Transcriptome Regulation of Carotenoids in Five Flesh-Colored Watermelon (Citrullus lanatus)

Pingli Yuan
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Muhammad Jawad Umer
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Nan He
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Shengjie Zhao
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Xuqiang Lu
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Hongju Zhu
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Chengsheng Gong
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Weinan Diao
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Haileslassie Gebremeskel
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
Hanhui Kuang
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute
liu wenge (liuwenge@caas.cn)
Chinese Academy of Agricultural Sciences Zhengzhou Fruit Research Institute https://orcid.org/0000-0003-2435-6370

Research article

Keywords: Citrullus lanatus, Flesh color, Transcriptional regulation, Transcription factors, WGCNA

Posted Date: July 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-42816/v1

Abstract

Background: Fruit flesh color in watermelon (*Citrullus lanatus*) is a great index for evaluation of the appearance quality and a key contributor influencing consumers preferences, but the molecular mechanisms of this intricate trait remain largely unknown. Here, the carotenoids and transcriptome dynamics during fruit development in watermelon cultivars with 5 different flesh colors were analyzed.

Results: A total of 13 carotenoids and 16,781 differentially expressed genes (DEGs) including 1,295 transcription factors (TFs) were detected during the development of five watermelon genotypes. A number of structural genes and transcription factors were found to be involved in the carotenoid biosynthesis pathway using comparative transcriptome analysis. Furthermore, we performed weighted gene co-expression network analysis and predicted hub genes in 6 main modules determining carotenoids contents. *Cla018406* (a Chaperone protein dnaJ-like protein) maybe a candidate gene for β-carotene and highly expressed in orange flesh colored fruit. *Cla007686* (a zinc finger CCCH domain-containing protein) was highly expressed in the red color watermelon, maybe a key regulator for lycopene accumulation. *Cla003760* (merbrane protein) and *Cla007686* (photosystenI reaction center subunit II) are predicted to be hub genes and play an important role in yellow flesh color formation.

Conclusions: These results provide an important resource for dissecting the molecular basis and candidate genes governing flesh color formation in watermelon fruit.

Background

Watermelon (*Citrullus lanatus*) is originally cultivated in Africa, belongs to the Cucurbitaceae family. Now, watermelon has become one of the top 5 freshly consumed fruits, with China at the top in production and consumption of watermelon worldwide. Watermelon flesh contains many nutrients, such as lycopene, citrulline, and other health-promoting compounds related to human diet[1], carotenoids that are involved in antioxidation, quenching free radicals, and anti-cancer, are necessary for human life and health. Lycopene has been reported to be involved in the prevention of cancers and cardiovascular diseases. Carotenoids are converted to Vitamin A, which plays an essential role in vision[2].

In plants, carotenoids are mainly involved in photosynthesis, light-harvest, and photoprotection[3]. Carotenoids are also the essential precursors of phytohormones (abscisic acid and strigolactones), which are key regulators for plant development and stress responses[4]. Moreover, apocarotenoids, which are carotenoid oxidative and enzymatic cleavage derivatives, are crucial for plant photosynthesis and photoprotection and regulates plant growth and development[5, 6], apocarotenoids also serve as signaling molecules and contribute to the flavor and aroma of flower petals or fruits[7].

The cultivated watermelons have the ability to synthesize various kinds of carotenoids in fruit, responsible for the vivid flesh colors, including white, yellow, orange, pink, red, and mixed color[8]. Watermelon is a good model plant for studying the regulatory mechanisms and biosynthesis of carotenoids in fleshy fruit owing to variously colored flesh. Lycopene is the main pigment in red flesh

color watermelons[9], while, xanthophylls (zeaxanthin and its derivatives, Neoxanthin, violaxanthin, and neochrome) were the main pigments in yellow-fleshed watermelons[10]. β-carotene, ζ-carotene, prolycopene are main pigments in orange-fleshed watermelons[11].

Some researches focus on the inheritance of flesh color in watermelon. The canary yellow(C) is dominant to red/pink/orange(c), the white flesh (*Wf*) is epistatic to the yellow flesh trait[12]. The *py* gene generated pale yellow flesh[13]. Scarlet red flesh, Y^{scr}, is dominant to the coral red flesh color[8]. Some quantitative trait loci (QTLs) associated with flesh color in watermelon have been reported previously. Initially, two QTLs related to red flesh color were identified on LG II and VIII using an integrated genetic linkage map[14]. Bang's research found a Clcyb.600 marker was perfectly co-segregated with flesh color phenotypes[13]. QTL related to the lycopene content and red flesh color was reported on chromosome 4 in a genetic population derived from red and pale yellow flesh colors by Liu[15]. The locus Y^{scr} was first mapped for the scarlet red flesh color on chromosome 6 using a segregated population derived from scarlet- and coral- red flesh varieties[16]. The QTL associated with β-carotene accumulation in watermelon fruit was mapped on chromosome 1[17]. The elevated chromoplast-localized phosphate transporter *ClPHT4;2* expression level is necessary for carotenoid accumulation and flesh color formation[18]. Recently, The *ClLCYB* gene contributes to the red flesh color through decrease in its protein level instead of transcript level[19].

Few researches focused on the relationship of lycopene contents and transcription expression of lycopene metabolic genes spanning the period from young to mature fruits in watermelon [20, 21]. Comparative transcriptome analysis of red vs pale-yellow watermelons had been published by Zhu [22]. However, the comprehensive molecular mechanisms underlying flesh color formation in various watermelon cultivars remain ambiguous, and no regulators linked to watermelon fruit flesh color have been reported yet on the basis of co-expression network analysis. Here, for the first time we performed a comparative transcriptome combined with co-expression network analysis using flesh at different development stages in five flesh-colored cultivated watermelon, the data set provides a comprehensive view on the dynamic gene expression networks and their potential roles in controlling flesh color. Using pair-wise comparisons and weighted gene coexpression network analysis (WGCNA), we identified modules of co-expressed genes and candidate hub genes for each carotenoid. This work provides important insights into the molecular networks underlying watermelon flesh color formation.

Results

Flesh Color Assessment and Carotenoid Content Variation During Watermelon Fruit Development of Five Genotypes

Watermelon flesh features at different developmental stages have been shown in Fig. 1, we further determined the color space values to confirm flesh color variations. At the 10 DAP stage, the fruits were small, white and hard fleshed, there was no significant difference on color space parameters among

varieties at early stages (Additional file1: Table S1). At the 20 DAP stage, the flesh started to become soften with flesh colors ranging from white to pink and white to yellow, this is the key period of continuous and rapid accumulation of pigments. At the 34 DAP, the fruits were matured and characterized by fleshy water-storing, the flesh water is saturated and has the brightest colors except for white flesh cultivar. The significant differences of L*, a*, b*, and Chroma (C) can be observed in different fleshed fruits at 20 DAP and 34 DAP. The difference in flesh color appeared at the 20 DAP and become more evident at 34 DAP.

Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white flesh color varieties, respectively; DAP, days after pollination. Scale bars = 5 cm

Flesh color in watermelon fruits is determined by carotenoids composition and content. Generally, the lycopene, β-carotene, violaxanthin are responsible for the red, orange, and yellow flesh colors, respectively [10]. The major carotenoids metabolic changes in three stages with different flesh colors were measured using liquid chromatography-mass spectrometry (Additional file2: Table S2). The PCA result of the time-course and color comparative analyses suggesting the reliable metabolic data (Additional file3: Fig. S1). At 10 DAP,There were no pigments or only trace phytoene could be detected in the young fruits. At 20 DAP, lycopene, β-carotene, and violaxanthin are the major pigments in red/pink fruits, orange fruits and yellow fruit, low level carotenoids were detected for the white flesh. At 34 DAP□the fruit is ripe, the highest level of lycopene was noted in the red fleshed fruits, Orange flesh has the highest level of β-carotene, the highest violaxanthin content was noted in the yellow fleshed fruits. For the white fruits, only trace levels of violaxanthin, antheraxanthin and lutein were observed (Additional file2: Table S2, Fig. 1).

Taken together, the color space values, and carotenoid levels revealed substantial variations among red, pink, orange, yellow, and white genotypes. It is conceivable that the DEGs at the three stages among the 5 different flesh colored watermelons fruits may play important roles in determining flesh color formation.

Transcript Sequencing of Developing Watermelon Fruits with Different Flesh Colors

To explore the potential molecular mechanisms underlying the flesh coloration during the development of watermelon fruit among different genotypes, RNA-Sequencing analyses were conducted on fruit flesh to generate transcriptome profiles. Samples of fruit flesh at three critical stages (10 DAP, 20 DAP, and 34 DAP) were obtained from five genotypes (Fig. 1), all samples were analyzed in three independent biological replicates.

In total, 45 libraries were constructed and analyzed. After removing low quality reads, the average number of reads per library was over 51.9 million for 45 samples, with an average GC content of 44.15% (Additional file 4: Table S3). The RNA-Seq reads were aligned with the reference map of the watermelon (97103) genome (http://cucurbitgenomics.org/organism/1), using HISAT(version 2.0.4) [23]. More than 97% of the total clean reads had Phred-like quality scores at the Q20 level (Supplementary Table S3).

Ultimately, 24,794 genes (including 1354 novel genes) were identified by Cufflinks v2.1.1 [24]. Via comparative transcriptome analysis, this high-quality RNA-Seq data provided a solid foundation for identifying key genes participating in carotenoid syntheses during watermelon development and ripening.

The numbers of transcripts identified in each sample, were expressed in FPKMs. Approximately 39.08% of expressed genes were in the 0–1 FPKM range, and 13.48% of expressed genes were above 60 FPKM, exhibited high expression level (Additional file 5: Table S4). Genes with normalized reads lower than 1 FPKM were removed from the subsequent analysis. These data showed sufficient coverage of the transcriptome during fruit development in five cultivars. The gene expression level among different experimental groups are compared in Additional file 6: Fig. S2a. The expression patterns among biologically repeated samples are highly consistent similarity (Additional file 6: Fig. S2b) and the correlation coefficient is close to 1(Additional file 7: Table S5), it indicates that the high quality of the replicates and reliable results of subsequent differential gene analysis.

