BMC Genomic Data

DATA NOTE

Open Access

The full-length transcriptional of the multiple spatiotemporal embryo-gonad tissues in chicken (*Gallus gallus*)

Kai Jin[1,2,3*], Qisheng Zuo[1,2,3], Jiuzhou Song[4], Ahmed Kamel Elsayed[5], Hongyan Sun[1,2,3], YingJie Niu[1,2,3], Yani Zhang[1,2,3], Guobin Chang[1,2,3], Guohong Chen[1,2,3] and Bichun Li[1,2,3]

Abstract

Objectives Chicken (*Gallus gallus*), as the most economically important poultry, is a classical and ideal model for studying the mechanism of vertebrate developmental biology and embryology. However, the sex determination and differentiation in chicken is still elusive, which limited the application and slowed down many basic studies in chicken.

Data description We applied PacBio Iso-seq to multiple spatiotemporal embryo-gonad tissues in the male and female chicken, which contain the blastoderm (E0, un-differentiation stage), genital ridge (E3.5–6.5, sex-differentiation stage) and gonads (E18.5, full-sex-differentiation stage). We obtained 51,479 and 48,356 full-length transcripts in male and female chicken embryo, respectively. The comprehensive annotated and evaluated these transcripts. The 1,293 and 1,556 candidate lncRNAs, 5,766 and 4,211 AS events in male and female. Collectively, our data constitutes a grand increase in the known number of lncRNA, AS (Alternative splicing) and Poly(A) during chicken embryo sex-differentiation and plays an important role in improving current genome annotation. In the meantime, the data will be enriched the functional studies in other birds.

Keywords Chicken, Sex determination, Embryo-gonad tissues, Full-length transcript

Objective

The Chicken (*Gallus gallus*) is widely used in developmental biology and embryology due to its economic value in the poultry industry [1–4]. Understanding sex determination and differentiation is crucial as it impacts traits like growth and reproduction [5–7]. While female chickens are preferred in layer breeding, males are favored for meat production [8]. Despite the clear Z and W sex chromosomes, the mechanisms behind sex determination in chickens is still elusive [9, 10]. The embryonic gonad originates from the genital ridge at day 3.5 (E3.5) and undergoes sex-specific changes by E6.5, developing into either ZZ testis or ZW ovary [11, 12]. Although the chicken genome was sequenced in 2004 and RNA-seq has advanced, identifying genes involved in sex differentiation is still challenging [13–16]. CA Smith identified

*Correspondence:
Kai Jin
007838@yzu.edu.cn
[1] Key Laboratory of Animal Breeding Reproduction and Molecular Design for Jiangsu Province, College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China
[2] Institutes of Agricultural Science and Technology Development, Yangzhou University, Yangzhou 225009, China
[3] Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou 225009, China
[4] Animal & Avian Sciences, University of Maryland, College Park, MD 20741, USA
[5] Anatomy and Embryology Department, Faculty of Veterinary Medicine, Suez Canal University, Ismailia 41522, Egypt

