SCIENTIFIC DATA

DATA DESCRIPTOR

Global tissue-specific transcriptome analysis of *Citrus sinensis* fruit across six developmental stages

Guizhi Feng, Juxun Wu & Hualin Yi

Citrus sinensis fruit is a type of nonclimacteric fruit that mainly consists of four tissues: the epicarp, albedo, segment membrane and juice sac. The fruit quality is determined by the characteristics of these four tissues. However, our knowledge of the molecular processes that occur in these four tissues during citrus fruit development and ripening is limited. Tissue-specific transcriptomes provide a comprehensive and detailed molecular regulatory network of citrus fruit development and ripening. In our study, we collected four types of tissue from *C. sinensis* fruits at six developmental stages. A total of 72 libraries were constructed from 24 samples (each sample had three replicates), and the transcriptomes were sequenced by an Illumina HiSeq 4000. The comprehensive analyses of the transcriptomes from the four tissues and six developmental stages presented here provide a valuable resource for the discovery of the molecular networks underlying citrus fruit development and ripening.

Background & Summary

Citrus, a nonclimacteric fruit, is widely cultivated worldwide. Citrus is mainly composed of the inedible epicarp (EP) and albedo (AL) and the edible segment membrane (SM) and juice sac (JS). The development and ripening of citrus fruit is a complex and sophisticated regulatory process that can be divided into three stages: cell division stage; expansion stage involving cell enlargement and water, sugar accumulation; and ripening stage[1]. At present, most studies investigating citrus fruit development and ripening are based on the organ-wide or mixed-tissue level, which inevitably obscures many tissue-specific phenomena. Therefore, a systematic study of the different tissues and periods within the same citrus fruit can help us to more clearly understand the developmental characteristics of different tissues and their interrelationships.

The analysis of tissue-specific expression profiles avoids the potential dilution of location-specific regulation and has been proven to effectively reveal otherwise undetectable biological pathways and regulatory networks[2]. In recent years, tissue-specific transcriptomes have been used in many types of plants. Through the transcriptome analysis of multiple tissues in young tomato plants, the unique developmental regulation of tomato seed was discovered[3]. The association between the transcriptome of 23 tissues of cucumber and volatile organic compounds in cucumber was screened to select a group of candidate genes that may be involved in the synthesis of volatile substances in cucumber[4]. Through the transcriptome analysis of different structures of maize seed, the key role of MRP-1 (MYB-Related Protein-1) in maize endosperm development was found[5]. By analyzing the transcriptome data of five developmental stages and five tissues of strawberry, it was found that the role of the endosperm and seed coat in auxin and gibberellin biosynthesis for fruit set is more prominent than that of the embryo[6]. The mechanisms underlying fruit development and ripening are unique among different fruit types. A comprehensive transcriptome profile of the tissue types and stage types can reveal the enormous diversity in gene expression associated with tissue type and developmental stage. Anatomically, we usually divide citrus into four tissues: EP, AL, SM, and JS. Different tissues play different roles in the development and maturation of citrus fruits. The development of the EP determines the appearance of the fruit, and the metabolism of carotenoids plays an important role in this process[7,8]; the development of the AL determines how easy it is to peel the fruit, the hardness of the fruit, etc. SM directly determines the chewiness of the fruit, and cell wall metabolism plays an important role in this process[9]; the JS stores large amounts of sugars and organic acids. Sucrose enters the JS after unloading at the SM phloem[10,11]. A large amount of organic acid is accumulated in the JS, 90% of which is citric acid[12]. Currently, our understanding of the underlying molecular mechanisms involved in these processes is limited.

Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan, 430070, China. Correspondence and requests for materials should be addressed to J.W. (email: wjxun@mail.hzau.edu.cn)
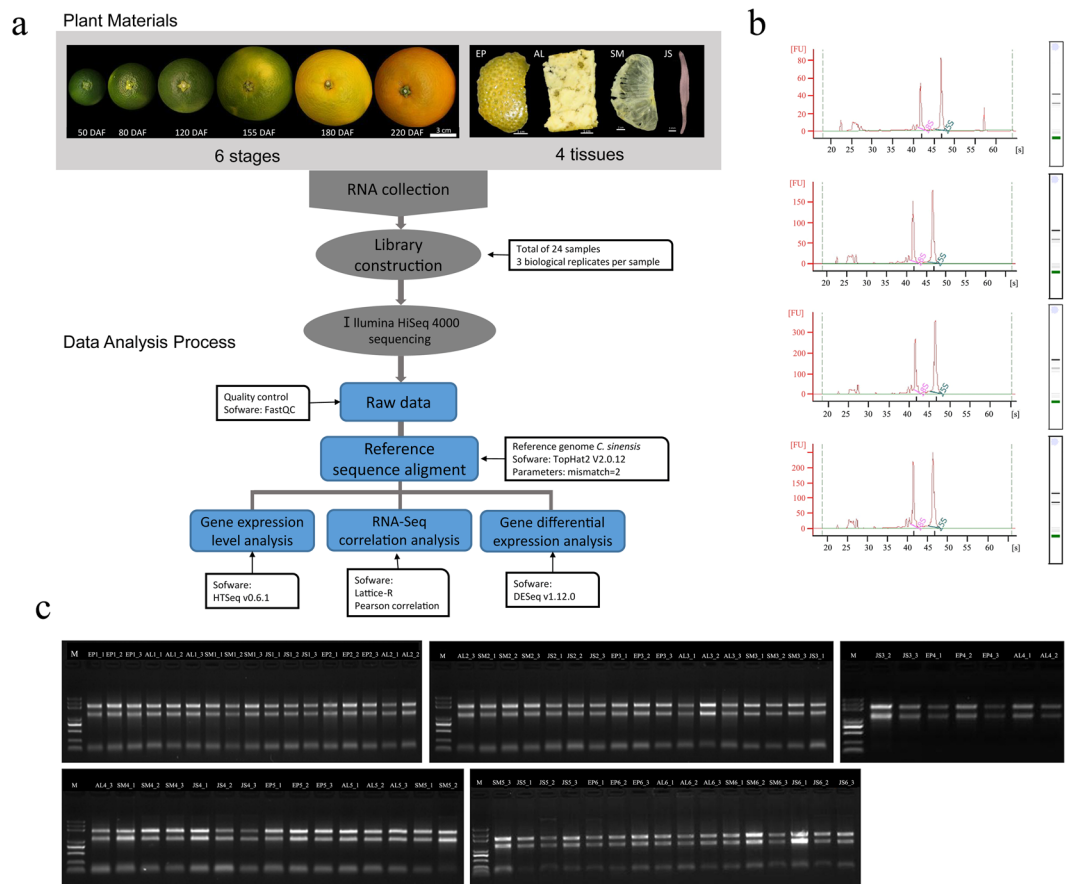
Fig. 1 Study work flow from sample preparation through sequence processing. (**a**) Schematic overview of the study. (**b**) Presequencing RNA quality control check with a Bioanalyzer 2100 system, and example results for EP, AL, SM, and JS samples show good quality RNA for library preparation. (**c**) RNA from 72 tissues analyzed by agarose gel electrophoresis. M, Marker DL2000. In all figures that follow, the samples are named "Tissue_stage_replicate" or "Tissue_stage". EP_1_1 means replicate 1 of the epicarp at stage 1. AL_2 means the albedo sample at stage 2.

In this study, we performed transcriptome sequencing of four tissues (EP, AL, SM and JS) throughout the development and ripening stages (50, 80, 120, 155, 180 and 220 days after flowering (DAF)) of the navel orange fruit. After we removed low-quality sequencing data, we obtained a total of 549 Gb (G bases) of transcriptome data from the collected samples (no less than 6 Gb per sample). We describe how we collected and processed samples, extracted mRNA, built a cDNA library, controlled the quality of transcriptome data, aligned the data to a reference genome, and performed correlation analysis between samples. Finally, we describe how we obtained reliable data for use in subsequent analysis and future research. All the experimental processes involved in the paper are shown in Fig. 1a.

Methods

Overview of experimental design. Fruit samples of the 'Fengjie 72-1' navel orange (*C. sinensis* L. Osbeck) were harvested after the second physiological fruit-falling period (50 DAF), expansion period (80, 120 and 155 DAF), coloring period (180 DAF) and full-ripening period (220 DAF). Four fruit tissues (EP, AL, SM and JS) were manually dissected at every stage, rapidly frozen in liquid nitrogen and kept at −80 °C. The experimental design and analysis pipeline are shown in Fig. 1a. After the extraction of the total RNA, a total of 72 libraries were constructed from 24 samples (each sample had three replicates). A total of 72 transcriptome profiles were obtained by RNA-seq using an Illumina HiSeq[TM] 4000 sequencing platform. Subsequently, the clean reads filtered from the raw reads were mapped to the reference genome of *C. sinensis*[13,14] (http://citrus.hzau.edu.cn/orange/index.php). The gene expression analyses were performed with HTSeq[15], and the differential expression analyses were performed with DESeq2[16,17].

