A Developmental Transcriptional Network for Maize Defines Coexpression Modules\[1][C][W][OA]

Gregory S. Downs, Yong-Mei Bi, Joseph Colasanti, Wenqing Wu, Xi Chen, Tong Zhu, Steven J. Rothstein, and Lewis N. Lukens*

Department of Plant Agriculture (G.S.D., L.N.L.) and Department of Molecular and Cellular Biology (Y.-M.B., J.C., W.W., S.J.R.), University of Guelph, Guelph, Ontario, Canada N1G2W1; and Syngenta Biotechnology, Inc., Research Triangle Park, North Carolina 27709 (X.C., T.Z.)

Here, we present a genome-wide overview of transcriptional circuits in the agriculturally significant crop species maize (Zea mays). We examined transcript abundance data at 50 developmental stages, from embryogenesis to senescence, for 34,876 gene models and classified genes into 24 robust coexpression modules. Modules were strongly associated with tissue types and related biological processes. Sixteen of the 24 modules (67%) have preferential transcript abundance within specific tissues. One-third of modules had an absence of gene expression in specific tissues. Genes within a number of modules also correlated with the developmental age of tissues. Coexpression of genes is likely due to transcriptional control. For a number of modules, key genes involved in transcriptional control have expression profiles that mimic the expression profiles of module genes, although the expression of transcriptional control genes is not unusually representative of module gene expression. Known regulatory motifs are enriched in several modules. Finally, of the 13 network modules with more than 200 genes, three contain genes that are notably clustered ($P < 0.05$) within the genome. This work, based on a carefully selected set of major tissues representing diverse stages of maize development, demonstrates the remarkable power of transcript-level coexpression networks to identify underlying biological processes and their molecular components.

Systems biology approaches recently have begun to elucidate the patterns of transcriptome organization. In contrast to analyses that compare whole transcriptomes of samples and those that compare mean levels of gene expression differences between samples, the systems biology strategy integrates expression patterns of single genes to infer their common biological function. Genes with coordinated expression across samples are hypothesized to be coregulated in response to external and internal cues and to be regulated by similar transcription factors (Moreno-Risueno et al., 2010). Inferring gene regulatory networks from transcriptome data and subsequently testing the attributes of the network provides a system-wide view of developmental processes.

A number of studies have pooled diverse assortments of publicly available microarray data to identify clusters of plant genes with shared patterns of expression (Fierro et al., 2008; Ficklin et al., 2010; Mochida et al., 2011). A number of modules are conserved across species (Ficklin and Feltus, 2011; Movahedi et al., 2011; Mutwil et al., 2011). For example, modules associated with drought stress responses and cellulose biogenesis are common to barley (Hordeum vulgare), Arabidopsis (Arabidopsis thaliana), and Brachypodium distachyon (Mochida et al., 2011).

The great functional and morphological variation in plant tissue types arises from the differential regulation of a finite set of genomic transcripts. Microarray technology has been used to compare gene transcript abundances between different tissues (Ma et al., 2005; Schmid et al., 2005; Benedito et al., 2008; Jiao et al., 2009; Sekhon et al., 2011). These studies have identified genes that are transcribed in specific organs and examined the relationships among tissue expression patterns using principal component transformation. These studies have also noted that similar tissues had more highly correlated gene transcript abundances than less similar tissues (e.g., the correlation coefficient of two developmental stages of leaf tissue is greater than the correlation coefficient of leaf with another tissue).

Here, we have constructed a developmental gene expression network from microarray transcriptome profiles of 50 maize (Zea mays) tissues across different stages of development and identified modules of putative coregulated genes within this network. We characterized the attributes of modules to begin to understand transcriptome organization. Specifically, we investigated whether network modules are associated with specific tissue types and are enriched for specific biological

---

1 This work was supported by the Ontario Research Fund and the Natural Sciences and Engineering Research Council of Canada.

* Corresponding author; e-mail llukens@uoguelph.ca.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Lewis N. Lukens (llukens@uoguelph.ca).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[C] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.112.213231
processes. Furthermore, we determined whether modules are specifically excluded from tissues and if modules reflect developmentally responsive processes. Moreover, we investigated the centrality of transcription factors within modules and if modules share common cis-regulatory motifs. Finally, we determined whether modules contain genes that are clustered within the genome. This work explores the gene expression network throughout maize development for an inbred genotype grown under controlled conditions and describes the remarkably discrete functionalities of modules within the network.

RESULTS
Gene Networks of the Maize Developmental Transcriptome

We set out to investigate the organization of the maize transcriptome throughout development by analyzing a microarray data set generated from three biological replicates of 50 tissue types (Table I). Samples were derived from many tissues, from early embryo to senescence-stage leaves, including anthers, cob, ear, embryo, endosperm, husk, leaf, ovule, pericarp, root, silk, stalk, and tassel. Several of these tissues, including ear, leaf, and tassel, were sampled at multiple stages of development (Table I; Fig. 1). All processed RNAs were hybridized to a custom microarray with 82,661 probe sets and 1,322,576 probes. Of these, 55,672 probe sets were expressed in at least one tissue (Fig. 2) and 33,664 mapped to the filtered gene set of the maize gene models, release 4a.53 (Schnable et al., 2009). An additional 9,919 probe sets that were not annotated as genes mapped to the maize genome, and 12,089 probe sets did not match the genome using our criteria (Fig. 2). To ensure the highest level of data quality, among the probe sets that matched to the genome, we removed redundant and nonspecific probes prior to data analysis. In the end, we examined 34,876 probe sets (Fig. 2; Supplemental Table S1). For clarity, we refer to these probe sets as genes.

We clustered all sample transcriptome profiles using the flashClust function of WGCNA (Langfelder and Horvath, 2008) to obtain an overview of transcriptome relationships. With few exceptions, biological replicate arrays cluster within a group containing only the replicates of the tissue at that stage, and arrays from the same tissue cluster together (Supplemental Fig. S1). Figure 3 shows a dendrogram of all 50 tissues constructed from average transcript abundances across replicate arrays. Distinct groups contain leaf, root, seed, and silk expression profiles. Transcriptomes from tissues harvested at different developmental stages largely cluster together, as do tissues that have strong developmental similarity, such as the V7 tassel and V7 ear. Nonetheless, some groups contain mixed tissues or do not contain all arrays from a tissue type. For example, the R1 stalk is grouped with leaf transcriptomes, and the prephotosynthetic VE leaf did not cluster with other leaves (Fig. 3). The innermost husk, a modified leaf, clustered with inflorescence tissues (cob, ovule, silk, and tassel) rather than leaves (Fig. 3).

