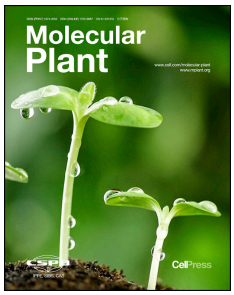
Journal Pre-proof



MicroTom Metabolic Network: Rewiring Tomato Metabolic Regulatory Network throughout the Growth Cycle

Yan Li, Yang Chen, Lu Zhou, Shengjie You, Heng Deng, Ya Chen, Saleh Alseekh, Yong Yuan, Rao Fu, Zixin Zhang, Dan Su, Alisdair R. Fernie, Mondher Bouzayen, Tao Ma, Mingchun Liu, Yang Zhang

PII: S1674-2052(20)30183-0
DOI: https://doi.org/10.1016/j.molp.2020.06.005
Reference: MOLP 946

To appear in: *MOLECULAR PLANT*
Accepted Date: 10 June 2020

Please cite this article as: **Li Y., Chen Y., Zhou L., You S., Deng H., Chen Y., Alseekh S., Yuan Y., Fu R., Zhang Z., Su D., Fernie A.R., Bouzayen M., Ma T., Liu M., and Zhang Y.** (2020). MicroTom Metabolic Network: Rewiring Tomato Metabolic Regulatory Network throughout the Growth Cycle. Mol. Plant. doi: https://doi.org/10.1016/j.molp.2020.06.005.

MicroTom Metabolic Network: Rewiring Tomato Metabolic Regulatory Network throughout the Growth Cycle

Yan Li[1,4], Yang Chen[1,4], Lu Zhou[1,4], Shengjie You[1,4], Heng Deng[1], Ya Chen[1], Saleh Alseekh[2], Yong Yuan[1], Rao Fu[1], Zixin Zhang[1], Dan Su[1], Alisdair R. Fernie[2], Mondher Bouzayen[1,3], Tao Ma[1], Mingchun Liu[1]* & Yang Zhang[1]*

[1]Key Laboratory of Bio□resource and Eco□environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China.

[2]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany and Center of Plant Systems Biology and Plant Biotechnology, 4000 Plovdiv, Bulgaria.

[3]GBF, University of Toulouse, INRA, Castanet-Tolosan, France.

[4]These authors contributed equally: Yan Li, Yang Chen, Lu Zhou, Shengjie You.

Correspondence and requests for materials should be addressed to Yang Zhang (email:yang.zhang@scu.edu.cn) and Mingchun Liu (email:mcliu@scu.edu.cn)

Running Title: Metabolic Network Throughout MicroTom Tomato Growth Cycle

Short summary: High-resolution spatio-temporal metabolome and transcriptome data were generated to cover 20 major stages and tissues of MicroTom tomato. This MicroTom Metabolic Network dataset presents the global figure of tomato metabolic regulatory network during its life cycle and helps to verify new regulators controlling important metabolic pathways.

Abstract

Tomato (*Solanum lycopersicum*) is a major horticultural crop worldwide and has emerged as a preeminent model for metabolic research. Although many research efforts have focused on analyzing metabolite variants between varieties and species, the dynamics of metabolic changes during the tomato growth cycle and the regulatory networks underlying these changes are poorly understood. In this study, we integrated high-resolution spatio-temporal metabolome and transcriptome data to systematically explore the metabolic landscape across 20 major tomato growth tissues and stages. In the resulting MicroTom Metabolic Network (MMN), the 540 detected metabolites and their co-expressed genes could be divided into ten distinct clusters that were based on their biological functions. Using this dataset, we constructed a global map of the major metabolic changes that occur throughout the tomato growth cycle and dissected the underlying regulatory network. In addition to verifying previously well-established regulatory networks for important metabolites, we identified novel transcription factors that regulate the biosynthesis of important secondary metabolites, such as steroidal glycoalkaloids and flavonoids. Our findings provide insight into spatio-temporal changes in tomato metabolism and generate a valuable resource for studying metabolic regulatory processes in model plants.

Keywords: tomato; metabolome; transcriptome; transcription factor; co-expression; flavonoids; steroidal glycoalkaloids;

Introduction

Because of its unique flavor and high nutritional value, tomato (*Solanum lycopersicum*) has become one of the world's favorite fruits and is an important source of micronutrients in the human diet (USDA, 2017). Its short life cycle (90–120 days) and self-compatibility make it an important cash crop for both small- and large-scale growers (Carqueijeiro et al., 2018). During the past 50 years, molecular biology studies of tomato have been driven by demands for increased yield and improved nutritional value (fiber, carbohydrates, and phytonutrient compounds) and by consumer preferences (fruit size, color, flavor, and aroma), making tomato a major research focus for fruit biology (Azzi et al., 2015; Klee and Giovannoni, 2011; Tieman et al., 2017). Many genetic resources such as natural varieties (Tieman et al., 2017; Zhao et al., 2019; Zhu et al., 2018), introgression lines (Eshed Y, 1995; Schauer et al., 2006), bioinformatic research tools (Shinozaki et al., 2018; Zouine et al., 2017), mutants generated by gene editing (Shimatani et al., 2017; Uluisik et al., 2016; Zhang et al., 2006), and multi-omics tools (Sonawane et al., 2016) are available to systematically understand metabolic regulatory networks in tomato.

Nowadays, tomato is considered as a model to study plant secondary metabolism. Major classes of plant metabolites, such as flavonoids, carotenoids, terpenoids, and alkaloids, are all present in tomato and are strongly associated with its growth habits and fruit characteristics (Butelli et al., 2008; Lewinsohn et al., 2001; Liu et al., 2015; Xu et al., 2018; Zhang et al., 2015a; Zhu et al., 2018). Flavonoids belong to a group of compounds rich in fruit and are associated with health benefits (Adato et al., 2009). These secondary metabolites are ubiquitous in plants yet have tremendous chemical diversity as well as diverse roles (Grotewold, 2016; Jin et al., 2000; Sheehan et al., 2015; Tohge et al., 2016). In addition, flavonoids are reported to benefit health by having anti-cancerous, antioxidant, and anti-osteoporotic activities (Echeverry et al., 2015; Kaakoush and Morris, 2017). Besides health benefit compounds such as flavonoids, in recent decades, many studies have been done to the regulation of plant

defensive compounds such as steroidal glycoalkaloids (SGAs). SGAs are cholesterol-derived molecules and mainly produced by *Solanaceous* species (Cárdenas et al., 2019; Cárdenas et al., 2016). Although SGAs contribute to plant resistance to a wide range of pathogens and predators, including bacteria, fungi, oomycetes, viruses, insects, and other animals, some of them are considered as antinutritional factors for humans due to their disruption of membranes and inhibition of acetylcholine esterase activity (Itkin et al., 2013).

To date, systematic studies of tomato growth and quality have focused on fruit, because of the commercial importance of fruit traits. Before the tomato genome sequence was released, complementary DNA (cDNA) microarrays were used to study gene expression dynamics during tomato fruit development (Alba et al., 2005). As the genome sequence became available and second generation sequencing techniques were developed, transcriptome and epigenome profiling were used to provide global, quantitative measurements of gene expression during fruit development (Hua et al., 2009; Zhong et al., 2013). More recently, improved sampling techniques, such as laser-capture microdissection, and high-throughput RNA sequencing methods have been used to analyze gene expression at high spatial resolution (Espina et al., 2006; Shinozaki et al., 2018).

During the past 15 years, the integration of metabolic profiling with transcriptome data has proven to be highly effective for identifying gene functions and elucidating metabolic pathways in plants (Luo, 2015). For example, pioneering work by Carrari et al. combined gas chromatography-mass spectrometry (GC-MS) with parallel transcriptome analysis to dissect metabolic changes during tomato fruit development (Carrari et al., 2006). This approach was expanded to compare the peel and flesh tissues of different tomato introgression lines during fruit development (Mintz-Oron et al., 2008). Later, a more detailed metabolic quantitative trait loci (mQTL) analysis was carried out using the well-characterized *Solanum pennellii* introgression lines (Alseekh et al., 2015; Schauer et al., 2006). To evaluate how breeding has changed

the tomato fruit metabolome, different research groups have recently generated and analyzed large datasets encompassing genomes, transcriptomes, and metabolomes from hundreds of tomato genotypes (Tieman et al., 2017; Zhao et al., 2019; Zhu et al., 2018). Through metabolite genome-wide association studies (mGWAS), researchers have identified a large number of genetic loci that affect the concentrations of important metabolites in tomato fruit (Tieman et al., 2017; Zhao et al., 2019). Furthermore, GC- and LC-MS based metabolomics and publically available RNA sequencing data were carried out to investigate polyphenolic metabolism in the lycopersicum complex across eight different species of tomato (Tohge et al., 2020).

Although some multi-omic studies of tomato have examined vegetative tissues (Balcke et al., 2017; Jennifer, 2013; Zouine et al., 2017), most have focused on the effect of genetic variation among varieties and species on fruit quality (Alseekh et al., 2015; Alseekh et al., 2017; Dan et al., 2019; Garbowicz et al., 2018; Tieman et al., 2017; Zhang et al., 2016; Zhao et al., 2019; Zhu et al., 2018). Therefore, the changes that occur to metabolic and regulatory networks throughout the tomato growth cycle have remained largely unknown.

To establish how the metabolism is rewired during the tomato growth cycle and the underlying transcriptional regulation, we developed the MicroTom Metabolic Network (MMN), a high-temporal-resolution metabolome and transcriptome dataset of different tissues for MicroTom tomato at different development stages. We detected 540 annotated metabolites and 31,256 expressed genes in 20 different tissues and stages. Co-expression analysis integrating metabolome and transcriptome data provides comprehensive knowledge of the dynamics of major metabolites during tomato development and the transcriptome profiles underlying them. Using this dataset, we not only verified the previously known regulatory networks for major metabolites, but also identified novel transcription factors associated with important compounds, such as flavonoids and steroidal glycoalkaloids. Our MMN dataset and experimental strategy can be a useful resource for identifying key regulators of

important metabolites in other crops.

Results

Generation of the MMN dataset

To create a comprehensive and accurate record of changes to metabolic regulatory networks during the life cycle of MicroTom tomato, we generated the MMN dataset (Figure 1). This dataset consists of parallel metabolic profiling and transcriptome analysis for samples collected from all major growth stages and tissues of MicroTom tomato. We collected 20 tissues: roots, stems, leaves from the bud stage (when the first flower bud emerges, 30 days post-germination [DPG]), flowering stage (when 50% of flowers reach anthesis, 45 DPG), breaker stage (when the first fruit reaches the breaker stage, 85 DPG). Bud and flower samples were collected from bud stage seedlings and flowering stage seedlings, respectively. In addition, pericarps at nine stages of fruit development stages (10 days post-anthesis [DPA], 20 DPA, immature green [IMG], mature green [MG], breaker [Br], breaker plus 3 days [Br3], Br7, Br10, and Br15) (Figure 1). Three biological replicates, each of which was a pooled sample from 10 plants, were analyzed for all 20 time points/tissues.

For metabolome analysis, samples were analyzed by a broadly targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based metabolic profiling method (Zhu et al., 2018) (see Methods). A total of 540 distinct annotated metabolites were identified in at least one tissue, including 70 flavonoids, 76 amino acids and derivatives, 54 lipids, 52 organic acids, 50 nucleotides and derivatives, 14 phenolamides, 32 alkaloids, 43 hydroxycinnamoyls and derivatives, 9 polyphenols, 13 carbohydrates, 12 vitamins, 10 benzoic acids and derivatives, 18 polyamines, and 87 additional compounds which did not fit into these 13 main classes (Figure 2A and Supplemental Data 1). Analyses of the 540 metabolites among the different tissues showed that the metabolites can be divided into three big groups: metabolites present in vegetative tissues (root, stem, leaf), flower tissues (bud and flower) and fruit tissues

(pericarp throughout fruit development) (Figure 2A). This result indicates that vegetative tissues at different development stages have similar metabolic patterns, which are significantly different from those of the fruit tissues. For example, metabolites such as lipids, alkaloids, hydroxycinnamoyl derivatives, and phenolamides accumulate preferentially during vegetative phases, whereas the fruit tissues have substantially higher levels of amino acids and their derivatives, flavonoids, nucleotides and their derivatives, organic acids, polyphenols, and vitamins (Figure 2A). These specific metabolites likely reflect the spatial differentiation of tomato metabolism between the two major phases: vegetative growth and fruit development. In addition, we performed principal component analysis (PCA) and established that the 540 metabolites can be further divided into five groups, each associated with a specific tissue (root, stem, leaf, flower, and fruit) (Figure 2C). In line with the PCA, the cluster dendrogram can also be divided into five independent subgroups (Figure 2E). All these data indicate that the accumulation of metabolites is tissue specific during tomato development.

