Partitioned Local Depth analysis of time course transcriptomic data reveals elaborate community structure

Maleana G. Khoury[1,2¶], Kenneth S. Berenhaut[3¶], Katherine E. Moore[3], Edward E. Allen[4], Alexandria F. Harkey[1,2], Joëlle K. Muhlemann[1], Courtney N. Craven[3], Jiayi Xu[3], Suchi S. Jain[1], David J. John[5], James L. Norris[3], Gloria K. Muday[1,2*]

[1]Department of Biology and Center for Molecular Signaling, Wake Forest University, Winston-Salem, NC 27109

[2]Molecular Genetics and Genomics Program, Wake Forest University, Winston-Salem, NC 27109

[3]Department of Statistical Sciences, Wake Forest University, Winston-Salem, NC27109

[4]Department of Mathematics, Wake Forest University, Winston-Salem, NC 27109

[5]Department of Computer Science; Wake Forest University, Winston-Salem, NC 27109

¶These authors made equal contributions

*__Corresponding Author__: muday@wfu.edu (GKM)

Keywords: Systems biology, community structure, Arabidopsis thaliana, auxin, ethylene, salt stress, roots

Abstract

Transcriptome studies which provide temporal information can be valuable for identifying groups of similarly-behaving transcripts and provide insight into overarching gene regulatory networks. Nevertheless, inferring meaningful biological conclusions is challenging, in part because it is difficult to holistically consider both local relationships and global structure of these complex and overlapping transcriptional responses. To address this need, we employed the recently-developed method, Partitioned Local Depth (PaLD), which reveals community structure in large data sets, to examine four time-course transcriptomic data sets generated using the model plant *Arabidopsis thaliana.* As this is the first paper in systems biology to implement the PaLD approach, we provide a self-contained description of the method and show how it can be used to make predictions about gene regulatory networks based on temporal responses of transcripts. The analysis provides a global network representation of the data from which graph partitioning methods and neighborhood analysis can reveal smaller, more well-defined groups of like-responding transcripts. These groups of transcripts that change in response to hormone treatment (auxin and ethylene) and salt treatment were shown to be enriched in common biological function and/or binding of transcription factors that were not identified with prior analyses of this data using other clustering methods. These results reveal the potential of PaLD to predict gene regulatory networks within large transcriptomic data sets.

Author Summary:

This paper applies a newly developed approach to identify community structure in distance-based data and applies it for the first time to time-course transcriptomic datasets. The application of Partitioned Local Depth analysis (PaLD) to four previously published datasets measuring transcript abundance changes over time in response to hormone or stress treatment identified unique features of these datasets. This method was used to build networks revealing relationships between transcripts based on temporal responses and to identify groups of transcripts with similar temporal responses. These groups were examined revealing enrichment in functional annotation and in targets of specific transcription factors that provided additional insight beyond commonly used clustering approaches.

Introduction

Transcriptome remodeling is a highly dynamic, tightly controlled process during which transcript abundance is altered in response to stimuli with precise temporal responses. Information on the transcriptional machinery driving this process can be hidden within the transcript abundance data itself when changes over time are monitored, as a rapid increase in the synthesis of transcription factors (TFs) suggests that these TFs may control the expression of downstream target genes [1–3]. Transcriptome studies that measure changes in gene expression over time provide valuable temporal information that can be harnessed to identify groups of genes with similar temporal transcript abundance patterns, as well as gene regulatory networks composed of sequentially-induced transcriptional cascades [4–7]. While transcriptome studies attempt to predict intricate gene regulatory networks through measuring changes in transcript abundance, inferring meaningful biological conclusions from the data can be challenging. Ultimately, identification of overarching patterns of changes in transcript abundance has the potential to facilitate interpretation of large data sets in meaningful ways, which is valuable for uncovering the mechanisms that drive transcriptome remodeling [8,9].

A common approach to identify patterns and relationships within temporal transcriptomic data is to sort transcripts into clusters or groups based on kinetic responses. For overviews of clustering methods, see for instance [10–13]. At their core, clustering techniques group entities (or objects) together into classes (or collections of objects). While clustering techniques often provide class labels at a global level they do not focus on local-level information regarding ways in which the underlying entities (transcripts) and classes (clusters) are related. Recently, Berenhaut, Moore, and Melvin introduced a new technique,

Partitioned Local Depth (PaLD), that provides both local- and global-level information [14]. In particular, PaLD reveals structure in the form of a weighted network, in which edge weights (and resulting proximity in the network) convey relationship strengths between data points. This approach is motivated by a community-level social perspective in which an underlying assumption is that an object (transcript) is relatively similar to its *nearby* neighbors. Here, the concept of "nearby" reflects local perspective and adapts to varying density without the need for extraneous parameters [14]. Community detection and other network-based methods can then be applied to the weighted networks obtained from PaLD to reveal further structure and patterns in data [15,16].

In this study, the PaLD approach is employed to search for relationships between transcripts whose abundance changes in several time-course microarray or RNA-Seq data sets. The patterns reveal groups of transcripts with closely related temporal responses and provide insight into the relationships between these groups, while accounting for heterogeneity and varying density in the data. The data sets chosen for this analysis measured changes in transcript abundance in root samples from the model plant species *Arabidopsis thaliana* across a broad time series. Two of these data sets resulted from treatments that elevated the levels of the gaseous plant hormone ethylene, with transcript changes defined by microarray analyses [17] or RNA-Seq [18]. Two other data sets examined transcriptional response to elevated levels of auxin [19] or salt [20]. These data sets were chosen because of the extensive sampling across time and because both auxin and salt have been shown to interact with ethylene to modulate root development [21,22].

Applying the PaLD perspective on these transcriptomic data sets yields relational networks with local and global structure. These networks can then be viewed in lower dimensions, where genes with common temporal responses are proximate. PaLD analyses of these data sets effectively sorted genes into biologically meaningful subnetworks, which we call PaLD groups. These groups were enriched in specific biological functions and TF targets (meaning they contained more of these designations than expected, compared to the entire genome). Furthermore, the combination of functional enrichments and TF targets were subnetwork-specific, meaning these unique combinations were not shared between groups. We compared these PaLD networks to clusters and groups identified by other approaches, including *k*-means clustering and the Short Time Series Expression Miner (STEM) [17,19,20], which revealed greater enrichment in functional annotations and groups of genes targeted by common TFs in PaLD groups for some data sets. These findings support the utility of PaLD as a transparent and effective method to search for patterns within time course transcriptomic data.

Results

PaLD analysis of transcriptomic data

In Fig. 1, we apply the PaLD method to reveal community structure in a subset of seven transcripts from the ACC microarray data set (see Materials and Methods and Figure 2 for discussion of the complete data). The sole input for PaLD is a set of pairwise distances (or, more generally, distance comparisons). The pairwise distances between the transcripts are displayed in Fig. 1A; for details on how these distances are derived from fitted curves for the time course data, see Materials and Methods.

A Distance Matrix

	1	2	3	4	5	6	7
1	0.0	38.4	42.4	33.1	35.9	17.4	13.6
2	38.4	0.0	13.8	21.9	16.4	23.6	41.9
3	42.4	13.8	0.0	18.5	8.5	32.1	43.5
4	33.1	21.9	18.5	0.0	12.5	25.8	32.2
5	35.9	16.4	8.5	12.5	0.0	29.0	35.2
6	17.4	23.6	32.1	25.8	29.0	0.0	28.4
7	13.6	41.9	43.5	31.5	35.2	28.4	0.0

B Cohesion Matrix

	1	2	3	4	5	6	7
1	**0.234**	0.000	**0.000**	0.000	**0.000**	0.095	0.151
2	0.000	**0.214**	0.117	0.075	0.075	**0.024**	0.000
3	**0.000**	0.113	**0.252**	0.071	**0.169**	0.000	0.000
4	0.000	0.075	0.075	**0.214**	0.117	0.024	0.000
5	**0.000**	0.071	**0.169**	0.113	**0.252**	0.000	0.000
6	0.103	**0.033**	0.000	0.033	0.000	**0.192**	**0.048**
7	0.129	0.000	0.000	0.000	0.000	**0.048**	**0.212**

C

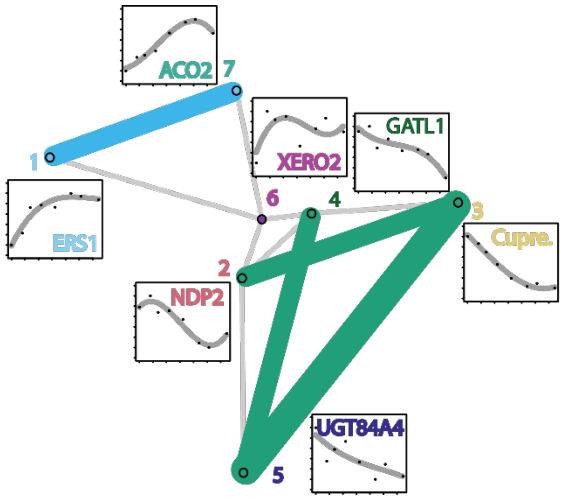


Figure 1. An illustrative example of PaLD applied to a small set of seven transcripts. In A, the matrix of distances between transcripts is given. In B, the matrix of cohesion values is displayed, see Eq. (2)]. Larger values of cohesion indicate stronger pairwise relationships; those which are particularly strong (Eq. (3)) are indicated in bold. In C, the resulting community graph is displayed together with insets that show the temporal response of the transcripts.

The output cohesion matrix, indicating relationship strength, is provided in Fig 1B, and the resulting community structure is displayed in Fig 1C. Relationships which are particularly strong (see Eq (3) below) are indicated in bold in Fig. 1B and emphasized with wider colored edges in Fig 1C; for full details see Materials and Methods. For step-by-step computation of values in Fig 1B see Supplementary Materials.

131 **PaLD network of 447 transcripts that have altered gene expression in response to**
132 **elevated levels of ethylene**

133 The PaLD method was used to generate a network using 447 transcripts whose abundance
134 changed in response to treatment with 1-aminocyclopropane-1-carboxylic acid (ACC),
135 which is the immediate precursor to the plant hormone ethylene. This treatment increases
136 ethylene synthesis and signaling [17]. The resulting network structure illustrated local and
137 global relationships among the transcripts, which is shown with each node representing a
138 gene and each edge connecting the transcripts with a similar kinetic response (Fig. 2). The
139 network contains two large groups of transcripts that were largely downregulated (at the
140 top of the diagram) or upregulated (in the lower half of the diagram), while a smaller group
141 in the middle of the network contained transcripts with mixed responses. We selected 7
142 transcripts to represent different parts of the network and show fiber plots of the
143 abundance of those 7 transcripts as a function of time after ACC treatment. Additionally,
144 smaller subnetworks, or neighborhoods, were determined for each of these 7 transcripts.
145 These neighborhoods consist of a transcript and all of their direct connections in the
146 network, which are shown in gray on the fiber plots. Some neighborhoods consisted of
147 more than 20 genes, while others had fewer than 10. These results illustrated the strength
148 of PaLD in grouping genes with similar behavior, as the fibers in each of these plots
149 followed the same trends.

