

# Transcriptomic and metabolomic analysis provides insights into anthocyanin and pyocyanin accumulation in pear

Shouqian Feng (✉ [shouqianlove@sdau.edu.cn](mailto:shouqianlove@sdau.edu.cn))

Shandong Agricultural University

Zhen Zhang

Shandong Agricultural University

Changping Tian

Yantai Agricultural Institute

Ya Zhang

Shandong Agricultural University

Chenzhiyu Li

Shandong Agricultural University

Xi Li

Shandong Agricultural University

Qiang Yu

Yantai Agricultural Institute

Shuo Wang

Shandong Agricultural University

Xinyu Wang

Shandong Agricultural University

Xuesen Chen

Shandong Agricultural University

---

## Research article

**Keywords:** anthocyanin, procyanidin, metabolome, transcriptome, pear, PcGSTF12

**Posted Date:** December 12th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.18800/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Version of Record:** A version of this preprint was published at BMC Plant Biology on March 27th, 2020.  
See the published version at <https://doi.org/10.1186/s12870-020-02344-0>.

# Abstract

## Background

Pear is one of the most important fruit crops worldwide. Anthocyanins and procyanidins (PAs) are important secondary metabolic substances that affect the appearance and nutritive quality of pear. However, few studies have focused on the molecular mechanism underlying anthocyanin and PA accumulation in pear.

## Results

We conducted metabolome and transcriptome analyses to identify candidate genes involved in anthocyanin and PA accumulation in young pear fruits of 'Clapp Favorite' (CF) and its red mutation variety 'Red Clapp Favorite' (RCF). Gene–metabolite correlation analyses revealed a core set of 20 genes for pear anthocyanin and PA accumulation. One gene in the core set, PcGSTF12, was confirmed to be involved in anthocyanin and PA accumulation by complementation of the tt19-7 Arabidopsis mutant. Interestingly, PcGSTF12 was found to be responsible for procyanidin A3 accumulation, but not petunidin 3, 5-diglucoside accumulation, opposite to the function of AtGSTs in Arabidopsis. Transformation with PcGSTF12 greatly promoted or repressed genes involved in anthocyanin and PA biosynthesis, regulation, and transport. Electrophoretic mobility shift and luciferase reporter assays confirmed positive regulation of PcGSTF12 by PcMYB114.

## Conclusion

These findings identify a core set of genes for pear anthocyanin and PA accumulation. Of these, PcGSTF12, was confirmed to be involved in anthocyanin and PA accumulation. Furthermore, an important anthocyanin and PA regulation node was constructed by two core genes, PcGSTF12 and PcMYB114. These results provide novel insights into anthocyanin and PA accumulation in pear and represent a valuable data set to guide future functional studies and pear breeding.

## Background

Anthocyanins and procyanins (PAs) are the two largest classes of flavonoids. To date, more than 635 anthocyanins and 680 PAs have been characterized in numerous plant species [1]. Anthocyanins and PAs are abundant in colored plant tissues, and contribute to the color, taste, and nutrition of many fruits, vegetables, and teas [2, 3].

Anthocyanins and PAs are initially biosynthesized from phenylalanine, and share most steps in the flavonoid biosynthetic pathway. They are biosynthesized in the cytosol by enzymes including phenylalanine ammonia-lyase (PAL), chalcone isomerase (CHI), chalcone synthase (CHS), flavonoid 3-hydroxylase (F3H), and dihydroflavonol reductase (DFR) [4]. Leucoanthocyanidins and anthocyanidins are two important branch points between the anthocyanin and PA biosynthesis pathways. Downstream

of these branch points, anthocyanins are synthesized by anthocyanidin synthase (ANS) and UDP-glucose flavonoid 3-Oglucosyl transferase (UFGT), and PAs are synthesized by leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR) [5, 6]. O-methyltransferase (OMT) and glycosyltransferase (GT) are responsible for the elaboration of diverse anthocyanin and PA compositions [7, 8].

After anthocyanins and PAs are synthesized in the cytosol, they are transported to their final destination, the vacuole. Although much is known about anthocyanin and PA synthesis, less is known about the transport machinery for these compounds and their precursors. Three models have been proposed for the sequestration of anthocyanins in vacuoles: (I) direct transport of non-modified anthocyanins [9]; (II) transport after their modification [10]; and (III) transport mediated by a carrier protein [11]. Some glutathione transferases (GSTs) are believed to act as anthocyanin carrier proteins through binding to anthocyanins to sequester them into vacuoles [11].

Several studies have revealed details of the vacuolar sequestration mechanisms for PAs. Some PA-deficient *Arabidopsis* mutants have been isolated, such as *tt12*, *tt18*, and *tt19* [12]. The putative constituents of the transport machinery for PA precursors have also been identified. *TT12* encodes a putative MATE-type transporter, which is at least partly responsible for the vacuolar sequestration of PA precursors but not anthocyanins [13]. *TT18* encodes a putative LDOX that catalyzes a biosynthetic step common to both the anthocyanin and PA synthesis pathways [14, 15]. *TT19* encodes a putative GST-type transporter that is responsible for the vacuolar sequestration of PA precursors and anthocyanins [12]. Homologs of the known anthocyanin and PA-related GST and MATE genes have been isolated from several plants and characterized. These genes show extensive sequence and functional diversification [12, 16]. Thus, the proteins involved in the vacuolar sequestration of anthocyanins and PAs in pear have not yet been identified, and the steps involved in anthocyanin and PA sequestration remain understudied.

Anthocyanins and PAs are transcriptionally regulated by transcription factors (TFs). The MBW complex, comprising a R2R3-MYB TF, a bHLH TF, and a WD40 TF, plays a key role in controlling anthocyanin and PA accumulation. Among them, the R2R3-MYB TF determines the specific recognition of downstream target anthocyanin- and PA-related genes. Many R2R3-MYB TFs have been identified in various plants, including anthocyanin-related R2R3-MYB TFs such as *PAP1* in *Arabidopsis*, *MdMYB1* in apple, and *VvMYBA1* in grape [17, 18, 19]. Some R2R3-MYB TFs regulate both anthocyanins and PAs, including *VvMYB5a* and *VvMYB5b* in grape, and *MdMYB9* and *MdMYB11* in apple, *MYB2* in *Medicago truncatula*, and *PpMYB18* in peach [20, 21, 22, 23, 24]. Other R2R3-MYB TFs are related to the regulation of PAs, including *AtTT2* in *Arabidopsis*, *VvMYBPA1* in grape, *PpMYBPA1* in peach, and *PtMYB134* in poplar [20, 25, 26, 27]. The R2R3-MYB TFs bind to promoter motifs and determine the specificity of gene transcription for anthocyanin and/or PA biosynthesis. In addition to bHLH and WD40, other TFs such as JAZ, AUX, and ERF also regulate anthocyanin/PA biosynthesis by directly or indirectly interacting with R2R3-MYB TFs and structural genes [22, 28, 29, 30, 31].

Pear is an important economically important fruit in temperate regions. With the continuous improvement of living standards, consumers are paying increasing attention to fruit quality. Since anthocyanins and

PAs are important compounds determining pear fruit quality, their regulation mechanisms have attracted much attention from researchers. Recent studies have advanced our understanding of the molecular mechanisms involved in anthocyanin biosynthesis in pear. Anthocyanin related genes, such as the structural genes PAL, CHS, CHI, F3H, DFR, ANS, UFGT, ANR, and the TF genes ERF, R2R3-MYB, bHLH, and WD40 have been isolated and identified from pear [32, 33]. Pear has complex patterns of anthocyanin regulation. Anthocyanin biosynthesis may be subject to differential regulation mechanisms, and may be controlled by the different co-expression of anthocyanin-related genes [34]. Compared with anthocyanins, less is known about PA regulation in pear.

A series of anthocyanin and PA metabolites have been identified in pear. For example, five anthocyanins (cyanidin 3-galactoside, cyanidin 3-glucoside, cyanidin 3-arabinoside, peonidin 3-galactoside, and peonidin 3-glucoside) and two PAs (procyanidin B1 and procyanidin B2) have been identified in the red pear cvs. 'D'Anjou' [35] and 'Red Zaosu' [36]. With progress in new metabolic technologies, many more anthocyanins and PAs may be characterized in pear. Considering the large number of anthocyanin and PA varieties, the molecular mechanisms underlying anthocyanin and PA metabolites in pear might be more complex than expected. However, the regulatory networks controlling anthocyanin and PA accumulation in pear are still poorly understood.

Recent new technical advancements in transcriptome and metabolome analysis have provided effective ways to identify new genes and metabolites, and to elucidate complex bioprocesses in plants. Integrated transcriptome and metabolome analyses have revealed details of developmental changes and complex responses to environmental factors and abiotic stresses [37, 38, 39]. Recent studies using combined transcriptome and metabolome datasets successfully constructed expression–metabolite networks in potato [1], but there are no similar reports on pear.

Red pear mutations are rich in anthocyanins and PAs. They are ideal materials to study anthocyanin and PA accumulation because of their highly similar genetic backgrounds [32]. 'RCF' is a typical red pear sport of 'CF' that was discovered in the USA. The fruit of 'RCF' is initially green, then changes quickly to red within 1 week after full bloom, and remains red until the fruit ripens. The coloration pattern of 'RCF' differs from that of most pear species, which color at the ripening stage. Thus, analyzing the differences in anthocyanins, PAs, and gene expression in young fruits between 'CF' and 'RCF' may provide insights into the molecular mechanism of anthocyanin and PA accumulation in pear. Here, we carried out metabolome and transcriptome analyses to identify candidate genes involved in anthocyanin and PA accumulation in pear using the young fruits of 'CF' and 'RCF'. We identified a core set of 20 candidate genes for pear anthocyanin and PA accumulation, including seven genes with previously characterized roles in anthocyanin accumulation. One member of the core set, the GST gene PcGSTF12, was confirmed to play a role in anthocyanin and PA accumulation in pear by functional complementation analysis. PcGSTF12 was found to be directly and positively regulated by PcMYB114, a well-known TF regulating anthocyanin accumulation in pear. However, the function of PcGSTF12 was found to differ from that of known anthocyanin- and PA-related GSTs in *Arabidopsis*. These results greatly extend our knowledge of the mechanism of anthocyanin/PA accumulation in pear.