To investigate transcriptional regulation over the course of the MicroTom tomato growth cycle, we built a spatio-temporal dynamic transcriptome landscape for all tissues. Approximately 453 Gb of raw data were generated and the statistics of the sequencing libraries are summarized in Supplemental Data 2. Unique mapped reads were used to calculate the expression level of transcripts per million (TPM) (Bo et al., 2010; Wagner et al., 2012). To reduce the influence of transcriptional noise, we defined a gene as expressed if its average TPM value was > 0. In total, 31,256 genes were found to be expressed in at least one sample (Supplemental Data 3). To form a global picture of gene expression across the 20 tissues, we made a Z-score normalized expression heatmap clustered by genes. A correlation matrix of the transcriptome indicates that the global gene expression pattern is tissue-specific, regardless of the development stage (Figure 2B and Supplemental Data 3). As we established for the metabolome, further PCA and the cluster dendrogram of the transcriptome also showed clustering of samples from the same tissue among the different stages (Figure

2D and 2F). All these results suggest that gene expression and metabolite accumulation show significant tissue specificity during tomato development.

The tomato metabolome and transcriptome are co-regulated in ten clusters corresponding to different tissues and developmental stages

To gain further insights into metabolic changes during the MicroTom growth cycle, we divided all 540 annotated metabolites into ten clusters based on their accumulation patterns using the *k*-means clustering algorithm (Howe et al., 2010) (Supplemental figure 1A, Table 1, Supplemental Data 4). By analyzing the ten clusters, we can identify metabolites enriched in specific tissues such as root (Cluster I), stem (Cluster III), leaf (Cluster IV), and flower (Cluster V) (Figure 3 and Supplemental figure 1A). We also identified compounds that decrease during the period between the flowering and ripe fruit stages (Cluster VI) and compounds that increase during the green fruit stages (Cluster VII) and during the fruit ripening stages (Cluster X). In addition, there were compounds that were enriched in more than two tissues or stages (Clusters II, VIII, and IX) (Figure 3 and Supplemental figure 1A).

To further correlate the gene expression pattern with metabolite accumulation, we applied co-expression analysis to our metabolome and transcriptome data (Giovannoni, 2018; Serin et al., 2016). A rigorous multiple test correction ($r \geq 0.8$) was used to filter the genes that significantly correlated with each metabolite. A total of 17,003 genes that are co-regulated with at least one metabolite were identified (Supplemental figure 1B, Supplemental Data 4 and 5).

Next, the 540 metabolites and 17,003 genes were divided into ten co-expression clusters based on Pearson Correlation Coefficient and the fact that they had a consistent and clear expression pattern during tomato development (Figure 3). Interestingly, after removing the genes not highly correlated with any of the 540 metabolites, PCA (Supplemental figure 2A) and clustering analysis (Supplemental figure 2B) of the remaining 17,003 highly co-expressed genes still showed clustering

patterns similar to the global gene expression clusters (Supplemental figure 1B). This indicates that during the normal growth cycle, the pattern of tomato gene expression is largely paralleled by the dynamics of major metabolic pathways.

Spatio-temporal insights into the regulation of key metabolic pathways

To examine whether MMN can provide spatio-temporal insights into the metabolic regulatory network in tomato, we first tested it with previously known regulatory networks. Flavonoid biosynthesis branches from the phenylpropanoid pathway, which provides precursors in the form of phenylalanine. Following the reactions catalyzed by the gateway enzymes phenylalanine ammonia lyase (PAL) and chalcone synthase (CHS), the flavonoid pathway consists of a series of enzymes that catalyze diverse downstream reactions leading to the biosynthesis of aglycone backbones (Figure 4D). Flavonoid biosynthesis has been a growing area of research in the past two decades and much progress has been made in the identification of both the biosynthetic genes and their regulators (Vogt, 2010). Multiple MYB family transcription factors have been identified as regulatory factors in flavonoid biosynthesis (Ballester et al., 2010; Borevitz et al., 2000; Mehrtens et al., 2005). It has been suggested that the expression of *Flavonoid-3',5'-hydroxylase* (*SlF3'5'H*), a flavonoid modification gene, is correlated with *SlMYB75* in vegetative tissues, an anthocyanin biosynthetic transcription factor, rather than *SlMYB12*, the predominant regulator of flavonoid biosynthesis (Ana-Rosa Ballester, 2010; Jian et al., 2019). *SlF3'5'H* expression is also crucial for activation of anthocyanin synthesis in tomato fruit because tomato dihydroflavonol 4-reductase (DFR) prefers dihydromyricetin (3 -OHs on the B ring) over dihydrokaempferol (1 -OH on the B ring) (Silvia et al., 2009).

In line with the previous results, our RNA-seq data for flavonoid biosynthetic genes and associated transcription factors revealed that *SlF3'5'H* is strongly co-expressed with *SlMYB75* but not with *SlMYB12* (Figure 4A and 4B). A quantitative real-time polymerase chain reaction (RT-qPCR) analysis of these genes produced results that

were highly consistent with the RNA-seq data (Supplemental figure 3, Supplemental Data 6). A dual luciferase reporter assay revealed that *SlMYB75* can interact with the promoter region of *SlF3'5'H* to induce much higher expression than that induced by *SlMYB12* (Figure 4C). In summary, our spatiotemporal data shows good correlation between metabolite production and transcriptome changes, suggesting that MMN can be utilized to study the regulation of important metabolic pathways in tomato.

SGA biosynthesis begins with glycolysis, followed by the mevalonate and cycloartenol pathways. The cholesterol pathway provides cholesterol, a precursor of the SGA pathway, and the phytosterol pathway overlaps with the cholesterol pathway (Sonawane et al., 2016) (Supplemental figure 4A, Supplemental figure 5). From cholesterol, there are a series of enzymes that catalyze diverse downstream reactions in the pathway leading to the biosynthesis of SGAs (Supplemental figure 4A). Previous co-expression analyses identified clustered *GLYCOALKALOID METABOLISM* (*GAME*) genes that form the biosynthetic pathway from cholesterol to *α*-tomatine (Itkin et al., 2013). In addition, SlGAME9, an AP2/ERF transcription factor, was found to positively regulate SGA biosynthesis (Cárdenas et al., 2016). Among the 540 metabolites, we detected 29 SGAs which are mainly present in clusters III, V, VI, and X (Supplemental Data 7). Interestingly, we identified three SGAs (dehydrofilotomatine, dehydrotomatine, and dehydrotomatine isomer (25R)) and 15 SGA-related genes in cluster V (Supplemental figure 6A, Supplemental figure 4). Among these, the SGA biosynthetic genes (*SlGAME1*, *SlGAME4*, *SlGAME6*, *SlGAME11*, *SlGAME17*, and *SlGAME25*) co-express well with the regulatory gene *SlGAME9* (Supplemental figure 6B). This result is in perfect concordance with previous co-expression analysis of the *GAME* gene cluster (Cárdenas et al., 2016; Itkin et al., 2013). We also performed RT-qPCR analysis of these genes and established that they were mainly increased in vegetative tissue and decreased during fruit development (Supplemental figure 7). To sum up, using MMN we can perfectly verify previously known metabolic regulatory networks throughout the tomato growth cycle.

Identification of a novel transcription factor regulating steroidal glycoalkaloids metabolism

We next attempted to use MMN to identify new regulators controlling the SGA biosynthetic pathway, given that these defensive compounds are also important for tomato growth and fruit quality (Itkin et al., 2013; Zhu et al., 2018). Our clustering data indicated that 15 SGA-related genes (*SlGAME9* and 14 biosynthetic genes) are co-expressed with three SGAs in cluster V. We searched this cluster and found that a gene (*Solyc01g096370*) encoding a bHLH transcription factor is also strongly co-expressed with these SGAs and metabolic genes (Figure 5A). Further investigation indicated that *Solyc01g096370* shares a large number of co-regulated metabolites and genes with the known regulator *SlGAME9* (Cárdenas et al., 2016) (Supplemental figure 8A). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the common genes and metabolites revealed that the molecular functions are mainly enriched in steroid and alkaloid biosynthesis, which suggests that *Solyc01g096370* can potentially modulate the SGA pathway (Supplemental figure 8B). Intriguingly, the level of correlation between *Solyc01g096370* and the SGA pathway is as good as that of *SlGAME9* (Figure 5A).

Phylogenetic analysis revealed that the bHLH family protein encoded by *Solyc01g096370* belongs to the MYC family (Supplemental figure 9A-B), which is mainly involved in jasmonate signaling (Chen et al., 2016; Deng et al., 2015; Du et al., 2017; Du et al., 2018). Further analysis indicated Solyc01g096370 is SlbHLH114, which belongs to bHLH subfamily 15 (Supplemental figure 9C) (Hua et al., 2015) and is related to jasmonate signaling and development of type VI glandular trichomes in *Solanum lycopersicum* (Kemparaju, 2018). Moreover, SlbHLH114 was clustered with subfamily 8 from *Arabidopsis thaliana* (Supplemental figure 9D), which functions in jasmonate signal transduction pathways and affects formation of root hairs and trichomes (Tominaga-Wada et al., 2011). These data suggest that SlbHLH114 might be involved in SGA metabolism and JA signaling.

We next analyzed the expression levels of *SlbHLH114* at different developmental stages and in different tissues. Our transcriptome analysis indicated that expression of *SlbHLH114*, like that of known SGA biosynthetic genes and metabolites, is high in vegetative tissue and gradually decreases during fruit development to a very low level at fruit ripening stages (Supplemental figure 4B, 9E). RT-qPCR analysis confirmed this expression pattern (Supplemental figure 9F). Subcellular localization showed that SlbHLH114 is exclusively located in the nucleus (Supplemental figure 10).

To investigate the potential function of *SlbHLH114*, we generated transgenic tomato plants overexpressing *SlbHLH114* driven by the fruit specific *E8* promoter. We obtained seventeen T0 plants in which the expression levels of *SlbHLH114* in ripening stage fruit were significantly higher than that in MicroTom (Supplemental figure 11). Two of these lines (Line C and Line I) were chosen for further characterization of T1 (Figure 5B and 5D). We performed RNA-seq analysis of Br3 stage pericarp samples from WT and T1 generation *E8:bHLH114* Line C and Line I plants (Figure 5B). A large proportion of the induced genes were shared by *E8:bHLH114* Line C and Line I, including many genes and metabolites in the SGA biosynthetic pathway (Figure 5C). Compared to MicroTom tomato fruit at the same stage, there were 2,851 genes that were upregulated and 2,153 genes that were downregulated in both overexpressing lines (FC$\geqslant$1.5) (Supplemental figure 12A, Supplemental data 8 and 9).

KEGG enrichment analysis indicated that the upregulated genes are involved in signal transduction, plant-pathogen interaction, and biosynthesis of secondary metabolites that are related to SGAs (Supplemental figure 12B, Supplemental Data 10). RT-qPCR analysis confirmed this observation; in general, genes involved in glycolysis and the mevalonate, cycloartenol, cholesterol, phytosterol, and SGA pathways were all upregulated in both *E8:SlbHLH114* lines (Supplemental figure 13). As a result, metabolic profiling of pericarp samples at the Br10 stage showed that both transgenic lines accumulated significantly higher amounts of SGAs (predominantly tomatidine, hydroxytomatidenol, hydrotomatidine, γ-tomatine, β2-tomatine, and α-tomatine)

(Figure 5C). Taken together, these results indicate that overexpression of *SlbHLH114* in tomato fruit enhances SGA accumulation.

To further investigate whether SlbHLH114 directly induces the expression of SGA biosynthetic genes, we performed dual luciferase reporter assays using *Arabidopsis* protoplast. We cloned the promoters of three SlbHLH114-induced biosynthetic genes involved in the cholesterol precursor pathway (*SlSSR2* and *Sl7-DR2*) and the SGA biosynthetic pathway (*SlGAME4*). Using SlGAME9 as a positive control, the results showed that the activities of these promoters were all significantly higher when *SlbHLH114* is expressed than in the control (Figure 5E and 5F), indicating that SlbHLH114 is a positive regulator of the SGA biosynthetic pathway.

Previously, SlMYC2 (Solyc08g076930) was reported as a key regulator of SGA biosynthesis (Cárdenas et al., 2016). We conducted co-expression analysis of *SlMYC2* and SGA-related genes and found *SlMYC2* cannot co-expressed well with known SGA genes. Instead, it has strong correlations with JA-signaling genes, which is consistent with its important role in regulating JA signaling (Supplemental figure 14) (Du et al., 2017; Liu et al., 2019). We further did co-expression analysis of *SlMYC2* and *SlbHLH114* using our MMN dataset, as well as previously published SGN-TEA (http://tea.solgenomics.net/overview) (Shinozaki et al., 2018) and TomExpress datasets (http://tomexpress.toulouse.inra.fr/) (Zouine et al., 2017). In all three datasets, there is no significant co-expression pattern between *SlMYC2* and *SlbHLH114* (Supplemental figure 15).

