq data is shown in Fig. 1b. These data will provide comprehensive insight into gene expression profiles and transcriptional regulation of animal cells in the response to ZEA exposure, which may aid the detection of biomarkers and drug targets for predicting and controlling ZEA contamination.

Library preparation for mRNA sequencing. Total RNA of the experimental samples was extracted using the Trizol method following the manufacturer’s protocols (Tiangen, Beijing, China). A total amount of 3 $\mu$g of RNA per sample was used for mRNA sequencing library preparations using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, MA, USA) following the manufacturer’s protocols. Index codes were added to attribute sequences to each sample. The library was quantified using the Qubit 2.0 Fluorometer (Thermo Scientific, MA, USA) and diluted into 1 ng/$\mu$L, and the library quality was assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Clustering of the index-coded samples was conducted on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumia, CA, USA) according to the manufacturer’s guidelines. Following cluster generation, the library preparations were then sequenced on the Illumina Hiseq. 2500 platform and 150 bp paired-end reads were yielded.
Library preparation for miRNA sequencing. Total RNA of the experimental samples was extracted using the Trizol method following the manufacturer’s protocols (Tiangen, Beijing, China). A total amount of 3 μg of RNA for each sample was used for sequencing library preparation using the NEBNext Small RNA Library Prep Set for Illumina (NEB, MA, USA) following the vendor’s instructions (Illumina, CA, USA). The library was quantified using the Qubit 2.0 Fluorometer (Thermo Scientific, MA, USA) and diluted into 1 ng/μL, and the library quality was assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) using DNA High Sensitivity Chips. Clustering of the index-coded samples was conducted on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumina, CA, USA) according to the manufacturer’s guidelines. The library preparations were then sequenced on the Illumina Hiseq. 2500 platform and 50 bp single-end reads were produced.

Library preparation for lncRNA sequencing. Total RNA of the experimental samples was extracted using the Trizol method following the manufacturer’s protocols (Tiangen, Beijing, China). A total amount of 3 μg of RNA of each sample was used for library construction, and ribosomal RNA was removed by Epicentre Ribo-zero rRNA Removal Kit (Epicentre, WI, USA). Sequencing libraries were prepared using the rRNA-depleted RNA by NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, MA, USA) following manufacturer’s recommendations. The library was quantified using the Qubit 2.0 Fluorometer (Thermo Scientific, MA, USA) and diluted into 1 ng/μL, and the library quality was evaluated on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Clustering of the index-coded samples was conducted on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumina, CA, USA) according to the manufacturer’s guidelines. The library preparations were then sequenced on the Illumina Hiseq. 2500 platform and 150 bp paired-end reads were produced.

Transcripts expression quantification. Paired-end clean reads of mRNA sequencing (Table 2) were aligned to the Sscrofa11.1 genome assembly (https://www.ncbi.nlm.nih.gov/genome/?term= pig) using HISAT2. More than 93% of the clean reads of each sample were mapped to the reference genome (Table 2). Read numbers mapped to each gene were counted using featureCounts. The FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) value of each gene was determined by the length of the gene and read counts mapped to this gene and used for estimating gene expression levels. Gene expression data have been uploaded in Figshare.

| Sample ID | Raw reads | Clean reads (%) | Q20 (%) | Q30 (%) | GC content (%) | Total mapped (%) | Accession |
|-----------|------------|----------------|---------|---------|----------------|------------------|-----------|
| mZEAT1    | 65442452   | 64092924 (97.94) | 97.59   | 93.64   | 53.82          | 60140448 (93.83) | SRR9945711 |
| mZEAT2    | 49134000   | 47847138 (97.38) | 97.47   | 93.41   | 53.64          | 44644114 (93.31) | SRR9945710 |
| mZEAT3    | 53309044   | 52070580 (97.68) | 97.64   | 93.79   | 53.35          | 48811296 (93.74) | SRR9945713 |
| mCTR1     | 49591798   | 48587854 (97.97) | 97.66   | 93.79   | 53.67          | 45639499 (93.93) | SRR9945712 |
| mCTR2     | 49852674   | 48856766 (98.00) | 97.74   | 94.00   | 54.13          | 45976517 (94.1)  | SRR9945715 |
| mCTR3     | 46269648   | 45523266 (98.39) | 97.66   | 93.39   | 54.43          | 42748329 (93.9)  | SRR9945714 |

Table 2. Summary statistics for mRNA, miRNA, and lncRNA sequencing data.