Sample collection. We selected a total of 9 trees (3 trees as one biological replicate) to collect samples. These nine trees were grafted on the same rootstock *Poncirus trifoliata* (L.) Raf and cultivated in the same orchard (Fengjie, Chongqing City, China). At each stage, four fruits were sampled from each tree, and a total of 12 fruits were mixed as one biological replicate for each sample. After the fruit was sampled from the tree, the EP was quickly cleaned with distilled water and dried with a sterile gauze. Subsequently, the EP was gently scraped with
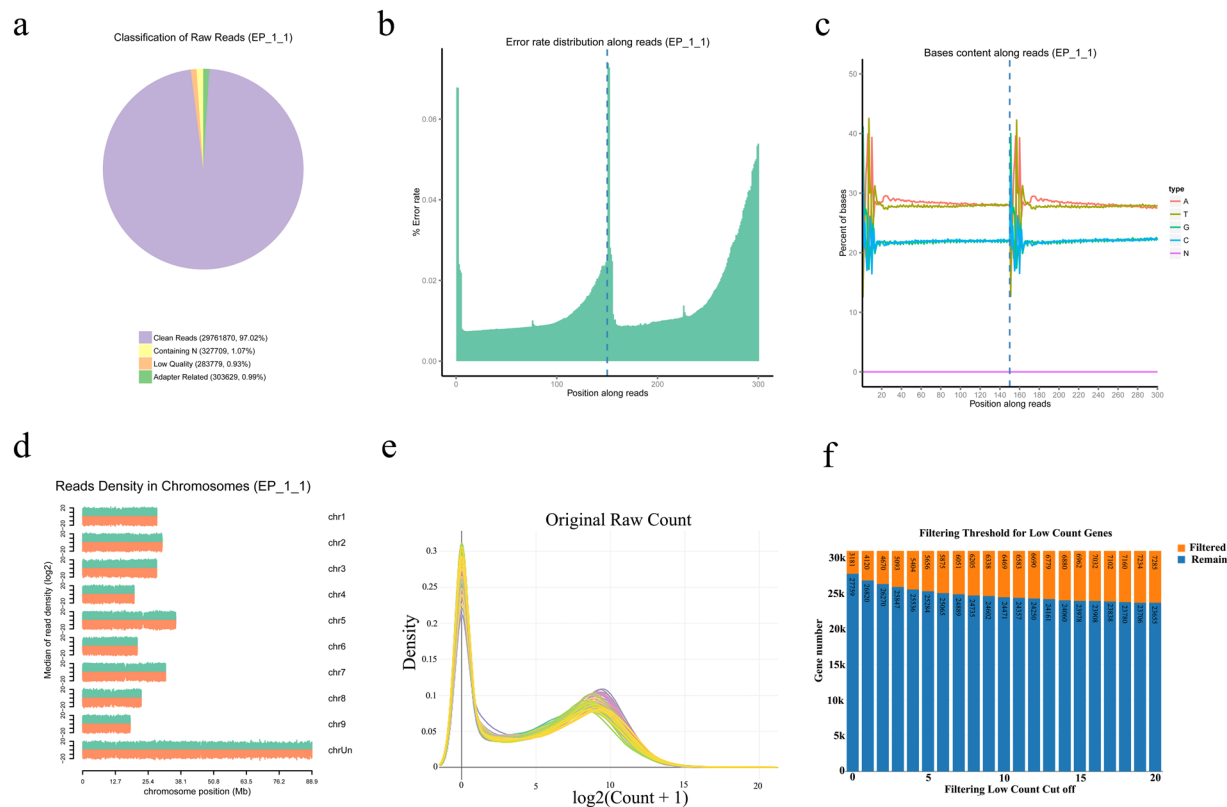
Fig. 2 Global assessment of transcriptome data. Example results for (**a**) sequencing error rate distribution, (**b**) A/T/G/C content distribution, and (**c**) composition of raw data, and (**d**) the density distribution of reads on the chromosome, which show the good quality of the transcriptome libraries. (**e**) Distribution density of original raw counts. The x axis is $\log_2$ (count $+1$), and the y axis is the density. (**f**) Filtered genes with a total read count below the thresholds. Max threshold $= 20$.

a scalpel. Then, the residual EP was removed, and the AL was separated. This step ensured that there were no residual colored substances or internal tissues on the AL. Finally, the SM and JS were separated. The JS was gently scraped from the SM, and the SM was rinsed with distilled water to remove the residual JS liquid. All of the above anatomical and tissue acquisition steps were rapidly performed in a low temperature environment created by crushed ice, and the separated tissues were quickly treated with liquid nitrogen and stored at $-80\,°C$.

RNA extraction, library construction, and RNA sequencing. In total, 72 materials (3 biological replicates per sample) were used to extract total RNA, as previously described[18]. In brief, the RNA extraction protocol includes two parts: rough extraction and purification extraction. Rough extraction: approximately 0.5 g sample powder ground in liquid nitrogen is transferred to a 10 ml centrifuge tube with 5 ml Trizol Buffer; add 3 ml chloroform, shake 30 s, then stand at room temperature for 10 min; subsequently, add an equal volume of pre-cooled isopropanol, mix gently and stand for 10 min; discard the supernatant and soak in 3 ml of pre-cooled 75% ethanol for 1 h; next, centrifuge at 12000 rpm for 4 min at 4 °C, then discard the ethanol and dry it in air. Purification extraction: add 800 ul TESAR into the 10 ml centrifuge tube with RNA precipitate and vortex it; then, add 800 ul AQ/CTAB and 800 ul Bu/CTAB and vortex for 15 min; next, transfer it into two 1.5 ml centrifuge tubes and centrifuge at room temperature for 6 min; the supernatant is transferred to a 1.5 ml centrifuge tube, and add 350 ul NaCl (0.2 mol/L), vortex for 1 min, centrifuge at room temperature for 15 min; after carefully pipette the supernatant into a 1.5 ml centrifuge tube, add 50 ul NaAC (3 mol/L) and 1 ml pre-chilled ethanol, mix gently, and freeze at $-20\,°C$ for 1 h; at the end, centrifuge at 13200 rpm for 30 min at 4 °C, discard the supernatant, and dissolve it in 40 ul DEPC water. The quality of RNA was detected by agarose gels (Fig. 1c). RNA purity (OD260/280 ratio) was checked using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA); RNA concentration was measured using a Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA). RNA integrity was assessed using a RNA Nano 6000 Assay Kit on a Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Total RNA with RIN ≥ 6.5 was used for cDNA library construction (Fig. 1b and Online-only Table 1).