As both SlMYC2 and SlbHLH114 belong to the bHLH subfamily 15 (Supplemental figure 16) (Hua et al., 2015), to investigate the possible interaction between SlMYC2 and SlbHLH114, we conducted dual luciferase reporter assays in *Arabidopsis* protoplast and found *probHLH114* can be activated by SlMYC2 (Supplemental figure 17A), indicating that *SlbHLH114* is a target of SlMYC2. On the other hand, we checked previously published SlMYC2 related work and found in *SlMYC2*-RNAi

plants, the expression of *SlbHLH114* is significantly lower than WT plants (Du et al., 2017), which confirmed the potential regulation of SlMYC2 on *SlbHLH114* (Supplemental figure 17B). By contrast, no significant induction of *proSlMYC2* by SlbHLH114 was observed in the dual luciferase reporter assays using *Arabidopsis* protoplast (Supplemental figure 17C). In addition, in our RNA-seq data of *E8:bHLH114* and WT fruit, there is no significant differences of *SlMYC2* expression levels (Supplemental figure 17D). All these data indicate SlbHLH114 cannot directly regulate the expression of *SlMYC2*.

Prediction of a novel transcription factor regulating flavonoid metabolism

Using the same strategy described above, we used MMN to expand our search for transcription factors regulating flavonoid metabolism. Interestingly, when we conducted a co-expression analysis of flavonoid compounds and biosynthetic genes, we found, in addition to the known gene *SlMYB12*, a gene (*Solyc02g077790*) encoding an APETALA2/Ethylene Response Factor (AP2/ERF) transcription factor that is better co-expressed with the flavonoid metabolic pathway (Figure 6A). Further investigation showed that *Solyc02g077790* shares large amount of co-expressed genes and metabolites with *SlMYB12*. KEGG analysis of the common genes and metabolites revealed that the molecular functions are mainly enriched in flavonoid and phenylpropanoid biosynthesis, which suggests that *Solyc02g077790* could potentially modulate the flavonoid pathway (Supplemental figure 18).

Phylogenetic analysis of the AP2/ERF family in *Solanum lycopersicum* indicated that the Solyc02g077790 protein belongs to Ethylene Response Factor (ERF) subgroup G and can be designated as SlERF.G3-like (Supplemental figure 19A) (Liu et al., 2016). RNA-seq and RT-qPCR analyses both indicated that *SlERF.G3-like* is specifically expressed in fruit tissues and reaches its highest expression at Br3 (Supplemental figure 19B), similar to the expression pattern of major flavonoid biosynthetic genes and metabolites. Subcellular localization further indicated that SlERF.G3-like is

located in the nucleus and the cytoplasm (Supplemental figure 10).

To investigate the function of *SlERF.G3-like*, we generated transgenic MicroTom plants overexpressing *SlERF.G3-like* from the fruit-specific *E8* promoter. We obtained 14 independent *E8: SlERF.G3-like* overexpression lines, of which two representative lines (lines 12 and 36) were selected for further study (Figure 6B, 6C and Supplemental figure 20). We then performed RT-qPCR analysis of *SlERF.G3-Like-12*, *SlERF.G3-Like-36*, and WT pericarp samples at the Br3 stage and found that the expression of flavonoid biosynthesis genes, including *SlCHS1*, *SlCHS2*, *SlCHI*, *SlF3H*, *SlF3'H*, and *SlFLS*, was significantly increased in the two transgenic lines (Figure 6B and 6D). In line with this increased expression of biosynthetic genes, LC-MS analysis of Br7 stage fruits showed a significant increase in the contents of major flavonoid compounds and intermediates (naringenin, eriodicytol, kaempferol-3-*O*-glucoside, kaempferol-3-*O*-rutinoside, quecertin-3-*O*-glucoside and quecertin-3-*O*-rutinoside) (Figure 6D). As both *E8: SlERF.G3-like* line 12 and 36 showed orange color (Figure 6B), we also checked the expression of carotenoid biosynthetic genes. Compared to MicroTom, there is no significant reduction of carotenoid biosynthesis genes in *E8: SlERF.G3-like*-36 fruit (Supplemental figure 21). This matches our previous observation on flavonoid-enriched tomato (Ying et al., 2020; Zhang et al., 2015a). All these data indicate instead of the inhibition of carotenoid biosynthesis, the orange color of *E8: SlERF.G3-like* fruit is mainly due to the accumulation of flavonoid compounds.

To directly compare the functions of SlERF.G3-like with known regulator of flavonoid biosynthesis, SlMYB12, we analyzed the promoters of major flavonol biosynthesis genes (*SlCHS1*, *SlF3H* and *SlFLS*) by dual luciferase reporter assays in *Arabidopsis* protoplasts. The results showed that all three promoters can be significantly induced by SlMYB12 and SlERF.G3-like (Supplemental figure 22). In addition, when both TFs were combined, the activities of tested promoters can be further increased (Supplemental figure 22). However, since the expression levels of

SlMYB12 displays no significant change in *E8:SlERF.G3-like* fruit compared to WT (Supplemental figure 23A), and the transcript abundance of *SlERF.G3-like* is invariant in *AtMYB12* (*Arabidopsis thaliana* homologous of *SlMYB12*) overexpressing fruits (Supplemental figure 23B), All these data suggest although both TFs can upregulate flavonoid biosynthesis, there is no direct interaction between SlMYB12 and SlERF.G3-like.

In particular, we checked the expression of *SlMYB12* and *SlERF.G3-like* in both peel and flesh of WT pericarp at Br3. Compared to *SlMYB12*, which is mainly expressed in fruit peel (Adato et al., 2009), the expression of *SlERF.G3-like* is significantly induced in the flesh (Supplemental figure 23C-D). On the other hand, low expression correlation of the two genes was observed in MMN, SGN-TEA (Shinozaki et al., 2018) and TomExpress (Zouine et al., 2017) databases (Supplemental figure 24). All these data indicate that the two TFs may act independently to regulate the biosynthesis of flavonols in tomato fruit.

Taken together, these data indicate that SlERF.G3-like is a potential transcription factor regulating flavonoid biosynthesis in tomato. Using the MMN dataset, we can uncover the metabolic regulatory networks that operate during the tomato growth cycle and verify novel regulators of major metabolic pathways.

Discussion

We present here a comprehensive analysis of the metabolome and transcriptome of 20 samples covering the major tissues and growth stages of tomato. The MMN dataset generated for this study includes data for 540 metabolites and 31,256 expressed genes that were detected in at least one tissue or stage. Based on our analysis of the global patterns of metabolite and transcript abundance among the tissues and stages, we established that the 540 metabolites can be divided into ten clusters and that 17,003 genes were co-expressed ($r \geq 0.8$) with at least one of the 540 metabolites, accounting for 54.4% of total predicted protein-encoding genes in tomato genome. In addition,

the expression pattern of these 17,003 genes was found to represent the global gene expression pattern (Figure 2D, 2F and Supplemental figure 2).

Because of the detection limit of the broadly targeted LC-MS/MS we used in this study, there are still compounds, such as carotenoids and terpenoids, that have not been quantified in our samples. Therefore, in addition to the ten metabolite clusters we identified, future metabolome analysis of the same samples may reveal novel accumulation patterns for compounds and metabolic pathways that were not included in this study. Once we have this information, the number of genes co-expressed with metabolites is likely to increase further, providing a rich resource for thorough investigation of the genetic networks regulating metabolic pathways in tomato.

Our MMN dataset is also a good addition to current tomato multi-omics resources. Previous studies have focused largely on comparing specific tissues among different individuals, which provides high-resolution correlations between genetic variation and metabolic changes (Alseekh et al., 2015; Tieman et al., 2017; Zhu et al., 2018). However, if a compound or a pathway is not expressed in the sampled tissues, it is difficult to analyze its regulatory network. By covering most tissues and developmental stages of tomato, the MMN dataset provides a high-resolution map of metabolic and transcriptional dynamics throughout the tomato growth cycle. The strategy we used to analyze the MicroTom growth cycle can be used to guide the design of future experiments comparing different species and accessions. Using the MMN dataset, we constructed a global map of metabolic changes during tomato growth, including the following: significant metabolic changes between the vegetative phase and the reproductive phase (Figure 2A); increases in nutritional compounds, such as flavonoids, during fruit ripening to high levels at the ripe stage (Cluster X, Figure 3 and Table 1); and high levels of anti-nutritional compounds, such as SGAs, in vegetative tissues and unripe fruit, which decline significantly during fruit ripening (Cluster V, Figure 3, Supplemental figure 4 and Supplemental figure 6). The transcriptional changes underlying these and other significant metabolic changes can

be captured by MMN and provide a valuable resource for comprehensive investigation of metabolic regulation.

Using the MMN dataset, we were able to verify some previously known metabolic regulatory networks and to identify novel transcription factors that may regulate metabolism in tomato. Previous studies had identified *SlMYB12* and *SlMYB75* as regulators of the flavonoid pathway. Expression of maize (*Zea mays* L.) *Leaf colour* (*LC*) and *Colourless1* (*C1*) genes, or *AtMYB12* (a homolog of *SlMYB12*) in tomato fruit prevented induction of *SlF3'5'H*, resulting in fruit that accumulated higher levels of flavonols but no anthocyanins (Bovy et al., 2002; Luo et al., 2008). However, when *SlMYB75* or *AmDel/AmRos1* was expressed in tomato, *SlF3'5'H* expression was activated and anthocyanins were produced in the fruit (Butelli et al., 2008; Jian et al., 2019). In our study, we verified that *SlMYB75* is mainly present in leaves and controls production of 3-OH flavonoids, which are reported to contribute to UV protection and pathogen resistance (Zhang et al., 2015b).

The MMN dataset also provided an unbiased resource to screen for other regulators of the flavonoid pathway. This led to the identification of *SlERF.G3-like*, a novel ERF transcription factor. We found overexpression of *SlERF.G3-like* can activate the expression of major flavonoid biosynthesis genes (such as *SlCHS1*, *SlF3H* and *SlFLS*) to enhance the production of flavonoids (Figure 6 and Supplemental figure 22). In addition, it seems like SlERF.G3-like acts independently to known flavonoid regulator SlMYB12 (Supplemental figure 22 and 23). All these data indicate SlERF.G3-like is a new regulator of flavonoid biosynthesis in tomato fruit.

The MMN dataset has also proven to be a useful tool for investigating the regulatory networks for biosynthesis of defensive compounds, such as SGAs. The enzyme GLYCOALKALOID METABOLISM1 (GAME1) catalyzes glycosylation of steroidal alkaloids and regulates their toxicity (Itkin et al. 2011). Recent studies in tomato found that Jasmonate-Responsive ERF transcription factors (such as JRE4) or

SlGAME9 are involved in the regulation of SGA levels (Cárdenas et al., 2016; Nakayasu et al., 2018). In MMN, the accumulation of metabolites and the transcription factors (TFs) regulating their biosynthesis perfectly match their biological functions through tomato growth cycle (Figure 3, Supplemental figure 4). Screening of the MMN dataset allowed us to identify SlbHLH114. Similar to SlGAME9, SlbHLH114 can significantly induce the expression of SGA biosynthesis genes and overexpression of *SlbHLH114* in tomato fruit can enhance the production of major SGAs (Figure 5). All these data indicate bHLH114 is a novel transcription factor involved in regulating the SGA pathway

In summary, our MMN dataset provides a high-resolution spatio-temporal representation of the metabolome and transcriptome of 20 different tissues/stages of the tomato growth cycle. It presents a global view of the major metabolic changes during the tomato growth cycle and the transcriptional regulation that underlies these changes. Using this dataset, we will able to verify other important patterns of metabolic regulation and identify novel transcription factors controlling these pathways. Taken together, the MMN dataset provides insights into the transcriptional control of major tomato metabolites and gives guidance for future quality improvement.

Methods

Plant materials and tissue preparation

Tomato (*Solanum lycopersicum cv* MicroTom) seeds were obtained from PanAmerican Seed™. MicroTom plants were grown in a climate-controlled growth room with a light/dark photoperiod of 16/8 h at 23 □. Four to five-week-old plants with five to six expanded true leaves and one bud were used for bud stage (30 DPG, days post germination). Six to seven-week-old plants with more than 50 % flowers were used for flowering stage (45 DPG). Twelve to thirteen-week-old plants with one breaker fruit were used for breaker stage (85 DPG). All these three stages were used to collect root, stem and leaf tissues (the leaf tissue is collected from the fifth true leaf to all healthy young leaves). Flowers at anthesis (0 DPA) were tagged and expanding fruit were harvested at 10DPA, 20DPA, 30DPA (Immature green). Mature and ripening fruits were harvested based on the tomato color chart "USDA Visual Aid TM-L-1"(USDA Agricultural Marketing Service, 1975) as follows: MG (Mature green) stage (full-size green fruit, approximately 35 DPA), Br (Breaker) stage (approximately 40 DPA, definite break in color from green to tannish-yellow with less than 10 % of the surface), and 3, 7, 10, 15 day later (Br3, Br7, Br10, and Br15, respectively)). All samplings were collected before the greenhouse light off (around 7:00–9:00 p.m.). Samples from 10 individual plants were pooled together as one biological replicate and immediately frozen in liquid nitrogen. Three biological replicates were applied to later transcriptome and metabolome analysis.

