## DATA NOTE

# Full-length transcriptome profiling for fruit development in *Diospyros oleifera* using nanopore sequencing

Yang Xu, Cui-yu Liu, Wen-qiang Cheng, Kai-yun Wu and Bang-chu Gong*

## Abstract

**Objectives**  *Diospyros oleifera*, one of the most economically important *Diospyros species*, is an ideal model for studying the fruit development of persimmon. While, the lack of whole-transcriptome has hindered the complex transcriptional regulation mechanisms of sugar and tannin during fruit development.

**Data description**  We applied Oxford Nanopore Technologies to six developmental stage of fruit from *D. oleifera* for use in transcriptome sequencing. As a result of full-length transcriptome sequencing, 55.87 Gb of clean data were generated. After mapping onto the reference genome of *D. oleifera*, 51,588 full-length collapsing transcripts, including 2,727 new gene loci and 43,223 transcripts, were obtained. Comprehensively annotated, 38,086 of new transcripts were functional annotation, and 972 lncRNAs, 7,159 AS events were predicted. Here, we released the transcriptome database of *D. oleifera* at different stage of fruit development,which will provide a fundamention of to investigatethe transcript structure, variants and evolution of persimmon.

**Keywords**  *Diospyros oleifera*, Fruit development, Nanopore sequencing, Full-length transcriptome

## Objective

There are approximately 500 species in the genus *Diospyros*, which range in ploidy level from diploid $(2n = 2x = 30)$ to nonaploid $(2n = 9x = 135)$ [1, 2]. Among these species, *Diospyros oleifera* and *Diospyros kaki* have been cultivated as important fruit crops in east Asia for centuries., these edible fruitstare rich in vitamins, sugars, nutrients, and antioxidants that are important for optimum health [3, 4]. Furthermore, *D. oleifera* is diploid $(2n = 2 \times = 30)$ and is closely related to *D. kaki* $(2n = 6 \times = 90)$ [4, 5]. As an added advantage, *D. oleifera* could be used as a model plant for studies of *Diospyros* [4, 6, 7].

Fruit development plays an important role in the life cycle of higher plants. *D.oleifera* will also be a potential model plant for studies of sugar synthesis and transformation, tannin formation and deastringency, coordination network of tannin and sugar during fruit development. Although we have reported the *D. oleifera* genome [6], transcript profile data on *Diospyros* during fruit development is insufficient compared with those of other fruit [8–10]. Even, no full-length transcriptome of *D.oleifera* has been reported. In this study, the ONT was used to generate large-scale full-length transcripts and collect the gene expression profile of *D. oleifera* fruit development.These data will provide gene sequence information and comprehensive understanding of the fruit development of persimmon.

*Correspondence:
Bang-chu Gong
gongbc@126.com
Research Institute of Subtropical Forestry, Chinese Academy of Forestry,
No. 76, Daqiao Road, Fuyang District, Hangzhou, China

## Data description

The fruit flesh of *D. oleifera* were obtained from 10 years-old plant in LanXi Plant Nursery (E, 119°28′27.274″; N, 29°8′48.946″), which located in LanXi City, Zhejiang Province. Three biological replicates were harvested at six development stages: (10 days after pollination (DAP) (T01-T03), 40 DAP(T04-T06), 100 DAP(T07-T09), 160 DAP(T10-T12), 180 DAP(T13-T15) and 200 DAP(T16-T18)). An RNeasy Plant Mini kit (Qiangen, 74,904) was used to extract total RNA, which was then treated with RNase-free DNase I (TAKARA, D2215). Nanodrop 2000 and Agilent 2100 were used to assess RNA quality (Data file 1). 1ug of total RNA was used for cDNA libraries with the protocol of Oxford Nanopore Technologies (ONT)(Oxford Nanopore Technologies, Oxford, UK). FLO-MIN109 flowcells were used to run the final cDNA libraries at Biomarker Technology Company (Beijing, Chinai), using the PromethION platform.