We constructed a weighted gene coexpression network with the R software WGCNA by transforming the 34,876 genes’ pairwise Pearson correlation coefficients into a weighted adjacency matrix (Langfelder and Horvath, 2008). We created a signed network, which allows modules to contain both positively and negatively correlated genes, since transcripts involved in one process may be up- or down-regulated. The topological overlap measure (TO; Li and Horvath, 2007) was used to transform the adjacency matrix into a coexpression distance matrix. Genes were clustered hierarchically, and a dynamic tree-cutting algorithm cut the dendrogram and defined 49 modules. Gene module assignments are given in Supplemental Table S2. The modules range in size from 30 to 4,370 genes (mean, 712; median, 123). To validate modules, we compared the mean TO value for each module with a distribution of mean TO values for 50,000 iterations of modules composed of a randomly selected group of genes (Ravasz et al., 2002; Yip and Horvath, 2007). We focused on the 24 of the 49 modules that were validated as significant ($P < 0.05$; Supplemental Table S3). These 24 modules contain 30,768 genes. Module eigengenes (MEs) were calculated for each module as the first principal component of the gene expression matrix for the module, and these can be considered as a vector of gene expression values characteristic of the module. Correlations between the MEs for each module indicate that most modules have an eigengene with similar correlation patterns as the eigengene from one or more other modules (Supplemental Fig. S2). The use of more permissive criteria for module identification could group these modules together, but subsequent analyses revealed that they have distinct attributes.

Many Modules Correlate with Specific Tissue Types

We investigated whether each module’s eigengene had significantly higher expression in specific tissues relative to all other tissues. Tissue-specific modules may also contain genes with low expression in one tissue type relative to others. Sixteen of the 24 modules (67%) are moderately to highly correlated with tissue type ($r > 0.4$; Fig. 4). One or more modules are correlated with anthers, ear, embryo, endosperm, leaf, pericarp, root, and tassel. No module is correlated with cob, floret, husk, ovule, stalk, and silk (Fig. 4). Of the 24 modules, eight had eigengenes that are moderately to highly negatively correlated ($r < -0.4$) with anthers, endosperm, leaf, root, and stalk (Fig. 4). Two of these eight modules also had positive correlations with tissue types, so only two of the 24 modules were not associated with a specific tissue type.

To investigate the robustness of tissue-associated modules, we cross referenced gene modules with the list of tissue-specific genes reported by Sekhon et al. (2011) in a survey of maize transcriptomes. Their study...
Table 1. Description of developmental stages and tissues sampled for microarray analyses

All samples were collected around noon. The VE stage is germination and emergence. The Vn leaf stage refers to when the collar of the nth leaf is visible.

| Developmental Stage<sup>a</sup> | Tissue/Organ | Tissue Group | No. of Visible Leaves at Sampling | Details of the Harvested Sample |
|--------------------------------|--------------|--------------|----------------------------------|-------------------------------|
| VE                             | Leaf         | Leaf         | 0                                | Coleoptile                    |
| VE                             | Seminal root | Root         | 0                                | Root                          |
| V1                             | Leaf         | Root         | 2                                | First and second leaves       |
| V1                             | Seminal root | Root         | 2                                | Root                          |
| V2                             | Seminal root | Root         | 4                                | Seminal root                  |
| V2                             | Nodal root   | Root         | 4                                | Nodal root                    |
| V2                             | Stalk        | Stalk        | 4                                | Stalk                         |
| V2                             | Leaf         | Leaf         | 4                                | Leaf (actively growing leaf: fourth leaf) |
| V4                             | Tassel       | Tassel       | 6                                | 1-mm tassel meristem and 1-mm uppermost stem below tassel |
| V5                             | Seminal root | Root         | 8                                | Seminal root                  |
| V5                             | Nodal root   | Root         | 8                                | Nodal root                    |
| V5                             | Stalk        | Stalk        | 8                                | Stalk below tassel (2 cm)     |
| V5                             | Leaf         | Leaf         | 8                                | Leaf (actively growing leaf: eighth leaf, 15 cm including tip) |
| V5                             | Tassel       | Tassel       | 8                                | Tassel 3–5 mm                 |
| V7                             | Ear          | Ear          | 12                               | Top ear shoot                 |
| V7                             | Tassel       | Tassel       | 12                               | Tassel 2 cm                   |
| V8–V9                          | Tassel       | Tassel       | 13–14                            | Tassel 12–14 cm               |
| V8–V9                          | Ear          | Ear          | 13–14                            | Top ear 3–5 mm                |
| V10–V11                        | Tassel       | Tassel       | 15–16                            | Top 10 cm of tassel (approximately 20 cm) |
| V10–V11                        | Ear          | Ear          | 15–16                            | Top ear 1–1.5 cm              |
| V13–V15                        | Tassel       | Tassel       | 15–16                            | Spikelet of tassel (approximately 22 cm) |
| V13–V15                        | Ear          | Ear          | 15–16                            | Top ear 3–3.5 cm              |
| V15–V16                        | Floret       | Floret       | 15–16                            | Top ear (5 cm) floret         |
| V15–V16                        | Cob          | Cob          | 15–16                            | Top ear (5 cm) cob            |
| V15–V16                        | Silk         | Silk         | 15–16                            | Top ear (5 cm) silk           |
| V15–V16                        | Tassel       | Tassel       | 15–16                            | Spikelet of tassel (top 10 cm) |
| VT                             | Anthers      | Anthers      | 15–16                            | Anther                        |
| R1                             | Ovule        | Ovule        | 15–16                            | R1, ovule of top ear          |
| R1                             | Cob          | Cob          | 15–16                            | R1, cob of top ear            |
| R1                             | Silk         | Silk         | 15–16                            | R1, silk of top ear           |
| R1                             | Husk         | Husk         | 15–16                            | R1, most inner husk of top ear |
| R1                             | Leaf         | Leaf         | 15–16                            | R1, 15-cm tip of second leaf above top ear |
| R1                             | Nodal root   | Root         | 15–16                            | R1, adult root                |
| R1                             | Stalk        | Stalk        | 15–16                            | R1, 15-cm stalk below tassel  |
| 5 DAP                          | Ovule        | Ovule        | 15–16                            | Ovule of top ear              |
| 5 DAP                          | Cob          | Cob          | 15–16                            | Cob of top ear                |
| 10 DAP                         | Embryo       | Embryo       | 15–16                            | Embryo of top ear             |
| 10 DAP                         | Endosperm    | Endosperm    | 15–16                            | Endosperm of top ear          |
| 10 DAP                         | Leaf         | Leaf         | 15–16                            | 15-cm tip of second leaf above top ear |
| 17 DAP                         | Embryo       | Embryo       | 15–16                            | Embryo of top ear             |
| 17 DAP                         | Endosperm    | Endosperm    | 15–16                            | Endosperm of top ear          |
| 17 DAP                         | Pericarp     | Pericarp     | 15–16                            | Pericarp of top ear           |
| 17 DAP                         | Leaf         | Leaf         | 15–16                            | 15-cm tip of second leaf above top ear |
| 24 DAP                         | Leaf         | Leaf         | 15–16                            | 15-cm tip of second leaf above top ear |
| 24 DAP                         | Nodal root   | Root         | 15–16                            | Nodal root                    |
| 24 DAP                         | Pericarp     | Pericarp     | 15–16                            | Pericarp of top ear           |
| 24 DAP                         | Embryo       | Embryo       | 15–16                            | Embryo of top ear             |
| 24 DAP                         | Endosperm    | Endosperm    | 15–16                            | Endosperm of top ear          |
| 31 DAP                         | Leaf         | Leaf         | 15–16                            | 15-cm tip of second leaf above top ear |
| 31 DAP                         | Embryo       | Embryo       | 15–16                            | Embryo of top ear             |

<sup>a</sup> We also measured days after seeding for developmental stages (VE, 4; V1, 7; V2, 14; V4, 17; V5, 27; V7, 34; V8 to V9, 41; V10, 48; V15, 51; V16, 54; VT, 54). R1, 1 d before pollination. All DAP samples were harvested based on the date of pollination.

