Short Title: Genetic Architecture of Gene Expression in Tomato

Corresponding author:

Neelima R. Sinha

Department of Plant Biology

1002 Life Sciences

One Shields Ave.

Phone: (530) 754-8441

Fax: (530) 752-5410

e-mail: nrsinha@ucdavis.edu

Research area: **Genes, Development and Evolution**

Title: eQTL in a Precisely Defined Tomato Introgression Population Reveal Genetic Regulation of Gene Expression Patterns Related to Physiological and Developmental Pathways

Aashish Ranjan[#a¶], Jessica M. Budke[¶], Steven D. Rowland[¶], Daniel H. Chitwood[#b], Ravi Kumar[#c], Leonela Carriedo, Yasunori Ichihashi[#d], Kristina Zumstein, Julin N. Maloof, and Neelima R. Sinha

Department of Plant Biology, University of California at Davis, Davis, California, United States of America

[¶] These authors contributed equally to this work.

Summary: Genetical genomics approach in tomato identified genetic hotspots that regulate gene expression patterns relating to diverse biological processes such as plant development, photosynthesis and defense.

Footnotes:

Funding information: This work is supported through a National Science Foundation grant (IOS-0820854) awarded to NRS and JNM. DHC was a fellow of the Life Sciences Research Foundation funded through the Gordon and Betty Moore Foundation. JMB is a recipient of Katherine Esau Postdoctoral Fellowship at UC Davis.

Present Address:

[#a] : National Institute of Plant Genome Research, New Delhi, India.

[#b] : Donald Danforth Plant Science Center, St. Louis, Missouri, United States of America.

[#c] : Novozymes, Davis, California, United States of America.

[#d] : RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, Japan.

Corresponding author email: nrsinha@ucdavis.edu

Abstract

Variation in gene expression, in addition to sequence polymorphisms, is known to influence developmental, physiological and metabolic traits in plants. Genetical genomics approaches on genetic mapping populations have facilitated the identification of expression Quantitative Trait Loci (eQTL), the genetic determinants of variation in gene expression patterns. We used an introgression population developed from the wild desert-adapted *Solanum pennellii* and domesticated tomato *Solanum lycopersicum* to identify the genetic basis of transcript level variation. We established the effect of each introgression on the transcriptome through differential gene expression analysis, and identified ~7,200 eQTL regulating the expression of 5,300 genes. Barnes-Hut *t*-distributed stochastic neighbor embedding clustering identified 42 modules revealing novel associations between gene expression patterns and biological processes. The results showed a complex genetic architecture of global gene expression pattern in tomato. Several genetic hotspots regulating a large number of gene expression patterns relating to diverse biological processes such as plant defense and photosynthesis were identified. We identified important eQTL regulating gene expression patterns related to leaf number and complexity, and hypocotyl length. Genes associated with leaf development showed an inverse correlation with photosynthetic gene expression but their regulation was dispersed across the genome. This is the first comprehensive insight into the global regulation of transcript abundance in tomato and its influence on plant

phenotypes, which sets the stage for identifying gene/s underlying these regulatory loci.

Introduction

The genetic basis of many qualitative and quantitative phenotypic differences in plants has been associated with sequence polymorphisms and the corresponding changes in gene function. However, differences in the levels of gene expression, without underlying changes in coding sequences, also significantly influence plant phenotypes. Closely related plant species often have little coding sequence divergence, nonetheless they often develop unique physiological, metabolic, and developmental characteristics indicating that patterns of gene expression are important in species-level phenotypic variation (Kliebenstein, 2009; Koenig et al., 2013). Phenotypic differences attributed to variations in gene expression patterns have been found to influence disease resistance, insect resistance, phosphate sensing, flowering time, circadian rhythm, and plant development (Kroymann et al., 2003; Werner et al., 2005; Clark et al., 2006; Zhang et al., 2006; Svistoonoff et al., 2007; Chen et al., 2010; Hammond et al., 2011).

Global gene expression changes across defined genetic backgrounds have been used to identify expression Quantitative Trait Loci (eQTL) through genetical genomics approaches (Jansen and Nap, 2001; Kliebenstein, 2009; Druka et al., 2010; Chitwood and Sinha, 2013). An eQTL is a chromosomal

region that drives variation in gene expression patterns (i.e., transcript abundance) between individuals of a genetic mapping population and can be treated as a heritable quantitative trait (Brem et al., 2002; Kliebenstein, 2009; Cubillos et al., 2012). Recent advances in next-generation sequencing have enabled high-throughput genotyping and transcript abundance estimation to provide direct readouts of the regulatory changes in mapping populations, allowing identification of thousands of eQTL in a single experiment (Pickrell et al., 2010; Chitwood and Sinha, 2013; Battle et al., 2015). Global eQTL studies in animals, fungi, and plants have provided genetic insights into the transcriptional regulation of these organisms (Brem and Kruglyak, 2005; Keurentjes et al., 2007; West et al., 2007; Schadt et al., 2008; Hammond et al., 2011; Holloway et al., 2011; Zhang et al., 2011; Cubillos et al., 2012).

Depending upon the proximity to the gene being regulated, eQTL can be classified into two groups: *cis*-eQTL when the physical location of an eQTL coincides with the location of the regulated gene, and *trans*-eQTL when an eQTL is located at a different position from the gene being regulated (Kliebenstein, 2009). eQTL studies with the model plant Arabidopsis underlined the significance and contributions of *cis*- and *trans*-eQTL (DeCook et al., 2006; West et al., 2007; Holloway and Li, 2010). *Cis*-eQTL have a significant effect on local expression levels, whereas *trans*-eQTL often have global influences on gene regulation. Individual genes can have multiple *trans*-eQTL contributing to the regulation of their expression pattern. Identification of *trans*-acting eQTL hotspots, defined as regions affecting the expression of a significantly larger number of genes than

expected, have been found in many eQTL studies. Genes underlying an eQTL hotspot could be master transcription factors controlling the expression of a suite of genes that act in the same biological process or pathway. For example, eQTL hotspots in Arabidopsis co-locate with the *ERECTA* locus, which has been shown to pleiotropically influence many traits including those regulating morphology (Keurentjes et al., 2007). In addition, the eQTL identified using pathogen challenged tissues in barley were enriched for genes related to pathogen response (Chen et al., 2010; Druka et al., 2010). Similarly the rice *sub1* locus, which regulates submergence tolerance, controls the activity of an ethylene response factor with significant *trans* effects (Fukao et al., 2006; Xu et al., 2006). Thus eQTL analyses have the potential to reveal a genome-wide view of the complex genetic architecture of gene expression regulation, the underlying gene regulatory networks, and may also identify master transcriptional regulators.

Cultivated tomatoes, along with their wild-relatives, harbor broad genetic diversity and large phenotypic variability (Moyle, 2008; Ranjan et al., 2012). Wide interspecific crosses bring together divergent genomes and hybridization of such diverse genotytpes leads to extensive gene expression alterations compared to either parent. Gene expression in hybrids often deviates from purely additive effects and may contribute to hybrid vigor and in nature can act as a source of variation for selection to act upon (Hegarty et al., 2008). Introgression lines (ILs), developed by crosses between wild-relatives and the cultivated tomato to bring discrete wild-relative genomic segments into the cultivated background, have proved to be a useful genetic resource for genomics and

molecular breeding studies. One such highly characterized introgression population was developed from the wild desert-adapted species *Solanum pennellii* and domesticated *Solanum lycopersicum* cv. M82 (Eshed and Zamir, 1995; Liu and Zamir, 1999). These two species exhibit large phenotypic differences for both developmental and metabolic traits. This IL population has been successfully used to map numerous QTL for metabolites, enzymatic activity, yield, fitness traits, and developmental features, such as leaf shape, size, and complexity (Frary et al., 2000; Holtan and Hake, 2003; Fridman et al., 2004; Chitwood et al., 2013; Muir et al., 2014). Though these studies led to fine mapping of candidate genes underlying identified QTL, phenotypic QTL mapping does not provide insight into the molecular regulation and gene networks underlying a trait. Comparative transcriptomics using cultivated and wild species enables identification of transcript abundance variation potentially underlying trait differences, such as salt tolerance, between species (Koenig et al., 2013). However, the genetic regulators of these transcriptional differences between the species still need to be elucidated. Therefore, we used a genetical genomics approach to identify the genetic basis of transcript level variation in tomato using the *S. pennellii* introgression lines.

Using both genomic DNA and RNAseq reads we recently pinpointed the exact gene content harbored in the introgression regions for the *S. pennellii* ILs and identified more than a thousand QTL underlying plant developmental traits, including leaf traits (Chitwood et al., 2013). Here we report on a comprehensive transcriptome profile of the ILs, a comparison between the gene expression

patterns of the ILs and the cultivated M82 background (differential gene expression – DE), as well as a global eQTL analysis to identify patterns of genetic regulation of transcript abundance in the tomato shoot apex. We have identified more than 7,200 *cis*- and *trans*-eQTL in total, which regulate the expression patterns of 5,300 genes in tomato. Additional analyses using Barnes-Hut *t*-distributed stochastic neighbor embedding (BH-SNE) (van der Maaten, 2013) identified 42 modules revealing novel associations between gene expression patterns and biological processes. The gene expression patterns under strong genetic regulation are related to plant defense, photosynthesis, and plant developmental traits. We also report important eQTL regulating gene expression pattern associated with leaf number, complexity, and hypocotyl length phenotypes.

Results and Discussion

Transcriptome Profiling of Introgression Lines

RNAseq reads obtained from the tomato shoot apex with developing leaves and hypocotyl were used to identify differentially expressed (DE) genes between each *S. pennellii* IL and the cultivated M82 (Supplemental Dataset 1). The total number of genes differentially expressed for each IL both in *cis* (in this population reflecting “local” level regulation either from within a gene itself or other genes in the introgression) and *trans*, along with the number of genes in the introgression regions, are presented in Figure 1 and Supplemental Table I. There was a strong correlation between the number of genes in the introgression

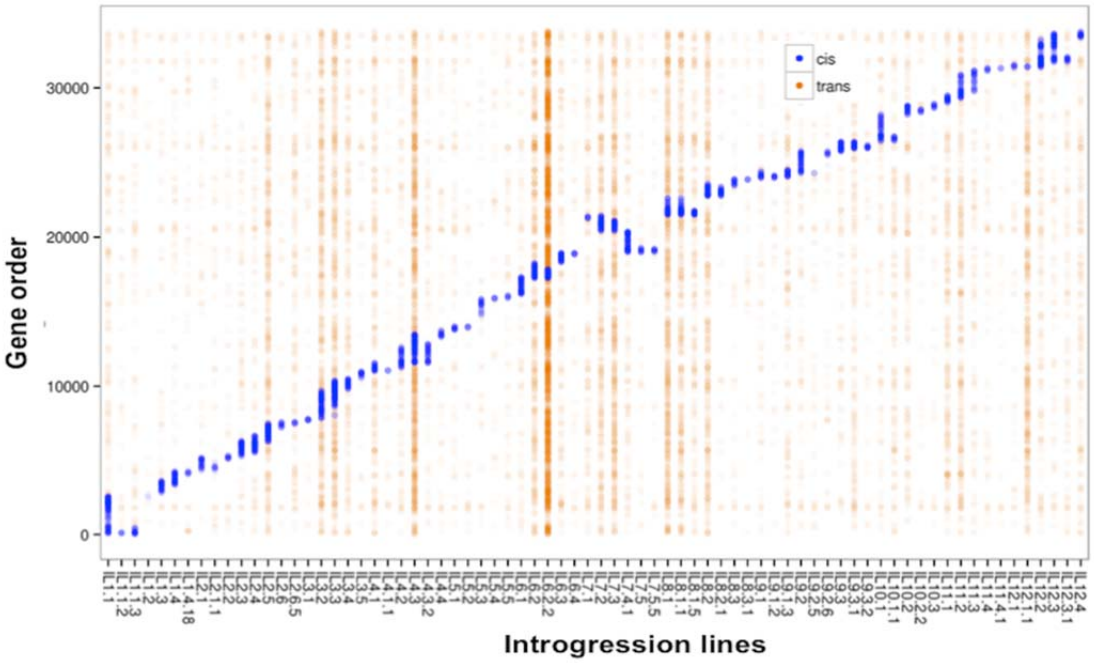

Figure 1. Transcriptome profile of the tomato introgression lines. Differentially expressed genes for the ILs compared to cultivated parent M82. Y-axis shows all the tomato genes starting from the first gene on chromosome 1 to the last gene on chromosome 12, and X-axis depicts the individual ILs. Genes differentially expressed within the introgression regions (in *cis*) are shown as blue points and differentially expressed genes in *trans* (outside) the introgression region are shown as orange points.