First, raw reads were filtered under the standard of an average read quality score is not lower than 7 and a read length is not lower than 500 bases [11]. Ribosomal RNA (rRNA) were discarded after mapping to rRNA database. Full-length transcripts (FLs) were identifiedusing the primers at both ends of cleaned reads. Full-length andnon-chemiric (FLNC) transcripts were clustered via mapping to *D. oleifera* reference genome [6] with mimimap2 [12]. Then consensus isoforms were obtained from each cluster using pinfish. Mapped reads were further collapsed to remove redundant FLs with 85% of min-coverage and 90% of min-identity by cDNA_Cupcake package. 5' difference was not considered when collapsing redundant transcripts. A single transcript of fusion candidates must conform the following criteria: (1) map loci must be more than or equal to 2, (2) coverage for each loci is > =5% and minimum coverage in bp is more than or equal to 1 bp, (3) total coverage is > =95%, (4) distance between the loci is not shorter than 10 kb.

Alternative splicing (AS) events and alternative polyadenylation (APA) events were identified by AStalavista tool (v3.2) [13] and TAPIS [14], respectively. The coding sequences and corresponding amino acid sequences was predicted by TransDecoder v3.0.0 [15]. GMAP (http://research-pub.gene.com/gmap/, v2017-11–15) was used to identify new transcripts. Four computational approaches include Coding Potential Calculator (CPC) [16], Coding-Non-Coding Index (CNCI) [17], Coding Potential Assessment Tool (CPAT) [18], and Pfam reference protein databases [19] were combined to sort non-protein cosubsequent to filtering. Long Non-coding RNAs (lncRNAs) were identified under the standard of at least 200 nt and two exons. Target genes regulating by identified lncRNAs were predicted using LncTar (v1.0) [20].

The annotations of transcripts were performed with e-values of $1e^{-5}$ on eight databases, including non-redundant protein sequence database(NR) [21], the database of Homologous protein family (Pfam) [19], eukaryotic Ortholog Groups(KOG) [22], Clusters of Orthologous Groups of proteins [23], evolutionary genealogy of genes: Non-supervised Orthologous Groups(eggNOG) [24], a manually annotated, non-redundant protein sequence database(Swiss-Prot) [25], Kyoto Encyclopedia of Genes and Genomes (KEGG) [26] and Gene Ontology(GO) [27].

Full-length reads were mapped to the reference transcriptome sequence, and then reads with match quality above 5 after mapping were further used to quantify. The absolute CPM (counts per million) value more than 0.1 was considered as a reliable expression. Differential expression analysis of two samples was performed using the DESeq R package (1.18.0) [28] with the following criteria: FDR < 0.01 and fold-change ≥ 2.

We applied Oxford Nanopore Technologies on six developmental stages of *D. oleifera* fruits for transcriptome sequencing (Data file 1). As a result, a total of 55.87 Gb clean data were generated (Data file 2, Data set 1- Data set 18). After mapping onto the reference genome of *D. oleifera* and discarding rRNA, we obtained 1,190,459 to 3,046,317 full-length reads (FL reads) from each sample (Data file 3, Data set 1-Data set 18). Though clustering, we obtained 51,588 full-length collapsing transcripts with an average length of 1,311 bp. And then, 43,223 new transcripts were identified among these collapsing redundant transcripts. Comprehensively annotated, 38,086 of new transcripts were functional annotation. In total, 35, 243 genes were detected, including 32,406 genes with functional annotation and 2,727 newly identified genes (Data file 4). 7,159 Alternative Splicing (AS) events were detected, as shown in Data file 5 and Data file 6 including 100 mutually exclusive exons, 2,115 intron retention (IR) events, 1,698 exon skipping (ES) events, 1,553 5' AS (Alt. 5') sites and 1,693 3'AS (Alt. 3') sites. We further detected 9274–13,034 APA events (Data file 7) and 14 -52 fusion genes (Data file 8) in each sample. 972 lncRNAs were screened and classified as shown in Data file 9. And, Data file 10 shows the target genes for 933 lncRNAs. We also found that 19, 276 genes and 39,969 transcripts were diferentially expressed during fruit development. Moreover, differentially expressed genes (DEGs) and differentially expressed transcripts (DETs) between all pairs of adjacent stages were also shown in Data file 11 and Data file 12. The dataset not only can offer fundamental genetic information to investigate transcript structure, variants and evolution of persimmon, but also can offer a reference to further analyse