reported 863 tissue-specific genes, of which we could trace 276 to our experiment, due to differences between microarray platforms and our stringent oligonucleotide-to-gene mapping criteria (Fig. 2). Remarkably, 75% (206) of the genes identified by Sekhon et al. (2011) as tissue specific were within network modules that were significantly correlated with the same tissue (Table II; Supplemental Table S4). The 70 (25%) tissue-specific genes reported by Sekhon et al. (2011) that did not map to a module may be explained, in part, by
Figure 1. Images of selected tissues at the time of sampling. Images of all 50 tissues can be found in Supplemental Figure S6. A, VE leaf and root. B, V1 leaf and root. C, V2 leaf, stalk, seminal, and nodal root. D, V4 tassel. E, V4 tassel primordium. F, V5 tassel. G, V8 tassel and top ear shoot. H, V10 tassel and top ear shoot. I, V13 top ear. J, V15 silk. K, V15 cob and floret. L, V15 tassel. M, VT anthers and pollen. N, R1 root. O, R1 stalk. P, R1 leaf. Q, 10 DAP embryo and endosperm. R, 17 DAP milky endosperm. S, 24 DAP leaf. T, 24 DAP root. U, 24 DAP embryo, endosperm, and pericarp. V, 31 DAP leaf. W, 31 DAP embryo, endosperm, and pericarp.
environmental effects that altered transcription profiles between experiments. In our study, all plants were grown in a controlled environment, whereas Sekhon et al. (2011) harvested young plants from the greenhouse and older plants from the field.

Modules Are Highly Enriched for Biological Processes

Tissue-specific modules are often characterized by specific biological functionalities. We used the topGO package in R to identify Gene Ontology (GO) terms that appear in modules more frequently than expected by chance. Nine of the 16 modules that positively correlate with specific tissues are highly enriched with GO biological processes (Fisher’s exact test, $P < 0.0001$; Supplemental Table S5). These modules are associated with anther, ear, embryo, endosperm, leaf, root, pericarp, and tassel. The overrepresented GO terms are often consistent with known tissue attributes. Module Zm_mod12 is associated with anthers (1,010 genes, tissue correlation $r = 0.55$, $P = 4 \times 10^{-5}$) and is overrepresented by genes related to ‘sexual reproduction’ (GO:0019953; Supplemental Table S5). Module Zm_mod11 (1,109 genes, $r = 0.96$, $P = 2 \times 10^{-29}$) is correlated with roots and has an overrepresentation of genes involved in ‘response to oxidative stress’ (GO:0006979), ‘oxidation reduction’ (GO:0055114), and ‘hydrogen peroxide catabolic process’ (GO:0042744; Supplemental Table S5). A number of tissues are associated with different modules, suggesting that functionally distinct modules can share tissue specificity. Leaves have two modules with significant GO terms: Zm_mod06 and Zm_mod07. Zm_mod06 (2,069 genes, $r = 0.83$, $P = 6 \times 10^{-14}$) is enriched for six GO biological process terms relating to photosynthesis (Supplemental Table S5). This module contains almost every enzyme in three pathways important for chlorophyll biosynthesis: biosynthesis of chlorophyllide $a$, biosynthesis of phytol diphosphate, and biosynthesis of chlorophyll $a$ (Supplemental Fig. S3). Phytol diphosphate is the source of the phytol chain in
chlorophylls, and chlorophyllide a is an intermediate compound in chlorophyll biosynthesis. The biological process “glycolipid transport” (GO:0046836) is over-represented within module Zm_mod07 (1,808 genes, $r = 0.44$, $P = 1 \times 10^{-3}$; Supplemental Table S5). Together, the annotations provide strong evidence of the functionalities of many tissue-specific modules.

Four of the eight modules negatively correlated with a tissue type are significantly enriched for GO biological process terms. “Glycine betaine biosynthetic process” (GO:0031456) was enriched in module Zm_mod02 (4,081 genes, $r = -0.99$, $P = 7 \times 10^{-45}$), which is negatively associated with anther tissue. Zm_mod01 is also negatively associated with anthers (4,370 genes, $r = -0.47$, $P = 6 \times 10^{-5}$) and was enriched for “translation” (GO:0006412) and “ribosome biogenesis” (GO:0003743; Supplemental Table S5). Genes within modules that were neither positively nor negatively correlated with a tissue did not have significant enrichment for any GO term (Fig. 4; Supplemental Table S5).

Figure 3. Average transcript abundances from 50 arrays (three biological replicates of each tissue/stage) were clustered using the flashClust module from WGCNA. Developmental stages of the same tissue tend to cluster together. For definitions of developmental stages, see Table I.

Modules Capture Developmental Stages within Tissues

The relationships of leaf transcriptomes and to a smaller degree embryo transcriptomes reflected the developmental age of the samples (Fig. 3). Leaf samples include two samples of juvenile leaves and one sample of adult leaf prior to flowering (V1, V2, and V5; Table I). We also sampled the second leaf above the top ear 1 d before pollination and at four time points following pollination (10, 17, 24, and 31 d after pollination [DAP]). A number of modules are correlated or anticorrelated either with green leaves sampled prior to anthesis/silking (V1, V2, and V5) or with green
leaves sampled after flowering (R1 and 10, 17, 24, and 31 DAP; Fig. 5). Other modules, including Zm_mod06, enriched for genes related to photosynthesis, have correlations with all leaf samples. Unlike leaves before and after flowering, no modules differentiated the two juvenile-stage leaves from the single adult leaf. A heat map of ME correlations with individual tissues is shown in Supplemental Figure S4.

The ME for Zm_mod13 (691 genes) is very strongly correlated with embryo age (10, 17, 24, or 31 DAP; \( r = 0.96, P = 4 \times 10^{-27} \)). This module is enriched for the GO term “embryonic development” (GO:0009790; Supplemental Table S5). Chloroplast THIAMINE THIAZOLE SYNTHASE2 (THI1-2; GRMZM2G074097), an enzyme in the thiamine biosynthetic pathway, is within the Zm_mod13 module. The transcript abundance

---

### Figure 4. A heat map of ME and tissue correlations. Boxes contain Pearson correlation coefficients and their associated \( P \) values. A strong positive correlation (red) indicates that the ME has higher expression in the given tissue relative to all other tissues. A strong negative correlation (blue) indicates low expression in the given tissue relative to all other tissues. Tissues were classified into groups as described in Table I.