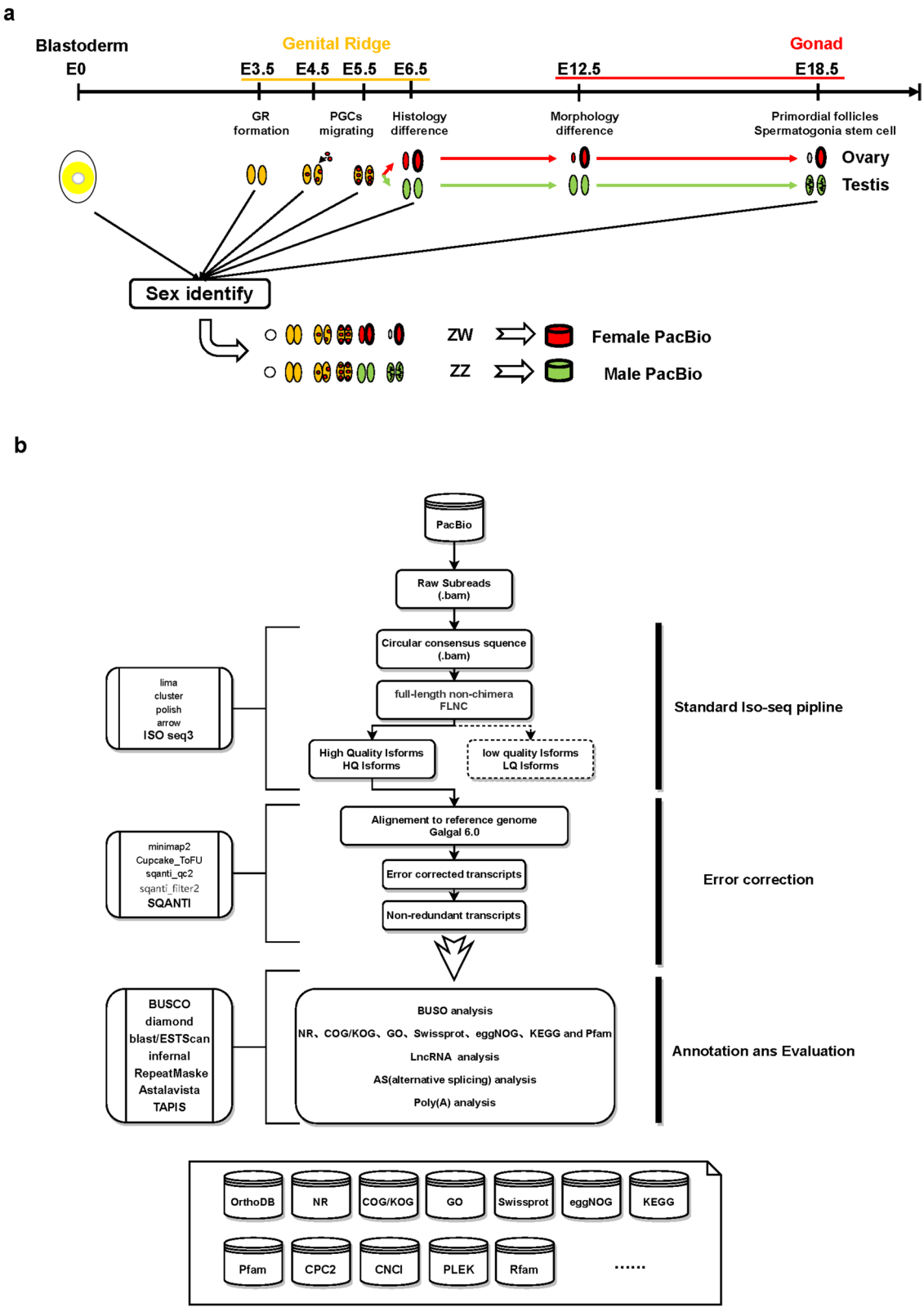


Fig. 1 Experimental design and standard Iso-Seq pipeline for raw data processing

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Figure_1 Experimental design and standard Iso-Seq pipeline for raw data processing	Image file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.26841016.v1 [33]
Data file 2	Figure_2 The annotation statistics of male and female	Image file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.26841022.v1 [34]
Data file 3	Figure_3 KEGG pathway and GO functional annotations of the male and female full-length transcripts	Image file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.26841028.v1 [35]
Data file 4	Figure_4 Characterization of identified novel lncRNAs	Image file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.26841034.v1 [36]
Data file 5	Figure_5 The total number of AS events and Poly(A) Sites	Image file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.26841058.v1 [37]
Data file 6	Table 1 The purity and completeness of RNA for library	Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.26841139.v4 [38]
Data file 7	Table_2 Read number and length distribution after ISO-Seq analysis	Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.26841187.v1 [39]
Data file 8	Table_3 BUSCO analysis of Transcrpt completeness	Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.26841229.v1 [40]
Data file 9	Table_4 Annotation Statistics	Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.26841259.v1 [41]
Data file 10	Table_5 The annotation of male-biased uniq-transcripts	Excel file (.xls)	Figshare, https://doi.org/10.6084/m9.figshare.26841274.v1 [42]
Data file 11	Table_6 The annotation of female-biased uniq transcripts	Excel file (.xls)	Figshare, https://doi.org/10.6084/m9.figshare.26841280.v1 [43]
Data file 12	Table_7 The annotation of common uniq-transcripts in male and female	Excel file (.xls)	Figshare, https://doi.org/10.6084/m9.figshare.26841286.v1 [44]
Data set 1	Pacbio of male chicken:multiple spatiotemporal embryo-gonad tissues	SAM file (*.bam)	NCBI Sequence Read Archive (SRA), http://identifiers.org/insdc.sra:SRX9530712 [45]
Data set 2	Pacbio of female chicken:multiple spatiotemporal embryo-gonad tissues	SAM file (*.bam)	NCBI Sequence Read Archive (SRA), http://identifiers.org/insdc.sra:SRX9530713 [46]

