Running head:
Network analyses of tomato fruit shape regulation

Author for correspondence:
Esther van der Knaap
Ohio State University, Ohio Agricultural Research and Development Center
Department of Horticulture and Crop Science
1680 Madison Ave
Wooster OH 44691
330-263-3822
Vanderknaap.1@osu.edu
Title:
Network analyses reveal shifts in transcript profiles and metabolites that accompany the expression of SUN and an elongated tomato fruit

Josh P. Clevenger¹, Jason Van Houten¹, Michelle Blackwood², Gustavo R. Rodriguez¹, Yusuke Jikumaru³, Yuji Kamiya³, Miyako Kusano³, Kazuki Saito³,⁴, Sofia Visa² and Esther van der Knaap¹

¹ Department of Horticulture and Crop Science, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH 44691, USA
² Department of Mathematics and Computer Science, The College of Wooster, Wooster OH 44691, USA
³ RIKEN Center for Sustainable Resource Science, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
⁴ Department of Molecular Biology and Biotechnology, Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan.

Summary:
High expression of SUN leads to elongated tomatoes and this is accompanied by dramatic shifts in gene expression, and metabolite and hormone accumulation during the early stages of fruit development.
Footnotes:

Financial source: This work was supported by the National Science Foundation grant IOS 0922661 to SV and EvdK. JPC was supported in part by an Ohio State University Presidential Fellowship.

Present addresses:

JPC: Institute of Plant Breeding, Genetics, & Genomics, University of Georgia, Tifton GA 31793

JVH: Basic Medical Science Program, Wayne State School of Medicine, Detroit MI 48201

GRR: Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Rosario, Argentina. grodrig@unr.edu.ar

YJ: Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572, Japan. miyako.kusano@riken.jp

Corresponding author:

Esther van der Knaap; Vanderknaap.1@osu.edu
Abstract

*SUN* controls elongated tomato (*Solanum lycopersicum*) shape early in fruit development through changes in cell number along the different axes of growth. The gene encodes a member of the IQD family characterized by a calmodulin-binding motif. To gain insights into the role of *SUN* in regulating organ shape, we characterized genome-wide transcriptional changes, and metabolite and hormone accumulation following pollination and fertilization in wild type (WT) and *SUN* fruit tissues. Pericarp, seed/placenta and columella tissues were collected at 4, 7 and 10 days after anthesis. Pairwise comparisons between *SUN* and WT identified 3,154 significant differentially expressed genes that cluster in distinct Gene Regulatory Networks (GRN). GRNs that were enriched for cell division, calcium/transport, lipid/hormone, cell wall, secondary metabolism and patterning processes contributed to profound shifts in gene expression in the different fruit tissues as a consequence of high expression of *SUN*. Promoter motif searches identified putative cis-elements recognized by known transcription factors and motifs related to Mitotic Specific Activator sequences. Hormone levels did not change dramatically, but some metabolite levels were significantly altered, namely participants in glycolysis and the TCA cycle. Also, hormone and primary metabolite networks shifted in *SUN* compared to WT fruit. Our findings imply that *SUN* indirectly leads to changes in gene expression most strongly of those involved in cell division, cell wall and patterning-related processes. When evaluating global co-regulation in *SUN* fruit, the main node represented genes involved in calcium-regulated processes, suggesting that *SUN* and its calmodulin-binding domain impact fruit shape through calcium signaling.
Introduction

Tomato is an important vegetable that is extensively studied for aspects related to plant and fruit development, fruit ripening and quality (de Jong et al., 2011; Goulet et al., 2012; Matas et al., 2011; Mazzucato et al., 2013; Osorio et al., 2013; Pan et al., 2013; Ruan et al., 2012; van der Knaap et al., 2014; Zhong et al., 2013). As a member of the Solanaceae family, tomato is also an excellent genetic system due to its diploid and highly inbred nature, as well as the availability of genetic and genomic resources such as a reference genome sequence (Tomato et al., 2012). Fruit development initiates at anthesis (flower opening) which in tomato is marked by the release of pollen and pollination. After successful fertilization of the ovules, fruit growth commences. In tomato and many other plants, the initial growth stages are characterized by cell division followed by the cell expansion of the maternally derived fruit tissues (Gillaspy et al., 1993; Xiao et al., 2009). Crosstalk among the various hormones play key roles during the initial stages of fruit growth (de Jong et al., 2009; Gillaspy et al., 1993; Montoya et al., 2005; Ozga et al., 2003).

Growth of plant organs including the fruit occurs along three axes: the proximal-distal, the medio-lateral and the abaxial-adaxial axis. Fruit length is determined by the degree of growth along the proximal-distal axis whereas fruit width is determined by the degree of growth along the medio-lateral axis. The degree of the pericarp thickness and other internal tissues is determined along the abaxial-adaxial axis. These growth axes are defined much earlier, namely during the formation of the floral meristem and gynoecium primordia in the developing flower. Correct initiation of the gynoecium requires the specification of organ and tissue identity as well as the establishments of the boundaries between primordia to ensure that the appropriate identities and division patterns are initiated and maintained throughout organ growth (Balanza et al., 2006; Dinneny et al., 2005; Girin et al., 2009; van der Knaap et al., 2014). Hormones such as auxin play a critical role in the setting up the patterns of cell division by regulating the expression of transcription factors that control ovary development along the three axes of growth (Nole-Wilson et al., 2010; Staldal and Sundberg, 2009; Wang et al., 2011).

The role of patterning genes in the initiation and growth of organ primordia is relatively well understood. Except in fruit ripening (Vrebalov et al., 2009), the role of patterning genes in fruit ontogeny is largely unknown. Soon after initiation, the different tissue types within the developing gynoecium arise by reactivation of the cell identity and cell division in the so called
“quasi meristems” (zones of active cell proliferation within an organ) (Girin et al., 2009). The “quasi meristems” give rise to medial tissues such as the ovules, the septum, style, stigma and a structure specific to fruit of members in the Brassicaceae family, the replum (Girin et al., 2009). Other than the embryo and the placenta surrounding the developing seeds, fruit development is not typically marked by newly arising tissues. Yet rates of cell division and cell expansion are enhanced following fertilization (Gillaspy et al., 1993; Xiao et al., 2009) and it is therefore conceivable that the patterning genes also play important roles during the initial stages of fruit development.

The tomato fruit shape gene SUN regulates proximal-distal patterning by controlling the elongation of the fruit (Xiao et al., 2008). The mutation resulted from an interchromosomal gene duplication event whereby the coding region was placed in a different genome context resulting in high expression of the derived version of SUN in tissues and at developmental time points when the ancestral version of this gene is barely expressed (Wu et al., 2011; Xiao et al., 2009). SUN encodes a protein of the IQD family characterized by the IQ67 motif which is known to bind calmodulin (CaM) (Levy et al., 2005; Xiao et al., 2008). CaM-binding proteins have extremely diverse functions in plants, including metabolism regulation and hormone signaling (Kim et al., 2009). The effect on fruit shape by SUN is noticeable at anthesis but is most dramatic seven to 10 days post-anthesis (dpa) (Van der Knaap and Tanksley, 2001; Wu et al., 2011; Xiao et al., 2009). SUN does not control fruit weight but instead controls the redistribution of fruit mass and the degree of elongation is positively correlated to the level of gene expression (Wu et al., 2011; Xiao et al., 2008). Throughout development from the ovary at anthesis until the breaker stage, SUN is expressed highly in the SUN mutant compared to wild type (WT) fruit (Xiao et al., 2009). The cellular basis of this elongated shape is clearly visible at 7 dpa, where an increase in cell number is found in columella and septum tissues in the proximal-distal direction and a decrease in the medio-lateral direction in plants carrying the SUN mutation compared to WT. Cell size is not significantly different in SUN compared to WT fruit. Based on these results, it was proposed that SUN results in changes of the plane of cell division towards more cells in the proximal-distal direction (Wu et al., 2011). Little is known about the function of the IQD family despite their widespread occurrence in plant genomes (Abel et al., 2005; Huang et al., 2013). In Arabidopsis, increased expression of AtIQD1 leads to higher glucosinolate levels which coincide with increased expression of certain enzymes in the shared auxin/glucosinolate
pathway involving the intermediate metabolite IAOx (Levy et al., 2005). In tomato however, glucosinolates are not produced and the IAOx pathway is only found in the Brassicaceae family (Glawischnig et al., 2004; Sugawara et al., 2009; Won et al., 2011; Zhao, 2010) suggesting that the role of these proteins is not specific to glucosinolate production. In support of that notion, AtIQD1 was found to localize to microtubules in planta and physically interacting with a kinesin motor protein, KLCR1 (Burstenbinder et al., 2013). These results imply that members of the IQD family may provide a scaffold for protein transport along the microtubules via kinesin motor proteins (Abel et al., 2013; Burstenbinder et al., 2013). In tomato, SUN overexpression leads to extreme phenotypes such as seedless fruit with a pointed shape, twisted stems and leaf rachis, altered vascular patterning as well as increased leaf serration (Wu et al., 2011). Combined, these traits are associated with altered auxin homeostasis in tomato and other plants (Bouchard et al., 2006; de Jong et al., 2009; Ge et al., 2014; Molesini et al., 2009; Sawchuk et al., 2013; Wu et al., 2011). SUN expression does not lead to extensive changes in expression of AUX/IAA or ARF, therefore, SUN has been proposed to function in auxin–regulated processes by either increasing its biosynthesis or altering polar transport (Wu et al., 2011; Xiao et al., 2009). However, direct links between SUN expression and the auxin pathway have not been established.