## Results

### Anthocyanin and PA profiles of 'CF' and its red mutant 'RCF'

No significant morphological differences were observed between the fruits of 'CF' and its red mutant 'RCF', except for skin color. The young fruits of both 'CF' and 'RCF' initially had a deep green appearance. The color difference between 'CF' and 'RCF' became visible from about 6 DFAB. The 'RCF' fruits quickly turned dark red, and retained their strong color until maturity. In contrast, the fruits of 'CF' only developed a slight red blush on the sun-exposed surface (Fig. 1A).

We identified and quantified the individual anthocyanins and PAs in 'CF' and its red mutant 'RCF'. We identified and quantified 12 anthocyanins and seven PAs from 'CF1', 'RCF1', 'CF2', and 'RCF2'. The anthocyanin metabolites included pelargonin, cyanidin 3-rutinoside, pelargonidin 3-Glu, malvidin-3-galactoside chloride, cyanidin 3-O-malonylhexoside, oenin, delphinidin, peonidin O-hexoside, cyanidin, and rosinidin O-hexoside; and the PAs included procyanidin A, procyanidin B3, procyanidin B, procyanidin A1, and procyanidin A2, which were all detected for the first time in pear. As shown in Fig. 1B, most anthocyanins and PAs were significantly up-regulated in 'RCF' compared with 'CF'. The levels of anthocyanins and PAs in 'CF' and 'RCF' were initially low and then sharply increased during fruit coloration, except for delphinidin and rosinidin O-hexoside (Fig. 1B). These results were consistent with the strikingly different fruit color phenotypes of 'RCF' and 'CF'.

The thresholds for significant differences in metabolite levels between the two cultivars were variable importance in projection (VIP) value  $\geq 1$  and  $\log_2(\text{fold change}) \geq 1$ . Against these criteria, six, eight, six, and eight anthocyanin metabolites, and four, seven, seven, and seven PA metabolites were significantly differentially accumulated in the four comparison groups: 'RCF1' vs. 'CF1', 'RCF2' vs. 'CF2', 'RCF2' vs. 'RCF1', and 'CF2' vs. 'CF1', respectively (Supplementary Table S2). Therefore, these anthocyanins and PAs were selected for further metabolite and transcript correlation analyses.

### Transcriptome analysis

The RNA-seq analysis yielded 95.6 G clean bases and 637 million clean reads. The mean number of clean reads per sample was 53 million. Of the clean reads, 93.54% were mapped in total, and 90.72% were mapped uniquely against the improved apple reference genome sequence. In total, 14514 genes were expressed with FPKM  $\geq 10$  (Supplementary Table S3).

We identified 4065 DEGs in the four comparison groups. There were 340 DEGs in 'RCF1' vs. 'CF1', 252 in 'RCF2' vs. 'CF2', 2379 in 'RCF2' vs. 'RCF1', and 3055 in 'CF2' vs. 'CF1' (Supplementary Table S4): in those comparison groups, 123, 155, 858, and 1060 genes were up-regulated, and 217, 97, 1521, and 1995 genes were down-regulated, respectively (Fig. 2A). Only 12 DEGs were common to all four comparison groups (Fig. 2B).

The DEGs between group 2 – 1 and group RCF-CF were subjected to GO (Supplementary Table S5) and KEGG functional pathway analyses (Supplementary Table S6). For correlation tests with anthocyanins and PAs, we chose DEGs in group 2 – 1 categorized into DNA binding, plant hormone signal transduction, flavonoid biosynthesis, phenylpropanoid biosynthesis, flavonoid metabolism, phenylalanine metabolism, glutathione metabolism, and DEGs in group RCF-CF categorized into DNA binding, plant hormone signal transduction, phenylpropanoid biosynthesis, flavonoid biosynthesis, glutathione metabolism, and phenylalanine metabolism (Fig. 2C, D). In total, we selected 203 DEGs. Of these, there were 12 DEGs that overlapped between group 2 – 1 ('RCF2' vs. 'RCF1', and 'CF2' vs. 'CF1') and group RCF-CF ('RCF1' vs. 'CF1', 'RCF2' vs. 'CF2'). These 12 DEGs encoded a  $\beta$ -glucosidase (PCP011059), two peroxidases (PCP024451 and PCP017906), a CHS (PCP023048), a homeodomain protein (PCP024513), an ERF (PCP044584), a RBR (retinoblastoma-related protein, PCP007207), an AUX (PCP036703), a NAC (PCP028501), a GST (PCP025171), and two SAURs (PCP037299 and PCP040169) (Fig. 2E). Of these, PcGST (PCP025171) was the most up-regulated gene in the comparison groups 'RCF1 vs. CF1' and 'RCF2 vs. CF2' (Fig. 2F).

## Correlation analysis between selected transcripts and anthocyanins/PAs

To identify the candidate genes in anthocyanin and PA accumulation in pear, we conducted correlation analyses between selected transcripts and metabolites. In total, we detected 420 significant correlations (correlation coefficient,  $R^2 > 0.8$ ) between 203 transcripts and 17 metabolites, including 10 anthocyanins (kuromanin chloride, pelargonin, cyanidin 3-rutinoside, pelargonidin 3-Glu, malvidin-3-galactoside chloride, cyanidin 3-O-malonylhexoside, oenin, delphinidin, peonidin O-hexoside, cyanidin, and rosinidin O-hexoside) and seven PAs (procyanidin A, procyanidin A1, procyanidin A2, procyanidin B, procyanidin B1, procyanidin B2, and procyanidin B3) (Supplementary Table S7). Each metabolite was correlated with many different transcripts. Malvidin-3-galactoside chloride, oenin, and delphinidin were correlated with the fewest transcripts: five, seven, and three transcripts, respectively. Kuromanin chloride and pelargonin were correlated with the highest numbers of transcripts: 184 and 56 transcripts, respectively. Interestingly, kuromanin chloride and pelargonin shared the largest number of common transcripts: 36 transcripts. This suggested that kuromanin chloride and pelargonin might have evolved similar accumulation mechanisms.

The 203 transcripts were annotated with descriptions from the SwissProt and NR databases. Six transcripts have been functionally characterized to play roles in anthocyanin accumulation in pear previously: PcMYB10, PcMYB114, PcCHS, PcCHI, PcF3H, and PcANS (Supplementary Table S7). The rest were newly identified as candidate genes involved in anthocyanin and PA accumulation in pear. The 203 transcripts were grouped into two clusters (I-II) (Supplementary Table S7). Genes in cluster I were strongly correlated with anthocyanins. Cluster I comprised 183 genes (90.1%). Of these, 147 genes were correlated with a single anthocyanin: 142 genes were correlated with kuromanin chloride, three genes were correlated with cyanidin, one gene was correlated with malvidin-3-galactoside chloride, and one gene was correlated with pelargonin. The remaining genes in cluster I were closely correlated with two or more

anthocyanins: 31 genes were commonly correlated with kuromanin chloride and pelargonin, two genes were correlated with pelargonin and cyanidin, two genes were correlated with malvidin-3-galactoside chloride and oenin, and one gene was correlated with cyanidin 3-rutinoside, oenin, and cyanidin. Cluster II contained 20 genes (9.9%) that were strongly correlated with both anthocyanins and PAs. Of these genes, we found two phenylpropanoid structural genes (encoding 4CL1 and 4CL2), six flavonoid structural genes (encoding CHS, 3 CHIs, F3H, and ANS), six TF genes (encoding bZIP1, MYB3, MYB86, MYB111, MYB114 and KNAT1), two phytohormone signal transduction genes (encoding IAA13 and ERF003), two DNA-directed RNA polymerase genes (encoding rpoB and Rpb1) and one GST transporter gene (encoding GSTF12) that were positively correlated with anthocyanins and PAs. One WRKY TF gene, WRKY28, was negatively correlated with anthocyanins and PAs (Supplementary Table S7, S8). Each gene in cluster II was strongly correlated with many metabolites. We found that these 20 genes were strongly correlated with all 17 anthocyanin and PA metabolites (Fig. 3A). Thus, they represent the core genes for anthocyanin and PA accumulation in pear. Of these, PcRBP1 (PCP004386) was correlated with the fewest metabolites: one anthocyanin and three PAs; and PcGSTF12 (PCP025171) was correlated with the most metabolites: seven anthocyanins and seven PAs (Fig. 3B).

## qPCR analysis of DEGs related to anthocyanin and PA accumulation

To validate the RNA-seq data, we conducted qPCR analyses of 10 of the anthocyanin and/or PA candidate genes (for primers, see Supplementary Table S1). The transcript profiles of all select genes were very similar to those detected from the RNA-seq data (Fig. 4). The results showed that PcGSTF12 was most up-regulated in comparison groups 'RCF1 vs. CF1' and 'RCF2 vs. CF2'. This result was highly consistent with the results of RNA-seq, and suggested that it may play a core role in anthocyanin and PA accumulation in pear. Thus, we conducted further analyses to confirm the function of PcGSTF12.

## PcGSTF12-mediated anthocyanin and PA accumulation in pear

Our combined metabolite and transcriptomic analyses revealed a core set of genes closely correlated with pear anthocyanins and PAs, which strongly suggested that they play key roles in anthocyanin and PA accumulation in pear. To test this, we focused on the most up-regulated gene among the core set of anthocyanin and PA candidate genes, PcGSTF12, for functional analysis.