a series of homologous genes in the Z/W chromosome and described the discrepancy via RNA-seq [17]. However, the Short-read sequencing is insufficient for accurately identifying long non-coding RNAs (lncRNAs) and alternatively spliced (AS) genes, which may critical for this process [18–20]. The full-length (FL) RNA-seq provides a more accurate representation of lncRNA and gene isoforms, improving genome annotation related to sex differentiation.

In this study, we performed the PacBio Iso-seq for multiple spatiotemporal embryo-gonad tissues of chicken at different times of the developmental stages including; the blastoderm (E0, un-differentiated stage), genital ridge (E3.5–6.5, sex-differentiation stage) and gonads (E18.5, full-sex differentiation stage). Collectively, we obtained 51,479 and 48,356 full-length transcripts from male and female chicken embryonic reproductive organs, respectively. The subsequent systemic functional annotation of these full-length transcripts detected 1,293 and 1,556 candidate LncRNA as well as 5,766 and 4,211 AS events in male and female sex-determining tissues, respectively. This comprehensive dataset provides valuable insights into the roles of lncRNAs and AS events during sex differentiation in chickens and is a critical resource for future studies on sex determination in birds. These data also provide a valuable resource for genomic annotation at different specific chicken embryological developmental stages.

Data description

The Fertilized eggs from Rugao Yellow Chicken were obtained from the Poultry Institute, Chinese Academy of Agricultural Sciences. Eggs were incubated at 37 °C with 75% humidity using a Brinsea Incubator Ova-Easy 100. Embryonic tissues from both male and female chickens were collected at three developmental stages: blastoderm (E0), genital ridge (E3.5–6.5), and gonads (E18.5) (Fig. 1). Tissues were flash-frozen in liquid nitrogen and stored at -80 °C for RNA extraction and sex identification. The RNA was extracted using TRIzol reagent, following the manufacturer's protocol. RNA integrity was assessed using a NanoDrop2000 spectrophotometer and an Agilent 2100 Bioanalyzer (Data file 6). Sex identification was conducted via PCR amplification of the Chd1 gene, with males showing a single 580 bp band and females showing two bands (580 bp and 423 bp). The two