Generation of MicroTom Metabolic Network (MMN) dataset: Metabolome Profiling and Transcriptome Profiling

Metabolome profiling was carried out using a widely targeted metabolome method by Wuhan Metware Biotechnology Co., Ltd. (Wuhan, China) (http://www.metware.cn/). Briefly, the tomato tissues were lyophilized and ground into fine powder using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 *min* at 30 *Hz*. 100 *mg* tissue

powder was weighed and extracted overnight with 1.0 *mL* 70% aqueous methanol at 4℃, followed by the centrifugation for 10 *min* at 10,000 *g*, all the supernatants were collected and filtered with a membrane (SCAA-104, 0.22 mm pore size; ANPEL, Shanghai, China, http://www.anpel.com.cn/) before LC-MS analysis. Quantification of metabolites was carried out using a scheduled multiple reaction monitoring method (Wei et al., 2013; Zhu et al., 2018).

Transcriptome profiling was performed as described previously (Ying et al., 2020). Briefly, the mapped clean reads were calculated using Hiseq-X-ten sequencing platform, and then mapped to tomato reference genome (Version 3.0) (The Tomato Genome Consortium, 2012) using Hisat 2 (Daehwan et al., 2015), and then normalized to reads TPM by StringTie (Pertea et al., 2015). All raw data were deposited into the Genome Sequence Archive in Big Data Center, Beijing Institute of Genomics, Chinese Academy of Science, under accession number CRA001723 and CRA001712 that were publicly accessible at http://bigd.big.ac.cn/gsa (Members, 2018; Wang et al., 2017).

Co-expression/co-regulation cluster identification and Analysis

Co-expression/ co-regulation analysis was done for 20 different time points/tissues samples by the MeV (Version 4.9) with the *k*-means method (Gasch and Eisen, 2002). The normalized expression values of genes and metabolites were calculated by dividing their expression level at different time points/tissues. Hierarchical clustering (HCL) and principal component analysis (PCA) was performed using the prcomp function in R software (Team, 2013) with default settings to facilitate graphical interpretation of relatedness among 20 different time points/tissues samples. The transformed and normalized gene and metabolites expression values with z-scores were used for HCL and PCA.

Network building for metabolome and transcriptome of metabolic pathway

We used the Pearson's correlation algorithm method (Bishara and Hittner, 2012) to construct a transcription factor-related gene and metabolites regulatory network. Mutual information for calculating the expression similarity between the expression levels of transcription factor and gene, metabolites pairs were calculated by R software. All the associations among transcription factor and gene, metabolites were shown by Cytoscape software (Kohl et al., 2011).

Sequence alignment and phylogenetic analyses

Sequence alignments were performed using the MUSCLE algorithm in MEGA7 (Edgar, 2004). Molecular phylogenetic analysis was done on the basis of the NJ matrix-based model. Bootstrap values were calculated using 1,000 replicates. All phylogenetic analyses were conducted in MEGA7.

Sample extraction and metabolomic analysis

Preparation of extracts and the profiling of flavonoids in tomato fruit tissues were performed with same methods as described previously (Ying et al., 2020). For measurement of steroidal glycoalkaloids, tomatoes were harvested at 10 days after breaker (Br10). Fruit pericarp was freeze-dried and ground into fine powder using an automatic sample rapid grinder (JXFSTPRP-24, Shanghai jingxin industrial development CO., LTD) with 3 times for 1 *min* at 50 *Hz*. Extraction was performed as previously described (Itkin et al., 2011). Briefly, 100 *mg* freeze-dried tissue was extracted with 1 *mL* 80 *%* methanol/water (*v*/*v*) containing 0.1 *%* formic acid. The mixture was vortexed for 30 *s*, sonicated for 30 *min* at 4 □, vortexed again for 30 *s*, centrifuged (20,000 *g*, 10 *min*, 4 □), and filtered through a 0.22-mm polytetrafluoroethylene membrane filter. The profiling of steroidal glycoalkaloids of tomato tissues was performed by LC-MS analysis using the SCIEX Triple Quad™ 5500 LC-MS/MS System with the UPLC column connected online to a photo diode array detector (Shimadzu), separation of metabolites and detection of the eluted compound masses was performed as described (Zhu et al., 2018). All samples were

performed in biological triplicate.

RNA extraction and RT-qPCR

Both MicroTom and transgenic fruits were harvested at Br3 and fruit pericarp was ground into fine powder using liquid nitrogen. RNA extraction and cDNA synthesis were done as described previously. RT-qPCR was done by BIO-RAD CFX384 Real-Time System, following the manufacturer's instruction. With *SlUBI* as an internal control, the relative expression of each genes was calculated by the ΔCt method. The primer pairs for RT-qPCR were blasted at NCBI database to ensure primer specific (see Supplemental Data 6).

Vector construction and generation of transgenic lines

The overexpression constructs used for *Agrobacterium*-mediated tomato plant transformation were built using the Goldenbraid cloning (Sarrion-Perdigones et al., 2013) and Gateway Cloning Technology (Curtis and Grossniklaus, 2003). For fruit specific expression, full length of coding sequence (CDS) was introduced into pBin18-E8-GW through Gateway Cloning (Ying et al., 2020). We used the Goldenbraid system for generating 35S promoter- transcription factor -35S terminator as a one transcriptional unit, kanamycin resistance gene under the control of NOS promoter and NOS terminator were additional genes. The plasmid with the correct insertion was introduced into *Agrobacterium tumefaciens* strain EHA105 and tomato transformation was done as described previously (Ying et al., 2020).

Subcellular localization

Tobacco (*Nicotiana benthamiana*) used in this study was both grown in the greenhouse with a light/dark photoperiod of 16/8 h at 25 □. Determination of the subcellular localization of individual transcription factors (TF) fused to the GFP fluorescent reporter is performed in *Nicotiana benthamiana* leaves protoplasts as

described earlier (Deng et al., 2018). Briefly, the released protoplast cells from leaves of 3-4 week-old tobacco were isolated and transformed via the PEG-mediated protoplast transfection method (Deng et al., 2018). Full-length cDNA is fused with GFP in the pBI101.3, the pBI101.3-SlTF and pBI101.3 were individually transient transformation into protoplasts of *N. benthamiana* leaves. After one night grown, DAPI was used to stain the nucleic acids of the protoplasts. Signals from GFP and DAPI were visualized with a laser scanning confocal microscopy (Zeiss CELL Observer SD).

Transient dual luciferase reporter assay

Promoter sequences were amplified from genomic DNA using PCR and inserted upstream of the *Luciferase* (*LUC*) CDS by Goldenbraid 2.0 cloning strategy to yield the promoter-LUC reporter vectors. For an internal control, the expression of the *Renlia* (*REN*) gene was driven by the CaMV35S promoter in a reporter vector. The CDS of *SlbHLH114* was cloned into the pUPD2 vector using the Goldenbraid 2.0 cloning strategy, to create an overexpression construct named *35S-SlbHLH114*. The empty vector P35S-T35S was used as the negative control (CK).

Arabidopsis thaliana (Columbia-0, Col-0) used in this study was grown in the greenhouse with a light/dark photoperiod of 16/8 h at 22 ℃. Dual luciferase reporter assay was performed as described previously (Deng et al., 2018). Briefly, protoplasts used for transfection were isolated from 4-5 week-old *Arabidopsis thaliana* leaves. Protoplast co-transfection assays were performed using the reporter plasmids and the internal control vectors. Results were analyzed and quantified by flow cytometry 16 h following protoplast transfection. Luciferase activity was detected using the dual-luciferase reporter assay system (Promega) with a Synergy™ H1 hybrid multimode microplate reader (BioTek). Expression was expressed as the ratio of LUC to REN activity.

Statistics

For comparison of individual treatments with their relevant controls, unpaired two-tailed Student's t-tests were used, and $P \leq 0.05$ was considered significant. To compare measurements of multiple treatments with each other, we performed univariate ANOVA followed by the posthoc Tukey's test of multiple pairwise comparisons to determine group differences using GraphPad Prism version 8 (Swift, 1997).

Accession numbers

Tomato genome sequence data for this article was downloaded from the SGN (https://solgenomics.net/). The raw values obtained from the metabolomic data sets were available in Supplemental data 1 (associated figures: Figure 2A, 2C, 2E and 3, Supplemental figure 1A, and Table 1). RNA-seq data sets for the transcriptome analysis were available at the Genome Sequence Archive in Big Data Center, under accession number CRA001723 and CRA001712 that are publicly accessible at http://bigd.big.ac.cn/gsa (Wang et al., 2017) (associated figures: Figure 2B, 2D, 2F, 4A and 4B, Supplemental figure 1B, 2, 3, 4, 5, 6 and 11). Sequence data of phylogenetic analysis can be found in Supplemental Table 2. If any data sets are unavailable through the links stated above, they can be obtained from the corresponding author on request.

Funding

This study was funded by grants from the National Natural Science Foundation of China (31701255, 31772372, 31772372), YZ was supported by the Fundamental Research Funds for the Central Universities (2017SCU04A11 and SCU2019D013). MCL acknowledges the support from National Key R&D Program of China (2016YFD0400100). SA and ARF acknowledges the support from the PlantaSYST project by the European Unions Horizon 2020 research and innovation programme (SGA-CSA No 664621 and No 739582 under FPA No. 664620).

Authors' contributions

YZ, MCL and TM conceived and designed the study. YL, LZ and YC grew MicroTom tomato plants and collected plant materials. YL, YC and TM performed the analysis of the RNA-seq data. SA and ARF analyzed the metabolome data. YL, LZ, SJY, HD and DS carried out validation functional assays with the help of YY, RF and ZXZ. YL, YZ, MCL, ARF, and MB wrote the manuscript. All authors have discussed the results and approved the manuscript.

Acknowledgements

We would like to thank the Mass Spectrometry Core Facility in College of Life Science, Sichuan University for assistance in metabolic analysis. And the Bioinformatics Center in College of Life Science, Sichuan University for assistance in -omics and statistics analysis.

References

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., Domı´nguez, E., Wang, Z., Vos, R.C.H.D., Jetter, R., et al. (2009). Fruit-Surface Flavonoid Accumulation in Tomato Is Controlled by a SlMYB12-Regulated Transcriptional Network. Plos Genetics 5:1-23.

Alba, R., Payton, P., Fei, Z., McQuinn, R., Debbie, P., Martin, G.B., Tanksley, S.D., and Giovannoni, J.J. (2005). Transcriptome and Selected Metabolite Analyses Reveal Multiple Points of Ethylene Control during Tomato Fruit Development. The Plant Cell 17:2954-2965.

Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., Kleessen, S., Giavalisco, P., Pleban, T., and Muellerroeber, B. (2015). Identification and Mode of Inheritance of Quantitative Trait Loci for Secondary Metabolite Abundance in Tomato. Plant Cell 27:485-512.

Alseekh, S., Tong, H., Scossa, F., Brotman, Y., Vigroux, F., Tohge, T., Ofner, I., Zamir, D., Nikoloski, Z., and Fernie, A.R. (2017). Canalization of Tomato Fruit Metabolism. Plant Cell 29:2753-2765.

Ana-Rosa Ballester, J.M., Ric de Vos, Bas te Lintel Hekkert, Diego Orzaez, Josefina-Patricia Fernández-Moreno, Pasquale Tripodi, Silvana Grandillo, Cathie Martin, Jos Heldens, Marieke Ykema, Antonio Granell, Arnaud Bovy. (2010). Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. Plant Physiology 152:71-84.

Azzi, L., Deluche, C., Gévaudant, F., Frangne, N., Delmas, F., Hernould, M., and Chevalier, C. (2015). Fruit growth-related genes in tomato. Journal of Experimental Botany 66:1075-1086.

Balcke, G.U., Bennewitz, S., Bergau, N., Athmer, B., Henning, A., Majovsky, P., Jiménez-Gómez, J.M., Hoehenwarter, W., and Tissier, A. (2017). Multi-Omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. The Plant cell 29:960-983.

Ballester, A.-R., Molthoff, J., de Vos, R., Hekkert, B.t.L., Orzaez, D., Fernández-Moreno, J.-P., Tripodi, P., Grandillo, S., Martin, C., Heldens, J., et al. (2010). Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. Plant Physiology 152:71-84.

Bishara, A.J., and Hittner, J.B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. Psychological Methods 17:399-417.

Bo, L., Victor, R., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26:493-500.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid

Biosynthesis. The Plant Cell 12:2383-2393.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., Collins, G., Robinson, S., Verhoeyen, M., Hughes, S., et al. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. The Plant cell 14:2509-2526.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnology 26:1301-1308.

Cárdenas, P.D., Sonawane, P.D., Heinig, U., Jozwiak, A., Panda, S., Abebie, B., Kazachkova, Y., Pliner, M., Unger, T., Wolf, D., et al. (2019). Pathways to defense metabolites and evading fruit bitterness in genus Solanum evolved through 2-oxoglutarate-dependent dioxygenases. Nature Communications 10:1-13.

