

# A co-expression gene network associated with developmental regulation of apple fruit acidity

Yang Bai · Laura Dougherty · Lailiang Cheng · Kenong Xu

Received: 28 September 2014 / Accepted: 30 December 2014 / Published online: 11 January 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Apple fruit acidity, which affects the fruit's overall taste and flavor to a large extent, is primarily determined by the concentration of malic acid. Previous studies demonstrated that the major QTL *malic acid* (*Ma*) on chromosome 16 is largely responsible for fruit acidity variations in apple. Recent advances suggested that a natural mutation that gives rise to a premature stop codon in one of the two aluminum-activated malate transporter (ALMT)-like genes (called *Mal*) is the genetic causal element underlying *Ma*. However, the natural mutation does not explain the developmental changes of fruit malate levels in a given genotype. Using RNA-seq data from the fruit of 'Golden Delicious' taken at 14 developmental stages from 1 week after full-bloom (WAF01) to harvest (WAF20), we characterized their transcriptomes in groups of high ( $12.2 \pm 1.6$  mg/g fw, WAF03–WAF08), mid ( $7.4 \pm 0.5$  mg/g fw, WAF01–WAF02 and WAF10–WAF14) and low ( $5.4 \pm 0.4$  mg/g fw, WAF16–WAF20) malate concentrations. Detailed analyses showed that a set of 3,066 genes (including *Mal*) were expressed not only differentially ( $P_{\text{FDR}} < 0.05$ ) between the high and low malate groups (or between the early and late

developmental stages) but also in significant ( $P < 0.05$ ) correlation with malate concentrations. The 3,066 genes fell in 648 MapMan (sub-) bins or functional classes, and 19 of them were significantly ( $P_{\text{FDR}} < 0.05$ ) co-enriched or co-suppressed in a malate dependent manner. Network inferring using the 363 genes encompassed in the 19 (sub-) bins, identified a major co-expression network of 239 genes. Since the 239 genes were also differentially expressed between the early (WAF03–WAF08) and late (WAF16–WAF20) developmental stages, the major network was considered to be associated with developmental regulation of apple fruit acidity in 'Golden Delicious'.

**Keywords** Apple · Developing fruit · Acidity · Transcriptome · RNA-seq · Gene network

## Introduction

Acidity, which affects fruit overall taste and flavor, has long been considered one of the most important quality attributes in fruits, such as apple, grape and tomato (Sweetman et al. 2009; Etienne et al. 2013). As such, titratable acidity and pH are routinely measured to evaluate fruit acidity levels for advancing selections in fruit breeding. Many organic acids contribute to acidity levels in fruit, but malic acid (or its conjugate base malate) and citric acid (or citrate) are regarded the major contributors. Given their roles in determining fruit acidity, extensive efforts have been made to understand how malate and citrate accumulate in fruit cells, which has been reviewed recently in detail (Sweetman et al. 2009; Etienne et al. 2013).

In apple, acidity in mature fruit is primarily determined by the concentration of malic acid as it usually makes up 90 % or more of total acidity (Hulme and Wooltorton 1957;

---

Communicated by S. Hohmann.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00438-014-0986-2) contains supplementary material, which is available to authorized users.

---

Y. Bai · L. Dougherty · K. Xu (✉)  
Horticulture Section, School of Integrative Plant Science, Cornell University, New York State Agricultural Experiment Station, Geneva, NY, USA  
e-mail: kx27@cornell.edu

L. Cheng  
Horticulture Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

Ulrich 1970; Beruter 2004; Zhang et al. 2010). Obviously, biochemical and physiological processes that affect malate accumulations are relevant in determining fruit acidity levels. These processes may include malate synthesis, degradation, intracellular transport, and storage (Beruter 2004; Sweetman et al. 2009; Etienne et al. 2013). Malate is primarily synthesized in cytosol through glycolysis of hexoses derived from sorbitol and sucrose, which are imported from leaves. It could also be synthesized through pathways pertaining to photosynthesis in chloroplast, the tricarboxylic acid (TCA) cycle in mitochondrion, and the glyoxylate cycle in glyoxysome at varying fruit growth and developmental stages. For malate degradation, the main pathways are gluconeogenesis in cytosol and the TCA cycle in mitochondrion. Since vacuole is the major repository of malate, intracellular transport of malate into and out of the vacuole is surely an integral part for cell malate homeostasis.

To understand these biological processes important for fruit acidity in apple, genes and/or enzymes involved in malate synthesis and degradation are logical targets of investigations. However, studies on the relevance of these genes or enzymes in determining acidity levels had drawn inconsistent conclusions. On one hand, no difference in measuring the activity of phosphoenol pyruvate carboxylase (PEPC), NAD-dependent malate dehydrogenase (NAD-MDH) or NADP dependent malic enzyme (NADP-ME), the key enzymes in malate metabolism, was found between a low acid variety ‘Usterapfel’ and its high acid mutant (Beruter 2004), suggesting a limited role, if any, of these enzymes in fruit acidity. On the other hand, profiling the expression patterns and their corresponding enzyme activities for genes *MdPEPC* (EU315246, for PEPC) and *MdcyME* (DQ280492, for NADP-ME) underscored that there were differences between low and high acid genotypes (Yao et al. 2009), implicating their roles in fruit acidity. In addition, the functionality of NAD-MDH (DQ221207) had been demonstrated in malate synthesis in apple (Yao et al. 2011).

Although the study (Beruter 2004) did not find significant differences between the low acid variety ‘Usterapfel’ and its high acid mutant in the catalytic activity for enzymes PEPC, NAD-MDH and NADP-ME, it demonstrated that the uptake of [ $^{14}\text{C}$ ] malate was significantly lower in excised tissue of ‘Usterapfel’ than in that of the mutant. This suggested that the low malate content in ‘Usterapfel’ fruit was the result of a restricted ability to transport malate into its vacuoles and then maintain the malate levels in them. Indeed, several vacuolar transporters, such as the vacuolar pumps, e.g. V-ATPase (Schumacher and Krebs 2010) and MdVHA-A (EF128033, for subunit A of vacuolar H<sup>+</sup>-ATPase) (Yao et al. 2009), tonoplast dicarboxylate transporter, e.g. AtTD1 (Emmerlich et al. 2003), and members of the ALMT1 (aluminum-activated malate

transporter1) family proteins (Barbier-Brygoo et al. 2011), e.g. AtALMT9 (Kovermann et al. 2007) and AtALMT6 (Meyer et al. 2011), had been implicated of critical roles in maintaining the homeostasis of malate in plant cells. The most important and direct evidence of the role of vacuolar transporters in determining malate levels in apple came from the isolation of *Ma* (malic acid) (Bai et al. 2012; Khan et al. 2013), a major gene or QTL controlling fruit acidity on linkage group 16 (Maliepaard et al. 1998; Liebhard et al. 2003; Kenis et al. 2008; Xu et al. 2012). The *Ma* locus encodes two ALMT1-like genes, called *Mal1* and *Mal2*, and a single nucleotide mutation that leads to a premature stop codon in *Mal1* is attributed to the low acidity phenotype in apple although the expression levels of *Mal1* is also important (Bai et al. 2012).

In developing apple fruit of a given genotype, the malate concentrations could vary by threefold between the highest and the lowest. The change commonly begins with a rapid increase in the first four or five weeks after full-bloom and then progressively decreases through maturity (Hulme and Wooltorton 1957; Ulrich 1970; Beruter 2004; Zhang et al. 2010). Although the *Ma* locus largely governs fruit acidity variations in diverse apples, it does not genetically explain the developmental changes of fruit acidity in a given genotype as the genotype at the *Ma* locus never changes, suggesting the existence of developmental regulation of fruit acidity. We hypothesized that such developmental regulation involves a complex network consisting of many genes. The goal of this study is to shed light on the gene network by (1) identifying the genes that were differentially expressed in fruit not only varying in malate contents but also in developmental stages through RNA-seq based transcriptome analysis, (2) identifying functional classes or MapMan (sub-) bins that were either co-enriched or co-suppressed in a manner dependent on both malate levels and development stages, and (3) constructing a co-expression gene network from the genes encompassed in the functional classes identified in (2).

## Materials and methods

### Fruit growth evaluation and malate quantification

Fruit samples of ‘Golden Delicious’ (GD) were collected in a previous study (Wang and Xu 2012). Briefly, 5–10 fruits were taken at each of 14 time points from 1 week after full-bloom (WAF01) through harvest (WAF20) in 2010. The fruits were first measured for their diameters and then were immediately frozen in liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$  until use. For WAF01–03 (three time points), the whole fruit was pooled with the pedicel removed, and there were at least ten fruits for each sampling time point. For the rest

11 time points WAF04–20 (WAF04, 05, 06, 08, 10, 12, 14, 16, 18, 19 and 20), at least five fruits were pooled for each time point, but only the cortex tissues were used. In addition, three biological replicates at WAF20 were also prepared from one, two and two fruit, respectively. As a result, a total of 17 fruit samples were prepared for the 14 time points (one pooled sample for each time point of WAF01–20 plus three additional replicates for WAF20).

Malate concentrations of the 14 pooled fruit samples were determined on an Agilent 7890A GC/5975C MS (Agilent Technology, Palo Alto, CA, USA) with three technical replicates from the pooled fruit samples. Organic acids were extracted and derivatized following a protocol described previously (Lisec et al. 2006) with minor modifications. Briefly, homogenized apple fruit tissues of 100 mg (per sample per replicate) were extracted in 1.4 ml of 75 % methanol with 600 ppm ribitol added as internal standard. After fractionation of non-polar metabolites into chloroform, 2.5  $\mu$ l of the polar phase was transferred into a 2.0 ml Eppendorf tube and was dried under vacuum without heating and then derivatized with 40  $\mu$ l of 5 mg/ml methoxyamine hydrochloride and 40  $\mu$ l of *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (MSTFA) sequentially. To measure malate, 1  $\mu$ l of the derivatized sample was injected at 230 °C in splitless mode with helium carrier gas flow set to 1 ml/min. Chromatography was performed on a DB-5MS capillary column (20 m  $\times$  0.18 mm  $\times$  0.18  $\mu$ m) with a 5 m Duraguard column (Agilent Technology). The temperature program was isothermal at 70 °C for 2.471 min, followed by a 10.119 °C/min ramp to 330 °C and a final 2.471 min heating at 330 °C. Cooling was performed as fast as possible. The system was then equilibrated at a temperature of 70 °C for 5 min before the next injection. Mass spectra were collected at 5.6 scans s<sup>-1</sup> over an *m/z* 50–600 scanning range. The transfer line temperature and the ion source temperature were set to 250 and 230 °C, respectively. Metabolites were identified by comparing fragmentation patterns with those in a mass spectral library generated by the GC/MS system and an annotated quadrupole GC–MS spectral library downloaded from the Golm Metabolome Database ([http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd\\_msri.html](http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd_msri.html)), and were quantified based on standard curves generated for each metabolite and internal standard.