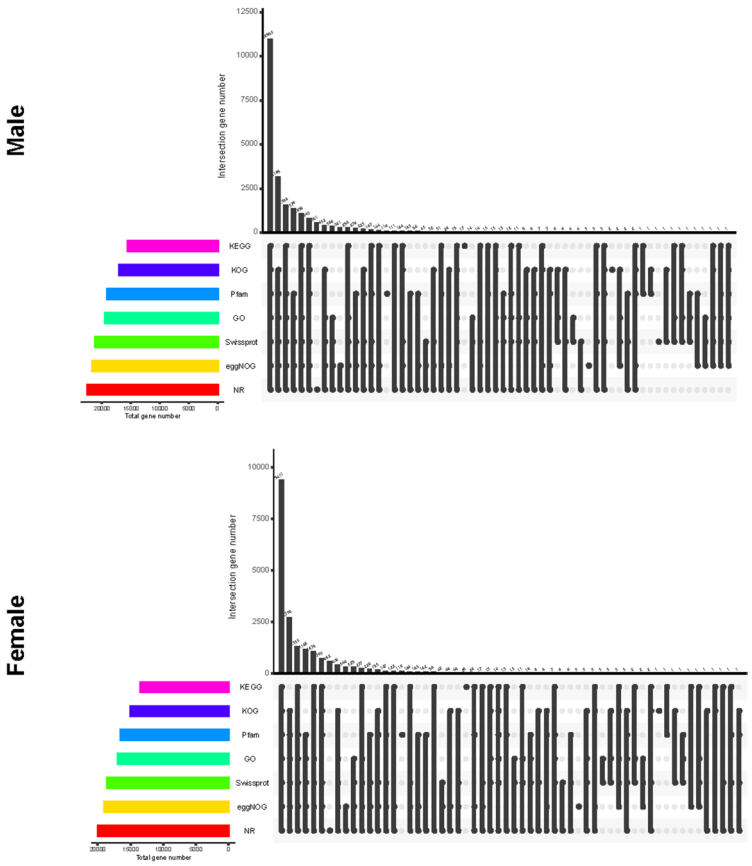


Fig. 2 The annotation statistics of male and female

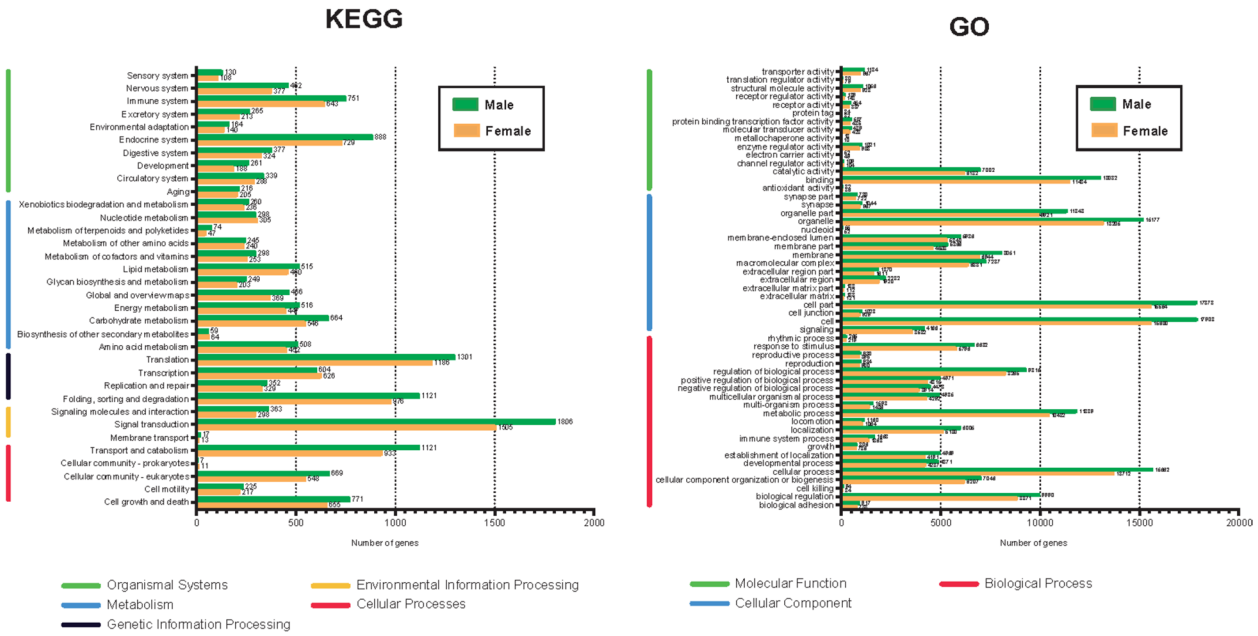


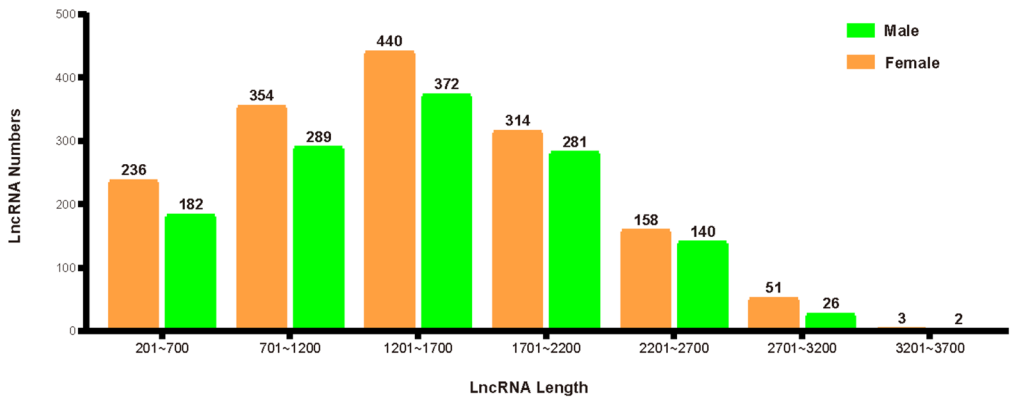
Fig. 3 KEGG pathway and GO functional annotations of the male and female full-length transcripts