| Sample ID | Raw reads | Clean reads (%) | Q20 (%) | Q30 (%) | Non-mitochondrial mapped (%) | Accession |
|-----------|------------|----------------|---------|---------|-----------------------------|-----------|
| atacZEAT1 | 71215928   | 71214222 (99.99) | 98.88   | 95.99   | 61560564 (86.44)            | SRR945701 |
| atacZEAT2 | 53119698   | 53112547 (99.99) | 98.51   | 95.11   | 46843372 (88.20)            | SRR945700 |
| atacCTR1  | 61879862   | 61879361 (99.99) | 98.69   | 95.49   | 49919250 (80.67)            | SRR945720 |
| atacCTR2  | 44434012   | 44433766 (99.99) | 97.66   | 93.02   | 35316004 (79.48)            | SRR945721 |

Table 3. Summary statistics for ATAC-seq data.
For miRNA clean reads, length filter was first processed for all samples (Table 1). The filtered reads were then mapped to the Suscrofa11.1 genome assembly using Bowtie18 without mismatch to analyze their expression and distribution (Table 2). Mapped small RNA tags were utilized to identify known miRNAs using miRDeep219 based on miRBase 22 (http://www.mirbase.org/). Custom scripts were used to quantify the miRNA counts and base bias on the first position of identified miRNA with certain length. For novel miRNA prediction, miREvo20 and miRDeep2 19 were integrated to predict novel miRNAs by exploring the second structure, Dicer cleavage site, and minimum free energy of the unannotated small RNA tags. miRNA expression levels were normalized as follows: normalized expression = mapped read count × 10^6/library size^21. Target gene prediction of miRNAs

Fig. 2 Volcano plot of differential expression profiles between ZEA-treated and control groups. (a) Differential expression of mRNAs. (b) Differential expression of miRNAs. (c) Differential expression of lncRNAs. Green dots represent significant down-expression, and red dots represent significant up-expression.

Fig. 3 Reads distribution analysis and peak calling of ATAC-seq data. (a) Read density within 3 kb upstream and 3 kb downstream of the transcription start site. (b) Motif enrichment differences of known transcription factors in ZEA-treated and control samples. (c) Hierarchical clustering of peak enrichment patterns between ZEA-treated and control groups. TSS: transcription start site.
was performed using miRanda. The expression levels of known miRNAs and novel miRNAs, and the predicted target genes are available at Figshare.

The paired-end clean reads of lncRNA sequencing (Table 2) were aligned to the Sscrofa11.1 genome assembly using HISAT2. The mapped reads of each sample were assembled using StringTie via a reference-based method. FPKM of lncRNAs in each sample was then calculated using StringTie. The assembled transcripts were selected based on following criteria: number of exons ≥ 2; the length > 200 bp nucleotides; non-overlap with the annotated exons in the reference genome. Four programs including Pfam-scan (v1.3), CPC2 (v0.1), PhyloCSF (v20121028), and CNCI (v2) with default parameters were used to assess the coding potential of transcripts. Transcripts predicted with coding potential by any of the four tools were removed, and those without coding potential were considered as candidate lncRNAs. Prediction of lncRNA-mRNA co-location networks was conducted with the parameters of upstream and downstream 100 kb of the location of lncRNAs. LncRNA-mRNA co-expression networks were predicted with R function "cor.test", and mRNAs with absolute value of the correlation coefficient greater than 0.95 were retained. The expression levels of lncRNAs, lncRNA-mRNA co-location networks, and lncRNA-mRNA co-expression networks are available at Figshare.

Differential expression analysis. Following expression quantification, differential expression analyses of miRNAs, lncRNAs, and mRNAs between ZEA-treated and control groups were performed using DESeq. Benjamini and Hochberg's method was applied to correct the resulting P-values for controlling false discovery rate. The mRNAs with a corrected P-value < 0.05 and |log2 fold change| ≥ 1 were defined as differentially expressed (Fig. 2a). The miRNAs with |log2 fold change| ≥ 1 and a P-value < 0.05 were defined as differential expression miRNAs (Fig. 2b). The lncRNAs with a corrected P-value < 0.05 were defined as differential expression lncRNAs (Fig. 2c). The differential expression data of miRNA, lncRNA, and mRNAs are available at Figshare.

Library construction for ATAC-seq. ATAC-seq was conducted according to the protocols previously reported. In brief, the nuclei were extracted and resuspended in the Tn5 transposase reaction mix. The transposition reaction was incubated at 37 °C for 30 min. Post transposition, the equimolar adapter1 and adapter 2 were added, and then PCR was performed to amplify the library. The library was purified with the AMPure beads and measured with Qubit 2.0 Fluorometer for quality assessment (Thermo Scientific, MA, USA). Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumia, CA, USA) following the manufacturer's instructions. The library preparations were sequenced on the Illumina Hiseq. 2500 platform by Novogene Bioinformatics Institute (Novogene, Beijing, China) and 150 bp paired-end reads were generated.