regions and the number of DE genes in *cis* (Supplemental Figure S1A). In contrast, the number of DE genes in *trans* was poorly correlated with introgression size (Supplemental Figure S1B). ILs showing higher total number of DE genes, such as IL6.2.2 and IL4.3, showed regulation of most of the DE genes in *trans* (Figure 1, Supplemental Table I, Supplemental Figure S2). IL12.1.1, despite having one of the smallest introgressions, showed 96% of ~500 DE genes regulated in *trans* (Supplemental Table I, Supplemental Figure S2). In contrast, IL1.1 and IL12.3, the ILs with highest number of genes in the introgression regions, showed smaller numbers of total and *trans* DE genes (Figure 1, Supplemental Table I, Supplemental Figure S2). Together these examples suggest that specific loci and not the introgression size determine gene regulation in *trans*. For example, *S. pennellii* chromosomal regions introgressed

in IL6.2.2 and IL12.1.1 have stronger genetic influence on global transcript abundance than the regions introgressed in IL1.1 and IL2.3. This could, in part, be due to the presence of genes encoding key transcription factors or developmental regulators in these regions with strong influence on gene expression pattern as is seen in the *ERECTA* containing genomic region in Arabidopsis (Keurentjes et al., 2007). A total of 7,943 unique tomato genes were DE between the ILs and cv. M82, representing approximately one third of the ~21,000 genes with sufficient sequencing depth to allow DE analysis. This suggests that in addition to protein coding differences, transcriptional regulation of less than one third of all genes accounts for most of the phenotypic and trait differences between the ILs and the cultivated parent. Among the 7,943 DE genes, 2790 (35%) genes were under *cis*-regulation only, 3603 (45%) genes were under *trans*-regulation only, and 1550 (20%) genes were under both *cis*- and *trans*- regulation. Of these DE genes 4057 (51%) were DE in only one IL, 840 genes were DE in 5 or more ILs, 8 genes were DE in over 50 ILs, and a single gene was DE in 70 ILs (Supplemental Dataset 2, Supplemental Figure S3A). The genetic regulation of many transcripts by more than one introgression and in *trans* suggests a complex genetic architecture of global transcript regulation in tomato, a phenomenon noticed in many other eukaryotes such as Arabidopsis, yeast, mice, and humans (Brem and Kruglyak, 2005; West et al., 2007; Schadt et al., 2008).

Global eQTL analysis at IL-bin level

Identifying eQTL localized to subsets of the introgressions, based on overlaps between them, enabled us to narrow down the regions that contain the regulatory loci. This analysis brings us one step closer to identifying potential candidates that influence gene expression patterns in tomato. The 20,795 genes with sufficient read depth in the RNAseq libraries were analyzed across 74 ILs with 112 bins defined based on the overlap of introgression regions from *S. pennellii* in the cv. M82 background (Chitwood et al., 2013) (Supplemental Figure S4). We found 7,225 significant eQTL involving 5,289 unique genes across the 74 ILs (Figure 2; Supplemental Dataset 3). These 7,225 significant eQTL (located in bins) were assigned the following designations (Supplemental Figure S4): *cis* (defined as local gene regulation within the same bin) - if the gene was located on the bin it is correlated with, *trans* (distant) - if the gene was correlated with a bin that is neither the bin it is on nor a bin that shares an overlapping IL with the correlated bin, and *chrom0* - if the gene is not located on any of the chromosomes and thus lies in the unassembled part of the genome. When a gene has a designation *cis*-eQTL and a secondary correlation was found with a bin that shares an overlapping introgression this secondary correlation was not designated as an eQTL (Supplemental Figure S4). When a gene does not have a designated *cis*-eQTL and a correlation was found with a bin that shares an overlapping introgression this correlation was designated as a *trans*-eQTL (Supplemental Figure S4). This resulted in a total of 1,759 *cis*-up and 1,747 *cis*-down eQTL, 2,710 *trans*-up and 920 *trans*-down eQTL, and 51 *chrom0*-up and 38 *chrom0*-down eQTL (Spearman's rho values, Supplemental Figure S5,

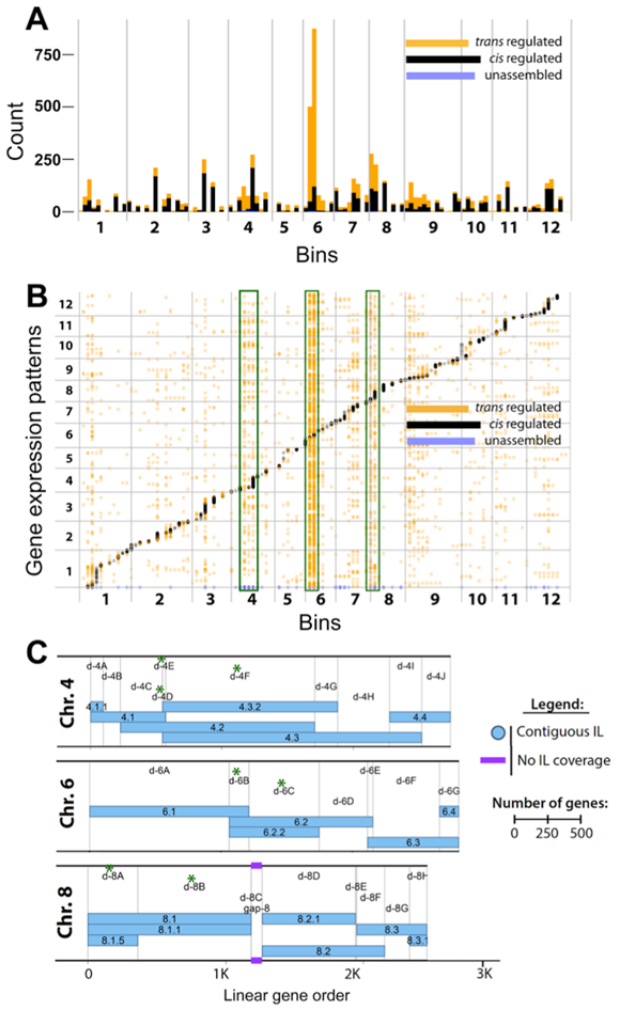

Figure 2. *Cis*- and *Trans*-eQTL plotted by bin across the 12 chromosomes of *S. lycopersicum* cv. M82. A) Stacked bar graph showing the sum of the number of eQTL mapping to each bin. B) Dotplot showing each eQTL arranged vertically by bin and horizontally by the location of the gene expression pattern it regulates. Bins with the largest numbers of *trans*-eQTL (4D, 4E, 4F, 6B, 6C, 8A, 8B) are highlighted by green boxes. C) Map of chromosomes 4, 6, and 8 showing the overlapping IL regions, which define the bins (Modified from Chitwood et al., 2013). Bins with the largest numbers of *trans*-eQTL are indicated by green asterisks.

Supplemental Table II). The majority of genes (over 4,000 out of 5,289) are under the regulation of a single eQTL (3,134 *cis*-, 1,014 *trans*-, and 19 *chromo0*-; Supplemental Figure S3B). This shows the predominance of *cis*-eQTL for genetic regulation of transcript expression patterns in the tomato ILs. Similar correlation between transcript-level variation and genome-wide sequence divergence within seven Arabidopsis accessions was reported to be due to *cis* control of a majority of the detected variation (Kliebenstein et al., 2006).

Several bins harbor a large number and/or large proportion of significant eQTL (Supplemental Dataset 4). The bins with over 100 significant *trans*-eQTL are on chromosomes 6, 8, and 4 (respectively 6C = 753, 6B = 452, 8A = 169, 8B = 128, 4D = 117). The number of genes in each bin is highly variable ranging from 1 to over 1890. Thus another metric to assess the influence of a particular bin is the number of significant *trans*-eQTL divided by total number of genes per bin. Bins with over 1.5 significant *trans*-eQTL per gene include 4D (117.0), 7D (14.7), 8C (8.0), 6B (3.0), 4E (2.5), and 3B (1.8). As expected, bins containing over 100 significant *cis*-eQTLs are scattered across the genome (Supplemental Dataset 4). Three bins, 1F, 3I and 8G, that each contain over 100 genes, have no significant *trans*- or *cis*-eQTL and are transcriptionally silent (Supplemental Dataset 4). The abundance of *trans*-eQTL on chromosomes 4, 6, and 8 strengthens the idea that *trans*-eQTL are clustered in *trans*-eQTL hotspots, as reported in other organisms, and these hotspots control the expression levels of a large number of transcripts (Brem et al., 2002; Schadt et al., 2003). The resolution in this analysis is at the level of bin, and these significant eQTL likely map to a smaller number of genes within the bins. However functional classification of genes being regulated by these eQTL and phenotypic association with the relevant ILs provides important insights into the candidate genes in the bin.

Clustering eQTL targets into modules defined by expression patterns

In order to functionally categorize the eQTL regulated genes, Barnes-Hut *t*-distributed stochastic neighbor embedding (BH-SNE, van der Maaten, 2013) was performed on the target genes to detect novel associations between gene expression patterns. The 5,289 genes with significant eQTL were mapped using BH-SNE (as explained in Materials and Methods) based on similarities in their expression patterns across the ILs. The exaggerated separation of non-neighboring clusters in this method improves 2D resolution, allowing identification of novel groupings not readily apparent in other clustering methods. This resulted in 42 distinct modules containing 3,592 genes (Figure 3). Seventeen of these modules had significant GO enrichment (p-value <0.05) with each module consisting of gene expression patterns either predominately regulated by *cis*- or *trans*-eQTL (Supplemental Table III). To determine which ILs are important for module regulation, the median expression value of module genes for each IL was calculated and used to identify ILs with significantly altered module gene expression.

Three modules were present in all mappings of the BH-SNE (van der Maaten and Hinton, 2008) determined through iterations of DBscan analysis and GO enrichment, and were designated as landmark modules (Figure 3B; Supplemental Figure S6; Supplemental Dataset 5; Supplemental Table III). The largest module had a GO enrichment for photosynthesis and related processes, and significant *trans*-eQTL scattered widely across the genome with no bin or IL identified as the primary regulating region (Figure 4B; Supplemental Figure S6A; Supplemental Dataset 5 & 6). The second landmark module was enriched for

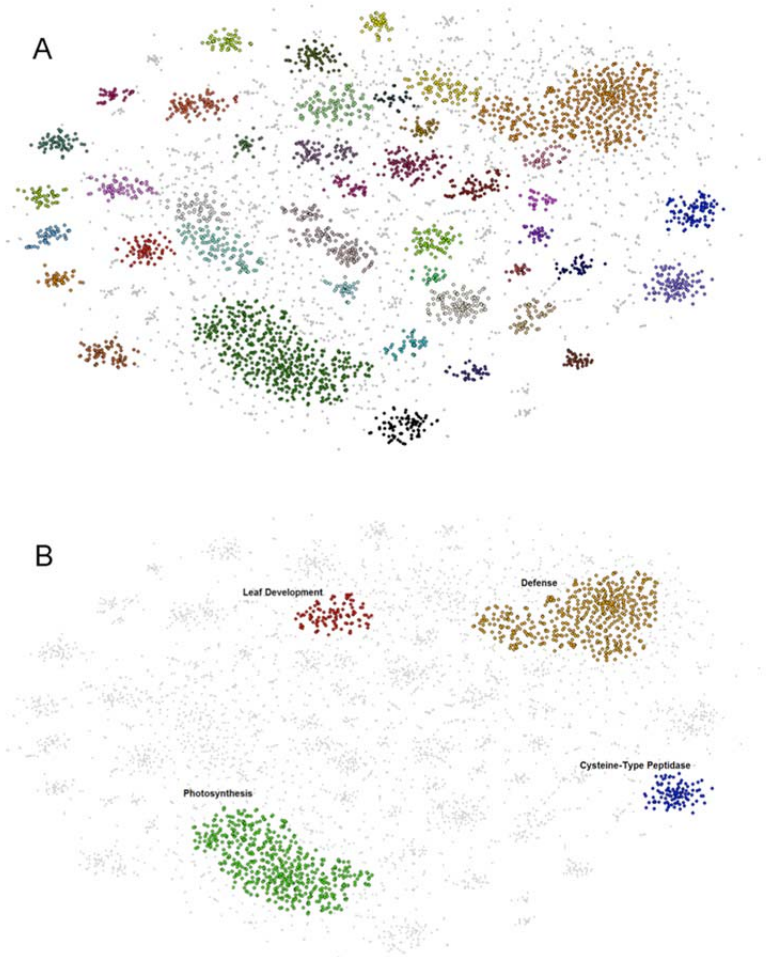

Figure 3. BH-SNE 2D mapping of eQTL. (A) Forty-two distinct modules identified by DBscan from the mapping generated by BH-SNE analysis. (B) The three modules defined as landmark modules: photosynthesis, defense and cysteine-type peptidase activity and the leaf development module's position within the mapping. Modules are false colored.