iso-seq libraries were created by pooling RNA from male and female tissues separately. Equal amounts of RNA from each developmental stage (10 μg per tissue) were combined. cDNA was synthesized using the SMARTer PCR cDNA Synthesis Kit, and SMRT bell libraries were constructed using the Pacific Biosciences DNA Template Prep Kit. Sequencing was performed using the PacBio Sequel System(Pac Bio's Iso-seq™). The PacBio data were processed and evaluated following the figured standard pipeline of the Iso-seq analysis (Fig. 1). Briefly, the raw data were processed using the SMRTlink software to generate circular consensus sequences (CCS). Sequences were refined using IsoSeq3, generating high-quality (HQ) non-chimeric sequences[21, 22]. HQ isoforms were mapped to the Galgal6 reference genome using *minimap2* [23], and redundant sequences were collapsed using *Cupcake-ToFU*. The resulting non-redundant transcripts were analyzed with *SQANTI2* for quality control [24]. High-quality transcripts were further annotated using the OrthoDB database, and sequences were aligned with NR, SwissProt, and COG/KOG databases for functional annotation (Fig. 2 and Data file 9). Gene Ontology (GO) classification and KEGG pathway analysis were conducted for deeper insights [25]. The-Functional annotation was performed using databases like NR, SwissProt, and Pfam [26, 27]. GO terms were assigned to each isoform, identifying key biological processes, cellular components, and molecular functions.

a

Term	All(>=200bp)	>=2000bp	>=6000bp	N50	Total_Length	Max_Length	Min_Length	Average_Length
Female	1556	327	0	1727	2239317	3521	202	1439.15
Male	1293	259	0	1724	1864240	3474	200	1441.79