### 1. Functional analysis of PcGSTF12

The phylogenetic analysis showed that PcGSTF12 is a homolog of FvRAP in strawberry, Riant2 in peach, and MdGST in apple, all of which are in the phi subfamily [40] (Fig. 5A). Members of the phi subfamily are anthocyanin transporters. To test the potential role of PcGSTF12 in anthocyanin accumulation, 35S::

PcGSTF12 was transformed into the Arabidopsis mutant tt19-7 (for primers, see Supplementary Table S1). The tt19-7 plants showed a green hypocotyl phenotype, while the tt19-7-OE transgenic plants showed the red hypocotyl phenotype, like that of the wild type (WT) (Fig. 5B). However, the brown color of seed coats was not rescued in the tt19-7-OE lines (Fig. 5B). This result was consistent with the fresh seed phenotype obtained by transferring 35S::RAP-RFP into Arabidopsis tt19-7 [40].

To explore the role of PcGSTF12 in anthocyanin and PA accumulation, we conducted a metabolite analysis using Arabidopsis seedlings. Three PAs and nine anthocyanins were significantly up-regulated, and three anthocyanins were significantly down-regulated in tt19-7-OE compared with tt19-7. Procyanidin A3, cyanidin O-acetylhexoside, delphinidin 3-O-rutinoside, and cyanidin 3-p-hydroxybenzoylsophoroside-5-glucoside were specifically up-regulated in tt19-7-OE compared with tt19-7 (Fig. 5C). Five anthocyanins: malvidin 3-acetyl-5-diglucoside, pelargonidin 3-O-beta-D-glucoside, delphinidin 3-O-rutinoside, pelargonin, cyanidin 3-O-galactoside, and one PA: procyanidin A3 were significantly up-regulated, and two anthocyanins: petunidin 3, 5-diglucoside and delphinidin O-malonyl-malonylhexoside were significantly down-regulated in tt19-7-OE compared with WT. Interestingly, a large amount of petunidin 3, 5-diglucoside was detected in WT but not in tt19-7-OE. In contrast, procyanidin A3 was detected only in tt19-7-OE. These results confirmed that PcGSTF12 is responsible for anthocyanin and PA accumulation. Interestingly, its affinity for anthocyanins and PAs differed from that of AtGSTs in Arabidopsis. In particular, PcGSTF12 was responsible for the accumulation of procyanidin A3 but not petunidin 3, 5-diglucoside, opposite to the function of AtGSTs in Arabidopsis. PcGSTF12 is a newly identified member of the phi GST family involved in anthocyanin and PA accumulation.

Next, we analyzed RNA-seq data to identify which genes were affected by PcGSTF12 in the seedlings of tt19-7-OE vs. tt19-7. In total, we found 28 strongly affected genes, which encoded proteins involved in anthocyanin and PA biosynthesis, regulation, and transport (Supplementary Table S9). These results showed that PcGSTF12 might not only be an anthocyanin and PA transporter, but may also participate in many other steps of anthocyanin and PA accumulation.

## 2. Upstream regulation of PcGSTF12

Correlation analyses showed that the transcript level of PcGSTF12 was significantly correlated with that of PcMYB114 (Fig. 5D). Further, several MYB-binding sites were found in the PcGSTF12 promoter, indicating that PcGSTF12 might be directly bound by, and regulated by, MYB transcription factors (Supplementary Table S9). Several R2R3-MYB genes are known to bind to MBS sites (), and we found a MBS site within an 801-bp region upstream of the start codon. Thus, this MBS site was used in an EMSA assay (for primers, see Supplementary Table S1). The biotinylated probe was able to bind PcMYB114 protein, and the addition of a high concentration of cold probe significantly reduced the binding affinity of the biotinylated probe. To test whether PcGSTF12 could be regulated by PcMYB114, we further carried out luciferase reporter assay (for primers, see Supplementary Table S1). The relative LUC activity of the PcGSTF12 promoter was about 6-fold that of the empty control. These results showed that PcMYB114

could directly bind to the MBS site in the PcGSTF12 promoter (Fig. 5E) and positively regulate its activity (Fig. 5F).

We also found many cis-acting elements involved in auxin-, ethylene-, and gibberellin-signaling in the PcGSTF12 promoter. These results indicated that PcGSTF12 might be the common downstream target of R2R3-MYBs, and auxin, ethylene, and gibberellin signals that regulate the anthocyanin and PA pathways (Supplementary Table S10). Together, these results provide new clues about PcGSTF12-mediated anthocyanin and PA accumulation in pear.

## Discussion

We present the first genome-wide examination of anthocyanins, PAs and the gene expression profiles of pear using young fruits of cv. 'CF' and its red mutant 'RCF'. Through combined transcriptomic and metabolic analyses, we found a core set of 20 candidate genes related to anthocyanin and PA accumulation in pear. These findings increase our understanding of the molecular mechanism of anthocyanin and PA accumulation in pear, especially during the early stage of fruit development.

The core set of candidate genes for pear anthocyanin and PA accumulation included six flavonoid structural genes: PcCHI (PCP027877, PCP021141, and PCP039844), PcCHS (PCP023048), PcF3H (PCP013732), and PcANS (PCP027029). CHI, CHS, and F3H are common to both the anthocyanin and PA biosynthetic pathways [4]. ANS can catalyze the conversion of (+)-catechin to cyanidin and a procyanidin [41]. Functional analyses have confirmed the roles of PcCHI, PcCHS, and PcF3H in anthocyanin accumulation [42]. The results of those studies and our study provide evidence that the patterns of anthocyanin and PA accumulation are conserved among different plants.

In plants, TFs play important roles in flavonoid regulation. The R2R3-MYBs make up one of the largest TF families [43], and most of them play important roles in flavonoid accumulation [44]. PyMYB10 was the first R2R3-MYB TF identified to be involved in anthocyanin accumulation in pear [32]. Functional analyses have confirmed the roles of PyMYB10.1 and PyMYB114 in the regulation of anthocyanin accumulation in pear [33, 45], and the roles of PbMYB10b and PbMYB9 in both anthocyanin and PA accumulation in pear [36]. In this study, the core set of candidate genes for anthocyanin and PA accumulation included four R2R3-MYB genes: PcMYB3, PcMYB86, PcMYB111 and PcMYB114, which were strongly positively correlated with anthocyanins and PAs. This result provides evidence that PcMYB114 functions in anthocyanin accumulation in pear. Also, our results provide a new clue about the function of PcMYB114 in PA accumulation in pear. While PcMYB114 was already known to function in anthocyanin accumulation, PcMYB3, PcMYB86, and PcMYB111 are newly identified candidate genes involved in anthocyanin and PA accumulation in pear. Interestingly, these R2R3-MYBs were correlated with different anthocyanins and PAs, suggesting that they have undergone sub-specialization to play different and specific roles in anthocyanin and PA accumulation in pear.

Other TFs are also involved in anthocyanin and PA accumulation in pear. The HD family of TFs is unique to plants, and its members are proposed to play key roles in developmental processes such as root

development, plant cell differentiation, fruit ripening, and leaf and flower senescence [46, 47, 48]. Members of the HD-Zip I and HD-Zip IV TF subfamilies play key roles in anthocyanin accumulation. ANTHOCYANINLESS2 (ANL2) was the first HD-Zip IV gene found to be involved in the tissue-specific accumulation of anthocyanin. In Arabidopsis, ANL2 affects anthocyanin accumulation in subepidermal tissue on the abaxial side of rosette leaves and in epidermal tissue on the abaxial side of the leaves [49]. Recently, two HD-Zip I genes, MdHB1 and RhHB1 have been shown to influence anthocyanin accumulation in apple and rose, respectively. Overexpression of MdHB1 led to reduced anthocyanin accumulation in apple flesh. MdHB1 was found to repress the transcription of MdDFR and MdUFGT indirectly by interacting with MdMYB10, MdbHLH3, and MdTTG1 [50]. Consistent with the results in apple, silencing of RhHB1 in rose led to higher anthocyanin levels in the petals [51]. In this study, the core set of genes involved in anthocyanin and PA accumulation in pear included an HD TF gene, PckNAT1. KNAT1 is a member of the Class I KNOX HD gene family and is thought to play a role in meristem development and leaf morphogenesis [52]. Our results showed that PckNAT1 was strongly positively correlated with five anthocyanins and seven PAs, implying that its function differs from the known functions of HD-Zip I and HD-Zip IV TFs in anthocyanin accumulation. Our results suggest that the Class I KNOX HD gene family might play important roles in anthocyanin and PA accumulation; this expands our knowledge of the function of the Class I KNOX HD gene family.

In our study, a bZIP TF gene, PcbZIP1, was positively and closely correlated with six anthocyanins and five PAs. The bZIP TFs harbor a highly conserved bZIP domain [53]. They are diverse transcriptional regulators, playing crucial roles in plant development, physiological processes, and biotic/abiotic stress responses [54]. Recently, two bZIP TFs, MdHY5 and MdbZIP44, were shown to promote anthocyanin accumulation in apple [55, 56]. Our results provide further evidence that bZIP TFs act as positive regulators in anthocyanin accumulation. Thus, as well as functioning in anthocyanin accumulation, bZIP TFs may also play important roles in PA accumulation.

A WRKY TF gene, WRKY28, was closely negatively correlated with one anthocyanin and four PAs. WRKY41-1 in Brassica napus and WRKY75 in Arabidopsis are known repressors of anthocyanin biosynthesis [57, 58], while MdWRKY40 in apple is a positive regulator in wounding-induced anthocyanin biosynthesis [59]. Our results were consistent with findings in Brassica napus and Arabidopsis and opposite to those in apple. It is possible that WRKYs have evolved different functions in anthocyanin accumulation, like the R2R3-MYB TFs. For example, some R2R3-MYB TFs are positive regulators of anthocyanin accumulation [32, 33], while others are negative regulators [24]. Further studies are required to elucidate the complex roles of WRKYs in PA accumulation in fruit trees and other plants.