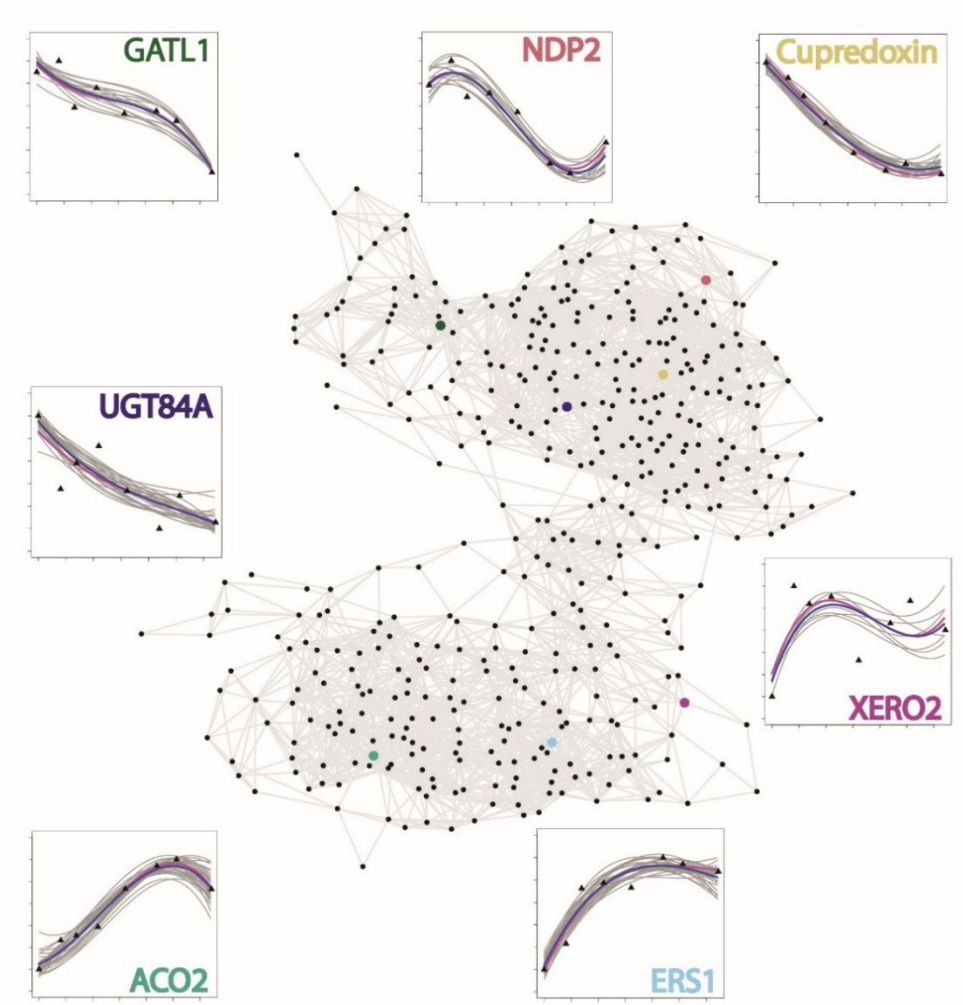


Figure 2. PaLD network of the 447 transcripts in the ACC microarray data set [17]. Each node represents a single transcript, and each edge represents a pairing between nearest neighbors. Fiber plots illustrate transcript abundance over time for seven representative genes (magenta lines) and their nearest neighbors (gray lines) averaged across three biological replicates, with the position of each transcript indicated in the same color as gene names. Averages for each neighborhood are shown as blue lines. Black triangles indicate the average abundance at each time point. Data is considered in log time and has been min-max scaled.

To divide this network into larger neighborhoods of like-responding genes, we used the Louvain method for community detection [15]. This method maximizes modularity, a

quantity that is larger for partitions with stronger within-group ties relative to that which would be expected by chance. When used in combination with PaLD, this approach revealed seven regions of temporally similar transcripts, which are shown with color coded nodes on a network diagram (Fig. 3A). The transcripts in each of these PaLD groups are listed in Supplemental Data Set 1, along with the abundance changes in these transcripts reported as the SLR (signal log ratio) over the time course of ACC treatment. Fiber plots of transcript abundance over time, with the averages shown in blue and the behavior of each gene in the network shown in gray, were overlaid on the network revealing that this method identifies transcripts with similar kinetic response (Fig. 3B).

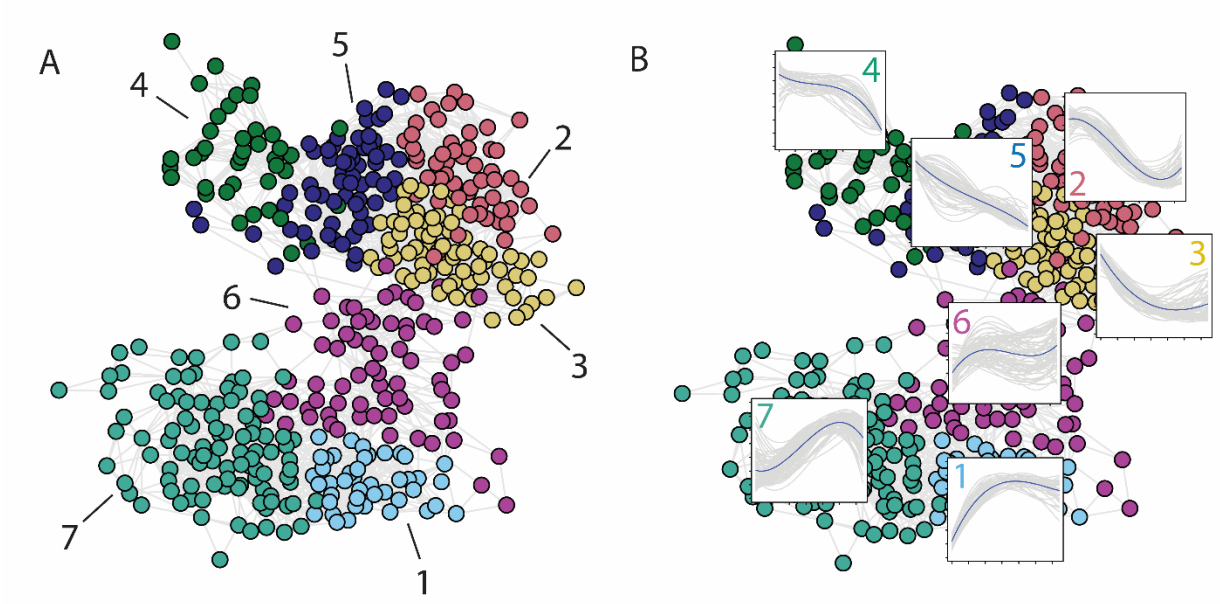


Figure 3. The PaLD community detection algorithm was used to partition the ACC microarray data set into groups with similar kinetic responses. (A) Seven groups, or subnetworks, were revealed by application of the Louvain method for community detection to the PaLD network. Each group is color-coded and labeled 1–7. (B) Bundles of fiber plots illustrate temporal responses of all transcripts within a group with the averages of all transcripts in that group shown in blue.

Comparison of PaLD groups to *k*-means clusters

One goal of this work is to understand how PaLD groups containing transcripts with similar kinetic response compare to other methods used to sort transcripts into like-responding groups or clusters. To evaluate the ability of PaLD to organize transcripts as compared to widely-used methods, we compared the PaLD groups to previously generated *k*-means clusters using this same set of data. *K*-means clustering was previously applied to this group of 447 transcripts yielding 72 clusters ranging in size from 1 to 111 genes, with 48 clusters of size 1 and 24 clusters of size greater than or equal to 2 [17]. To compare the *k*-means clusters to PaLD groups, we colored the nodes in the PaLD network to correspond to each of the 10 largest *k*-means clusters, which we refer to as A-J (corresponding to Clusters 1-10 in [17]), with this network shown in Figure 4.

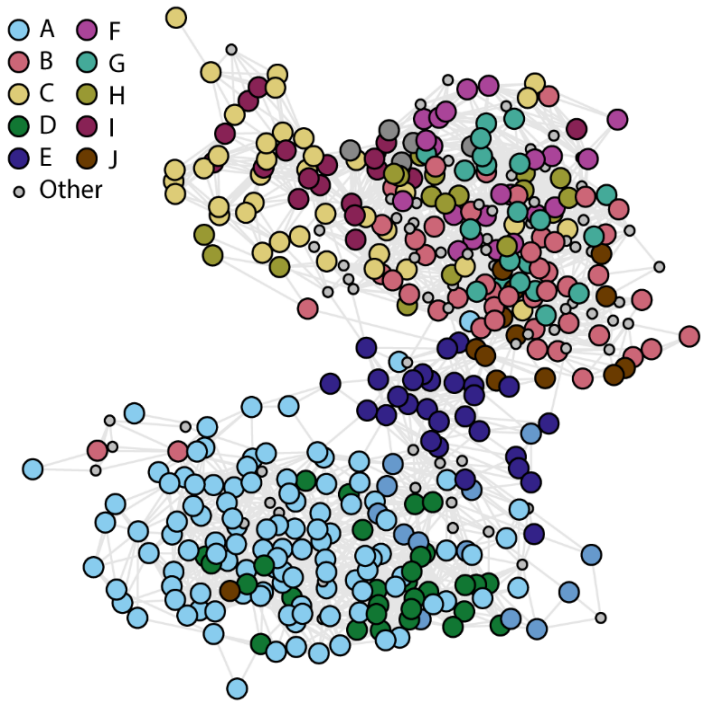


*Figure 4. Clusters of like-responding transcripts identified by k-*means *clustering were overlaid on the PaLD ACC network. The position of transcripts from the top ten k-*means *clusters determined previously* [17] *were denoted by distinct colors within the PaLD ACC network. Transcripts that were not part of the top ten k-*means *clusters are shown as smaller gray circles in the network.*

We also quantitatively compared the *k*-means clusters to the groups previously defined for the PaLD network using a table to illustrate overlaps between PaLD groups and *k*-means clusters (Table 1). *K*-means clusters A, D, and E had the least spread across the PaLD groups, which is consistent with the diagram in Figure 4. The matrix illustrates that 100% of the genes in cluster E were in PaLD group 6, 72% of the genes in cluster A (80/111 transcripts) were in group 7, and 67% of the genes in cluster D were in group 1, while other PaLD groups showed broader distribution between *k*-means clusters with cluster J containing genes in five PaLD groups with no more than 33% of those genes in any single PaLD group. This analysis revealed that some sets of transcripts have consistent behavior and are grouped together by both *k*-means clustering and PaLD, while for other genes these two methods detect relationships in different ways.

*Table 1. A comparison of PaLD groups and the k-*means *clusters in* [17]*. The PaLD approach revealed seven groups while 18 k-*means *clusters were originally considered (the largest 10 highlighted here). PaLD groups and k-*means *clusters (of size less than 15) are combined as "Other" in the table (comprising 60 k-*means *clusters with a total of 105 transcripts and 7 PaLD groups with a total of 8 transcripts, respectively). As an example, PaLD group 3 consists of 1 element of k-*means *cluster A, 24 elements of B, 2 elements of C, etc. The "Other" category is much larger for k-*means *as many transcripts were not in the top 10 clusters and had fewer than 15 transcripts.*

		***k*-means clusters**											
		A	B	C	D	E	F	G	H	I	J	Other	Sum
PaLD Groups	1	16	0	0	20	0	0	0	0	0	0	7	43
	2	0	10	4	0	0	9	9	5	1	2	16	56
	3	1	24	2	0	0	5	10	2	0	5	20	69
	4	0	0	25	0	0	0	0	0	10	0	15	40
	5	0	10	8	0	0	8	2	11	7	0	17	63
	6	13	0	0	4	28	0	0	0	0	5	19	69
	7	80	2	0	6	0	0	0	0	0	1	10	99
	Other	1	2	0	0	0	0	0	0	0	2	3	8
	Sum	111	48	39	30	28	22	21	18	18	15	105	447

222 **Enrichment analysis and conserved TF binding in a PaLD-generated ACC network**

223 We hypothesized that transcripts which respond similarly in time may work together to
224 drive the same biological responses and their synthesis may be regulated by the same TFs.
225 Since genes often work together in smaller subnetworks, we reasoned that the groups
226 revealed by the community detection algorithm would be large enough to yield enrichment
227 results yet small enough to represent biologically meaningful subnetworks. Therefore, to
228 assess the ability of PaLD to predict genes that function, and are regulated, in the same
229 biological network, we examined whether there were enriched functions and TF targets
230 within PaLD Louvain groups (Table 2). Using AgriGo v2.0, we looked for enriched function
231 in these groups, finding that six out of seven PaLD groups were enriched in gene ontology
232 annotations. We used TF DEACoN [23], which looks for significant enrichment in genes that
233 have been shown to be targets of TFs defined by a publicly available DNA Affinity
234 Purification (DAP)-Seq data set [24]. This analysis revealed that 5 out of 7 groups were
235 enriched in targets of known TFs. Of the 7 groups, 4 had enrichment in both annotated
236 function and TF binding, which we predicted to be the more biologically meaningful
237 subnetworks. Interestingly, the group that was enriched in transcripts annotated with
238 "negative regulation of ethylene signaling" was also enriched in targets of the ethylene
239 signaling master transcriptional regulator ETHYLENE INSENSITIVE 3 (EIN3), which
240 validated our prediction that this was a biologically meaningful subnetwork.