The goals of this study were to investigate the molecular mechanisms by which SUN controls fruit shape in tomato. To do so, we carried out a network analyses approach to gain insights into SUN’s role in fruit development. Three tissues at different stages of fruit development were collected from WT and SUN plants and evaluated for gene expression, primary metabolite and hormone profiles. By clustering gene expression based on similar profiles in the different tissues over time, and evaluating metabolite and hormone accumulation, putative regulatory networks were identified that accompanied the early stages of fruit development when the changes in fruit shape mediated by SUN became apparent. The results demonstrated that the early stages of tomato fruit development were highly dynamically controlled. Moreover, clusters of genes that were co-differentially expressed in SUN compared to WT share common promoter elements which might explain the co-differential expression patterns. The findings provided novel information as to how SUN may regulate elongated fruit shape, while also providing a framework for the metabolite, hormone and gene networks that are operating during the early stages of tomato fruit development.
Results

Clustering of co-regulated genes in developing fruit tissues

Fruit growth following pollination consists of a rapid increase in cell division and the initiation of seed development, followed by a period of cell expansion and organ enlargement (Gillaspy et al., 1993; Xiao et al., 2009). The effect of SUN on fruit shape is most dramatic immediately following pollination and, in cultivated tomato, the shape is final at 10 to 12 dpa (Fig. 1) (Van der Knaap and Tanksley, 2001). The tissue that is most affected in the change in shape is the central part of the fruit, namely the columella and septum (Wu et al., 2011). To identify key processes that accompanied the initial stages of fruit development following pollination, we first evaluated global variation in gene expression in WT fruits in the following tissues and time points: pericarp, columella and the combined seed/placenta tissues at 4, 7 and 10 dpa (Tables S1 and S2). Using fuzzy C-means, we selected a subset of clusters out of 100 based on the most dynamic expression changes in a spatial and temporal manner (Fig. 2, Table S3). The selected clusters represented dynamic developmental changes in expression programs during early fruit development and included enrichment of different bin ontologies. Cell division (Fig. 2A; 604 genes) exhibited peak expression in 7 dpa pericarp tissue and shares a similar pattern with RNA (Fig. 2C; 224 genes), a cluster enriched for transcription factors. Showing peak expression during 4 and 7 dpa pericarp, the cluster enriched for secondary metabolism (Fig. 2D; 396 genes) was also enriched for lipid metabolism, cell division, and transport (Table S4A). The cell wall (Figure 2B; 188 genes) enriched gene cluster showed peak expression at 7 dpa in all three fruit tissues. Interestingly, a cluster enriched for transport/signaling showed peak expression in 4 dpa columella and decreased as development of this tissue proceeded (Fig. 2F; 224 genes). This cluster’s expression program was in contrast to a cluster enriched for transport/hormone which showed peak expression at 10 dpa columella (Fig. 2E; 353 genes) implying that signaling led to hormonal and transport responses during columella development. A cluster of genes enriched for photosynthesis (Fig. 2G; 201 genes), showed peak expression in 4 and 7 dpa columella. This cluster was also enriched for TCA cycle, CHO metabolism, OPP, and protein demonstrating an emphasis on metabolism in columella tissue during the earliest stages of fruit development. At 10 dpa, the latest developmental stage we assayed, a cluster of protein-related enriched genes showed peak expression in all fruit tissues, presumably setting the stage for the next phase of fruit development.
Figure 1. Tomato fruit growth following pollination. (A) wild type fruit at anthesis, 4 dpa, 7 dpa, 10 dpa and 16 dpa. (B) SUN fruit at anthesis, 4 dpa, 7 dpa, 10 dpa and 16 dpa. Size bar represents 1 cm. (C) Expression of SUN in WT and SUN fruit tissues and developmental time points. Per, pericarp; col, columella; se, seed and placenta.
Collectively the temporal and spatial...
clusters demonstrated a coordinated program of gene expression that was dynamically executed during early fruit development. To investigate whether SUN expression altered overall gene expression programs, we next evaluated those that were differentially regulated in the same tissues and time points.

Clustering of the co-differential expressed genes in SUN and WT fruit

To provide insights into how SUN regulate shape and whether this is accompanied with shifts in transcript profiles, we identified differentially expressed genes in the eight pairwise comparisons of SUN and WT fruit tissues and time points. The total number of significant differentially expressed genes in at least one comparison was 3,154 based on DESeq analysis and multiple Bonferroni corrections (Anders and Huber, 2010; Anders et al., 2013). Seven genes showed robust differential expression in all tissues at all developmental time points. Three of the genes were expected, namely SUN itself (Solyc10g079240) (Fig. 1C), and the two defensin genes DEFL1 and DEFL2 (Solyc07g007760; Solyc07g007750) whose expression was perturbed by the retrotransposition of SUN from chromosome 10 to chromosome 7 (Jiang et al., 2009; Xiao et al., 2008). Of the other four genes, three were lower expressed in SUN fruit and one was higher expressed. The higher expressed gene, Solyc07g064380, encoded a serine/threonine phosphatase. Of the three genes that showed lower expression in SUN, two encoded proteins that were involved in secondary metabolism. Solyc06g035940 was a likely paralog of PDF2, encoding a transcription factor regulating cutin biosynthesis (Nadakuduti et al., 2012). Solyc07g006670 encoded a hydroxycinnamoyl CoA quinate transferase participating in chlorogenic acid synthesis.

The differential gene expression appeared dynamic over the course of fruit development and in different tissues (Table S5). To further investigate how elevated SUN expression impacts overall pathways in the developing tomato fruit, we evaluated the enrichment of GO categories in the significantly differentially expressed gene data set (Table S4B). Collectively, the enrichment for genes involved in cell wall, photosynthesis, hormone metabolism, transport, biotic stress, as well as secondary metabolism including lipid metabolism, metal handling, and cell division indicated that these processes were most dramatically affected by high expression of SUN. Linear factorial modeling of the data with genotype and genotype × time point interactions
in the model also supported the importance of many of the same processes in differential fruit growth that is conditioned by SUN (Table S4C and S6).

We clustered the differentially expressed genes based on the log2 fold change of WT and SUN using fuzzy C-means. This analysis resulted in the identification of several gene regulatory networks (GRN) of which 14 exhibited dynamic differences in expression between SUN and WT fruit in the developmental time points and tissue types (Fig. S1 and Table S7). GRN8 showed one of the highest significant enrichment for only one category, namely “cell division” (Table S4D). This cluster included 9 cyclins and 17 cell organization genes including 15 kinesin/microtubule motor genes (Table S7). Of the several networks that were enriched for “cell wall”, GRN4 was most significantly enriched for genes in this category (16% of all genes in the network, Tables S4D and S7). The network included genes involved in cell wall synthesis, degradation, modification, UDP GLUCOSYL and GLUCORONYL TRANSFERASES, INVERTASES, PECTATE LYASES and PECTIN ESTERASES. Another interesting cluster was represented by GRN10 and 14, showing similar dynamics in differential expression (Fig. S1). Combined, they represented genes such as three PIN1-like involved in auxin transport, and enrichment for genes encoding transport related proteins, cell wall, hormone metabolism, several leucine-rich repeat receptor-like kinases and transcription factors putatively involved in patterning. Additional GRNs showed genes enriched in the secondary metabolism category, and included those involved in the isoprenoid, polypropanoid and lignin pathways (GRN2, Table S4D). GRN3 had the most genes related to calcium, in particular those involved in its transport and signaling, and included a member of the IQD family, SISUN25. GRN9 was enriched for genes involved in lipid, hormone and secondary metabolism. In this cluster, four were predicted to act in the brassinosteroid pathway, and four other genes were predicted to act in the steroid metabolism pathway. Several genes involved in isoprenoid and phenylproponoid pathways were found in this network as well.