Identification of Differentially Expressed Genes Among the Genotypes and Different Fruit Developmental Stages

To identify the genes correlating with the flesh color in different development stages, we conducted pairwise comparison at each developmental stage among five genotypes. The DEGs were screened with FDR < 0.05, |log2(FoldChange)| >1 as a threshold, the number and list of significantly differentially expressed gene (up-regulated and down-regulated) of each pairwise comparisons are shown in Additional file 8: Table S6 and Additional file 9–12: dataset 1–4, 16,781 genes were differentially expressed between at least one comparison.

At the early development stage (10 DAP), 5318 significantly differentially expressed genes were identified (Fig. 3a, Additional file 9: dataset 1). Specifically, 510,262□588□349 significantly differentially expressed genes were identified in red flesh variety compare to pink, orange, yellow, and white flesh variety. The numbers of other pairwise comparisons were listed in Additional file 8: Table S6. The fruit shape candidate gene ClFS1 (*Cla011257*) were differentially expressed in different fruits, consistent to the previous research [25]. *Cla019403* encode Xyloglucan endotransglucosylase/hydrolase, is related to the growth of plant cell [26], and highly expressed at this stage (Additional file 13: Fig. S3a, Additional file 14: Table S7). An Auxin response factor(ARF, *Cla009800*), A Growth-regulating factor 5 (GRF, *Cla006802*), An Auxin-induced SAUR-like protein (*Cla016617*), were obviously higher expressed at early development stage (Additional file 13: Fig. S3a, Additional file 14: Table S7) and associated with the fruit development and expansion [27]. There are less DEGs at 10 DAPs compared to later stages, due to little phenotypic differences among varieties at this stage.

At the pigment accumulation stage (20 DAP), 11814 significantly differentially expressed genes were identified (Fig. 3b, Additional file 10: dataset 2). Specifically, 2498,4830□3123□4876 significantly differentially expressed genes were identified in red flesh variety compare to pink, orange, yellow, and white flesh variety. The numbers of other pairwise comparisons were listed in Supplementary Table 6. Geranylgeranyl pyrophosphate synthase (*Cla015797*), Phytoene synthase protein (*Cla005425*), Phytoene

desaturase (Cla010898), Carotenoid isomerase (*Cla017593*), Lycopene cyclase (*Cla016840*), Violaxanthin de-epoxidase-related protein (*Cla007759*), 9-cis-epoxycarotenoid dioxygenase (*Cla015245*) are carotenoid biosynthesis genes and differentially expressed among materials(Additional file 13: Fig. S3b, Additional file 14:Table S7), directly control the flesh color formation of fruit. AP2-EREBP (*Cla000701, Cla017389*) and bHLH (*Cla020193, Cla022119*) were differentially expressed among materials(Additional file 13: Fig. S3b, Additional file 14:Table S7) and maybe the potential color regulator in watermelon [28, 29].

At the mature stage (34 DAP), 10779 significantly differentially expressed genes were identified (Fig. 3c, Additional file 11: dataset 3). Specifically, 2097,2572，2429，3316 significantly differentially expressed genes were identified in red flesh variety compared to pink, orange, yellow, and white flesh variety. The numbers of other pairwise comparisons were listed in Additional file 8: Table S6. Other comparison combinations are shown in the Additional file 9–11: dataset 1–3. Most of the carotenoids pathway genes and TFs are differentially expressed among genotypes. Geranylgeranyl reductase (*Cla003139, Cla019109*) Geranylgeranyl pyrophosphate synthase (*Cla015797, Cla020121*), Phytoene synthase protein (*Cla005425, Cla009122*), Phytoene desaturase (*Cla010898*), Carotenoid isomerase *(Cla017593, Cla011810)*, Lycopene cyclase (*Cla005011, Cla017416, Cla016840*), 9-cis-epoxycarotenoid dioxygenase (*Cla015245, Cla009779, Cla005404, Cla005453, Cla019578*), Beta-carotene hydroxylase (*Cla011420, Cla006149*), Zeta-carotene desaturase (*Cla003751*), Zeaxanthin epoxidase (*Cla020214*), and many TFs including (AP2-EREBP, MADS, MYB, G2-like, NAC, AUX), were differentially expressed at 34 DAP (Additional file 13: Fig. S3c, Additional file 14: Table S7). *Cla015245* and *Cla005404* were highly expressed in white flesh compared to yellow flesh, which may lead to the degradation of xanthophyll colorless flesh (Additional file 13: Fig. S3c, Additional file 14: Table S7). In particular, *Cla017389* and *Cla015515* (AP2-ERFBP) are highly expressed in red and pink color fruit. *Cla006599* and *Cla022119* (bHLH) are decreased at 20 and 34DAP compared to 10 DAP, similar to the expression pattern of CpbHLH1/2 in papaya [30], these genes may be important regulator gene.

For each genotype, there are 6430, 9019, 9605, 9147, and 10881 developmental DEGs were obtained in red, pink, orange, yellow, and white fleshed watermelon fruit, respectively, during the fruit development (Fig. 3d, Additional file 12: dataset 4). These results indicate that many genes are involved in the regulation of fruit development. More genes are differentially expressed in the white color flesh variety, suggesting a more complicated regulatory network of gene expression in this variety. We also identified some genes related to fruit development using comparative transcriptome analysis. A Cytokinin dehydrogenase gene (*Cla022463*) that was highly expressed at 10 DAP (Additional file 14: Table S7), which may contribute to the early fruit development. The pyrabactin resistance 1-like protein (PYL8) can regulate the plant growth and stress responses by mediate ABA signaling in Arabidopsis [31]. Here, we found that the expression of four Abscisic acid receptor PYL8 (*Cla004235, Cla004904, Cla015009, Cla021167*) were significantly different in different watermelon flesh (Additional file 13: Fig. S3d, Additional file 14: Table S7). These genes involved in fruit development and ripening.

The global hierarchical clustering (Additional file 15: Fig. S4a) and principal component analysis (Additional file 15: Fig. S4b) were performed based on the FPKM values for all the DEGs. The results

revealed that the 45 samples could be generally assigned into 3 main groups corresponding to development stages based on gene expression pattern, the samples from 10 DAP were distinctly clustered as one group, but the samples from 20 DAP and 34 DAP are relatively near, except for the samples at 34 DAP of red flesh (Additional file 15: Fig. S4a), suggesting that very complex regulatory networks in the middle and late stages of fruit development among different genotypes. In the PCA, three biological repeats for red color at 34 DAP are not together (Additional file 15: Fig. S4b), may be caused by a distinct regulatory mechanism in each fruit of one variety and that susceptible to environmental impact, but we used as valid data owing to the Pearson-correlation coefficient among the three repeats is up to 0.899 (Additional file 7: Table S5). These results indicate that the Differentially expressed genes between each pairwise comparison may be associated with the difference in watermelon fruit flesh coloration.

Clustering of DEGs into six Groups based on Gene Expression Patterns

Based on the pattern of gene expression, the 16,781 DEGs are grouped into 6 different subclusters using the h-cluster clustering method (Fig. 4a, Additional file 16: Table S8), the genes which grouped in the same cluster shared the similar expression pattern and have a similar function or participate in the same biological process. In the subcluster 1, 8,931 genes display a relatively stable expression levels across different stages and varieties. In the subclusters 2 and 3, gene expression exhibited a general downward trend according to the development stage. Genes in subclusters 4, 5, and 6 exhibited a general upward trend (Fig. 4a).

To functionally characterize the biological roles of these DEGs in the subcluster 1, GO enrichment analyses were performed. GO term analysis enriched almost the same proportion of genes into biological process, cellular component, and molecular function, GO terms related to various basic life activities, such as binding, cellular macromolecule metabolic process, intracellular, and cell (Additional file 17: Fig. S5a, Additional file 18: dataset 5). In addition, DEGs are enriched to the Spliceosome, RNA transport, and Ribosome biogenesis in eukaryotes pathways by KEGG analysis (Fig. 4b, Additional file 18: dataset 5). This result suggests that these DEGs are involved in basic growth and development of fruits, are to maintain the basic growth of plants.

3282 genes have a slightly higher expression at 20 DAP and 34DAP in the subcluster 2 (Fig. 4a). GO classification showed that GO terms such as single-organism metabolic process, small molecular metabolic process, and organonitrogen compound metabolic process were enriched (Additional file 17: Fig. S5b, Additional file 18: dataset 5). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis showed that the most significantly enriched pathways are proteasome, biosynthesis of amino acids, carbon metabolism pathway, TCA-cycle, Glycolysis/ Gluconeogenesis proteasome and pyruvate metabolism pathway (Fig. 4b, Additional file 18: dataset 5). The proteasome pathway containing 26S protease regulatory subunit genes and proteasome subunit type genes. The pyruvate metabolism pathway including Acetyl-CoA carboxylase biotin carboxylase, Pyruvate kinase, malate dehydrogenase, and others.

Subcluster 3 represents the genes with high expression at 34 DAP, the change ranges were more marked than subcluster1 (Fig. 4a). 754 DEGs in this cluster mainly allocated into molecular function and biological process according to GO term analysis, with 310 and 287 DEGs were classified into the metabolic process and catalytic activity (Additional file 17: Fig. S5c, Additional file 18: dataset 5). Notably, the DEGs were significantly involved in pathways associated with photosynthesis-antenna proteins biosynthesis, photosynthesis, protein processing in endoplasmic reticulum and biosynthesis of unsaturated fatty acids (Fig. 4b, Additional file 18: dataset 5). *Cla006149* (Beta-carotene hydroxylase), *Cla011420* (Beta-carotene hydroxylase), *Cla009122* (Phytoene synthase), and *Cla009779* (9-cis-epoxycarotenoid dioxygenase) are related to carotenoid pathway, these two genes showed increased expression at 20DAP and 34DAP (Additional file 13: Figs. S3b and S3c, Additional file 14:Table S7), indicating that accumulation of pigments occurs at this stage. In addition, some MYBs, AP2-ERFBPs, bHLHs, NACs, and WRKYs are also in subcluster 3 (Additional file 18: dataset 5), they are important regulators in fruit development and ripening. Two MYB (*Cla018631* and *Cla006739*) and two WRKY (Cla002243 and Cla002084) are highly expressed in yellow and pink color flesh (Additional file 13: Fig. S3e, Additional file 14:Table S7), respectively, may be involved in the regulation of yellow and pink color formation.