gene expression patterns with roles in defense, metabolism, and signaling with the majority of their *trans*-eQTL mapped to IL6.2 and 6.2.2 (Figure 4A; Supplemental Figure S6B; Supplemental Dataset 5 & 7). The third module, which is enriched for gene expression patterns with cysteine-type peptidase activity,

was predominately composed of genes regulated by *cis*-eQTL on IL 4.2, 4.3, and 4.3.2 (Bins 4E & 4F) (Figure 4C; Supplemental Figure S6C; Supplemental Dataset 5 & 8). A cluster of genes enriched for “peptidase regulation” also emerged from a transcriptome study of leaf development for three tomato

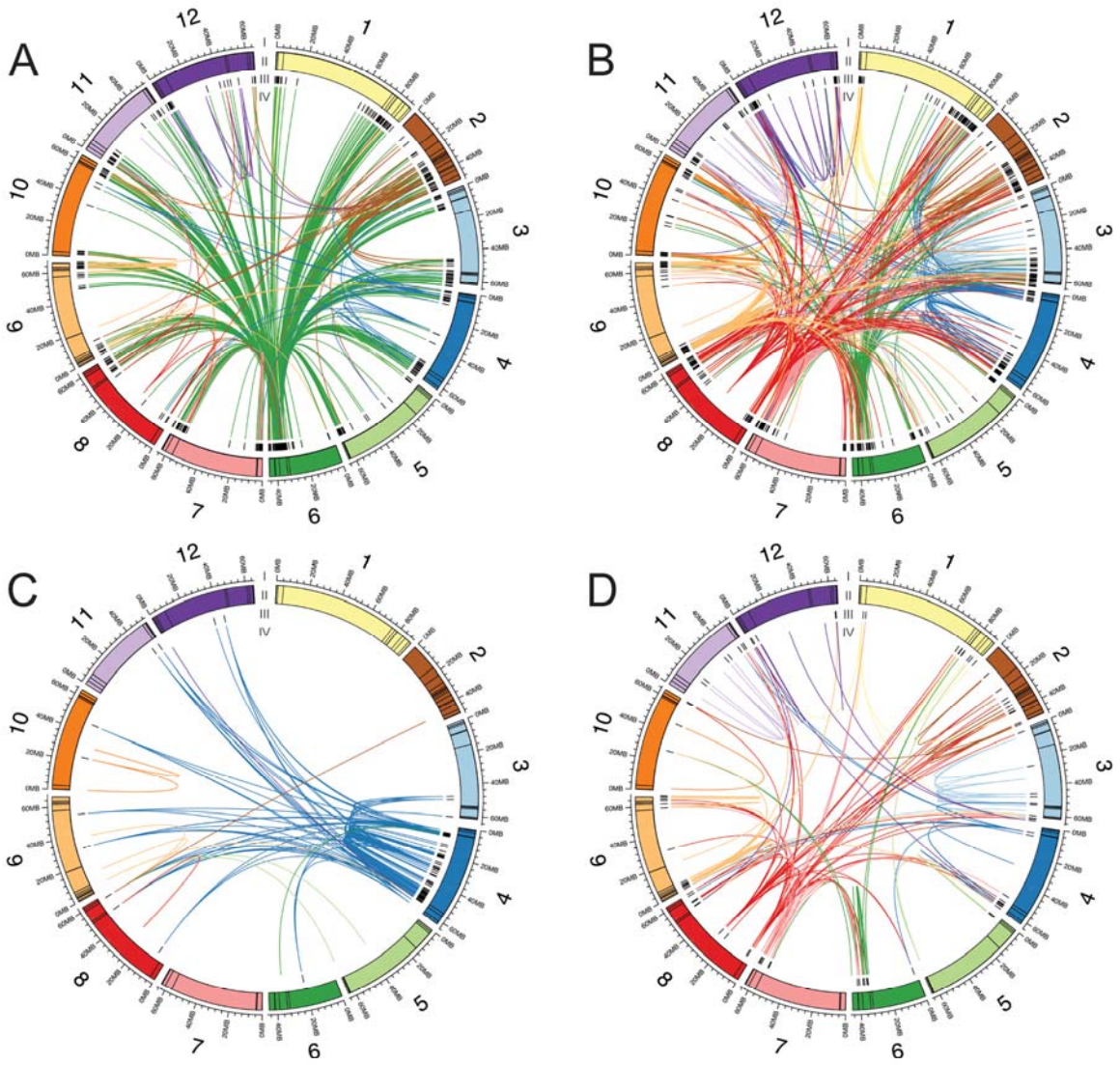

Figure 4. Connections between eQTL and the genes they regulate. Each plot includes the genes with eQTL that were clustered together into a module based on expression patterns. A) Defense module. B) Photosynthesis module. C) Cysteine peptidase module. D) Leaf development module. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

species; this cluster was uniquely associated with *S. pennellii* orthologs at the P5 stage of leaf development, indicating that this species has a unique pattern of gene expression, which involves peptidase regulation (Ichihashi et al., 2014) and may be related to leaf senescence processes (Diaz-Mendoza et al., 2014).

Genetic regulation of transcriptional responses associated with plant defense

One of the landmark modules from the clustering analysis was enriched for gene expression patterns related to plant defense (Figure 3B; Supplemental Dataset 7). Therefore we explored the genetic basis of transcriptional changes associated with plant defense. IL6.2 and IL6.2.2, in particular, influence the transcriptional responses of genes associated with plant defense and signaling (Supplemental Dataset 1). The overlapping regions between IL6.2 and IL6.2.2, bins 6B and 6C, also have the largest numbers of *trans*-eQTL (Figure 4A). The genes showing increased expression in both ILs compared to cv. M82, as well as the genes regulated by the corresponding bins, show enrichment of the GO categories response to stress and stimulus, cell death, defense response, and plant-type hypersensitive response (Supplemental Dataset 9 and 10). Promoter enrichment analysis for these genes showed enrichment of a W-box promoter motif that is recognized by WRKY transcription factors and influences plant defense response (Supplemental Dataset 11 and 12) (Yu et al., 2001). Both the bins, in particular bin 6C, contain genes involved in pathogen, disease, and defense response: such as *NBS-LLR resistance genes*, *WRKY transcription factors, Multidrug resistance genes, Pentatricopeptide repeat-containing genes, Chitinase,* and *Heat Shock Protein* coding genes. These genes are known to influence plant defense responses at a global transcriptome level. The transcriptional response in the ILs is also reflected in the morphology of IL6.2.2;

the plants are necrotic and dwarfed (http://tgrc.ucdavis.edu/pennellii_ils.aspx, (Sharlach et al., 2013). A bacterial spot disease resistance locus from *S. pennellii*, *RXopJ4*, that confers hypersensitive response in IL6.2 and 6.2.2, has been mapped to 190Kb region in bin 6C (Sharlach et al., 2013). In addition IL 4.1, 5.2, 8.1.1, and 9.1.3 also showed DE of genes involved in plant defense response, though the effect of was not as strong as that of IL6.2 and 6.2.2. Taken together, these findings suggest bins 6B and 6C contain master genetic regulators of plant defense response genes, though identification of the causal gene/s that influence so many other genes in *trans* will need further genetic dissection of these bins.

Genetic regulation of transcriptional responses associated with leaf development

Given the striking differences in leaf features between *Solanum pennellii* and cv. M82 that are manifested in many ILs (Chitwood et al., 2013), the IL population provides an excellent system for determining the extent of genetic regulation on leaf developmental genes. Moreover the tissues used for transcriptomic analyses included developing leaves and leaf primordia. Therefore we examined the effect of individual introgressions and associated bins on leaf developmental genes. Previous phenotypic and QTL analyses identified many ILs, such as IL4.3, IL8.1.5, IL8.1.1, and IL8.1, harboring loci regulating leaf and plant developmental traits (Holtan and Hake, 2003; Chitwood et al., 2013; Muir et al., 2014). IL4.3 harbors loci with the largest contribution to leaf shape, which manifests as larger epidermal cell size and reduced leaf complexity (Chitwood et

al., 2013). This IL shows decreased expression of many genes encoding proteins involved in cell division, such as Cyclin-dependent protein kinase regulator-like protein (CYCA2;3), Cyclin A-like protein (CYCA3;1), CYCLIN B2;3, and F-box/LRR-repeat protein 2 SKP2A (Supplemental Dataset 9). In addition, IL4.3 regulated genes were enriched for the promoter motifs MSA (M-specific activators that are involved in M-phase specific transcription) and the E2F binding site (Supplemental Dataset 11). Down-regulated genes in ILs 8.1.5, 8.1.1, and 8.1, also included leaf development and morphology genes, including genes encoding WD-40 repeat family protein LEUNIG, Homeobox-leucine zipper protein PROTODERMAL FACTOR 2, and the transcription factor ULTRAPETALA (Supplemental Dataset 9; Abe et al., 2003; Cnops et al., 2004; Carles et al., 2005). However target genes for none of the individual eQTL showed GO enrichment associated with leaf development categories (Supplemental Dataset 10). This could, in part, be due to complexity of the tissue used for RNAseq analysis, as the mixture of shoot apical meristem, growing leaves and hypocotyl could have diluted the significant leaf developmental genes.

As an alternative, we investigated the expression dynamics of a set of literature-curated leaf developmental genes (Ichihashi et al., 2014) across the ILs, to locate the hotspots regulating these genes (Supplemental Dataset 13). A number of canonical leaf developmental genes such as *SHOOT MERISTEMLESS* (Solyc02g081120, *STM*), *GROWTH-REGULATING FACTOR* 1 (*GRF1,* Solyc04g077510), *ARGONAUTE 10* (*AGO10*, Solyc12g006790), *BELL* (*BEL1*, Solyc08g081400) *LEUNIG* (Solyc05g026480) and *SAWTOOTH 1* (*SAW1,*

Solyc04g079830) were differentially expressed in more than five ILs. At the level of bins, genes involved in leaf development were seen to be regulated by eQTL scattered widely across the genome (Figure 4D). This underscores the highly polygenic regulation of leaf development (Chitwood et al., 2013) as multiple loci, residing in many different chromosomal locations, regulate the expression of key leaf-developmental genes. Fifty-eight of literature-curated leaf developmental genes have expression patterns under eQTL regulation; three bins (2G, 7H, 8A) each have four significant eQTL regulating leaf developmental genes, which is the maximum number for any bin (Figure 4D, Supplemental Dataset 14). Bin 8A, the overlapping region of IL8.1, 8.1.1, and 8.1.5, contains *LEUNIG* that showed a 200-fold down-regulation in the differential gene expression analysis for the three ILs (Supplemental Dataset 1). Bin 8A, further, contains *trans*-eQTL that regulate the expression pattern of key auxin-related genes, such as *AINTEGUMENTA (ANT;* Solyc04g077490), and *YUCCA4* (Solyc06g065630; Supplemental Dataset 14). Consistent with this, IL 8.1.1, 8.1.5 and 8.1 have strong reduction in leaf complexity in the direction of *S. pennellii* (Chitwood et al., 2013). Similarly IL4.3 contains loci with the largest contributions to leaf shape variation (Chitwood et al., 2013), and also regulates the *TCP5* gene in *trans*.