241 We also compared the PaLD groups to a set of core ethylene response genes, called the *gold*
242 *standards*, which were identified in a meta-analysis of root ethylene responses examining a
243 single time point of treatments to elevate ethylene [17]. Of the 447 transcripts used in the

244 PaLD analysis, 45 were part of the group of gold standards. The number of gold standards
245 in each of the PaLD groups or *k*-means clusters were reported in the summary table. It was
246 not surprising that the group annotated with ethylene signaling had the highest number
247 and proportion of gold standards. To determine whether the gold standard genes were
248 concentrated in neighborhoods, we overlaid the 45 gold standard genes [25] onto the PaLD
249 network (as shown in Supplemental Figure 1). The core ethylene genes were found in 6 of
250 the 7 PaLD groups. Furthermore, areas of the network that did not contain any core
251 ethylene genes could be observed. The distribution of these gold standards across the PaLD
252 groups and *k*-means clusters are also shown in Tables 2 and 3, which illustrate that they
253 are concentrated in 3 PaLD groups, with 80% found in Groups 1, 3, and 7. The partitioning
254 of these gold standard genes into limited PaLD groups is particularly striking, since the
255 temporal response to treatments was not part of the identification of this set of ethylene-
256 responsive transcripts.

257

258 *Table 2. Summary of enrichment results for the PaLD groups generated from the ACC*
259 *microarray data set* [17]*. The top 3 GO annotations with the lowest FDRs were reported. TFs*
260 *with targets enriched logFC ≥1 relative to the genome were reported. * Indicates groups*
261 *whose size of some groups is larger in the enrichment analysis due to microarray probes*
262 *recognizing transcripts from more than one gene.*

PaLD group	Number of transcripts	Number of gold standards (%)	Enriched GO annotations	TFs with enriched targets
1	44*	15 (33)	**Negative regulation of ethylene-activated signaling pathway**	EIN3, MGP, IDD(1, 5, 7, 11)
2	59*	2 (3)	Single organism metabolic process	None
3	69	10 (22)	None	AT5G18450
4	40	0	Phenylpropanoid biosynthetic process; cellular or organic substance catabolic process	MYB98
5	66*	2 (4)	Secondary cell wall biogenesis; cell wall organization; cellular polysaccharide biosynthetic process	ANAC(31, 37, 54, 79, 98)
6	70*	4 (9)	Response to wounding or endogenous stimulus; defense response, incompatible interaction	ANAC29, ANAC42, CAMTA1, DDF1, FAR1, WRKY20, WRKY33
7	102*	12 (27)	Hydrogen peroxide metabolic process; regulation of signal transduction; response to acid chemical	None

263

264 We also examined functional annotation and TF binding in the *k*-means clusters to ask
265 whether these groups were similarly enriched in function and regulation (Table 3). In
266 contrast, 7 out of 10 *k*-means clusters were enriched in annotations, and only 4 out of 10
267 clusters were enriched in targets of TFs. Clusters A, D, and E are the 3 groups that highly

overlap with PaLD clusters, while those without enriched TF regulators are those with less overlap with the PaLD clusters. The fourth cluster enriched in TF targets is cluster I, which shows a 56% overlap with PaLD group 4, resulting in enrichment in binding sites for MYB98 in both. There is only a single cluster with function linked to ethylene signaling, which is Cluster D, and that is highly conserved (67%) with Group 1. These results suggested that for this microarray data set, PaLD followed by the Louvain method revealed biologically meaningful subnetworks to a higher degree than *k*-means clustering.

276 *Table 3. Summary of enrichment results for the top 10 k-*means *clusters defined in Harkey et*
277 *al., 2018. The top three GO annotations with the lowest FDRs were reported. TFs with targets*
278 *enriched logFC ≥1 relative to the genome were reported. *Indicates groups whose size is*
279 *larger in the enrichment analysis due to microarray probes recognizing transcripts from*
280 *more than one gene or smaller due to probes that were not associated with a gene identifier*
281 *being removed.*

***k*-means cluster**	**Number of transcripts**	**Number of gold standards (%)**	**Enriched GO annotations**	**TFs with enriched targets**
A	115*	12 (10)	None	FAR1, IDD11, JKD
B	50*	5 (10)	None	None
C	41*	0	Cellular or organic substance catabolic process	None
D	30	16 (53)	**Ethylene-activated signaling pathway**; regulation of signal transduction; response to inorganic substance	AT3G10580, AT3G42860, HAT2, MYB3R4, MYB(73, 81, 119), WRKY(20, 31, 33, 40, 55)
E	29*	0	Response to other organism	CAMTA1, CAMTA5, CBF3
F	21*	6 (29)	Phenylpropanoid metabolic process	None
G	23*	4 (17)	Cellular component assembly	None
H	18	0	None	None
I	18	0	Secondary cell wall biogenesis; cellular polysaccharide biosynthetic process; cell wall organization	ANAC45, ANAC96, MYB98, MYB118
J	14*	1 (1)	None	None

282

283 **PaLD network of transcripts with altered expression in response to elevated auxin**

284 We also examined a larger microarray data set which was simultaneously generated with
285 the ACC data set described above, to contrast with our first analysis. PaLD was used to
286 generate a network of 1284 transcripts that have altered transcript abundance in roots in
287 response to treatment with the auxin indole-3-acetic acid (IAA). Interestingly, the general
288 structure forms a torus, or doughnut shape, which is quite different from the "hourglass"
289 shaped ACC network. The Louvain approach was applied to this data, yielding 8 groups
290 which are shown in distinct colors in Fig. 5 and are reported in Supplemental Dataset 1. We
291 overlaid the fiber plots onto the IAA PaLD groups and found groups positioned at opposing
292 ends of the structure had opposite patterns of change. Group 6 consisted of genes that were
293 upregulated most rapidly in response to IAA with a unimodal curve returning to starting
294 levels by the end of 24 hours. Groups 1, 4, and 5, had progressively slower rates of
295 upregulation. Group 2, which is on the opposite side of the network, also shows a unimodal
296 curve, but with decreased transcripts in response to auxin, followed by a return to baseline.
297 Fiber plots of each group reveal that the patterns of transcript abundance over time were
298 slightly shifted from one group to the next. We reasoned that these results might be
299 indicative of waves of transcripts that were activated at various times from the start of
300 hormone treatment.

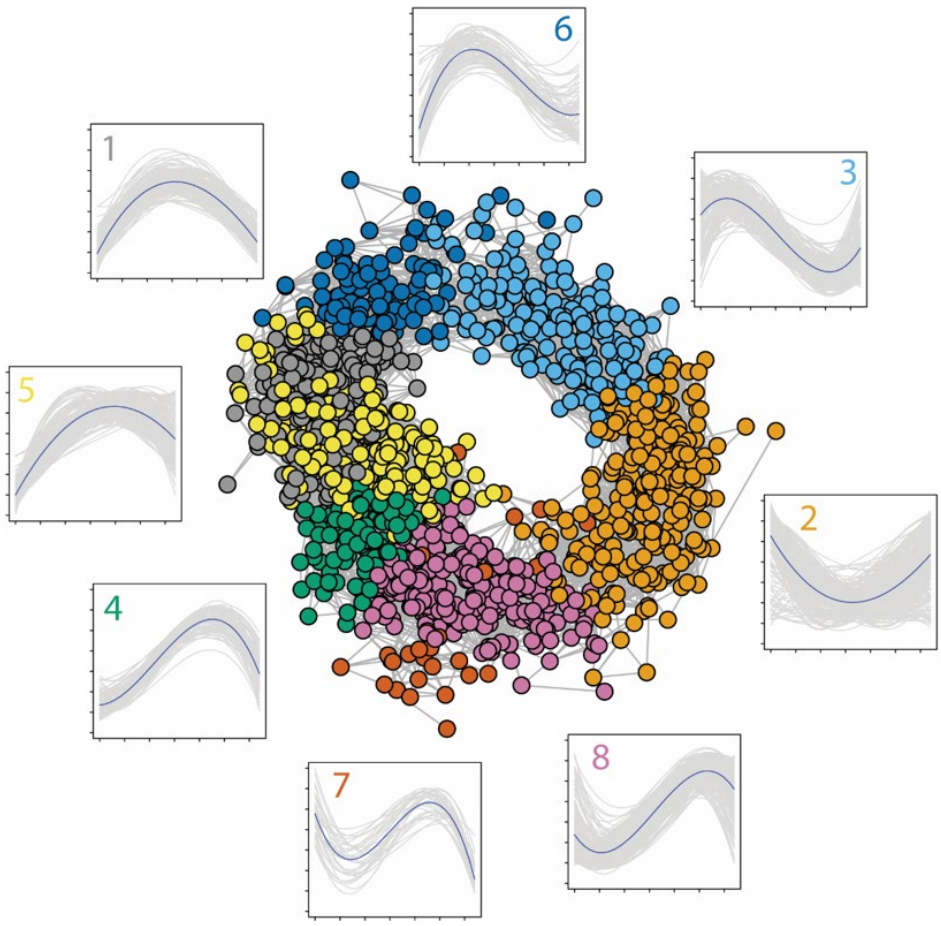


Figure 5. PaLD analysis of the 1284 transcripts with altered transcript abundance in response to treatment with IAA, an auxin [19]. The network diagram shows each PaLD group in a different color. Fiber plots of the Louvain groups illustrate transcript abundance over time with the number of each group matching the color of each group. Data is considered in log time and has been min-max scaled.

To further contrast the differences in network structure between the ACC and IAA microarray data sets, we generated videos showing the structure from multiple angles (Supplemental Movies 1 and 2). From those videos we produced snapshots of the networks from representative angles (Fig. 6). From these images it was clear that many of the groups that appear to overlap with one another in the two-dimensional networks were spatially, and thus temporally, distinct.