There are 2274, 1174, and 366 genes in the subclusters 4, 5, and 6, respectively, are highly expressed at 10 DAP and decreased to a low gene expression levels at the later stages, but the changing extents were different (Fig. 4a). Go enrichment analysis of subcluster 4 indicated that biological processes were most enriched. (Additional file 17: Fig. S5d, Additional file 18: dataset 5). KEGG analysis showed that genes significantly involved in the pathway of plant hormone signal transduction, such as signal transduction histidine kinase (*Cla000685*, *Cla005808*), auxin responsive protein (*Cla003635*), Ein3-binding f-box protein (*Cla020970*), and so on (Fig. 4b, Additional file 18: dataset 5). which are important for fruit early development. GO enrichment of subcluster 5 genes assigned to the GO terms of biological process and molecular function with no significance, such as protein phosphorylation, protein kinase activity (Additional file 17: Fig. S5e, Additional file 18: dataset 5). Genes are enriched into KEGG plant hormone signal transduction, Alanine, aspartate and glutamate metabolism, and others (Fig. 4b, Additional file 18: dataset 5). These genes are associated to the fruit growth. Subcluster 6, the enriched Go term were predominantly related to molecular function and biological process, such as “enzyme inhibitor activity” and “endopeptidase regulator activity” (Additional file 17: Fig. S5f, Additional file 18: dataset 5). KEGG enrichment analysis mostly related to the pathway of plant hormone signal transduction pathway, phenylpropanoid biosynthesis, and Phenylalanine metabolism (Fig. 4b, Additional file 18: dataset 5). *Cla019806*, *Cla0004102*, *Cla002975*, and *Cla016617* were involved in hormone synthesis, responsible for fruit enlargement and highly expressed during early stages of fruit development as compared to later fruit developmental stages (Additional file 13: Fig. S3f, Additional file 18: dataset 5, Additional file 14:Table S 7).

Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the “_1, _2, _4” represent the sample group at 10, 20, 34 DAP, respectively.

Co-expression network analysis identified carotenoid-related DEG s

To identify the potential genes (structural genes and putative transcription factors (TFs) that are highly associated with different kinds of carotenoids accumulation, the carotenoids content in each sample were used as a phenotypic data and 16,781 DEGs were used to perform the weighted gene co-expression network analysis (WGCNA).

A sample dendrogram and trait heatmap was constructed to illustrate the expression of each phenotypic parameter at different developmental stages (Additional file 19: Fig. S6a). The best parameter value determination for module construction is 7.7 for this dataset (Additional file 19: Fig. S6b). A total of 40 distinct co-expression modules were formed according to the pairwise correlations of gene expression across all samples and co-expression patterns of individual genes, as shown in the cluster dendrogram (Additional file 20: Fig. S7a), the modules labeled with different colors. Moreover, a network heatmap of all the DEGs in gene-modules was also drawn to exhibit the correlation between modules (Additional file 20: Fig. S7b). Notably, 6 coexpression modules (indicated with red underlines) have a high positive correlation with most of the carotenoids (Fig. 5a), indicating that the genes in these modules play an important role in the carotenoid accumulation.

'Yellow' module contains 846 Genes (including 34 TFs)(Additional file 21: Table S9) exhibited a stronger positive relationship with Zeaxanthin (correlation coefficient, cc = 0.91), Neoxanthin(cc = 0.91), Antheraxanthin (cc = 0.84), Violaxanthin (cc = 0.79), Phytofluene (cc = 0.81), apocarotenal (cc = 0.76), β-cryptoxanthin (cc = 0.68), lutein (cc = 0.69), and α-carotene (cc = 0.62) (Fig. 5a). In this module, a set of genes related to cellular metabolic process, and involved in pyruvate metabolism and proteasome pathway (Additional file 22: Fig. S8). Pyruvate is an important mediator of carbohydrate, fat and protein metabolism, and participates in several important metabolic functions in vivo. The proteasome is related to the regulation of carotenoid content in tomato [32]. This module containing genes that highly differentially expressed in the yellow/orange sample (Additional file 23: Fig. S9a), maybe are positive factors involved in yellow pigment accumulation. According to gene function annotation, *Cla005011* is lycopene beta-cyclase in watermelon [19]. *Cla003751* encodes a Zeta-carotene desaturase, *Cla020214* encode s zeaxanthin epoxidase, they are involved in the carotenoid pathway (Additional file 13: Fig. S3c, Additional file 14 and 21: Table S7 and S9). The *Cla014416* (Plastid-lipid-associated protein, ClPAP) is homologous to SlPAP(NCBI Reference Sequence: NP_001234183.1) that affect carotenoid content in tomato [33]. The expression level of *Cla014416* significantly higher in yellow and orange flesh color materials in this study (Additional file 13: Fig. S3g, Additional file 14 and 21: Table S7 and S9), different from the previous report that ClPAP highly expressed in red and orange color lines [34]. *Cla000655* encodes a cytochrome P450, is related to the protein lutein deficient 5 (CYP97A3) in Arabidopsis thaliana [35] and highly expressed in yellow and orange color fruits (Additional file 13: Fig. S3g, Additional file 14 and 21: Table S7 and S9). *Cla010839* is homologous to 15-cis-zeta-carotenoid isomerase in Arabidopsis thaliana [36]. *Cla018347* encodes a cytochrome P450, is related to carotenoid epsilon-monooxygenase (CYP97C1) in Arabidopsis thaliana [37] (Additional file 13: Fig. S3g, Additional file 14 and 21: Table S7 and S9). The MYB transcription factor can regulate the carotenoid content in *Mimulus lewisii* flowers [38],

Cla013280 and *Cla010722* belongs to MYB family and highly expressed in yellow color fruit (Additional file 13: Fig. S3g, Additional file 14 and 21: Table S7 and S9). The hub genes linked to this module were further analyzed using Cytoscape cytoHubba (Fig. 5b) and ATP synthase protein I, (*Cla013542*), Cysteine desulfuration protein SufE (*Cla003340*), Membrane protein (Cla003760), and others are important candidate gene for yellow color formation (Additional file 24: Fig. S10, Additional file 21 and 25: Table S9 and S10).

The 'darkred' module containing 111 genes were highly associated with the content of Antheraxanthin and Violaxanthin, having a correlation coefficient of 0.76 and 0.71 respectively. Heatmaps (Additional file 23: Fig. S9b) show that the 'darkred' module-specific genes were over-represented in samples (yellow, orange, white) be rich in Antheraxanthin and Violaxanthin. *Cla004704* encodes a Photosystem II oxygen evolving complex protein PsbP, *Cla005429* encode an Oxygen-evolving enhancer protein 2, chloroplastic, PsbP. *Cla004746* encodes a Chlorophyll a-b binding protein 6A (Additional file 13: Fig. S3g, Additional file 14 and 21: Table S7 and S9). *Cla021635* encodes a Photosystem I reaction center subunit II, rank as the top hub gene in this module (Fig. 5c, Additional file 24: Fig. S10, Additional file 25: Table S10). Many genes in this module are also related to chloroplast or photosystem I, II (Additional file 14: Table S7).

The "mediumpurple 3" module, with 32 identified genes, was highly correlated to α-carotene, Violaxanthin, Neoxanthin, lutein, and Zeaxanthin with correlation coefficient of 0.55, 0.70, 0.72, 0.66, and 0.73 respectively (Fig. 5a). The heatmap analysis were shown in (Additional file 23: Fig. S9c). *Cla005637*, *Cla017046*, and *Cla011297* were identified as candidate hub genes for this module (Additional file 24: Fig. S10, Additional file 25: Table S10). The black, and steelblue module were specific to the lycopene (cc = 0.57), and γ-Carotene content (cc = 0.55), respectively (Fig. 5a). The hub genes were listed in Additional file 25: Table S10, and expression level are showed in Supplementary Fig. 10.

By WGCNA, we found that most carotenoid pathway genes are included in the yellow module, which was highly correlated with the yellow and orange color samples (Additional file 23: Fig. S9a). Through network analysis, we identified hub genes for watermelon flesh carotenoid content (Additional file 25: Table S10).

Validation of the Expression of Key DEGs by qRT-PCR

The qRT-PCR analysis was used to validate the quality of RNA-Seq data. Five genes that previously reported and involved in the regulation of carotenoid synthesis during watermelon fruit development and seven key genes selected in this study were used for qRT-PCR analysis. Here, we found the strong correlation between the RNA-Seq and qRT-PCR data, indicating the reliability of our transcriptomic profiling data (Additional file 26: Fig. S11), and the selected gene from this study can be used as candidate genes for further research on watermelon fruit flesh color formation.

Discussion

Carotenoids in different flesh-colored watermelon fruits

Carotenoids are the second most abundant natural pigments worldwide [39], widely exist in carrot, sweet potatoes, red peppers, tomato (*Lycopersicon esculentum*), citrus fruit (*Citrus* spp.), peach (*Prunus persica*), *Cucumis* melons and watermelon (*Citrullus lanatus*). Carotenoids are divided into two subgroups, namely, carotenes (non-oxygenated, β-carotene and lycopene) and xanthophylls (oxygenated, lutein, violaxanthin, and neoxanthin). The different compositions and contents of carotenoid lead to the flesh color range from white to yellow and red. The watermelon flesh color is an important appearance quality and is closely related to consumers' preferences. The watermelon flesh color changed gradually during fruit ripening, and carotenoids content increased rapidly after 12 days of pollination [40]. In this study, the flesh color was observed all white at the early fruit developmental stages and then changed to vivid color at the later stages, among the five genotypes. Lycopene was the main pigment in red and pink flesh genotypes, these results were consistent with the previous report [34, 40]. The orange flesh color is largely determined by the content of β-carotene, the same result as the previous report [10], moreover, the orange flesh has a higher content of violaxanthin zeaxanthin, neoxanthin, lutein, and β-cryptoxanthin than other genotypes. The previous study reported that violaxanthin and lutein, or neoxanthin are the dominant carotenoids in yellow flesh [10], in this study, we also detected the higher level of antheraxanthin, violaxanthin zeaxanthin, neoxanthin, lutein, and β-cryptoxanthin in yellow fleshed samples. In addition, violaxanthin, antheraxanthin, and lutein were trace observed in white flesh samples. The α-carotene availability may partially explain the lutein content in the watermelon. Remarkable apocarotenal level was detected in orange and yellow flesh, which may produce a special flavor for the fruits [28]. These results indicated the specific variation in carotenoid biosynthesis leads to variations in flesh colors.