The log2 fold (SUN/WT) profiles of the total set of 3,154 differentially expressed genes were plotted in the PCA space (Fig. 3A). The highlighting of selected GRNs showed that they were tightly co-differentially regulated in SUN fruit in distinct programs (Fig. 3B). The most intriguing co-differentially regulated cluster was GRN8 enriched for “cell division”-related genes (Fig. 3E). This cluster showed decreased expression in SUN fruit at 7 dpa pericarp and increased expression at 7 dpa columella. At 10 dpa, the genes in this cluster were not differentially
expressed between SUN and WT fruit, coinciding with when the shape controlled by SUN became finalized. GRN2, enriched for secondary metabolism, showed reduced expression in SUN fruit in the pericarp at 4 dpa and increased expression in SUN fruit in seed and columella at 7 dpa (Fig. 3C). The opposite regulation was found in GRN10 and 14, enriched for cell wall/RNA and hormone metabolism/development/transport respectively, demonstrating increased expression in SUN fruit in pericarp at 4 dpa, decreased expression in SUN fruit in seed at 7 dpa, and increased expression in SUN fruit at 4 dpa columella (Fig. 3G). GRN9, enriched for lipid and hormone metabolism, showed increased expression in SUN fruit at 7 dpa in seed and columella and at 10 dpa in seed and pericarp (Fig. 3F).

To investigate in more detail the level of co-expression among selected genes, we focused on those that were known from other studies to participate in shared pathways or were involved in patterning. For brassinosteroid biosynthesis, eight tomato putative orthologs were higher expressed in SUN compared to WT in seed and columella at 7 dpa; and in seed, columella, and pericarp at 10 dpa (Fig. S2A). The putative ortholog of the auxin biosynthesis gene YUCCA4,
auxin influx gene **AUX1**, and auxin efflux gene **PIN1** were higher expressed at 4 dpa columella and pericarp and the putative auxin conjugation gene **JAR1** showed an opposite expression pattern similar to findings from other systems (Fig. S2B). Calcium-related genes were correlated with **SLSUN25** and higher expressed in 4 dpa and 7 dpa columella tissue in SUN fruit while being lower expressed in 7 dpa pericarp (Fig. S2C). **GRF1**, **GRF-INTERACTING FACTOR (GIF)1**, and 12 cyclin genes were also higher expressed at 4 dpa pericarp and 7 dpa columella tissues in SUN fruit (Fig. S2D).

Among the patterning genes, the Arabidopsis genes **PIN1** and **KANADI2** are involved in regulating organ polarity and are co-regulated with one another (Izhaki and Bowman, 2007). In SUN compared to WT tomato fruit, the putative orthologs of **PIN1** and **KANADI2** showed similar differential expression dynamics with a strong increased expression in 4 dpa columella tissue (Fig. S3A). **JACKDAW** and **MAGPIE** control **SHORTROOT (SHR)** activity in Arabidopsis roots (Welch et al., 2007). The putative tomato orthologs showed a similar expression dynamic as in Arabidopsis: when **JACKDAW** and **MAGPIE** were not differentially expressed, **SHR** was differentially expressed in fruit tissues (Fig. S3B). The putative orthologs of the transcription factors, **REPLUMLESS (RPL)**, **AGAMOUS (AG)**, **SHOOT MERISTEMLESS (STM)**, and **PHABULOSA (PHB)** showed co-regulated expression in SUN fruit (Fig. S3C). **RPL** and **APETALA2 (AP2)** showed an opposite differential expression pattern in tomato fruit where **RPL** was most differentially expressed at 4 and 7 dpa in the columella (Fig. S3D). **RPL** and **AG** were expected to be coregulated because they are both repressed by AP2 (Drews et al., 1991).

**Promoter analysis of co-differentially expressed genes**

To gain insights into the molecular mechanism that drives the regulation of the differentially expressed genes found in the same network, we evaluated the promoter sequences in a subset of the GRNs. The selected GRNs were enriched for genes related to the cell cycle (GRN8), RNA/cell wall (GRN10), lipid/hormones (GRN9), and one network featuring genes involved in calcium signaling and transport (GRN3). GRN10 contained several transcription factors representing those putatively involved in patterning such as **AIL5**, **STM**, **KANADI2**, **PHB** as well as three **PIN1-like** genes and **AUX1** involved in auxin transport. Commonly used motif finding programs such as MEME, Cosmo and Wordseeker resulted in few putative cis-elements and
most of them resembled microsatellite sequences (Blackwood et al., 2013). Therefore, we newly
designed and implemented an exhaustive search and frequency-based analysis for 6-mers in 1 kb
of the promoter using Python scripts ran on Hadoop (Hadoop, 2013). For each network and 6-
mer, the occurrence of the motif was counted in the actual and shuffled promoters yielding a
promoter enrichment score (Table S8). For the cell cycle network, 13 putative 6-mers were
recognized based on the enrichment score of 3 or higher. These motifs were then clustered
manually into four consensus motifs (Table I). For the RNA/cell wall network, two 6-mers were
identified which clustered into one consensus motif. For the lipid and hormone network, 18 6-
mers were identified, resulting in six consensus motifs. In the calcium signaling and transport
network, the five 6-mers identified were clustered into three consensus motifs (Table I). Some of
the identified motifs were unknown whereas others had been described previously to bind MYB
or TCP transcription factors, involved in cell cycle regulation or represented putative ABA
response elements (Abe et al., 1997; Ito, 2000; Ito et al., 1998; Iwasaki et al., 1995; Schommer
et al., 2014; Schommer et al., 2008; Zhang et al., 2010) (Table S8).

To test the validity of the computationally identified 6-mers, the enrichment score for each
was evaluated in the promoters of the genes in the other three networks. For the cell cycle
cluster, all four consensus elements were occurring at much higher enrichment scores in the cell
cycle than in any of the other networks (Table I). In fact, two of its consensus motifs, TRRCCGT
and CCACGGYYA, were 22- to 5- fold higher enriched in the cell cycle network than in any of
the others. On the other hand, the single consensus motif found in the RNA/cell wall network,
TGGACCA, was highest in its own network but only slightly lower in the lipid/hormone
network. This result implied that the genes in the RNA/cell wall and lipid/hormone network may
be co-regulated via the same putative 6-mer. Indeed, the RNA/cell wall GRN was represented by
several genes involved in auxin transport. All six consensus elements from the lipid/hormone
network showed higher enrichments scores compared to the other three networks (a 25- to 3-
fold higher representation). Finally, the three consensus motifs from the calcium network were 3-
to 2- fold higher enriched in this compared to the other three networks.

The genes whose promoters carried the “cell cycle” network consensus motif,
CCAACGGYYA expected to recognize MYB transcription factors or the mitotic-specific
activator sequences, the “RNA/cell wall” consensus motif TGGACCA possibly recognized by
TCP transcription factors, the “lipid/hormone” consensus motif, CAYRTG expected to be an
ABA response element, and the unknown “calcium/transport” consensus motif AGGTSATG, were further investigated (Fig. 4). The expression profiles of the genes carrying these motifs in their promoters clustered according to genotype (SUN or WT), clearly demonstrating that these genes were co-differentially expressed. To validate these results, we analyzed a random set of differentially expressed genes which did not cluster based on genotype (Fig. 4C and D). Thus,
SUN expression is associated with extensive reprogramming of genes in the developing fruit that is coinciding with the changes in fruit growth patterns.

SUN modulates metabolite networks to form new associations with the TCA cycle and abolishes associations with amino acids

To investigate whether metabolite accumulation also differed in SUN and WT fruit, we used the same tissues that were analyzed for gene expression analysis and evaluated the metabolite levels using gas chromatography–time-of-flight–mass spectrometry (GC-MS). Pair-wise comparisons were made between SUN and WT fruit and placed into a global metabolic context by constructing a pathway map of significant changes using a relaxed statistical threshold (p < 0.05) (Fig. 5). The pathway map of significant changes between SUN and WT showed that columella tissue at 7 and 10 dpa exhibited the most significant changes (Fig. 5). Globally, decreased accumulation of intermediates in the TCA cycle, and increased accumulation in certain sugars and amino acids were observed in SUN although these changes were less than two-fold. Certain metabolites also showed significant changes between SUN and WT in seed tissue 10 dpa in the same pattern as columella tissue 10 dpa (Fig. S4). However, these findings should be interpreted with caution since pair-wise comparisons may oscillate around the mean which could lead to false positive and false negative results and the fold differences are relatively low.