![Heat map of ME and tissue correlations](image)

### Table I

| MEZm_mod13 | MEZm_mod18 | MEZm_mod23 | MEZm_mod02 | MEZm_mod19 | MEZm_mod03 | MEZm_mod01 | MEZm_mod04 | MEZm_mod21 | MEZm_mod22 | MEZm_mod07 | MEZm_mod10 | MEZm_mod11 | MEZm_mod20 | MEZm_mod24 | MEZm_mod17 | MEZm_mod08 | MEZm_mod09 | MEZm_mod12 | MEZm_mod15 | MEZm_mod06 | MEZm_mod14 |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| -0.13 (1)  | -0.06 (1)  | -0.04 (1)  | -0.02 (1)  | -0.11 (1)  | -0.08 (1)  | -0.10 (1)  | -0.11 (1)  | -0.13 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  |
| -0.09 (1)  | -0.06 (1)  | -0.04 (1)  | -0.02 (1)  | -0.11 (1)  | -0.08 (1)  | -0.10 (1)  | -0.11 (1)  | -0.13 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  |
| 0.08 (1)   | 0.06 (1)   | 0.04 (1)   | 0.02 (1)   | -0.11 (1)  | -0.08 (1)  | -0.10 (1)  | -0.11 (1)  | -0.13 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  | -0.10 (1)  |

**Anthers, Cobs, Ear, Embryo, Endosperm, Floret, Husk, Leaf, Ovule, Pericarp, Root, Silk, Stalk, Tassel**
Expression of Transcription-Related Genes within Modules

The transcription of specific sets of genes triggers molecular cascades that determine developmental fates (Kaufmann et al., 2010). Modules contain transcription-related genes that are correlated with MEs. For example, the leaf-associated module Zm_mod14 (315 genes, \( r = 0.61, P = 2 \times 10^{-5} \)) contains \( \sigma \) factor SIG2A of RNA polymerase (GRMZM2G143392), with an eigengene correlation, or module membership (MM), of 0.74 (Supplemental Table S2). The \( \sigma \) factor is a nucleoside-encoded gene whose product is transported to the chloroplast, where it facilitates plastid RNA polymerase binding to chloroplastic promoters, predominantly in leaf tissue containing mature chloroplasts (Lysenko, 2007). The eigengene of the endosperm-associated module Zm_mod09 (1,403 genes, \( r = 0.69, P = 3 \times 10^{-5} \)) is highly correlated both with GRMZM2G118205, which encodes a protein similar to the Polycomb group factor SIG2A of RNA polymerase, and with CC(A/G)CCC, which is overrepresented in the upstream sequences of module genes. Six of the 24 modules are enriched for 10 motifs (Table III). An interesting motif is CC(A/G)CCC, which is overrepresented in Zm_mod01 and Zm_mod06. These modules are negatively and positively correlated with leaf tissue, respectively. The MITOCHONDRIAL NUCLEOID FACTOR1 transcription factor is associated with CC(A/G)CCC and initiates the transcription of PPC1 (a C4-type phosphoenolpyruvate carboxylase) in maize mesophyll cells exposed to light (Morishima, 1998). While plant promoters are often described as compact, there are exceptions to this general rule. An examination of 1-, 1.5-, and 2-kb upstream sequences identified some novel overrepresented promoter motifs within modules. Of the 10 motifs identified as overrepresented in the 500-bp upstream regulatory sequences of certain modules, six were shared when 1 kb of upstream sequence was examined, and four were shared when 2 kb was examined (Supplemental Table S7).

Module Genes Are Significantly Clustered within the Genome

We hypothesized that genes related to transcriptional control would have expression patterns more highly similar to each module’s eigengene than to other genes. In 10 modules, transcription-related genes have one of the top five ranks when genes within the module are sorted by descending MM (Supplemental Table S6). Nonetheless, the top rank of the transcription-related gene is not significantly higher than expected for other genes within any module (\( P < 0.01; \) Supplemental Table S6). We also evaluated whether genes classified with GO terms related to transcription have on average higher topological overlap scores than expected. The observed connectivity of transcription-related genes was not significantly greater than expected for any module (data not shown).

Coordinated regulation of module genes may be due in part to shared transcription factor regulation. Transcription factors bind to specific promoter motifs upstream of the transcription initiation site. We used Fisher’s exact test to determine if any one of 106 previously reported maize regulatory motifs are overrepresented in the upstream sequences of module genes. Six of the 24 modules are enriched for 10 motifs (Table III). An interesting motif is CC(A/G)CCC, which is overrepresented in Zm_mod01 and Zm_mod06. These modules are negatively and positively correlated with leaf tissue, respectively. The MITOCHONDRIAL NUCLEOID FACTOR1 transcription factor is associated with CC(A/G)CCC and initiates the transcription of PPC1 (a C4-type phosphoenolpyruvate carboxylase) in maize mesophyll cells exposed to light (Morishima, 1998). While plant promoters are often described as compact, there are exceptions to this general rule. An examination of 1-, 1.5-, and 2-kb upstream sequences identified some novel overrepresented promoter motifs within modules. Of the 10 motifs identified as overrepresented in the 500-bp upstream regulatory sequences of certain modules, six were shared when 1 kb of upstream sequence was examined, and four were shared when 2 kb was examined (Supplemental Table S7).

Table II. Numbers of tissue-specific genes reported by Sekhon et al. (2011) that are classified into tissue-specific modules

| Tissue | Sekhon Genes | Module Genes | No. in Agreement | Proportion |
|--------|--------------|--------------|------------------|------------|
| Cob    | 4            | 0            | 0                | 0          |
| Embryo | 48           | 23           | 15               | 0.65       |
| Endosperm | 168      | 48           | 35               | 0.73       |
| Internode | 12        | 3            | 0                | 0.00       |
| Leaf   | 334          | 109          | 92               | 0.84       |
| Root   | 151          | 39           | 29               | 0.74       |
| Silk   | 12           | 5            | 0                | 0.00       |
| Tassel | 134          | 49           | 35               | 0.71       |
| Total  | 863          | 276          | 206              | 0.75       |

\(^{a}\) Sekhon et al. (2011) reported tissue-specific genes in these eight tissues. \(^{b}\) Count of genes reported by Sekhon et al. (2011) in each tissue. \(^{c}\) Count of genes in our study that were mapped to tissue-specific genes in Sekhon et al. (2011). \(^{d}\) Number of genes in our study that were in tissue-specific modules that corresponded to tissue in Sekhon et al. (2011).
et al., 2001; Lercher et al., 2002; Zhan et al., 2006). We investigated whether module genes tended to have nonrandom genomic positions. Under the null hypothesis, the genomic position of a module’s gene in the genome is independent of all other module genes, and the distribution of module genes per chromosomal interval is expected to follow a Poisson distribution with an equal mean and variance. For the 13 modules that contain more than 200 genes (Table IV), we calculated the module dispersion score: the average number of module member genes within a 300-kb segment of chromosomal DNA divided by the variance. Three of the 12 modules have genes that are significantly (P < 0.05) clustered (Table IV), although the mean number of genes per interval is lower than the variance for every module. Two of these, Zm_mod10 (1,147 genes, r = 0.53, P = 9 × 10⁻²) and Zm_mod11 (described above), are positively associated with roots (Fig. 4). The other, Zm_mod01, is negatively associated with leaf tissue.

DISCUSSION

The concept of modularity in transcriptome analyses is that transcript abundance data can be partitioned into a collection of discrete and informative modules. Each module is self-contained and presumably functions to perform a distinct task separate from the tasks of other modules. At the same time, the components of a transcriptome are dynamically interconnected, so a complex web of interactions defines transcript patterns and abundances. To investigate the interconnections and modular structure of plant developmental transcriptomes, we constructed a transcriptional network from a high-quality microarray data set derived from 50 maize tissue types and developmental stages that represent the range of maize morphogenesis and span developmental time from embryogenesis to senescence. We clustered transcripts into a hierarchy with nested modules of increasing size and decreasing interconnectedness, and we identified 49 modules, 24 of which have robust interconnectedness.