A total of 3 μg of RNA per sample was used as the input material to construct the cDNA library. The sequencing libraries generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) were added to the attribute sequence for each sample according to the manufacturer's recommendations[19,20]. Briefly, mRNA was purified from total RNA using magnetic beads attached to a poly-T oligo and cleaved using a divalent cation in the NEBNext strand synthesis reaction buffer (5X) in the a first elevated temperature. First-strand cDNA
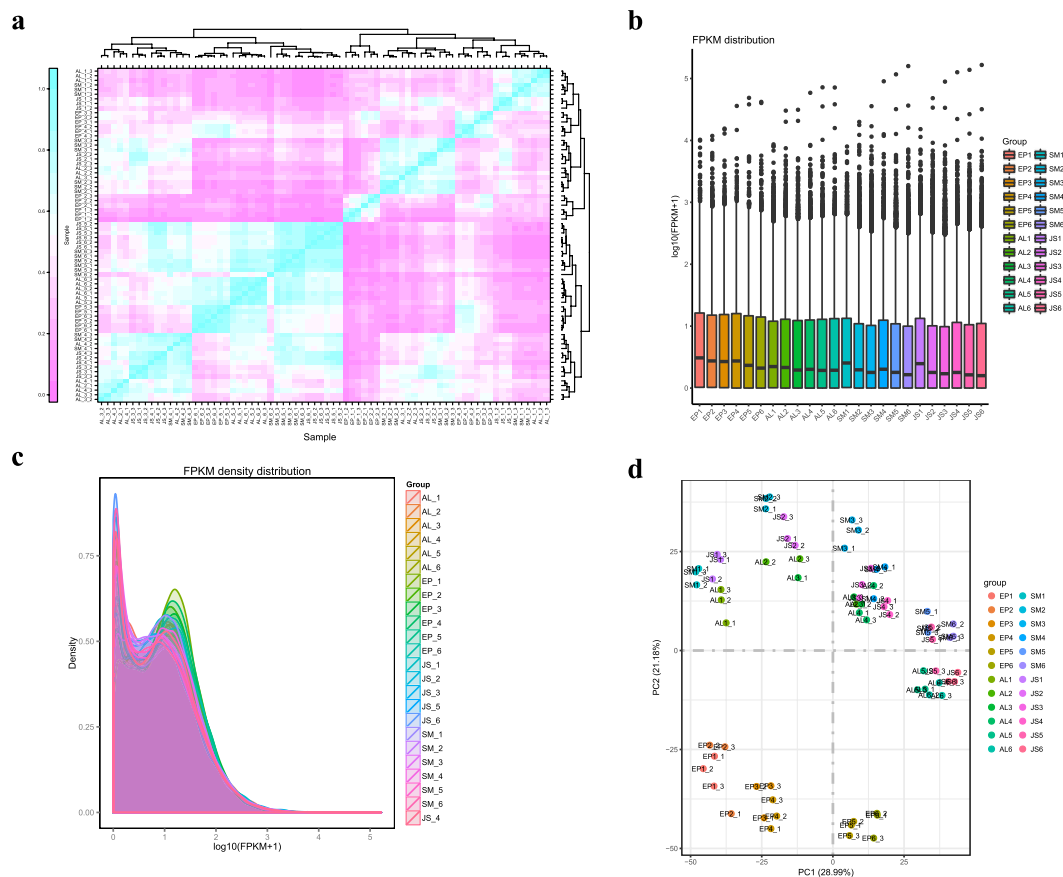
Fig. 3 Analysis of global gene expression among fruit tissues. (**a**) Biological replicate sample correlation analysis. (**b**) Boxplot shows different sample FPKM distributions. The y axis is $\log_{10}$ (FPKM +1). (**c**) Intuitive display genes expressed at different FPKM levels. The x axis is $\log_{10}$ (FPKM +1), and the y axis is the density of the gene. (**d**) Two-dimensional PCA analysis with nonmetric multidimensional scaling (NMDS).

was synthesized using random hexamer primers and M-MuLV reverse transcriptase (RNase H-). Next, second strand cDNA synthesis was performed with DNA polymerase I and RNase H. Exonuclease/endonuclease/polymerase activity joined the protruding end into blunt end. NEBNext 3′ end polyadenylation of the DNA fragments resulted in the hairpin loop structure of the connection adapter in preparation for hybridization. To select cDNA fragments of preferably 150–200 bp in length, the library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA)[21]. The cDNA ligated with the size-selected linker was then ligated with 3 µl of USER enzyme (NEB, USA) for 15 minutes at 37 °C and then at 95 °C for 5 minutes before PCR. Then, polymerase Phusion high-fidelity DNA, universal PCR primers and index (X) primer were added for PCR. Finally, PCR enrichment was performed to obtain a final cDNA library. After the library was constructed, preliminary quantification was performed using Qubit 2.0, and the library was diluted to 1 ng/µl. Then, the insert size of the library was detected using Agilent 2100. After the insert size was verified, Q-PCR was performed. The effective concentration of the library was accurately quantified (effective library concentration >2 nM) to ensure library quality.