Auxin is known to suppress anthocyanin accumulation, and has been shown to decrease the expression of anthocyanin regulatory and structural genes in apple and Arabidopsis [60, 61]. A recent study showed that the auxin factor MdARF13 negatively regulates the anthocyanin pathway in apple through interacting with MdMYB10 and binding to the promoter of MdDFR. The auxin/IAA protein MdIAA121 was shown to inhibit the recruitment of MdARF13 to the MdDFR promoter and weaken the inhibitory effect of MdARF13 on anthocyanin accumulation [30]. Contrary to auxin, ethylene enhances anthocyanin and PA

accumulation in pear and apple. In pear, the ethylene-responsive factor PyERF3 enhances anthocyanin accumulation via interacting with PyMYB114 [33]. In apple, MdERF1B regulates anthocyanins and PAs through interacting with MdMYB1, MdMYB9, and MdMYB11 [31]. In this study, we found that an ethylene-responsive gene, PcERF003, was positively correlated with six anthocyanins and seven PAs, and an auxin-responsive gene, PcIAA13, was positively correlated with two anthocyanins and four PAs. These results are consistent with the previous findings that auxin represses anthocyanin accumulation via IAA genes, and ethylene enhances anthocyanin accumulation via ERF genes. Phytohormones play important roles in young fruit development. Young pear fruits contain high level of auxin, and low level of ethylene [62]. Further functional analyses of PcERF003 and PcIAA13 may help to elucidate the links between the effects of auxin and ethylene on anthocyanin and/or PA accumulation, and fruit development in young pear fruit.

Plant GSTs are encoded by a large gene family, and are soluble and abundant in the cytosol [40, 63, 64]. They can be divided into eight subgroups, among which the tau and phi classes play key roles in flavonoid transport [65]. Bronze 2 (bz2) in maize was the first tau class GST reported to be involved in anthocyanin accumulation; bz2 produces yellow skin kernels because of disabled anthocyanin transport into the vacuole [66]. Anthocyanin deposition is also affected by genes in the phi class, such as AtGSTF12 in Arabidopsis [12, 67, 68], FvRAP in strawberry [40], CsGSTF1 in tea [65], and VvGST4 in grapevine [69]. Interestingly, these phi-class GSTs show extensive functional diversification. For example, AtGSTF12 plays a key role in both anthocyanin and PA accumulation in Arabidopsis [67], while CsGSTF1 in tea functions only in anthocyanin accumulation [65]. Functional divergence of GSTs has arisen through nonsynonymous mutations, especially at key amino acid sites [64]. For example, a single amino acid mutation (Arg39 to Trp39) was found to be responsible for the high enzymatic activity of *Populus euphratica* PeGSTU30 [64]. Li et al. [70] showed that one Trp to Leu substitution at amino acid 205 in AtTT19 resulted in an anthocyanin-deficient phenotype in Arabidopsis. Recently, Luo et al. reported that a single nucleotide polymorphism (C to T) in the second exon of FvRAP resulted in a premature stop codon and dramatically lowered the anthocyanin level in the fruit and petiole of strawberry [40]. In our study, a phi-class GST gene, PcGSTF12, was among the core genes for anthocyanin and PA accumulation in pear. PcGSTF12 was strongly associated with most of the studied metabolites in pear: seven anthocyanins and seven PAs. Similar to AtGSTF12, PcGSTF12 was functionally characterized as an anthocyanin and PA carrier. Interestingly, we detected different affinities for anthocyanin and PA between PcGSTF12 and AtGSTs in Arabidopsis. PcGSTF12 plays an opposite role to that of AtGSTs in procyanidin A3 and petunidin 3, 5-diglucoside accumulation. These results suggest that phi-class GSTs have undergone extensive functional diversification during evolution. Furthermore, transformation with PcGSTF12 affected genes encoding proteins involved in anthocyanin and PA biosynthesis, regulation, and transport. This functional analysis of PcGSTF12 deepens our understanding of the roles of phi-class GSTs in anthocyanin and PA accumulation in pear and other plants.

We found that PcMYB114 can positively regulate PcGSTF12 activity by directly binding to the MBS motif. This result provides evidence that PcMYB114 functions both in anthocyanin and PA accumulation in pear via regulating PcGSTF12 activity. This greatly extends previous knowledge of the role of PcMYB114 in

anthocyanin accumulation in pear [33]. Furthermore, we detected many cis-acting elements involved in auxin, ethylene, and gibberellin signaling in the promoter of PcGSTF12. Thus, we propose that PcGSTF12 might be the common downstream target of auxin-, ethylene-, and gibberellin-mediated anthocyanin and PA accumulation pathways.

## Conclusions

In conclusion, we identified a core set of 20 candidate genes for pear anthocyanin and PA accumulation. Most of them are newly identified to play roles in pear anthocyanin and PA accumulation. Furthermore, we functionally characterized a core gene, PcGSTF12, in pear anthocyanin and PA accumulation. Interestingly, we found that its function in anthocyanin and PA accumulation differs from the known functions of GST genes in *Arabidopsis*. In addition, we found that PcGSTF12 is directly positively regulated by the upstream R2R3-MYB, PcMYB114. These results provide novel insights into pear anthocyanin and PA accumulation. The core set of candidate genes in pear anthocyanin and PA accumulation presented here represent a valuable data set to guide future functional studies.

## Methods

### Plant materials

The pear cultivars 'CF' and 'RCF' were cultivated in the experimental orchard of Yantai Academy of Agricultural Science, Shandong province, Yantai, China (37°5'N, 122°1'W). The fruits of 'CF' and 'RCF' were collected in 2017 from 6-year-old trees grafted onto *Pyrus betulaefolia* rootstocks. The different red phenotypes of 'CF' and 'RCF' were visible from a very early stage of fruit development. The red color of 'RCF' fruits was maintained until fruits ripened. The regulation of anthocyanin and PA accumulation at the early developmental stage of fruits is important for fruit coloration. Therefore, we collected fruits of 'CF' and 'RCF' at two early developmental stages for further analyses. Briefly, fruits of 'CF' and 'RCF' with a similar green color were first sampled on 19 April 2017; these samples were collected at 2 days after full bloom (DFAB), and were named 'CF1' and 'RCF1', respectively. Fruits with significant differences in color between 'CF1' and 'RCF' were sampled on 22 April 2017; these samples collected at 12 DFAB were named 'CF2' and 'RCF2', respectively. In each experiment, skins from 100 fruits per replicate were collected. Three independent biological replicates were collected for analyses. The fruit skin samples were immediately frozen in liquid nitrogen and stored at - 80 °C until further metabolite, RNA-sequencing (RNA-Seq), and qPCR analyses.

### Metabolite extraction and separation

Metabolite extraction and separation were carried out following Wang et al. [71]. Briefly, the freeze-dried fruit skin was crushed into powder and then extracted overnight at 4 °C in 1.0 ml 70% aqueous methanol. Following centrifugation at 10,000 g for 10 min, the extracts were filtered before HPLC analysis. The

analytical conditions were as follows, HPLC column: Waters ACQUITY UPLC HSS T3 C18 (1.8  $\mu\text{m}$ , 2.1 mm  $\times$  100 mm) (Waters, Milford, MA, USA); solvent system: water (0.04% acetic acid), acetonitrile (0.04% acetic acid); gradient program: 100:0 V/V at 0 min, 5:95 V/V at 11.0 min, 5:95 V/V at 12.0 min, 95:5 V/V at 12.1 min, 95:5 V/V at 15.0 min; flow rate: 0.40 ml/min; temperature: 40  $^{\circ}\text{C}$ ; injection volume: 5  $\mu\text{l}$ .

## Anthocyanin and PA identification and quantification

Anthocyanins and PAs were identified and quantified as described by Wang et al. [71]. The HPLC eluent was connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS. The LIT and triple quadrupole (QQQ) scans were acquired using a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), API 4500 Q TRAP LC/MS/MS System, equipped with an ESI Turbo ion-spray interface, operating in the positive ion mode and controlled by Analyst 1.6 software (AB Sciex, Redwood City, CA, USA). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 550  $^{\circ}\text{C}$ ; ion spray voltage (IS) 5500 V; ion source gas I (GSI), gas II (GSII), and curtain gas (CUR) set at 55, 60, and 25.0 psi, respectively; collision gas (CAD), high. Instrument tuning and mass calibration were performed with 10 and 100  $\mu\text{mol/L}$  polypropylene glycol solutions in the QQQ and LIT modes, respectively. The QQQ scans were acquired as described for multiple reaction monitoring (MRM) experiments with the collision gas (nitrogen) set to 5 psi. The parameters of declustering potential and collision energy were optimized for each compound. A specific set of MRM transitions was monitored for each period according to the metabolites eluted within that period.

Metabolites were quantified using MRM, and were annotated by comparisons against public databases including MassBank (<http://www.massbank.jp/>), KNAPSACk (<http://kanaya.naist.jp/KNApSACk/>), HMDB (<http://www.hmdb.ca/>), MoToDB (<http://www.ab.wur.nl/moto/>), and METLIN (<http://metlin.scripps.edu/index.php>).

## Total RNA isolation and RNA-Seq analysis

Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and its integrity was evaluated using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The mRNA was purified from high-quality total RNAs using oligo (dT) magnetic beads and then broken into short fragments with fragmentation buffer. The cDNA was synthesized using a cDNA Synthesis Kit (TaKaRa, Dalian, China) and linked to sequencing adapters at both ends. The cDNA libraries were sequenced using the Illumina sequencing system (HiSeq™ 2000, Illumina, San Diego, CA, USA). Clean reads were obtained using the NGS QC Toolkit [72](Patel and Jain 2012). Differential expression analysis was carried out using the DESeq R package (2012). The threshold for significant differential expression was  $P < 0.05$ , and  $|\log_2 \text{fold change}| \geq 1$  was used to identify the differentially expressed genes (DEGs) between two different cDNA libraries. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DEGs were performed using R as described elsewhere [73].