Regulation of carotenoid biosynthesis pathway in different flesh-colored watermelon

The flesh color is due to the accumulation of pigments, which are regulated and controlled by a complicated network consisting of a series of biosynthesis-, degradation-, and stable storage-related genes. To determine the potential regulatory networks underlying pigment content in flesh, we performed comparative transcriptome analysis combine with WGCNA to identify highly correlated genes with the carotenoid accumulation pattern. In our study, 44 Carotenoid pathway genes were differentially expressed in different samples (Additional file 27: Table S11). The geranylgeranyl reductase (*Cla019109*), geranylgeranyl transferase type subunit beta (*Cla004679*), phytoene desaturase (*Cla010898*), 15-cis-zeta-carotene isomerase (*Cla010839*), 9-cis-epoxycarotenoid dioxygenase (*Cla015245*, *Cla005404*), were differentially expressed in the samples (Fig. 6, Additional file 27: Table S11). The phytoene synthase (*Cla009122*), a rate-limiting enzyme in carotenoid biosynthesis flux, is highly expressed in the later development stage, consistent with the carotenoid accumulation. The lycopene beta-cyclase (*Cla005011*) exhibit the similar expression level in the different sample, corresponding to the previous conclusion that the content of lycopene was not related to the gene expression level, but related to the abundance of protein level [19]. Beta-carotene hydroxylase (*Cla006149*) was highly expressed in the orange, yellow, pink flesh, which may lead to xanthophylls synthesis (Fig. 6). The orange gene (*BoOr*) encodes a plastidial DNA J cysteine-rich domain-containing protein and is an important regulator for carotenoid biosynthesis

in cauliflower [41], in the current study, *Cla018406* (on chromosome 7) which is homologous to *BoOr* gene and highly expressed is orange flesh, is a strong candidate gene for orange flesh color in watermelon (Additional file 26: Fig. S11, Additional file 27: Table S11), different from the previously identified QTL of on chromosome 1[17]. *SlBBX20* (zinc-finger transcription factor) is a positive regulator of carotenoid accumulation in tomato [42], here we identified a zinc finger CCCH domain-containing protein (*Cla007686*) which is highly expressed in the red color watermelon (Additional file 26: Fig. S11, Additional file 27: Table S11), maybe a key regulator for lycopene accumulation. Previous studies showed that HY5 (ZIP) and Golden2-Like2 (MYB), are involved in chloroplast biogenesis in Arabidopsis and tomato [43, 44]. Consistent with this, we found that two ZIP TF (*Cla016581*, *Cla020795*), two GLK2 TF (*Cla010265*, *Cla020369*), were differentially expressed in different samples (Additional file 28: Fig. S12a, Additional file 27: Table S11). These TFs may be important for flesh color formation and maintains white color. The R2R3-MYB protein family act as a regulatory function in carotenoid pathway [45]. Here, five R2R3-MYB related genes (*Cla020633, Cla007790, Cla009263, Cla017995*, and *Cla019223*) were identified and found to be upregulated in colored flesh compared to light color flesh (Additional file 28: Fig. S12a, Additional file 27: Table S11), and these might be the candidate genes for regulating color formation in watermelon. Two MADS (*Cla000691*, *Cla010815*) homologous to SlMADS1, which plays an important role in fruit ripening as a repressive modulator in tomato [46]. Two MADS (*Cla009725*, *Cla019630*) homologous to CsMADS6, which coordinately expressed with citrus fruit development and coloration [47]. Besides the carotenoid biosynthesis genes, several other biological pathways and genes (TFs) might be involved in the regulation of carotenoid pathway. We identified 26 DEGs related to chlorophyll biosynthesis (Additional file 28: Fig. S12b, Additional file 27: Table S11), including chlorophyll synthase (*Cla018095*) highly expressed in orange and yellow fleshed samples, chlorophyll a-b binding protein (*Cla015680*) were highly expressed in the colored flesh, we speculated that they are important for flesh coloration. Moreover, plastid is the place where carotenoids are synthesized and stored, plastid development is closely related to the accumulation of carotenoids, 22 DEGs were annotated to be involved in the plastid biogenesis (Additional file 28: Fig. S12c, Additional file 27: Table S11).

Considering the fact that molecular mechanisms underlying flesh color formation are still not very well understood, and needs deep mining of the regulatory mechanisms our results will assist further understanding of the color-specific molecular mechanisms and will help to validate the functions of candidate genes we predicted that are involved in flesh coloration during fruit development in different flesh colored watermelon.

Conclusions

In this study, we performed transcriptomic comparison among five different flesh color cultivated watermelon to understand the carotenoid accumulation patterns and their regulation during watermelon fruit development. Colored watermelon genotypes possess a high content of carotenoids as compared to the colorless genotypes specifically at the maturity stage. Our results provide a comprehensive information on genes involved in the fruit development and color formation. The WGCNA is a useful method in identifying trait-specific modules and key genes. Our multiple screening steps suggests that

these might be the true candidate genes. We speculate *Cla018406* (Chaperone protein dnaJ-like protein), *Cla007686* (a zinc finger CCCH domain-containing protein), *Cla003760* (membrane protein) and *Cla021635* (photosystem I reaction center subunit II) are candidate genes for orange, red, and yellow flesh in watermelon. Further researches like transgene or genome editing are required to elucidate the underlying molecular mechanisms controlling the accumulation of carotenoids in watermelon.

Methods

Plant materials and sampling

Five watermelon inbred lines seeds were provided by the polyploidy watermelon research group (Zhengzhou, China), Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences. The fruits with different flesh colors were identified by our lab using vision and colorimeter: the orange fleshed inbred Qitouhuang, the yellow fleshed inbred Xihua, the scarlet red-fleshed inbred Zhengzhou No. 3, the coral red-fleshed inbred Hualing, and the white fleshed inbred Bingtangcui (Fig. 1).

Watermelon seeds were sown in pots (filled with nutritional media) in a greenhouse in April 2018, one-month-old watermelon seedlings were transplanted in the open field at the Xinxiang experimental farm (Xinxiang, Henan, China), with spacing as 30 cm between plants and 150 cm between rows. The plants were separated by genotype and replication and placed under field management followed common horticultural practices (fertilization, irrigation, pathogen prevention, and pest control) for open-field watermelon growing.

In addition, flowers were hand-pollinated and tagged to record the number of days after pollination (DAP). Flesh samples were collected from uniform injury-free watermelon fruits at three critical development stages(10, 20 and 34 DAP)respectively, then these samples were immediately frozen in liquid nitrogen and stored at -80 °C until use. The pooled samples from three fruits as one biological replicate, three individual biological replicates for each treatment. Approximately 10 g and 50 g of flesh samples were collected for RNA-seq analysis and carotenoid profiles determination.

Phenotyping

The fruits were picked, cut open longitudinally and visually scored for flesh color first. Images were taken from all fruits, allowing further confirmation of flesh color phenotypes during data analysis. Flesh color was also measured using a Chroma-meter Konica-Minolta CR-400 (Japan) as an unbiased method to fix the visualized color of the developing fruits samples (nine reads per fruit). CIE color space L*, a*, and b* values were obtained at nine points of each fruit cross section. L* represents lightness (ranging from 0, black to 100, white), a* represents red (positive)to green (negative) axis, and b* represents yellow (positive) to blue (negative) axis. The colorimeter was calibrated on a white plate before each use. Chroma (C) = $[(a^*)^2 + (b^*)^2]^{1/2}$ measures color saturation or intensity and the hue angle (h) = arc tan b*/a* determines the red, yellow, green, blue, purple, or intermediate colors between adjacent pairs of these basic colors.

Quantitation of carotenoid

100 mg flesh powder samples (after lyophilization) were used for carotenoid measurement. Carotenoid extraction using normal hexane–acetone–ethanol (2:1:1, V/V/V) containing 0.01% butyl hydroxytoluene (BHT). The extracted carotenoids were analyzed using an UPLC-APCI-MS/MS system (Ultra-high-performance liquid chromatography; ExionLC™ AD); The API 6500 Q TRAP LC/MS/MS system, equipped with an APCI Turbo Ion-Spray interface). The measurement conditions: YMC C30 column (3 μm, 2 mm × 100 mm); mobile phase A: acetonitrile:methanol (3:1, v/v) and 0.01% butylated hydroxytoluene(BHT); mobile phase B: methyl tert-butyl ether and 0.01% BHT; gradient program: 85:5 (v/v) at 0 min, 75:25 (v/v) at 2 min, 40:60 (v/v) at 2.5 min, 5:95 (v/v) at 3 min, 5:95 (v/v) at 4 min, 85:15 (v/v) at 4.1 min, 85:15 (v/v) at 6 min; flow rate, 0.8 mL/min; temperature: 28 °C; injection volume: 5 μL. The APCI source operation parameters were as follows: ion source: turbo spray; source temperature: 350 °C; curtain gas (CUR): 25.0 psi; collision gas (CAD): medium. The declustering potential (DP) and collision energy (CE) for individual multiple reaction monitoring transitions were determined with further DP and CE optimization. The authentic carotenoid standards (Sigma Aldrich, St. Louis, MO, USA) were used for the qualitative analysis of MS data. The relative contents of each sample were corresponding to the spectral peak intensity values. The absolute content was calculated using linear equations of standard curves.

RNA extraction and sequencing

For different watermelon flesh samples, total RNA was extracted using Plant Total RNA Purification Kit (GeneMark, Beijing, China) following the manufacturer's instructions. The RNA degradation and contamination were monitored on 1% agarose gels. The RNA purity, concentration and integrity were checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively.