The libraries were combined into paired pools, and double-end sequencing was performed by sequencing the cDNA product that was reverse transcribed into a small fragment of 150–200 bp using a HiSeq 4000 instrument (Illumina, San Diego, USA) at Novogene Company (Beijing, China). The raw data files obtained from high-throughput sequencing were converted to the original sequence by CASAVA Base Calling analysis[22]. The results were stored in a FASTQ file.

Preprocessing of sequencing data. The raw data obtained by sequencing was filtered by removing the reads with adapters, reads with N (N means that the base information cannot be determined) greater than 10%, and low-quality reads (Qphred <=20 bases account for more than 50% of the entire read length The reads) to get clean reads[23,24] (Fig. 2a). Then, the quality-controlled reads (clean reads) of the 72 transcriptome libraries were mapped to the reference genome of *C. sinensis*[13]. TopHat (v2.0.12)[14] was used as the mapping tool. First, the entire obtained sequence was aligned to the genomic exon, and the obtained sequence was then segmentally aligned with the two exons of the genome[24]. The statistical comparisons of the reads with the reference genome are shown in Online-only Table 1. Because the reference genome is generated from the dihaploid DNA of sweet orange and the assembled sequence only covers 87.3% of the estimated orange genome[13], the rates of uniquely mapped reads of 72 navel orange transcriptome libraries ranged from 67.74% to 76.46% are in the normal range. Samtools
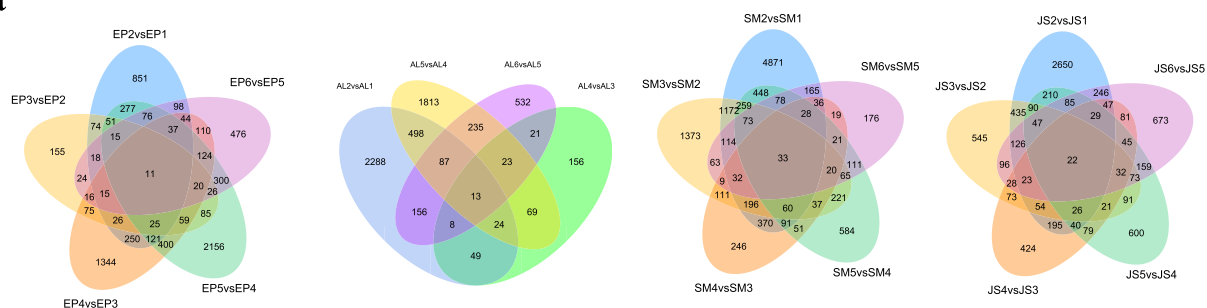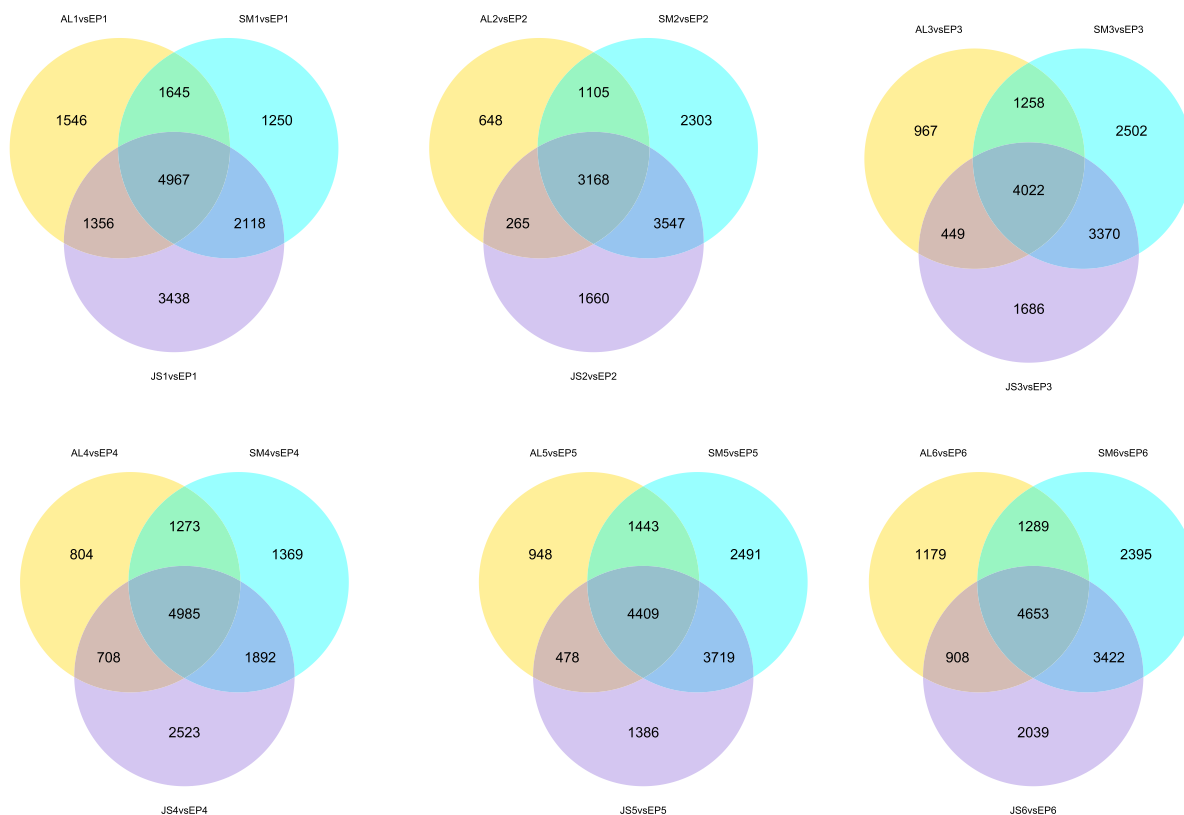
Fig. 4 Veen plot shows differentially expressed genes in different tissues at the same stage (**a**) and among different stages in the same tissue (**b**).