Xu *et al. BMC Genomic Data*          (2023) 24:17

Page 3 of 5

**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 | Summary of sequencing sample and strategies in this study | MS Word fle(.docx) | figshare(https://doi.org/10.6084/m9.figshare.19314470) [29] |
| Data file 2 | Statistic of ONT-sequencing in this study | MS Word fle(.docx) | figshare(https://doi.org/10.6084/m9.figshare.19314515) [30] |
| Data file 3 | Read number and length distribution of FLNC and Collapse transcripts after ONT-Seq analysis | MS Word fle(.docx) | figshare(https://doi.org/10.6084/m9.figshare.19314524) [31] |
| Data file 4 | Gene information and database annotations | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314536) [32] |
| Data file 5 | The total number of AS events in detected genes and transcripts | MS Word fle(.docx) | figshare(https://doi.org/10.6084/m9.figshare.19314539) [33] |
| Data file 6 | The characteristics of AS events in each sample | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314545) [34] |
| Data file 7 | The statistical lists of APA events for each sample | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314548) [35] |
| Data file 8 | The statistical list of all fusion gene for each sample | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314563) [36] |
| Data file 9 | The result of LncRNAs classifications | Portable Document Format file (.pdf) | figshare(https://doi.org/10.6084/m9.figshare.19314566) [37] |
| Data file 10 | The information of target genes of these 933 lncRNAs | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314569) [38] |
| Data file 11 | The quantitative gene expression of all DEGs | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314572) [39] |
| Data file 12 | The quantitative gene expression of all DETs | MS Excel file(.xls) | figshare(https://doi.org/10.6084/m9.figshare.19314584) [40] |
| Data set 1 | Nanopore sequencing reads of T01 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918124) [41] |
| Data set 2 | Nanopore sequencing reads of T02 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918123) [42] |
| Data set 3 | Nanopore sequencing reads of T03 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918114) [43] |
| Data set 4 | Nanopore sequencing reads of T04 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918113) [44] |
| Data set 5 | Nanopore sequencing reads of T05 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918112) [45] |
| Data set 6 | Nanopore sequencing reads of T06 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918111) [46] |
| Data set 7 | Nanopore sequencing reads of T07 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918110) [47] |
| Data set 8 | Nanopore sequencing reads of T08 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918109) [48] |
| Data set 9 | Nanopore sequencing reads of T09 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918108) [49] |
| Data set 10 | Nanopore sequencing reads of T10 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918107) [50] |
| Data set 11 | Nanopore sequencing reads of T11 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918122) [51] |
| Data set 12 | Nanopore sequencing reads of T12 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918121) [52] |
| Data set 13 | Nanopore sequencing reads of T13 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918120) [53] |
| Data set 14 | Nanopore sequencing reads of T14 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918119) [54] |
| Data set 15 | Nanopore sequencing reads of T15 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918118) [55] |
| Data set 16 | Nanopore sequencing reads of T16 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918117) [56] |

Xu *et al. BMC Genomic Data*      (2023) 24:17

Page 4 of 5

**Table 1** (continued)

| Label | Name of data file/data set | File types (file extension) | Data repository and identifier (DOI or accession number) |
|---|---|---|---|
| Data set 17 | Nanopore sequencing reads of T17 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918116) [57] |
| Data set 18 | Nanopore sequencing reads of T18 | fastq (.fastq.gz) | NCBI Sequence Read Archive(https://identifiers.org/ncbi/insdc.sra:SRR14918115) [58] |

the transcriptome in persimmon fruit development (Table 1).

## Limitations

Here, we describe the transcriptomic profile of *D. oleifera* during different stages of fruit development. One limitation of our study is that qRT-PCR analysis should be conducted to validate the identified patterns of differential gene expression here. Long-read sequencing platforms have the capability to sequence entire cDNA molecules end-to-end, while, the accuracy of these long reads is usual lower than Illumina sequencing. So, high-accuracy and short reads obtained from Illumina sequencing should supply to offset the reduced accuracy of these long reads from nanopore sequencing.