BH-SNE clustering corroborated the strong leaf phenotypes for ILs 4.3, 8.1, 8.1.1, and 8.1.5. We compared a curated list of leaf developmental genes combined with a set of co-expressed genes (LC+; (Ichihashi et al., 2014)) to the 3,592 genes found within the identified modules in the BH-SNE mapping (Figure 3). A total of 175 genes out of 697 from the list were found within the modules;

the highest concentrations were 108 genes located in the photosynthesis module and 19 in the leaf development module (Supplemental Figure S7B; Supplemental Dataset 15 and 16), suggesting a relationship between these two modules. The eQTL regulating expression patterns of these leaf developmental genes were distributed broadly across the genome (Supplemental Figure S7A, Supplemental Dataset 17). This is consistent with the differential expression and eQTL analyses, which found that no one bin or IL regulates leaf developmental genes. Over one third of the gene expression patterns in the leaf development module have significant eQTL that map to bins 4D, 8A, and 8B (5.4%, 16.2%, and 15.5%, respectively), suggesting that these bins contain important regulators of leaf development. The canonical leaf developmental genes within this module include: *GROWTH-REGULATING FACTOR*7 (*GRF*7; Solyc03g082430); GRF10 (Solyc09g009200); *ERECTA-LIKE 1* (Solyc03g007050, *ERL1*); *ARGONAUTE 10* (Solyc12g006790, *AGO10*); *AUXIN RESPONSE FACTOR 9* (Solyc08g008380, *ARF9*); and *AINTEGUMENTA* (Solyc04g077490, *ANT*). Altogether DE, eQTL and BH-SNE results indicate that while there is no obvious master regulatory bin for leaf developmental genes, many are under strong genetic regulation by eQTL distributed throughout the genome (Figure 4D).

Genetic regulation of transcriptional responses associated with photosynthesis

Since photosynthesis GO terms were enriched for the largest module from the clustering analysis (Figure 3B) and there was a correlation between

photosynthesis and leaf developmental modules (Supplemental Figure S7B), we examined the genetic regulation of photosynthetic genes by specific ILs and corresponding bins. Genes related to photosynthesis show increased expression across 21 ILs distributed on all chromosomes except chromosome 5 (Supplemental Dataset 9), showing multigenic regulation of these critical traits. Many of these ILs, including 8.1.5, 8.1.1, and 8.1, which share an introgressed region, and IL4.3, have up-regulated genes enriched for the GO-categories photosynthesis, chlorophyll biosynthesis, and response to light stimulus (Supplemental Dataset 9). The eQTL target genes of corresponding bins 4D and 8A also have GO enrichments for photosynthesis (Supplemental Dataset 10). These ILs also had significant effects on leaf shape and morphology (Chitwood et al., 2013). Previously we reported a link between leaf development and photosynthesis by meta-analysis of developmental and metabolic traits (Chitwood et al., 2013). This indicates that ILs may also differ from each other and from cultivated M82 background in photosynthetic efficiency. However no studies, so far, has investigated the photosynthetic phenotype of these ILs.

To analyze this relationship between the leaf development and photosynthesis modules the median expression value of all genes in each module was compared resulting in a significant negative correlation (adj r2 = 0.77; Figure 5). The transition from leaf development to leaf maturation is potentially reflected in our data. The genes found in the leaf development module may promote developmental processes such as cell division and maintenance or meristematic potential, whereas the leaf development related genes found in the

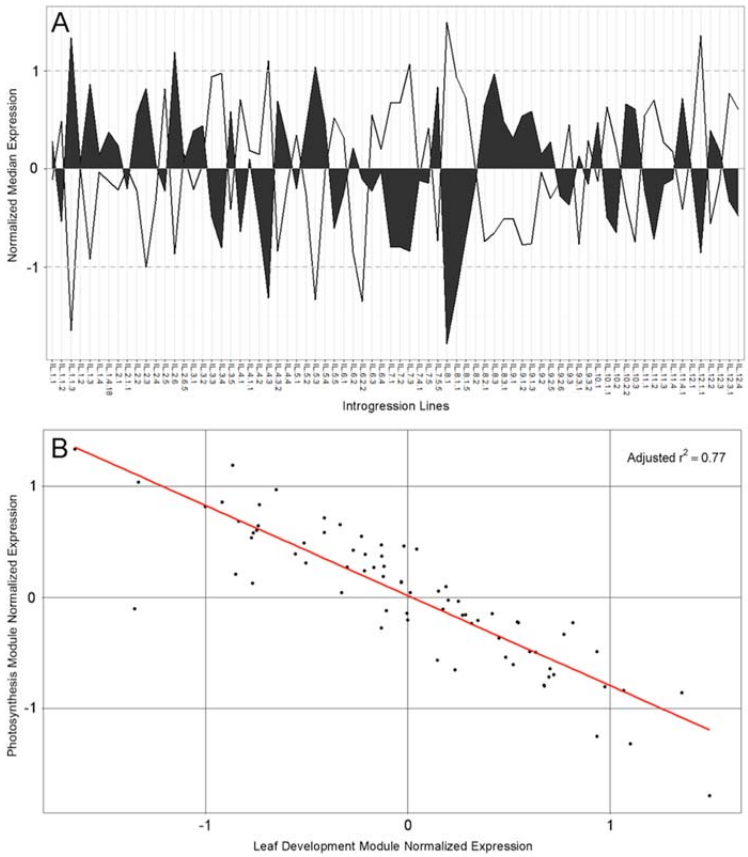

Figure 5. Median expression values for leaf development and photosynthesis related modules and expression correlation. (A) The median expression values of a module for each IL are shown. A consistent negative correlation between photosynthesis and Leaf development transcript expression is evident across nearly all 74 ILs. Dashed lines indicate one significant deviation from the module mean expression. Filled areas represent the median expression of the leaf development module, while open areas indicate the photosynthesis module median expression. (B) Leaf development median expression versus photosynthesis median expression values for each IL show a distinct negative correlation with an adjusted R-squared value of 0.77 (calculated by linear regression in R).

photosynthesis module may act to suppress this process to allow for maturation of the leaf. The strong negative correlation between the expression patterns of these two modules (Figure 5) and the overlap between their most influential eQTLs (Supplemental Figure S7A; Supplemental Table III) suggests that leaf development and photosynthetic genes not only have expression levels in opposition but also likely share common regulatory loci. Each module has eQTLs on bins 4D, 8A, and 8B which may indicate these regions drive the regulatory "switch" from development to maturation and photosynthetic activity.

Dissection of identified eQTL to spatially- and temporally- regulated development

The significant eQTL detected in this study represent the genetic regulation of gene expression in the tomato shoot apex that includes shoot apical meristem (SAM) and developing leaves. Using previous gene expression data, we spatio-temporally resolved the detected eQTL to specific tissues and temporally-regulated development. We analyzed gene expression in laser micro-dissected samples representing the shoot apical meristem (SAM) + P0 (the incipient leaf) vs. the P1 (the first emerged leaf primordium) (Figure 6A) and hand dissected samples of the SAM + P0-P4 vs. the P5 collected over time (Figure 6B-C), representing genes regulated by vegetative phase change (heteroblasty) (Chitwood et al., 2015). The former dataset represents gene expression in the meristem (SAM) and the first differentiated leaf (P1), whereas the latter dataset informs about gene expression changes during the temporal development of the shoot apex.

Using a bootstrapping approach, we identified bins statistically enriched for genetically regulating genes with previously identified gene expression patterns (Figure 6D-F). Except for one instance (cis-regulated genes with high SAM/P0 expression located in bin 2I), bins enriched for gene expression patterns represented *trans* regulation, hinting at predominately transcription factor-based regulation of gene expression patterns. Most SAM/P0 vs. P1 enriched bins were enriched for P1 gene expression (Figure 6D). We previously showed that genes with high P1 expression are enriched for photosynthetic-related GO terms,

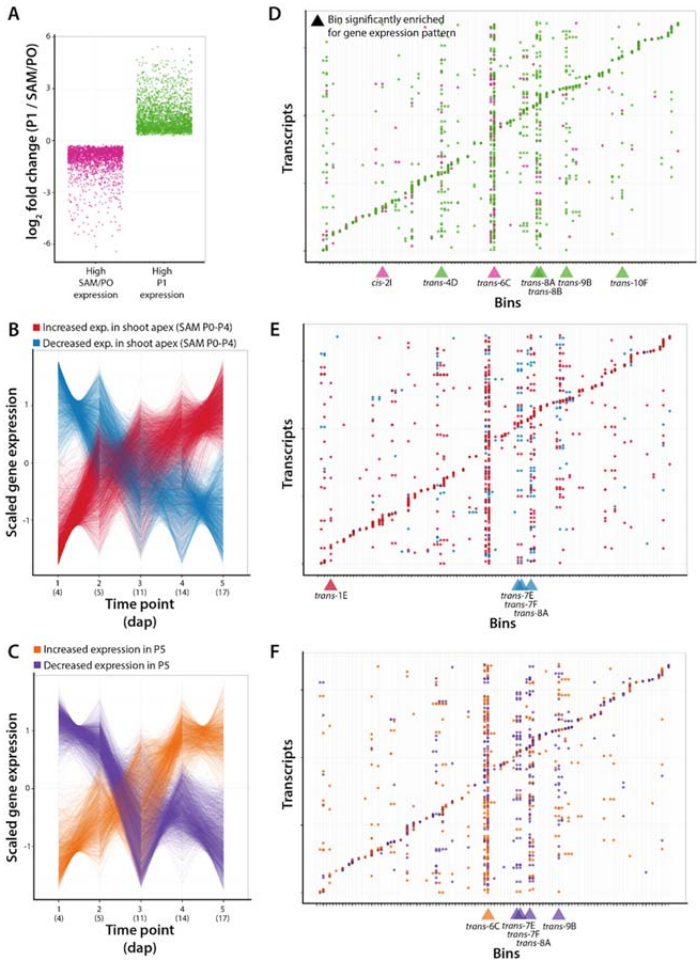

Figure 6. Enriched expression patterns for genes genetically regulated by bins.
A) Log fold change values (P1/SAM+P0) for previously identified differentially expressed genes with high expression in the SAM + P0 (magenta) vs. P1 (green). **B)** Scaled gene expression values for previously identified differentially expressed genes with increasing (red) and decreasing (blue) expression over developmental time in the SAM + P0-P4. **C)** Scaled gene expression values for previously identified differentially expressed genes with increasing (orange) and decreasing (purple) expression over developmental time in P5. **D)** Transcripts (y-axis) and bins (x-axis) showing the genetic regulation of gene expression (eQTL). Colors indicate SAM + P0 (magenta) and P1 (green) transcripts. Bins enriched for genetically regulating genes with specific expression patterns are indicated below with triangles. **E)** Same as in D), except showing genes with increasing (red) and decreasing (blue) expression over temporal time in the SAM + P0-P4. **F)** Same as in D), except showing genes with increasing (orange) and decreasing (purple) expression over temporal time in P5. Previously determined gene expression patterns are previously published (Chitwood et al., 2015).

compared to SAM/P0 genes enriched for transcription, cell division, and epigenetics-related GO terms (Chitwood et al., 2015), suggesting a genetic basis at both a functional and tissue-specific level for genes related to photosynthesis expressed preferentially in the P1 compared to the SAM/P0. Of bins enriched for SAM/P0 vs. P1 gene expression patterns, bins 4D, 8A, and 9B were significantly enriched for photosynthetic-related GO terms, indicating that these bins in

499 particular contain eQTL important for the regulation of photosynthetic processes
500 (Supplemental Dataset S18).