A total amount of 5 μg total RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). The clustering of the samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia). The library preparations were sequenced on an Illumina Hiseq platform and 125 bp/150 bp paired-end reads were generated. The high-quality data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. The watermelon reference genome (97103 V1) was downloaded from the genome website directly (http://cucurbitgenomics.org/organism/1). Paired-end clean reads were aligned to the reference genome using Hisat2 v2.0.4.

Quantification of gene expression level

HTSeq v0.9.1 was used to count the reads numbers mapped to each gene and then FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene[24].

Differential expression analysis

Differential expression analysis was performed using the DESeq R package (1.18.0). Genes with an adjusted P-value < 0.05 found by DESeq were assigned as differentially expressed.

GO term and KEGG enrichment analysis of differentially expressed genes

Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the GOseq R package, in which gene length bias was corrected. GO terms with corrected P value less than 0.05 were considered significantly enriched by differential expressed genes. We used KOBAS software to test the statistical enrichment of differential expression genes in KEGG pathways.

Co-expression networks analysis

Coexpression networks analysis was performed using R package WGCNA [48] and visualized using Cytoscape software [49], based on 16,781 normalized FPKM values and the trait data representing carotenoid levels in different samples. The key gene in each module was analyzed using cytohubba.

Validation of DEG expression by qRT-PCR

The first-strand cDNA was synthesized from 1 μg RNA using a Prime ScriptTM RT reagent kit with gDNA Eraser (TaKaRa, Kusatsu, Shiga, Japan) based on the manufacturer's protocol. The cDNA was synthesized from 1 ug of total RNA with PrimeScript™ RT reagent Kit with gDNA Eraser following the manufacturer's instructions (Takara, Japan). For quantitative reverse transcription polymerase chain reaction (qRT-PCR), relative gene expression levels of the target gene were measured using a Roche LightCycler480 RT-PCR system (Roche, Swiss). The SYBR Green real-time PCR mix was added to the reaction system according to the manufacturer's instructions. The primers were designed using Primer premier 6 based on Cucurbit Genomics Database (http://cucurbitgenomics.org/) and listed in Additional file 29: Table S12. All genes were run in triplicate from the three biological replicates. The raw data of qRT-PCR were analyzed using LCS 480 software 1.5.0.39 (Roche, Swiss) and the relative expression was determined by using the $2^{-\Delta\Delta CT}$ method. The watermelon *ClCAC* and *ClACTIN* genes were used as the internal control[50].

Statistical analysis

Statistical analysis for color parameters was conducted using SPSS 19.0.

Abbreviations

DAP
days after pollination; R:red; P:pink, O:orange; Y:yellow; W:white; DEGs:differentially expressed genes; TFs:transcription factors; WGCNA:weighted gene coexpression network analysis; C:Chroma; PCA:Principal Component Analysis; FPKM:Reads Per Kilobase per Million mapped reads; qRT-

PCR:Quantitative real-time polymerase chain reaction; GO:Gene Ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; TCA cycle:tricarboxylic acid cycle; cc:correlation coefficient; BHT:butyl hydroxytoluene; Metabolites background are colored according to their compound colors, Enzymes are red font. DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; GGPS, geranylgeranyl diphosphate synthase; GGPR, Geranylgeranyl diphosphate reductase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-carotene desaturase; CRTISO, carotenoid isomerase; LCYE, lycopene ε-cyclase; LCYB, lycopene β-cyclase; CHYB, β-carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, Violaxanthin de-epoxidase; NXS, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase.

Declarations

Availability of data and materials

The transcriptome raw reads have been deposited as a BioProject under accession: PRJNA644468. The materials are available from the corresponding author on reasonable request after the publication of the work.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Funding

In this work, the materials planting, sample collection was supported by the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-ZFRI-07). The RNA sequencing was supported by the National Key R&D Program of China (2018YFD0100704), the China Agriculture Research System (CARS-25-03) and the National Nature Science Foundation of China (31672178 and 31471893). The qRT-PCR experiment was supported by Scientific and Technological Project of Henan Province (202102110197). The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors’ contributions

W.L. and P.Y. conceived and designed the experiments. P.Y., N.H., S.Z., X.L., H.Z., W.D. C.G., and H.G. prepared the materials and take the samples. P.Y. performed the lab experiments. P.Y. and J.U. carried out data analysis. P.Y. make the figures/tables and wrote the whole manuscript. P.Y. and J.U. checked the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Perkins-Veazie P, Collins K, Clevidence J. B: **WATERMELONS AND HEALTH**. *Acta Horticulturae* 2007(731):121–128.
2. G ZR: **A Review of Epidemiologic Evidence that Carotenoids Reduce the Risk of Cancer**. *Journal of Nutrition* 1989(1):1.
3. Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, Li L. Carotenoid Metabolism in Plants: The Role of Plastids. Mol Plant. 2018;011(001):58–74.
4. Alder A, Jamil M, Marzorati M, Bruno M, Vermathen M, Bigler P, Ghisla S, Bouwmeester H, Beyer P, Al-Babili S. The Path from ?-Carotene to Carlactone, a Strigolactone-Like Plant Hormone. Science. 2012;335(6074):1348–51.
5. Cazzonelli CI, Pogson BJ. Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci. 2010;15(5):0–274.
6. Michel H. **Carotenoid oxidation products as stress signals in plants**. *Plant Journal* 2014.
7. Walter MH, Strack D. Carotenoids and their cleavage products: Biosynthesis and functions. Nat Prod Rep. 2011;28(4):663.
8. Gusmini G, Wehner TCJJoH. Qualitative Inheritance of Rind Pattern and Flesh Color in Watermelon. J Hered. 2006;97(2):177–85.
9. Perkins-Veazie P, Collins JK, Davis AR, Roberts W. Carotenoid content of 50 watermelon cultivars. J Agric Food Chem. 2006;54(7):2593–7.
10. Liu C, Zhang H, Dai Z, Liu X, Liu Y, Deng X, Chen F, Xu J. Volatile chemical and carotenoid profiles in watermelons [Citrullus vulgaris(Thunb.) Schrad (Cucurbitaceae)] with different flesh colors. Food Science Biotechnology. 2012;21(2):531–41.
11. Tadmor Y, King S, Levi A, Davis A, Meir A, Wasserman B, Hirschberg J, Lewinsohn E. Comparative fruit colouration in watermelon and tomato. Food Res Int. 2005;38(8–9):0–841.
12. WR H, GH S, TC W. **Interaction of flesh color genes in watermelon**. *Journal of Heredity* 1998(1):1.
13. Bang HJ, Davis AR, Kim SG, Leskovar DI, King SR. Flesh color inheritance and gene interactions among canary yellow, pale yellow, and red watermelon. Infection Immunity. 2010;65(11):4620–3.

14. Hashizume T, Shimamoto I, Hiral M. Construction of a linkage map and QTL analysis of horticultural traits for watermelon [Citrullus lanatus (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. Tagtheoretical Applied Geneticstheoretische Angewandte Genetik. 2003;106(5):779.
15. Liu S, Gao P, Wang X, Davis AR, Baloch AM, Luan F. Mapping of quantitative trait loci for lycopene content and fruit traits in Citrullus lanatus. Euphytica. 2015;202(3):411–26.
16. Li N, Shang J, Wang J, Zhou D, Li N, Ma S. **Discovery of the Genomic Region and Candidate Genes of the Scarlet Red Flesh Color (Yscr) Locus in Watermelon (Citrullus Lanatus L.)**. *Frontiers in Plant Science* 2020, 11(116).
17. Branham S, Vexler L, Meir A, Tzuri G, Frieman Z, Levi A, Wechter WP, Tadmor Y, Gur A. Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. Mol Breeding. 2017;37(12):146.
18. Zhang J, Guo S, Ren Y, Zhang H, Gong G, Zhou M, Wang G, Zong M, He H, Liu F, et al. High-level expression of a novel chromoplast phosphate transporter ClPHT4;2 is required for flesh color development in watermelon. New Phytol. 2017;213(3):1208–21.
19. Zhang J, Sun H, Guo S, Ren Y, Li M, Wang J, Zhang H, Gong G, Xu Y. **Decreased Protein Abundance of Lycopene β-Cyclase Contributes to Red Flesh in Domesticated Watermelon**. *Plant Physiology* 2020:pp. 01409.02019.
20. Dou JL, Yuan PL, Zhao SJ, Nan HE, Zhu HJ, Gao L, Wan-Li JI, Xu-Qiang LU, Liu WG. Effect of ploidy level on expression of lycopene biosynthesis genes and accumulation of phytohormones during watermelon (Citrullus lanatus) fruit development and ripening. Journal of Integrative Agriculture. 2017;16(9):1956–67.
21. Yuan PL, Liu WG, Zhao SJ, Lu XQ, Yan ZH, He N, Zhu HJ, Sari N, Solmaz I, Aras V. **Lycopene content and expression of phytoene synthase and lycopene å°¾-cyclase genes in tetraploid watermelon**. In: *Cucurbitaceae Xth Eucarpia Meeting on Genetics & Breeding of Cucurbitaceae: 2012*.
22. Zhu Q, Gao P, Liu S, Zhu Z, Amanullah S, Davis AR, Luan F. Comparative transcriptome analysis of two contrasting watermelon genotypes during fruit development and ripening. BMC Genom. 2017;18(1):3.
23. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.
24. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
25. Dou J, Zhao S, Lu X, He N, Zhang L: **Genetic mapping reveals a candidate gene** (**ClFS1**) **for fruit shape in watermelon** (**Citrullus lanatus L**.). *Theoretical & Applied Genetics* 2018.
26. Saladié M, Rose JKC, Cosgrove DJ, Catalá C. Characterization of a new xyloglucan endotransglucosylase/hydrolase (XTH) from ripening tomato fruit and implications for the diverse modes of enzymic action. Plant Journal for Cell Molecular Biology. 2006;47(2):282–95.