b



c

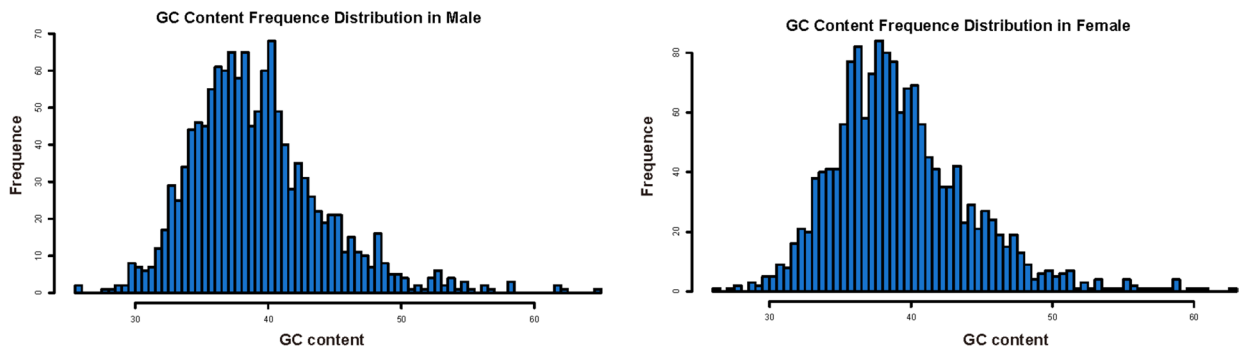


Fig. 4 Characterization of identified novel lncRNAs

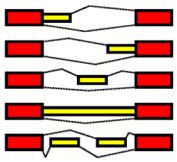
KEGG pathway analysis classified transcripts into cellular processes, metabolism, and organismal systems, among others (Fig. 3). A total of 1,293 male and 1,556 female candidate lncRNAs were identified using a customized pipeline based on CPC2, CNCI, Pfam, and PLEK databases [27–29].As a final outcome, the interesting isoforms without coding potentials are considered as our candidate lncRNA(Fig. 4) [30, 31].The Alternative splicing (AS) events were identified using *Astalavista*, with exon skipping being the most common type (Data file 10, 11 and 12) [32]. Male tissues showed more AS events than female, suggesting a role in sex differentiation. Poly(A) site analysis revealed differences in distribution between male and female tissues, which exhibit the different pattern of Poly(A) site like UBP1, indicating that alternative polyadenylation may play a critical role during sex differentiation (Fig. 5). All the results indicating that the complexity and diversity of transcription is enhanced by AS and other post-transcription regulation during chicken sex differentiation. Finally, the transcripts were clustered in to total 26,089 male and 23,889 female non-redundant transcripts which were used for further analysis. All data obtained from the Iso-Seq3 analysis were listed in Data file 7. The BUSCO orthologs software was used to assess the transcript completion. As we concerned, the percentage of complete and single-copy BUSCO genes (vertebrata_odb9 dataset, 65 species, 2586 sequences) is 59.1% and 52.6% in male and female full-length transcripts, respectively (Data file 8).

Collectively, our data represent the first comprehensive full-length transcriptomic resource for chicken embryo sex differentiation, including predicted lncRNA, alternative splicing (AS) events, and Poly(A) signals identified through Iso-seq technology. These findings significantly expand the known repertoire of lncRNAs, AS events, and Poly(A) signals involved in chicken embryo sex differentiation, contributing to improved genome annotation. Importantly, this data may also inform breeding strategies by providing molecular insights into sex determination, offering potential applications to address challenges in the poultry industry related to sex differentiation. Furthermore, the findings enrich functional studies in other bird species and provide a valuable resource for broader vertebrate developmental biology research.

Limitations

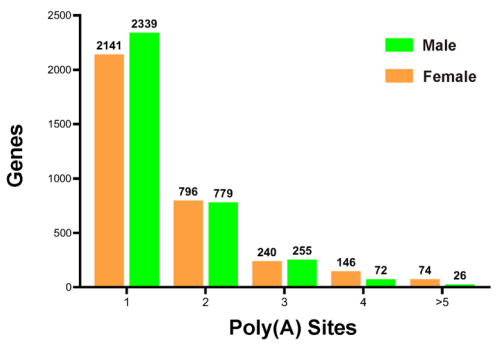
The long-read sequencing technology can accurately identify the full length of transcripts, but it still has a significant disadvantage of a high error rate. This issue can be mitigated by combining it with RNA-seq. Therefore, in this study, the results obtained from analyzing samples of both sexes independently still need to be verified through molecular biology techniques (such as qPCR, Northern Bolting, Western Blotting, and in situ Hybridization) on these transcripts in the future.

a



AS class		female		male	
		Numbers	Percentage(%)	Numbers	Percentage(%)
Alternative 5' donor sites	A5SS	457	10.83%	596	10.34%
Alternative 3' accptor sites	A3SS	585	13.86%	878	15.23%
Exons skipping/inclusion	ES	2116	50.13%	2748	47.66%
Intron retention	IR	993	23.53%	1439	24.96%
Mutually exclusive exons	MXE	70	1.66%	105	1.82%
Total		4221		5766	

b



c

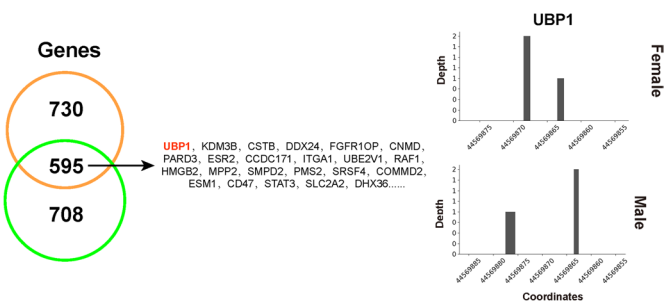


Fig. 5 The total number of AS events and Poly(A) Sites

Code availability

The version and parameters of main software tools are described below:

(1) SMARTLink: version (v6.0), parameters: pbccs.task options.max.length=20,000 pbccs.task options. min_length=300.
(2) Cupcake-ToFU: version (v4.1), parameters: -i 0.85.
(3) BUSCO: version (v3.0.1), parameters: default.
(4) diamond: version (v0.9.7), parameters: --more-sensitive-e 1e-5.
(5) kobas: version (v3.0), parameters: default.
(6) blast+: version (v2.6.0), parameters: -evalue 1e-10.
(7) CPC2: version (v2), parameters: default.
(8) CNCI: version (v2.0), parameters: -m ve.
(9) Pfam: version (v2015-06-02), parameters: -e_seq 0.001.
(10) Astalavista (v4.0), parameters: default.
(11) TAPIS (v1.2.1), parameters: default.