# Integrated metabolome and transcriptome analyses

Pearson's correlation coefficients were calculated between the metabolome and transcriptome data. The coefficients were calculated from  $\log_2$  (fold change) of each metabolite and  $\log_2$  (fold change) of each transcript by the EXCEL program. Correlations with a coefficient of  $R^2 > 0.8$  were selected. Metabolome and transcriptome relationships were visualized using Cytoscape (version 2.8.2).

## qRT-PCR validation

The total RNA samples used for RNA-Seq were also used for cDNA synthesis using the PrimeScript™ RT Reagent Kit (TaKaRa) according to the manufacturer's instructions. The qRT-PCR analyses were conducted as described previously [32] using the primers shown in Supplementary Table 1. PcActin was used as the reference gene. Three biological replicates were analyzed.

## Phylogenetic analysis

The protein sequences were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>) according to relevant accession numbers or were obtained from the literature. Sequence alignments were performed using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo>). An unrooted phylogenetic tree was constructed using MEGA 7 (<http://www.megasoftware.net/>) with the neighbor-joining statistical method and bootstrap analysis (1000 replicates).

## PcGSTF12 promoter analysis

The 2000-bp upstream sequence of PcGSTF12 was analyzed using tools at the PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and PLACE Signal Scan Search databases (<https://www.dna.affrc.go.jp/PLACE/>). The functional motifs are listed in Supplementary table 10. A MYB-binding site (MBS) sequence CAACTG was found at - 801 bp in the promoter of PcGSTF12 by PlantCARE. This MBS element is known to be a binding sequence of R2R3-MYBs [74]. Therefore, we selected this MBS element sequence to further analyze the interaction between PcMYB114 and the promoter of PcGSTF12 using electrophoretic mobility shift assays (EMSA) as described below.

## Electrophoretic mobility shift assays

The EMSAs were performed using a LightShift Chemiluminescent EMSA Kit (Thermo Scientific, Waltham, MA, USA). Oligonucleotide probes were synthesized and labeled with biotin at the 3' ends (Invitrogen, Carlsbad, CA, USA). The full-length PcMYB114 coding sequence (CDS) was cloned into the expression vector pET32a (EMD Biosciences, Novagen, San Diego, CA, USA). The PcMYB114 recombinant protein was expressed in *Escherichia coli* strain BL21 (Tiangen, Beijing, China) and purified using a Ni-agarose

His-Tagged Protein Purification Kit (CWbiotech, Beijing, China). All probes were synthesized and labeled by the Sangon Biotech Co. Ltd. (Shanghai, China). Double-stranded probes were synthesized using annealing buffer (Beyotime, Shanghai, China). Briefly, biotin-labeled probes were incubated in 1 × binding buffer containing 2.5% glycerol, 5 mM MgCl<sub>2</sub>, 50 mM KCl, and 10 mM EDTA with or without proteins at room temperature for 20 min. Unlabeled probe was added to the reactions for unlabeled probe competition.

## Luciferase reporter assay

The CDS of PcMYB114 was recombined into the pHBT-AvrRpm1 effector. Each promoter fragment of PcGSTF12 was inserted into the pFRK1-LUC-nos reporter. Protoplasts derived from 'Orin' apple callus were used as the materials for transient transfection. Each transfection assay contained the GUS plasmid for normalization. For transient transfection, 1 μl GUS plasmid, 3 μl LUC reporter, and 6 μl effector were mixed and transformed into apple protoplasts using 40% polyethylene glycol. After incubation at 24 °C for 6 h, the LUC and GUS activities were detected by measuring absorbance using a Multimode Plate Reader (Victor X4, PerkinElmer, <http://www.perkinelmer.com/>). The promoter activity was calculated as the ratio of LUC to GUS activity.

## Ectopic expression of PcGSTF12 in Arabidopsis

For gene transformation, the CDS of PcGSTF12 was recombined into the pRI101-AN vector and then transformed into *Agrobacterium tumefaciens* GV3101 using the floral-dip method [75]. The T1 transgenic plants were selected on half-strength Murashige and Skoog (MS) solid medium with kanamycin. Kanamycin-resistant seedlings were grown in soil in a light incubator under a 16-h light/ 8-h dark photoperiod at 24 °C. Seven-day-old T2 seedlings were used for RNA-seq analysis and for anthocyanin and PA identification and quantification.

## Abbreviations

ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; CF, Clapp Favorite; CHI, chalcone isomerase; CHS, chalcone synthase; DEGs, differentially expressed genes; DFR, dihydroflavonol reductase; F3H, flavonoid 3-hydroxylase; GSTs, glutathione S-transferases; LAR, leucoanthocyanidin reductase; PAs, procyanidins; RCF, Red Clapp Favorite; TFs, transcription factors; UFGT, UDP-glucose flavonoid 3-Oglucosyl transferase); WT, wild-type.

## References

1. Cho K, Cho KS, Sohn HB, Ha IJ, Hong SY, Lee H, et al. Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J Exp Bot.* 2016; 67: 1519-33.
2. Lesschaeve I, Noble AC. Polyphenols: factors influencing their sensory properties and their effects on food and beverage preferences. *Am J Clin Nutr.* 2005; 81: 330-35.

3. Pang Y, SAbeysinghe ISB, He J, He XZ, Huhman D, Mewan M, et al. Functional characterization of proanthocyanidin pathway enzymes from tea (*Camellia sinensis*) and their application for metabolic engineering. *Plant Physiol.* 2013; 161: 1103-16.
4. Koes R, Verweij W, Quattrocchio F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 2005; 10: 236-42.
5. Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR. Proanthocyanidin biosynthesis in plants: purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *J Biol Chem.* 2003; 278: 31647-56
6. Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science.* 2003; 299: 396–99
7. Lam KC, Ibrahim RK, Behdad B, Dayanandan S. Structure, function, and evolution of plant *O*-methyltransferases. *Genome.* 2007; 50: 1001-13.
8. Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiol.* 2010; 153: 1729-46.
9. Klein M, Weissenböck G, Dufaud A, Gaillard C, Kreuz K, Martinoia E. Different energization mechanisms drive the vacuolar uptake of a flavonoid glucoside and a herbicide glucoside. *J Biol Chem.* 1996; 271: 29666-71.
10. Hopp W, Seitz HU. The uptake of acylated anthocyanin into isolated vacuoles from a cell suspension culture of *Daucus carota*. *Planta.* 1987; 170: 74-85.
11. Mueller LA, Goodman CD, Silady RA, Walbot V. AN9, a petunia glutathione *S*-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol.* 2000; 123: 1561-70.
12. Kitamura S, Shikazono N, Tanaka A. *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* 2004; 37: 104-14.
13. Debeaujon I, Peeters AJM, Léon-Kloosterziel KM, Koornneef M. The *TRANSPARENT TESTA 12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell.* 2001; 13: 853-71.
14. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 2001; 126: 485-93.
15. Shikazono N, Yokota Y, Kitamura S, Suzuki C, Watanabe H, Tano S, et al. Mutation rate and novel tt mutants of *Arabidopsis thaliana* induced by carbon ions. *Genetics.* 2003; 163: 1449-55.
16. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* 2015; 20: 576-85.
17. Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell.* 2000. 12: 2383-94.
18. Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP Walker AR. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 2006; 142: 1216-32.

19. Walker AR, Lee E, Bogs J, McDavid DA, Thomas MR, Robinson SP. White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J.* 2007; 49: 772-85.
20. Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, et al. Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiol.* 2006; 140: 499-511.
21. Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, et al. The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiol.* 2008; 147: 2041-53.
22. An XH, Tian Y, Chen KQ, Liu XJ, Liu DD, Xie XB, et al. MdMYB9 and MdMYB11 are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. *Plant Cell Physiol.* 2015; 56: 650-62.
23. Jun JH, Liu C, Xiao X and Dixon RA. The transcriptional repressor MYB2 regulates both spatial and temporal patterns of proanthocyanidin and anthocyanin pigmentation in *Medicago truncatula*. *Plant Cell.* 2015; 27: 2860-79.
24. Zhou H, Lin-Wang K, Wang F, Espley RV, Ren F, Zhao J, et al. Activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation. *New Phytol.* 2018; 221: 1919-34.
25. Ravaglia D, Espley RV, Henry-Kirk RA, Andreotti C, Ziosi V, Hellens RP, et al. Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biol.* 2013; 13: 68.
26. Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AM, de Vos RC, Jonker HH, et al. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa*) fruits. *New Phytol.* 2013; 197: 454-67.
27. Gesell A, Yoshida K, Tran LT, Constabel CP. Characterization of an apple TT2-type R2R3 MYB transcription factor functionally similar to the poplar proanthocyanidin regulator PtMYB134. *Planta.* 2014; 240: 497-511.
28. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell.* 2006; 18: 283-94.
29. Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, et al. The Jasmonate-ZIM-Domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell.* 2011; 23: 1795-1814.
30. Wang YC, Wang N, Xu HF, Jiang SH, Fang HC, Su MY, et al. Auxin regulates anthocyanin biosynthesis through the Aux/IAA-ARF signaling pathway in apple. *Hortic Res.* 2018; 5: 59.
31. Zhang J, Xu H, Wang N, Jiang S, Fang H, Zhang Z, et al. The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. *Plant Mol Biol.* 2018; 98: 205-18.
32. Feng S, Wang Y, Yang S, Xu Y, Chen X. Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. *Planta.* 2010; 232: 245-55.