## Abbreviations

| | |
|---|---|
| DAP | Days after pollination |
| ONT | Oxford Nanopore Technologies |
| rRNA | Ribosomal RNA |
| FLs | Full-length transcripts |
| FLNC | Full-length, non-chemiric |
| AS | Alternative splicing |
| APA | Alternative polyadenylation |
| CPC | Coding Potential Calculator |
| CNCI | Coding-Non-Coding Index |
| CPAT | Coding Potential Assessment Tool |
| NR | Non-redundant protein sequence database |
| Pfam | The database of Homologous protein family |
| KOG | Eukaryotic Ortholog Groups |
| COG | Clusters of Orthologous Groups of proteins |
| eggNOG | Evolutionary genealogy of genes: Non-supervised Orthologous Groups |
| Swiss-Prot | A manually annotated, non-redundant protein sequence database |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| GO | Gene Ontology |
| lncRNAs | Non-coding RNAs |
| CPM | Counts per million |
| FL reads | Full-length reads |
| IR | Intron retention |
| ES | Exon skipping |
| DEGs | Differentially expressed genes |
| DETs | Differentially expressed transcripts |

## Authors' contributions
Y.X. processed and analysed data. Y.X. and C.Y.L wrote the draf manuscript. C.Y.L and W.Q.C. performed library preparation and assisted in drafing the manuscript. K.Y.W. processed the samples. B.C.G. designed and supervised the project. The author(s) read and approved the final manuscript.

## Availability of data and materials
Data described in this Data note can be freely and openly accessed on NCBI under Bioproject ID PRJNA736836, accession number SRR14918107-SRR14918124. Please see The details were showed in Table 1 and references [29–58].

## Declarations

### Ethics approval and consent to participate
Not applicable.

### Consent for publication
Not applicable.

### Competing interests
The authors declare no competing interests.

## References
1. Luo ZR, Wang RZ. Persimmon in China: domestication and traditional utilizations of genetic resources. Adv Hortic Sci. 2008;22:239–43.
2. Zhuang DH, Kitajima A, Ishida M, Sobajima Y. Chromosome numbers of Diospyros kaki cultivars. J Jpn Soc Hort Sci. 1990;59:289–97.
3. Wang RZ, Yang Y, Li GC. Chinese persimmon germplasm resources. Acta Hortic. 1997;436:43–50. https://doi.org/10.17660/ActaHortic.1997.436.3.
4. Kanzaki S, Nara NJ. The origin and cultivar development of Japanese persimmon (Diospyros kaki Thunb.). J Jpn Soc Food Sci Technol. 2016;63:328–30. https://doi.org/10.3136/nskkk.63.328.
5. Fu J, Liu H, Hu J, Liang Y, Liang J, Wuyun T, Tan X. Five complete chloroplast genome sequences from diospyros: genome organization and comparative analysis. PLoS ONE. 2016;11(7): e0159566.
6. Zhu QG, Xu Y, Yang Y, Guan CF, Zhang QY, Huang JW, Grierson D, Chen KS, Gong BC, Yin XR. The persimmon (Diospyros oleifera Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution. Hortic Res. 2019;6:138. https://doi.org/10.1038/s41438-019-0227-2.
7. Suo Y, Sun P, Cheng H, Han W, Diao S, Li H, Mai Y, Zhao X, Li F, Fu J. A high-quality chromosomal genome assembly of Diospyros oleifera Cheng. Gigascience. 2020;9(1):p.giz164. https://doi.org/10.1093/gigascience/giz164.

Xu *et al. BMC Genomic Data*        (2023) 24:17

Page 5 of 5

8.    Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni JJ. Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. Plant Cell. 2005;17(11):2954–65. https://doi.org/10.2307/3872422.

9.    Yu K, Xu Q, Da X, Guo F, Ding Y, Deng X. Transcriptome changes during fruit development and ripening of sweet orange (Citrus sinensis). BMC Genomics. 2012;13:10. https://doi.org/10.1186/1471-2164-13-10.

10.   Zhang S, Shi Q, Albrecht U, Shatters RG Jr, Stange R, McCollum G, Zhang S, Fan C, Stover E. Comparative transcriptome analysis during early fruit development between three seedy citrus genotypes and their seedless mutants. Hortic Res. 2017;4:17041. https://doi.org/10.1038/hortres.2017.41.