was used to convert the *.sam* file to a *.bam* file with the default parameters. The count matrix file was imported into DESeq2[16,17] for differential expression analysis. The Benjamini and Hochberg method was used to adjust the *P* value obtained to control the false discovery rate. The gene expression levels were calculated by the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) method using HTSeq[15] (v0.6.1).

Data Records
The RNA-Seq raw data of 72 samples were deposited in NCBI Sequence Read Archive with accession number SRP182638[25]. The file of gene expression level was deposited in NCBI Gene Expression Omnibus (GEO) with accession number GSE125726[26].

Technical Validation
Quality control. The original sequence obtained after sequencing was used for subsequent analysis after a series of quality control analyses. First, the distribution of the sequencing error rate was evaluated (Fig. 2b). The error rate of each sequenced base was determined by the Phred score (Q_{phred}) with the formula (1: $Q_{phred} = -10$ $\log_{10}(e)$), and the Phred value is passed through a base calling analysis. The probability model showed that this model can accurately predict the error rate of base discrimination. Subsequently, the distribution of A/T/G/C
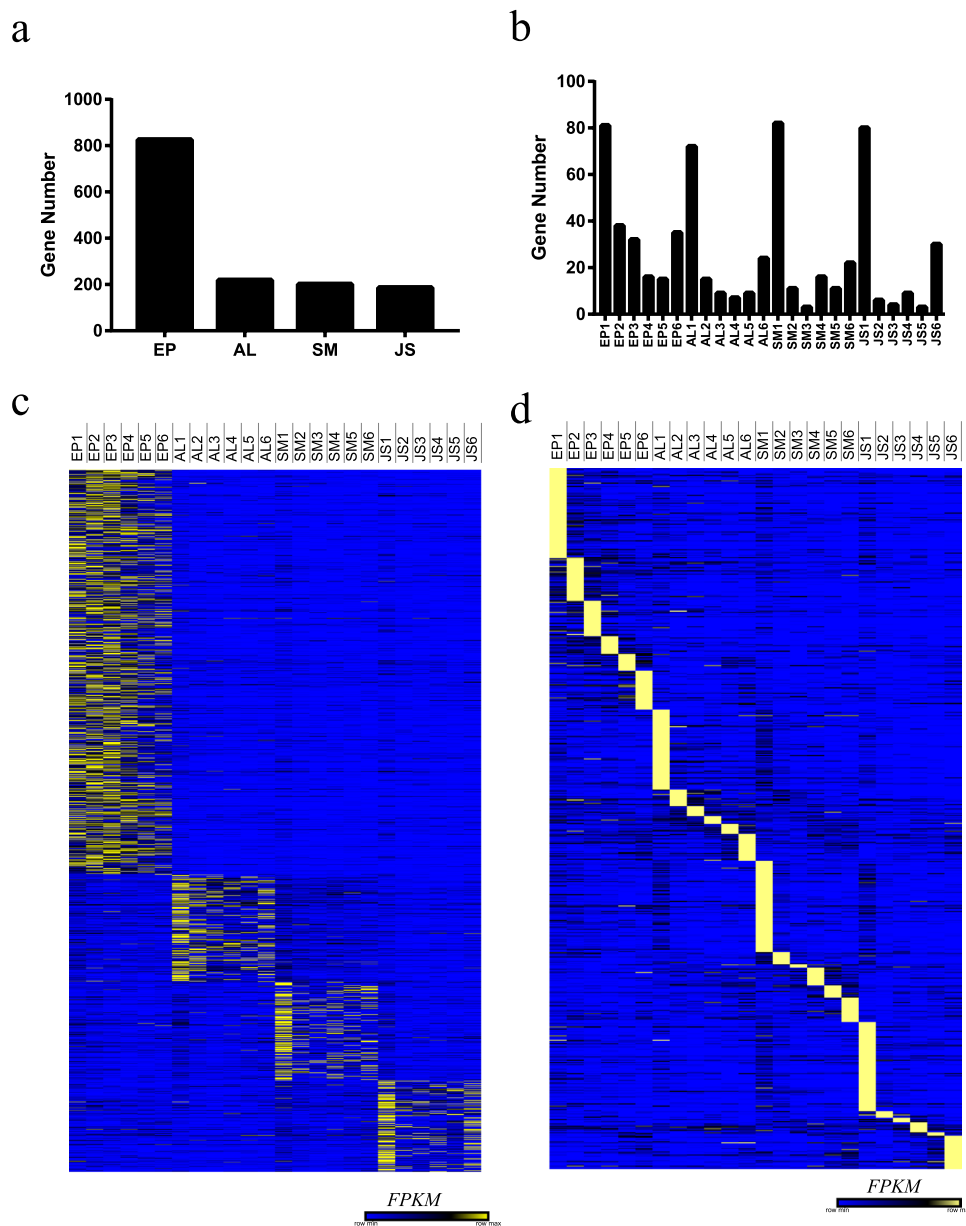
Fig. 5 Screening of the spatiotemporal expression genes. (**a**) The number of tissue-specific genes. (**b**) The number of sample-specific genes. (**c**) Heat-map analysis of tissue-specific genes. (**d**) Heat-map analysis of sample-specific genes.