We next evaluated correlations of the metabolites with one another in SUN and WT fruit. The relative accumulation for each primary metabolite that was evaluated in SUN and WT tissues showed that many did not differ dramatically except certain sugars and organic acids (Fig. S5). Sucrose levels were lower in SUN fruit at 4, 7, and 10 dpa columella. In contrast, glucose-6-phosphate was higher in 10 dpa columella in SUN fruit. Another sugar, trehalose, was higher in 10 dpa pericarp in SUN fruit. The organic acids, fumaric acid and succinic acid, were lower in 7 and 10 dpa columella whereas fumaric acid was lower in 10 dpa pericarp. Pair-wise correlations of each metabolite were conducted separately. The analysis suggested positive and negative correlations between metabolites that were only found in SUN (Table S9 and Fig. S6). Conversely, positive and negative correlations were also found only in WT fruit (Table S10 and Fig. S7). We subsequently extracted the new and abolished metabolite connections and created...
subnetworks of those metabolites exhibiting ten or more correlations (Fig. S8). These hubs mirror the results of the pairwise comparisons, suggesting that these metabolites appeared to be affected in SUN fruit, at least in the later stage of fruit development that were assayed.

Profiling of hormone levels in SUN and WT fruit

Hormone levels were profiled in SUN and WT fruit in 4 dpa columella/seed and pericarp tissue; and in 10 dpa columella, seed, and pericarp tissue (Table S11). Eleven hormones were profiled including auxin (IAA); gibberellic acid (GA1 and GA4); brassinolide (BL); the cytokinins (CK) trans-zeatin (tZ), dihydrozeatin (DHZ), isopentenyl adenine (iP); jasmonic acid; and the ABA-related hormones: abscisic acid (ABA), and their conjugates. The levels of these hormones were compared between SUN and WT. Some of the changes in hormone levels were significant at the p < 0.05 level (Fig. S9).
acid (JA) and its activated conjugate JA-isoleucine (JA-Ile); salicylic acid (SA); and abscisic acid (ABA). We identified those hormonal networks that were only found in SUN or only in WT (Fig. S9A-B). Two new subnetworks emerged in SUN fruit (Fig. S9A and C). One was a network with ABA as the hub and positive correlations with IAA and the cytokinins iP, DHZ, and tZ. iP and tZ also were highly correlated in SUN compared to WT (Pearson’s Correlation Coefficient (PCC) > 0.98). Additionally, GA4 and JA-Ile had formed a new subnetwork in SUN fruit. One network with DHZ as the hub was only found in WT and was therefore abolished in SUN fruit (Fig. S9B and D). Negative correlations between IAA and GA4, and IAA and DHZ were only found in WT as well as negative correlations between DHZ and GA1.

Hierarchical clustering of hormone correlations showed three distinct groups in WT fruits: GA; the defense hormones JA, JA-Ile, SA, and ABA; and IAA, BL and the CK (Fig. S9D). In SUN compared to WT, the grouping broke down as ABA became correlated with IAA, iP, and tZ; and JA-Ile became correlated with GA (Fig. S9C). However, overall hormone levels did not change dramatically in these experiments, and therefore the results should be interpreted with caution.
Discussion

Tomato fruit initiation and associated transcription profiles

By conducting transcriptome analyses of separate tissues at an early stage of fruit development, we identified co-regulated gene networks that differed in SUN and WT fruit. Using different approaches, a cluster of genes involved in cell cycle regulation and cell division (“cell division” cluster, Fig. 2A; GRN8, Fig. S1) was found to differ most dramatically during fruit development while also being differentially expressed in SUN and WT fruit. Genes in GRN8 showed the highest co-regulation, the most striking genotype-specific regulation of all the co-regulated gene clusters, and the highest enrichment score of cluster-specific promoter motifs identified in this study. For example, six of the eight cyclin genes in the cluster contain the promoter consensus motif CCAACGGYYA, including CYCB1;2 (solyc01g009040, solyc10g078330), CYCB2;4 (solyc02g082820, solyc04g082430) and FZR3 (solyc06g043150). Thus, this consensus motif represents a potential target for regulation of cell division patterns. The data show that the cell cycle is one of the key processes that is most dynamically altered by SUN, which is consistent with its effect on increasing cell number on the proximal-distal axis while decreasing cell number on the medio-lateral axis (Wu et al., 2011).

Prior studies into gene expression regulated by SUN were less informative because the RNA was isolated from whole organs as opposed to separate tissue types (Xiao et al., 2009). This demonstrates that expression analyses of specific tissues are more insightful than expression analysis of whole organs.

A second cluster of genes, encoding those involved in cell wall-related processes, is dynamically and differentially expressed in developing fruit tissues, and highly enriched based on GO terms (“cell wall” cluster, Fig. 2; GRN4/10 and others, Fig S1; Table S4). The shift in expression of the cell wall-related genes as early as 4 days after fruit initiation implied that high expression of SUN leads to changes in cell wall properties that may precede the differential expression of genes related to cell division. The cell wall, its composition and structure are known to vary during tissue growth (Hernandez-Hernandez et al., 2014; Mirabet et al., 2011). Two models explain organ growth that result in divergent shapes: 1) Stress-based axiality whereby genes that alter patterning lead to cellular stresses influencing the cytoskeleton which then drives the changes in the orientation of growth. 2) Polarity-based axiality whereby genes
that affect the distribution of signaling molecules define the polarity within the tissue (Kennaway et al., 2011; Uyttewaal et al., 2010). In both cases, mechanical constraints within the tissue caused by the properties of cell walls together with stress and/or polarity-based drivers are thought to result in the final morphology of plant organs. Thus, it may be expected that changes in expression of cell wall related genes are associated with changes in organ shapes.

Another group of differentially expressed genes were putative orthologs of patterning genes. The role of these genes during growth of plant organs after their initiation is largely unknown. Reactivation of meristematic cells in otherwise differentiated zones has been proposed for the formation of ovules and other tissues in the developing ovary (Girin et al., 2009). Whether this also applies to the reactivation of cell division in fruit tissues following pollination and fertilization is unclear. GRN10 was moderately enriched for RNA regulation (Table S4D). Since there is no organ patterning category using the Mapman bin ontology, we annotated the set of tomato ITAG2.4 gene models using the Trinotate pipeline (http://trinotate.sourceforge.net/) and associated GO terms with gene models (data not shown). Using these Trinotate GO terms, GRN10 was enriched for regulation of meristem structural organization, polarity specification of adaxial/abaxial axis, radial pattern formation, and asymmetric cell division. Overall, our data showed that putative organ patterning genes were strongly differentially regulated in SUN fruit. This might be directly causative to the change in fruit shape by modulating the expression of cell cycle-related genes. Thus, in addition to cell division and cell wall, SUN expression also leads to dramatic shifts in the expression of putative patterning-related genes.

The transcriptome data were also analyzed using linear factorial modeling approaches (van Leeuwen et al., 2007). Several GO term categories that were found in the GRNs were also identified using linear modeling. Importantly, the linear factorial modeling identified approximately 80% of the genes found in GRN10 (“cell wall” and “RNA regulation”) and GRN14 (most significantly enriched for “development”), validating the genes identified in the clusters (Table S6). These two GRNs were specifically higher expressed in SUN fruit in 4 dpa pericarp and 4 dpa columella, indicating significant genotype x time point interactions. However, the factorial approach did not identify GRN8, the most dramatically altered “cell division” cluster (Table S4). Since our experimental set up was not representative of a full factorial (e.g. 4 dpa columella and seed/placenta were combined), we were not able to evaluate the interaction of genotype x time point x tissue type for all the tissues collected. Therefore, genes that are
affected by *SUN* at only one level as is the case for GRN8 would not be identified using the linear model approach. Instead, to correct for testing errors in pairwise comparisons, we conducted a multiple testing correction on the eight comparisons on top of the multiple testing correction within each comparison and only selected the genes that passed both. With the understanding that few genes in the GRN clusters still represent false positives, the clustering based on the pairwise comparisons gave us excellent insights into GO term pathways that were most clearly affected by SUN in a dynamic interaction of genotype, time point, and tissue type.