#### RNA isolation and strand specific RNA-seq library construction and sequencing

For total RNA isolation, ground fruit tissues of 2 g (young fruit)–3 g (mature or near mature fruit) were used following a previous protocol (Gasic et al. 2004) with modifications, i.e. 1 ml Sarkosyl of 20 % (w/v) was added to the extraction buffer before tissue tearer homogenization. The

isolated total RNA was stored in EB buffer (Qiagen, Germantown, MD) with addition of 1 $\times$  Ambion RNA secure (Invitrogen/Life Technologies, Carlsbad, CA). Activation of RNAsecure was carried out by incubating the samples at 60 °C (in a water bath) for 10 min and then immediately put on ice. Evaluation of RNA quality and concentration was performed using a combination of Nanodrop 1000 (Thermo Scientific, Waltham, MA), Bioanalyzer 2100 with RNA 6000 Nano Chip (Agilent, Santa Clara, CA) and electrophoresis of a 2 % agarose gel (using 1/10 RNA dilutions in EB buffer with 1 $\times$  Ambion RNA secure). Immediately prior to mRNA isolation, the RNA samples were treated with DNase I (amplification grade, Invitrogen) at 37 °C for 30 min and then followed by heat inactivation at 65 °C for 15 min.

The procedure for RNA-seq library construction was previously described (Bai et al. 2014). Briefly, we used NEBNext Poly (A) mRNA Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) for mRNA isolation and strand specific RNA-seq library preparation. The starting amount of total RNA for each library was 5  $\mu$ g and the manufacturer's protocols were followed with minor modifications (Bai et al. 2014). The 14 libraries constructed from pooled fruit samples were multiplexed with 60 ng each for single-end sequencing of 101 bases without replication in one lane of Illumina HiSeq 2000 (Illumina, San Diego, CA) at the Cornell University Biotechnology Resource Center (Ithaca, NY). The three libraries constructed from the three WAF20 replicates were multiplexed with other samples and sequenced similarly in another lane.

#### RNA-seq data analysis

17 sequence files with a total of 214.2 million (data not shown) raw reads in total were generated by the Illumina pipeline in software CASAVA v1.8 in Sanger FASTQ format (available under NCBI SRA experiment number SRX392051), but only those (199.8 million, i.e. 93.3  $\pm$  1.2 % of the total raw reads) that passed the chastity filter (i.e. no more than one base call in the first 25 cycles has a chastity higher than 0.6) in the pipeline were used (Supplementary Table S1).

Data analyses were performed using CLC Genomics Workbench (CLC GW) v6.5 (CLCBio, Cambridge, Massachusetts). To remove reads derived from rRNA, the ribosomal RNA database SILVA Release 115 (<http://www.arb-silva.de/>) was downloaded and used as reference sequences. Mapping of the reads to the rRNA references was conducted using the minimum length fraction of 1.0 and the minimum similarity of 0.95. The reads that could not be mapped to the reference SILVA Release 115 were collected and then filtered by quality (using the CLC GW default settings) to

further remove low quality reads and/or bases (Supplementary Table S1). For mapping of RNA-seq reads against the improved apple reference transcriptome (Bai et al. 2014), the minimum length fraction of 0.8 and the minimum similarity of 0.98 were used (our empirical sequence identity threshold often effective in differentiating paralog sequences in the apple genome). The improved reference transcriptome is available at the Genome Database for Rosaceae (Jung et al. 2014), which includes 53,654 of the 63,541 genes or MDPs (MDPs are the three letters prefixing gene IDs in apple) predicted originally and 17,524 novel transcripts (Bai et al. 2014). For convenience, hereafter the 17,524 novel transcripts will be referred to ‘gene’ and named ‘G#####’ as in ‘G101234’ and ‘MDP0000’ (e.g. MDP0000252114) will be abbreviated to ‘M’ (e.g. M252114).

For expression analysis, the count of reads mapped for a given gene was expressed as RPKM (reads per kilobase exon model per million mapped reads) (Mortazavi et al. 2008). Genes were considered expressed if their RPKM values were  $>0.3$ , a threshold similarly used previously (Kang et al. 2013). For time point WAF20, the mean RPKM value of the four samples was used.

The cutoff for a gene said to be differentially expressed between malate groups is  $P_{\text{FDR}} < 0.05$  in Baggerley’s test (Baggerly et al. 2003). In the subsequent gene co-enrichment or co-suppression analysis, K-mean clustering and network construction, RPKM values were square root transformed.

#### Identification of genes and MapMan (sub-) bins associated with the variation of malate concentrations

We define that malate associated genes are those that meet the following two criteria: (1) differently expressed ( $P_{\text{FDR}} < 0.05$ ) between the high and low malate samples in Baggerley’s test (Baggerly et al. 2003) (conducted using CLC GW), and (2) expressed in significant correlation ( $P_r < 0.05$ ) with malate concentrations and/or the expression of *Mal* (Bai et al. 2012) in the 14 sampling time points (conducted using MS Excel). Association of MapMan (sub-) bins or functional classes with malate was established using the PageMan tool (Usadel et al. 2006) embedded in the MapMan software, where the significant threshold  $P_{\text{FDR}} < 0.05$  was set in Wilcoxon rank sum test (Wilcoxon 1945) followed by Benjamini–Hochberg’s false discovery rate (FDR) correction (Benjamini and Hochberg 1995).

To assign the MapMan (sub-) bins or functional classes, the sequences of the 71,178 genes in the improved apple reference transcriptome (Bai et al. 2014) were submitted for BLAST search of multiple databases using the web-based search tool Mercator (<http://mapman.gabipd.org/web/guest/mercator>) (Lohse et al. 2014). The databases searched include: TAIR-Arabidopsis TAIR proteins (release

10), PPAP-SwissProt/UniProt Plant Proteins, CHLAMY-JGI Chlamy release 4 Augustus models, ORYZA-TIGR5 rice proteins, KOG: Clusters of orthologous eucaryotic genes database (KOG), CDD-conserved domain database, and IPR-interpro scan. The output file of Mercator lists the best hits in databases and assigns MapMan’s (sub-) bins (Thimm et al. 2004) for each input sequences if possible, which were used as the ‘mapping file’ required by the PageMan tool. The ‘experiment file’ contained only the malate associated genes (met with the two criteria above) and their RPKM values with square root transformation.

#### Gene clustering and co-expression network inferring

Genes in the MapMan (sub-) bins that were associated with malate variations were clustered by K-means using MeV4.9 (Saeed et al. 2006) with default parameters except for the number of clusters, which was set to six. Network inferring from these genes was conducted using the Cytoscape (Cline et al. 2007) Network Inference (Cyni) tool based on the Pearson correlation coefficient (PCC)  $r > 0.95$ , an optimal threshold obtained using a method described in Aoki et al. (2007) (Supplementary Fig. S2). For network analysis, the Cytoscape plugin NetworkAnalyzer (Assenov et al. 2008) was used.

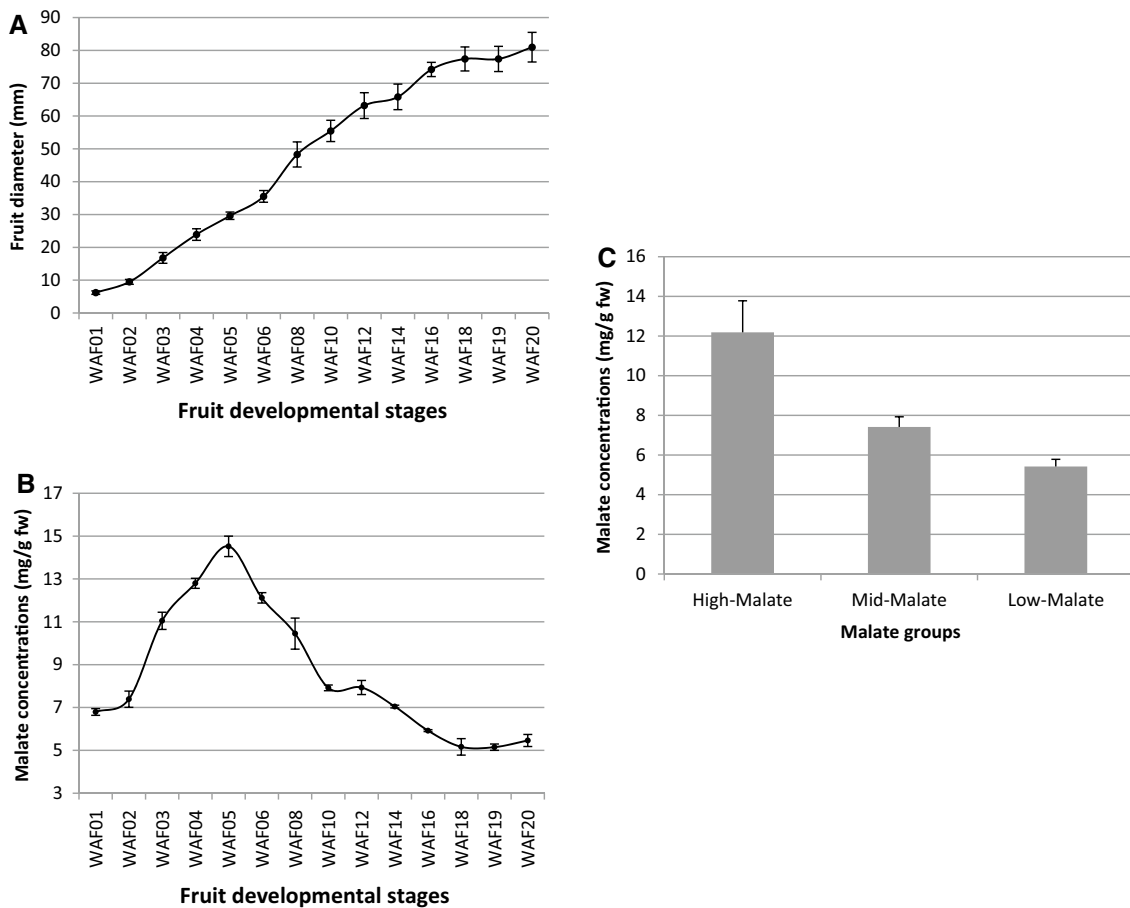
#### qRT-PCR analysis

The same set of total RNA samples used for RNA-seq libraries was also used for qRT-PCR analysis. Three independent reverse transcription reactions were carried out for each sample with 2  $\mu\text{g}$  of total RNA using the Superscript III RT (Invitrogen, Carlsbad, CA, USA). The resultant first strand cDNA was diluted by fivefold, and then used as templates for qRT-PCR analysis, in which an apple actin gene (EB136338, part of M774288) was served as a reference. *Mal* (M252114) and five randomly selected genes M196894, M127123, M163222, M190273 and M219042 were tested, and their primer sequences were available (Supplementary Table S2). qRT-PCR was performed with the Roche (Indianapolis, IN, USA) LightCycler 480 Real-Time PCR System as described previously (Bai et al. 2012). Expression quantification and data analysis were performed via LightCycler 480 Software (Version 1.5) using the comparative cycle threshold method (Pfaffl 2001).