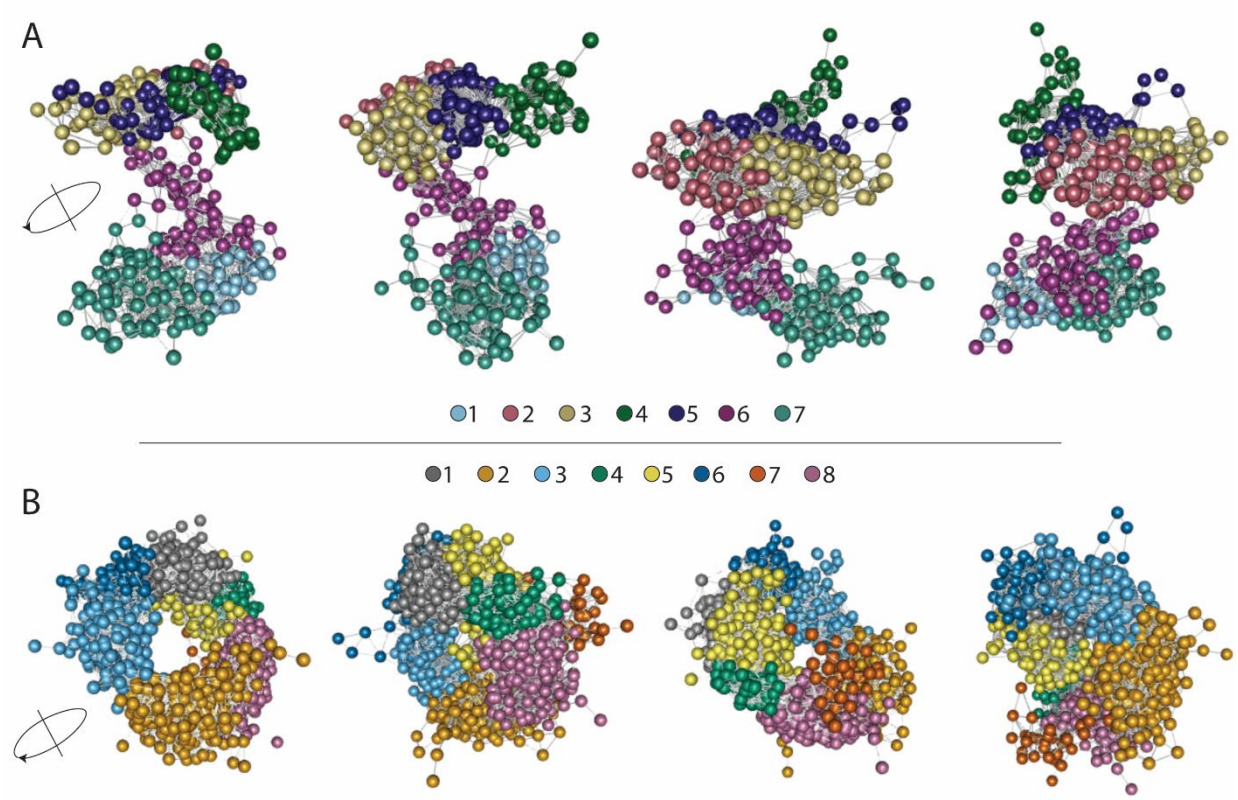


Figure 6. Three-dimensional illustrations of the PaLD networks for the ACC and IAA microarray data sets. (A) The ACC PaLD network is shown from 4 representative angles depicted as rotating counterclockwise around the axis (left) from left to right. (B) The IAA PaLD network is shown from 4 representative angles illustrated as rotating clockwise around the axis (left) from left to right. The colors represent the PaLD groups.

PaLD groups and *k*-means clusters were similarly enriched for the IAA microarray data set

We compared the distribution of transcripts between the PaLD groups and the top ten *k*-means clusters for the IAA data set [19] in a matrix (Table 4). The PaLD and *k*-means methods appeared to sort temporally-linked transcripts in different ways, providing a basis for comparing enrichment between the PaLD groups and *k*-means clusters. We looked for enriched function and enriched TF binding to genes in these PaLD groups and compared

327 them to the *k*-means clusters. Enriched GO annotations were found in all 8 PaLD groups
328 and 8 out of 10 (80%) *k*-means clusters (Tables 5 & 6). On the other hand, 5 out of 8 (63%)
329 and 7 out of 10 (70%) PaLD groups and *k*-means clusters, respectively, were enriched in
330 targets of TFs. As predicted, each set of groups/clusters had groups enriched in auxin
331 signaling, with the PaLD IAA group 5 and 6 and *k*-means cluster G having this annotation. It
332 is also possible to examine this network in other ways that can provide additional insight,
333 including identification of the next neighbors of each transcript, as described below.

334 *Table 4. Matrix of the overlap between the PaLD and the top 10 k-*means *clusters for the IAA*
335 *data set [19]. The sum of the transcripts in each PaLD or STEM group is listed in the last row*
336 *or column, respectively. The "Other" column indicates transcripts that were not in the largest*
337 *component or the top ten clusters.*

		k means Clusters											
		A	B	C	D	E	F	G	H	I	J	Other	Sum
PaLD Groups	1	160	0	0	0	1	0	1	0	0	0	24	194
	2	0	113	15	2	0	45	0	6	0	0	65	246
	3	3	1	0	83	26	15	0	0	0	0	58	186
	4	18	0	8	0	0	0	7	0	0	0	69	102
	5	57	0	2	0	1	0	43	1	0	8	81	193
	6	32	0	0	7	29	0	1	0	0	1	24	94
	7	0	3	0	0	0	0	0	0	18	0	11	32
	8	0	12	97	0	0	0	0	36	3	5	79	232
	Other	0	1	0	0	1	0	0	0	1	0	2	5
	Sum	278	130	122	92	58	60	52	43	22	14	413	1284

338

339 *Table 5. Summary of enrichment results for the PaLD groups generated from the IAA*
340 *microarray data set [19]. The top 3 GO annotations with the lowest FDRs were reported. TFs*
341 *with targets enriched logFC ≥1 relative to the genome were reported. *The indicated group*
342 *size is smaller than expected due to probes that were not associated with a gene identifier*
343 *being removed.*

PaLD group	Number of transcripts	Enriched GO annotations	TFs with enriched targets
1	193*	Ribosomal large subunit biogenesis; ribosome assembly; pyrimidine nucleotide biosynthetic process	TRP1
2	246	Cell wall organization; response to water deprivation; sulfur compound metabolic process	AT1G01250, BEH3, bZIP51, CAMTA5, FAR1, LEPBEH3, MYB96, RAP2.1
3	185*	Glycosyl or organonitrogen compound catabolic process; response to wounding	ERF2, ERF5, RAP2.9, WIND4
4	101*	Cell cycle; covalent chromatin modification; regulation of protein modification process	None
5	193	Auxin-activated signaling pathway; root morphogenesis; root meristem growth	None
6	94	**Cellular response to auxin stimulus**	None
7	32	Cellular lipid metabolic process	ANAC(7,37,38,45,50,53, 54,57,79,87,92,96), AT4G26030, MYB(49,74,80,99,107)
8	228*	Cell division; **hormone-mediated signaling pathway**; response to abiotic stimulus	DDF2, DOF4.2, JKD, LBD23, LOB, MYB3R4

344

345 *Table 6. Summary of enrichment results for the 10 largest k-*means *clusters defined in Lewis*
346 *et al., 2013. The top three GO annotations with the lowest FDRs were reported. TFs with*
347 *targets enriched logFC ≥1 relative to the genome were reported. *The indicated groups are*
348 *larger in the enrichment analysis due to microarray probes recognizing transcripts from*
349 *more than 1 gene or smaller due to removal of probes not associated with a gene identifier.*

***k*-means cluster**	**Number of transcripts**	**Enriched GO annotations**	**TFs with enriched targets**
A	277*	Megagametogenesis; ribosomal large subunit biogenesis; rRNA processing	None
B	130	Cell division; response to osmotic stress or water deprivation	bZIP28, bZIP51, CAMTA1, ERF014, FAR1, LBD23, LEP, LOB, NLP7, SPL13B, TINY
C	122	Cell division; regulation of cell cycle	None
D	91*	Organonitrogen or cellular catabolic process; cell communication	AT5G18450
E	58	**Auxin-activated signaling pathway**; negative regulation of cellular process; protein ubiquitination	ABF2, AHDP, ANAC042, AT1G47655, BEH3, bHLH31, bHLH46, BMY2, bZIP(2,44,48,66), BZR1, ERF(11,19, 43), ESE3, FAR1, HB9, IDD11, JKD, MGP, MYB98, MYB118, NLP7, PDF2, RAP2.1, RAP2.9 TCP(3, 7, 9, 15, 20)
F	60	Cell wall organization or biogenesis; polysaccharide metabolic process; secondary metabolic process	bZIP22, ERF(3, 34, 38, 43), HB20, RAP2.1, RAP2.9
G	52	**Auxin-activated signaling pathway**; regulation of hormone levels; response to organic cyclic compound	ANAC079, ASIL2, AT3G25990, AT3G42860, bHLH31, bZIP(2,28,44, 48), FAR1, GT-1, HB20, IDD(1,7,11), MYB(3R4,3R5,63), RAP2.9, WRKY31
H	42*	None	None
I	22	None	ANAC(7,37,38,45,50,53,54,57,58, 70,75,76,78,79,87,92,96,98,101), AT4G26030, MYB(74,80,96,99,121)
J	14	Regulation of biological quality; **response to hormone** or abiotic stimulus	ANAC040, CRC, HB(9,13,24,53), IDD1, IDD7, JKD, MGP, MYB3R1, RVE5, RVE6, SPL13B

350 **PaLD neighborhoods can be used to identify biologically meaningful subnetworks**

351 We used a second approach to look for neighborhoods of like-responding genes in the PaLD
352 network, by generating lists of nearest neighbors for each gene. It is important to note that
353 neighborhoods can be of varying size reflecting homogeneity in the data. To illustrate this
354 approach, we selected the rapidly upregulated gene encoding Auxin Response Factor 19
355 (ARF19), which drives auxin gene regulatory networks in roots. The neighborhood of
356 ARF19 was highlighted on the network (Fig. 7A) and the identity of direct neighbors were
357 defined and the abundance behavior of these transcripts is shown in the Fiber plot in Fig.
358 7B and are listed in Supplemental Data Set 1. Interestingly, most of the transcripts in the
359 ARF19 neighborhood encoded proteins known to be involved in auxin signaling and
360 transport, including PINFORMED 3 (PIN3), PIN7, and AUXIN-INDUCIBLE 2 (IAA2). In
361 addition, we reveal transcripts whose next neighbor groups were associated with
362 particular GO annotations. A different enrichment analysis software, Metascape [26], was
363 selected for this purpose. Since each transcript has its own neighborhood, there were over
364 1200 neighborhoods processed through Metascape, revealing significantly enriched GO
365 terms associated with the neighborhood of each transcript, including enrichment in
366 response to auxin, auxin transport, and auxin homeostasis. Figure 7G shows the enriched
367 biological functions for each of the direct neighbors of ARF19.

368 We used these data with annotations for each transcript's neighborhood and asked which
369 transcript's neighborhoods were annotated with key word fragments such as "auxin" or
370 "root". The position of those genes were then marked on the network, as shown in Figure
371 7C and E. The transcripts involved in root-related processes formed a localized niche in the

network, which overlapped significantly with the ARF19 neighborhood, while transcripts associated with auxin-related processes were more dispersed throughout a larger region of the network (Fig. 7B). The fiber plots of these groups of next neighbors are also shown in Fig. 7D and F.