Cárdenas, P.D., Sonawane, P.D., Jacob, P., Robin, V.B., Veena, D., Efrat, W., Lior, T., Sagit, M., Ilana, R., and Sergey, M. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. Nature Communications 7:1-16.

Carqueijeiro, I.T., Brown, S., Chung, K., Dang, T.T., Walia, M., Besseau, S., Dugé, d.B.T., Oudin, A., Lanoue, A., and Billet, K. (2018). Two tabersonine 6,7-epoxidases start synthesis of lochnericine-type alkaloids in Catharanthus roseus. Plant Physiology.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M.-I., Nunes-Nesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M., et al. (2006). Integrated Analysis of Metabolite and Transcript Levels Reveals the Metabolic Shifts That Underlie Tomato Fruit Development and Highlight Regulatory Aspects of Metabolic Network Behavior. Plant Physiology 142:1380-1396.

Chen, S.-P., Kuo, C.-H., Lu, H.-H., Lo, H.-S., and Yeh, K.-W. (2016). The Sweet Potato NAC-Domain Transcription Factor IbNAC1 Is Dynamically Coordinated by the Activator IbbHLH3 and the Repressor IbbHLH4 to Reprogram the Defense Mechanism against Wounding. PLOS Genetics 12:1-26.

Curtis, M.D., and Grossniklaus, U. (2003). A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. Plant Physiology 133:462.

Daehwan, K., Ben, L., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12:357-360.

Dan, L., Liang, Y., Jin-zhe, Z., Guang-tao, Z., Hong-jun, L., Ya-qing, L., Yan-ling, W., Xue, C., Tian-shu, S., and San-wen, H. (2019). Domestication and breeding changed tomato fruit transcriptome. Journal of Integrative Agriculture 19:120-132.

Deng, H., Pirrello, J., Chen, Y., Li, N., Zhu, S., Chirinos, X., Bouzayen, M., Liu, Y., and Liu, M. (2018). A novel tomato F-box protein, SlEBF3, is involved in tuning ethylene signaling during plant development and climacteric fruit

ripening. The Plant Journal 95:648-658.

Deng, X., Guo, D., Yang, S., Shi, M., and Chao, J. (2015). Jasmonate signalling in the regulation of rubber biosynthesis in laticifer cells of rubber tree, Hevea brasiliensis. Frontiers in Plant Science 6:1-13.

Du, M., Zhao, J., Tzeng, D.T.W., Liu, Y., and Li, C. (2017). MYC2 Orchestrates a Hierarchical Transcriptional Cascade That Regulates Jasmonate-Mediated Plant Immunity in Tomato. Plant Cell 29:tpc.00953.02016.

Du, T., Niu, J., Su, J., Li, S., Guo, X., Li, L., Cao, X., and Kang, J. (2018). SmbHLH37 Functions Antagonistically With SmMYC2 in Regulating Jasmonate-Mediated Biosynthesis of Phenolic Acids in Salvia miltiorrhiza. Frontiers in Plant Science 9:1-13.

Echeverry, C., Arredondo, F., Martínez, M., Abin-Carriquiry, J.A., Midiwo, J., and Dajas, F. (2015). Antioxidant Activity, Cellular Bioavailability, and Iron and Calcium Management of Neuroprotective and Nonneuroprotective Flavones. Neurotoxicity Research 27:31-42.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792-1797.

Eshed Y, Z.D. (1995). An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147-1162.

Espina, V., Wulfkuhle, J.D., Calvert, V.S., VanMeter, A., Zhou, W., Coukos, G., Geho, D.H., Petricoin, E.F., and Liotta, L.A. (2006). Laser-capture microdissection. Nature Protocols 1:586-603.

Garbowicz, K., Liu, Z., Alseekh, S., Tieman, D., Taylor, M., Kuhalskaya, A., Ofner, I., Zamir, D., Klee, H.J., and Fernie, A.R. (2018). Quantitative trait loci analysis identifies a prominent gene involved in the production of fatty-acid-derived flavor volatiles in tomato. Molecular Plant 11:1147-1165.

Gasch, A.P., and Eisen, M.B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. Genome Biology 3:research0059.0051.

Giovannoni, J. (2018). Tomato Multiomics Reveals Consequences of Crop Domestication and Improvement. Cell 172:6-8.

Grotewold, E. (2016). Flavonols drive plant microevolution. Nature Genetics 48:112.

Howe, E., Holton, K., Nair, S., Schlauch, D., Sinha, R., and Quackenbush, J. (2010). MeV: MultiExperiment Viewer. In: Biomedical Informatics for Cancer Research--Ochs, M.F., Casagrande, J.T., and Davuluri, R.V., eds. Boston, MA: Springer US. 267-277.

Hua, S., Hua-Jie, F., and Hong-Qing, L. (2015). Genome-wide identification and characterization of the bHLH gene family in tomato. BMC Genomics 16:1-12.

Hua, W., Nicolas, S., Bjoern, U., Pierre, F., Mohamed, Z., Michel, H., Alain, L., Jean-Claude, P., Fernie, A.R., and Mondher, B. (2009). Regulatory features underlying pollination-dependent and -independent tomato fruit set revealed by transcript and primary metabolite profiling. Plant Cell 21:1428-1452.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza,

S.E., Unger, T., Malitsky, S., and Finkers, R. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341:175-179.

Itkin, M., Rogachev, I., Alkan, N., Rosenberg, T., Malitsky, S., Masini, L., Meir, S., Iijima, Y., Aoki, K., de Vos, R., et al. (2011). GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. Plant Cell 23:4507-4525.

Jennifer, L. (2013). A quantitative genetic basis for leaf morphology is revealed in a set of precisely defined tomato introgression lines. Plant Cell 25:2465-2481.

Jian, W., Cao, H., Yuan, S., Liu, Y., Lu, J., Lu, W., Li, N., Wang, J., Zou, J., Tang, N., et al. (2019). SlMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits. Horticulture Research 6:1-15.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B., and Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. The EMBO Journal 19:6150-6161.

Kaakoush, N.O., and Morris, M.J. (2017). More Flavor for Flavonoid-Based Interventions? Trends in Molecular Medicine 23:293-295.

Kemparaju, C.C. (2018). Identification and Characterization of Key Regulatory Components Involved in the Development of Type VI Glandular Trichomes in Solanum Lycopersicum: Martin-Luther-Universität Halle-Wittenberg.

Klee, H.J., and Giovannoni, J.J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. Annual Review of Genetics 45:41-59.

Kohl, M., Wiese, S., and Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. Methods in Molecular Biology 696:291-303.

Lewinsohn, E., Schalechet, F., Wilkinson, J., Matsui, K., Tadmor, Y., Nam, K.-H., Amar, O., Lastochkin, E., Larkov, O., Ravid, U., et al. (2001). Enhanced Levels of the Aroma and Flavor Compound S-Linalool by Metabolic Engineering of the Terpenoid Pathway in Tomato Fruits. Plant Physiology 127:1256-1265.

Liu, L., Shao, Z., Zhang, M., and Wang, Q. (2015). Regulation of carotenoid metabolism in tomato. Molecular plant 8:28-39.

Liu, M., Gomes, B.L., Mila, I., Purgatto, E., Peres, L.E.P., Frasse, P., Maza, E., Zouine, M., Roustan, J.-P., Bouzayen, M., et al. (2016). Comprehensive Profiling of Ethylene Response Factor Expression Identifies Ripening-Associated ERF Genes and Their Link to Key Regulators of Fruit Ripening in Tomato. Plant Physiology 170:1732.

Liu, Y., Du, M., Deng, L., Shen, J., Fang, M., Chen, Q., Lu, Y., Wang, Q., Li, C., and Zhai, Q. (2019). MYC2 Regulates the Termination of Jasmonate Signaling via an Autoregulatory Negative Feedback Loop. The Plant Cell 31:106-127.

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. Current Opinion in Plant Biology 24:31-38.

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., Weisshaar, B., and Martin, C. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant Journal 56:316-326.

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis Transcription Factor MYB12 Is a Flavonol-Specific Regulator of Phenylpropanoid Biosynthesis. Plant Physiology 138:1083-1096.

Members, B.D.C. (2018). Database Resources of the BIG Data Center in 2018. Nucleic Acids Research 46:D14-D20.

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldberg, L., Lotan, O., Yativ, M., Wang, Z., Jetter, R., Venger, I., Adato, A., et al. (2008). Gene Expression and Metabolism in Tomato Fruit Surface Tissues. Plant Physiology 147:823-851.

Nakayasu, M., Shioya, N., Shikata, M., Thagun, C., and Shoji, T. (2018). JRE4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. The Plant Journal 94:975-990.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33:290-297.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2013). GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. Plant Physiology 162:1618-1631.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C., Kopka, J., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology 24:447-454.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M., and Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. Frontiers in Plant Science 7.

Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L., and Kuhlemeier, C. (2015). MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. Nature Genetics 48:159.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nature Biotechnology 35:441.

Shinozaki, Y., Nicolas, P., Fernandezpozo, N., Ma, Q., Evanich, D.J., Shi, Y., Xu, Y., Zheng, Y., Snyder, S.I., and Martin, L.B.B. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nature Communications 9:1-13.

Silvia, G., Andrea, M., and Pierdomenico, P. (2009). Purple as a tomato: towards high anthocyanin tomatoes. Trends in Plant Science 14:237-241.

Sonawane, P.D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., Unger,

T., Malitsky, S., Arendt, P., and Pauwels, L. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. Nature Plants 3:16205.

Swift, M.L. (1997). GraphPad Prism, Data Analysis, and Scientific Graphing. Journal of Chemical Information & Modeling 37:411-412.

Team, R.C. (2013). R: A language and environment for statistical computing.

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641.

Tieman, D., Zhu, G., Jr, R.M., Lin, T., Nguyen, C., Bies, D., Rambla, J.L., Beltran, K.S., Taylor, M., and Zhang, B. (2017). A chemical genetic roadmap to improved tomato flavor. Science 355:391-394.

Tohge, T., Scossa, F., Wendenburg, R., Frasse, P., Balbo, I., Watanabe, M., Alseekh, S., Jadhav, S.S., Delfin, J.C., Lohse, M., et al. (2020). Exploiting the natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the lycopersicum complex. Molecular Plant.

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., et al. (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. Nature Communications 7:12399.

Tominaga-Wada, R., Iwata, M., Nukumizu, Y., and Wada, T. (2011). Analysis of IIId, IIIe and IVa group basic-helix-loop-helix proteins expressed in Arabidopsis root epidermis. Plant Science 181:471-478.

Uluisik, S., Chapman, N.H., Smith, R., Poole, M., Adams, G., Gillis, R.B., Besong, T.M.D., Sheldon, J., Stiegelmeyer, S., Perez, L., et al. (2016). Genetic improvement of tomato by targeted control of fruit softening. Nature Biotechnology 34:950.

USDA. (2017). National Agricultural Statistics Service - Vegetables 2016 Summary.

Vogt, T. (2010). Phenylpropanoid Biosynthesis. Molecular Plant 3:2-20.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Bioences 131:281-285.

Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., Tang, B., Dong, L., Ding, N., Zhang, Q., et al. (2017). GSA: Genome Sequence Archive*. Genomics, Proteomics & Bioinformatics 15:14-18.

Wei, C., Liang, G., Guo, Z., Wang, W., Zhang, H., Liu, X., Yu, S., Xiong, L., and Jie, L. (2013). A Novel Integrated Method for Large-Scale Detection, Identification, and Quantification of Widely Targeted Metabolites: Application in the Study of Rice Metabolomics. Molecular Plant 6:1769-1780.

Xu, H., Lybrand, D., Bennewitz, S., Tissier, A., Last, R.L., and Pichersky, E. (2018). Production of trans-chrysanthemic acid, the monoterpene acid moiety of natural pyrethrin insecticides, in tomato fruit. Metabolic Engineering 47:271-278.

Ying, S., Su, M., Wu, Y., Zhou, L., Fu, R., Li, Y., Guo, H., Luo, J., Wang, S., and Zhang, Y. (2020). Trichome regulator SlMIXTA-like directly manipulates

primary metabolism in tomato fruit. Plant Biotechnology Journal 18:354-363.

Zhang, J., Tang, X., Alba, R., and Giovannoni, J. (2006). Tomato Expression Database (TED): a suite of data presentation and analysis tools. Nucleic Acids Research 34:D766-D770.

Zhang, J., Zhao, J., Liang, Y., and Zou, Z. (2016). Genome-wide association-mapping for fruit quality traits in tomato. Euphytica 207:439-451.

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., Kawar, P.G., Hill, L., Santino, A., Fernie, A.R., et al. (2015a). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. Nature Communications 6:219-246.

Zhang, Y., De Stefano, R., Robine, M., Butelli, E., Bulling, K., Hill, L., Rejzek, M., Martin, C., and Schoonbeek, H.-j. (2015b). Different reactive oxygen species scavenging properties of flavonoids determine their abilities to extend the shelf life of tomato. Plant physiology 169:1568-1583.

Zhao, J., Sauvage, C., Zhao, J., Bitton, F., Bauchet, G., Liu, D., Huang, S., Tieman, D.M., Klee, H.J., and Causse, M. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. Nature Communications 10:1-12.

Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nature Biotechnology 31:154.

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., and Yang, C. (2018). Rewiring of the Fruit Metabolome in Tomato Breeding. Cell 172:249-261.

Zouine, M., Maza, E., Djari, A., Lauvernier, M., Frasse, P., Smouni, A., Pirrello, J., and Bouzayen, M. (2017). TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. The Plant Journal 92:727-735.

Figure and Table Legends

Figure 1. Schematic representation of the design for MicroTom Metabolic Network (MMN). 20 samples of 3 key development stages were collected for metabolic profiling and RNA-seq. Axis indicates sample harvest date (days post germination, DPG). Leaf (L), root (R), stem (S), bud (F30) and flower (F45) samples were harvest at bud stage (30 DPG), flowering stage (45 DPG) and breaker stage (85 DPG), respectively. Fruit samples were harvested at 10 days post anthesis (10DPA), 20DPA, Immature Green (IMG), Mature Green (MG), Breaker (Br), Breaker plus 3 Days (Br3), Br7, Br10, Br15.

Figure 2. Summary of metabolome and transcriptome data of MMN. Overview of 540 annotated metabolites (A) and hierarchical clustering analysis of gene expression profiles with 31,256 genes (B) from 20 tomato samples. Principal component analysis (PCA) (C and D) and Cluster dendrogram (E and F) of metabolome (C and E) and transcriptome (D and F) in the 20 MicroTom samples. For metabolome data (A), the metabolite per row is Z-score standardized to -3 to 3. For transcriptome data (B), the color scale 0-1 represents Spearman's correlation coefficients. Axis numbers in (A) indicate tissues: 1-R30, 2-R45, 3-R85, 4-S30, 5-S45, 6-S85, 7-L30, 8-L45, 9-L85, 10-F30, 11-F45, 12-10DPA, 13-20DPA, 14-IMG, 15-MG, 16-Br, 17-Br3, 18-Br7, 19-Br10, 20-Br15.

Figure 3. Dynamic of metabolite and gene expression during MicroTom growth cycle. *K*-means clustering grouped the expression profile of the tomato metabolome (red) and transcriptome (blue) into 10 clusters. The x axis depicts 20 samples from 3 key development stages and the y axis depicts Z-score standardized per metabolite (red) and gene (blue). The numbers shown in each box (for example, 67 metabolites and 4,543 genes for cluster I) were derived based on the number of metabolites and genes across all 20 samples in each cluster. Axis numbers indicate tissues: 1-R30, 2-R45, 3-R85, 4-S30, 5-S45, 6-S85, 7-L30, 8-L45, 9-L85, 10-F30, 11-F45, 12-10DPA, 13-20DPA, 14-IMG, 15-MG, 16-Br, 17-Br3, 18-Br7, 19-Br10, 20-Br15. Ten clusters were identified.

Figure 4. SlMYB75 is associated with the biosynthesis of flavonoids with 3 - OHs on the B - ring. (A) Network built on correlation among structure genes and TFs. Pearson correlation coefficient (PCC) values were calculated for each pair of genes, the color scale has been normalized to range from -1 to 1, where -1 is negative correlation and 1 corresponds to the positive correlation. (B) Expression pattern of flavonoid biosynthetic genes and TFs in 20 samples. Expression data were Z-score standardized to -3 to 3 per gene. (C) Dual luciferase reporter assay indicated SlMYB75 can better induce the activity of the promoter of *SlF3'5'H* than SlMYB12. Error bars represent the standard deviation (n=3). Different letters indicate significantly different values at $P < 0.05$ (one-way ANOVA, Tukey's posthoc test). (D) Schematic representation of flavonoids biosynthesis and regulation in tomato. Purple arrows indicate genes and metabolites regulated by SlMYB75 and orange arrows indicate genes and metabolites regulated by SlMYB12.

Figure 5. SlbHLH114 is a new transcription factor in steroidal alkaloids pathway. (A) Co-expression network of steroidal alkaloids biosynthetic pathway. Metabolites,

structure genes and transcription factors were marked in pink, light blue and red, respectively. (B) Phenotypes of transgenic and MicroTom tomato fruit at different stages. Br3, three days post breaker; Br10, ten days post breaker. (C) Summarized gene and metabolic changes in tomato fruits expressing SlbHLH114. Based on transcriptome and metabolome data, genes and metabolites which are significantly increased in the transgenic fruits (Line C and Line I) were colored. Data was represented as log_2fold-change compared to MicroTom. (D) Expression level of *SlbHLH114* in transgenic fruit from two independent T1 generation lines and MicroTom at Br3. Error bars represent the standard deviation (n=3). (*P < 0.05; ** P < 0.01; *** P < 0.001; Student's t test). (E) Schematic representation of promoter activation test. The promoters were cloned into the dual-luciferase reporter vector to activate the expression of luciferase (LUC). The renilla (REN) driven by the CaMV 35S promoter served as an internal control. T35S, CaMV 35S terminator. CK, the empty vector. Asterisks indicate significant differences to CK. (F) SlbHLH114 directly interact with the promoter of pathway genes. Promoter sequences can be found in Supplemental Table 1. Different letters indicate significantly different values at $P < 0.05$ (one-way ANOVA, Tukey's posthoc test).

Figure 6. SlERF.G3-like is a new regulator for flavonoid biosynthesis. (A) Co-expression network of key genes and metabolites in flavonoid pathway. Metabolites, structure genes and transcription factors were marked in light pink, blue and gray, respectively. Pearson correlation coefficient (PCC) values were calculated for each pair of genes/metabolites. (B) Phenotypes of transgenic and MicroTom tomato fruit at seven days post breaker. (C) RT-qPCR of transgenic plants in Br3. Error bars represent the standard deviation (n=3). (*P < 0.05; ** P < 0.01; *** P < 0.001; Student's t test). (D) Up-regulated genes and metabolites in *E8: SlERF.G3-Like* overexpression lines.

Table 1. Distribution of the Compounds and Genes Identified in This Study in Different Clusters.

Supplemental Information

Supplemental figure 1. Metabolites accumulation and gene expression pattern over tomato growth cycle.

Supplemental figure 2. Transcriptome relationships of 17,003 genes co-expressed with metabolites among 20 tissues/time points.

Supplemental figure 3 Expression of *SlMYB12* (A), *SlMYB75* (B) and *SlF3'5'H* (C) expression in RNA-seq data (left) and validation by RT-qPCR (right).

Supplemental figure 4. Co-expression of major SGA compounds and pathway genes among different tissues and stages.

Supplemental figure 5. Expression pattern of SGA pathway genes in RNA-seq data.

Supplemental figure 6. *GAME* genes and steroidal alkaloid compounds are co-expressed in cluster V.

Supplemental figure 7. Validation of SGA pathway genes expression through RT-qPCR.

Supplemental figure 8. Coexpression analysis of *SlbHLH114* and *SlGAME9*.

Supplemental figure 9. Molecular Characterization of SlbHLH114.

Supplemental figure 10. Subcellular localization of SlbHLH114 and SlERF.G3-like.

Supplemental figure 11. Screening of pBin19-*E8:SlbHLH114* T0 tomato.

Supplemental figure 12. Summarization of *E8:SlbHLH114* fruit transcriptome data.

Supplemental figure 13. Verification of differently expressed SGA related genes in *SlbHLH114* overexpression lines.

Supplemental figure 14. Coexpression analysis *SlMYC2, SlbHLH114, SlGAME9* with SGAs related and JA-signaling related genes in MMN.

Supplemental figure 15. Expression pattern of *SlMYC2* and *SlbHLH114* in (A) SGN-TEA, (B) Tomexpress and (C) MMN (This study).

Supplemental figure 16. Phylogenetic analysis indicates both SlbHLH114 and SlMYC2/SlbHLH147 belong to the bHLH subfamily 15.

Supplemental figure 17. SlMYC2 may activate the expression of *SlbHLH114*.

Supplemental figure 18. Coexpression analysis of *SlERF.G3-like* and *SlMYB12*.

Supplemental figure 19. Characterization of SlERF.G3-Like.

Supplemental figure 20. Screening of *E8: SlERF.G3-Like* T0 tomato.

Supplemental figure 21. Expression of carotenoid pathway genes in *E8: SlERF.G3-Like* and MicroTom fruit.

Supplemental figure 22. SlERF.G3-like could directly activate the major expression of flavonoid biosynthetic genes.

Supplemental figure 23. *SlMYB12* and *SlERF.G3-Like* have different expression pattern in fruit.

Supplemental figure 24. Expression pattern of *SlERF.G3-like* and *SlMYB12* in (A) SGN-TEA, (B) Tomexpress and (C) MMN (This study).

Supplemental Table 1. Gene and promoter sequence used in this study.

Supplemental Table 2. Protein sequence of bHLH and ERF family used for phylogenetic analysis.

Supplemental Dataset 1. Summary of metabolome profiling for tomato tissues.

Supplemental Dataset 2. Summary of transcriptome mapping for tomato tissues.

Supplemental Dataset 3. Gene expression in tomato tissues.

Supplemental Dataset 4. Summary of coexpression clusters.

Supplemental Dataset 5. Summary of coexpression gene clusters.

Supplemental Dataset 6. Summary of PCR and qRT-PCR primers used in this study.

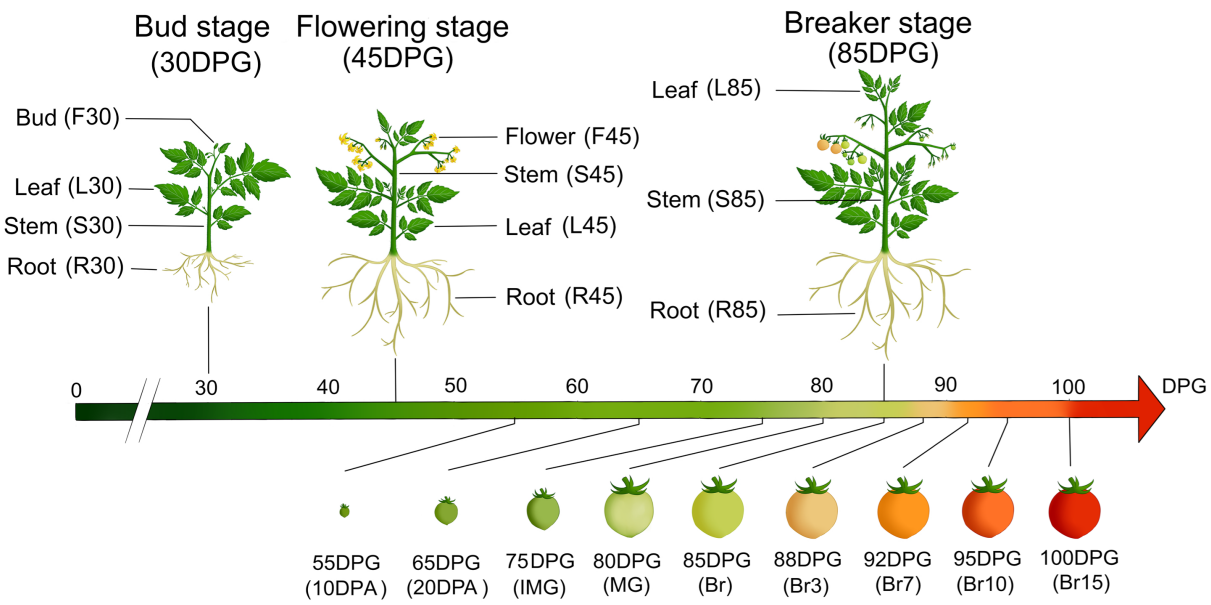
Supplemental Dataset 7. Summary of metabolome profiling for steroidal glycoalkaloids.