501 Bins enriched for regulation of genes with temporally-dependent
502 expression were mostly associated with genes with decreasing expression over
503 time, for both the SAM + P0-P4 and P5 (Figure 6E-F). Interestingly, 3 bins (7E,
504 7F, and 8A) share enrichment for genes with decreasing expression patterns
505 over time in both the SAM + P0-P4 and P5 (Figure 6E-F), suggesting true
506 temporal *trans* regulation, regardless of tissue, by these loci. Broadly, genes with
507 increasing expression over time are associated with transcription and small RNA
508 GO terms in both the SAM + P0-P4 and P5, whereas decreasing expression over
509 time is associated with translation associated GO terms in the SAM + P0-P4 and
510 photosynthetic activity in the P5. Specifically bin 6C, which was enriched for
511 genes with increasing expression over time for P5, was also significantly
512 enriched for defense processes GO terms; whereas bins 8A and 9B, that showed
513 enrichment for genes with decreasing patterns over time for P5, were enriched
514 for photosynthesis-related GO terms (Supplemental Dataset S18).

515

516 **Linking leaf and hypocotyl phenotypes to detected eQTL**

517 To better understand genetic regulation of gene expression itself, eQTL
518 provide crucial links between genetic loci and organism level phenotypes, often
519 generated in diverse tissue/organ types and environments. The tissue samples
520 used for our transcriptomic analyses contained developing leaves and hypocotyl.
521 Therefore, in order to connect detected eQTL with leaf and hypocotyl phenotypes

522 under two different environmental conditions, we correlated gene expression with
523 leaf number, leaf complexity (as measured in Chitwood et al., 2014) and
524 hypocotyl length phenotypes of the ILs grown under simulated sun and shade
525 conditions. Significant correlations with gene expression patterns were identified
526 for all three phenotypes analyzed under both treatments (Supplemental Table IV).
527 Focusing on a subset of these gene expression patterns that had associated
528 eQTL enabled us to connect the phenotypes to their regulatory loci
529 (Supplemental Table IV).

530 Genes negatively correlated with leaf number showed enrichment of leaf
531 development GO terms, whereas positively correlated genes showed enrichment
532 of photosynthesis-related GO terms (Supplemental Figure S8A-B; Supplemental
533 Dataset 2 in Chitwood et al., 2014). The expression pattern of these genes
534 associated with leaf number was predominantly regulated by eQTL on
535 chromosomes 7 and 8 (Supplemental Figure S8C-D) and reveals developmental
536 transitions reflecting a shift from leaf development to photosynthetic activity with
537 the increase in leaf number and plant maturity. For the leaf complexity trait,
538 correlations were reversed compared to leaf number, and a positive correlation
539 with genes enriched for leaf development and negative correlation with
540 photosynthesis genes was seen (Supplemental Figure S9A-B; Supplemental
541 Dataset 19). Moreover a larger number of correlations with leaf developmental
542 gene expression patterns were observed for leaf complexity in response to shade
543 (Supplemental Figure S9C-D; Supplemental Table IV). This reflects the higher
544 leaf complexity observed in shade-grown tomato plants (Xu et al., 2009;

Chitwood et al., 2015). eQTL on chromosomes 4, 7 and 8 were primarily involved in regulation of gene expression patterns associated with leaf complexity (Supplemental Figure S9C-D). These results, in combination with DE, eQTL and BH-SNE, highlight bins on chromosomes 4 and 8 as important genetic regulators of leaf developmental genes.

Five genes were positively correlated with hypocotyl length under simulated shade and only one gene (Solyc10g005120) was negatively correlated with hypocotyl length under both sun and shade (Figure 7A**;** Supplemental Dataset 20). eQTL for the positively correlated genes are located on chromosomes 3, 7, and 11, whereas the single cis-eQTL for the negatively correlated gene, Solyc10g005120 (an uncharacterized *Flavanone 3-hydroxylase-like* gene**)**, was located in bin 10A.1 (Supplemental Figure S10; Figure 7B). The gene is only expressed in IL 10.1, which has the *S. pennellii* copy of the gene, and not in other ILs with the M82 version of the gene, such as IL 10.1.1. Further, M82 and IL 10.1.1 showed significant elongation of hypocotyls in response to shade, whereas IL 10.1 had an attenuated shade avoidance response (Supplemental Figure S11). This indicates the influence of genes exclusively in this introgression on hypocotyl length, specifically those located in bin 10A that includes Solyc10g005120. A set of BIL (Backcross Inbred Lines) lines generated in a new mapping population developed from the same parents, cv. M82 and *S. pennellii*, provides higher resolution gene mapping as bin sizes are significantly smaller (Muller et al., 2016; Fulop et al., In Review). BIL-128 harbored a smaller introgression on chromosome 10 compared to IL 10.1, carried the *S. pennellii*

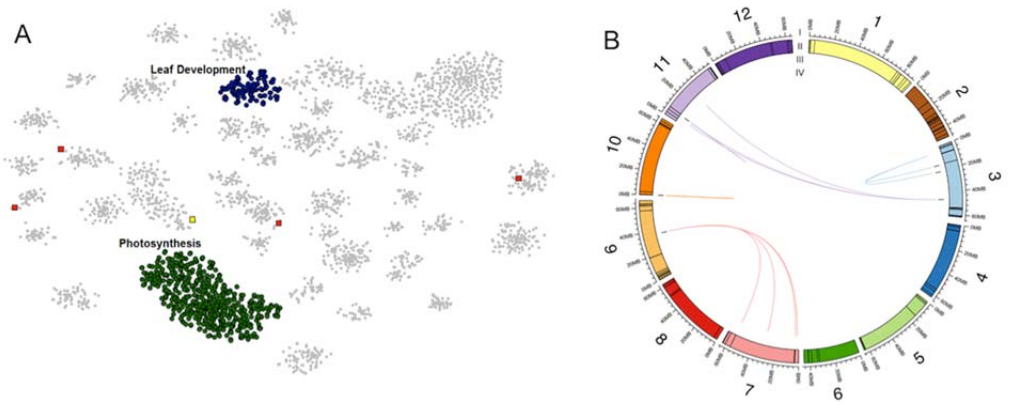

Figure 7. eQTL regulation of gene expression patterns that correlate with hypocotyl length.
A) Forty-two distinct modules identified by DBscan from the eQTL mapping generated by BH-SNE analysis. Modules enriched for genes with leaf development and photosynthesis GO terms are labeled in blue and green, respectively. Genes with expression patterns correlated with hypocotyl length under simulated shade are indicated by squares with positive correlations in red and negative correlations in yellow.
B) Genes with expression patterns correlated with hypocotyl length under simulated shade are shown connected to their respective eQTL with chords. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

version of Solyc10g005120, and resembled IL10.1 in its hypocotyl length in response to shade (Supplemental Figure S10; S11). However, it had an additional introgression on chromosome 2. BIL-033 that shared an overlapping introgression with BIL-128 on chromosome 2 showed a hypocotyl phenotype similar to M82, ruling out the possibility of influence of chromosome 2 genes on the hypocotyl phenotype (Supplemental Figure S10; S11). Though the function of Solyc10g005120 has not been described for Arabidopsis or tomato so far, our gene expression and phenotypic observations present it as a new potential candidate for further exploring the regulation of hypocotyl length and other plant phenotypes in response to shade.

Transgressive expression of genes among ILs

Interspecific crosses combining divergent genomes can sometimes result in improved performance. Transcriptomic studies of wild and resynthesized hybrids show extensive gene expression alterations in the hybrids compared to their parents (Hegarty et al., 2008). The IL population allows us to study the combination of divergent genomes across many genetically distinct lines. Many of the changes we see are “transgressive”, falling outside the range seen in either parent. A total of 2,286 genes, more than one fourth of unique DE genes between the ILs and cv. M82, showed transgressive expression patterns, i.e. genes were differentially expressed for the IL but not for *S. pennellii* compared to cv. M82 (Supplemental Dataset 21). These genes were distributed across all ILs but the extent of transgressive gene expression varied among the ILs. ILs with high number of differentially expressed genes, such as IL6.2.2, IL4.3, and IL8.1, also showed higher number of transgressively expressed genes (Supplemental Figure S12A). While synthetic allotetraploids generated from interspecific crosses show large scale down regulation of gene expression (likely due to silencing of duplicate gene copies (Wang et al., 2006)) and the affected genes fall in the categories of cell defense, ageing and hormonal regulation, the transgressive genes in this IL population generated by interspecific crossing show mostly upregulation of gene expression with genes falling in the broad categories of photosynthesis, energy metabolism, microtubule movement, and extrinsic membrane components (Supplemental Dataset 22). In order to estimate the relative effect of the introgression on transgressive gene expression, we quantified the ratio of the number of genes with transgressive expression to the

number of genes with *S. pennellii* like expression (transgressive_to_*S. pennellii*-like) for each IL (Supplemental Figure S12B). A few ILs, such as IL1.4.18, IL4.3, IL7.3, IL8.1, IL11.1, and IL12.1.1, showed a transgressive_by_*S. pennellii* value greater than 0.5, suggesting a higher effect of the corresponding introgression on gene expression in M82 background. Most of the transgressively expressed genes for each IL were located in *trans*, beyond the introgression region (Supplemental Figure S12A). These may result from novel combinations of regulatory factors (Riddle and Birchler, 2003), recombination of alleles present at different loci in the parent species through complementary gene action (Rieseberg et al., 1999), more abundant production of small interfering (si) RNAs in progeny of interspecific crosses than in either parent with concomitant suppression of the corresponding target genes, and hypermethylation of the corresponding genomic DNA (Shivaprasad et al., 2012).

Conclusion

In this study we have investigated the regulation of gene expression in the progeny of crosses between cultivated tomato (*Solanum lycopersicum* cv. M82) and a wild relative (*Solanum pennellii*, accession LA716). A combination of differential gene expression, eQTL, and clustering analyses provide a comprehensive picture of genetic regulation of transcript expression patterns in this IL population. We show differential genetic effects of discrete genomic regions that regulate gene expression patterns. Certain regions are highly active, particularly those on chromosome 4, 6, and 8, and they influence the expression

patterns of genes in their own introgression regions and beyond. Clustering of genes by expression patterns and GO/promoter motif enrichment analyses provided further insights into the genetic regulation of expression patterns of genes related to specific pathways. Our data show that some biological pathways, such as plant defense, are under the regulation of a limited number of loci with strong effects, whereas loci regulating other pathways, such as photosynthesis and leaf development, are scattered throughout the genome most likely with weaker individual effects. Using published gene expression atlases for tomato we were able to partition the eQTLs into tissue specific and age dependent patterns of expression. We correlated gene expression with leaf and hypocotyl phenotypes and identified the regulatory regions driving these gene expression patterns. In the case of hypocotyl length we identified a strong candidate gene that may regulate hypocotyl length under shade avoidance. This is the first comprehensive insight into the global regulation of transcript abundance in tomato and sets the stage for identification of gene/s underlying these regulatory loci. Coupled with comprehensive phenotyping on these ILs this data set provides insights into how the expression pattern of key genes or gene modules can be genetically manipulated to achieve a desirable plant phenotype. We report on differential gene expression between the ILs and the recurrent parent, M82, eQTLs that regulate thee gene expression patterns and transgressive changes in gene expression. Together these facets of change in gene expression may explain hybrid performance, and could provide genotypes that have an enhanced ability to survive in habitats not accessible to the parents.

While they may provide avenues for enhancing breeding efforts, they also reveal unexpected consequences of wide crosses. A study of transcriptional variation is only one component of many others, such as protein and metabolite levels, to influence downstream phenotypes. Given a number of important plant traits have complex inheritance, transcript abundance can function as an intermediate between genomic DNA sequence variation and complex traits. Our ability to predict and understand the downstream effects of genes introgressed from wild relatives on gene expression patterns and ultimately phenotypes will be a critical component of crop plant enhancement.

Materials and Methods

Plant Materials, Growth Conditions, and Experimental Design

Plant materials, growth conditions, and experimental design were described in (Chitwood et al., 2013), but are outlined here briefly. Seeds of *Solanum pennellii* ILs (Eshed and Zamir, 1995; Liu and Zamir, 1999) and *Solanum lycopersicum* cv. M82 were obtained either from Dani Zamir (Hebrew University, Rehovot, Israel) or from the Tomato Genetics Resource Center (University of California, Davis). Seeds were stratified in 50% bleach for 2 min., grown in darkness for 3 d for uniform germination before moving to a growth chambers for 5 days. Six seedlings of each genotype were planted per pot for each replicate. The 76 ILs (and two replicates each of cv. M82 and *S. pennellii*) were divided into four cohorts of 20 randomly assigned genotypes. These cohorts were placed across

four temporal replicates in a Latin square design as described in (Chitwood et al., 2013). The seedlings were harvested 5 d after transplanting (13 d of growth in total). Cotyledons and mature leaves >1 cm in total length were excluded, and remaining tissues (including the shoot apical meristem) above the midpoint of the hypocotyl were pooled, for all individuals in a pot, into 2-mL microcentrifuge tubes and immediately frozen in liquid nitrogen. Two ILs, IL7.4 and IL12.4.1 were not included in the final analysis due to seed contaminations.