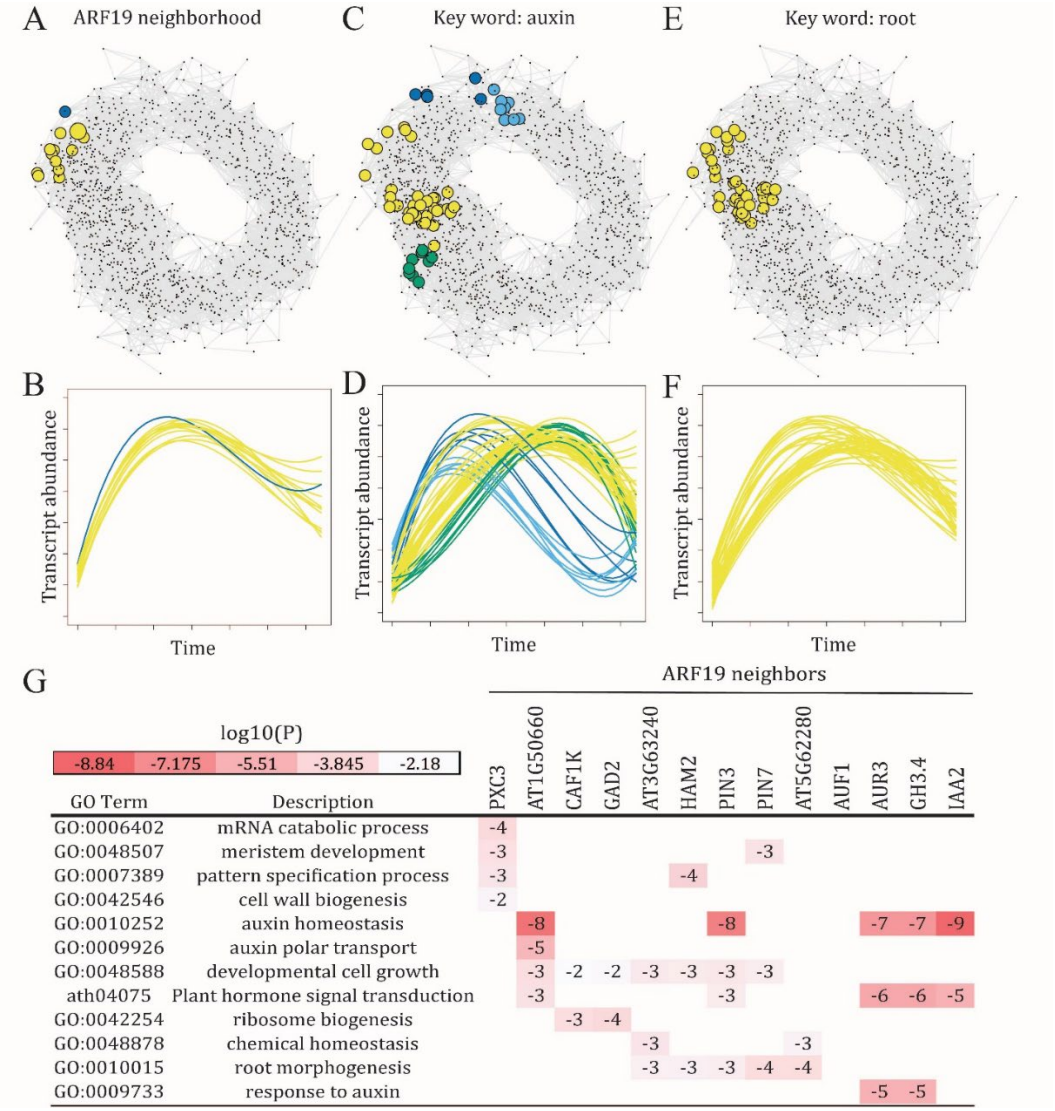


Figure 7. Functionally relevant neighborhoods in the IAA network. (A) On the top, the neighborhood of Auxin Response Factor 19 (ARF19), with each node colored according to PaLD groups as displayed in Figure 5. (B) This fiber plot shows transcript abundance over time for all transcripts in the ARF19 neighborhood. Data is considered in log time and has been min-max scaled. (C and E) Neighborhoods for each transcript were queried for conserved function using Metascape and transcripts annotated with the GO term "auxin" or "root" are highlighted on the network with colors corresponding to PaLD groups as displayed in Figure 5. (D and F) The fiber plots for these annotated groups are shown with the same color coding. (G) The enrichment output for the neighborhoods of the neighbors of ARF19. The names of the genes or locus identifier of ARF19 neighbors are shown along the top of the heatmap. Values indicate the p values in log base 10 rounded to the nearest whole number for enrichment of that term in each gene's neighborhood relative to the whole genome. Neighborhoods that were not enriched in a GO term do not have a value shown.

391 **PaLD network of transcripts that have altered expression in response to the plant**
392 **hormone ethylene**

393 We also applied PaLD to an RNA-seq data set consisting of transcripts that respond to the
394 plant hormone ethylene [18]. Interestingly, the PaLD network of this data set did not take
395 on either of the shapes of the PaLD networks of the ACC and IAA microarray data sets (Fig.
396 8A). With 957 transcripts, the RNA-Seq data set was in between the microarray data sets in
397 terms of size. The fiber plots for the PaLD groups conveyed diversity in transcriptional
398 response despite the seemingly high interconnectedness of the groups (Fig. 8B) and the
399 transcripts in each group are shown in Supplemental Dataset 1. To examine subnetwork
400 functionality, an enrichment analysis was performed. We found enriched GO annotations
401 and targets of TFs in 8 and 7 (out of 9) groups, respectively. (supplemental Table 1).
402 Moreover, we found 3 groups with enriched functions related to ethylene signaling. These
403 groups had distinct kinetic responses and were regulated by distinct sets of TFs, suggesting
404 that the same hormone can activate multiple waves of transcriptional responses that are
405 part of different subnetworks. Group 4 had an unusually high amount of TF regulators.
406 Examining the fiber plot for this group revealed a slow increase in expression in response
407 to the treatment, suggesting that a higher number of TFs upstream could act together to
408 elicit the observed response. These findings suggest a link between transcriptional kinetics,
409 regulation, and biological function. The RNA-Seq data set had previously been analyzed
410 using the Dynamic Regulatory Events Miner (DREM). However, the distribution of
411 transcripts among the DREM groups was not published, and thus enrichment analysis of
412 the DREM groups was not performed. Taken together, these results suggest PaLD is a good
413 tool to identify biologically meaningful subnetworks.

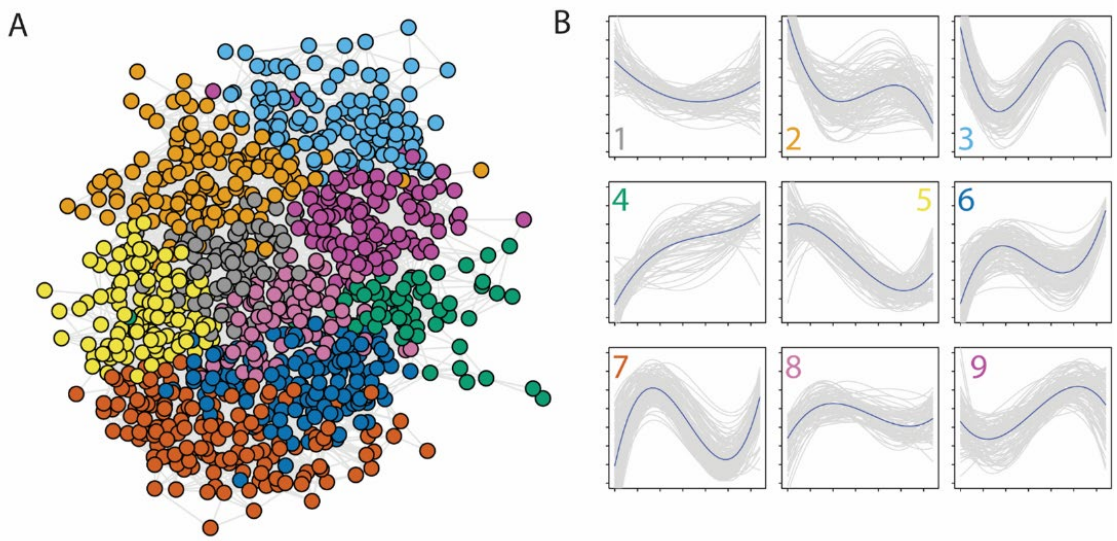
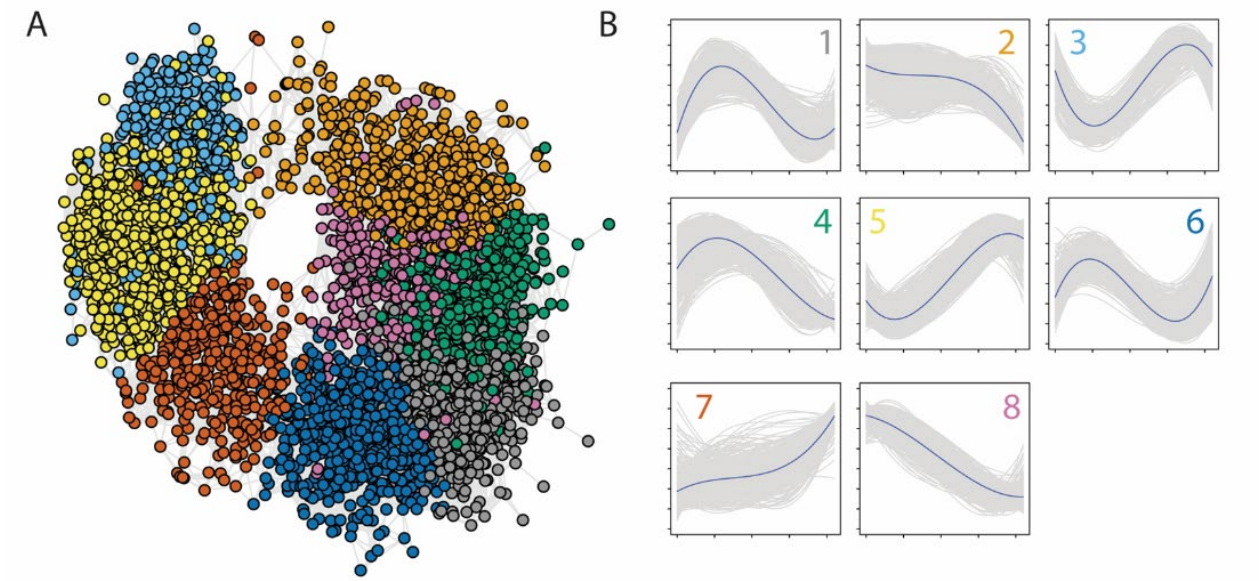


Figure 8. PaLD analysis of transcripts with altered gene expression in response to the plant hormone ethylene [18]*. (A) PaLD network of the 957 transcripts in the ethylene RNA-Seq data set with each PaLD group having a distinct color. (B) Fiber plots of the PaLD groups illustrating transcript abundance over time color coded to match the network diagram. Data is considered in log time and has been min-max scaled. Colors are matched to those in the network in (A).*

PaLD network of transcripts that have altered gene expression in response to salt stress

We also examined one salt stress transcriptomic dataset to encapsulate the gene regulatory mechanisms happening in a plant cell during stress conditions, to contrast with hormone treatments using PaLD. These transcripts were derived from an RNA-Seq data set that had been processed using the short time-series expression miner, or STEM [20]. The PaLD network of these stress-altered transcripts highlighted the diversity in kinetic response, as suggested by the circularity of the network (Fig. 9A). Additionally, PaLD groups were generated as described above and the genes in each group are summarized in Supplemental Dataset 1. The fiber plots of the PaLD groups revealed slightly shifted

431 kinetics from one group to an adjacent group in the network, suggesting a wave-like
432 pattern of temporal response (Fig. 9B).



433
434 *Figure 9. PaLD analysis of transcripts with altered gene expression in response to salt stress*
435 *[20]. (A) PaLD network of the 3,808 transcripts in the salt stress RNA-Seq data set* with PaLD
436 group having a distinct colors*. (B) Fiber plots of the PaLD groups illustrating transcript*
437 *abundance over time. Data is considered in log time and has been min-max scaled. The PalD*
438 *group fiber plots for each group are colored to match the network in (A).*

439

440 **Enrichment analysis revealed new information about transcripts with similar**
441 **temporal responses**

442 The transcripts in the salt stress RNA-Seq data set had previously been sorted into STEM
443 groups based on their transcriptional response patterns. We sought to compare the
444 functional annotation in the top 10 STEM groups to the PaLD groups to determine whether
445 the PaLD algorithm could reveal new information about temporally-linked transcripts. A
446 matrix of the overlapping transcripts between the PaLD and top 10 STEM groups revealed

differences in how the transcripts were sorted (Table 9). These differences were reflected in the enrichment results. While both methods yielded similar quantities of output for enriched GO annotations, a higher proportion of PaLD groups (5 out of 8, or 63%; Table 10) were enriched in targets of TFs as compared to STEM groups (5 out of 10, or 50%; Table 11). On the other hand, all of the STEM groups that were enriched in targets of TFs had multiple TFs targeting them, while most of the PaLD groups with enriched targets of TFs had one TF targeting them. Overall, the PaLD and STEM algorithms yielded similar amounts of enrichment outputs with variation in which functions and TF binding targets are enriched in these groups. Taken together, these results suggested that PaLD can identify biologically meaningful subnetworks in an RNA-Seq data set and yields different information about temporally-linked transcripts than STEM.