Abbreviations

AS	Alternative splicing
lncRNA	Long non-coding RNA
FL	Full-length
LN_2	Liquid nitrogen
RIN	RNA concentration and its integrity number value
Chd1	Chromo-helicase DNA binding 1
CCS	Circular consensus sequence
FLNC	Full length non-chimeric sequences
ICE	Iterative cluster merging
HQ	High quality
NR	NCBI non-redundant
GO	Gene Ontology
KEGG database	Kyoto Encyclopedia of Genes and Genomes
ORF	Open Reading Frame
IR	Intron retention
ES	Exons skipping/inclusion
A5	Alternative 5′donor sites
A3	Alternative 3′acceptor sites
MXE	Mutually exclusive exons

Acknowledgements
We thank Dr. Jing Wang and Prof. Jiuzhou Song for assisting in the preparation of this manuscript.

Authors' contributions
Kai Jin and Qisheng Zuo collected samples, analyzed data and drafted the manuscript. Jiuzhou Song involved in the data analysis. Kai Jin and Ahmed Kamel Elsayed wrote the manuscript, suggested the analysis pipeline. Hongyan Sun, YingJie Niu and Yani Zhang revised and improved the manuscript draft. Guohong Chen and Bichun Li conceived and supervised the project. All authors read and approved the final manuscript.

Funding
The work was supported by the earmarked fund for National Natural Science Foundation of China (32202655, 32172718 and 32372864), "JBGS" Project of Seed Industry Revitalization in Jiangsu Province (JBGS【2021】029), The Project funded by China Postdoctoral Science Foundation(2022M722697), and the High level talents support program of Yangzhou University and Priority Academic Program Development of Jiangsu Higher Education Institutions.

Data availability
The data described in this Data note can be freely and openly accessed on NCBI GenBank under the BioProject PRJNA670545 which include Biosample SAMN16512697 and Biosample SAMN16512680. Associated Data files are available on Figshare. Please see Data set 1, Data set 2 and Refs [33–46] for details and links to the data.

Declarations

Ethics approval and consent to participate
The animal experiments were approved by the Institutional Animal Care and Use Committee of the Yangzhou University Animal Experiments Ethics Committee (Permit Number: SYXK [Su] IACUC 2012–0029). All experimental procedures were performed in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals approved by the State Council of the People's Republic of China.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 14 June 2024 Accepted: 17 October 2024
Published online: 25 October 2024

References

1. Davey MG, Tickle C. The chicken as a model for embryonic development. Cytogenet Genome Res. 2007;117:231–9.
2. Vergara MN, Canto-Soler MV. Rediscovering the chick embryo as a model to study retinal development. Neural Dev. 2012;7:22.
3. Swanberg SE, et al. Telomere biology of the chicken: a model for aging research. Exp Gerontol. 2010;45:647–54.
4. Kim YM, Han JY. The early development of germ cells in chicken. Int J Dev Biol. 2018;62:145–52.
5. Douglas C, Turner JMA. Advances and challenges in genetic technologies to produce single-sex litters. PLoS Genet. 2020;16:e1008898.
6. Galli R, et al. Sexing of chicken eggs by fluorescence and Raman spectroscopy through the shell membrane. PLoS ONE. 2018;13:e0192554.
7. Tran HT, Ferrell W, Butt TR. An estrogen sensor for poultry sex sorting. J Anim Sci. 2010;88:1358–64.
8. Tizard ML, et al. Potential benefits of gene editing for the future of poultry farming. Transgenic Res. 2019;28:87–92.
9. Smith CA, Sinclair AH. Sex determination in the chicken embryo. J Exp Zool. 2001;290:691–9.
10. Hirst CE, Major AT, Smith CA. Sex determination and gonadal sex differentiation in the chicken model. Int J Dev Biol. 2018;62:153–66.
11. Chue J, Smith CA. Sex determination and sexual differentiation in the avian model. FEBS J. 2011;278:1027–34.
12. Lambeth LS, Cummins DM, Doran TJ, Sinclair AH, Smith CA. Overexpression of aromatase alone is sufficient for ovarian development in genetically male chicken embryos. PLoS ONE. 2013;8:e68362–e68362.
13. Dequéant ML, Pourquié O. Chicken genome: new tools and concepts. Dev Dyn an Off Publ Am Assoc Anat. 2005;232:883–6.
14. LDW Hiller et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432:695–716.
15. Wolf JB, Bryk J. General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq. BMC Genomics. 2011;12:91.
16. Ayers KL, et al. Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. BMC Genomics. 2015;16:704.
17. Ayers KL, et al. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. Genome Biol. 2013;14:R26–R26.
18. Dey BK, Mueller AC, Dutta A. Long non-coding RNAs as emerging regulators of differentiation, development, and disease. Transcription. 2014;5:e944014.