## Results

### Evaluation of fruit growth and malate concentrations

Fruit growth was evaluated with the diameters of the collected fruit (5–10 fruit per time point) over the entire



**Fig. 1** Fruit diameter and malate concentration measurements in developing fruit of ‘Golden Delicious’ (GD). Standard deviations are shown with the *error bars*. *fw* fresh weight. **a** Fruit diameters (mm) at 14 developmental stages from 1 week after full-bloom (WAF01)

through harvest (WAF20). **b** GC–MS measurements of malate concentrations (mg/g fw) in fruit from WAF01 through WAF20. **c** Mean malate concentrations of fruit groups of high (WAF03–06 and WAF08), mid (WAF01–02, WAF12–16) and low (WAF16–20) malate

development process (Fig. 1a). The fruit diameter grew continuously from 6.2 mm at WAF01 (1 week after full-bloom) to 81.0 mm at the harvest (WAF20). Fruit malate concentrations were measured by GC–MS at the same 14 stages from WAF01 to WAF20. The malate concentrations increased slowly initially from WAF01 ( $6.79 \pm 0.16$  mg/g fw) to WAF02 ( $7.39 \pm 0.38$  mg/g fw), but rapidly from WAF02 through WAF05 ( $14.52 \pm 0.48$  mg/g fw) when the peak was reached (Fig. 1b). Decline in malate concentration began at WAF06 and continued until harvest (WAF20) although there were minor upticks at WAF12 and WAF20. The lowest malate concentration was recorded at WAF19 ( $5.15$  mg/g fw), equivalent to 35.5 % of the peak at WAF05.

Based on the overall malate concentrations, the samples from the 14 time points were categorized into three groups: (1) high malate of five samples WAF03–WAF6 and WAF08, (2) mid malate of another five WAF01, WAF02, WAF10, WAF12 and WAF14, and low malate of four WAF16 and WAF18–WAF20 (Fig. 1c). ANOVA

analysis indicated that the differences in group means were significant between the high and mid malate groups ( $P = 1.8E-12$ , *z* test) as well as between the mid and low malate groups ( $P = 7.8E-12$ , *z* test). Co-incidentally, fruit in the high malate group were from WAF03–08, 8 weeks earlier than those in the low malate group, which were from WAF16–20. This provided a unique opportunity for investigating developmental regulation of fruit malate levels.

#### Expression analysis in groups of varying malate concentrations

Gene expression analyses were conducted based on the improved apple reference transcriptome (Bai et al. 2014). After removing reads derived from rRNA and those of low quality, the total reads input for mapping was 179.9 million, and the total mapped and uniquely mapped reads were 142.5 million (79.0 %) and 118.9 million (65.9 %), respectively (Supplementary Table S1). The mean of the total

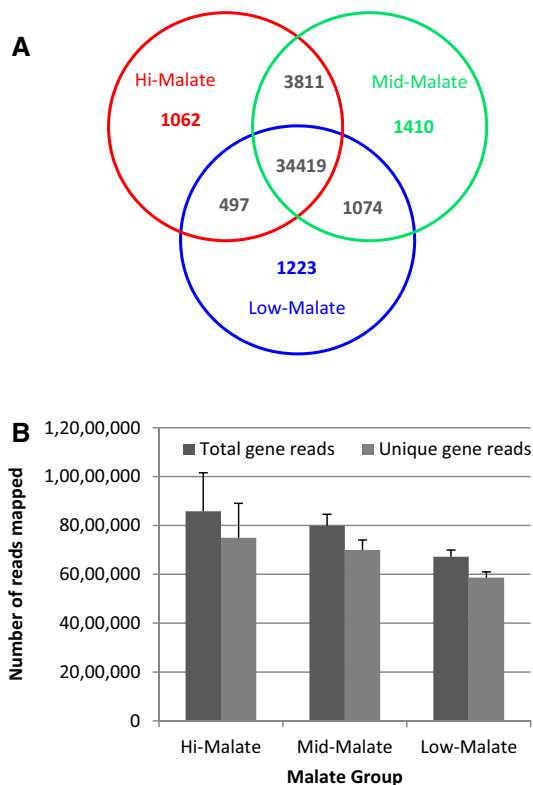


mapped reads per sample was  $8.4 \pm 1.5$  million, of which  $7.0 \pm 1.3$  million were mapped uniquely.

Overall, there were 50,700 genes expressed (i.e. with an RPKM  $>0.3$  at least in one of the 14 time points, data not shown). However, if the mean RPKM values in the three groups were used, the number of expressed genes (i.e. with an RPKM  $>0.3$  at least in one of the three groups of high, mid and low malate) was 43,496 (Fig. 2a). In individual malate groups, there were 39,789 genes expressed in the high, 40,714 in the mid, and 37,213 in the low. The genes that were uniquely expressed were 1,062, 1,410 and 1,223 in the high, mid and low malate groups, respectively (Fig. 2a). For the expressed genes, the mean mapped reads per time point were  $8.6 \pm 1.6$  million in total and  $7.5 \pm 1.4$  million in unique in the high malate group,  $8.0 \pm 0.5$  million in total and  $7.0 \pm 0.4$  million in unique in the mid, and  $6.7 \pm 0.3$  million in total and  $5.9 \pm 0.2$  million in unique in the low (Fig. 2b).

#### Identification of genes associated with variations of malate concentrations

To identify genes putatively responsible for the variations in malate concentrations in developing fruit, we



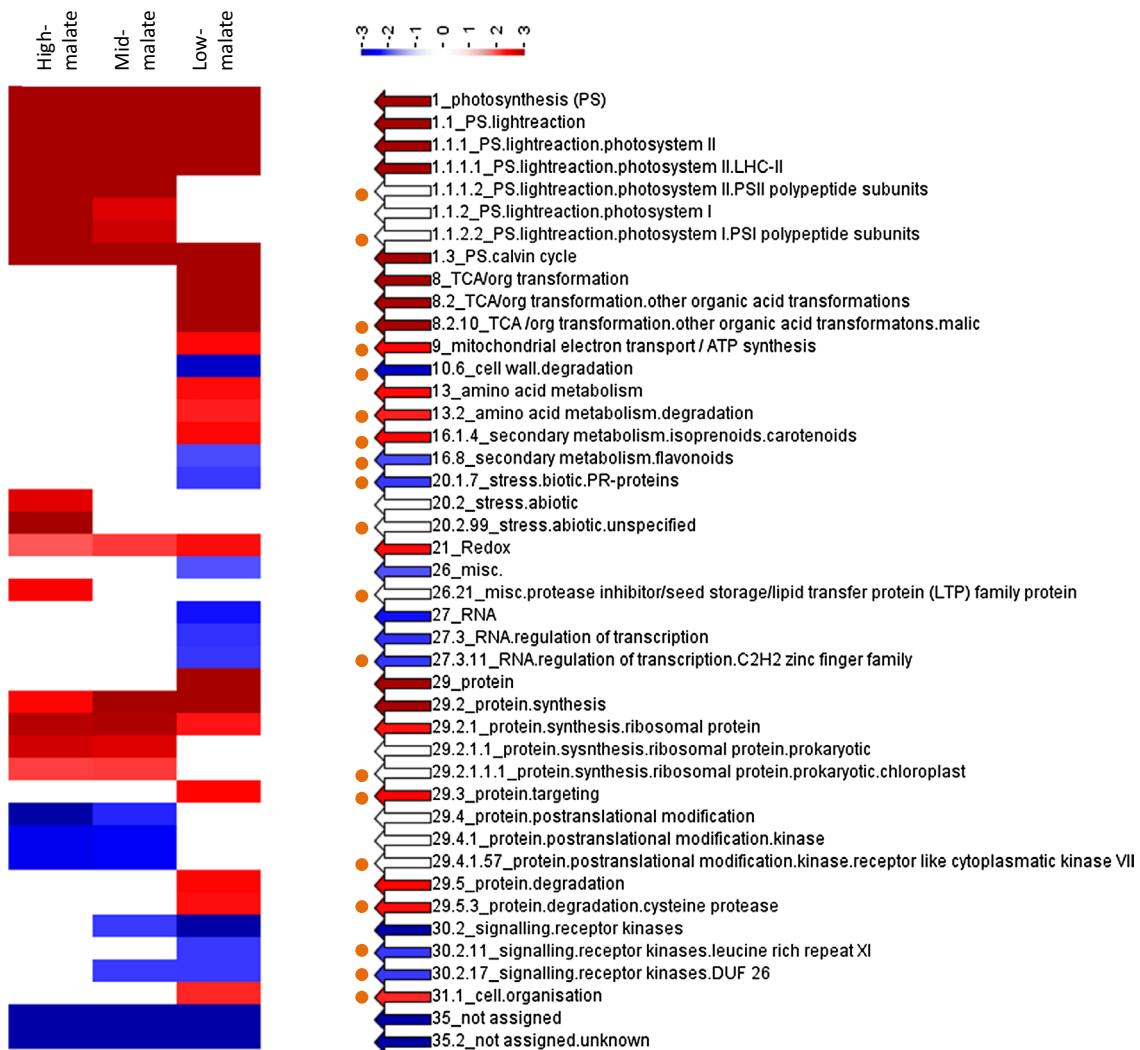
**Fig. 2** Overview of RNA-seq data analysis in the fruit groups of high, mid and low malate. **a** Venn diagram representation of the number of genes expressed (RPKM  $> 0.3$ ). **b** Mean number of the mapped reads (per sample). Standard deviations were shown with the error bars

conducted two comparisons of expression patterns: the first was between the high and low malate groups (or between groups of the early and late developmental stages), which identified 4,041 genes or transcripts differentially expressed ( $P_{\text{FDR}} < 0.05$ ) based on Baggerley's test (Baggerley et al. 2003). Notably, *Mal* had a  $P_{\text{FDR}}$  value of 0.678, suggesting *Mal* was not differentially expressed between the high and low malate groups under this cutoff. The second was among the 14 individual time points, which was started with a large set of 22,870 genes that showed minimal difference between any of the two time points, i.e. RPKM absolute difference  $>3$  and RPKM absolute fold change  $>1.2$ . (Note: These low stringency filters were intended to include a maximum number of genes initially, and was not used for the identification of genes associated with malate). The genes in the large set were each calculated for their correlations with malate concentrations as well as the expression of *Mal*. This identified 7,150 genes significantly ( $P_{r > 0.538} < 0.05$ ) correlated with either malate concentrations and/or the expression of *Mal*. The correlation coefficient ( $r$ ) between malate and *Mal* was 0.572 ( $P < 0.05$ ), indicating the expression of *Mal* was significantly correlated with malate variations in developing fruit. Combining the two lists of genes (4,041 and 7,150) identified 3,066 genes in common (including *Mal*) (Supplementary Table S3). These genes were therefore not only differentially expressed ( $P_{\text{FDR}} < 0.05$ ) between the high and low malate groups (or between fruit in the early and late stages), but also significantly ( $P_{r > 0.538} < 0.05$ ) correlated with malate concentrations (2,180 or 71.1 %), or with the expression of *Mal* (100 or 3.3 %), or both (786 or 25.6 %).