Growth conditions and quantification of hypocotyl length

Seeds 76 ILs along with the parents were sterilized using 70% ethanol, followed by 50% bleach, and finally rinsed with sterile water. This experiment was replicated three times each in 2011 and 2012. Ten to twelve seeds of each IL were sown into Phytatray II (Sigma-Aldrich) containers with 0.5x Murashige and Skoog minimal salt agar. Trays were randomized and seeds germinated in total darkness at room temperature for 48h. Trays of each IL were randomly assigned to either a sun or shade treatment consisting of 110μMol PAR with a red to far-red ratio of either 1.5 (simulated sun) or 0.5 (simulated shade) at 22°C with 16 hour light / 8 hour dark cycles for 10d. Three genotypes were excluded from the analyses due to poor germination (IL3.3) or their necrotic dwarf phenotypes (IL6.2, 6.2.2). After 10d, seedlings were removed from the agar and placed onto transparency sheets containing a moistened kimwipe to prevent dehydration and scanned using an Epson V700 at 8-bit grayscale at 600 dpi. Image analysis was carried out using the software ImageJ (Abramoff et al., 2004).

For hypocotyl length analysis of backcross inbred lines between *S. pennellii* and *S. lycopersicum* cv. M82, seeds were sterilized in 50% bleach and then rinsed with sterile water. They were then placed in Phytatrays in total dark at room temperature for 72 hours, and then moved to 16 hour light / 8 hour dark for 4d. Seedlings were transferred to soil using a randomized design and assigned to either a sun or shade treatment (as described above) for seven days. Images were taken with a HTC One M8 Dual 4MP camera and hypocotyl lengths measured in ImageJ (Abramoff et al., 2004) using the Simple Neurite Tracer (Longair et al., 2011) plugin.

Correlation of phenotype with gene expression pattern

Gene expression patterns were correlated with three phenotypes collected from the ILs along with the parents. Normalized estimated read counts with 3-4 independent replicates per IL were log2 transformed prior to the analyses. Leaf number and leaf complexity were collected from the ILs as outlined in Chitwood et al. (2014) under both sun and shade treatments. Hypocotyl lengths were measured as detailed above. To test whether a gene's expression pattern was correlated with a particular phenotype boostrapping analyses were performed. Expression and phenotype data were randomly permuted (with replacement) using the sample() function against IL and then merged. For each analysis, 1000 replications were performed and the p-values were calculated from the Spearman's rho value distributions. P-values were adjusted for multiple comparisons using the BH correction (Benjamini & Hochberg, 1995). Significant

correlations were identified as those with an adjusted p-value < 0.05 and the mean rho value (the correlation coefficient) was used to designate the correlation as either positive (positive slope) or negative (negative slope). All analyses were implemented using the statistical software R and custom scripts (R Core Team, 2015).

RNA-Seq Library Preparation and Preprocessing RNA-Seq Sequence Data

RNAseq libraries were prepared and the reads were preprocessed as described in (Chitwood et al., 2013), and are outlined here. mRNA isolation and RNA-Seq library preparation were performed from 80 samples at a time using a high-throughput RNA-Seq protocol (Kumar et al., 2012). The prepared libraries were sequenced in pools of 12 for replicates 1 and 2 (one lane each) and in pools of 80 for replicates 3 and 4 (seven lanes) at the UC Davis Genome Centre Expression Analysis Core using the HiSeq 2000 platform (Illumina). Preprocessing of reads involved removal of low quality reads (phred score < 20), trimming of low-quality bases from the 39 ends of the reads, and removal of adapter contamination using custom Perl scripts. The quality-filtered reads were sorted into individual libraries based on barcodes and then barcodes were trimmed using custom Perl script.

Read Mapping and Gene Expression Analysis

Mapping and normalization were done on the iPLANT Atmosphere cloud server (Goff et al., 2011). *S. lycopersicum* reads were mapped to 34,727 tomato cDNA

741 sequences predicted from the gene models from the ITAG2.4 genome build
742 (downloadable from
743 ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.4_release/). A
744 pseudo reference list was constructed for *S. pennellii* using the homologous
745 regions between *S. pennellii* scaffolds v.1.9 and *S. lycopersicum* cDNA
746 references above. Using the defined boundaries of ILs, custom R scripts were
747 used to prepare IL-specific references that had the *S. pennellii* sequences in the
748 introgressed region and *S. lycopersicum* sequences outside the introgressed
749 region. The reads were mapped using BWA (Li and Durbin, 2009; Roberts and
750 Pachter, 2013) using default parameters except for the following that were
751 changed: bwa aln: -k 1 -l 25 -e 15 -i 10 and bwa samse: -n 0. The bam alignment
752 files were used as inputs for express software to account for reads mapped to
753 multiple locations (Roberts and Pachter, 2013). The estimated read counts
754 obtained for each gene for each sample from express were treated as raw counts
755 for differential gene expression analysis. The counts were then filtered in R using
756 the Bioconductor package EdgeR version 2.6.10 (Robinson and Oshlack, 2010)
757 such that only genes that have more than two reads per million in at least three
758 of the samples were kept. Normalization of read counts was performed using the
759 trimmed mean of M-values method (Robinson and Oshlack, 2010), and
760 normalized read counts were used to identify genes that are differentially
761 expressed in each IL compared to cv. M82 parent as well as in between two
762 parents, *S. pennellii* and M82. The DE genes for each IL were compared to those
763 between the two parents to identify genes that were differentially expressed for

the IL but not for *S. pennellii* compared to cv. M82. Those genes were considered to show transgressive expression pattern for the specific IL, whereas other DE genes were considered to show *S. pennellii* like expression pattern.

Methods for eQTL Analyses

eQTL mapping analyses were performed to determine whether the expression pattern of a gene is correlated with the presence of a specific introgression from *S. pennellii* into *S. lycopersicum* cv. M82. This was examined at the level of "bin", with a bin defined as a unique overlapping region between introgressions. Examining eQTL at the bin level enables them to be mapped to considerably smaller intervals than the ILs themselves (Liu and Zamir, 1999). eQTL mapping analyses were performed on the normalized estimated read counts with 3-4 independent replicates per IL, which were log2 transformed prior to the analyses. To test whether a gene's expression pattern is correlated with the presence of a particular bin a Spearman's rank correlation test was used with ties resolved using the midrank method. P-values were adjusted for multiple comparisons using the BH correction (Benjamini and Hochberg, 1995). Significant eQTL were identified as those with an adjusted p-value < 0.05 and Spearman's rho (the correlation coefficient) was used to designate the eQTL as up (positive slope) or down (negative slope). All analyses were implemented using the statistical software R and custom scripts (R Core Team, 2015).

Methods for eQTL clustering analysis

Data Preparation: In preparation for analysis using the Barnes-Hut-SNE algorithm the data set was log2 transformed. Each gene's expression profile was then normalized across all 74 introgression lines so that the profile had a mean of zero and a standard deviation of one. Normalization of the data allowed for comparison of the relative relationship between each gene expression profile (Bushati et al., 2011).

Barnes-Hut-SNE: *t*-SNE or *t*-distributed stochastic neighbor embedding (van der Maaten and Hinton, 2008) is a non-linear dimensionality reduction method, which faithfully maps objects in high dimensional space (H-space) into low dimensional space (V-space). Crowding is avoided through the long-tailed *t*-distribution, which forces non-neighbor clusters farther away from each other in V-space than they actually are in H-space (van der Maaten and Hinton, 2008). The exaggerated separation of non-neighboring clusters improves 2D resolution, allowing identification of novel groupings not readily apparent in other clustering methods. However, this method is resource intensive and with higher dimensionality the number of genes that can be analyzed is limited. Because of this, Barnes-Hut-SNE a newer implementation of *t*-SNE was used, which greatly increases the speed and number of genes that can be analyzed (van der Maaten, 2013). It accomplishes this through the use of a Vantage Point tree and a variant of the Barnes-Hut algorithm (van der Maaten, 2013). For clustering, 2D maps were generated using a perplexity of 30 and without the initial PCA step from the Barnes-Hut-SNE R implementation (Rtsne package; (Krijthe, 2014)). Theta was

set to 0.3 based on (van der Maaten, 2013) in order to maintain an accurate dimensionality reduction without sacrificing processing speed.

Clustering for Module Selection: The DBscan algorithm (Density Based spatial clustering of applications with noise) was used to select modules from the Barnes-Hut-SNE results (fpc package; (Hennig, 2014)). This had the advantage of both selecting modules and removing any genes that fell between modules. The scanning range (*epsilon*) and minimum seed points (*minpts*) were selected manually, and used to determine if any one point is a member of a cluster based on physical positioning within the mapping relative to neighboring points.. A *minpts* of 25 was used to capture smaller modules on the periphery and an *epsilon* of 2.25 was used to avoid the overlapping of internal and closely spaced modules.

Plots: Boxplots were generated from normalized expression values for each module. The ribbon plot was generated from correlated expression values from leaf development and photosynthesis related modules. These plots were generated using ggplot form the ggplot2 R Package (Wickham, 2009). The median expression of the genes mapped to a module was calculated for each IL and replicated for all modules. Significant ILs were identified as those with a median expression greater than 1 standard deviation from the mean of all genes across all ILs in the module.

GO Enrichment analysis

Differentially expressed genes for individual ILs and Genes with significant eQTL were analyzed for enrichment of Gene Ontology (GO) terms at a 0.05 false discovery rate cutoff (goseq Bioconductor package; Young et al., 2010).

Promoter enrichment analysis

Promoter enrichment analysis was performed by analyzing the 1000 bp upstream of the ATG translational start site for genes with significant eQTL using 100 motifs represented in the AGRIS AtTFDB (http://arabidopsis.med.ohio-state.edu/AtTFDB). The Biostrings package was used to analyze the abundance of 100 motifs in groups of genes with significant eQTL compared to motif abundance in promoters of all analyzed genes using a Fisher's exact test ($p < 0.05$) with either zero or one mismatch (Ichihashi et al., 2014).

Dissection of eQTL to different stages and time of development at shoot apex

Differentially expressed genes with enriched expression in laser micro-dissected SAM/P0 vs. P1 samples or hand-dissected samples of the SAM + P0-P4 or P5 sampled over developmental time were obtained from Chitwood et al., 2015. Genes for which a differential expression call could be made (i.e., had enough reads and passed quality filters) were merged with detected eQTL using the merge() function in R (R Core Team, 2015). For bootstrapping, *cis*- and *trans*-regulated transcripts were analyzed separately. Merged gene expression patterns were randomly permuted (without replacement) using the sample()

function against bin identity. For each test, 10,000 permutations were sampled to count the times that a particular expression pattern was assigned to a bin more than the actual count. Resulting frequencies, representing a probability value, were multiple test adjusted using the Benjamini-Hochberg (Benjamini and Hochberg, 1995) method using p.adjust(). Those bins with multiple test adjusted probability values <0.05 were analyzed further using visualizations created with ggplot2 (Wickham, 2009).

Sequence submission

The quality filtered, barcode-sorted and trimmed short read dataset, which was used to get the normalized read counts and for differential gene expression analysis, was deposited to the NCBI Short Read Archive under accessions SRR1013035 - SRR1013343 (Bioproject accession SRP031491).

Supplemental files

Supplemental figures:

Supplemental Figure S1. Number of genes in the introgression region for an IL and the number of differentially expressed genes compared to cv. M82. Strong correlation was observed for differentially expressed genes in *cis* (A), whereas a weak correlation was observed for genes in *trans* (B).

Supplemental Figure S2. Histograms for differentially expressed genes for the ILs. Up-regulated and down-regulated genes for the ILs in *cis* (A), and in *trans* (B).