With a custom-designed Affymetrix microarray chip to assay transcript levels, we were able to map 60% (33,664 of 55,672) of the probe sets for which we detected target hybridization to the high-quality, filtered gene set originally reported by Schnable et al. (2009; Fig. 2). Similarly, Sekhon et al. (2011) designed probe sets for a NimbleGen microarray with maize transcript assemblies and FGENESH gene models of the B73 genome sequence and found that about 70% of the probes matched filtered maize gene models. The maize filtered gene set is a conservative list of maize genes, and of the 55,672 probe sets for which we detected target hybridization, 9,919 (18%) map to the B73 genome and do not map to the maize gene models (Fig. 2). Twenty-two percent of the expressed probes did not match the genome, perhaps because some probes arise from misassembled Unigenes. The maize

Figure 5. ME correlations with leaves sampled before and after anthesis and silking. All_Leaf_Stages represents all nine leaf samples. Vegetative leaves are the V1, V2, and V5 leaves. Reproductive leaves are samples from the second leaf above the top ear (R1 and 10, 17, 24, and 31 DAP). For definitions of developmental stages, see Table I.
Network Modules Have Strong Associations with Specific Tissues and Biological Processes

Modules are composed of genes that have similar patterns of expression across all tissues. Nonetheless, 22 of our 24 robust modules (92%) are characterized by transcripts that are preferentially expressed or repressed within a specific tissue type relative to all other tissue types (Fig. 4). Our results indicate that organ identity is a primary factor that explains transcriptome variation throughout plant development and suggest that organ identity is the key determinant of cellular function. This discovery echoes the identification of numerous mutants that have aberrant cell structures but nonetheless demonstrate normal organ development (Smith et al., 1996). Whole-transcriptome comparisons consistently show that age, cell type, and environmental stimuli have relatively minor effects on transcriptional profiles relative to organ type (Ma et al., 2005; Schmid et al., 2005; Druka et al., 2006; Jiao et al., 2009; Sekhon et al., 2011). In addition, large numbers of tissue-specific transcripts have been identified in plants (Ma et al., 2005; Druka et al., 2006; Sekhon et al., 2011). For example, of 18,481 detected transcripts from barley, 650 were expressed in only a single tissue type (Druka et al., 2006). Seventy-five percent of the maize tissue-specific genes reported by Sekhon et al. (2011) that we examined are found in the appropriate, tissue-specific modules. Although environmental stimuli may rewire underlying network architectures (Luscombe et al., 2004), the congruence of these results indicates that the rough outlines of developmental network topologies are highly robust. Key expectations of tissue-specific modules are, first, that mutations within genes most highly connected, or central, to a tissue-specific module will have a phenotypic effect that preferentially affects that tissue. Second, as a number of elicitor molecules drive organ change (e.g. GA can induce floral feminization), genes differentially expressed in response to elicitor treatment should be overrepresented in specific developmental modules (Zhan and Lukens, 2010).

Maturation signals and switches in cell identity also activate distinct transcriptional regulatory modules within a single organ. By examining leaf tissues of different ages, we identified modules preferentially expressed in leaves prior to anthesis and silking and modules preferentially expressed after flowering. We also identified one ME with a strong correlation with embryo age. The detection of modules correlated with different stages of a single organ was possible because of the breadth of the transcription profiles collected in this study.

We find a high congruence between module associations with specific tissue types and biological processes. Twelve of the 24 verified modules were enriched for genes involved in specific biological processes, and all were positively or negatively associated with specific tissue types. Ficklin and Feltus (2011) developed a maize transcriptome network using 297 microarray data sets from various maize tissues and genotypes grown in a number of conditions, including 48 arrays from pulvinus and nine arrays of methylation-filtered genomic DNA. They identified clusters of genes enriched for ribosome and translation, seed storage activity, and photosynthesis (Ficklin and Feltus, 2011). It is likely that the robust signal of the different tissues within these samples contributed to the functional enrichment of modules. Nonetheless, it would be interesting to identify if modules clustered around functions independently of tissue type.

We expect that modules are a valuable resource for predicting gene function. Many of the genes with high MM have unknown functions. For these sparsely annotated genes, their hub status in a particular tissue-specific module generates novel hypotheses through the principle of guilt by association. For example, the transcript detected by probe set Zm028519_at is highly correlated with the Zm_mod13 eigengene (MM = 0.98) and with embryo age, with a similar expression profile in embryo to GRMZM2G074097 (Supplemental Fig. S5) discussed previously. The Unigene sequence on which the Zm028519_at probe set was based is homologous (BLASTN versus EST database; e-value = 3 x 10^{-118}) to an EST from a sorghum (Sorghum bicolor) embryo library. Nonetheless, the transcript does not map to the maize filtered gene set, has no known functions, and is not detected in any tissue other than embryo (Supplemental Fig. S5).

Transcriptional Modules in Maize Development

Genes within a module are likely transcriptionally regulated. A number of transcription-associated genes have high MM (Supplemental Table S2). These genes include a chloroplastic RNA polymerase protein, a chromatin-remodeling enzyme, and a DNA-binding-with-one-finger transcription factor. Nonetheless, transcription factors and other genes involved in transcription were not unusually central to modules relative to genes with other functions (Supplemental Table S6). We propose three explanations for this observation. First, transcription factors that are expressed in only one tissue may be rare. Transcription factors likely are active in more than a single condition and can alter their regulatory interactions between conditions (Luscombe et al., 2004; Brady et al., 2011). Second, precise regulation of module genes may arise
through the combinatorial protein interactions of transcription factors (Smaczniak et al., 2012). Finally, transcription factors that direct organ cell identity may act transiently to establish the identity early in development, and our study did not capture this time point. For example, heritable silencing of the expression of transcription factor \textit{FLOWERING LOCUS C} (\textit{FLC}) is accomplished by the transient expression of \textit{VERNALIZATION-INSENSITIVE3} that guides the heritable, epigenetic modification of \textit{FLC} (Sung and Amasino, 2004).

We note that binding sites of transcription factors can be overrepresented among module genes. Six modules have significant overrepresentation of 10 known promoter motifs (Table III). The genome-wide, verified binding sites of key developmental transcription factors (Bolduc et al., 2012; Morohashi et al., 2012) may further elucidate how genes within modules are coregulated.

Some Module Members Are Physically Clustered in the Maize Genome

Of the 13 modules with more than 200 genes, genes within three modules are significantly more colocated in chromosomal regions than expected by chance (Table IV). Clusters of functionally related genes may arise because of a shared chromatin environment that promotes their coregulation (Udvardy et al., 1985). Alternatively, epistasis among genes within modules seems feasible, as modules contain genes that act in biochemical pathways (Supplemental Fig. S3). Selection may have favored linkage of epistatic genes to reduce recombination between favorable alleles, thus contributing to the variability in linkage disequilibrium decay across the maize genome (Van Inghelandt et al., 2011).

Here, we begin to investigate factors that explain maize transcriptome variation across development and the regulatory basis for that variation. Future work will improve the resolution of the maize transcriptome modules by incorporating in-depth expression data of diverse RNAs. The functional relationships among genes across development also will likely be improved through the integration of other data (Zhu et al., 2008). Finally, a major objective will be to functionally characterize modules and to investigate how to alter modules to drive developmental changes.