11.   Yu X, Yu K, Chen B, Liao Z, Huang W. Nanopore long-read RNAseq reveals regulatory mechanisms of thermally variable reef environments promoting heat tolerance of scleractinian coral Pocillopora damicornis. Environ Res. 2021;195(8):110782. https://doi.org/10.1016/j.envres.2021.110782.

12.   Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191.

13.   Foissac S, Sammeth M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res. 2007;35(Web Server issue):W297-299. https://doi.org/10.1093/nar/gkm311.

14.   Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Hur AB, Reddy ASN. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. 2016;7:1–11. https://doi.org/10.1038/ncomms11706.

15.   Haas B, Papanicolaou AJGS: TransDecoder (find coding regions within transcripts). Google Scholar https://github.com/TransDecoder/TransDecoder/wiki (2016).

16.   Kong L. ZhangY, Ye ZQ, Liu XQ, Gao G: CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(S2):345–9. https://doi.org/10.1093/nar/gkm391.

17.   Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long noncoding transcripts. Nucleic Acids Res. 2013;41(17):e166. https://doi.org/10.1093/nar/gkt646.

18.   Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41:74. https://doi.org/10.1093/nar/gkt006.

19.   Finn RD, Bateman AA, Clements J, Coggill P, Ruth Y, Sean ER, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30. https://doi.org/10.1093/nar/gkt1223.

20.   Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, Cui Q. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Brief Bioinform. 2015;16(5):806–12. https://doi.org/10.1093/bib/bbu048.

21.   Deng YY, Li JQ, Wu SF. ZhuY, Chen Y, Fuchu HE: Integrated nr database in protein annotation system and its localization. Comput Eng. 2006;32:71–4. https://doi.org/10.1109/INFOCOM.2006.241.

22.   Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 2004;5(2):R7. https://doi.org/10.1186/gb-2004-5-2-r7.

23.   Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33–6. https://doi.org/10.1093/nar/28.1.33.

24.   Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14. https://doi.org/10.1093/nar/gky1085.

25.   Soudy M, Anwar AM, Ahmed EA, Osama A, Ezzeldin S, Mahgoub S, Magdeldin S. UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). J Proteomics. 2020;213:103613. https://doi.org/10.1016/j.jprot.2019.103613.

26.   Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32(Database issue):D277-280. https://doi.org/10.1093/nar/gkh063.

27.   Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat Genet. 2000;25(1):25–9. https://doi.org/10.1038/75556.

28.   Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. https://doi.org/10.1038/npre.2010.4282.2.

29.   Data file 1: Summary of sequencing sample and strategies in this study. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314470

30.   Data file 2: Statistic of ONT-sequencing in this study. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314515 .

31.   Data file 3: Read number and length distribution of FLNC and Collapse transcripts after ONT-Seq analysis. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314524 .

32.   Data file 4: Gene information and database annotations. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314536 .

33.   Data file 5: The total number of AS events in detected genes and transcripts. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314539 .

34.   Data file 6: The characteristics of AS events in each sample. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314545 .

35.   Data file 7: The statistical lists of APA events for each sample. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314548 .

36.   Data file 8: The statistical list of all fusion gene for each sample. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314563 .

37.   Data file 9: The result of LncRNAs classifications. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314566 .

38.   Data file 10: The information of target genes of these 933 lncRNAs. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314569 .

39.   Data file 11: The quantitative gene expression of all DEGs. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314572 .

40.   Data file 12: The quantitative gene expression of all DETs. (2022). Figshare. https://doi.org/10.6084/m9.figshare.19314584 .

41.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918124 .

42.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918123 .

43.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918114 .

44.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918113 .

45.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918112 .

46.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918111 .

47.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918110 .

48.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918109 .

49.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918108 .

50.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918107 .

51.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918122 .

52.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918121 .

53.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918120 .

54.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918119 .

55.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918118 .

56.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918117 .

57.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918116 .

58.   NCBI Sequence Read Archive. (2021). https://identifiers.org/ncbi/insdc.sra:SRR14918115 .

## Publisher's Note

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