Supplemental Figure S3. Frequency and distribution of differentially expressed genes for the IL population at the introgression and the bin level. (A) Frequency of genes differentially expressed in one or more ILs. (B) Frequency of genes being regulated by one or more BINs/eQTL. (C) Distribution of genes under one or more eQTL regulation for the IL population on different chromosomes

Supplemental Figure S4. eQTL and the gene expression pattern they regulate. Each blue bar is a unique introgression. When the gene expression pattern of gene 1 is correlated with bin-A, then bin-A contains a *cis*-eQTL. When the gene expression pattern of gene 1 is correlated with bin-E then bin-E contains a *trans*-eQTL. When gene 2 has a *cis*-eQTL designated for bin-D and the gene expression pattern of gene 2 is also correlated with bin-B, then this secondary correlation is not designated as an eQTL, since these bins share overlapping introgression regions. When the gene expression pattern of gene 2 is correlated with bin-B and gene 2 does not have a *cis*-eQTL designated for bin-D, then bin-B is designated as a *trans*-eQTL. All eQTLs for genes that lie in the unassembled portion of the genome (not on any chromosome) cannot be designated as either *cis*- or *trans*- and are designated *chromo0*-eQTL.

Supplemental Figure S5. *Cis*- and *Trans*-eQTL. Histograms plotting the numbers of significant eQTL mapped to each bin across the 12 chromosomes of *S. lycopersicum* (M82). A. *cis*-eQTLs. B. *trans*-eQTLs.

Supplemental Figure S6. Boxplots of the normalized expression patterns for the three landmark modules. The relative expression of all genes found in each module for the 74 IL's. The y-axis is the relative expression of all eQTL for each IL. (A) Photosynthesis module. (B) Defense, metabolism, and signaling module. (C) Cysteine-type peptidase activity module. Asterisks represent ILs with a median expression significantly different from the module mean.

Supplemental Figure S7. Normalized expression of the leaf development module and leaf developmental genes within the mapping. (A) Boxplot of normalized expression pattern for the leaf development module. The 108 genes contained in the leaf development module and their relative median expression for each IL. Asterisks represent ILs with a median expression significantly different from the module mean. (B) LC+ list of leaf developmental genes overlaid on the Leaf Development and Photosynthesis modules. False colored orange, dark blue, and light blue respectively.

Supplemental Figure S8. eQTL regulation of gene expression patterns that correlate with leaf number.

A, B) Forty-two distinct modules identified by DBscan from the eQTL mapping generated by BH-SNE analysis. Modules enriched for genes with leaf development and photosynthesis GO terms are labeled in blue and green, respectively. Genes with expression patterns correlated with leaf number under simulated sun (A) and shade (B) are indicated by squares with positive correlations in red and negative correlations in yellow.

C, D) Genes with expression patterns correlated with leaf number under simulated sun (C) and shade (D) are shown connected to their respective eQTL with chords. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

Supplemental Figure S9. eQTL regulation of gene expression patterns that correlate with leaf complexity.

A, B) Forty-two distinct modules identified by DBscan from the eQTL mapping generated by BH-SNE analysis. Modules enriched for genes with leaf development and photosynthesis GO terms are labeled in blue and green, respectively. Genes with expression patterns correlated with leaf complexity under simulated sun (A) and shade (B) are indicated by squares with positive correlations in red and negative correlations in yellow.

C, D) Genes with expression patterns correlated with leaf complexity under simulated sun (C) and shade (D) are shown connected to their respective eQTL

with chords. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

Supplemental Figure S10. Distributions of introgressions from *S. pennellii* into *S. lycopersicum* cv. M82. Map of chromosomes 2, 10, and 11 showing the locations of the introgressions for BIL 033 and 128, as well as the overlapping IL regions, which define the bind (Modified from Chitwood et al., 2013).

Supplemental Figure S11. Tomato hypocotyl length under sun and shade treatments. M82 shows a typical shade response with a significantly longer hypocotyl in the shade (Δ of 7 mm). IL 10.1 and BIL 128, which share an overlapping introgression (**Supplemental Fig. BIL**), do not significantly respond to the shade treatments. The presence of a response in BIL 033 in combination with the shared introgression with BIL 128 on chromosome 2, indicates that the gene region responsible for the lack of shade response in BIL 128 is located in the introgression on chromosome 10. Bars represent means +/- SE with a minimal of N = 22 for each (ANOVA, $F_{7,182}$ = 44.6, p < 0.001). Letters indicate differences at the p < 0.05 significance level for Tukey pairwise tests.

Supplemental Figure S12. Quantification of genes with transgressive expression pattern in the IL population (A) Number of genes showing

transgressive expression for each IL along with their location in *cis* or *trans*. (B) Ratio of number of genes with transgressive expression to number of genes with *S. pennellii* like expression (transgressive to *S. pennellii*-like) for each IL**.**

<u>Supplemental tables:</u>

Supplemental Table I. Number of differentially expressed (DE) genes in *cis*, *trans*, and the total number of DE genes for the ILs along with number of genes in the introgression region.

Supplemental Table II. Correlation coefficients (Spearman's rho) for significant eQTLs. Divided into *trans*-, *cis-,* and *chromo0-,* then designated as up (positive slope) or down (negative slope) based on the correlation coefficients.

Supplemental Table III. GO enrichment and *cis* or *trans* regulation of the 42 identified modules. All 42 distinct modules are listed with the total number of genes present in each module. The GO enrichment (if one is present) is given for each module, along with whether that module is predominantly *cis* or *trans* regulated. Only nine of the forty two module show *trans* correlation, which includes the leaf development module.

Supplemental Table IV. Significant correlations between gene expression patterns and phenotypes. Bootstrapping analyses correlated gene expression

patterns across the 74 ILs with three phenotypes in under both sun and shade treatments. Genes with significant correlations that also have eQTL are listed.

Supplemental datasets:

Supplemental Dataset 1. List of Differentially Expressed Genes. List of significant (FDR < 0.05) differentially expressed genes for each Introgression line (IL) compared to cultivated parent *Solanum lycopersicum* cv. M82. For each IL, gene ID, log Fold Change (logFC), log Counts Per Million (logCPM), P-value, False Discovery Rate (FDR) as well as annotation of the differentially expressed genes are presented.

Supplemental Dataset 2. Frequency of differentially expressed genes among the ILs (Genes that show differential expression in at least one IL are listed).

Supplemental Dataset 3. All eQTL. All significant eQTL with AGI and ITAG annotations added.

Supplemental Dataset 4. Number of eQTL per bin.

Supplemental Dataset 5. Number of eQTL and the bin on which they reside for each of the landmark modules. The Photosynthesis, Defense and Cysteine-Type Peptidase Activity modules are listed with the bins on which their eQTL reside. The number eQTL per bin and the percentage of total eQTL within each module is listed.

Supplemental Dataset 6. Module Gene List. Listing of all genes within the Photosynthesis module with Gene ID and Description.

Supplemental Dataset 7. Module Gene List. Listing of all genes within the Defense module with Gene ID and Description.

Supplemental Dataset 8. Module Gene List. Listing of all genes within the Cysteine-type Peptidase module with Gene ID and Description.

Supplemental Dataset 9. GO Enrichment for DE Genes. Enriched GO and GOslim categories for the up- and down-regulated genes in each IL compared to cultivated parent *Solanum lycopersicum* cv. M82.

Supplemental Dataset 10. GO Enrichment for eQTL. GO and GO SLIM enrichment results for all the eQTL mapped to a bin, as well as the *cis*- and *trans*-eQTL separately.

Supplemental Dataset 11. Promoter Motif Enrichment for DE Genes. Enriched promoter motifs with no mismatch for the up- and down-regulated genes in each IL compared to cultivated parent *Solanum lycopersicum* cv. M82.

Supplemental Dataset 12. Enriched promoter motifs for *trans*-eQTLs mapped to each bin.

Supplemental Dataset 13. Differentially expressed leaf Developmental Genes. Frequency of differentially expressed literature-curated leaf-developmental genes among the ILs (Genes that show differential expression in at least one IL are listed).

Supplemental Dataset 14. Curated list of leaf developmental genes with eQTLs.

Supplemental Dataset 15. Literature-curated plus list of leaf development genes present in the leaf development modules. Nineteen genes from the literature-curated plus list (Ichihashi et al., 2014) were present within the leaf development module, representing approximately 3% of the curated genes and 11% of the curated genes found in the eQTL data set. The Gene ID, *Arabidopsis* orthologue, and a description for each gene is provided.

Supplemental Dataset 16. Literature-curated plus list of leaf development genes that are present in all modules. A total of 175 genes from the curated list plus of leaf developmental genes (Ichihashi et al., 2014) were present in the 5289 genes with significant eQTL. Matching gene number is the number of genes in the module, which were found in the curated list. The percent of curated

genes defines the percentage of matched genes in a module out of the total curated list plus. The percent total in eQTL represents the percentage of matched genes in a module out of the 175 present only in the eQTL. The Photosynthesis and leaf development modules contain the highest proportion of leaf development curated genes.

Supplemental Dataset 17. Module Gene List. Listing of all genes within the Leaf Development module with Gene ID and Description.

Supplemental Dataset 18: GO enrichment results for bins statically enriched for genes expressed spatio-temporally across tissues.

Supplemental Dataset 19: Total leaf complexity data for all the ILs under simulated sun and shade.

Supplemental Dataset 20: Hypocotyl length data for all the ILs under simulated sun and shade for two years.

Supplemental Dataset 21. Genes with Transgressive Expression. List of the genes showing transgressive expression in the IL population along with details of their expression and annotation.

Supplemental Dataset 22. GO Enrichment for Transgressive Expression. Enriched GO-categories for the genes showing transgressive expression in the IL population.

Acknowledgments

We thank Lauren R. Headland, Jason Kao and Paradee Thammapichai for help in generating plant materials. We also thank Mike Covington for his advice on bioinformatic analyses. We thank the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303), and computational resources/cyber infrastructure provided by the iPlant Collaborative (www.iplantcollaborative.org), funded by the National Science Foundation (Grant DBI-0735191).

Authors' contributions

DHC, JNM and NRS conceived and designed the experiments. AR, DHC, RK, LC, YI and KZ performed the experiments. AR, JMB and SDR analyzed the data. AR, JMB, SDR, DHC and JNM contributed reagents/materials/analysis tools. AR, JMB, SDR and NRS wrote the paper. All authors read and approved the final manuscript.

Figure legends:

Figure 1. Transcriptome profile of the tomato introgression lines. Differentially expressed genes for the ILs compared to cultivated parent M82. Y-axis shows all the tomato genes starting from the first gene on chromosome 1 to the last gene on chromosome 12, and X-axis depicts the individual ILs. Genes differentially expressed within the introgression regions (in *cis*) are shown as blue points and differentially expressed genes in *trans* (outside) the introgression region are shown as orange points.

Figure 2. *Cis*- and *Trans*-eQTL plotted by bin across the 12 chromosomes of *S. lycopersicum* cv. M82. A) Stacked bar graph showing the sum of the number of eQTL mapping to each bin. B) Dotplot showing each eQTL arranged vertically by bin and horizontally by the location of the gene expression pattern it regulates. Bins with the largest numbers of *trans*-eQTL (4D, 4E, 4F, 6B, 6C, 8A, 8B) are highlighted by green boxes. C) Map of chromosomes 4, 6, and 8 showing the overlapping IL regions, which define the bins (Modified from Chitwood et al., 2013). Bins with the largest numbers of *trans*-eQTL are indicated by green asterisks.

Figure 3. BH-SNE 2D mapping of eQTL. (A) Forty-two distinct modules identified by DBscan from the mapping generated by BH-SNE analysis. (B) The three modules defined as landmark modules: photosynthesis, defense and

cysteine-type peptidase activity and the leaf development module's position within the mapping. Modules are false colored.

Figure 4. Connections between eQTL and the genes they regulate. Each plot includes the genes with eQTL that were clustered together into a module based on expression patterns. A) Defense module. B) Photosynthesis module. C) Cysteine peptidase module. D) Leaf development module. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

Figure 5. Median expression values for leaf development and photosynthesis related modules and expression correlation. (A) The median expression values of a module for each IL are shown. A consistent negative correlation between photosynthesis and Leaf development transcript expression is evident across nearly all 74 ILs. Dashed lines indicate one significant deviation from the module mean expression. Filled areas represent the median expression of the leaf development module, while open areas indicate the photosynthesis module median expression. (B) Leaf development median expression versus photosynthesis median expression values for each IL show a distinct negative correlation with an adjusted R-squared value of 0.77 (calculated by linear regression in R).