**Comparison between metabolite and transcriptome profiling in developing tomato fruit**

Investigations into metabolite accumulation in SUN fruit showed that some TCA cycle metabolites were lower in columella tissues. To link these results with the transcriptome data, we found that GRN5 was enriched for genes encoding TCA cycle enzymes and included malate dehydrogenase, citrate synthase, and citrate lyase. In tomato, inhibition of 2-oxoglutarate dehydrogenase results in decreased respiration, increased accumulation of oxoglutarate and succinic acid, but no difference in fruit size or biomass (Araujo *et al.*, 2012). In contrast, inhibition of succinate dehydrogenase leads to increased photosynthate assimilation and fruit biomass (Araujo *et al.*, 2011). Our data suggest that these changes in TCA intermediates are a product of differential flux through the pathway possibly as an effect of increased cell division in the columella resulting in altered fruit shape (Wu *et al.*, 2011). Other links between metabolites and their biosynthesis pathways were not evident, suggesting that the pairwise comparisons were not sufficiently robust or that the overlap in pathways identified from gene expression and metabolites is generally low.

**Genes, metabolites, and hormones interact to regulate elongated fruit shape**

To evaluate the combined effect of SUN on gene expression, metabolite and hormone accumulation, we constructed a “gene-metabolite-hormone interactome map” (Fig. 6). Genes and metabolites were grouped into pathways to demonstrate how SUN-affected pathways interact during early fruit development. The size of each node represented how many edges attributed to the pathway, i.e. how many significant correlations each pathway exhibited with
other pathways (PCC > 0.90 and < -0.90). The most represented pathways were calcium regulation/transport, lipid metabolism, cell cycle and division, amino acid metabolism, auxin biosynthesis/transport/signaling, brassinosteroid metabolism, and ethylene biosynthesis/signaling. A zoomed in view at the network (Fig. 6B) revealed that processes involved in calcium regulation/transport form the main hub of connections with other pathways that are involved in regulating elongated fruit shape resulting from high expression of SUN. The calcium pathway was positively correlated with patterning, cell cycle/division, GRF-GIF gene
expression, auxin regulation/transport, and the hormones IAA and BL. TCA cycle genes, as well as TCA cycle metabolites, were negatively correlated with the gene pathways controlling cell division/patterning as well as auxin.

**Conclusion**

It is likely that SUN sets up the pattern for differential fruit growth before anthesis because ovary shape is slightly different at the time the flower opens (van der Knaap *et al.*, 2014; Wu *et al.*, 2011). Thus, the fruit tissues used in this study represent later growth stages when shape changes mediated by SUN are being executed. Inferring from the data generated in this study, we propose that the manner by which SUN regulates fruit shape is by impacting a calcium signaling cascade via its CaM domain-interacting motif to set up the pattern of organ growth. Based on the interaction of a SUN family member AtIQD1 (Burstenbinder *et al.*, 2013) with a kinesin motor protein, SUN may be involved in transport of cargo throughout the cell (van der Knaap *et al.*, 2014), possibly of cell wall components. This then leads to dramatic shifts in the spatio-temporal gene expression of many cell wall, cell division and patterning genes at the early stages of fruit development. Even though plant phenotypes associated with high SUN expression suggest a role for auxin in regulating fruit shape and plant morphology (Wu *et al.*, 2011; Xiao *et al.*, 2008), there may be no direct role for any of the known plant hormone in the regulation of fruit shape mediated by SUN.
Materials and Methods

Plant Materials

Nearly Isogenic Lines (NILs) were constructed by repeated backcrosses of recombinants harboring approximately 70 kb introgression of the sun locus from *S. pimpinellifolium* LA1589 (WT for the sun locus). Sun1642 harboring the mutation at the sun locus (SUN) was used as the recurrent parent and backcrosses were monitored using background markers. Thus, in principle, these lines only differ for the allele of the tomato SUN gene (Wu et al., 2011). The NILs (WT; SUN) were grown in the greenhouse in two gallon pots and 16 hours day light. Flowers at anthesis were tagged and self-pollinated on successive days. Pollination of flowers was staggered such that fruit of all developmental stages would be harvested on the same day. Fruits at 4, 7, and 10 days post anthesis (dpa) were harvested on ice and brought to the laboratory for further dissection. Columella, seed, and pericarp tissues were separated using a razor blade, immediately frozen in liquid nitrogen and stored at -80°C. Four dpa columella represents a mixture of columella, seed and placenta tissues. Four dpa pericarp represents a mixture of pericarp and exocarp tissue. Seven and 10 dpa pericarp samples were peeled and lack exocarp tissues. The seed tissues include the developing gel tissue surrounding the seed (placenta). Four harvest dates were carried out on five days, 2/25, 2/26, 2/27, 3/6, and 3/7 in 2009 for a total of 359 samples. A subset of 116 samples were sent for metabolite and hormone analysis representing a minimum of eight biological replicates per tissue type per genotype for 10 dpa, and four biological replicates per tissue type per genotype for 7 and 4 dpa. The remaining samples were used for RNA extraction (see below). We collected two additional replicates for RNA extraction in the summer of 2012. These tissues were collected as described above except that they were pooled for each replicate for each genotype, tissue type and timepoint, and aliquots were used for RNA extraction. The two replicates were collected over two successive days.

For the metabolite and hormone analyses, samples were freeze dried and shipped to the RIKEN CSRS institute. They were divided for the hormone and metabolite using 9 - 450 mg dry weight (DW) for the hormone (Growth Regulation Research Group) and 2.0 – 2.7 mg DW for the metabolite (Metabolomics Research Group) analyses. Metabolite extraction procedures (2.5 mg DW/ml extraction solution) and identification/annotation of metabolites using GC-MS was
done as described previously (Redestig et al., 2009; Schauer et al., 2005). Hormone analyses were conducted as described previously (Katsumata et al., 2011; Yamamoto et al., 2007).

RNA isolation, RNA-seq library preparation, sequencing

Frozen tissues were ground in liquid nitrogen and total RNA was extracted using the Trizol (Invitrogen Inc. USA) method as described by the manufacturer. RNA quantity and quality were assessed using a Qubit 2.0 fluorometer RNA Assay Kit (Invitrogen Inc. USA) and an Agilent 2100 Bioanalyzer RNA 6000 Nano kit (Agilent, USA) housed at the Molecular and Cellular Imaging Center on the OSU-Wooster campus. Strand-specific RNA-seq libraries with insert size of approximately 250 bp were prepared using the protocol described previously (Zhong et al., 2011) using 10 μg of total RNA. Eight libraries with compatible barcodes were pooled and run on a single lane on a flowcell of the Illumina HiSeq2000 at the Genomics Resources Core Facility at Weill Cornell Medical College (New York, NY). Single end reads of 51 bp were generated.

Alignment and analysis of Illumina reads

After sequencing, Illumina reads were quality checked, de-multiplexed, and trimmed. The reads were aligned to a known ribosomal RNA database using Bowtie (Langmead and Salzberg, 2012) allowing for 2 mismatches. The filtered reads with the ribosomal reads removed were aligned with TopHat2 (Kim et al., 2013) against the Solanum lycopersicum genome ITAG 2.4 allowing for maximum intron lengths of 5000 bp, segment lengths of 22 bp, and only 1 mismatch per segment. All other parameters were set to default. Aligned sequences were separated into sense and antisense, and counted using an in-house Python script. Of the four replicates collected from the 16 samples, all but one showed a quality score of above 95% with the average of 14.4 million reads per library that mapped to the genome (Table S1). The correlations between replicates were in general very high, in particular the samples that were collected from WT fruit (Table S2). Reads per kilobase of exon model per million mapped reads (RPKM) were calculated by the formula RPKM_{ij} = (10^9 * C)/(N * L) where C is the reads mapped to gene i in sample j, N is the total number of reads mapped (taken as the sum of reads mapped to annotated genes in sample j), and L is the number 1000 bp of exon in the gene model of gene i. The reads are archived at the Short Read Archive (SRA) at NCBI under accession number SRA065144.

Differential gene expression analysis
Analysis of differentially expressed genes between SUN and WT fruit was conducted using the DESeq package (Anders et al., 2013). Raw counts data were normalized by library size and fit to a negative binomial model. We used the fit-only option of the SharingMode function, using the method of estimating the dispersions. Briefly, we assumed that the variation in dispersions from the predicted fit line is due to the sampling variance and not due to the true endemic variant expression of the gene. P values were adjusted using the Benjamini-Hochberg correction for multiple testing. All four biological replicates were used in the analysis and differentially expressed genes with an adjusted p < 0.05 were considered significant.