### Table III. Modules with overrepresented promoter motifs

| Promoter Motif | Zm_mod01 | Zm_mod02 | Zm_mod04 | Zm_mod06 | Zm_mod10 | Zm_mod23 |
|---------------|----------|----------|----------|----------|----------|----------|
| AAAG         |          |          | 6.471    |          |          |          |
| AATTTAA      | 0.000    |          | 2.099    |          |          |          |
| CC(A/G)CCC   | 0.525    | 0.003    |          |          |          |          |
| CCCCCCG      | 9.448    |          |          |          |          |          |
| CGCGCC       | 0.021    |          |          |          |          |          |
| CCCCCGG      | 0.744    |          |          |          |          |          |
| TGGTTT       |          |          | 4.376    |          |          |          |
| TTTAAAAA     |          |          | 8.545    |          |          |          |
| AG/CCGAC     |          |          |          | 1.665    |          |          |

*P* values are calculated from Fisher's exact test (*P* < 0.01).  
\(^{a}\)Module is negatively correlated with leaf tissue.  
\(^{b}\)Module is negatively correlated with leaf and positively correlated with ear.  
\(^{c}\)Module is positively correlated with leaf tissue.  
\(^{d}\)Module is positively correlated with root tissue.

### Table IV. Dispersion scores of network modules with more than 200 genes

| Module   | No. of Genes | Mean\(^{a}\) | Variance\(^{b}\) | Dispersion (Observed) | Mean Dispersion (Randomized) | \(I\)^a |
|----------|--------------|--------------|------------------|------------------------|-------------------------------|--------|
| Zm_mod01 | 1,839        | 0.2670       | 0.3143           | 0.8495                 | 0.8790                        | 0.0303d|
| Zm_mod02 | 1,424        | 0.2066       | 0.2271           | 0.9100                 | 0.9038                        | 0.6475 |
| Zm_mod03 | 1,587        | 0.2304       | 0.2590           | 0.8895                 | 0.8940                        | 0.3899 |
| Zm_mod04 | 1,526        | 0.2216       | 0.2431           | 0.9116                 | 0.8976                        | 0.8105 |
| Zm_mod05 | 641          | 0.0930       | 0.0983           | 0.9462                 | 0.9545                        | 0.3085 |
| Zm_mod06 | 963          | 0.1398       | 0.1540           | 0.9075                 | 0.9330                        | 0.0657 |
| Zm_mod07 | 807          | 0.1172       | 0.1263           | 0.9278                 | 0.9432                        | 0.1824 |
| Zm_mod08 | 533          | 0.0774       | 0.0807           | 0.9590                 | 0.9619                        | 0.4535 |
| Zm_mod09 | 580          | 0.0842       | 0.0871           | 0.9669                 | 0.9587                        | 0.6873 |
| Zm_mod10 | 497          | 0.0721       | 0.0773           | 0.9319                 | 0.9645                        | 0.0322d|
| Zm_mod11 | 484          | 0.0703       | 0.0755           | 0.9307                 | 0.9655                        | 0.0269d|
| Zm_mod12 | 417          | 0.0604       | 0.0631           | 0.9564                 | 0.9701                        | 0.2068 |
| Zm_mod13 | 251          | 0.0364       | 0.0376           | 0.9684                 | 0.9819                        | 0.2304 |

\(^{a}\)The mean number of module genes within a 300-kb region (100-kb step value).  
\(^{b}\)The variance of module gene counts within a 300-kb region.  
\(^{c}\)P value: the probability of finding an equal or lower dispersion in a sample of 100,000 networks where genes are assigned to modules at random.  
\(^{d}\)Significance level at \(P < 0.05\).
**MATERIALS AND METHODS**

**Growth Conditions and Tissue Sampling**

An elite Syngenta maize (Zea mays) inbred (SR200) was grown in a greenhouse at the University of Guelph during the summer of 2007. Growth conditions were 16-h days (approximately 600 \mu mol m^{-2} s^{-1}) at 28°C, 8-h nights at 23°C, and 50% relative humidity. Plants were grown semi-hydroponically in pots containing Turface clay and watered with a modified Hoagland solution containing 0.4 g L^{-1} 28-14-14 fertilizer, 0.4 g L^{-1} 15-15-30 fertilizer, 0.2 g L^{-1} NH\textsubscript{4}NO\textsubscript{3}, 0.4 g L^{-1} MgSO\textsubscript{4}·7H\textsubscript{2}O, and 0.03 g L^{-1} micro-nutrient mix (sulfur, cobalt, copper, iron, magnesium, molybdenum, and zinc).

Plant samples representing 50 developmental stages were sampled for RNA extraction (Table I; Fig. 1; Supplemental Fig. S6). Three biological replicates per sample were harvested in the middle of the day to minimize complications due to diurnal changes in carbon and nitrogen metabolism. Total RNA was isolated and used for complementary DNA (cDNA) synthesis. Complementary RNA synthesis and labeling followed a standard protocol recommended by Affymetrix. Labeled RNAs were fragmented and applied to a maize custom GeneChip microarray for molecular hybridization. The array images with hybridization signals were acquired and quantified by GeneChip Operation System software (Affymetrix). The quality of the hybridization was assayed using Expressionist (GeneData). Experiments were repeated for the arrays that failed to pass the quality assay.

**Microarray Attributes and Data Preparation**

RNA was hybridized to a custom Affymetrix Unigene array with 82,662 probe sets, each consisting of 16 probes of 25 nucleotides. The 150 CEL files were normalized using the Robust Multichip Average method from the Plant Physiol. Vol. 161, 2013 1841

The adjacency matrix, we

The microarray platform was annotated by determining homology between probe sequences and the cDNA sequences for predicted transcripts from the filtered gene set of the 4a.53 release of the B73 genome using BLASTN (Altshul et al., 1990). For a probe to match a transcript, either at least 23 contiguous nucleotides out of 25 were required to match or 24 of 25 matched with an identity of at least 85%. For probes matching more than 14 nucleotides but fewer than 25 nucleotides with 85% identity were noted as close matches. Close matches were used to identify cross hybridization among transcripts, as described below. If 12 of the 16 probes in a probe set matched the same transcript, the probe set corresponded to that gene. If fewer than 12 but more than one probe in a probe set were a match or a close match for the same gene, the probe set was identified as a partial match for that gene.

In order to identify expressed transcripts that did not correspond to the filtered gene set of gene models, we searched genomic DNA for the cDNA sequences from which the probes were designed. Exonerate (Slater and Birney, 2005), an aligner that uses more exhaustive heuristics than BLAST, was used to map the probe sets to genomic sequence. The “model est2genome” parameter was used, which allows for introns of reasonable length. These Unigene sequences and their genomic positions were combined with the previously identified B73 gene models. We eliminated probe sets that cross hybridize or are redundant. For redundant collections of probe sets, a single probe set was retained according to the following criteria: (1) has highest expression.