ATAC-seq data analysis. Paired-end clean reads (Table 3) were aligned to the Sscrofa11.1 genome assembly using BWA with default parameters. Read density (Fig. 3a) within 3 kb upstream and 3 kb downstream of the

![Fig. 4](https://example.com/fig4.png)

Fig. 4 Quality assessment of mRNA, miRNA, and lncRNA sequencing data. (a) Error rate distribution along mRNA sequencing reads. (b) Error rate distribution along miRNA sequencing reads. (c) Error rate distribution along lncRNA sequencing reads. (d) Read distribution in genomic contexts of exon, intron, and intergenic regions. (e) Length distribution of mapped miRNA sequencing reads.
transcription start site was calculated using computeMatrix of DeepTools. Peak calling was then performed using MACS2. All reads were shifted towards the 3′ direction to the length of insert fragments, and the dynamic λ of each 200 bp sliding window was calculated. P values of each window were calculated based on the Poisson distribution and corrected using the false discovery rate method. The regions with a corrected P-value < 0.05 were defined as peaks, and the peaks have been submitted to Figshare.

The Homer software suite was utilized to recognize motif sequence in the 250 bp upstream and 250 bp downstream of the peak summits. Motif sequences were matched to the known motifs of transcription factors (Fig. 3b). The distance of peak summits to the nearest

---

**Fig. 5** Pearson correlation analysis of experimental samples within ZEA-treated and control groups. (a) Pearson correlation between samples used for mRNA expression analysis. (b) Pearson correlation between samples used for miRNA expression analysis. (c) Pearson correlation between samples used for lncRNA expression analysis.
transcription start site and corresponding genes were analyzed using the PeakAnalyzer tool\textsuperscript{34}. Differential peaks between ZEA-treated and control groups were identified by calculating the ratio of fold rich between the two groups. Peaks with $|\log_2$ fold rich ratio$| \geq 1$ were defined as differential peaks. Hierarchical clustering analysis was performed to display the enrichment pattern of peaks in the two groups (Fig. 3c). Differential peaks between the two groups have been submitted to Figshare\textsuperscript{17}.

Data Records
The sequencing data of all experimental samples in the fastq format have been submitted to the Sequence Read Archive of NCBI under the accession number SRP218038. The files of gene expression level and differential expression data between the two groups have been deposited in Figshare\textsuperscript{17}.

Technical Validation
RNA quality control. RNA degradation and contamination was monitored on 1% agarose gels. The concentration and integrity of RNA samples were measured using the Qubit Fluorometer (Thermo Scientific, MA, USA) and Agilent 2100 Bioanalyzer platform (Agilent Technologies, CA, USA). Samples with rRNA ratio $(28S/18S) \geq 1.9$ and RNA integrity number $\geq 8$ were subjected to sequencing library construction.

Quality validation and analyses. We examined the error rate of mRNA (Fig. 4a), miRNA (Fig. 4b), and lncRNA (Fig. 4c) read sequence and found high-quality sequences across all read bases. Raw sequencing data of mRNA, miRNA, and lncRNA (Table 2) were filtered to remove the reads with 5' adapter contaminants, without 3' adapter or the insert tag, with the proportion of N base greater than 10%, with poly A/T/G/C, and low quality reads (proportion of the bases with Qphred $\leq 20$ greater than 30% of the total read bases) using FastQC\textsuperscript{36}. All samples produced >97% clean reads after quality control, and >90% of clean reads were mapped to the reference genome (Table 2). In parallel, Q20, Q30, and GC content of the clean data were calculated (Table 2). These analyses indicated the high-quality of library construction and sequencing data of experimental samples. Genomic distribution analysis showed that on average 86.59% of the mapped reads of control samples and 86.98% of the mapped reads of ZEA-treated samples were mapped to exons (Fig. 4d), suggesting the efficient reflection of genome-wide gene expressions. Length distribution analysis of miRNA read sequence showed that most of the reads (88.6%) were in the length of 21–24 nt (Fig. 4e), which was consistent with the biological features of small RNAs. Pearson correlation analysis was performed to further examine the reproducibility of biological replicates in different groups. Correlation coefficients of mRNA sequencing replicates within ZEA-treated and control groups were greater than 0.99 (Fig. 5a). Correlation coefficients of miRNA sequencing replicates within

---

**Fig. 6** ATAC-seq data quality control metrics of fragment size distribution and sequencing read enrichment around transcription start sites. TSS: transcription start site.
the two groups were greater than 0.98 (Fig. 5b), and those of lncRNA sequencing replicates within the two groups were greater than 0.85 (Fig. 5c).