MapMan gene ontology analysis showed that 1,949 (63.6 %) of 3,066 genes could be assigned to one or more MapMan Bins while 1,117 (36.4 %) could not (Supplementary Fig. S1). Bins '29\_protein' of 342 (11.2 %) genes, '26\_miscellaneous' of 215 (7.0 %) and '27\_RNA' of 199 (6.5 %) were the largest bins among the assigned, whereas '32\_microRNA...' of 0 (0 %), '24\_Biodegradation of Xenobiotics' of 1 (0.03 %) and '22\_Polyamine metabolism' of 2 (0.07 %) were the smallest bins. In the unassigned, 174 (5.7 %) were found of significant hits, but there were no ontology for them in MapMan; and the rest 938 (30.6 %) were either unknown or failed to be processed in Mecator-based database searches. Overall, the 3,066 genes were assigned into 648 MapMan (sub-) bins or functional classes (Supplementary Table S3).

#### Identification of MapMan (sub-) bins expressed in a malate dependent manner

To determine if the expression of any functional classes or MapMan (sub-) bins were co-enriched or co-suppressed as opposed to the overall expression of all genes, the 3,066



**Fig. 3** The functional classes (MapMan bins or sub-bins) that were significantly ( $P_{\text{FDR}} < 0.05$ ) co-enriched or co-suppressed in the fruit groups of high, mid and low malate. The FDR  $p$  values were converted into  $z$  scores (a  $p$  value of 0.05 equals a  $z$  score of 1.96) and represented with a color scale from blue (co-suppressed) to white

(expressed at normal levels) to red (co-enriched). The functional classes are annotated on the right panel and the  $z$  scores were listed in Supplementary Table S4. The orange dots indicate the 19 non-overlapping functional classes (color figure online)

genes in each of the three malate groups were subjected to a MapMan (sub-) Bin-wide Wilcoxon test (Fig. 3; Supplementary Table S4). It showed that a total of 43 functional classes were expressed significantly higher or lower than the rest in a given malate group. Nineteen of the 43 functional classes fell into the two primary MapMan bins, i.e. '1\_PS (photosynthesis)' of 8 and '29\_protein' of 11. The rest (24) functional classes were in 12 primary MapMan bins with three each in '8\_TCA/org transformation', '20\_stress', '27\_RNA' and '30\_signaling', two each in '13\_amino acid metabolism', '16\_secondary metabolism', '26\_miscellaneous' and '35\_not assigned', and one each in '9\_mitochondrial electron transport/ATP synthesis', '10\_cell wall', '21\_redox' and '31\_cell' (Fig. 3).

There were two groups of functional classes to be excluded from further analysis. The first was the ten MapMan (sub-) bins in codes 1, 1.1, 1.1.1, 1.1.1.1, 1.3, 21, 29.2, 29.2.1, 35 and 35.2 (Fig. 3; Supplementary Table S4) as they were either co-suppressed (in bins 35 and 35.2) or co-enriched (in the other eight) without discrimination in the three malate groups, indicating they were not causal for the variations of malate concentrations in developing fruit. The second was the 14 functional classes with MapMan bin-codes 1.1.2, 8, 8.2, 13, 20.2, 26, 27, 27.3, 29, 29.2.1.1, 29.4, 29.4.1, 29.5 and 30.2 (Fig. 3; Supplementary Table S4). These functional classes were in higher orders in the hierarchical nomenclature system of MapMan (sub-) bins and their detections were likely caused by the presence of those

**Table 1** The number (without *Mal*) and K-mean cluster of genes in MapMan (sub-) bins co-enriched or co-suppressed in varying malate groups

MapMan (sub-) bin code	MapMan (sub-) bin description	No. of genes	Percent (%)	K-means cluster					
				I	II	III	IV	V	VI
1.1.1.2	PS.lightreaction.photosystem II.PSII polypeptide subunits	29	8.0	22	2	5			
1.1.2.2	PS.lightreaction.photosystem I.PSI polypeptide subunits	18	5.0	17		1			
8.2.10	TCA/org transformation.other organic acid transformatons.malic	9	2.5		9				
9	Mitochondrial electron transport/ATP synthesis.NADH-DH. localisation not clear	9	2.5		8	1			
10.6	Cell wall.degradation	18	5.0		1	2	7	5	3
13.2	Amino acid metabolism.degradation	11	3.0	1	10				
16.1.4	Secondary metabolism.isoprenoids.carotenoids	8	2.2		8				
16.8	Secondary metabolism.flavonoids	30	8.3		5	2	20	2	1
20.1.7	Stress.biotic.PR-proteins	11	3.0	1	2	2	4	1	1
20.2.99	Stress.abiotic.unspecified	28	7.7		4	11	7	1	5
26.21	Misc.protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	10	2.8	1	1	3	2	2	1
27.3.11	RNA.regulation of transcription.C <sub>2</sub> H <sub>2</sub> zinc finger family	14	3.9	1	3	3	2	1	4
29.2.1.1.1	Protein.synthesis.ribosomal protein.prokaryotic.chloroplast	19	5.2	3	7	9			
29.3	Protein.targeting.nucleus	30	8.3	1	20	7	1	1	
29.4.1.57	Protein.postranslational modification.kinase.receptor like cytoplasmatic kinase VII	27	7.5		9	4	3	5	6
29.5.3	Protein.degradation.cysteine protease	19	5.2		13	3	1	1	1
30.2.11	Signalling.receptor kinases.leucine rich repeat XI	23	6.4	2	2	5	4	6	4
30.2.17	Signalling.receptor kinases.DUF 26	10	2.8	1	1	4	3	1	
31.1	Cell.organisation	39	10.8	2	27	4		5	1
Sum		362		52	132	66	54	31	27
%		100	100	14.4	36.5	18.2	14.9	8.6	7.5

in the (immediate) lower orders. For example, the detection of 1.1.2 was probably due to the presence of 1.1.2.2, and the detection of 8.2 due to the presence of 8.2.10 (Fig. 3; Supplementary Table S4). Removing the second group is necessary to eliminate the overlapping functional classes.

The remainder 19 functional classes, which encompassed a total of 362 genes (Fig. 3; Table 1; Supplementary Table S5), were considered to be important components in the gene network associated with developmental regulation of fruit malate levels. Given its known role in acidity, *Mal* was added to the list, making the total number of genes to be further analyzed at 363. Based on their putative functions, it could be postulated that: (1) several pathways might be critical for the variation of malate levels in developing fruit, including the malate and pyruvate inter-conversion reaction (8.2.10), photosynthesis (1.1.1.2 and 1.1.2.2), mitochondrial electron transport (9), amino acid degradation (13.2) and flavoloid metabolism, and (2) genes in functional classes C<sub>2</sub>H<sub>2</sub> zinc finger transcription factors (27.3.11, 14 genes), protein posttranslational modification (29.4.1.57, 27 genes), and signaling receptor kinases (30.2.11 and 30.2.17, 33 genes) were likely essential for

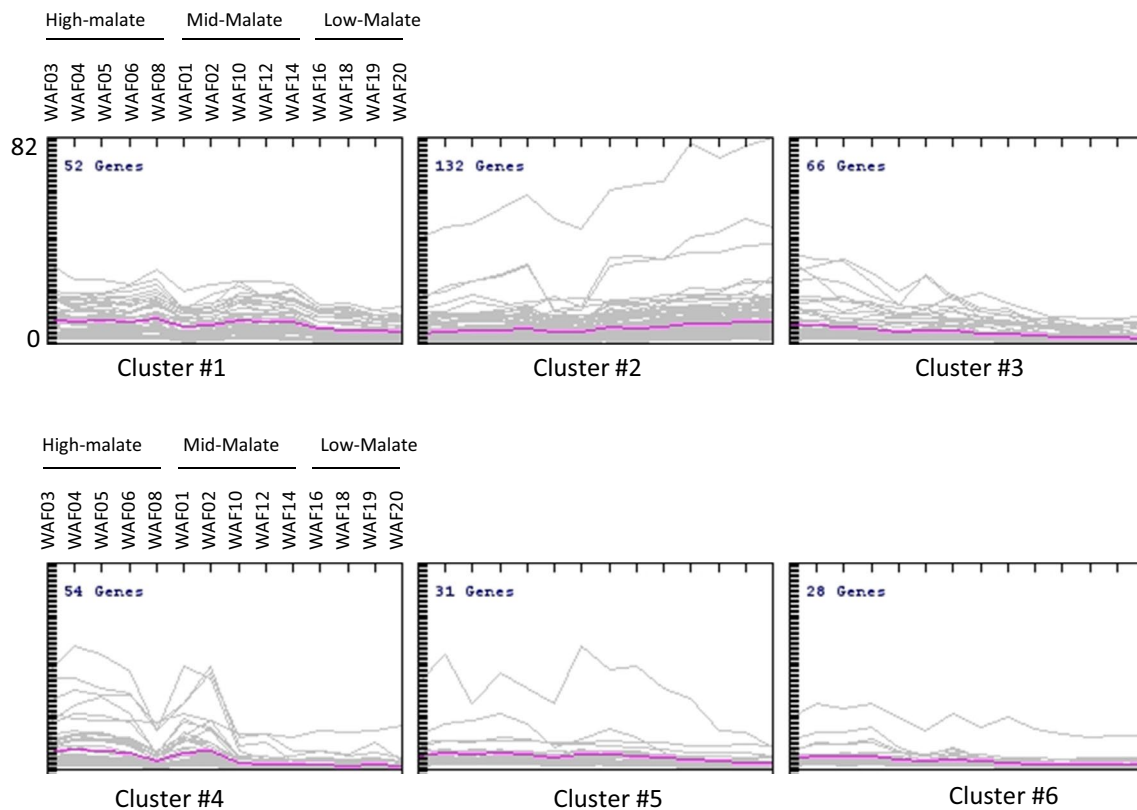
transcriptional regulation (Table 1; Supplementary Table S5).

K-means clustering of the 363 genes using software MeV 4.9 (Saeed et al. 2006) grouped them into six clusters of 52, 132, 66, 54, 32 and 28 genes (Fig. 4; Table 1; Supplementary Table S5). The 132 genes in Cluster 2 were down-regulated in high malate fruit (or in late developmental stage) while the rest 231 in the other five clusters were all up-regulated in the high malate fruit (Fig. 4).