Table 7. Matrix of the overlap between the PaLD (horizontal) and the top ten STEM (vertical) groups [20] *in the salt data set. The sum of the transcripts in each PaLD or STEM group is listed in the last row or column, respectively. The "Other" row includes transcripts that were not in the largest component.*

		STEM Groups										
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**Sum**
PaLD Groups	*1*	2	0	369	9	131	0	3	3	0	0	517
	2	261	0	12	293	1	0	0	0	0	2	569
	3	0	15	0	0	0	33	0	0	156	101	305
	4	230	2	141	106	17	0	3	0	0	0	499
	5	0	435	0	0	0	202	0	0	4	48	689
	6	2	1	25	2	175	3	198	141	0	0	547
	7	0	258	2	1	0	75	9	32	2	0	379
	8	245	0	0	15	0	0	19	0	0	0	279
	Other	4	2	4	1	0	1	1	4	4	3	24
	Sum	744	713	553	427	324	314	233	180	166	154	3808

462 *Table 8. Summary of enrichment results for the PaLD groups generated from the salt stress*

463 *RNA-Seq data set* [20]. *The top 3 GO annotations with the lowest FDRs were reported. TFs*

464 *with targets enriched logFC ≥1 relative to the genome were reported.*

PaLD groups	Number of transcripts	Enriched GO annotations	TFs with enriched targets
1	517	Cold acclimation; response to abscisic acid; response to high light intensity	DDF2
2	569	Auxin-activated signaling pathway; response to red or far red light; transmembrane receptor protein tyrosine kinase signaling pathway	LOB
3	305	Cellular response to hormone stimulus; response to organic cyclic compound; single organism cellular process	AT1G10250, AT4G18450, ERF5, LEP
4	499	Photosynthesis, light harvesting in photosystem I; protein-chromaphore linkage; response to cytokinin	TCP9
5	689	Maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8s rRNA, LSU-rRNA); ribosomal large subunit biogenesis; ribosomal small subunit biogenesis	None
6	547	Abscisic acid-activated signaling pathway; ethylene-activated signaling pathway; response to chitin	AT1G01250, bHLH46, DDF2, ERF2, GATA-19, LOB, TCP9
7	379	Defense response to bacterium; protein phosphorylation; response to endoplasmic reticulum stress	None
8	279	None	None

465

466 *Table 9. Summary of enrichment results for the top 10 STEM groups* [20]. *The top 3 GO*
467 *annotations with the lowest FDRs were reported. TFs with targets enriched logFC ≥1 relative*
468 *to the genome were reported.*

STEM group	Number of transcripts	Enriched GO annotations	TFs with enriched targets
A	744	Acyl-coA biosynthetic process; hydrogen peroxide catabolic process; photosynthesis, light harvesting	None
B	713	Cell wall pectin biosynthetic process; response to cadmium ion; tRNA aminoacylation for protein translation	None
C	553	Response to abscisic acid; **response to salt stress**; response to water deprivation	None
D	427	Protein-chromaphore linkage; response to far-red light; response to red light	None
E	324	Abscisic acid-activated signaling pathway; cuticle development; response to desiccation	AT1G01250, ERF2, LBD23, LOB
F	314	Maturation of LSU-rRNA; maturation of SSU-rRNA; protein targeting to mitochondrion	None
G	233	Defense response to bacterium; protein phosphorylation; response to chitin	AT1G01250, AT5G18450, bHLH31, bZIP28, ERF014, LOB, RAP2.1
H	180	Cell death; defense response; protein autophosphorylation	ANAC054, ANAC079
I	166	None	AT1G01250, bZIP28, CAMTA1, FAR1, LEP
J	154	Oxoacid metabolic process; response to toxic substance; small molecule biosynthetic process	bZIP22, bZIP26

469

470 # Discussion

471 A central challenge in systems biology is the inference of gene relationships from patterns
472 found in transcriptomic data. In particular, transcriptomic studies that include a temporal
473 dimension allow us to make predictions about like-responding transcripts, due to common
474 transcriptional controls or biological function of the proteins that they encode. These
475 transcriptomic data sets can be used to address questions including how hormones or
476 physiological stress alter the transcriptional landscape in plants. Such experiments reveal
477 the set of genes that have altered expression under elevated levels of plant hormones or
478 salinity, for instance, providing new information about the pathways and mechanisms that
479 are activated in response to these factors. In this paper, we analyzed several time-course
480 transcriptomic data sets, generated for the model plant *Arabidopsis thaliana,* to delve more
481 deeply into the mechanisms driving these complex regulatory processes by applying new
482 methods to query the data for predictive patterns. The data sets used in this analysis were
483 selected to encompass the types of research currently being done in Arabidopsis.

484 The wide variety of methods which can be employed in bioinformatics can make it
485 challenging to determine which methods to apply and, beyond this, may require the user to
486 make decisions that influence the results. Specifically, widely employed methods such as *k*-
487 means and hierarchical clustering, require the user to provide additional inputs (e.g., the
488 number of clusters) or select values from a parameter space (e.g., to minimize some cost
489 function). Many clustering methods provide only limited information about the global
490 structure of the data, and so dimension reduction techniques, such as principal component
491 analysis, are often utilized to provide a flat 2-dimensional display of the data. Since such

methods can compress valuable structural features, network-based methods can also be used to reveal informative structure. Yet, again, methods for creating such networks may force the user to select parameters (e.g., number of neighbors, or a threshold for the maximum distance between neighbors) and these choices can influence the results obtained. In light of these challenges, the need to share straightforward modeling algorithms for holistically reflecting local and global structure (including delineated groups or clusters), which can be easily applied to biological data sets, is more critical than ever before.

In this paper, we presented the first application of the Partitioned Local Depth (PaLD) method to transcriptomic data and provided a self-contained description of the approach. Our focus here was the shape of the curve for transcript abundance over time; we measured similarity in transcript response using fitted cubic polynomials to expression levels considered in log-time to reflect the rise/fall behavior typical of time course transcriptomic data; other application-dependent choices are possible. The networks defined by PaLD provide a natural set of temporally-linked transcripts. The method provides a transparent way to construct a network which does not require additional user input parameters. The recently introduced method was defined to reflect social interactions, and thus the neighbors and overall structure revealed by PaLD can be intuitively interpreted. For further discussion, see [14]. An R package for PaLD with detailed examples describing ways that it can be employed can be found at https://github.com/moorekatherine/pald (Katherine Moore, Kenneth Berenhaut, and Lucy D'Agostino McGowan, Github Package, 2022).

514 To contextualize the results and illustrate the value of this new method, we analyzed 3
515 time-course transcriptomic data sets that were generated in Arabidopsis. The datasets
516 include 3 that examine hormone response [17-19], while the other examined the stress
517 responses to elevated salt [20]. PaLD analysis of a microarray data set of root samples
518 treated with ACC, a precursor to the hormone ethylene [17], revealed proportionally more
519 groups with enrichment in functional annotations and targets of TFs as compared to the *k*-
520 means clustering method. Most striking was the significant enrichment in one of the PaLD
521 groups in ethylene signaling. This group was enriched in targets of the ethylene signaling
522 master transcriptional regulator ETHYLENE INSENSITIVE 3 (EIN3). In comparison, the *k*-
523 means method also yielded a group that was enriched in the function "ethylene signaling,"
524 but this group was not enriched in targets of EIN3. These findings suggested that for this
525 particular data set, PaLD more effectively grouped temporally-linked transcripts together
526 into biologically meaningful subnetworks. The application of PaLD to an ethylene response
527 transcriptional data set [18] also resulted in two groups enriched in ethylene signaling,
528 with one of these groups enriched in targets of EIN3.

529 Consistent with the ability of ACC to remodel root architecture [25,27], PaLD and *k*-means
530 clustering methods revealed groups that were enriched in secondary cell wall biogenesis,
531 which is necessary for driving primary root growth, lateral root and root hair formation.
532 The PaLD groups and *k*-means clusters with this annotation also showed enrichment in
533 targets of the NAC protein family of TFs, suggesting that the NAC family may drive the gene
534 regulatory networks remodeling root growth under conditions of elevated ethylene.

535 We also applied the PaLD method to analyze a microarray data set in which roots were
536 treated with IAA. Both the PaLD and *k*-means clustering methods resulted in similar
537 numbers of groups that were enriched in functional annotations and targets of TFs but
538 differ in distribution of genes between these groups and thus the content of these
539 enrichments. A benefit of using the PaLD approach is that the number of like-responding
540 transcripts can vary across the set and, in contrast to common nearest-neighbor methods,
541 is not a quantity that must be specified by the user. This allows us to identify meaningful
542 sets of temporally-similar transcripts that are directly linked to given genes of interest. For
543 illustrative purposes, we examined the neighborhood of Auxin Response Factor 19
544 (ARF19) revealing transcripts that were directly connected to ARF19 included the auxin
545 transporters PIN3 and PIN7. This finding is consistent with these transcripts having similar
546 kinetic responses to auxin, which may be due to ARF19 driving PIN3 and PIN7
547 transcription or with all 3 transcripts being controlled by a common TF. We also found
548 significant overlap between the ARF19 neighborhood and transcripts that were associated
549 with annotations including the key word "root." These results review the utility of the
550 localized neighborhoods that PaLD generates as a way to reveal transcriptional networks.

551 In addition to the role of plant hormones in controlling gene regulatory networks, we
552 examined an RNA-Seq data set in which seedlings were treated with NaCl to induce salt
553 stress. We compared the PaLD groups to the STEM groups that were defined in Wu et al.,
554 2021, and found that both methods had similar outputs of the number of groups with
555 enrichment in functional annotations and targets of TFs. We found using the PaLD method
556 that several groups were highly enriched in transcripts involved in auxin, ethylene and
557 abscisic acid signaling. This suggests that multiple hormone-induced signaling pathways

558 are activated in response to salt stress, consistent with hormone crosstalk central to salt
559 stress responses, an idea supported by other experimental results [28–30]. The PaLD
560 approach provides three-dimensional network models that are rich and informative spatial
561 representations of the relationships among temporally-linked transcripts. The resulting
562 local and global structure provides insight into the relationships between genes. We
563 showed how PaLD can be used to model biological networks and reveal groups of genes
564 with enriched function and regulation, which can be employed to predict biologically
565 meaningful subnetworks. PaLD revealed groups with enrichment in functional annotations
566 as well as targets of TFs, suggesting the ability of the algorithm to group temporally-linked
567 genes together in biologically meaningful ways. The PaLD networks can also allow one to
568 obtain a collection of neighboring genes, a feature which may be valuable in exploratory
569 analysis. Ultimately, the unique individual gene perspective offers a plethora of
570 opportunities for exploration into the functional and regulatory patterns of different
571 subnetworks. Finally, this network in combination with network analysis algorithms, such
572 as community detection, can result in informative groups of transcripts. These groups can
573 be examined for enriched functional annotation and transcriptional regulation which can
574 generate hypotheses about gene interactions and biological implications that can then be
575 tested experimentally. In particular, Arabidopsis T-DNA insertional mutants are widely
576 available to test which genes might be misregulated downstream of knockout mutant in a
577 transcription factor, thereby revealing transcriptional networks. Taken together, the PaLD
578 approach allows transcriptomic data, which is often rich and complex, to become more
579 interpretable and manageable in a research setting. In light of the increased specialization
580 of researchers across the sciences, the need to create straightforward modeling algorithms

that can be easily applied to data sets by biologists with limited training in bioinformatics is more critical than ever before. Ultimately, we demonstrated here that PaLD is a straightforward method that can be used to process transcriptomic data to reveal enrichment in function and regulation, which is the first step in modeling complex biological processes that can then be validated experimentally in the laboratory.