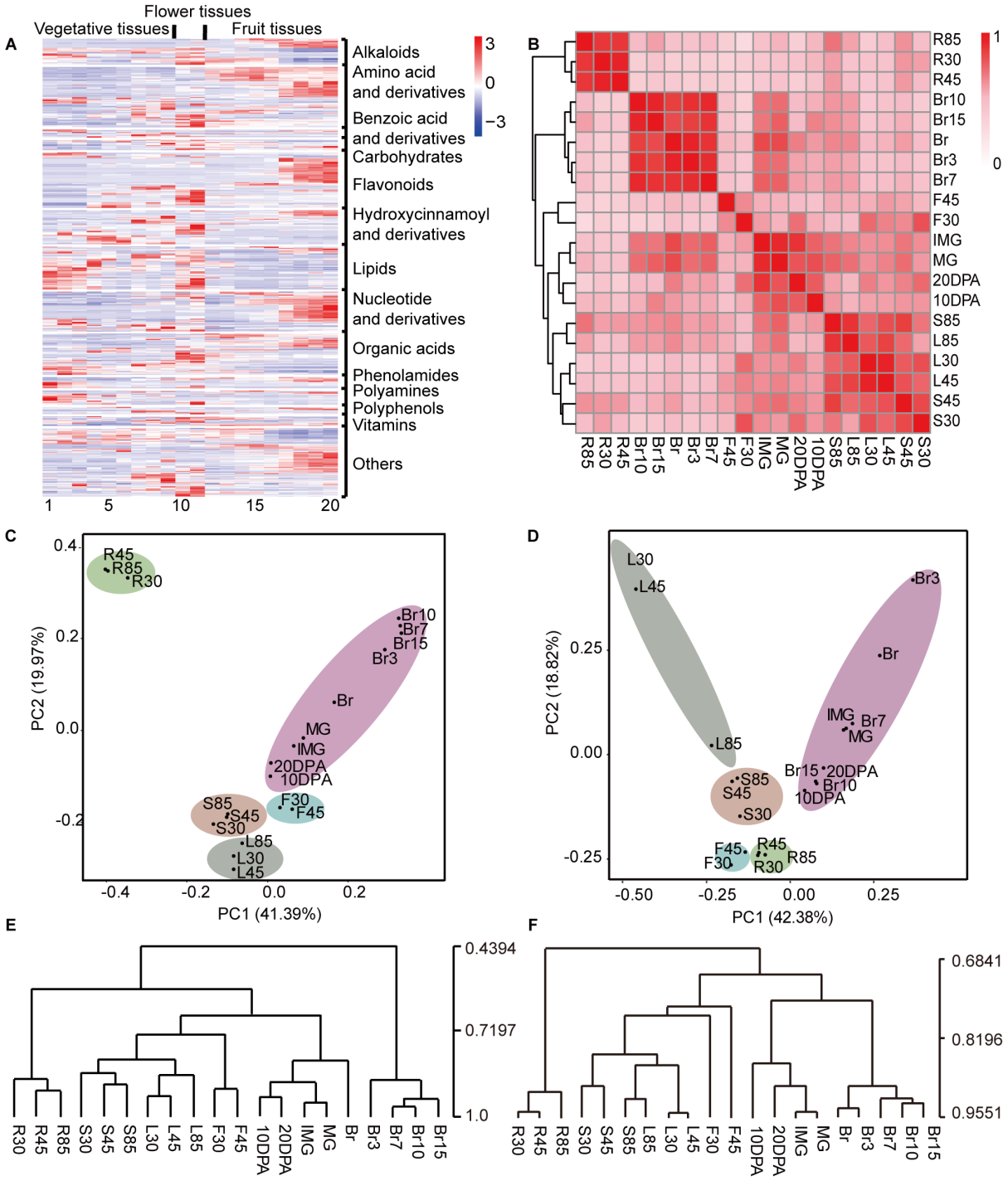
Supplemental Dataset 8. Up-regulated genes in *SlbHLH114-Ox* lines.

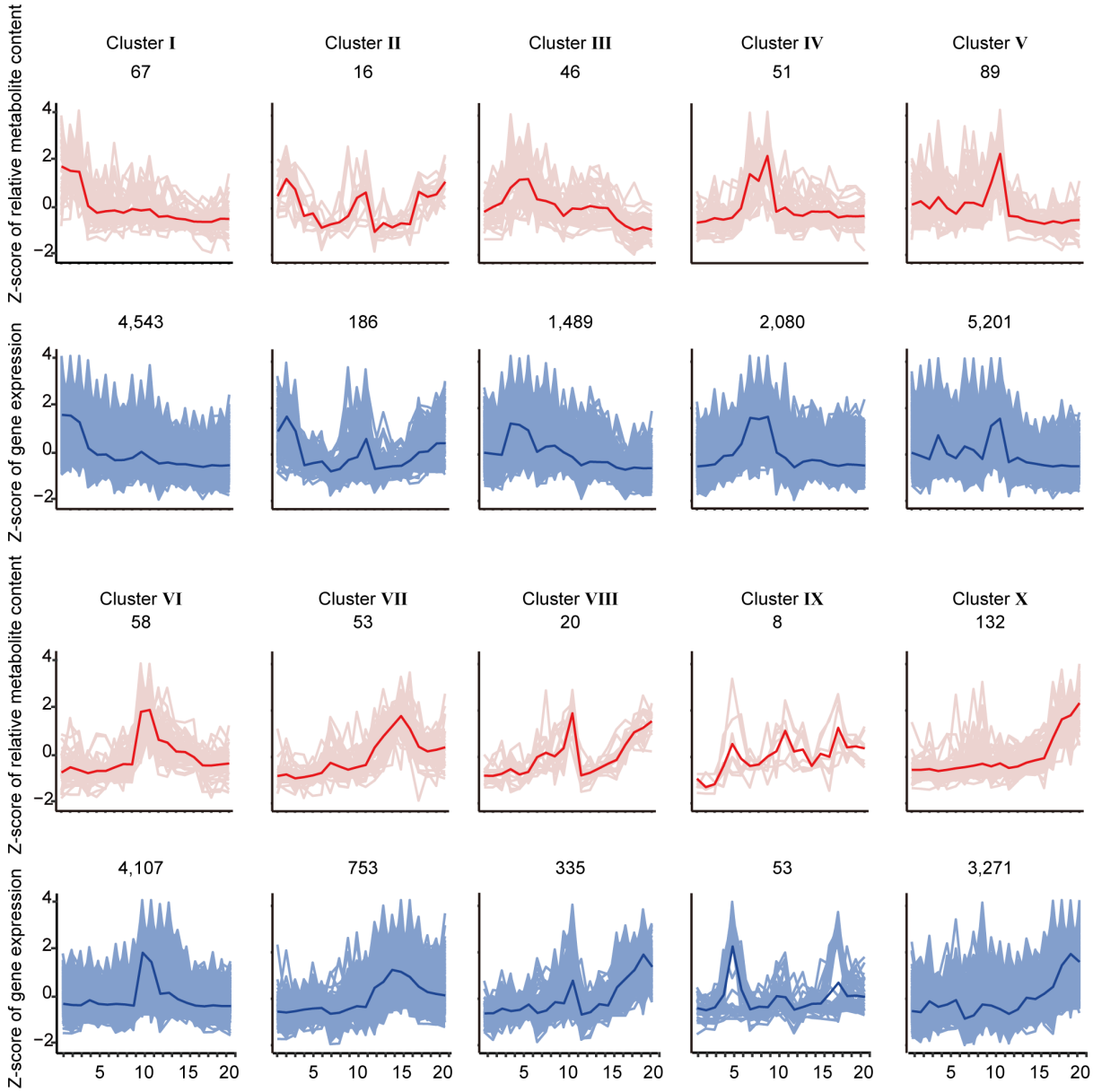
Supplemental Dataset 9. Differentially expressed genes in the *SlbHLH114-Ox* pericarp tissues.

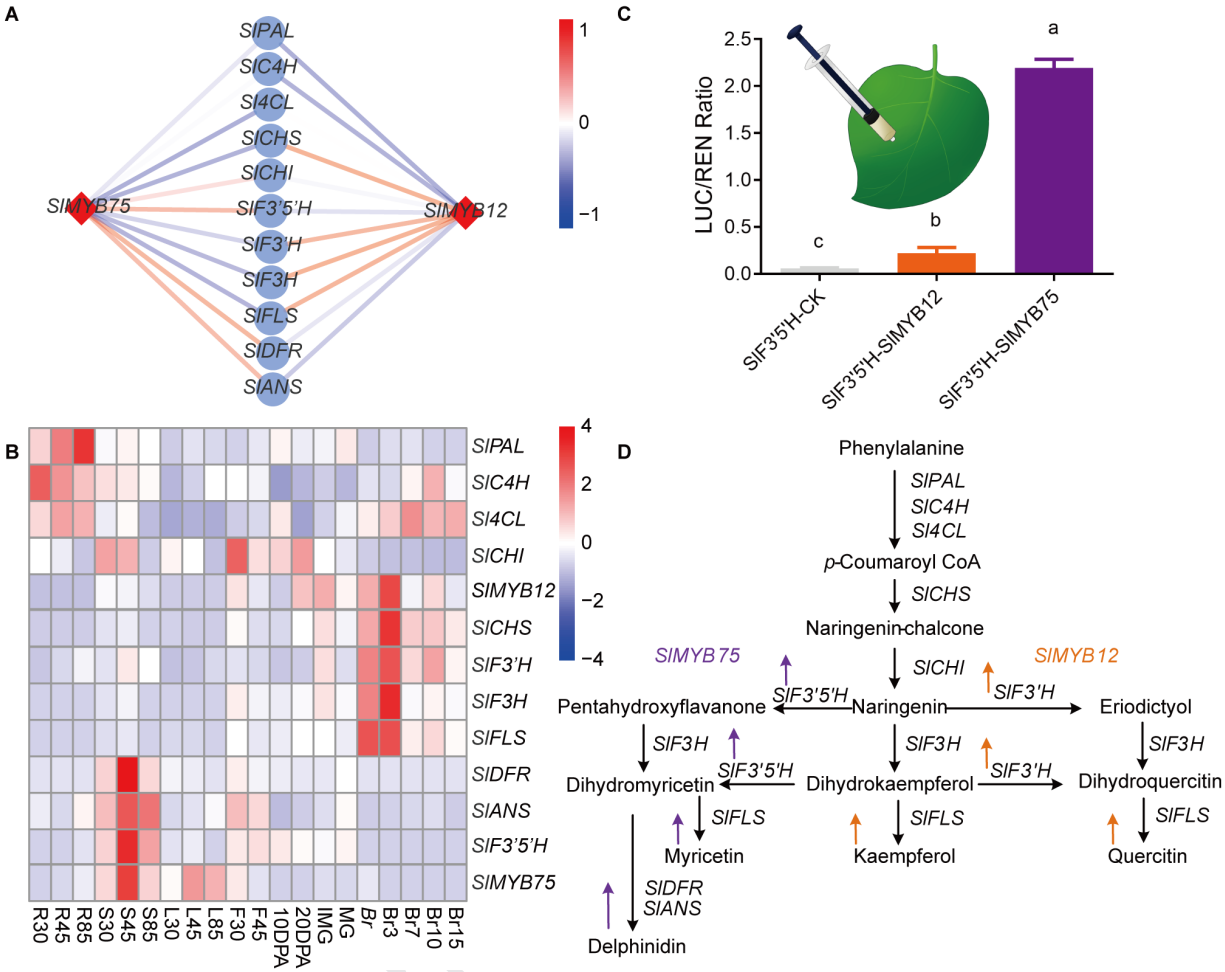
Supplemental Dataset 10. Kyoto Encyclopedia of Genes and Genomes enrichment in co-expression clusters.

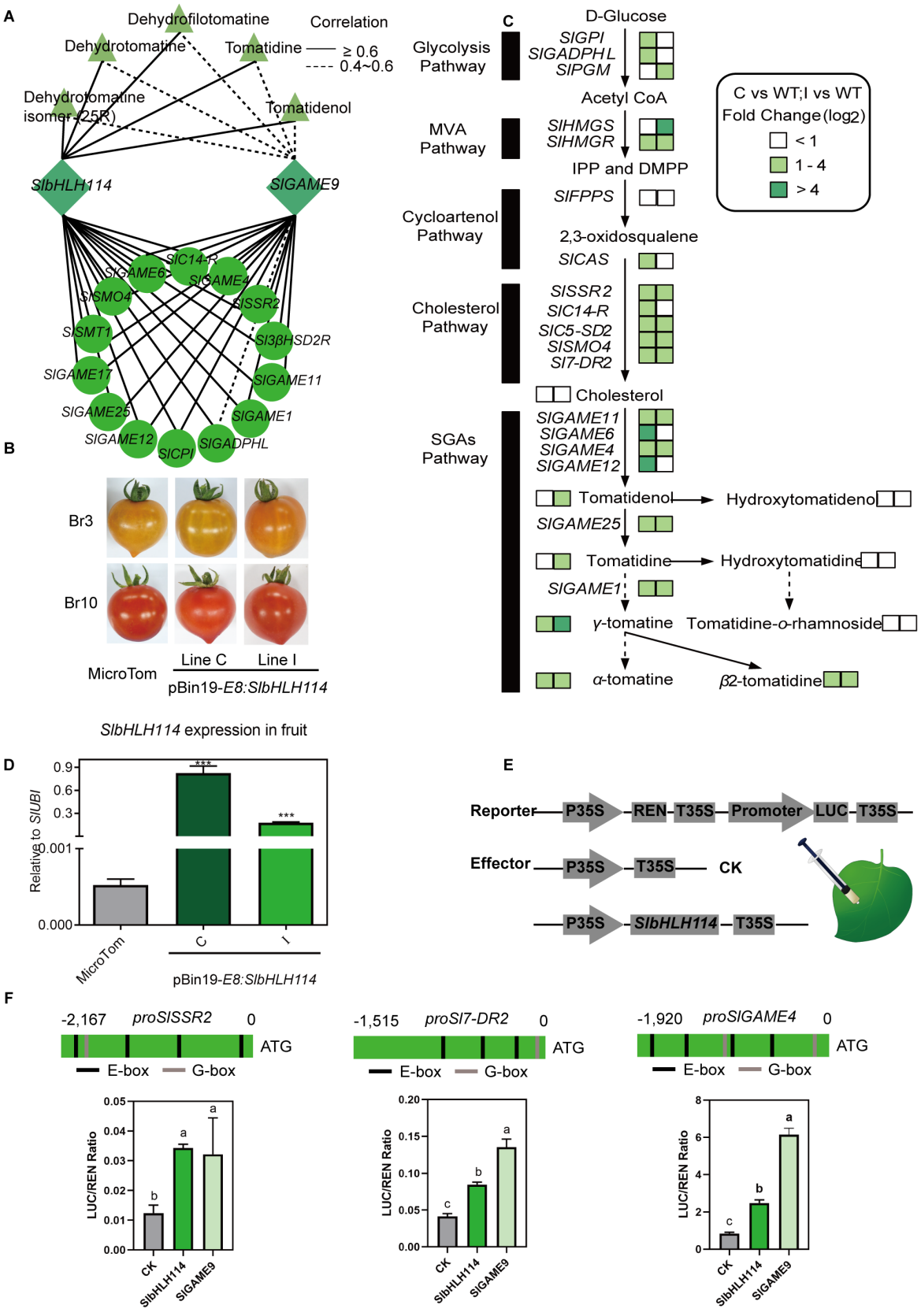
	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII	Cluster VIII	Cluster IX	Cluster X
Compounds	67	16	46	51	89	58	53	20	8	132
Genes	4,543	186	1,489	2,080	5,211	4,107	753	335	53	3,271
Alkaloids	3	0	4	1	5	8	3	0	0	8
Amino acids and derivatives	7	0	2	4	7	11	22	1	2	20
Benzoic acids and derivatves	2	2	1	4	1	0	0	0	0	0
Carbohydrates	0	0	1	3	2	3	1	0	1	2
Flavonoids	4	0	4	5	9	8	4	3	2	31
Hydroxycinnamoyls and derivatives	1	0	10	9	11	3	1	2	0	6
Lipids	16	4	2	2	24	1	2	1	0	2
Nucleotides and derivatives	4	4	1	3	4	1	3	5	1	24
Organic acids	8	1	5	3	8	12	5	0	0	10
Phenolamides	6	0	0	0	2	1	0	2	1	2
Polyamines	8	1	1	1	4	0	1	0	0	2
Polyphenols	0	0	1	3	1	1	0	0	0	3
Vitamins	0	1	1	0	3	3	3	1	0	0
Others	8	3	13	13	8	6	8	5	1	22