content was inspected to detect the presence or absence of AT and GC separation (Fig. 2c). At the same time, the Q20, Q30 and GC contents of the clean data were calculated (Online-only Table 1). The clean data obtained by filtering the raw data and the clean reads were compared to the chromosome (Fig. 2d).

The raw count was homogenized for all samples by $\log_2$ (count $+1$), and distribution density statistics were performed (Fig. 2e). The raw counts of all samples were screened with a maximum screening threshold of 20, and the number of genes under different screening thresholds was counted[27,28] (Fig. 2f). Additionally, the count matrix was clustered to generate clustered heat map with the lattice-R package, and the correlation coefficient was calculated with Pearson correlation. The results showed a good correlation between the three biological replicates of the different samples (Fig. 3a).

Analysis of RNA-seq data. First, we calculated the FPKM distribution density of all samples (Fig. 3b,c). Additionally, the FPKM value of the gene in the sample was used as the input data for nonmetric multidimensional scaling (NMDS)[29] to reduce and realign the data in a visual low-dimensional space, which was maximized in a plane scatter plot (Fig. 3d). The x-axis reflects the extent of the relationship between the biological replicates. Moreover, we found that the EP was more independent than the other three tissues, and the differences between

tissues were greater than those between developmental period, while AL, SM, and JS showed large differences between periods. The expression levels of individual genes were submitted to NCBI GEO[26].

Genes with adjusted P values < 0.05 as determined by DESeq2[16,17] were designated as differentially expressed genes[17,30,31]. The differentially expressed genes in different tissues and different stages during the development of navel orange were screened. The results of spatiotemporal differences in the genes are presented in the form of Venn diagram[32] (Fig. 4a,b). At different levels, approximately 100 to 5000 differentially expressed genes were screened in the different tissues (Fig. 4a). In the EP, the number of differentially expressed genes identified in the EP5 vs. EP4 comparison was the highest. In contrast, AL, SM and JS appeared to have the largest number of differentially expressed genes between Stage 2 and Stage 1. In particular, we did not screen for differentially expressed genes in AL3 vs. AL2 (Fig. 4a). Furthermore, at the tissue level, we compared the AL, SM, and JS at different developmental stages with EP (EP as a control), and different comparison combinations identified approximately 1000 to 3000 differential genes (Fig. 4b). From the distribution of the number of differential genes, the overall comparison with the EP was JS > SM > AL, which indicates that the spatial distance of the tissue affects the difference in the gene expression of the fruit. A larger spatial distance is correlated with a more obvious the difference between tissues. Finally, we also screened for tissue-specific genes and sample-specific genes. We identified 825 differentially expressed genes in the EP, which is much more than the 219, 201, and 186 genes identified for AL, SM, and JS, respectively (Fig. 5a,b), which further illustrates the specific function of the EP. In addition, the maximum number of all four tissue sample-specific genes was observed in Stage 1 (Fig. 5c,d).