19. Planells B, Gómez-Redondo I, Pericuesta E, Lonergan P, Gutiérrez-Adán A. Differential isoform expression and alternative splicing in sex determination in mice. BMC Genomics. 2019;20:202.
20. Bayega A, et al. Transcript Profiling Using Long-Read Sequencing Technologies. Methods Mol Biol. 2018;1783:121–47.
21. Gordon SP, et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. PLoS ONE. 2015;10:e0132628.
22. Abdel-Ghany SE, et al. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. 2016;7:11706.
23. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
24. Tardaguila M, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28:396–411.
25. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
26. Prakash A, Jeffryes M, Bateman A, Finn, RD. The HMMER Web Server for Protein Sequence Similarity Search. Curr. Protoc. Bioinforma. 2017;60:3.15.1–3.15.23.
27. Finn RD, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30.
28. Kang Y-J, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45:W12–6.
29. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics. 2014;15:311.
30. Gardner PP, et al. Rfam: updates to the RNA family's database. Nucleic Acids Res. 2009;37:D136–40.
31. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–2935.
32. Foissac S, Sammeth M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res. 2007;35:W297–9.
33. Jin K. Data file 1: Figure1 Experimental design and standard Iso-Seq pipeline for raw data processing. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841016.v1.
34. Jin K. Data file 2: Figure2 The annotation statistics of male and female. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841022.v1.
35. Kai Jin. Data file 3: Figure3 KEGG pathway and GO functional annotations of the male and female full-length transcripts. Figshare. 2024.https://doi.org/10.6084/m9.figshare.26841028.v1.
36. Jin K. Data file 4: Figure4 Characterization of identified novel lncRNAs. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841034.v1.
37. Jin K. Data file 5: Figure5 The total number of AS events and Poly(A) Sites. 2024. Figshare.https://doi.org/10.6084/m9.figshare.26841058.v1.
38. Kai Jin.Data file 6: Table1 The purity and completeness of RNA for library. Figshare. 2024.https://doi.org/10.6084/m9.figshare.26841139.v4.
39. Jin K. Data file 7: Table2 Read number and length distribution after ISO-Seq analysis. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841187.v1.
40. Jin K. Data file 8: Table3 BUSCO analysis of Transcrpt completeness. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841229.v1.
41. Jin K. Data file 9: Table4 Annotation Statistics. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841259.v1.
42. Jin K. Data file 10: Table5 The annotation of male-biased uniq-transcripts. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841274.v1.
43. Jin K. Data file 11: Table6 The annotation of female-biased uniq transcripts. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841280.v1.
44. Jin K. Data file 12: Table7 The annotation of common uniq-transcripts in male and female. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26841286.v1.
45. Kai Jin. Data set 1: Pacbio of male chicken:multiple spatiotemporal embryo-gonad tissues. NCBI. 2024. Identifier, http://identifiers.org/insdc.sra:SRX9530712.
46. Kai Jin. Data set 2: Pacbio of female chicken:multiple spatiotemporal embryo-gonad tissues. NCBI. 2024. Identifier, http://identifiers.org/insdc.sra:SRX9530713.

Publisher's Note