27. Horiguchi G, Kim GT, Tsukaya H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. The Plant journal: for cell molecular biology. 2005;43(1):68–78.
28. Yuan H, Zhang J, Nageswaran D, Li L. Carotenoid metabolism and regulation in horticultural crops. Hortic Res. 2015;2:15036.
29. Llorente B, D'Andrea L, Ruiz-Sola MA, Botterweg E, Pulido P, Andilla J, Loza-Alvarez P, Rodriguez-Concepcion M. Tomato fruit carotenoid biosynthesis is adjusted to actual ripening progression by a light-dependent mechanism. The Plant journal: for cell molecular biology. 2016;85(1):107–19.
30. Zhou D, Shen Y, Zhou P, Fatima M, Ming R. **Papaya CpbHLH1/2 regulate carotenoid biosynthesis-related genes during papaya fruit ripening**. *HORTICULTURE RESEARCH* 2019, 6.
31. Lim CW, Baek W, Han SW, Lee SC. Arabidopsis PYL8 Plays an Important Role for ABA Signaling and Drought Stress Responses. The plant pathology journal. 2013;29(4):471–6.
32. Tang X, Miao M, Niu X, Zhang D, Cao X, Jin X, Zhu Y, Fan Y, Wang H, Liu Y, et al. Ubiquitin-conjugated degradation of golden 2-like transcription factor is mediated by CUL4-DDB1-based E3 ligase complex in tomato. New Phytol. 2016;209(3):1028–39.
33. Kilambi HV, Kumar R, Sharma R, Sreelakshmi Y. Chromoplast-Specific Carotenoid-Associated Protein Appears to Be Important for Enhanced Accumulation of Carotenoids in hp1 Tomato Fruits. Plant Physiol. 2013;161(4):2085–101.
34. Fang X, Liu S, Gao P, Liu H, Wang X, Luan F, Zhang Q, Dai Z. Expression of ClPAP and ClPSY1 in watermelon correlates with chromoplast differentiation, carotenoid accumulation, and flesh color formation. Sci Hortic. 2020;270:109437.
35. Kim J, Dellapenna D. Defining the primary route for lutein synthesis in plants: The role of Arabidopsis carotenoid β-ring hydroxylase CYP97A3. Proc Natl Acad Sci USA. 2006;103(9):3474–9.
36. Yu, Chen, Faqiang, Li, Wurtzel, Eleanore T. **Isolation and Characterization of the Z-ISO Gene Encoding a Missing Component of Carotenoid Biosynthesis in Plants**. *Plant Physiology* 2010.
37. Tian L, Magallanes-Lundback M, Musetti V, DellaPenna D. Functional analysis of beta- and epsilon-ring carotenoid hydroxylases in Arabidopsis. Plant Cell. 2003;15(6):1320–32.
38. Janelle M, Sagawa, Lauren E, Stanley, Amy M, LaFountain H: **An R2R3-MYB transcription factor regulates carotenoid pigmentation inMimulus lewisiiflowers**. *New Phytologist* 2015.
39. Nisar N, Li L, Lu S, Khin NC, Pogson BJ. Carotenoid Metabolism in Plants. Mol Plant. 2015;8(1):68–82.
40. Lv P, Li N, Liu H, Gu H, Zhao W-e. **Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars**. *Food Chemistry* 2015, 174.
41. Lu S, Van Eck J, Zhou X, Lopez AB, O'Halloran DM, Cosman KM, Conlin BJ, Paolillo DJ, Garvin DF, Vrebalov J, et al. The cauliflower Or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. Plant Cell. 2006;18(12):3594–605.

42. Xiong C, Luo D, Lin A, Zhang C, Shan L, He P, Li B, Zhang Q, Hua B, Yuan Z, et al. A tomato B-box protein SlBBX20 modulates carotenoid biosynthesis by directly activating PHYTOENE SYNTHASE 1, and is targeted for 26S proteasome-mediated degradation. New Phytol. 2019;221(1):279–94.
43. Toledo-Ortiz G, Johansson H, Lee KP, Bou-Torrent J, Stewart K, Steel G, Rodríguez-Concepción M, Halliday KJ. The HY5-PIF Regulatory Module Coordinates Light and Temperature Control of Photosynthetic Gene Transcription. PLOS Genetics. 2014;10(6):e1004416.
44. Powell ALT, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons C, Fernández-Muñoz R, Vicente A, et al. Uniform ripening Encodes a Golden 2-like Transcription Factor Regulating Tomato Fruit Chloroplast Development. Science. 2012;336(6089):1711–5.
45. Meng X, Yang D, Li X, Zhao S, Sui N, Meng Q. Physiological changes in fruit ripening caused by overexpression of tomato SlAN2, an R2R3-MYB factor. Plant Physiology Biochemistry. 2015;89:24–30.
46. Tingting D, Zongli Hu, Lei, Deng, Yi, Wang. Mingku, Zhu: **A Tomato MADS-Box Transcription Factor, SlMADS1, Acts as a Negative Regulator of Fruit Ripening**. *Plant Physiology* 2013.
47. Lu SW, Zhang Y, Zhu K, Yang W, Ye JL, Chai L, Xu Q, Deng X. **The citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes**. *Plant Physiology* 2018:pp. 01830.02017.
48. Yang J, Yu H, Liu BH, Zhao Z, Liu L, Ma LX, Li YX, Li YY, Xing Y: **DCGL v2.0**: **An R Package for Unveiling Differential Regulation from Differential Co-expression**. *Plos One* 2013, **8**(11):e79729.
49. Yeung N, Cline MS, Kuchinsky A, Smoot ME, Bader GD. **Exploring Biological Networks with Cytoscape Software**. *Curr Protoc Bioinformatics* 2008, Chap. 8:Unit 8.13.
50. Kong Q, Yuan J, Gao L, Zhao L, Cheng F, Huang Y, Bie Z. Evaluation of Appropriate Reference Genes for Gene Expression Normalization during Watermelon Fruit Development. PLoS One. 2015;10(6):e0130865.

Supplementary Information

Additional file1: Table S1 Estimation of color coordinates values in the watermelon fruits of different development stages.

Additional file2: Table S2 The contents of carotenoids compounds (ug/g).

Additional file3: Fig. S1 PCA plots for all samples using log-transformation data of the relative content of carotenoid metabolites. T [1]and T [2] explain 69.9% and 20.4%, respectively. Each variety is represented by dots in different colors. In addition, a clear distinction between flesh color can be seen in the time-course. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow and white colored flesh, respectively; DAP, days after pollination; QC, quality control samples.

Additional file4: Table S3 A overview of the RNA-Seq data.

Additional file 5: Table S4 Numbers of detected transcripts in each sample.

Additional file 6: Fig. S2 Overview of the transcriptome sequencing. (a) Comparison of Gene Expression Levels among Different Experimental Groups. (b) Pearson correlation analysis based on global RNA-seq data from 45 libraries. R^2: the square of Pearson correlation coefficient. The correlation coefficient is close to 1, indicating that the similarity of expression patterns between biological duplicate samples is high. The "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively. For example: O_1, O_2, and O_4 represent the sample of 10 DAP, 20 DAP, 34 DAP of orange fleshed fruit; O_1_1, O_1_2, and O_1_2 mean the three biologically repeats.

Additional file 7: Table S5 Pearson correlation coefficient list between samples.

Additional file 8: Table S6 Numbers of differentially expressed genes (DEGs) in each Pairwise comparison.

Additional file 9: dataset 1 DEG at 10 DAP between the genotypes

Additional file 10: dataset2 DEG at 20 DAP between the genotypes

Additional file 11: dataset3 DEG at 34 DAP between the genotypes

Additional file 12: dataset4 DEG between different stage in each genotype

Additional file 13: Fig. S3 The expression selected key DEGs listed in this study. Each gene was analyzed separately. Red represents the high expression gene, and blue represents the low expression gene. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively.

Additional file 14: Table S7 FPKM of gene listed in this study

Additional file 15: Fig. S4 Hierarchical clustering analysis(a) and Principal component analysis (b) of the overall differentially expressed genes of transcriptomes in 45 samples. The log10 (FPKM + 1) value was normalized and transformed. Red represents the high expression gene, and blue represents the low expression gene. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively.

Additional file 16: Table S8 Gene list of each subcluster

Additional file 17: Fig. S5 Go term analysis of Subcluster1-6.

The asterisk on the column indicates significant enrichment, the figure related to supplementary dataset 5.

Additional file 18: dataset 5 Go and KEGG analysis of genes in six subclusters

Additional file 19: Fig. S6 (a) Sample dendrogram and module trait heatmap at each developmental stage. (b) The parameter, soft threshold, determination for module construction. The best value is 7.7 for this dataset.

Additional file 20: Fig. S7 (a) Genes cluster dendrogram (hierarchical clustering tree) of the transcriptome. Each leaf in the tree represents one gene. 40 modules were built based on gene expression value (labeled with different colors). (b) Network heatmap of selected genes.

Additional file 21: Table S9 Gene list of modules

Additional file 22: Fig. S8 Go term and KEGG analysis of genes in yellow module

Additional file 23: Fig. S9 Heat cluster analysis of (a) yellow module, (b) darkred module, and (c) purple module.

Additional file 24: Fig. S10 The heatmap of hub genes. The log10 (FPKM + 1) value was normalized and transformed. Red represents the high expression gene, and blue represents the low expression gene. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively.

Additional file 25: Table S10 Hub gene in modules calculated using cytohubba

Additional file 26: Fig. S11 Validation of selected DEGs expression by qRT-PCR.

Transcript abundance changes were detected by FPKM values according to RNA-Seq (left column). Red lines and column represent the FPKM values and RT-qPCR expression values, respectively. Values are reported as means ± SEs (n = 3). Three biological replications were included in the analysis.

Additional file 27: Table S11 The genes related to carotenoid biosynthesis

Additional file 28: Fig. S12 The heatmap of color related genes. The log10 (FPKM + 1) value was normalized and transformed. Red represents the high expression gene, and blue represents the low expression gene. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively.

Additional file 29: Table S12 Primers used for the quantitative real-time PCR analysis

Figures

Figure 1

The flesh color of five watermelon genotypes at 10 DAP, 20 DAP, and 34 DAP.