**Creating Modules of Coexpressed Genes**

All module construction was performed with WGCNA software (Langfelder and Horvath, 2008). A correlation network is fully specified by its adjacency matrix that contains the network connection strength between each gene pair. All 34,876 probe sets were analyzed as a single block. To calculate the adjacency matrix, we first calculated the Pearson correlation coefficient (r) between each pair of probe sets across all developmental time points. The adjacency of two genes is proportional to the absolute value of their correlation coefficients:

\[ a_{ij} = \frac{(1 + s_{ij})}{2} \beta \]

where \(s_{ij}\) is the adjacency value of gene i and gene j, \(s_{ij}\) is the Pearson correlation between gene i and gene j, and \(\beta\) is the weight. This coexpression similarity measure preserves information about negative correlations. The weight serves to highlight the strongest correlations while reducing the mean connectivity, the average number of connections per probe set, of the network (Supplemental Fig. S7). Weighted networks are robust with respect to the choice of power. Gene coexpression networks have been found to exhibit a scale-free topology; their connections follow a power-decay law such that a small number of very highly connected nodes exist (Barabási and Oltvai, 2004; Chung et al., 2006). Using the scale-free topology criterion as described by Langfelder and Horvath (2008), we selected a threshold value of 5 (Supplemental Fig. S7).

We used the TO to transform the adjacency matrix to a coexpression distance matrix using WGCNA in R. While a correlation considers each pair of genes in isolation, topological overlap considers each pair of genes in relation to all other genes in the network (Ravasz et al., 2002; Li and Horvath, 2007; Yip and Horvath, 2007). Two genes have a high TO if they share high correlations with a common set of other genes. The use of TO filters out spurious or isolated connections (Oldham et al., 2008). Network relationships among genes were identified using hierarchical clustering of the dissimilarity matrix (i.e. 1 – the coexpression distance matrix). A dynamic tree-coding algorithm was used to “cut” each dendrogram and define the modules. The tree-coding algorithm iteratively decomposes and combines branches until a stable number of clusters is reached (Langfelder and Horvath, 2008). A summary profile, or eigengene, was calculated for each module by performing principal component analysis for each module (Langfelder and Horvath, 2007). The first principal component of the gene expression matrix for each module was retained as the representative ME. Forty-nine coexpression modules resulted.

We validated the 49 modules by examining the average TO for all genes in the module. The mean TO values of all identified modules should be significantly higher than the TO values of modules composed of randomly selected genes. We calculated the average TO values of modules composed of randomly selected genes by randomly assigning the 34,876 genes to 49 modules that were the same sizes as the observed modules. This process was repeated 50,000 times to obtain 49 null distributions. The probability that a random set of genes could generate a TO greater than or equal to the observed TO values for a module is the fraction of 50,000 iterations where the random group of genes had a higher mean TO than the observed mean TO. Twenty-four of the 49 modules were verified as highly significant (P < 10^{-5}; Supplemental Table S3).

**Testing Tissue-Specific Transcript Abundance**

To test if modules were associated with preferential expression in distinct tissue arrays arising from the 50 developmental stages, genes were classified together (e.g. leaf, root, shoot, etc.; Table I). We created a binary indicator variable (tissue = 1; all other samples = 0) and determined if any MEs were significantly correlated with the indicator. Positive correlation between a ME and a tissue type indicates that probe sets in that module have low transcript levels in that tissue relative to all other tissues. Negative correlation between a ME and a tissue type indicates that probe sets in that module have high transcript levels in that tissue relative to all other tissues. Using the eigengene is similar to averaging the correlations between the expression profiles for each gene in the module with the tissue type but avoids the multiple testing problem. Because modules have varying extents of heterogeneity in gene expression, not all modules are represented equally well by the ME. The MM for each gene within a module is the Pearson correlation between the expression level of the gene and the ME (Horvath and Dong, 2008). MM is a quantitative measure of the degree to which a gene is central to a module.

We used a similar approach to identify modules specific to tissues at different developmental stages. We divided leaf tissues into two groups: leaves harvested prior to anthesis and silking (V1, V2, and V5) and leaves harvested after flowering (R1 and 10, 17, 24, and 31 DAP; Table I). We assigned all samples a value of 0 except those from the group under consideration, which were assigned 1, and we performed module correlations as described above. To investigate modules that correlate with embryo development, we assigned nonembryo tissues 0 and embryo samples were assigned 10, 17, 24, or 31, based on the number of days between pollination and the day they were sampled. Correlations were performed as above.
GO Enrichment Analysis

GO enrichment analysis of modules was performed using the topGO module in R. We used GO annotations derived from BLAST2GO (Conesa et al., 2005) using the cDNA sequences used for the design of the Affymetrix microarray platform and the National Center for Biotechnology Information nonredundant database (July 16, 2009). A total of 17,139 genes were assigned 2,684 unique GO terms. For each module, Fisher’s exact test was used to identify GO terms that occur more frequently than expected given the frequency of the GO terms among all of the genes in the analysis. The “elim” method was applied to remove higher level GO terms from probe sets with significant lower level annotations, which has been shown to reduce the rate of false positives (Alexa et al., 2006). Within each module, we selected genes with a MM greater than 0.5, as these genes most resemble the ME. We defined a GO biological process as significantly associated with a module at $P < 0.001$. This strict criterion was used to eliminate GO terms present one or two times within a module but that nonetheless were highly significant because of the low frequency of genes with that GO term within the data set.

The MM of Transcription-Related Genes

To investigate the centrality of transcription factors within modules, probe sets in each module were ordered by their MM score. Of 17,139 genes with GO annotations, we identified 1,063 genes with transcription-related GO terms (e.g. “transcription activator activity,” “transcription cofactor activity,” “transcription initiation, DNA-dependent,” etc.; Supplemental Table S6). The rank of the transcription-related annotated gene with the highest MM was recorded. To determine whether this rank was higher than expected, we compared each module’s rank to a distribution of the highest rank of transcription-related genes obtained by randomizing the order of genes within the module for 100,000 iterations. We also permuted the order of a random set of genes with the same size as the module and determined the highest rank of transcription-related gene in this data set.

Analysis of Promoter Motif Enrichment within Modules

To determine whether modules were enriched for genes containing particular cis-promoter motifs, we counted the number of genes within each module that contained each of 106 promoter motifs obtained from plantCARE (Rombauts et al., 1999), PLACE (Higo et al., 1999), and GRASSILS (Yilmaz et al., 2009). All motifs have been reported to be transcription factor-binding sites in maize. This analysis was limited to the 13,047 genes that had been mapped to the filtered set of maize gene models. We used Fisher’s exact test with a critical value of less than 0.01 to compare the number of genes in the module that contain the promoter sequence within 500 bp upstream of the transcription start site with the number of genes not in the module with that sequence. Transcription start sites were determined based on the transcript that mapped most upstream in cases where multiple transcripts were annotated. We also examined promoter regions of 1,000, 1,500, and 2,000 bp.

Physical Clustering of Module Genes

To determine whether the genes in our modules were clustered in the maize genome, we compared the observed dispersion of gene density with the expected dispersion of gene density for each module. The dispersion statistic (mean divided by variance) was calculated by counting module genes in 300-kb sliding windows with a step size of 100 kb and recording the mean gene density and the variance of gene density. To calculate the expected dispersion statistic, we randomly shuffled the module assignments of genes 100,000 times and determined a null distribution by recording the dispersion statistic for each permuted data set. This procedure was applied to the 12 modules that had at least 200 genes in the module for which the genomic position was known.

Microarray data have been deposited in NCBI’s Gene Expression Omnibus under accession number GSE44743 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44743).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Hierarchical clustering dendogram of all arrays.

Supplemental Figure S2. A dendogram and heat map of module eigen-gene correlations.

Supplemental Figure S3. Pathways contained within the leaf-associated Zm_mod06 module.