Differential expression analysis using linear factorial modelling

Using DESeq2 and the function nbinomLRT(), we tested three separate null hypotheses. Null hypothesis 1 tested whether each gene was significantly affected by genotype. Null hypothesis 2 tested whether each gene was affected by the interaction of genotype by time point. Because including time points and tissue types together in the model results in a model that is not full rank, for null hypothesis 2 we only tested genotype, time point, and the genotype x time point interaction. Null hypothesis 3 tested whether each gene was affected by the interaction of genotype by tissue type.

Clustering of the expression of genes during fruit development using fuzzy C-means

The expression dataset was filtered by excluding genes that were low expressed. This was done by calculating the RPKM values for all genes as described above and filtering using the criterion of at least four replicates being expressed above 2 RPKM. This filtering assures that in at least one tissue type, time point and genotype, we are confident that the gene is expressed. The filtering for low expression resulted in a reduced dataset to 19,963 genes. Gene expression values expressed in RPKM were mean centered and converted to their Z score which was calculated by dividing each mean-centered value by the standard deviation of all values for each gene. This calculation can be illustrated by the following formula:

\[ M_{ij} = \frac{m_{ij} - \mu_i}{\sigma_i} \]

Where \( M_{ij} \) is the normalized value for gene i in sample j, \( m_{ij} \) is the non-normalized value of gene i in sample j, \( \mu_i \) is the mean of gene i in all samples, and \( \sigma_i \) is the standard deviation of the gene i in all samples.
Expression profiles for WT fruit were clustered using fuzzy C-means by using the Mfuzz package (Futschik and Carlisle, 2005) in R with a C value of 100 to maximize dynamic differential clustering identification, and core clustered at 0.70 membership probability. Dynamically regulated gene clusters were identified by inspecting plots of the normalized expression profiles of each cluster which led to the identification of ten clusters that showed a dynamic range of expression in one tissue over another, as well as six ontogeny and spatially specific clusters. Based on genes found predominantly in each cluster and hypergeometric enrichment of Mapman ontology bins, the clusters were named accordingly: cell division, cell wall, RNA, secondary metabolism, transport/hormone, transport/signaling, photosynthesis, protein.

**Expression networks of differentially expressed genes using fold change SUN/WT**

For the network analysis, we filtered the dataset further by only including those that showed significant differential expression in at least one pairwise comparison using DESeq. All genes that were significantly differentially expressed (BH adjusted p < 0.05) were chosen (total of 5,623). An additional multiple testing correction using Bonferroni was performed to account for the 8 pairwise comparisons. The final list consisted of 3,154 non-redundant genes of which 1,455 were differentially expressed in more than one comparison. Any gene that was filtered out in a particular time point and tissue type because of low expression was set to 0 for that condition. Log-transformed fold-changes of SUN relative to WT for each gene were estimated using DESeq in R (Anders et al., 2013). Briefly, the means for each gene at a particular time point, tissue type and genotype were estimated based on all four biological replicates and the estimated dispersions were based on the sample dispersions. These estimated means were used to calculate the fold-change SUN/WT. A log₂ fold of 0 means no difference, a log₂ fold of 1 means 2-fold higher, a log₂ fold of 2 means 4-fold or higher and so on. Fold-change for each sample was used as differential expression profiles in order to include the information of differences in expression in SUN fruit in the same analysis. These profiles were clustered using fuzzy C-means with a C of 60 using the Mfuzz package in R (Futschik and Carlisle, 2005). We determined that a C of 60 properly incorporated all patterns in the data without forming spurious or random clustering. The results showed patterns that changed slightly but were still unique (differences below 1 log₂). The 14 clusters that showed the most dynamic response (i.e. clusters
with log₂ values above 1 and below -1) were chosen for further analysis (Fig. S1) and were grouped together based on the median profile of their members.

**Hypergeometric enrichment test of gene clusters**

Enrichment of Mapman ontology bin codes represented in each gene cluster was done using a hypergeometric test using the webserver, GeneProf (http://www.geneprof.org/GeneProf/tools/hypergeometric.jsp) and phyper() in R. P-values were adjusted for multiple testing by Benjamini-Hochberg correction in R (FDR < 5%)

**Promoter motif searches**

We selected four sets of genes from the identified GRNs above; GRN3 (calcium signaling/transport), GRN8 (cell cycle), GRN9 (lipid/hormone), and GRN10 (RNA/cell wall). Clusters were named based on hypergeometric enrichment tests and observations of putative orthologs present in each cluster. For example, GRN10 is enriched for RNA regulation transcription factors and contains a group of TFs that regulate patterns of cell proliferation. GRN clusters were further filtered for outliers by visually inspecting the log₂ SUN/WT profile. Filtering resulting in 5% reduction in GRN3 (54 remaining), 27.6% for GRN8 (118 remaining), 27.8% for GRN9 (57 remaining), and 25% for GRN10 (133 remaining).

To build the promoter region datasets we wrote a Python script which, using functionality from Biopython package (Cock *et al.*, 2009), accessed the tomato genome and extracted 1,000 bases upstream of each gene. For each of the four datasets, our script generated a text file with the corresponding promoter regions. These files were further used for identifying overrepresented motifs as described below.

For each of the four promoter datasets, an exhaustive search of all potential 5- to 8- base pair motifs was designed and implemented (Blackwood *et al.*, 2013). This computationally demanding step was conducted in Hadoop, a platform for big data analysis which supports distributed processing of large datasets. We wrote Python scripts, including mapper and reducer scripts for Hadoop, which we ran on the Cloudera Hadoop (www.cloudera.com) on Oracle VM VirtualBox for Mac OS. For each of the four datasets, these scripts computed the frequency of each possible motif of lengths 5, 6, 7, and 8. To investigate whether the motifs were occurring with similar frequencies in the randomly rearranged promoter sequences, the promoter regions
were shuffled 100 times and the averaged frequencies were computed from the shuffled
promoters. The ratio of the frequency in the actual promoter region over the shuffled one was
returned as the enrichment score. An enrichment score of 1 indicates that the corresponding
motif is found at the same rate in the original and the shuffled promoter. Therefore, the larger the
enrichment score, the more likely the corresponding motif is a true motif and not due to chance.
Further, only putative motifs with an enrichment score of 3 or larger were recorded. This
threshold ensured that only motifs occurring at least three times more frequently in the original
promoter region were selected for the next step of the analysis. In the next step unlikely motifs
such as microsatellite (e.g., ACACAC) or homopolymer runs (e.g., AAAAAA) were removed
from the set of results. For the reasons explained next, the results obtained for motifs of lengths
5, 7, and 8 were not as good as the ones for 6-bases long motifs. Many of the 5-bases long motifs
were found with similar frequencies in both the original and the shuffled promoter regions; in
addition, many 5-bases long motifs with high enrichment scores were repeats. Few motifs of
length 7 and 8 were identified, and these motifs were combinations of two or more motifs of
length 6. Therefore, to include all potential motifs, for each of the four datasets we continued the
analysis with the results of the 6-bases long motifs. In the next stage of our analysis, promoter
motifs of length 6 that shared a common core element (e.g., 4 out of 6 bases were conserved)
were grouped together into a consensus motif.

To further evaluate the likelihood of a consensus motif to be enriched in one cluster over
another, the enrichment scores (ratio of actual over shuffled frequency) were computed for each
consensus motif in all four datasets (Blackwood et al, 2013). More precisely, for each consensus
motif S found in a particular dataset \(D_i\), \(i=1,\ldots,4\), we first computed its relative frequency in that
dataset by using eq. (1), where S denotes a consensus motif, \(|S|\) denotes its number of (motif)
components, \(c\) denotes a (motif) component, and \(f_{ci}^{D_i}\) denotes the relative frequency of a
component c in the data set \(D_i\).

\[
f^{D_i}(S) = \sum_{c=1}^{\text{\(|S|\)}} f_{ci}^{D_i}, \quad i = 1, \ldots, 4
\]

(1)

Similarly, the frequency of a consensus motif S in the shuffled promoter region is computed by
eq (2).
Finally, the enrichment score for a consensus motif $S$ is calculated as the ratio of the two frequencies computed above:

$$\text{score}^D_i(s) = \frac{f^D_i(s)}{f^D_i(s)}$$

An enrichment score close to 1 indicates that the corresponding motif is equally represented in the original and the shuffled promoter regions, and therefore, not likely to be a true motif for a given dataset. An enrichment score of 2 or larger implies that the corresponding consensus motif is found more frequently in the original than in the shuffled dataset, and therefore such motifs are selected for further analysis.