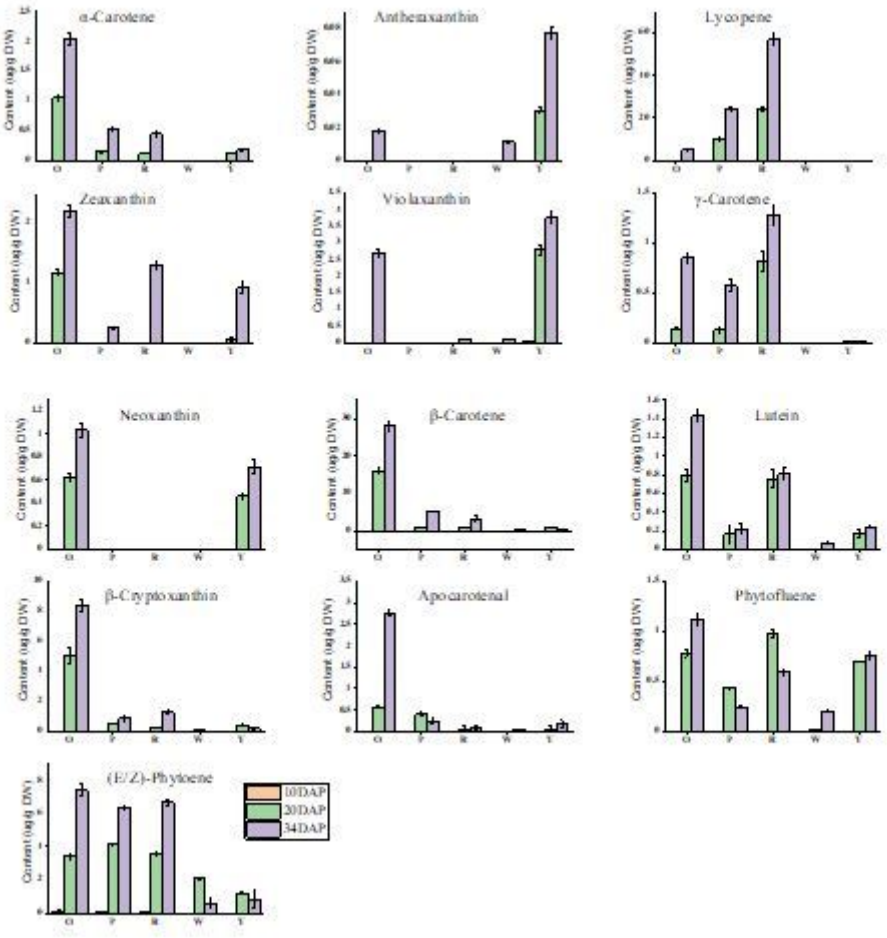

Figure 2

Content of 13 metabolites of the carotenoid pathway. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; DAP, days after pollination; Data are shown as the means ± s.e.m., n = 3.

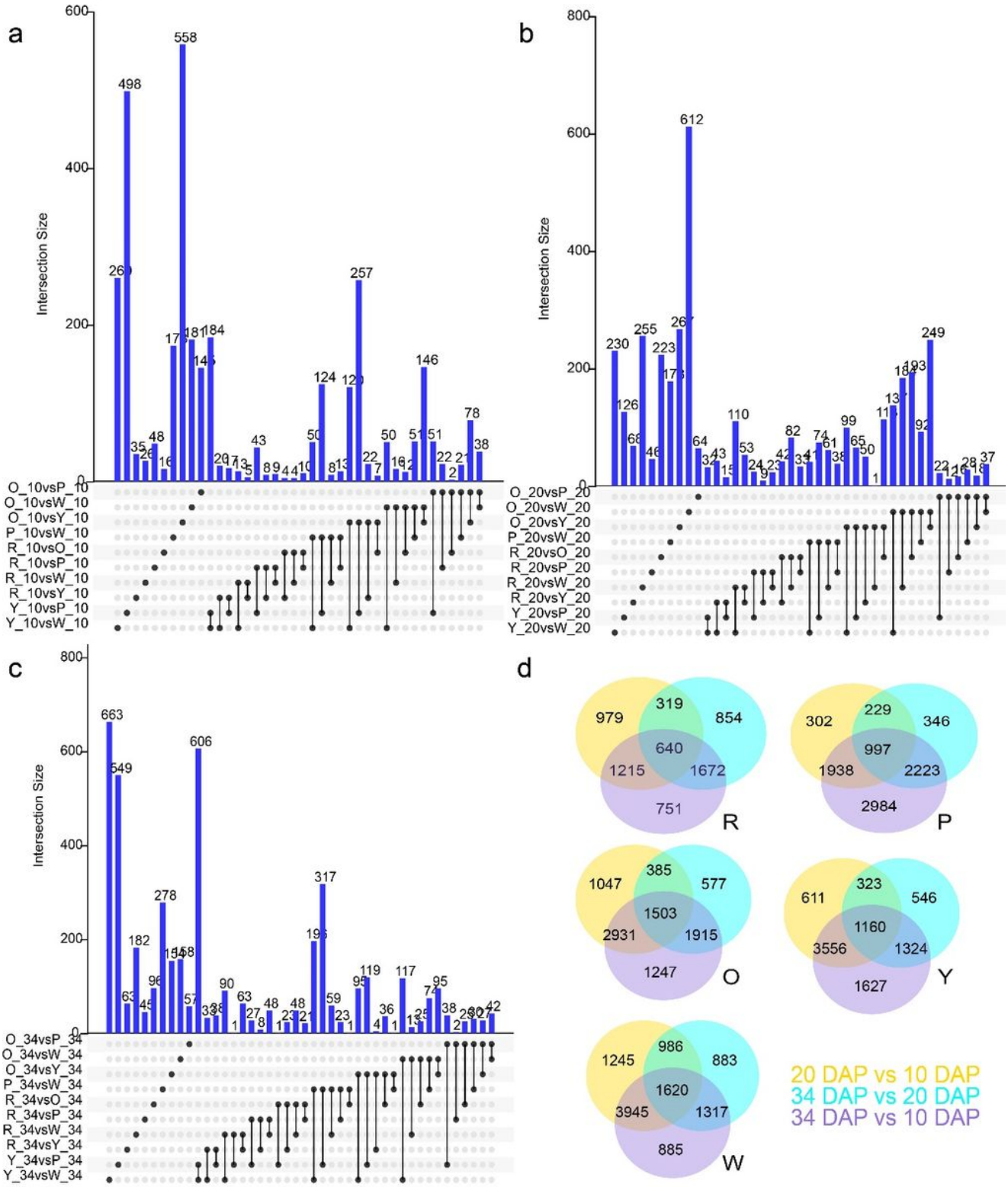

Figure 3

Venn diagrams of differentially expressed transcripts among 5 genotypes at 10 DAP(a), 20 DAP(b),34 DAP(c), and among the 3 stages of each genotype (d). Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed genotype, respectively.

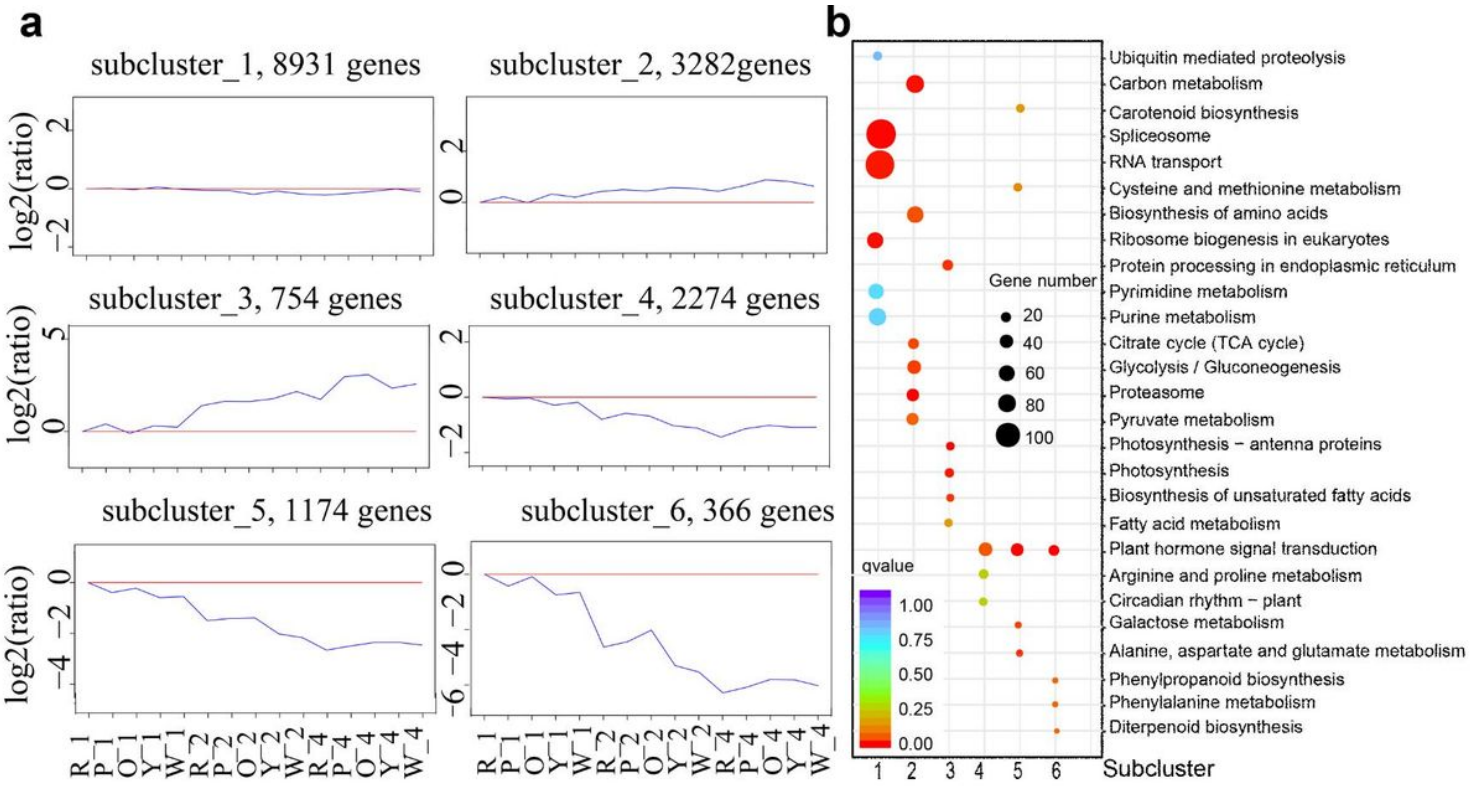

Figure 4

(a) h-clustering of DEGs and (b) KEGG enrichment analysis Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively.