33. Yao G, Ming M, Allan AC, Gu C, Li L, Wu X, et al. Map-based cloning of the pear gene *MYB114* identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J.* 2017; 92: 437-51.
34. Bai SL, Tao RY, Tang YX, Yin L, Ma YJ, Ni JB, et al. BBX16, a B-box protein, positively regulates light-induced anthocyanin accumulation by activating *MYB10* in red pear. *Plant Biotechnol J.* 2019; 17(10): 1985-97.
35. Thao Ngo, Yanyun Zhao. Stabilization of anthocyanins on thermally processed red D'Anjou pears through complexation and polymerization. *Food Sci Technol.* 2009; 42: 1144-52.
36. Zhai R, Wang Z, Zhang S, Meng G, Song L, Wang Z, et al. Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.). *J Exp Bot.* 2016; 67: 1275-84.
37. Glaubitz U, Li X, Schaedel S, Erban A, Sulpice R, Kopka J, et al. Integrated analysis of rice transcriptomic and metabolomic responses to elevated night temperatures identifies sensitivity and tolerance-related profiles. *Plant Cell Environ.* 2017; 40: 121-37.
38. Ru L, Osorio S, Wang L, Fernie AR, Patrick JW, Ruan YL. Transcriptomic and metabolomics responses to elevated cell wall invertase activity during tomato fruit set. *J Exp Bot.* 2017; 68: 4263-79.
39. An J, Sun M, van Velzen R, Ji C, Zheng Z, Limpens E, et al. Comparative transcriptome analysis of *Poncirus trifoliata* identifies a core set of genes involved in arbuscular mycorrhizal symbiosis. *J Exp Bot.* 2018; 69: 5255-64.
40. Luo HF, Dai C, Li YP, Feng J, Liu ZC, Kang CY. Reduced anthocyanins in petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. *J Exp Bot.* 2018; 69: 2595-2608.
41. Wellmann F, Griesser M, Schwab W, Martens S, Eisenreich W, Matern U, et al. Anthocyanidin synthase from *Gerbera hybrida* catalyzes the conversion of (+)-catechin to cyanidin and a novel procyanidin. *FEBS Lett.* 2006; 580:1642-48.
42. Wang ZG, Meng D, Wang A, Li TL, Jiang SL, Cong PH, et al. The methylation of the PcMYB10 promoter is associated with green- skinned sport in Max Red Bartlett pear. *Plant Physiol.* 2013; 162: 885-96.
43. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science.* 2000; 290: 2105-10.
44. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 2010; 15: 573-81.
45. Feng S, Sun S, Chen X, Wu S, Wang D, Chen X. PyMYB10 and PyMYB10.1 interact with bHLH to enhance anthocyanin accumulation in pears. *Plos One.* 2015; 10: e0142112.
46. Javelle M, Vernoud V, Rogowsky PM, Ingram GC. Epidermis: The formation and functions of a fundamental plant tissue. *New Phytol.* 2011; 189: 17-39.
47. Miao ZQ, Zhao PX, Mao JL, Yu LH, Yuan Y, Tang H, et al. HOMEBOX PROTEIN52 mediates the crosstalk between ethylene and auxin signaling during primary root elongation by modulating auxin transport-related gene expression. *Plant Cell.* 2018; 30: 2761-78.

48. Manavella PA, Arce AL, Dezar CA, Bitton F, Renou JP, Crespi M, et al. Cross-talk between ethylene and drought signaling pathways is mediated by the sunflower Hahb-4 transcription factor. *Plant J.* 2006; 48: 125-37.
49. Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M. ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. *Plant Cell.* 1999; 11: 1217-1226.
50. Jiang Y, Liu C, Yan D, Wen X, Liu Y, Wang H, et al. MdHB1 down-regulation activates anthocyanin biosynthesis in the white-fleshed apple cultivar 'Granny Smith'. *J Exp Bot.* 2017; 68: 1055-69.
51. Lü PT, Zhang CQ, Liu JT, Liu XW, Jiang GM, Jiang XQ, et al. RhHB1 mediates the antagonism of gibberellins to ABA and ethylene during rose (*Rosa hybrida*) petal senescence. *Plant J.* 2014; 78: 578-90.
52. Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S. A knotted1-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell.* 1994; 6: 1859-76.
53. Hurst HC. Transcription factors 1: bZIP proteins. *Protein Profile.* 1995; 2(2): 101-68.
54. Wang ZH, Cheng K, Wan LY, Yan LY, Jiang HF, Liu SY, et al. Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genomics.* 2015; 16: 1053.
55. An JP, Qu FJ, Yao JF, Wang XN, You CX, Wang XF, et al. The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. *Hortic Res.* 2017; 4: 17023.
56. An JP, Yao JF, Xu RR, You CX, Wang XF, Hao YJ. Apple bZIP transcription factor MdbZIP44 regulates abscisic acid promoted anthocyanin accumulation. *Plant Cell Environ.* 2018; 41(11): 2678-92.
57. Devaiah BN, Karthikeyan AS, Raghothama KG. WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. *Plant Physiol.* 2007; 143: 1789-1801.
58. Duan S, Wang J, Gao C, Jin C, Li D, Peng D, et al. Functional characterization of a heterologously expressed *Brassica napus* WRKY41-1 transcription factor in regulating anthocyanin biosynthesis in *Arabidopsis thaliana*. *Plant Sci.* 2018; 268: 47-53.
59. An JP, Zhang XW, You CX, Bi SQ, Wang XF, Hao YJ. MdWRKY40 promotes wounding-induced anthocyanin biosynthesis in association with MdMYB1 and undergoes MdbT2-mediated degradation. *New Phytol.* 2019; 224(1): 380-95.
60. Ji X, Wang Y, Zhang R, Wu S, An M, Li M, et al. Effect of auxin, cytokinin and nitrogen on anthocyanin biosynthesis in callus cultures of red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant Cell Tiss Org.* 2015; 120: 325-37.
61. Liu Z, Shi MZ, Xie DY. Regulation of anthocyanin biosynthesis in *Arabidopsis thaliana* red *pap1-d* cells metabolically programmed by auxins. *Planta.* 2014; 239: 765-81.
62. Cong L, RR Yue, HB Wang, JL Liu, R Zhai, J Yang, et al. 2, 4-D-induced parthenocarpy in pear is mediated by enhancement of GA4 biosynthesis. *Physiol Planta.* 2018; DOI: 10.1111/ppl.12835.

63. Dixon DP, Hawkins T, Hussey PJ, Edwards R. Enzyme activities and subcellular localization of members of the *Arabidopsis* glutathione transferase superfamily. *J Exp Bot.* 2009; 60: 1207-18.
64. Yang Q, Han X, Gu J, Liu Y, Yang M, Zeng Q. Functional and structural profiles of GST gene family from three *Populus* species reveal the sequence-function decoupling of orthologous genes. *New Phytol.* 2019; 221: 1060-73.
65. Wei K, Wang L, Zhang Y, Ruan L, Li H, Wu L, et al. A coupled role for *CsMYB75* and *CsGSTF1* in anthocyanin hyperaccumulation in purple tea. *Plant J.* 2018; 97: 825-40.
66. Marrs KA, Alfenito MR, Lloyd AM, Walbot V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. *Nature.* 1995; 375: 397-400.
67. Dixon DP, Skipsey M, Edwards R. Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry.* 2010; 71: 338-50.
68. Sun Y, Li H, Huang JR. *Arabidopsis* TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol Plant.* 2012; 5: 387-400.
69. Gomez C, Conejero G, Torregrosa L, Cheynier V, Terrier N, Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J.* 2011; 67: 960-70.
70. Li X, Gao P, Cui D, Wu L, Parkin I, Saberianfar R, et al. The *Arabidopsis* *tt19-4* mutant differentially accumulates proanthocyanidin and anthocyanin through a 3' amino acid substitution in glutathione S-transferase. *Plant Cell Environ.* 2011; 34: 374-88.
71. Wang Z, Cui Y, Vainstein A, Chen S, Ma H. Regulation of fig (*Ficus carica* L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front Plant Sci.* 2017; 8: 1990.
72. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *Plos One.* 2012; 7: e30619.
73. Feng S, Sun J, Sun S, Wang Y, Tian C, Sun Q, et al. Transcriptional profiles underlying the effects of methyl jasmonate on apple ripening. *J Plant Growth Regul.* 2017; 36: 271-80.
74. Mu RL, Cao YR, Liu YF, Lei G, Zou HF, Liao Y, et al. An R2R3-type transcription factor gene *AtMYB59* regulates root growth and cell cycle progression in *Arabidopsis*. *Cell Res.* 2009; 19: 1291-1304
75. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 1998; 16: 735-43.

## Declarations

### Acknowledgments

We thank Jennifer Smith, PhD, from Liwen Bianji, Edanz Group China ([www.liwenbianji.cn/ac](http://www.liwenbianji.cn/ac)), for editing the English text of a draft of this manuscript. We thank Dr Yufen Cao at Chinese Academy of Agricultural Sciences, and Mr Shaomin Wang at Shandong Institute of Pomology for suggestions about selection of fruit materials, Dr Chunying Kang at Huazhong Agricultural University for providing *Arabidopsis tt19-7* seeds and GST sequences of strawberry, Maiwei Daixie Co., Ltd. for technical support with HPLC and Lc-

Ms/Ms, KeGene Science & Technology Co., Ltd. for technical support with RNA-seq, Dr Zhiyong Pan at Huazhong Agricultural University, Dr Xueyong Yang at the Institute of Vegetables and Flowers Chinese Academy of Agricultural Sciences and Dr Fangfang Ma at Shandong Agricultural University for helpful discussions and suggestions.

## **Funding**

This work was supported by the National Key Research and Development Program of China (2018YFD1000105), the National Natural Science Foundation of China (31201593 and 31872940) and Agricultural Improved Seed Project of Shandong Province (2019LZGC008).

## **Author information**

Authors' affiliations

*State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, Shandong, China; College of Horticulture Sciences, Shandong Agricultural University, Tai'an, Shandong, China*

Zhen Zhang, Ya Zhang, Chenzhiyu Li, Xi Li, Shuo Wang, Xinyu Wang, Xuesen Chen & Shouqian Feng

*Cherry Research Department, Yantai Agricultural Science and Technology Institute, Yantai, Shandong, China*

Changping Tian & Qiang Yu

## **Contributions**

Conception and design of the research: SQF. Conducted the Experiments: ZZ, CPT, YZ, CZYL and XL. Analyzed the data: ZZ, CC, QY, SW, XYW and MYS. Contributed to writing the manuscript: ZZ, XSC and SQF. All the authors have read and approved the paper.