Investigation of gene expression of those containing CCAACGGYYA, CAYRTGG, TGGACCA, and AGGTSATG consensus motifs

Expression profiles from genes containing the motifs were extracted from SUN and WT fruit separately and principal components analysis was carried out using the prcomp() function in R. Loadings for PC2 were extracted for SUN and WT expression profiles and plotted using ggplot() in R (Table S12). Gene sets included 58 for CCAACGGYA motif containing set, 31 for CAYRTGG motif containing set, 24 for TGGACCA motif containing set, and 28 for AGGTSATG motif containing set.

Metabolite pair-wise comparisons

To compare levels of each metabolite in tissue types and time points we performed pair-wise t-tests between SUN and WT fruit. A relaxed significance threshold was used of $p<0.10$ to highlight patterns of change in SUN fruit. Average fold change of SUN/WT was used to construct a primary metabolite pathway heatmap that was constructed based on (Do et al., 2010) using pair-wise comparisons with $p < 0.10$.

Metabolite and hormone data analysis

To compare metabolite and hormone accumulation while taking into account large

$$f^D_i(s) = \sum_{c=1}^{4} f^D_i(s)$$
differences in magnitude and variance, the values for each were first mean centered by subtracting the mean value for a metabolite across all samples from the value for each sample. The data was further normalized for unequal variance by converting it into its Z score as described above.

**Metabolite and hormone correlation network comparison between genotypes**

In order to compare metabolite and hormone networks between WT and SUN fruit, the values were re-clustered separately and networks for only SUN were inferred. Basically, clustering was first constructed with fruit samples collected from SUN plants. These networks were then evaluated in wild type fruit for significant correlations. New correlations in SUN fruit (adj p < 0.05) that were not significant in wild type fruit were considered as new associations in SUN. Correlations that were significant in WT fruit (adj p < 0.05) and not significant in SUN fruit were considered abolished associations in SUN fruit. All p-values were adjusted for multiple testing by Benjamini Hochberg correction (FDR < 5%).

**Interactome network analysis of genes, metabolites and hormones**

Hormone data was not available for 7 dpa and therefore the genes-metabolite-hormone interactome network was generated using 4 dpa and 10 dpa data (five conditions). Hormone values were normalized to their Z score as described above. Normalized values from hormone and metabolite accumulation, and gene expression were used as described previously only using SUN fruit. Since less conditions were used to calculate correlations, a higher correlation (< -0.90 and > 0.90) was determined to be significant after multiple testing correction (adj p < 0.05). To assess how different pathways were interacting, genes were grouped into putative pathways based on closest Arabidopsis ortholog and S. lycopersicum genome annotation ITAG2.4 (www.solgenomics.net), metabolites were grouped into biochemical categories, and hormones were taken as hormone class. These identifiers were then used as the gene, metabolite, or hormone name when calculating correlations. When constructing the network, nodes represented a pathway, metabolite class, or hormone. The number of edges connecting the nodes was used to define the size of each node. The size of the nodes is indicative of the predominance of each pathway, metabolite category, or hormone in its interactions in SUN fruit and how each pathway is interacting with each other.

**Visual network creation**
Networks were constructed using Cytoscape software (Cline et al., 2007). Nodes are genes, metabolites and hormones, and edges are the significant Pearson Correlation Coefficients between the nodes. The layout of the nodes is based on the strength of the edge, and the proximity of two nodes is based on the strength of the correlation between the nodes.

**Graphics and statistical analysis**

Pearson’s Correlation Coefficients were calculated using R and corrected for 5% FDR using the Benjamini-Hochberg correction for multiple testing. All PCA plots and jigger box plots were created in R using the ggplot2() package.

**Acknowledgements**

We thank Meghan Fisher and Jenny Moyseenko for greenhouse assistance and plant care, Ms. Yumiko Takebayashi and Mr. Makoto Kobayshi for their technical support of hormone and GC-MS analysis, respectively.
Figure Legends

Figure 1. Tomato fruit growth following pollination. (A) wild type fruit at anthesis, 4 dpa, 7 dpa, 10 dpa and 16 dpa. (B) SUN fruit at anthesis, 4 dpa, 7 dpa, 10 dpa and 16 dpa. Size bar represents 1 cm. (C) Expression of SUN in WT and SUN fruit tissues and developmental time points. Per, pericarp; col, columella; se, seed and placenta.

Figure 2. Dynamically expressed and co-regulated genes during early fruit development in WT tomato. (A) 604 genes represented by cell cycle and cell division related processes. (B) 188 genes represented by cell wall related processes. (C) 224 genes represented by RNA and transcription related processes. (D) 396 genes represented to secondary metabolism related processes. (E) 353 genes related to transport and hormone related processes. (F) 224 genes represented by transport and signaling related processes. (G) 201 genes represented by photosynthesis related processes. (H) 284 genes represented by protein and translation related processes. Gene expression is on a relative mean-centered scale showing median expression level for all members in the respective cluster in SUN and WT fruit. Error bars indicate standard error for the genes in the respective cluster. per.4, pericarp 4 dpa; per.7, pericarp 7 dpa; per.10, pericarp 10 dpa; se.7, seed 7 dpa; se.10, seed 10 dpa; col.4, columella 4 dpa; col.7, columella 7 dpa; col.10, columella 10 dpa.

Figure 3. GRN expression dynamics in developing fruit. (A) PCA of all differentially expressed genes using the log2 SUN/WT expression over all tissue types and time points sampled. (B) Selected GRNs in PCA space. (C – G) Log2 SUN/WT for selected GRN per tissue type and time point. (C) GRN2 (D) GRN3 (E) GRN8 (F) GRN9 (G) GRN10/GRN14. per.4, pericarp 4 dpa; per.7, pericarp 7 dpa; per.10, pericarp 10 dpa; se.7, seed 7 dpa; se.10, seed 10 dpa; col.4, columella 4 dpa; col.7, columella 7 dpa; col.10, columella 10 dpa.

Figure 4. Principal components analysis of the expression profile of genes with detected motifs in the promoter region. (A) 58 genes in the cell-cycle cluster with the CCAACGGYYYA consensus motif (GRN8). (B) 24 genes in the RNA and cell wall cluster with the TGGACCA consensus motif (GRN10). (C) A group of 45 random genes. (D) 31 genes in the lipid and hormone cluster with the CAYRTGG consensus motif (GRN9). (E) 28 genes in the calcium
signaling and transport cluster with the AGGTSATG consensus motif (GRN3). (F) A second set of 45 random genes.

**Figure 5.** Metabolite changes in WT and SUN tomato fruit. Pair-wise comparisons of metabolite accumulation between SUN and WT fruit in the pericarp tissue 7 and 10 dpa, and septum tissue 4, 7, and 10 dpa. Highlighted boxes reflect a p < 0.05 using student’s T-test. Color legend represents fold change of SUN/WT fruit. Positive values indicate higher in SUN fruit. Negative values indicate higher in WT fruit.

**Figure 6.** Gene, metabolite, and hormone interactome in SUN fruit. The size of the nodes represents the number of edges to that node. Blue edges represent significant positive correlations. Red edges represent significant negative correlations. Nodes represent genes, metabolites, or hormones that participate in the pathway indicated. Squares represent genes. Diamonds represent metabolites. Hormones are represented as circles. Cell div indicates genes involved in cell division. GRF_GIF indicates genes in the GRF family and GIF. Cell pattern indicates transcription factors that regulate organ polarity and patterning. IQD are members of the SUN family. Calcium indicates genes involved in calcium transport, calmodulin binding, and proteins requiring calcium for function. Lipid indicates genes involved in lipid biosynthesis and regulation. BR indicates genes involved in brassinosteroid homeostasis. TCA indicates enzymes that participate in the TCA cycle. (A) Full interactome network with bundled edges. (B) Subnetwork of calcium hub.
Table I. Promoter enrichment scores for motifs found in four GRNs.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GRN8 motifs</th>
<th>GRN3 motifs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRRCCGT</td>
<td>GCCCAA</td>
</tr>
<tr>
<td>GRN8 cell cycle cluster</td>
<td>6.7</td>
<td>4.7</td>
</tr>
<tr>
<td>GRN3 calcium signaling/transport cluster</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GRN9 lipid/hormone cluster</td>
<td>1.2</td>
<td>2</td>
</tr>
<tr>
<td>GRN10 RNA/cell wall cluster</td>
<td>0.5</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GRN9 motifs</th>
<th>GRN10 motif</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAGAGGTG</td>
<td>CAYRTGG</td>
</tr>
<tr>
<td>GRN8 cell cycle cluster</td>
<td>2</td>
<td>1.4</td>
</tr>
<tr>
<td>GRN3 calcium signaling/transport cluster</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>GRN9 lipid/hormone cluster</td>
<td>6.5</td>
<td>5.4</td>
</tr>
<tr>
<td>GRN10 RNA/cell wall cluster</td>
<td>2.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

K corresponds to G or T; R corresponds to A or G; Y corresponds to C or T; S corresponds to C or G; and W corresponds to A or T.
Supplemental files:

Figure S1. The log2 fold dynamics of 14 GRNs in 8 fruit tissues and developmental time points.