Materials and Methods

Description of the PaLD Method

Suppose that S is a set of n points (i.e., $|S| = n$) and write $x \in S$ to denote membership in S. For each pair of points, $x, y \in S$, we denote the dissimilarity (or distance) of y from x by $d(x, y)$. Throughout, the only information leveraged is (anchored) dissimilarity comparisons of the form $d(z, x) < d(z, y)$ or $d(x, z) < d(y, z)$, for $x, y, z \in S$. The foundation in distance comparisons allows for an element of robustness and a broad interpretation of distance; the only metric requirement is that each x be closer to itself than to any other point. For example, one may employ application-specific measures of dissimilarity based on correlation coefficients or fixation indices, in which dissimilarity is interpreted non-linearly. The results are unchanged by any re-scaling of the distances (e.g., logarithmic or exponential).

Materials and Methods

Embedding S in a latent (social) space in which local comparisons (or "conflicts") arise, we define a measure of pairwise relationship strength by considering alignment in induced local regions. This perspective allows for consideration of interactions occurring across a variety of scales and gives a meaning of *local* that does not require the user to provide

additional inputs or search an underlying parameter space. We present the algorithm in a concrete computational form; in the note below, we briefly describe the probabilistic social perspective from which the PaLD approach arises; for further details see [14].

With $x, y \in S$, define the *local focus*, $U_{x,y}$, to be the set of points which are closer to x or y than x and y are to one another, that is ,

$$U_{x,y} = \{z \in S: d(z,x) \leq d(y,x) \text{ or } d(z,y) \leq d(x,y)\}. \tag{1}$$

We consider a point $w \in S$ to be aligned with x in opposition to y if w is in the local focus (i.e., $w \in U_{x,y}$) and is closer to x than to y (i.e., $d(w,x) < d(w,y)$). The strength of the alignment is then inversely proportional to the number of points in the local set (i.e., $|U_{x,y}|$). The *cohesion* of w to x is defined via

$$C_{x,w} \overset{def}{=} \frac{1}{n-1} \sum_{\substack{y \in S \\ y \neq x}} \frac{\mathbf{1}(d(w,x) < d(w,y) \text{ and } w \in U_{x,y})}{|U_{x,y}|} \tag{2}$$

i.e., it is the average strength of alignment, where $|U_{x,y}|$ is the number of elements in the set $U_{x,y}$ and $\mathbf{1}(Q) = 1$ if statement Q is true and $\mathbf{1}(Q) = 0$, otherwise. In the case that $d(w,x) = d(w,y)$, we resolve the tie by setting the numerator to $1/2$; for clarity of presentation this case is suppressed.

To identify important relationships between points, we will emphasize those pairs for which mutual cohesion is greater than what is typical for a point in an induced local set. Leveraging the symmetry of the local regions, we consider the relationship between x and y to be (*particularly) strong* whenever

$$\min\{C_{x,w}, C_{w,x}\} \geq \frac{1}{2n} \sum_{x \in S} C_{x,x} = \frac{1}{2n(n-1)} \sum_{x \neq y} \frac{1}{|U_{x,y}|}. \tag{3}$$

Further context and justification for the value of this threshold from a probabilistic perspective are given in (Berenhaut et al. 2022).

For data $S = \{s_1, s_2, \ldots, s_n\}$, we obtain an $n \times n$ matrix of cohesion values, $C = \left[C_{s_i,s_j}\right]$. For ease of notation, we write $C_{i,j}$ to represent the (i,j) entry in the matrix C. The cohesion matrix conveys strength of relationship between points and can be visualized as a network (or graph). The adjacency matrix defining the weighted undirected graph, G, is obtained by replacing each value, $C_{i,j}$, with the minimum of $C_{i,j}$ and $C_{j,i}$. Thresholding the symmetrized cohesion matrix at the value in Eq. (3), leads to an adjacency matrix defining the graph, $G^\star$, whose (weighted) edges are only those for which the cohesion is particularly strong.

For visualization purposes, we apply the Kamada-Kawai graph drawing algorithm to the graph $G^\star$. When further partitioning is desired (for instance, as in Section 3.2), we may employ a community detection algorithm, such as the Louvain algorithm, on the largest connected component. Louvain is one of several algorithms available to find a partition of a network into groups which maximizes modularity.

Note: (Probabilistic perspective and local depth). Aggregating the support for x from all points $w \in S$, we obtain a measure of local centrality, referred to as *local (community) depth*, and defined by

$$\ell(x) \stackrel{\text{def}}{=} \sum_{w \in S} C_{x,w}. \tag{4}$$

From the probabilistic perspective considered in [14], the local depth of x, $\ell(x)$, described above is the probability that, in conflict against a point Y selected uniformly at random

from the remaining points in S, a conflict-point, $Z \in U_{x,Y}$ selected uniformly at random is closer to x than it is to Y. We then have the following,

$$\ell(x) = P\big(d(Z,x) < d(Z,Y)\big) = \sum_{w \in S} P\left(Z = w, d(Z,x) < d(Z,Y)\right) = \sum_{w \in S} C_{x,w}. \qquad (5)$$

In particular, cohesion, as equivalently defined in Eq. (2), is obtained from partitioning the probability which defines local depth; thus, this approach is entitled *Partitioned Local Depth*.

Example (Example 1, revisited). Consider the distances in Figure 1A. To compute the cohesion of s_4 to s_2, we consider the induced local sets for $j = 1,3,4,5,6,7$. Among the six induced sets, s_4 is contained in $U_{s_2,s_1} = S$, $U_{s_2,s_4} = U_{s_2,s_5} = \{s_2, s_3, s_4, s_5\}$, $U_{s_2,s_6} = \{s_1, s_2, s_3, s_4, s_5, s_6\}$ and $U_{s_2,s_7} = S$. Across these sets, s_4 is closer to s_2 than to the opposing point, s_j, for $j = 1,6,7$. As a result, as in Eq. (2),

$$C_{s_2,s_4} = \frac{1}{6}\left(\frac{1}{7} + 0 + 0 + 0 + \frac{1}{6} + \frac{1}{7}\right) \approx 0.075.$$

The remaining pairwise cohesion values are computed similarly and are displayed in Fig. 1B. The threshold distinguishing particularly strong relationships is 0.112 (i.e., half the average of the entries on the diagonal; see Eq. (3). Relationships which are particularly strong are indicated in bold in the cohesion matrix in Fig. 1B. In Fig. 1C, we display the graph G; the subgraph graph, $G^\star$, of strong relationships is emphasized with bold colored edges.

Defining a Distance Between Transcripts

A goal of this project is to reveal network-based relationship information among transcripts from the shape of the curves associated with expression level over time. Here,

we take the average at each time point over trials, and then apply min-max scaling over the relevant time points. A min-max scaling linearly transforms the data to a [0, 1] range. Each of the data sets considered here were collected at less than 10 approximately exponentially spaced time points; to reflect decay in activity we consider the data in log time. Using least squares, we construct cubic polynomials, $f_i(t)$, $1 \leq i \leq n$ (one for each gene), to model the double-extrema behavior of interest. . To capture other behaviors, different non-linear or piecewise-linear functions could be used. To measure the distance between the fitted time-curves, f_i and f_j, (associated to two different genes), we consider similarity in functional value and first derivative (capturing the expression level and rate of change, respectively). In particular, we define the associated distance between the 2 curves via

$$d(f_i, f_j) \stackrel{def}{=} \int_0^{log(M+1)} |f_i(t) - f_j(t)| \; dt + \int_0^{log(M+1)} |f_i'(t) - f_j'(t)| \; dt, \qquad (6)$$

where f_i' is the first derivative of f_i and $M$is the maximal time point. Note that the first term above is the total (absolute) area between the curves; and the second term corresponds to the (absolute) area between the derivatives. In what remains, the distance between transcript abundance profiles is understood to be the dissimilarity of the associated fitted cubic polynomials as described in Eq. (6).

New—Choice of Distance Function

The distance function defined in Equation (6), defines the distance between 2 curves as the area between the graphs of the 2 curves plus the area between the graphs of their derivatives. There are many possible distance metrics, depending on the application at

686 hand. For instance, standardizing the coefficient of $\int_0^{\log(M+1)} \left| f_i(t) - f_j(t) \right| \, dt$ as 1, yields

687 distance functions

688 $$d(f_i, f_j) \stackrel{def}{=} \int_0^{log(M+1)} \left| f_i(t) - f_j(t) \right| \, dt + \sum_{k=1}^{m} c_k \int_0^{log(M+1)} \left| f_i^{(k)}(t) - f_j^{(k)}(t) \right| \, dt, \qquad (7)$$

689 where m is a positive integer, c_m are real coefficients and $f^{(k)}$ is the kth derivative of f are all

690 distance functions. By changing the weights, c_k, one can choose to emphasize or de-

691 emphasize the value of fundamental properties of the corresponding graph. In our

692 example, we chose to include the area between the graphs of the derivatives so as to give

693 additional weight reflecting whether the corresponding functions were increasing or

694 decreasing similarly. By leaving out the term corresponding to the second derivative, we

695 chose to ignore the effect of concavity. The number of time points in the data set as well as

696 the predicted underlying curves are important considerations in determining the definition

697 of the corresponding distance function. The definition of appropriate distance metrics for

698 the data is an important next question.

699 Description of the Transcriptomic Data sets

700 *Table 8. Summary of the time course transcriptomic data sets used in this study. For the*

701 *microarray and RNA-Seq data sets, the processed signals (log ratios) and differentially*

702 *expressed genes were used in our analysis, respectively.*

Data set	Type	Number of DE genes	Treatment	Time course (h)	Clustering method	Accession number
Harkey et al., 2018	Microarray	447	1 μM ACC	0, 0.5, 1, 2, 4, 8, 12, 24	*k*-means	GSE84446
Lewis et al., 2013	Microarray	1284	1 μM IAA	0, 0.5, 1, 2, 4, 8, 12, 24	*k*-means	GSE42007

Chang et al., 2013	RNA-Seq	957	10 µl l^{-1} ethylene gas	0, 0.25, 0.5, 1, 4, 12, 24	DREM	SRA063695
Wu et al., 2021	RNA-Seq	3808	100 mM NaCl	0, 1, 3, 6, 12, 24, 48, 72	STEM	GSE153103

703

704 **Microarray data sets**:

705 Microarray data sets generated in parallel and published in Lewis et al. 2013 ("auxin data
706 set") and Harkey et al. 2018 ("ACC data set"), were used in this study. Arabidopsis seedlings
707 at 5 days after germination were transferred to control growth media or media containing
708 either 1-amino-cyclopropane-1-carboxylic acid (ACC, 1 µM), which is a precursor to
709 ethylene, or indole-3-acetic acid (IAA, 1 µM), which is an auxin. RNA was isolated from
710 roots at 8 time points: 0, 0.5, 1, 2, 4, 8, 12, and 24 hours. Triplicates of RNA samples were
711 collected at each time point from root tissue. Transcript abundance was measured using an
712 ATH1 Arabidopsis genome array and normalized using systematic variation normalization.
713 The raw microarray data sets can be found in the Gene Expression Omnibus under
714 accession numbers GSE84446 and GSE42007. The response to ACC or IAA was reported as
715 Signal Log Ratio (of treatment over the time-matched control for each time point, $\log_2$ fold
716 change), and the data sets were filtered to identify transcripts that were above the
717 microarray background, had consistent magnitude and patterns of change and had an
718 SLR>0.5. This resulted in 447 DE genes for the ACC data set and 1246 for the IAA data set.
719 The SLRs of these samples at each time point were used for PaLD community network
720 analyses

721 **Gold Standard Data set**:

722 The ACC microarray data was reanalyzed in parallel with two additional transcriptomic
723 data sets (Harkey et al. 2019). Raw data was downloaded and combined for parallel
724 processing using *limma* and other packages in R. Only the 4-hour time point from Harkey et
725 al. 2018 was used to be consistent between all 3 data sets. Genes were deemed
726 differentially expressed (DE) if they had an absolute fold change greater than 0.5, and an
727 adjusted p-value less than 0.05. This meta-analysis identified 139 genes, which we refer to
728 as the *gold standard* of ethylene signaling in roots, that were DE in all three data sets.
729 Because of the parameters defined, only 45 of the gold standard genes were present in the
730 ACC 447. For more detail about the ethylene signaling gold standard genes, see (Harkey et
731 al. 2019).