#### Co-expression gene networks associated with developmental regulation of malate levels

Using the network Inferring tool of software Cytoscape 3.1 (Cline et al. 2007), co-expression gene networks were constructed with 246 of 363 genes under Pearson correlation coefficient (PCC)  $r > 0.95$ , which included a major network of 239 genes (Fig. 5a) and three mini networks of seven genes (Fig. 5b). Notably, *Mal* was not present in these networks. Analyzing the networks with a Cytoscape plugin NetworkAnalyzer v2.7 (Assenov et al. 2008) showed that there were 19 genes (nodes) of the highest network degrees (the





**Fig. 4** Expression of K-means clusters of the 363 genes or transcripts in MapMan functional classes that were significantly co-enriched or co-suppressed differentially in the fruit groups of high, mid and low malate. The vertical axis shows the square root transfor-

mation of RPKM values. Note that the expression of genes in Cluster #2 was negatively correlated with malate concentrations, and *Mal* was included in Cluster #6

number of edges) of 26–35 in the major network (Fig. 5c; Table 2) while 68 genes were of the lowest degrees with 24 of degree one, 20 of degree two and 24 of degree three (Fig. 5a–c). The remainder (majority) 159 had degrees 4–24.

The 19 genes of the highest degrees (Figs. 5a, 6a) were considered the core (or hub) of the major co-expression gene network, and they collectively had 106 immediate neighbor genes (Figs. 5a, 6b). Revealing the identity of the 19 genes (Table 2) suggested that there were four (M186555, M798156, M897253 and M157044) putatively encoding for leucine rich repeat (LRR) receptor kinases in signaling, three (M252536, M321839 and M491898) for cell wall degradation related enzymes, three (M575740, M252589 and G104117) for enzymes involved in flavonoid metabolism, one (M246502) for a C<sub>2</sub>H<sub>2</sub> zinc finger transcription factor, one (M258977) for a NADP malic enzyme (ME), and seven for others. These results largely reinforced the two postulations aforementioned.

#### qRT-PCR confirmation of gene expression

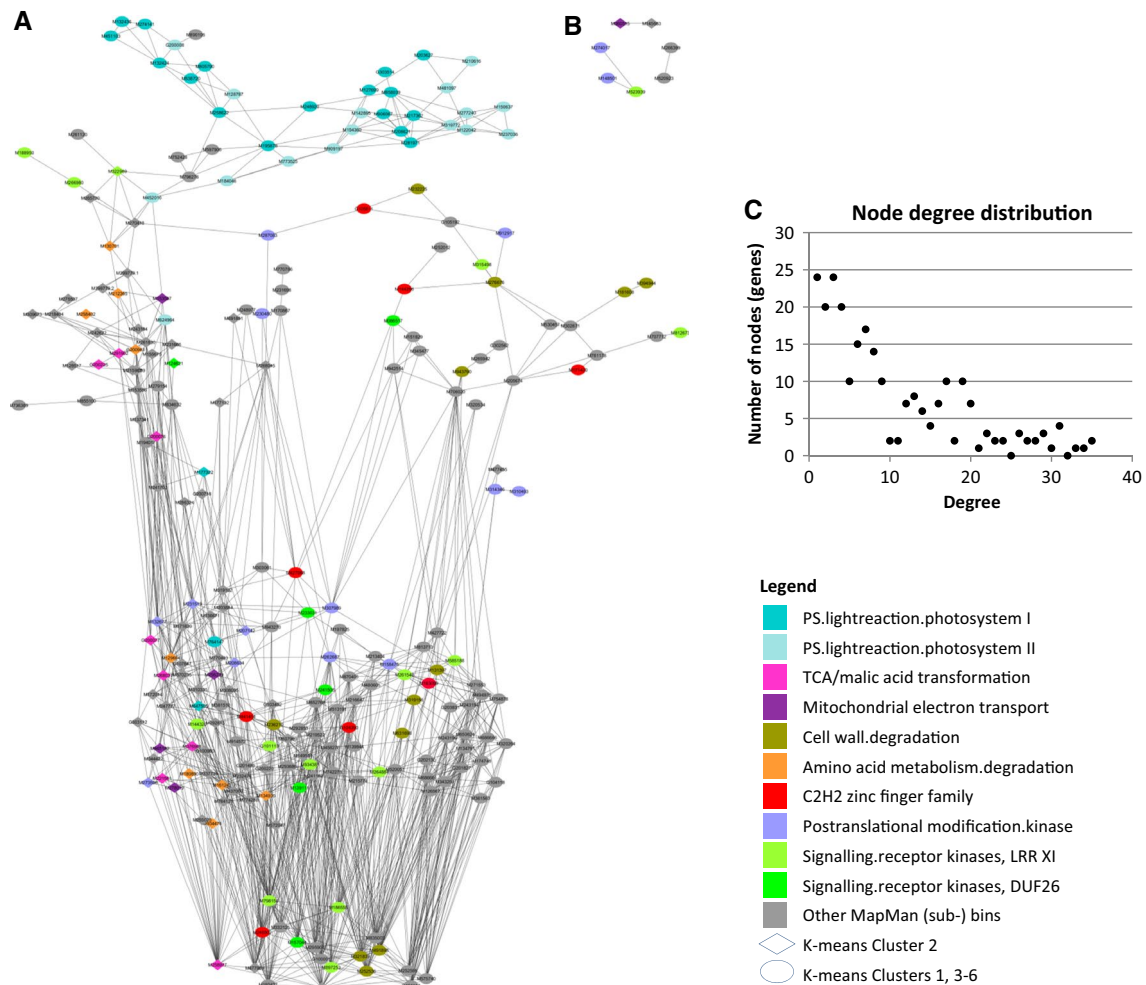
To confirm if the RPKM values reflect the expression levels, six genes *Mal* (M252114), M196894, M127123,

M163222, M190273 and M219042 were analyzed with qRT-PCR assay (Fig. 7). The relative expression of the six genes in qRT-PCR were highly correlated with their RPKM values with  $R^2$  ranged from 0.536 (M190273) to 0.913 (M127123), which exceeded the critical value ( $r^2 = 0.471$ ) at the significance level of  $P_{0.01}$ .

## Discussion

### Identification of genes associated with the variation of malate levels in developing fruit

Changes of malate levels in developing fruit has long been reported and are often characterized by a peak in a period from WAF04–08 (Hulme and Woollorton 1957; Ulrich 1970; Beruter 2004; Zhang et al. 2010). In this study, we found a similar developmental pattern of malate levels in fruit of ‘Golden Delicious’ (GD), i.e. it increased rapidly from WAF02 through WAF5 and then progressively decreased through harvest (WAF20). The malate concentrations varied from 5.15 mg/g fw (WAF19) to



**Fig. 5** A graphic representation of co-expression gene networks comprising 246 of the 363 genes (Supplementary Table S5) associated with developmental regulation of apple fruit acidity. The networks were constructed using the Cytoscape network inferring tool ( $r > 0.95$ ) and analyzed with NetworkAnalyzer (Assenov et al. 2008). *Node color* and *shape* keys are indicated in the legend. Note that the *diamond* and *oval* shapes represent for K-means cluster 2 (negatively correlated with malate levels, see also Fig. 4) and the remainder five

clusters (positively correlated with malate levels), respectively. **a** A major co-expression gene network of 239 members. The *bottom section* shows the 19 genes of the highest degrees at the network core, and the *mid-section* shows 106 immediate neighbors of the 19 genes at the core, and the *top* shows the rest (114 genes). **b** Three mini gene networks of seven members. **c** Degrees of nodes in the major and mini networks (color figure online)

$14.52 \pm 0.48$  mg/g fw (WAF05) in the 14 time points (Fig. 1b), suggesting a strong developmental regulation of fruit malate levels in apple. By analyzing a set of RNA-seq data obtained from the GD developing fruit using the improved apple reference transcriptome (Bai et al. 2014), we identified 3,066 genes that were expressed not only differentially ( $P_{\text{FDR}} < 0.05$ ) between the high and low malate groups and between the early (WAF03–08) and late (WAF16–20) developmental stages, but also in significant ( $P < 0.05$ ) correlation with malate concentrations and/or the expression of *Mal1*. These genes are considered to include a majority, if not all, of the genes that are responsible for malate variations in developing fruit. It should be pointed out that the RNA-seq data were obtained without biological

or technical replicates for 13 (WAF01–19) of the 14 sampling time points. However, each of these time points was represented by five or more pooled fruit, and the transcriptome comparisons were conducted among three groups of fruit comprised at least four sampling time points.

Pathways putatively involved in determining malate levels in developing fruit

To extract relevant biological information from the large set of 3,066 genes, we elected to use the MapMan annotation of functional classes that were designed for plants (Thimm et al. 2004) rather than the general Gene Ontology (GO) terms. Based on the comparative study on the

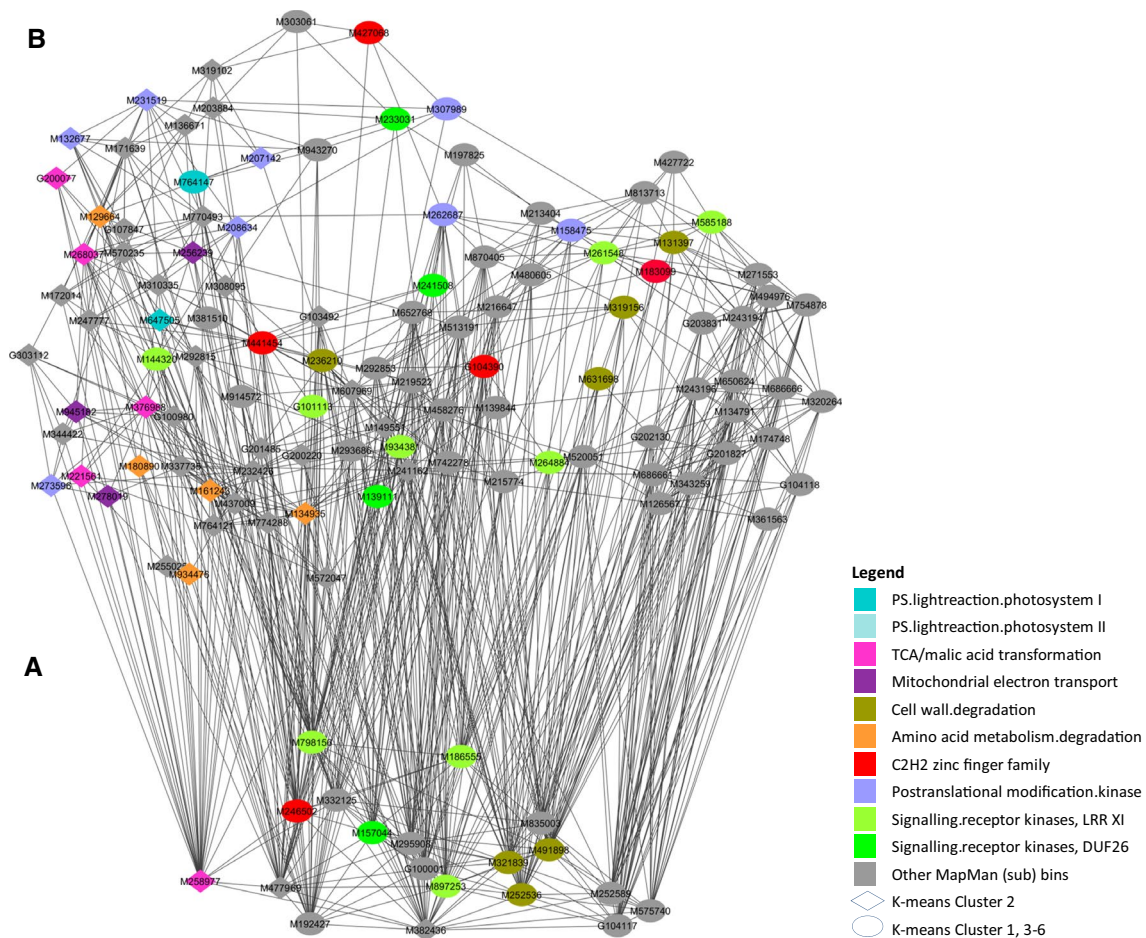
**Table 2** List of core (hub) genes of the major co-expression network