## **Corresponding author**

Correspondence to Shouqian Feng

## **Ethics declarations**

## **Ethics approval and consent to participate**

Not applicable

## **Consent for publication**

Not applicable

## **Competing interests**

The authors declare that they have no competing interests.

## Supplementary Data

**Table S1.** List of primers used for qPCR analysis and making constructs.

**Table S2.** Comparison of anthocyanin and PA compositions of 'CFs' and 'RCFs'. 'CF1' and 'RCF1' refer to fruits of 'CF' and 'RCF' collected at 2 DAFB, respectively, 'CF2' and 'RCF2' refer to fruits of 'CF' and 'RCF' collected at 12 DAFB, respectively.

**Table S3.** Overview of mapping of RNA-seq reads.

**Table S4.** Differently expressed genes in four comparison groups: 'RCF1' vs. 'CF1', 'RCF2' vs. 'CF2', 'RCF2' vs. 'RCF1', and 'CF2' vs. 'CF1'.

**Table S5.** Gene Ontology (GO) functional annotations of DEGs in group 2-1 and group RCF-CF.

**Table S6.** KEGG annotation of DEGs in group 2-1 and group RCF-CF.

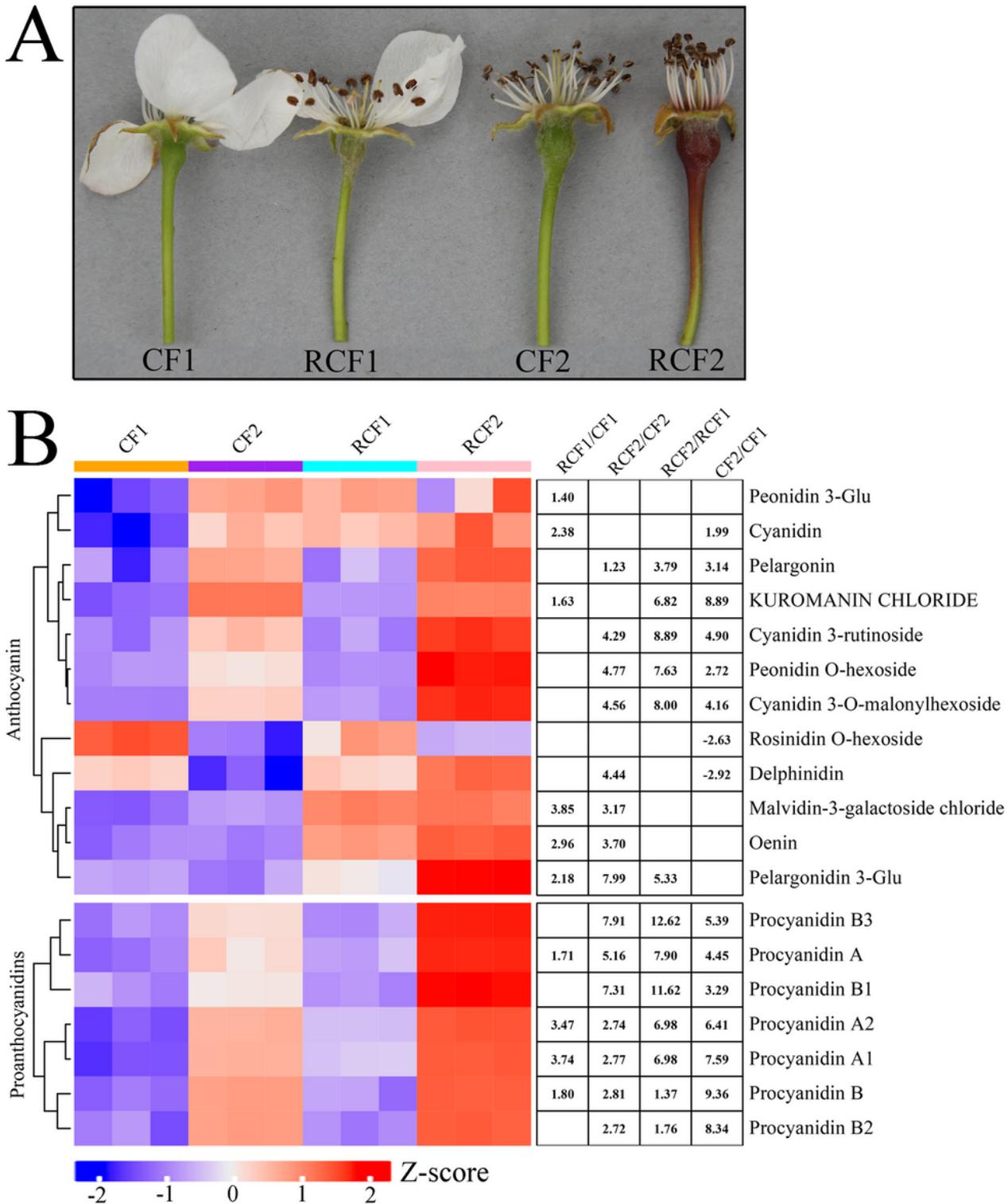
**Table S7.** Correlations between 203 DEGs and 17 metabolites with strong correlation coefficient ( $R^2 > 0.8$ ).

**Table S8.** Correlation analysis between 20 core genes and 10 anthocyanins and seven PAs.

**Table S9.** RNA-seq analysis of candidate genes for anthocyanin and PA accumulation in *tt19-7* and *tt19-7*-OE seedlings.

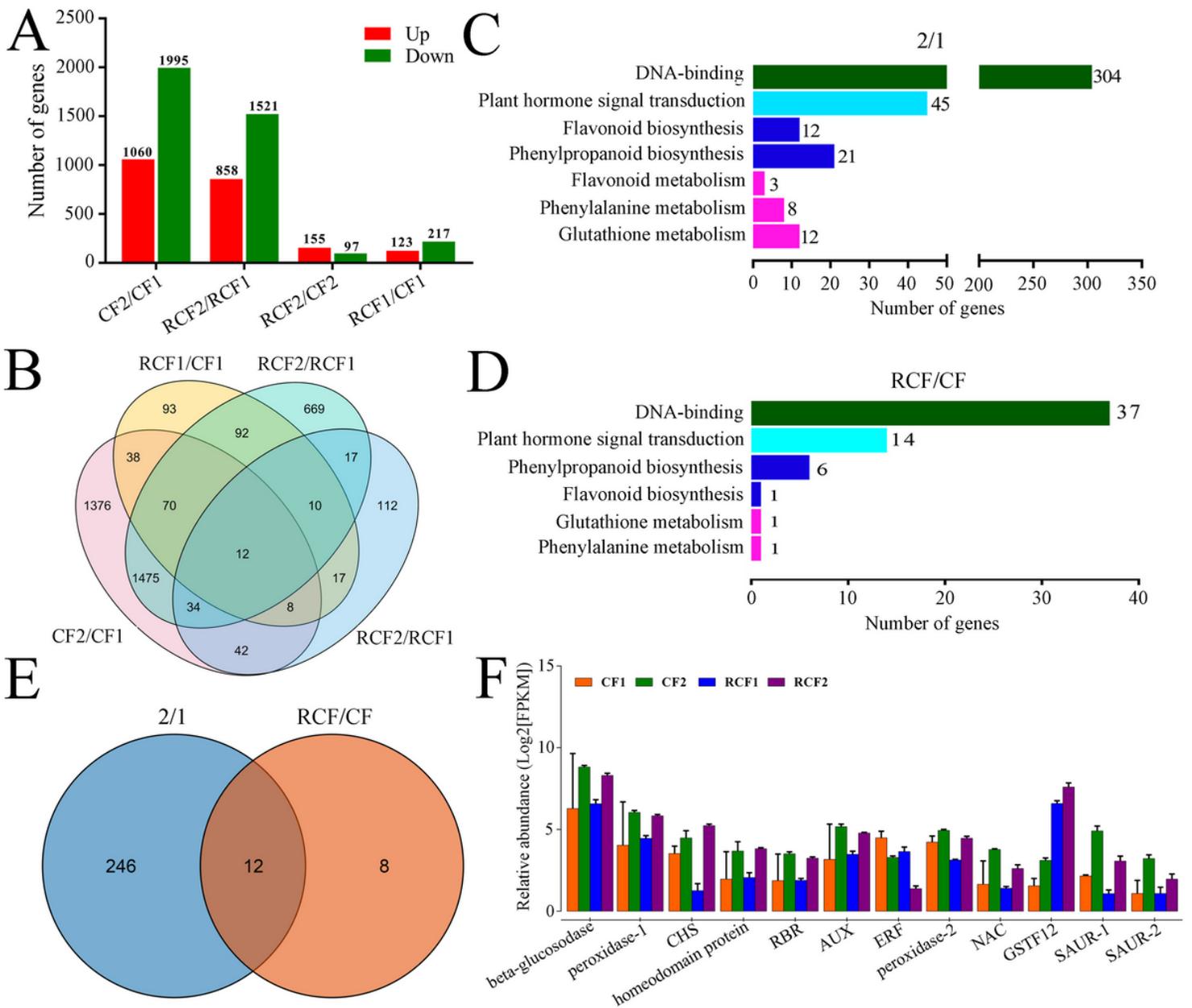
**Table S10.** *PcGSTF12* promoter analysis using PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and PLACE Signal Scan Search database (<https://www.dna.affrc.go.jp/PLACE/>).