Figure 6. Enriched expression patterns for genes genetically regulated by bins. A) Log fold change values (P1/SAM+P0) for previously identified differentially expressed genes with high expression in the SAM + P0 (magenta) vs. P1 (green). **B)** Scaled gene expression values for previously identified differentially expressed genes with increasing (red) and decreasing (blue) expression over developmental time in the SAM + P0-P4. **C)** Scaled gene expression values for previously identified differentially expressed genes with increasing (orange) and decreasing (purple) expression over developmental time in P5. **D)** Transcripts (y-axis) and bins (x-axis) showing the genetic regulation of gene expression (eQTL). Colors indicate SAM + P0 (magenta) and P1 (green) transcripts. Bins enriched for genetically regulating genes with specific expression patterns are indicated below with triangles. **E)** Same as in D), except showing genes with increasing (red) and decreasing (blue) expression over temporal time in the SAM + P0-P4. **F)** Same as in D), except showing genes with increasing (orange) and decreasing (purple) expression over temporal time in P5. Previously determined gene expression patterns are previously published (Chitwood et al., 2015).

Figure 7. eQTL regulation of gene expression patterns that correlate with hypocotyl length.

A) Forty-two distinct modules identified by DBscan from the eQTL mapping generated by BH-SNE analysis. Modules enriched for genes with leaf

development and photosynthesis GO terms are labeled in blue and green, respectively. Genes with expression patterns correlated with hypocotyl length under simulated shade are indicated by squares with positive correlations in red and negative correlations in yellow.

B) Genes with expression patterns correlated with hypocotyl length under simulated shade are shown connected to their respective eQTL with chords. I) The 12 tomato chromosomes in megabases. II) Colored boxes indicate the sizes of each bin. III) Black bars indicate the locations of the genes. IV) Chords connect eQTL to the genes whose expression patterns they regulate. Chords are colored by the chromosome location of the eQTL.

Parsed Citations

Abe M, Katsumata H, Komeda Y, Takahashi T (2003) Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development 130: 635-643
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Abramoff MD, Magalhaes PJ, Ram SJ (2004) Image Processing with ImageJ. Biophotonics International 11: 36-42
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y (2015) Genomic variation. Impact of regulatory variation from RNA to protein. Science 347: 664-667
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57: 289-300
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102: 1572-1577
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296: 752-755
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bushati N, Smith J, Briscoe J, Watkins C (2011) An intuitive graphical visualization technique for the interrogation of transcriptome data. Nucleic Acids Res 39: 7380-7389
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Carles CC, Choffnes-Inada D, Reville K, Lertpiriyapong K, Fletcher JC (2005) ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in Arabidopsis. Development 132: 897-911
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chen X, Hackett CA, Niks RE, Hedley PE, Booth C, Druka A, Marcel TC, Vels A, Bayer M, Milne I, Morris J, Ramsay L, Marshall D, Cardle L, Waugh R (2010) An eQTL analysis of partial resistance to Puccinia hordei in barley. PLoS One 5: e8598
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chitwood DH, Kumar R, Headland LR, Ranjan A, Covington MF, Ichihashi Y, Fulop D, Jimenez-Gomez JM, Peng J, Maloof JN, Sinha NR (2013) A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. Plant Cell 25: 2465-2481
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chitwood DH, Kumar R, Ranjan A, Pelletier JM, Townsley BT, Ichihashi Y, Martinez CC, Zumstein K, Harada JJ, Maloof JN, Sinha NR (2015) Light-Induced Indeterminacy Alters Shade-Avoiding Tomato Leaf Morphology. Plant Physiol 169: 2030-2047
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chitwood DH, Ranjan A, Kumar R, Ichihashi Y, Zumstein K, Headland LR, Ostria-Gallardo E, Aguilar-Martinez JA, Bush S, Carriedo L, Fulop D, Martinez CC, Peng J, Maloof JN, Sinha NR (2014) Resolving distinct genetic regulators of tomato leaf shape within a heteroblastic and ontogenetic context. Plant Cell 26: 3616-3629
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chitwood DH, Sinha NR (2013) A census of cells in time: quantitative genetics meets developmental biology. Curr Opin Plant Biol 16: 92-99
Pubmed: Author and Title
CrossRef: Author and Title

Google Scholar: Author Only Title Only Author and Title

Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nat Genet 38: 594-597

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cnops G, Jover-Gil S, Peters JL, Neyt P, De Block S, Robles P, Ponce MR, Gerats T, Micol JL, Van Lijsebettens M (2004) The rotunda2 mutants identify a role for the LEUNIG gene in vegetative leaf morphogenesis. J Exp Bot 55: 1529-1539

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cubillos FA, Coustham V, Loudet O (2012) Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. Curr Opin Plant Biol 15: 192-198

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cubillos FA, Yansouni J, Khalili H, Balzergue S, Elftieh S, Martin-Magniette ML, Serrand Y, Lepiniec L, Baud S, Dubreucq B, Renou JP, Camilleri C, Loudet O (2012) Expression variation in connected recombinant populations of Arabidopsis thaliana highlights distinct transcriptome architectures. BMC Genomics 13: 117

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in Arabidopsis. Genetics 172: 1155-1164

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Diaz-Mendoza M, Velasco-Arroyo B, Gonzalez-Melendi P, Martinez M, Diaz I (2014) C1A cysteine protease-cystatin interactions in leaf senescence. J Exp Bot 65: 3825-3833

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Druka A, Potokina E, Luo Z, Jiang N, Chen X, Kearsey M, Waugh R (2010) Expression quantitative trait loci analysis in plants. Plant Biotechnol J 8: 10-27

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Eshed Y, Zamir 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

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305: 1786-1789

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fukao T, Xu K, Ronald PC, Bailey-Serres J (2006) A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. Plant Cell 18: 2021-2034

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fulop D, Ranjan A, Ofner I, Covington MF, Chitwood DH, West D, Ichihashi Y, Headland L, Zamir D, Maloof JN Sinha NR A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. Plant Physiology In Review.

Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, Matasci N, Wang L, Hanlon M, Lenards A, Muir A, Merchant N, Lowry S, Mock S, Helmke M, Kubach A, Narro M, Hopkins N, Micklos D, Hilgert U, Gonzales M, Jordan C, Skidmore E, Dooley R, Cazes J, McLay R, Lu Z, Pasternak S, Koesterke L, Piel WH, Grene R, Noutsos C, Gendler K, Feng X, Tang C, Lent M, Kim SJ, Kvilekval K, Manjunath BS, Tannen V, Stamatakis A, Sanderson M, Welch SM, Cranston KA, Soltis P, Soltis D, O'Meara B, Ane C, Brutnell T, Kleibenstein DJ, White JW, Leebens-Mack J, Donoghue MJ, Spalding EP, Vision TJ, Myers CR, Lowenthal D, Enquist BJ, Boyle B, Akoglu A, Andrews G, Ram S, Ware D, Stein L, Stanzione D (2011) The iPlant Collaborative: Cyberinfrastructure for Plant Biology. Front Plant Sci 2: 34

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hammond JP, Mayes S, Bowen HC, Graham NS, Hayden RM, Love CG, Spracklen WP, Wang J, Welham SJ, White PJ, King GJ, Broadley MR (2011) Regulatory hotspots are associated with plant gene expression under varying soil phosphorus supply in Brassica rapa. Plant Physiol 156: 1230-1241
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ (2008) Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in Senecio. Philos Trans R Soc Lond B Biol Sci 363: 3055-3069
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hennig C (2014) FPC: Flexible procedures for Clustering. R Package Version: 2.1-9
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Holloway B, Li B (2010) Expression QTLs: applications for crop improvement. Molecular Breeding 26: 381-391
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Holloway B, Luck S, Beatty M, Rafalski JA, Li B (2011) Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics 12: 336
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Holtan HE, Hake S (2003) Quantitative trait locus analysis of leaf dissection in tomato using Lycopersicon pennellii segmental introgression lines. Genetics 165: 1541-1550
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ichihashi Y, Aguilar-Martinez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Peng J, Maloof JN, Sinha NR (2014) Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proc Natl Acad Sci U S A 111: E2616-2621
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17: 388-391
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci U S A 104: 1708-1713
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kliebenstein D (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. Annu Rev Plant Biol 60: 93-114
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kliebenstein DJ, West MA, van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St Clair DA (2006) Genomic survey of gene expression diversity in Arabidopsis thaliana. Genetics 172: 1179-1189
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci U S A 110: E2655-2662
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Krijthe J (2014) Rtsne: T-distributed Stochastic Neighbor Embedding using Barnes-Hut implementation. R Package Version: 0.9

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. Proc Natl Acad Sci U S A 100 Suppl 2: 14587-14592
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR (2012) A High-Throughput Method for Illumina RNA-Seq Library Preparation. Front Plant Sci 3: 202
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liu YS, Zamir D (1999) Second generation L. pennellii introgression lines and the concept of bin mapping. Tomato Genet. Coop. Rep. 49: 26-30
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Longair MH, Baker DA, Armstrong JD (2011) Simple Neurite Tracer: open source software for reconstruction, visualization and analysis of neuronal processes. Bioinformatics 27: 2453-2454
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Moyle LC (2008) Ecological and evolutionary genomics in the wild tomatoes (Solanum sect. Lycopersicon). Evolution 62: 2995-3013
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Muir CD, Pease JB, Moyle LC (2014) Quantitative genetic analysis indicates natural selection on leaf phenotypes across wild tomato species (Solanum sect. Lycopersicon; Solanaceae). Genetics 198: 1629-1643
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Muller NA, Wijnen CL, Srinivasan A, Ryngajllo M, Ofner I, Lin T, Ranjan A, West D, Maloof JN, Sinha NR, Huang S, Zamir D, Jimenez-Gomez JM (2016) Domestication selected for deceleration of the circadian clock in cultivated tomato. Nat Genet 48: 89-93
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464: 768-772
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing)
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ranjan A, Ichihashi Y, Sinha NR (2012) The tomato genome: implications for plant breeding, genomics and evolution. Genome Biol 13: 167
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Riddle NC, Birchler JA (2003) Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Genet 19: 597-600
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity (Edinb) 83 (Pt 4): 363-372
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 10: 71-

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297-302
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sharlach M, Dahlbeck D, Liu L, Chiu J, Jimenez-Gomez JM, Kimura S, Koenig D, Maloof JN, Sinha N, Minsavage GV, Jones JB, Stall RE, Staskawicz BJ (2013) Fine genetic mapping of RXopJ4, a bacterial spot disease resistance locus from Solanum pennellii LA716. Theor Appl Genet 126: 601-609
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Shivaprasad PV, Dunn RM, Santos BA, Bassett A, Baulcombe DC (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. EMBO J 31: 257-266
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Svistoonoff S, Creff A, Reymond M, Sigoillot-Claude C, Ricaud L, Blanchet A, Nussaume L, Desnos T (2007) Root tip contact with low-phosphate media reprograms plant root architecture. Nat Genet 39: 792-796
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

van der Maaten L (2013) Barnes-Hut-SNE. arXiv:1301.3342[cs.LG]
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

van der Maaten L, Hinton G (2008) Visualizing Data using t-SNE. Journal of Machine Learning Research 9: 2579-2605
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172: 507-517
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci U S A 102: 2460-2465
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175: 1441-1450
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer New York
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu F, Guo W, Xu W, Wei Y, Wang R (2009) Leaf morphology correlates with water and light availability: What consequences for simple and compound leaves?. Progress in Natural Science 19(12): 1789-1798
Pubmed: Author and Title

CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442: 705-708
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11: R14
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Yu D, Chen C, Chen Z (2001) Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. Plant Cell 13: 1527-1540
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhang L, Fetch T, Nirmala J, Schmierer D, Brueggeman R, Steffenson B, Kleinhofs A (2006) Rpr1, a gene required for Rpg1-dependent resistance to stem rust in barley. Theor Appl Genet 113: 847-855
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res 21: 725-733
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title