Gene ID	MapMan (sub-) bins	Putative function	Degree	Genomic contig/chromosomal locality
M258977	8.2.10	NADP-malic enzyme	31	MDC010772.65/Chr01_11311642_11314074
M252536	10.6.1	Glycosyl hydrolase, protein endoglucanase 1 precursor, putative	31	MDC016922.37/chr17_8305354_8308999
M321839	10.6.2	1-4-Beta-mannan endohydrolase, putative	31	MDC018286.199/Chr04_17599145_17601066
M491898	10.6.3	Pectate lyase family protein	35	MDC009663.119/chr13_7209618_7211439
M575740	16.8.2	naringenin-chalcone synthase	26	MDC011072.389/Chr09_16000950_16002755
M252589	16.8.2	chalcone-flavanone isomerase family protein	27	MDC016452.88/chr12_26160009_26162017
G104117	16.8.3.2	Flavonone-3-hydroxylase	29	MDC011373.244/chr02_11324973_11333001
M295908	20.2.99	Pollen Ole e 1 allergen and extensin family protein	29	MDC012046.326/Chr09_4585826_4587577
G100001	20.2.99	Pollen Ole e 1 allergen and extensin family protein	31	MDC021673.67/chr09_4996315_4998730
M835003	26.21	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	28	MDC020266.309/chr14_27933836_27934129
M246502	27.3.11	Zinc finger (C <sub>2</sub> H <sub>2</sub> type) family protein	26	MDC029184.30/chr11_7014798_7015913
M192427	29.3.4.1	ER lumen protein retaining receptor, putative	28	MDC009045.262/Chr07_26304595_26306870
M332125	29.3.4.99	SEC14 cytosolic factor family protein/phosphoglyceride transfer family protein	29	MDC013159.462/chr12_25250564_25252501
M798156	30.2.11	Leucine rich repeat (LRR) receptor kinase	33	MDC014159.224/chr15_13028580_13031710
M897253	30.2.11	Leucine rich repeat (LRR) receptor kinase	26	MDC012729.141/chr12_21126633_21129746
M186555	30.2.11	Leucine rich repeat (LRR) receptor kinase	27	MDC011209.233/Chr02_5517287_5518621
M157044	30.2.17	Leucine-rich repeat family protein/protein kinase family protein	30	MDC000636.612/chr15_24576085_24579174
M382436	31.1	Plastid-lipid associated protein PAP/fibrillin family-like protein	35	MDC022173.275/Chr04_10225653_10239011
M477969	31.1	Plastid-lipid associated protein PAP/fibrillin family-like protein	34	MDC000741.237/Chr06_2998593_3000895

two ontologies MapMan and GO (Klie and Nikoloski 2012), and the objective of this study, it appeared that MapMan based annotation was more appropriate. The 3,066 genes were assigned into 648 functional classes or MapMan (sub-) bins, of which 1,949 genes (63.6 %) were of a putative function while 1,117 (36.4 %) of unknown function (Supplementary Fig. S1). For cross-reference, we also identified the GO terms available for the 3,066 genes (Supplementary Table S3). The GO terms for MDPs, i.e. M#####s, are available in GDR. For the new genes (Bai et al. 2014), i.e. G#####s, their GO terms were obtained by BLAST2GO program (Götz et al. 2008) in conjunction with BLASTx search against the NCBI Reference Proteins database (cutoff  $E = 10^{-6}$ ). A hypergeometric test similar to the unconditional GOSTats test (Falcon and Gentleman 2007) against all the 43,496 expressed genes was conducted using CLC GW, and identified 13 GO categories significantly ( $P_{FDR} < 0.05$ ) over represented in the 3,066 genes relative to the full set in GO biological process (BF), 16 in GO cellular component (CC), and 11 in GO molecular function (MF) (Supplementary Table S6). There were zero GO categories significantly under represented.

The PageMan tool (Usadel et al. 2006) has been used widely in plants to condense large data set by identifying enriched pathways or MapMan functional classes (Usadel et al. 2009; Gowik et al. 2011; Klie and Nikoloski 2012; Secco et al. 2013). Using this tool, we found that 19 of the 648 MapMan functional classes were significantly ( $P_{FDR} < 0.05$ ) co-enriched or co-suppressed in a malate dependent manner (Fig. 3; Table 1; Supplementary Table S4). The 19 functional classes comprised 363 (362 plus *Mal*) genes, and K-means clustering of these 363 genes revealed that expression of 132 (36.5 %) of them was negatively correlated with malate levels while the expression of the rest 231 positively (Fig. 4; Table 1). The 19 functional classes included several known primary metabolic pathways for malate synthesis and degradation, such as the malate and pyruvate interconversion reaction and photosynthesis, as well as those that have indirect effects on malate, such as mitochondrial electron transport and amino acid metabolism.

The malate and pyruvate interconversion reaction is one of the key steps under the current model of potential metabolic pathways involving malate in fruit (Sweetman et al. 2009; Etienne et al. 2013). The interconversion



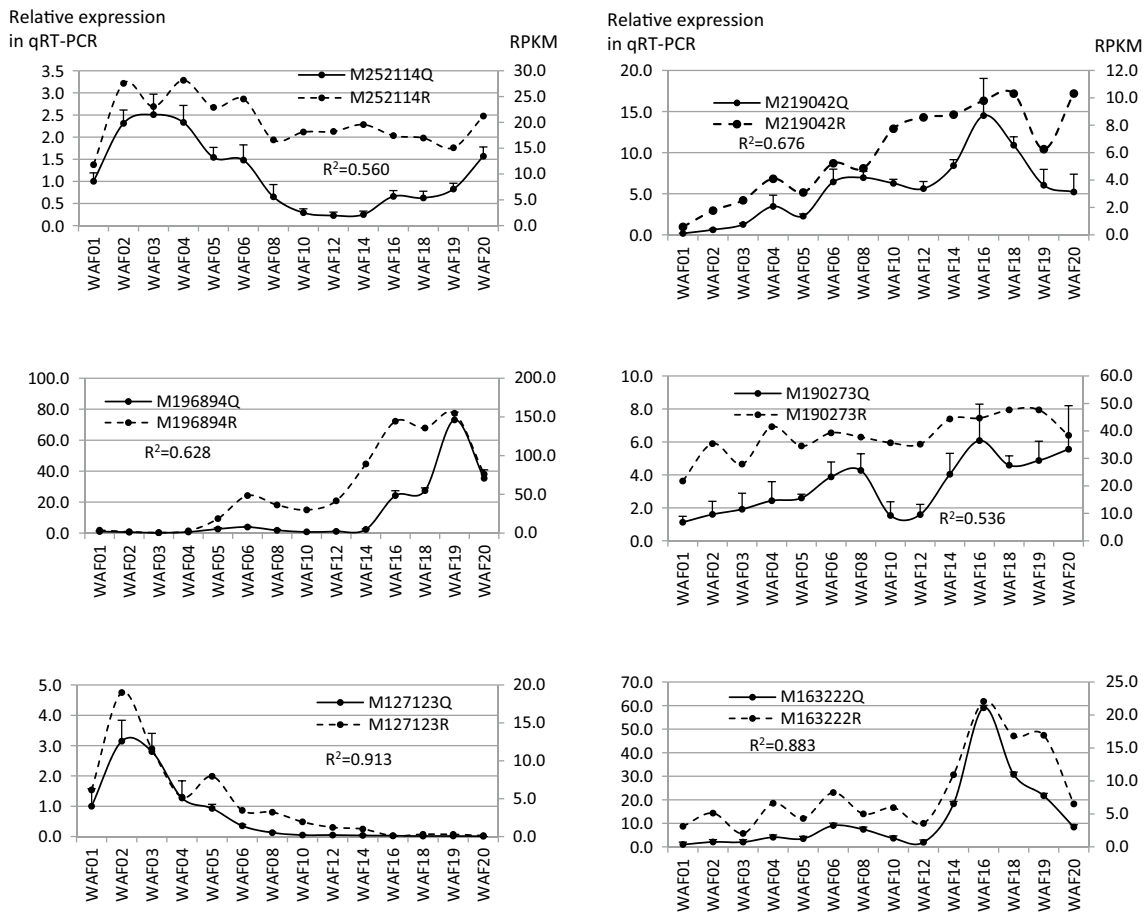
**Fig. 6** A detailed view of the 19 core or hub members (a) and their immediate 106 neighbors (b) in the co-expression network. The *node color and shape keys* are the same as noted in Fig. 5 (color figure online)

reaction occurs in cytosol and in chloroplast and is more favored from malate to pyruvate catalyzed by NADP dependent malic enzymes (NADP-ME). The identification of sub-bin 8.2.10 of nine NADP-ME encoding genes as one of the 19 functional classes strongly suggested the role of the malate and pyruvate interconversion reaction in malate levels. Since they were expressed significantly higher only in low malate group and were all fallen into K-means Cluster 2, the nine NADP-ME encoding genes appeared to negatively affect malate levels in fruit, i.e. higher expression leads to low malate levels. Based on the sequence similarity annotations of the nine NADP-MEs (Supplementary Table S5), six were similar to AtNADP ME1, two (including M258977 at the core of the gene network) to AtNADP ME3, and one to AtNADP ME4. Since enzymes AtNADP-ME2 and AtNADP-ME3 were cytosolic and AtNADP ME4 was plastidic (Wheeler et al. 2005), the negative correlation was likely a reflection of the cytosolic reaction rather than in plastids. Indeed, enzyme activities of a characterized cytosolic NADP-ME (DQ280492),

which is nearly identical to M192078, has been shown to negatively contribute to malate accumulation in apple fruit (Yao et al. 2009), although the difference in the expression of M192078 was insignificant between the high and low malate groups in this study. Moreover, a recent study on the plastidic NADP-ME that was knocked-out by RNAi in ripening tomato fruit showed that malic acid was reduced (Osorio et al. 2013), suggesting that disrupting the interconversion between pyruvate and malate in plastid would reduce rather than increase the malic acid levels in fruit.