For ATAC-seq data, raw reads (Table 3) were first trimmed using Skewer to remove the reads with sequencing adaptors, with proportion of N base greater than 10%, low quality reads (proportion of the bases with Qphred < 20 greater than 30% of the total reads bases), and the reads with the length smaller than 18 nt after trimming. To ensure the reliability of read mapping, reads with mapping quality > 13 and properly paired reads were retained for subsequent analysis. The size distribution of sequenced fragments displayed clear periodicity, and the regions around transcription start sites were enriched for ATAC-seq reads (Fig. 6). The two standard quality metrics demonstrated the ATAC-seq data quality to capture the accessible chromatin regions (Fig. 6). Moreover, the Pearson correlation coefficients are 0.969 of ZEA-treated replicates and 0.964 of control replicates (Fig. 7a), indicating high reproducibility of accessible chromatin regions between replicates within the two groups. We identified peaks by using the MACS2 program. Peak scores (−log10 (corrected P value)) were calculated and most of the peaks showed a peak score > 20 (Fig. 7b), indicating the high reliability of peak calling.

**Code availability**

HISAT2: [http://ccb.jhu.edu/software/hisat2/index.shtml](http://ccb.jhu.edu/software/hisat2/index.shtml). Version: 2.0.5. Parameters: --rna-strandness RF.

Bowtie: [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml](http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Version: 0.12.9. Parameters: -v 0 –k 1.

miRDeep2: [https://github.com/rajewsky-lab/mirdeep2](https://github.com/rajewsky-lab/mirdeep2). Version: 2.0.0.8. Parameters: quantifier.pl -p -m -r -y -g 0 -T 10.
References
1. Zinedine, A. et al. Review on the toxicity, occurrence, metabolism, detoxification, regulations and intake of zearalenone: An oestrogenic mycotoxin. Food Chem. Toxicol. 45, 1–18 (2007).
2. Zhao, F. et al. Multigenerational exposure to dietary zearalenone (zea), an oestrogenic mycotoxin, affects puberty and reproduction in female rats through the mechanism of premature activation of hypothalamic kisspeptin-gpr54 signaling. Mol. Cell. Endocrino. 437, 62–74 (2016).
3. Liang, Z. et al. Individual and combined effects of deoxynivalenol and zearalenone on mouse kidney. Environ. Toxicol. Pharmacol. 40, 686–691 (2015).
4. Al-Seeni, M., El-Sawi, N., Shaker, S. & Al-Amoudi, A. Investigation of the biochemical and histological changes induced by zearalenone mycotoxin on liver in male mice and the protective role of crude venom extracted from jellyfish Cassiopea andromeda. Food Nutr. Sci. 2, 314–322 (2011).
5. Vlata, Z. et al. A study of zearalenone cytotoxicity on human peripheral blood mononuclear cells. Toxicol. Lett. 165, 274–281 (2006).
6. Taranu, I. et al. Exposure to zearalenone mycotoxin alters in vitro porcine intestinal epithelial cells by differential gene expression. Toxicol. Lett. 232, 310–325 (2015).
7. Buenrostro, J. D. et al. Evaluation of cellular and molecular impact of zearalenone and Escherichia coli co-exposure on IPEC-1 cells using microarray technology. BMC Genomics 17, 576 (2016).
8. Thomos, D. W. & Dinger, M. E. Endogenous miRNA sponges: evidence and controversy. Nat. Rev. Genet. 17, 223–236 (2016).
9. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 200, i36–i43 (2004).
10. Birol, I. et al. Short read alignment with adaptive timing and exact seed handling. Genome Biol. 11, R110 (2010).
11. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
12. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
13. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell. 38, 576–589 (2010).
14. DeepTools Version: 3.0.2. Parameters: --cor Method Pearson.
15. MACS2: http://liulab.dfci.harvard.edu/MACS/. Version: 2.1.2. Parameters: --q 0.05--call-summits --nomodel --shift -100 --extsize 200 --keep-dup all.
16. Homer: http://homer.ucsd.edu/homer/. Version: 4.9.1. Parameters: -gc -len 8, 10, 12, 14.
17. Wang, H. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
18. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell. 38, 576–589 (2010).
19. Salmon-Divon, M., Dvinge, H., Tammioja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics 11, 415 (2010).
20. NCBI Sequence Read Archive, https://identifiers.org/insdc.sra:SRP218038 (2019).
21. Andrews, S. FastQC: a quality control tool for high throughput sequence data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2014).
22. Jiang, H. et al. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182 (2014).
Acknowledgements
This work was supported by grants from the National Natural Science Funds (Grant No. 31972535, No. 31702082), China Postdoctoral Science Foundation (2017M621842, 2018T110564), Qingnan Project of Yangzhou University, and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Author contributions
W.B. conceived and supervised the study. H.W., J.W. and H.Q. performed the experiments. H.W., J.J. and S.W. performed data analyses. H.W. and W.B. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to W.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2019