## Figures



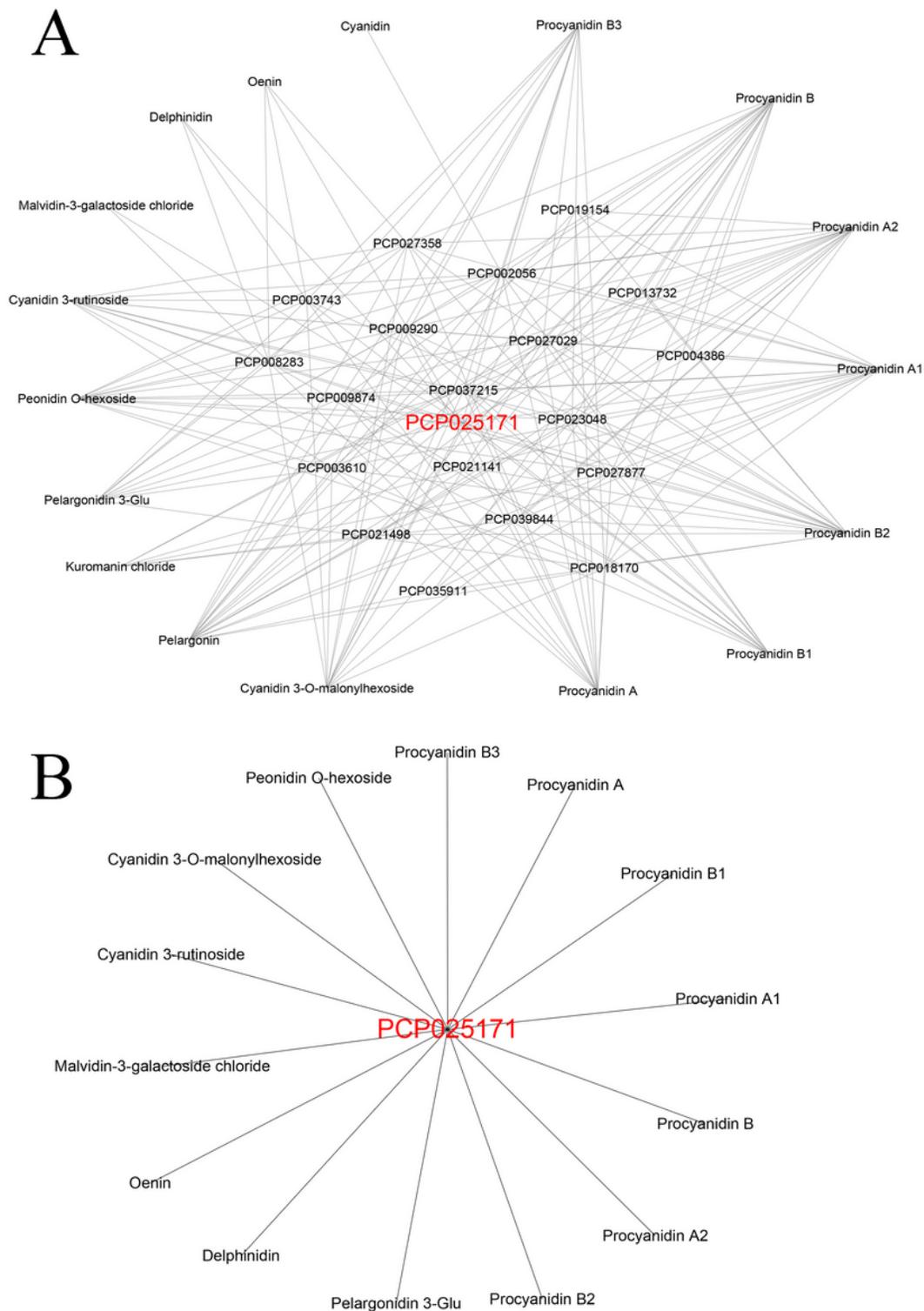
**Figure 1**

Changes in pigment (anthocyanins and procyanins; PAs) contents in 'CF' and 'RCF'. (A) Changes in fruit pigmentation of 'CF' and 'RCF'. (B) Concentrations of anthocyanins and PAs in peel of 'CF' and 'RCF'. 'CF1' and 'RCF1' refer to fruits of 'CF' and 'RCF' collected at 2 days after full bloom (DAFB), respectively, 'CF2' and 'RCF2' refer to fruits of 'CF' and 'RCF' collected at 12 DAFB, respectively, 'CF3' and 'RCF3' refer to fruits of 'CF' and 'RCF' collected at 126 DAFB, respectively. Numbers refer to -fold change of metabolites.



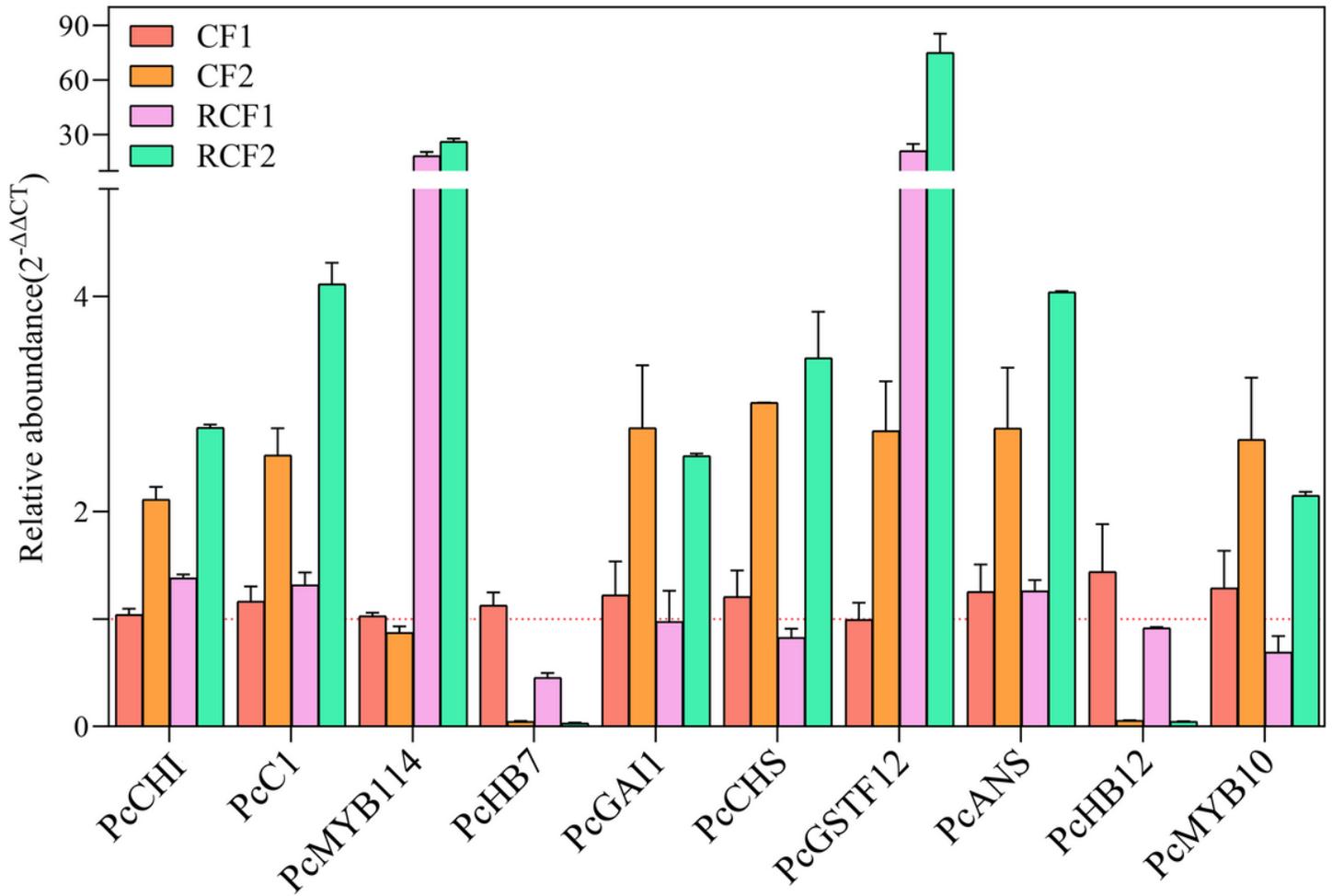
**Figure 2**

Number of differentially expressed genes (DEGs) identified by RNA-seq analysis. (A) Numbers of DEGs. (B) Venn diagram representing numbers of DEGs. (C) Number of DEGs in group 2-1 categorized into DNA binding, plant hormone signal transduction, flavonoid biosynthesis, phenylpropanoid biosynthesis, brassinosteroid biosynthesis, zeatin biosynthesis, flavonoid metabolism, phenylalanine metabolism, glutathione metabolism and ABC transporters. (D) Number of DEGs of group RCF-CF involved in DNA binding, plant hormone signal transduction, phenylpropanoid biosynthesis, flavonoid biosynthesis, glutathione metabolism and phenylalanine metabolism. (E) Overlapping DEGs between group 2-1 and group RCF-CF, (F) Transcript levels of overlapping DEGs between group 2-1 and RCF-CF determined by RNA-seq.



**Figure 3**

Connection network between core genes and anthocyanin and PA metabolites. (A) Networks between 20 core genes and 10 anthocyanins and seven PAs. (B) Networks between PcGSTF12 (PCP025171) and seven anthocyanins and seven PAs. PcGSTF12 is shown in red font.



**Figure 4**

Transcript levels of anthocyanin- and PA-related genes. Each experiment had three biological replicates.



Competitor probe was non-labeled probe. Mutant probe contained two nucleotide mutations. Competitors, and mutant probes at a 10×, 25×, and 50× molar excess were present (+) or absent (-) in each reaction. Black arrows indicate increasing multiples. (F) Effects of PcMYB114 on promoter activities as demonstrated by luciferase reporter assays. Empty-vector was used as a control. Values are means ± SD of three independent biological replicates. Statistical significance: \*\*P < 0.01.

## Supplementary Files

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

- [TableS2.xls](#)
- [TableS4.xls](#)
- [TableS8.xls](#)
- [TableS6.xls](#)
- [TableS5.xls](#)
- [TableS7.xls](#)
- [TableS1.xls](#)
- [TableS9.xlsx](#)
- [TableS3.xlsx](#)
- [TableS10.xlsx](#)