Supplemental Figure S4. Heat map of module eigengene and sample correlations.

Supplemental Figure S5. Expression of two genes in the embryo-associated Zm_mod13 module.

Supplemental Figure S6. Images of all sampled tissues.

Supplemental Figure S7. Analyses of network topology for various soft-thresholding powers.

Supplemental Table S1. Expression profiles across arrays for 34,876 probe sets.

Supplemental Table S2. Probe set annotations.

Supplemental Table S3. Validation of modules based on mean topological overlap.

Supplemental Table S4. Module memberships of tissue-specific genes reported by Sekhon et al. (2011).

Supplemental Table S5. Gene ontology terms overrepresented within modules.

Supplemental Table S6. The top module membership ranks of transcription-related genes for each module.

Supplemental Table S7. Promoter motifs overrepresented in modules.

ACKNOWLEDGMENTS

We thank the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca) and Compute/Calcul Canada for the use of their facilities.

Received December 21, 2012; accepted January 28, 2013; published February 6, 2013.

LITERATURE CITED

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22: 1600–1607

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Arabási A-L, Olivi ZN (2004) Network biology: understanding the cell’s functional organization. Nat Rev Genet 5: 101–113

Belanger FC, Leustek T, Chu B, Kriz AL (1995) Evidence for the thiamine biosynthetic pathway in higher-plant plastids and its developmental regulation. Plant Mol Biol 29: 809–821

Beló A, Beatty MK, Honedred D, Fengler KA, Li B, Rafalski A (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120: 355–367

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verder J, Zuber H, Ott T, et al (2008) A gene expression atlas of the model legume Medicago truncatula. Plant J 55: 504–513

Bolduc N, Yilmaz A, Mejia-Guerra MK, Moroshaki K, O’Connor D, Grotewold E, Hake S (2012) Unraveling the KNOTED1 regulatory network in maize meristems. Genes Dev 26: 1685–1690

Brady SM, Zhang I, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeple M, Kim D, et al (2011) A stel-like-enriched gene regulatory network in the Arabidopsis root. Mol Syst Biol 7: 459

Caron H, van Schaik B, van der Meer M, Baas F, Riggins G, van Sluis P, Hermus M-C, van Asperen R, Boon K, Voûte PA, et al (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science 291: 1298–1292

Chung W-Y, Albert R, Albert I, Nekrutenko A, Makova KD (2006) Rapid and asymmetric divergence of duplicate genes in the human gene expression network. BMC Bioinformatics 7: 46

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676
Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R, Close T, Kleinholts A, et al (2006) An atlas of gene expression from seed to seed through barley development. Funct Integr Genomics 6: 202–211

Ficklin SP, Feltus FA (2011) Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. Plant Physiol 156: 1244–1256

Ficklin SP, Luo F, Feltus FA (2010) The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. Plant Physiol 154: 13–24

Fierro AC, Vandenburgssche F, Engelen K, Van de Peer Y, Marchal K (2008) Meta analysis of gene expression data within and across species. Curr Genomics 9: 525–534

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genomes Biol 5: R80

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27: 297–300

Horvath S, Dong J (2008) Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4: e1000117

Jiao Y, Tsafta S, Gnadota N, Sun N, Liu T, Clay NK, Ceserani T, Chen M, Ma L, Holford M, et al (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet 41: 258–263

Kaufmann P, Kajoro A, Angenstein GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. Nat Rev Genet 11: 830–842

Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1: 54

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559

Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes with specific cellular, functional and developmental hierarchies. Nat Genet 31: 180–183

Li A, Horvath S (2007) Network neighborhood analysis with the multi-topological overlap measure. Bioinformatics 23: 222–231

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431: 308–312

Lysenko EA (2007) Plant sigma factors and their role in plastid transcription. Plant Cell Rep 26: 845–859

Ma L, Sun N, Liu X, Iiao Y, Zhao H, Deng XW (2005) Organ-specific expression of Arabidopsis genomce during development. Plant Physiol 138: 80–91

Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K (2011) Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticaceae crops. Plant Cell Physiol 52: 785–803

Moreno-Risueno MA, Busch W, Beney PN (2010) Omics meet networks: using systems approaches to infer regulatory networks in plants. Curr Opin Plant Biol 13: 126–131

Morishima A (1998) Identification of preferred binding sites of a light-inducible DNA-binding factor (MNFI) within 5′-upstream sequence of C4-type phosphoenolpyruvate carboxylase gene in maize. Plant Mol Biol 36: 633–646

Moroshiki K, Casas MJ, Falcone Ferreyra L, Mejía-Guerra MK, Pourcel L, Yilmaz A, Feller A, Carvalho R, Emiliani J, Rodriguez E, et al (2012) A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. Plant Cell 24: 2745–2764

Movahedi S, Van de Peer Y, Vandepoele K (2011) Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in Arabidopsis and rice. Plant Physiol 156: 1316–1330

Mutwil M, Klie S, Tohge T, Giorgi EM, Wilkins O, Campbell MM, Fennie AR, Usadel B, Nikoloski Z, Persson S (2011) PLA-Net: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell 23: 895–910

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL (1996) A mutation that allows endosperm development without fertilization. Proc Natl Acad Sci USA 93: 5319–5324

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. Nat Neurosci 11: 1271–1282

R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551–1555

Rombauts S, Déhais P, Van Montagu M, Rouze P (1999) PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Res 27: 295–296

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vinrong M, Schülkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501–506

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115

Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM (2011) Genome-wide atlas of transcription during maize development. Plant J 66: 553–563

Slater GS, Birney E (2005) Automated generation of heuristics for biologicai sequence comparison. BMC Bioinformatics 6: 31

Smaczniak C, Immink RG, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, Dinh QD, Liu S, Westphal AH, Boeren S, et al (2012) Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc Natl Acad Sci USA 109: 1560–1565

Smith LG, Hake S, Sylvestre AW (1996) The tangled-1 mutation alters cell division orientations throughout maize leaf development without altering leaf shape. Development 122: 481–489

Springer NM, Danilevskaya ON, Hermon P, Helentjaris TG, Phillips RL, Kaeppler HF, Kaeppler SM (2002) Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genes E(z), esc, and P(vg). Plant Physiol 128: 1332–1345

Springer NM, Ying K, Fu Y, Yi J, Tey C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5: e1000734

Sung S, Amasino RM (2004) Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. Nature 427: 159–164

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Fervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20: 1689–1699

Udvardy A, Maino S, Schedl P (1985) The 87A7 chromomere: identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J Mol Biol 185: 341–358

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123: 11–20

Vicente-Carbajosa J, Moos SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in boxe gene promoters and interacts with the basic leucine zipper transcriptional activator OpAque2. Proc Natl Acad Sci USA 94: 7685–7690

Yilmaz A, Nishiyma MY Jr, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E (2009) GRASSiUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol 149: 171–180

Yip AM, Horvath S (2007) Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 8: 22

Zhan S, Horrocks J, Lukens LN (2006) Islands of co-expressed neighboring genes in Arabidopsis thaliana suggest higher-order chromosome domains. Plant J 45: 347–357

Zhan S, Lukens L (2010) Identification of novel miRNAs and miRNA dependent developmental shifts of gene expression in Arabidopsis thaliana. Mol Plant 3: 1011–1017

Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet 40: 854–861