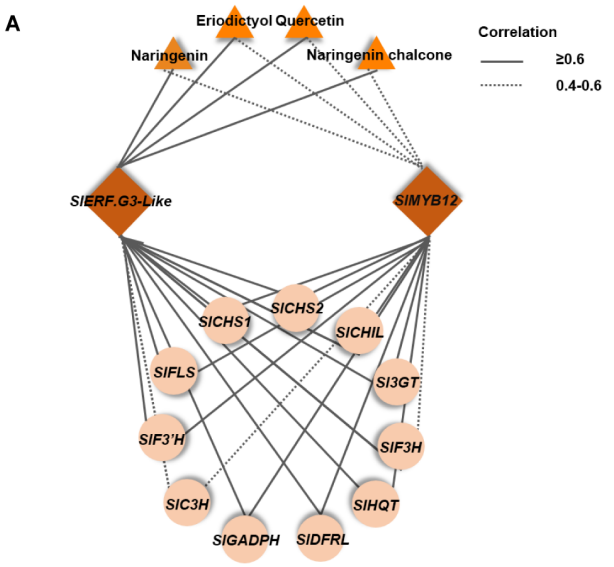


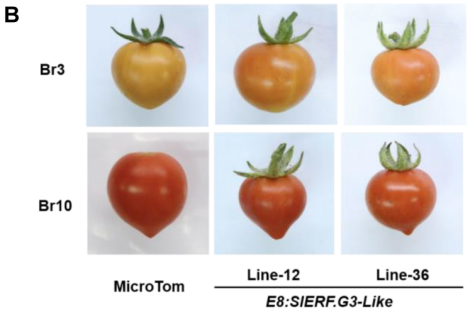

A
Flower tissues
Vegetative tissues
Fruit tissues
Alkaloids
Amino acid and derivatives
Benzoic acid and derivatives
Carbohydrates
Flavonoids
Hydroxycinnamoyl and derivatives
Lipids
Nucleotide and derivatives
Organic acids
Phenolamides
Polyamines
Polyphenols
Vitamins
Others
3
0
−3
1
5
10
15
20
B
R85
R30
R45
Br10
Br15
Br
Br3
Br7
F45
F30
IMG
MG
20DPA
10DPA
S85
L85
L30
L45
S45
S30
1
0
C
PC2 (19.97%)
PC1 (41.39%)
0.4
0.2
0.0
-0.2
-0.4
-0.2
0.0
0.2
R45
R85
R30
Br10
Br7
Br15
Br3
Br
MG
IMG
20DPA
10DPA
S85
S45
S30
F30
F45
L85
L30
L45
D
PC2 (18.82%)
PC1 (42.38%)
0.25
0.00
-0.25
-0.50
-0.25
0.00
0.25
L30
L45
Br3
Br
IMG
Br7
MG
L85
Br15
20DPA
Br10
10DPA
S85
S45
S30
F45
R45
F30
R30
R85
E
0.4394
0.7197
1.0
R30
R45
R85
S30
S45
S85
L30
L45
L85
F30
F45
10DPA
20DPA
IMG
MG
Br
Br3
Br7
Br10
Br15
F
0.6841
0.8196
0.9551
R30
R45
R85
S30
S45
S85
L85
L30
L45
F30
F45
10DPA
20DPA
IMG
MG
Br
Br3
Br7
Br10
Br15

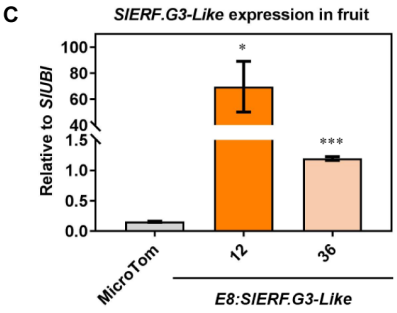

Cluster I
67
Cluster II
16
Cluster III
46
Cluster IV
51
Cluster V
89
4,543
186
1,489
2,080
5,201
Cluster VI
58
Cluster VII
53
Cluster VIII
20
Cluster IX
8
Cluster X
132
4,107
753
335
53
3,271
Z-score of relative metabolite content
Z-score of gene expression
Z-score of relative metabolite content
Z-score of gene expression
4
2
0
−2
5
10
15
20

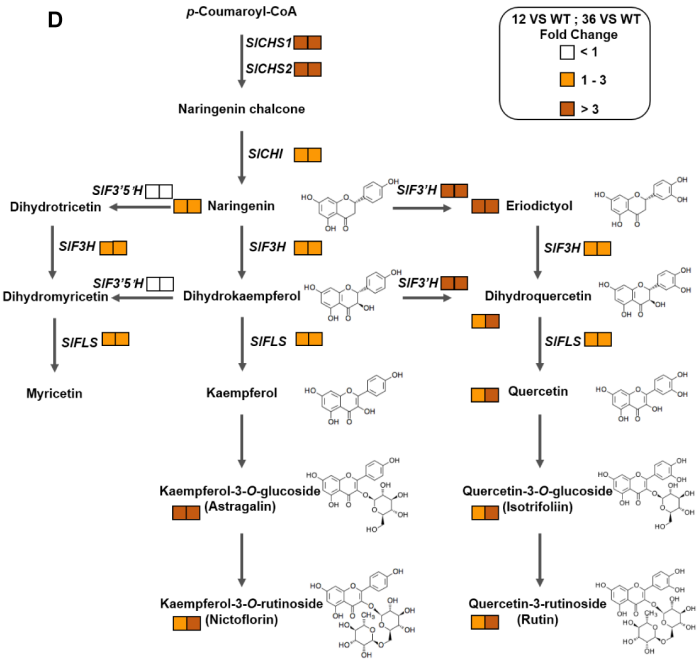

A
SlPAL
SlC4H
Sl4CL
SlCHS
SlCHI
SlF3'5'H
SlF3'H
SlF3H
SlFLS
SlDFR
SlANS
SlMYB75
SlMYB12
1
0
−1
B
SlPAL
SlC4H
Sl4CL
SlCHI
SlMYB12
SlCHS
SlF3'H
SlF3H
SlFLS
SlDFR
SlANS
SlF3'5'H
SlMYB75
4
2
0
−2
−4
R30
R45
R85
S30
S45
S85
L30
L45
L85
F30
F45
10DPA
20DPA
IMG
MG
Br
Br3
Br7
Br10
Br15
C
LUC/REN Ratio
2.5
2.0
1.5
1.0
0.5
0.0
a
b
c
SlF3'5'H-CK
SlF3'5'H-SlMYB12
SlF3'5'H-SlMYB75
D
Phenylalanine
SlPAL
SlC4H
Sl4CL
p-Coumaroyl CoA
SlCHS
Naringenin-chalcone
SlMYB75
SlCHI
SlMYB12
SlF3'5'H
SlF3'H
Pentahydroxyflavanone
Naringenin
Eriodictyol
SlF3H
SlF3H
SlF3H
SlF3'5'H
SlF3'H
Dihydromyricetin
Dihydrokaempferol
Dihydroquercitin
SlFLS
SlFLS
SlFLS
Myricetin
Kaempferol
Quercitin
SlDFR
SlANS
Delphinidin


A
Dehydrofilotomatine
Dehydrotomatine
Tomatidine
Correlation
≥ 0.6
0.4~0.6
Dehydrotomatine isomer (25R)
Tomatidenol
SlbHLH114
SlGAME9
SlC14-R
SlGAME6
SlGAME4
SlSMO4
SlSSR2
SlSMT1
Sl3βHSD2R
SlGAME17
SlGAME11
SlGAME25
SlGAME1
SlGAME12
SlCPI
SlGADPHL
B
Br3
Br10
MicroTom
Line C
Line I
pBin19-E8:SlbHLH114
C
D-Glucose
Glycolysis Pathway
SlGPI
SlGADPHL
SlPGM
Acetyl CoA
MVA Pathway
SlHMGS
SlHMGR
IPP and DMPP
SlFPPS
Cycloartenol Pathway
2,3-oxidosqualene
SlCAS
Cholesterol Pathway
SlSSR2
SlC14-R
SlC5-SD2
SlSMO4
Sl7-DR2
Cholesterol
SGAs Pathway
SlGAME11
SlGAME6
SlGAME4
SlGAME12
Tomatidenol
Hydroxytomatidenol
SlGAME25
Tomatidine
Hydroxytomatidine
SlGAME1
γ-tomatine
Tomatidine-α-rhamnoside
α-tomatine
β2-tomatidine
C vs WT;I vs WT
Fold Change (log2)
< 1
1 - 4
> 4
SlbHLH114 expression in fruit
D
Relative to SlUBI
0.9
0.6
0.3
0.001
0.000
***
***
MicroTom
C
I
pBin19-E8:SlbHLH114
E
Reporter
P35S
REN
T35S
Promoter
LUC
T35S
Effector
P35S
T35S
CK
P35S
SlbHLH114
T35S
F
-2,167
proSlSSR2
0
ATG
E-box
G-box
-1,515
proSl7-DR2
0
ATG
E-box
G-box
-1,920
proSlGAME4
0
ATG
E-box
G-box
LUC/REN Ratio
CK
SlbHLH114
SlGAME9


A
Eriodictyol
Quercetin
Naringenin
Naringenin chalcone
Correlation
≥0.6
0.4-0.6
SlERF.G3-Like
SlMYB12
SlCHS2
SlCHS1
SlCHIL
SlFLS
Sl3GT
SlF3'H
SlF3H
SlC3H
SlHQT
SlGADPH
SlDFRL


B
Br3
Br10
MicroTom
Line-12
Line-36
E8:SlERF.G3-Like


C
SlERF.G3-Like expression in fruit
Relative to SlUBI
100
80
60
40
1.5
1.0
0.5
0.0
*
***
MicroTom
12
36
E8:SlERF.G3-Like


D
p-Coumaroyl-CoA
SlCHS1
SlCHS2
Naringenin chalcone
SlCHI
SlF3'5'H
Dihydrotricetin
Naringenin
SlF3'H
Eriodictyol
SlF3H
SlF3H
SlF3H
SlF3'5'H
Dihydromyricetin
Dihydrokaempferol
SlF3'H
Dihydroquercetin
SlFLS
SlFLS
SlFLS
Myricetin
Kaempferol
Quercetin
Kaempferol-3-O-glucoside
(Astragalin)
Quercetin-3-O-glucoside
(Isotrifoliin)
Kaempferol-3-O-rutinoside
(Nictoflorin)
Quercetin-3-rutinoside
(Rutin)
12 VS WT ; 36 VS WT
Fold Change
< 1
1 - 3
> 3