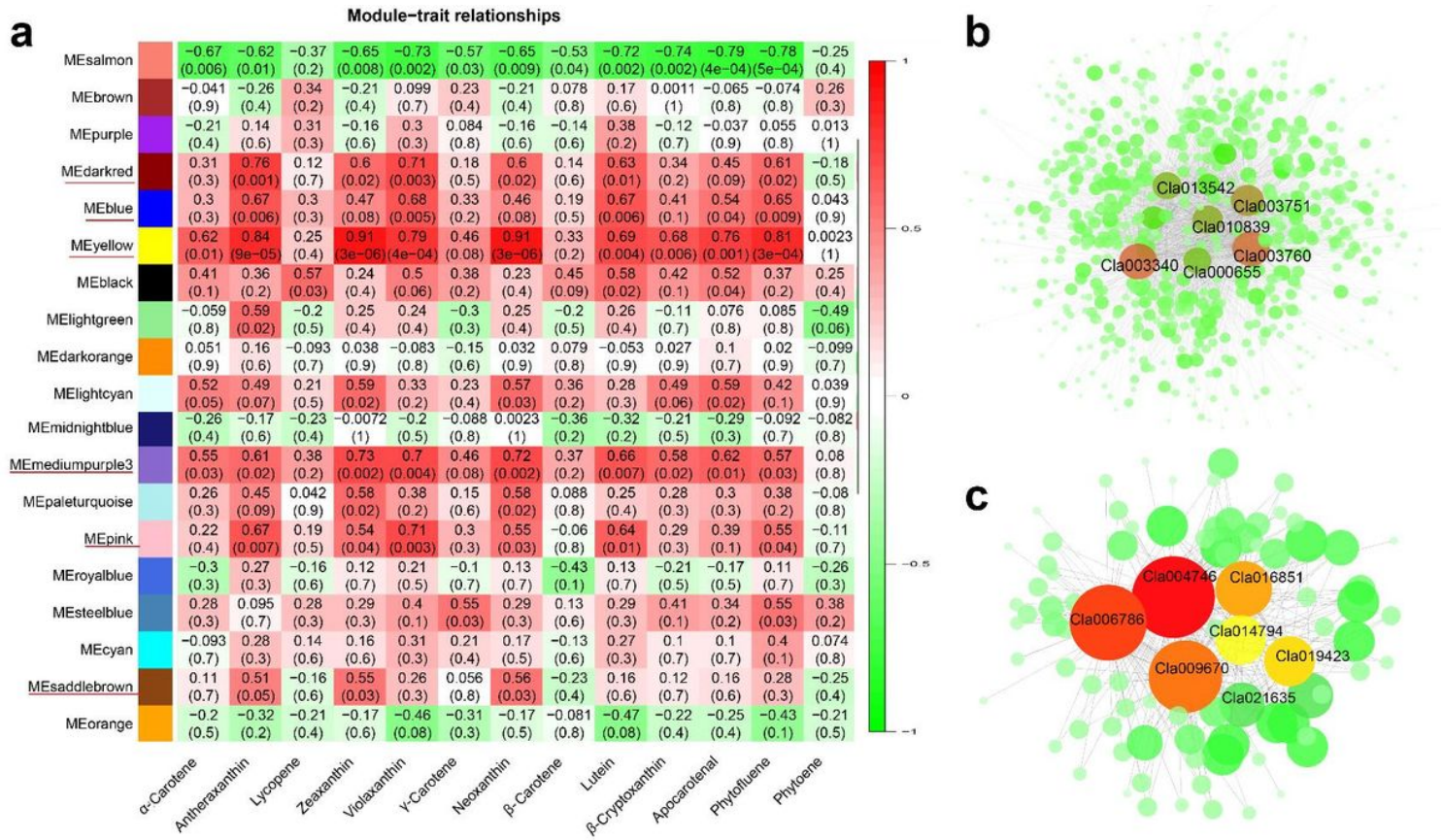

Figure 5

(a) Module-carotenoid relationship and (b, c) correlation network analysis of yellow and darkred module. Each row corresponds to a module, labeled with color as in Additional file 20: Fig. S7a. The value of each cell at the row-column intersection indicates the correlation coefficient between the module and the carotenoid and is displayed according to the color scale on the right. The value in parentheses in each cell represents the P value. Top10 genes with high degree of connectivity and their associated edges are displayed.

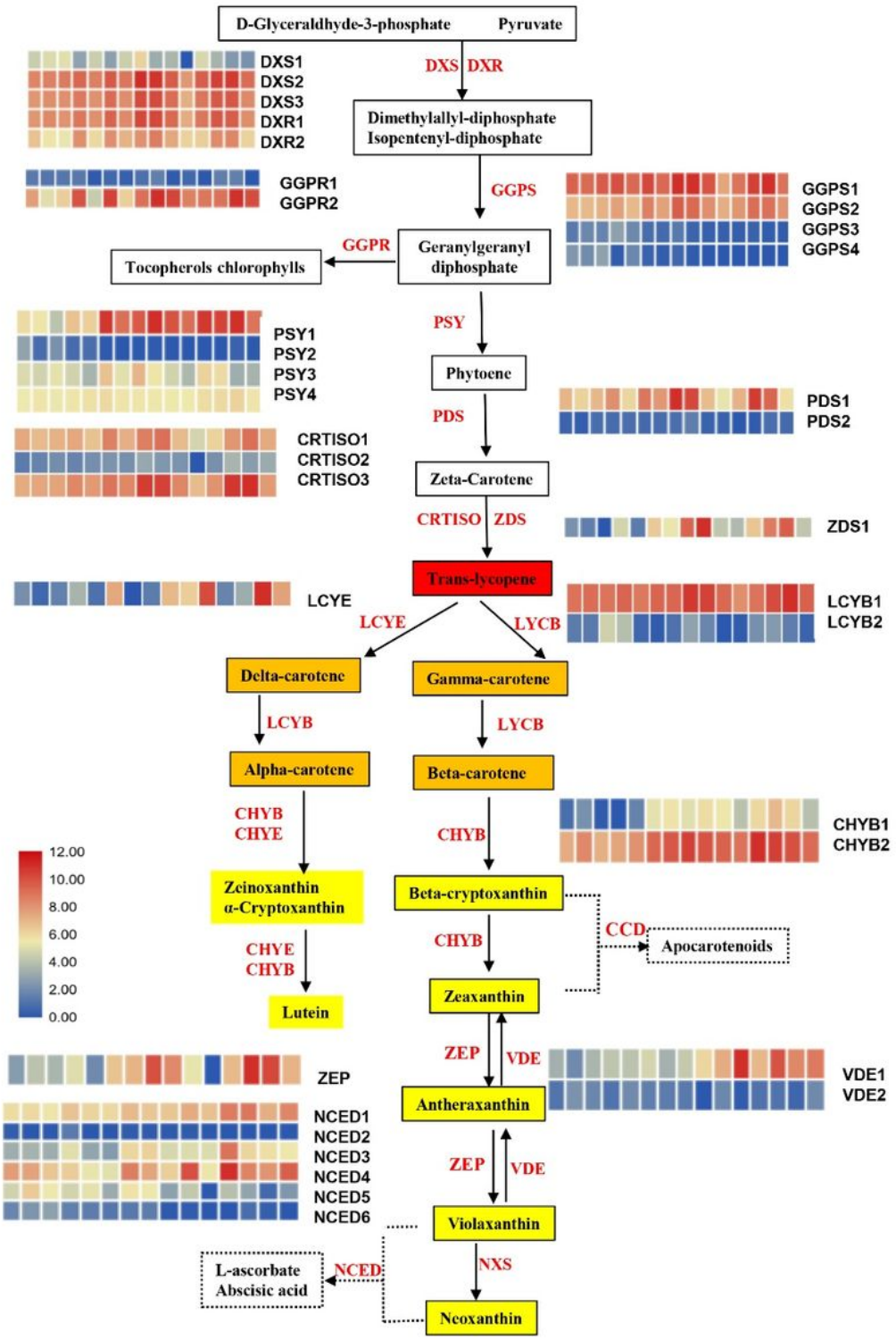

Figure 6

Expression profiles of genes involved in the carotenoid pathway of different flesh-colored watermelons. The FPKM of genes are listed in Supplementary Table S11. The heatmap cell from left to right represents R_10, P_10, O_10, Y_10,W_10, R_20, P_20, O_20, Y_20,W_20, R_34, P_34, O_34, Y_34,W_34. Genotype R, P, O, Y, and W represents the red, pink, orange, yellow, and white fleshed cultivar, respectively; the "_1, _2, _4" represent the sample group at 10, 20, 34 DAP, respectively. The colored cell represents the normalized

gene expression according to the color scale, red represents the high expression, and blue represents the low expression. Metabolites background are colored according to their compound colors, Enzymes are red font. DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; GGPS, geranylgeranyl diphosphate synthase; GGPR, Geranylgeranyl diphosphate reductase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-carotene desaturase; CRTISO, carotenoid isomerase; LCYE, lycopene ε-cyclase; LCYB, lycopene β-cyclase; CHYB, β-carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, Violaxanthin de-epoxidase; NXS, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydataset5202006yuan.xlsx
- Supplementarydataset4202006yuan.xlsx
- Supplementarydataset3202006yuan.xlsx
- Supplementarydataset2202006yuan.xlsx
- Supplementarydataset1202006yuan.xlsx
- Supplementarytablesforwatermelonfleshcolor20200630yuan.xlsx
- SupplementrayFig202007.pdf