Figure S2. Co-differential expression of genes in plant growth pathways.

Figure S3. Co-differential expression of the putative orthologs of patterning genes.

Figure S4. Significant differences in metabolite accumulation in seed tissue at 10 dpa.

Figure S5. Relative metabolite accumulation in WT and SUN fruit tissues.

Figure S6. New metabolite network of associations in SUN fruit.

Figure S7. Metabolite network of correlations in WT fruit.

Figure S8. Subnetworks of metabolite correlations in WT fruit.

Figure S9. Correlations of hormone levels in SUN and WT fruit.

Table S1. Summary statistics for each of the RNA seq libraries.

Table S2. Correlations among samples in fractions.

Table S3. Genes in eight co-regulated clusters in WT fruit.

Table S4. Hypergeometric GO term enrichments.

Table S5. RPKM values and pairwise differential expression of all genes.

Table S6. Genes found in the linear factorial modeling analyses.

Table S7. Membership of genes in the 14 GRNs.

Table S8. Promoter motifs enriched in GRN3, 8, 9 and 10.

Table S9. New metabolite correlations in SUN fruit.

Table S10. Abolished metabolite correlations in SUN fruit.

Table S11. Hormone levels in the different fruit tissues.
Table S12. Loadings for all PCA analyses.
Figure S1. The log2 fold dynamics of 14 GRNs in 8 fruit tissues and developmental time points. The black line shows the average log2 fold dynamics for the GRN.
Figure S2. Co-differential expression of genes in plant growth pathways. (A) Putative orthologs of brassinosteroid biosynthesis genes: FACKEL – solyc09g009040.2.1, HYDRA1 - solyc06g082980.2.1, DET2 - solyc11g006300.1.1, solyc10g086500.1.1, SMO1 - solyc01g091320.2.1, SMO2 - solyc06g005750.2.1, STE1 - solyc02g086180.2.1, DWF5 - solyc06g074090.2.1, DWF1 - solyc02g069490.2.1. (B) Putative orthologs of auxin biosynthesis, transport, and conjugation cluster: PIN1 - solyc03g118740.2.1, solyc09g014380.1.1, AUX1 - solyc09g014380.2.1, YUC4 - solyc06g065630.2.1, JAR1 - solyc05g050280.2.1, solyc10g009640.1.1, solyc10g011660.2.1. (C) Calcium cluster: Calcium binding - solyc02g079520.1.1, IQD18 (SLSUN25) - solyc08g083240.2.1, CAX7 - solyc09g072690.1.1, CAX3 - solyc06g006110.2.1, CNGC4 - solyc10g006800.2.1, solyc12g005400.1.1, Calmodulin-binding - solyc03g118810.1.1, CRK - solyc01g108400.2.1. (D) Putative orthologs of genes in the GRF and cell cycle pathway: GIF1 - solyc11g006230.1.1, solyc04g009820.2.1, GRF1 - solyc04g077510.2.1, Cyclin - solyc09g065200.2.1, solyc06g065680.2.1, solyc11g010460.1.1, solyc01g009040.2.1, solyc11g005090.1.1, solyc10g080950.1.1, solyc02g082820.2.1, solyc03g032190.2.1, solyc06g043150.2.1, solyc04g082430.2.1, solyc10g078330.1.1, solyc01g099270.2.1. When more than one gene is shown, for example in panel D for the cyclin genes, the expression pattern shown is the median pattern of the genes listed.
Figure S3. Co-differential expression of the putative orthologs of patterning genes. (A) Putative orthologs of KANADI2 and PIN1: KANADI2 - solyc08g005260.1.1, solyc08g076400.2.1, PIN1 - solyc03g118740.2.1, solyc10g080880.1.1, solyc10g078370.1.1. (B) Putative orthologs of JACKDAW, MAGPIE and SHORT ROOT: SHORT ROOT - solyc02g092370.1.1, JACKDAW - solyc10g084180.1.1, MAGPIE - solyc08g063040.2.1, solyc04g080130.2.1. (C) Putative orthologs of KANADI2 ASYMETRIC LEAVES 1: AS1 - solyc09g010840.1.1, KANADI - solyc08g005260.1.1, solyc08g076400.2.1. (D) Putative orthologs of APETALA2 and REPLUMLESS: APETALA2 - solyc03g044300.2.1, solyc10g086640.1.1, REPLUMLESS - solyc09g011380.2.1. When more than one gene is shown, for example log RPL in panel D, the expression pattern shown is the median pattern of the genes listed.
**Figure S4.** Significant differences in metabolite accumulation in seed tissue at 10 dpa. Values are mean-centered relative accumulation. P-values calculated by student’s T-test.
**Figure S5.** Relative metabolite accumulation in WT and SUN fruit tissues indicated below the graphs. Values are the mean centered Z-scores relative to genotype where 0 represents the average accumulation across all tissues.
Relative Accumulation

- Cysteine
- Leucine
- Linoleic acid
- Lysine
- Malic acid
- Mannose
- Methionine
- Myo-inositol
- Oleate
- Ornithine
- Oxalic acid
- Oxoglutaric acid

Genotype

SUN
WT
Figure S6. New metabolite network of associations in SUN fruit. Blue edges represent significant positive correlations. Red edges represent significant negative correlations (adjusted p < 0.05). Red square nodes represent organic acids, red octagons represent TCA cycle components. Orange nodes represent amino acids. Blue nodes represent sugars/sugar alcohols. Yellow nodes represent fatty acids. Green nodes represent “other”.
Figure S7. Metabolite network of correlations in WT fruit. These networks represent abolished metabolite correlations in SUN fruit. Blue edges represent significant positive correlations. Red edges represent significant negative correlations (adjusted p < 0.05). Red nodes represent organic acids and TCA cycle components. Orange nodes represent amino acids. Blue nodes represent sugars/sugar alcohols. Yellow nodes represent fatty acids. Green nodes represent “other”.
Figure S8. Subnetworks of metabolite correlations in WT fruit. These subnetworks represent abolished correlations in SUN fruit. Blue edges represent significant positive correlations. Red edges represent significant negative correlations (adjusted p < 0.05). (A) Inositol-1-phosphate subnetwork. (B) Oxalic acid subnetwork. (C) Caffeoyl quinic acid subnetwork. (D) Linoleic acid subnetwork. (E) Citric acid subnetwork. Layout reflects the strength of the edge, i.e. the stronger the correlation, the shorter the edge length. Red nodes represent organic acids and TCA cycle components. Orange nodes represent amino acids. Blue nodes represent sugars/sugar alcohols. Yellow nodes represent fatty acids. Green nodes represent “other”.

Downloaded from on December 30, 2017 - Published by Copyright © 2015 American Society of Plant Biologists
Figure S9. Correlations of hormone levels in SUN and WT fruit. (A) Network of hormone correlations (> 0.80, < -0.80) in WT fruit. (B) Network of hormone correlations (> 0.80, < -0.80) in SUN fruit. Solid edges reflect positive correlations. Dotted edges reflect negative correlations. Edges are Pearson’s Correlation Coefficient of normalized hormone levels in 4 and 10 dpa fruits. (C) Correlation heatmap and hierarchical clustering of hormones in WT fruit. (D) Correlation heatmap and hierarchical clustering of hormones in SUN fruit. IAA: indole-3-acetic acid, ABA: Abscisic Acid, DHZ: dihydrozeatin, tZ: trans-zeatin, iP: isopentenyl adenine, JA-ile: Jasmonic acid-isooleucine.