References

1. Bain, J. M. Morphological, anatomical, and physiological changes in the developing fruit of the Valencia orange, Citrus sinensis (L) Osbeck. *Aust. J. Bot.* **6**, 1–23 (1958).
2. Han, Y., Gao, S., Muegge, K., Zhang, W. & Zhou, B. Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights.* **9**, 29–46, https://doi.org/10.4137/bbi.s28991 (2015).
3. Pattison, R. J. *et al.* Comprehensive Tissue-Specific Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit Development. *Plant Physio.* **168**, 1684–U1002, https://doi.org/10.1104/pp.15.00287 (2015).
4. Wei, G. *et al.* Integrative Analyses of Nontargeted Volatile Profiling and Transcriptome Data Provide Molecular Insight into VOC Diversity in Cucumber Plants (Cucumis sativus). *Plant Physiol.* **172**, 603–618 (2016).
5. Zhan, J. P. *et al.* RNA Sequencing of Laser-Capture Microdissected Compartments of the Maize Kernel Identifies Regulatory Modules Associated with Endosperm Cell Differentiation. *Plant Cell* **27**, 513–531, https://doi.org/10.1105/tpc.114.135657 (2015).
6. Kang, C. *et al.* Genome-Scale Transcriptomic Insights into Early-Stage Fruit Development in Woodland Strawberry Fragaria vesca. *Plant Cell* **25**, 1960–1978, https://doi.org/10.1105/tpc.113.111732 (2013).
7. Ma, G. *et al.* Enzymatic formation of β-citraurin from β-cryptoxanthin and Zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. *Plant Physio.* **163**, 682–695 (2013).
8. Ma, G. *et al.* Accumulation of carotenoids in a novel citrus cultivar 'Seinannohikari' during the fruit maturation. *Plant Physiol. Biochem.* **129**, 349–356, https://doi.org/10.1016/j.plaphy.2018.06.015 (2018).
9. Lei, Y. *et al.* Comparison of cell wall metabolism in the pulp of three cultivars of 'Nanfeng' tangerine differing in mastication trait. *J Sci Food Agr.* **92**, 496–502 (2012).
10. Koch, K. E. The path of photosynthate translocation into citrus fruit. *Plant Cell Environ.* **7**, 647–653 (1984).
11. Koch, K. E. & Avigne, W. T. Postphloem, nonvascular transfer in citrus: kinetics, metabolism, and sugar gradients. *Plant Physiol.* **93**, 1405–1416 (1990).
12. Huang, D., Zhao, Y., Cao, M., Qiao, L. & Zheng, Z.-L. Integrated Systems Biology Analysis of Transcriptomes Reveals Candidate Genes for Acidity Control in Developing Fruits of Sweet Orange (Citrus sinensis L. Osbeck). *Front. Plant Sci.* **7**, https://doi.org/10.3389/fpls.2016.00486 (2016).
13. Xu, Q. *et al.* The draft genome of sweet orange (Citrus sinensis). *Nat Genet.* **45**, 59–66 (2013).
14. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, https://doi.org/10.1186/gb-2013-14-4-r36 (2013).
15. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, https://doi.org/10.1093/bioinformatics/btu638 (2015).
16. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138, https://doi.org/10.1093/bioinformatics/btp612 (2010).
17. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, https://doi.org/10.1186/s13059-014-0550-8 (2014).
18. Liu, Y., Liu, Q., Tao, N. & Deng, X. Efficient Isolation of RNA from Fruit Peel and Pulp of Ripening Navel Orange (Citrus sinensis Osbeck). *Journal of Huazhong Agricultural University* **25**, 300–304 (2006).
19. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
20. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, https://doi.org/10.1093/bioinformatics/bts635 (2013).
21. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303, https://doi.org/10.1101/gr.107524.110 (2010).
22. dataSiretskiy, A., Sundqvist, T., Voznesenskiy, M. & Spjuth, O. A quantitative assessment of the Hadoop framework for analyzing massively parallel DNA sequencing data. *Gigascience* **4**, https://doi.org/10.1186/s13742-015-0058-5 (2015).
23. Chen, C., Khaleel, S. S., Huang, H. & Wu, C. H. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source code for biology and medicine* **9**, 8–8, https://doi.org/10.1186/1751-0473-9-8 (2014).
24. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, https://doi.org/10.1186/gb-2009-10-3-r25 (2009).
25. *NCBI Sequence Read Archive*, https://identifiers.org/ncbi/insdc.sra:SRP182638 (2019).
26. Wu, J., Feng, G. & Yi, H. High-Spatiotemporal-Resolution Transcriptomes Insights into Fruit Development and Ripening in Citrus sinensis. *Gene Expression Omnibus*, https://identifiers.org/geo:GSE125726 (2019).
27. Sun, J., Nishiyama, T., Shimizu, K. & Kadota, K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **14**, https://doi.org/10.1186/1471-2105-14-219 (2013).
28. Hardcastle, T. J. & Kelly, K. A. BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**, https://doi.org/10.1186/1471-2105-11-422 (2010).
29. Gao, Y. *et al.* Vertical and horizontal assemblage patterns of bacterial communities in a eutrophic river receiving domestic wastewater in southeast China. *Environ Pollut.* **230**, 469–478 (2017).

30. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).
31. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, https://doi.org/10.1186/1471-2105-11-94 (2010).
32. Hulsen, T., de Vlieq, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics* **9**, 488 (2008).

Acknowledgements

Author Contributions

H.Y., J.W. conceived the project and designed the experiment. G.F. conducted the experiment. G.F., J.W. performed the data analyses. G.F. wrote the manuscript. All authors read the final manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

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