Developing apple fruit are capable of photosynthesis although the CO<sub>2</sub> assimilated is primarily from the mitochondrial respiration (Blanke and Lenz 1989). In this pathway, oxaloacetate is produced through the action of phosphoenolpyruvate carboxylase (PEPC) and then malate by NADP malate dehydrogenase (NADP-MDH). The expression of the two functional classes (sub-bins) 1.1.1.2 of 29 genes associated with photosystems I (PSI) and 1.1.2.2 of 18 from PSII were significantly higher in the high- and mid- malate groups while normal in the low malate group





**Fig. 7** Confirmation of gene expression of six selected genes using qRT-PCR. The normalized expression of target genes relative to a control gene (actin) in qRT-PCR is shown in *solid line* (against the

primary vertical axis), and their corresponding RPKM values from RNA-seq are in *broken line* (against the secondary vertical axis). The coefficient of determination ( $R^2$ ) was shown accordingly

(Fig. 3). K-means clustering of these two sub-bins indicated that 45 of the 47 genes falling into Clusters 1 and 3 that were positively correlated with malate levels (Table 1). These data suggest that a more active photosynthesis process in relatively young fruit may have facilitated the biosynthesis of malate, contributing to high malate levels in the high-malate group (WAF3–8).

Mitochondrial electron transport is one of the major respiratory pathways in which the reducing equivalents (e.g. NADH) generated from the TCA cycle are used to drive the synthesis of ATP (Fernie et al. 2004). While generating ATP, the complex I of electron transport chain also makes  $\text{NAD}^+$  available from NADH for the TCA cycle where the recycling of  $\text{NAD}^+$  is required. Therefore, the TCA cycle and the electron transport chain are in fact an integrated process, largely explaining the putative role of mitochondrial electron transport in fruit malate levels. There were nine genes in the mitochondrial electron transport functional class (bin 9), with three for complex I, two for complex III and four for ATP synthases (Table 1;

Supplementary Table S5). The nine genes were expressed at significantly higher levels in the low-malate group but normal in the high- and mid-malate groups (Fig. 3), and eight of them were in K-means Cluster II (Table 1; Fig. 4). These data suggested that high expression of genes in the electron transport chain was correlated with low malate levels. However, this is inconsistent with the findings that malate and citrate increased in the leaves of the CMSII mutant in *Nicotiana sylvestris* that lacks functional complex I (Gutierrez et al. 1997; Noctor et al. 2004).

The 11 genes identified in the functional class amino acid metabolism-degradation (sub-bin 13.2) were expressed at normal levels in the high- and mid-malate groups but at significantly higher levels in the low malate groups (Fig. 3; Table 1; Supplementary Table S5). Of these genes, four were for degradation of the aspartate family amino acids (asparagine, lysine, threonine, and methionine), five for the aromatic amino acids including tyrosine (4) and tryptophan (1), and two for branched chain amino acids (Leucine or isoleucine or valine). Although these amino acids have their



unique degradation pathways, the pathways could converge into intermediates acetyl CoA, fumarate, pyruvate, oxaloacetate, and succinyl-CoA (Nelson et al. 2008). Obviously, all of the intermediates can enter the TCA cycle, the known major pathway for malate metabolism, thereby accounting for the involvement of these amino acid degradation genes in fruit acidity.

#### Regulation of the major co-expression gene network associated with malate in developing fruit

The major co-expression gene network comprised 239 of 363 genes in the 19 MapMan functional classes that were co-enriched or co-suppressed in malate dependent manner (Fig. 5a). There were 19 genes of the highest network degrees (Figs. 5, 6; Table 2), which were considered to be the core (hub) genes of network and assumed to have greater roles in the network. The major gene *Mal* for fruit acidity was not present in the network, suggesting that the major co-expression gene network associated with developmental regulation of malate levels is not directly mediated by *Mal*. This may be attributed in part to the fact that *Mal* exerts its effect on fruit acidity by a combination of a premature stop codon leading mutation and expression levels, where the mutation seems to be more important (Bai et al. 2012). However, it is still possible that the variation in malate concentration actually involved *Mal*. This could be evidenced from the fact that 81 of the 239 genes in the network were initially identified through their significant correlations with *Mal* although 75 of them were also correlated with malate concentrations. In grape, a *Mal* like gene *VvALMT9* has been demonstrated to mediate both malate and tartrate accumulation in developing berries (Angeli et al. 2013). There are 114 co-expressed genes of *VvALMT9* in the ViTis Co-expression DataBase, called VTCdb (Wong et al. 2013) ([http://vtcdb.adelaide.edu.au/HRR\\_berry.aspx?Unique\\_ID=VIT\\_01s0011g03290](http://vtcdb.adelaide.edu.au/HRR_berry.aspx?Unique_ID=VIT_01s0011g03290)). Comparing the 886 (100 plus 776) co-expressed genes of *Mal* and the 114 genes of *VvALMT9* identified 27 grape genes that shared significant sequence and function similarities with 81 apple genes (Supplementary Table S7). Among the 27 grape genes, the four receptor kinases encoding genes (VIT\_14s0083g00330, VIT\_07s0031g01440, VIT\_08s0056g01290, and VIT\_08s0007g02290) are most notable, which show high levels of relatedness to a set of 31 apple genes encoding receptor kinases (30.2.11 and 30.2.17) or protein kinases (29.4.1.57) (Supplementary Table S7). Surprisingly, four of the 19 core or hub genes M798156, M897253, M186555 and M157044 (Table 2; Figs. 5, 6) were among the 31 apple genes, suggesting not only their potential roles but also the *Mal*'s in regulating the malate levels in developing apple fruit.

Members of the ALMT1 protein family are known for their critical role in aluminum tolerance in plants (Hoenkenga et al. 2006; Sasaki et al. 2004; Collins et al. 2008), where the efflux organic anions especially malate from roots is a common mechanism. Under the current transcriptional regulation model of aluminum tolerance in *Arabidopsis* (Delhaize et al. 2012), sensitive to proton rhizotoxicity 1 (STOP1), a C<sub>2</sub>H<sub>2</sub> zinc finger transcriptional factor, controls the expression of *AtALMT1*, multidrug and toxic compound extrusion 1 (*AtMATE1*), aluminum sensitive 3 (*AtALS3*), and other proton and Al<sup>3+</sup> responsive genes (Delhaize et al. 2012). It was evidenced that protein kinases might activate STOP1 by phosphorylation and the activated STOP1 would then bind to the promoters of the targeted genes to initiate transcription. However, the sensor elements interacting with Al<sup>3+</sup> and/or protons were unknown. A model comprising similar elements was also proposed to account for the Al<sup>3+</sup> tolerance in rice (Delhaize et al. 2012). The counterpart of STOP1 is Al<sup>3+</sup> resistance transcription factor 1 (ART1), another C<sub>2</sub>H<sub>2</sub> transcription factor, and ART1 had been proven to regulate the expression of at least 31 genes related to detoxification mechanisms (Yamaji et al. 2009). The 19 functional classes that were expressed in a malate dependent manner in developing fruit included important elements similar to those in the aluminum tolerance models, such as C<sub>2</sub>H<sub>2</sub> zinc finger transcriptional factors (14 genes in sub-bin 27.3.11), protein kinase for post translational modification (27 genes in sub-bin 29.4.1.57) and leucine rich repeat (LRR) receptor kinases for signaling (23 genes in sub-bin 30.2.11 and 10 genes in sub-bin 30.2.17). Interestingly, G105811, one of the 14 C<sub>2</sub>H<sub>2</sub>s in sub-bin 27.3.11 and a new gene identified from contig MDC016907.364 in the improved reference transcriptome shared the highest sequence similarities with STOP1 in *Arabidopsis* and is the only one annotated as an ortholog of STOP1 by Mercator (Lohse et al. 2014). It would be of interest to investigate if any of these genes would directly or indirectly interact with *Mal*.

The identification of the 18 genes (sub-bin 10.6) associated with cell wall degradation suggested that regulation of malate levels in developing fruit might also involve the roles of apoplast (e.g. pectin). There were two lines of evidence in the literature supporting the possible role of apoplast in fruit malate levels: (1) Pectin content and degree of methylation in root–apex was relevant for Al<sup>3+</sup> tolerance where malate efflux is an important mechanism (Wehr et al. 2003; Eticha et al. 2005; Smith et al. 2011; Yang et al. 2013). (2) The shift of fruit phloem unloading from symplastic to apoplastic pathway prior to maturity was coupled with sharper decline in malate in vacuole than in apoplast in ripening grape berries (Zhang et al. 2006; Keller and Shrestha 2014). In other fruits, such as cucumber (Hu et al. 2011) and tomato (Ruan et al. 1995), similar fruit phloem

unloading shift was reported as well. In apple ('Golden Delicious') developing fruit, phloem unloading was also observed to be via apoplast (Zhang et al. 2004).

The roles of genes in other seven functional classes related to secondary metabolism (16.1.4 and 16.8), stress (20.1.7 and 20.2.99), protein synthesis (29.2.1.1.1), protein targeting (29.3), protein degradation (29.5.3), cell organization (31.1) and miscellaneous (26.21) will not be discussed due to the complexity of their functions and the limited scope of the study. Obviously, whether or not and how any of these 363 genes would play a role in regulating fruit malate levels are questions that can only be answered by functional studies in the future. An important direction would be to investigate how gene *Mal* might interact with the genes putatively encoding receptor kinases (30.2.11 and 30.2.17), protein kinases (29.4.1.57) and C<sub>2</sub>H<sub>2</sub> transcription factors (27.3.11).

In conclusion, through analysis of RNA-seq data derived from 'Golden Delicious' fruit at 14 developmental stages, we identified 19 functional classes that were co-enriched or co-suppressed in a malate dependent manner, which encompassed 363 genes. The putative function of these genes suggested that several pathways might be critical for the variation of malate in developing fruit, including the malate and pyruvate interconversion reaction, photosynthesis, mitochondrial electron transport and amino acid degradation. The identification of functional classes of C<sub>2</sub>H<sub>2</sub> zinc finger transcription factors, protein posttranslational modification, and signaling receptor kinases suggested they are likely essential for transcriptional regulation in the major co-expression gene network associated with developmental regulation of fruit malate levels. The results provided transcriptomic clues relevant for better understanding the change of malate levels in developing apple fruit.

**Acknowledgments** We thank Dr. Zhangjun Fei and Dr. Yi Zheng for their helpful comments on the RNA seq data analysis. We also thank Drs Doreen Main and Stephen Ficklin for formatting the revised apple reference transcriptome and making it publicly available through the Genome Database for Rosaceae (GDR). This work was supported in part by the Agriculture and Food Research Initiative competitive grant no. 2014-67013-21660 of the USDA National Institute of Food and Agriculture.

## References

- Angeli A, Baetz U, Francisco R, Zhang J, Chaves M, Regalado A (2013) The vacuolar channel VvALMT9 mediates malate and tartrate accumulation in berries of *Vitis vinifera*. *Planta* 238:283–291
- Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol* 48:381–390
- Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24:282–284

- Baggerly KA, Deng L, Morris JS, Aldaz CM (2003) Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19:1477–1483
- Bai Y, Dougherty L, Li M, Fazio G, Cheng L, Xu K (2012) A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple. *Mol Genet Genomics* 287:663–678
- Bai Y, Dougherty L, Xu K (2014) Towards an improved apple reference transcriptome using RNA-seq. *Mol Genet Genomics* 289:427–438
- Barbier-Brygoo H, De Angeli A, Filleur S, Frachisse J-M, Gambale F, Thomine S, Wege S (2011) Anion channels/transporters in plants: from molecular bases to regulatory networks. *Annu Rev Plant Biol* 62:25–51
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B-Methodol* 57:289–300
- Beruter J (2004) Carbohydrate metabolism in two apple genotypes that differ in malate accumulation. *J Plant Physiol* 161:1011–1029
- Blanke MM, Lenz F (1989) Fruit photosynthesis. *Plant Cell Environ* 12:31–46
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang P-L, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2:2366–2382
- Collins NC, Shirley NJ, Saeed M, Pallotta M, Gustafson JP (2008) An ALMT1 gene cluster controlling aluminum tolerance at the *Alt4* locus of rye (*Secale cereale* L.). *Genetics* 179:669–682
- Delhaize E, Ma JF, Ryan PR (2012) Transcriptional regulation of aluminium tolerance genes. *Trends Plant Sci* 17:341–348
- Emmerlich V, Linka N, Reinhold T, Hurth MA, Traub M, Martinoia E, Neuhaus HE (2003) The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *P Natl Acad Sci USA* 100:11122–11126
- Eticha D, Stass A, Horst WJ (2005) Cell-wall pectin and its degree of methylation in the maize root-apex: significance for genotypic differences in aluminium resistance. *Plant Cell Environ* 28:1410–1420
- Etienne A, Génard M, Lobit P, Mbéguié-A-Mbéguié D, Bugaud C (2013) What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. *J Exp Bot* 64:1451–1469
- Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23:257–258
- Fernie AR, Carrari F, Sweetlove LJ (2004) Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol* 7:254–261
- Gasic K, Hernandez A, Korban SS (2004) RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol Biol Rep* 22:437–438
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435
- Gowik U, Bräutigam A, Weber KL, Weber APM, Westhoff P (2011) Evolution of C4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C4? *Plant Cell* 23:2087–2105
- Gutierrez S, Sabar M, Lelandais C, Chetrit P, Diolez P, Degand H, Boutry M, Vedel F, de Kouchkovsky Y, De Paeppe R (1997) Lack of mitochondrial and nuclear-encoded subunits of complex I and alteration of the respiratory chain in *Nicotiana*

- sylvestris* mitochondrial deletion mutants. P Natl Acad Sci USA 94:3436–3441
- Hoekenga OA, Maron LG, Pineros MA, Cancado GMA, Shaff J, Kobayashi Y, Ryan PR, Dong B, Delhaize E, Sasaki T, Matsumoto H, Yamamoto Y, Koyama H, Kochian LV (2006) AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis*. P Natl Acad Sci USA 103:9738–9743
- Hu L, Sun H, Li R, Zhang L, Wang S, Sui X, Zhang Z (2011) Phloem unloading follows an extensive apoplasmic pathway in cucumber (*Cucumis sativus* L.) fruit from anthesis to marketable maturing stage. Plant Cell Environ 34:1835–1848
- Hulme AC, Wooltorton LSC (1957) The organic acid metabolism of apple fruits: changes in individual acids during growth on the tree. J Sci Food Agr 8:117–122
- Jung S, Ficklin SP, Lee T, Cheng C-H, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S, Evans K, Peace C, Abbott AG, Mueller LA, Olmstead MA, Main D (2014) The Genome Database for Rosaceae (GDR): year 10 update. Nucleic Acids Res 42:D1237–D1244
- Kang C, Darwish O, Geretz A, Shahan R, Alkharouf N, Liu Z (2013) Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. Plant Cell 25:1960–1978
- Keller M, Shrestha P (2014) Solute accumulation differs in the vacuoles and apoplast of ripening grape berries. Planta 239:633–642
- Kenis K, Keulemans J, Davey M (2008) Identification and stability of QTLs for fruit quality traits in apple. Tree Genet Genomes 4:647–661
- Khan S, Beekwilder J, Schaart J, Mumm R, Soriano J, Jacobsen E, Schouten H (2013) Differences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. Tree Genet Genomes 9:475–487
- Klie S, Nikoloski Z (2012) The choice between mapman and gene ontology for automated gene function prediction in plant science. Front Genet 3:1–14
- Kovermann P, Meyer S, Hortensteiner S, Picco C, Scholz-Starke J, Ravera S, Lee Y, Martinoia E (2007) The *Arabidopsis* vacuolar malate channel is a member of the ALMT family. Plant J 52:1169–1180
- Liebhart R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003) Mapping quantitative physiological traits in apple (*Malus x domestica* Borkh.). Plant Mol Biol 52:511–526
- Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc 1:387–396
- Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ 37:1250–1258
- Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrieling-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73
- Meyer S, Scholz-Starke J, De Angeli A, Kovermann P, Burla B, Gambale F, Martinoia E (2011) Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. Plant J 67:247–257
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628
- Nelson DL, Lehninger AL, Cox MM (2008) Pathways of Amino Acid Degradation. Lehninger principles of biochemistry, 5th edn. W.H. Freeman, New York, pp 687–706
- Noctor G, Dutilleul C, De Paepe R, Foyer CH (2004) Use of mitochondrial electron transport mutants to evaluate the effects of redox state on photosynthesis, stress tolerance and the integration of carbon/nitrogen metabolism. J Exp Bot 55:49–57
- Osorio S, Vallarino JG, Szecowka M, Ufaz S, Tzin V, Angelovici R, Galili G, Fernie AR (2013) Alteration of the Interconversion of pyruvate and malate in the plastid or cytosol of ripening tomato fruit invokes diverse consequences on sugar but similar effects on cellular organic acid, metabolism, and transitory starch accumulation. Plant Physiol 161:628–643
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45
- Ruan Y, Mate C, Patrick J, Brady C (1995) Non-destructive collection of apoplast fluid from developing tomato fruit using a pressure dehydration procedure. Funct Plant Biol 22:761–769
- Saeed A, Bhagabati N, Braisted J, Liang W, Sharov V, Howe E, Li J, Thiagarajan M, White J, Quackenbush J (2006) TM4 microarray software suite. Methods Enzymol 411:134–193
- Sasaki T, Yamamoto Y, Ezaki B, Katsuhara M, Ahn SJ, Ryan PR, Delhaize E, Matsumoto H (2004) A wheat gene encoding an aluminum-activated malate transporter. Plant J 37:645–653
- Schumacher K, Krebs M (2010) The V-ATPase: small cargo, large effects. Curr Opin Plant Biol 13:724–730
- Secco D, Jabnourne M, Walker H, Shou H, Wu P, Poirier Y, Whelan J (2013) Spatio-temporal transcript profiling of rice roots and shoots in response to phosphate starvation and recovery. Plant Cell 25:4285–4304
- Smith E, Naik D, Cumming JR (2011) Genotypic variation in aluminum resistance, cellular aluminum fractions, callose and pectin formation and organic acid accumulation in roots of *Populus* hybrids. Environ Exp Bot 72:182–193
- Sweetman C, Deluc LG, Cramer GR, Ford CM, Soole KL (2009) Regulation of malate metabolism in grape berry and other developing fruits. Phytochemistry 70:1329–1344
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939
- Ulrich R (1970) Organic acids. In: Hulme A (ed) The biochemistry of fruit and their products. Academic Press, London and New York, pp 89–118
- Usadel B, Nagel A, Steinhauser D, Gibon Y, Blasing O, Redestig H, Sreenivasulu N, Krall L, Hannah M, Poree F, Fernie A, Stitt M (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics 7:535
- Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ 32:1211–1229
- Wang A, Xu K (2012) Characterization of two orthologs of reversion-to-ethylene sensitivity1 in apple. J Mol Biol Res 2:24–41
- Wehr JB, Menzies NW, Blamey FPC (2003) Model studies on the role of citrate, malate and pectin esterification on the enzymatic degradation of Al- and Ca-pectate gels: possible implications for Al-tolerance. Plant Physiol Bioch 41:1007–1010
- Wheeler MCG, Tronconi MA, Drincovich MF, Andreo CS, Flüge U-I, Maurino VG (2005) A comprehensive analysis of the nadp-malic enzyme gene family of *Arabidopsis*. Plant Physiol 139:39–51
- Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1:80–83

- Wong D, Sweetman C, Drew D, Ford C (2013) VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genom* 14:882
- Xu K, Wang A, Brown S (2012) Genetic characterization of the *Ma* locus with pH and titratable acidity in apple. *Mol Breeding* 30:899–912
- Yamaji N, Huang CF, Nagao S, Yano M, Sato Y, Nagamura Y, Ma JF (2009) A zinc finger transcription factor *art1* regulates multiple genes implicated in aluminum tolerance in rice. *Plant Cell* 21:3339–3349
- Yang XY, Zeng ZH, Yan JY, Fan W, Bian HW, Zhu MY, Yang JL, Zheng SJ (2013) Association of specific pectin methylsterases with Al-induced root elongation inhibition in rice. *Physiol Plantarum* 148:502–511
- Yao YX, Li M, Liu Z, You CX, Wang DM, Zhai H, Hao YJ (2009) Molecular cloning of three malic acid related genes *MdPEPC*, *MdVHA-A*, *MdcyME* and their expression analysis in apple fruits. *Sci Hortic* 122:404–408
- Yao Y-X, Li M, Zhai H, You C-X, Hao Y-J (2011) Isolation and characterization of an apple cytosolic malate dehydrogenase gene reveal its function in malate synthesis. *J Plant Physiol* 168:474–480
- Zhang LY, Peng YB, Pelleschi-Travier S, Fan Y, Lu YF, Lu YM, Gao XP, Shen YY, Delrot S, Zhang DP (2004) Evidence for apoplastic phloem unloading in developing apple fruit. *Plant Physiol* 135:574–586
- Zhang XY, Wang XL, Wang XF, Xia GH, Pan QH, Fan RC, Wu FQ, Yu XC, Zhang DP (2006) A shift of phloem unloading from symplasmic to apoplastic pathway is involved in developmental onset of ripening in grape berry. *Plant Physiol* 142:220–232
- Zhang YZ, Li PM, Cheng LL (2010) Developmental changes of carbohydrates, organic acids, amino acids, and phenolic compounds in ‘Honeycrisp’ apple flesh. *Food Chem* 123:1013–1018