732 **Ethylene RNA-Seq data set**:

733 The data set from Chang et al. 2013 used a time course to examine ethylene response in
734 Arabidopsis, but with some key differences compared to the Harkey et al. 2018 data set.
735 Plants were treated with ethylene gas (10 µl l^{-1}). RNA was then isolated from whole
736 Arabidopsis seedlings grown under different conditions (3 days old, etiolated), rather than
737 isolated root tissue. The time points after treatment were 0, 0.25, 0.5, 1, 4, 12, and 24 hr,
738 which are similar to the Harkey et al. 2018 time points, but with an additional early time
739 point (0.25 hr), and fewer later time points (Harkey et al. 2018 has 2 and 8 hr time points).
740 In addition, RNA-Seq was used to quantify transcripts, rather than microarray.

741 **Salt stress RNA-seq data set**:

742 This time series data set was generated by growing Col-0 seedlings for 7 d under
743 continuous light and seedlings were transferred to media containing 0 or 100 mM NaCl for

744 0, 1, 3, 6, 12, 24, 48 and 72 h [20]. RNA was isolated from root samples and RNA-Seq was
745 performed in triplicate. Differential expression analysis was performed using the R package
746 DESeq2 [31] followed by pairwise comparisons to the expression level at 0 h with the
747 threshold of $|\log2(FC)| \geq 1$ and FDR < 0.1. The differentially expressed genes were then
748 clustered by the short time-series expression miner (STEM) algorithm using the default
749 settings. We used the top 10 STEM groups in our enrichment analysis. All differentially
750 expressed salt stress genes were used in the PaLD analysis.

751 **Identification of transcription factor targets in PaLD groups, *k*-means and STEM**
752 **clusters**

753 Identification of enriched binding of transcription factors in the clusters and groups was
754 achieved using TF DEACoN , a program that uses DAP-Seq data [24] to report targets of a
755 subset of transcription factors in Arabidopsis [23]. The size of some groups is larger in the
756 enrichment analysis due to microarray probes recognizing transcripts from more than one
757 gene or smaller due to probes that were not associated with a gene identifier being
758 removed. The output of this analysis uses a log fold change cutoff greater than 1 and a p-
759 value cutoff of less than 0.05 and is such that the abundance in the group is compared to
760 the whole genome. (Note that adjusted p-values were used here, yielding slightly different
761 results than those obtained using raw values.)

762 **Gene Ontology enrichment in PaLD and other groups:**

763 Gene Ontology enrichment in PaLD groups, *k*-means clusters, or STEM groups, relative to
764 the whole genome, was performed in the Singular Enrichment Analysis tool in AgriGO [32].
765 In AgriGo, "statistical test method" was set to hypergeometric and "multi-test adjustment

method" was set to Yeketuli (FDR under dependence). The top 3 most significantly enriched GO annotations were reported.

Author contributions:

MGK: Formal analysis, Investigation, Writing-Original Draft Preparation and Review and Editing; KSB: Conceptualization, Formal analysis, Writing -Original Draft Preparation and Review and Editing; KEM: Formal analysis, Investigation, Writing-Original Draft Preparation; EEA: Formal analysis, Writing-Original Draft Preparation and Review and Editing, AFH: Investigation, Writing-Review and Editing, JKM: Investigation, Writing-Review and Editing, CNC: Formal Analysis, JX: Formal Analysis, SSJ: Investigation; DJJ: Formal Analysis, Writing-Review and Editing; JLN: Formal Analysis, Writing-Review and Editing, GKM: Conceptualization, Project Administration, Funding Acquisition, and Writing-Review and Editing.

Funding Support:

This work was supported by the Center for Molecular Signaling at Wake Forest University and a grant from the United States National Science Foundation (MCB 1716279) to GKM.

ORCID IDs:

Maleana G. Khoury:	0000-0001-9869-5688
Kenneth S. Berenhaut:	0000-0002-9651-852X
Katherine E. Moore	0000-0001-6943-2416
Edward E. Allen	0000-0002-7797-0799
Alexandria F. Harkey	0000-0001-9250-9851
Joëlle K. Muhlemann	0000-0003-2431-4357
Courtney N. Craven	0000-0003-2328-3640

789 Jiayi Xu 0000-0002-3916-0699
790 Suchi Jain 0000-0001-7570-1846
791 David J. John 0000-0002-3270-9669
792 James L. Norris 0000-0001-5376-6964
793 Gloria K. Muday 0000-0002-0377-4517

794 **References**

795 1. Petricka JJ, Winter CM, Benfey PN. Control of Arabidopsis Root Development. Annu
796 Rev Plant Biol. 2012;63: 563–590. doi:10.1146/annurev-arplant-042811-105501

797 2. Niklison-Chirou MV, Agostini M, Amelio I, Melino G. Regulation of Adult Neurogenesis
798 in Mammalian Brain. International Journal of Molecular Sciences. 2020;21.
799 doi:10.3390/ijms21144869

800 3. Yosef N, Regev A. Impulse Control: Temporal Dynamics in Gene Transcription. Cell.
801 2011;144: 886–896. doi:10.1016/j.cell.2011.02.015

802 4. de Luis Balaguer MA, Fisher AP, Clark NM, Fernandez-Espinosa MG, Möller BK, Weijers
803 D, et al. Predicting gene regulatory networks by combining spatial and temporal gene
804 expression data in *Arabidopsis* root stem cells. Proc Natl Acad Sci USA. 2017;114.
805 doi:10.1073/pnas.1707566114

806 5. Mochida K, Koda S, Inoue K, Nishii R. Statistical and Machine Learning Approaches to
807 Predict Gene Regulatory Networks From Transcriptome Datasets. Front Plant Sci.
808 2018;9: 1770–1770. doi:10.3389/fpls.2018.01770

809 6. Omranian N, Eloundou-Mbebi JMO, Mueller-Roeber B, Nikoloski Z. Gene regulatory
810 network inference using fused LASSO on multiple data sets. Sci Rep. 2016;6: 20533–
811 20533. doi:10.1038/srep20533

812 7. Van den Broeck L, Gordon M, Inzé D, Williams C, Sozzani R. Gene Regulatory Network
813 Inference: Connecting Plant Biology and Mathematical Modeling. Frontiers in Genetics.
814 2020;11. Available: https://www.frontiersin.org/article/10.3389/fgene.2020.00457

815 8. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding
816 modular structure in biological networks. Nat Rev Genet. 2013;14: 719–732.
817 doi:10.1038/nrg3552

818 9. Volpato V. Integration of functional genomics data to uncover cell type-specific
819 pathways affected in Parkinson's disease. Biochemical Society Transactions. 2021;49:
820 2091–2100. doi:10.1042/BST20210128

821 10. Jamail I, Moussa A. Current State-of-the-Art of Clustering Methods for Gene Expression
822 Data with RNA-Seq. Applications of Pattern Recognition. IntechOpen; 2020.
823 doi:10.5772/intechopen.94069

824 11. Everitt BS, Landau S, Leese M, Stahl D. Cluster Analysis: Everitt/Cluster Analysis.
825 Chichester, UK: John Wiley & Sons, Ltd; 2011. doi:10.1002/9780470977811

826 12. Kaufman L, Rousseeuw PJ. Finding Groups in Data: an Introduction to Cluster Analysis.
827 Hoboken: John Wiley & Sons, Inc.; 2009. Available:
828 http://www.SLQ.eblib.com.au/patron/FullRecord.aspx?p=469065

829 13. von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007;17: 395–416.
830 doi:10.1007/s11222-007-9033-z

14. Berenhaut KS, Moore KE, Melvin RL. A social perspective on perceived distances reveals deep community structure. Proceedings of the National Academy of Sciences. 2022;119: e2003634119. doi:10.1073/pnas.2003634119

15. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008;2008: P10008. doi:10.1088/1742-5468/2008/10/P10008

16. Newman MEJ. Communities, modules and large-scale structure in networks. Nature Phys. 2012;8: 25–31. doi:10.1038/nphys2162

17. Harkey AF, Watkins JM, Olex AL, DiNapoli KT, Lewis DR, Fetrow JS, et al. Identification of Transcriptional and Receptor Networks That Control Root Responses to Ethylene. Plant Physiology. 2018;176: 2095–2118. doi:10.1104/pp.17.00907

18. Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, et al. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Weigel D, editor. eLife. 2013;2: e00675. doi:10.7554/eLife.00675

19. Lewis DR, Olex AL, Lundy SR, Turkett WH, Fetrow JS, Muday GK. A Kinetic Analysis of the Auxin Transcriptome Reveals Cell Wall Remodeling Proteins That Modulate Lateral Root Development in Arabidopsis. The Plant Cell. 2013;25: 3329–3346. doi:10.1105/tpc.113.114868

849 20. Wu T-Y, Goh H, Azodi CB, Krishnamoorthi S, Liu M-J, Urano D. Evolutionarily
850 conserved hierarchical gene regulatory networks for plant salt stress response. Nature
851 Plants. 2021;7: 787–799. doi:10.1038/s41477-021-00929-7

852 21. Gamalero E, Glick BR. Recent Advances in Bacterial Amelioration of Plant Drought and
853 Salt Stress. Biology (Basel). 2022;11: 437. doi:10.3390/biology11030437

854 22. Chen H, Bullock DA, Alonso JM, Stepanova AN. To Fight or to Grow: The Balancing Role
855 of Ethylene in Plant Abiotic Stress Responses. Plants. 2022;11.
856 doi:10.3390/plants11010033

857 23. Harkey AF, Sims KN, Muday GK. A new tool for discovering transcriptional regulators
858 of co-expressed genes predicts gene regulatory networks that mediate ethylene-
859 controlled root development. in silico Plants. 2020;2: diaa006.
860 doi:10.1093/insilicoplants/diaa006

861 24. O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and
862 Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 2016;165: 1280–
863 1292. doi:10.1016/j.cell.2016.04.038

864 25. Harkey AF, Yoon GM, Seo DH, DeLong A, Muday GK. Light Modulates Ethylene
865 Synthesis, Signaling, and Downstream Transcriptional Networks to Control Plant
866 Development. Front Plant Sci. 2019;10: 1094. doi:10.3389/fpls.2019.01094

867 26. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape
868 provides a biologist-oriented resource for the analysis of systems-level datasets.
869 Nature Communications. 2019;10: 1523. doi:10.1038/s41467-019-09234-6

870 27. Muday GK, Rahman A, Binder BM. Auxin and ethylene: collaborators or competitors?
871 Trends in Plant Science. 2012;17: 181–195. doi:10.1016/j.tplants.2012.02.001

872 28. Kazan K. Auxin and the integration of environmental signals into plant root
873 development. Annals of Botany. 2013;112: 1655–1665. doi:10.1093/aob/mct229

874 29. Aerts N, Pereira Mendes M, Van Wees SCM. Multiple levels of crosstalk in hormone
875 networks regulating plant defense. The Plant Journal. 2021;105: 489–504.
876 doi:10.1111/tpj.15124

877 30. Choudhary P, Pramitha L, Rana S, Verma S, Aggarwal PR, Muthamilarasan M.
878 Hormonal crosstalk in regulating salinity stress tolerance in graminaceous crops.
879 Physiologia Plantarum. 2021;173: 1587–1596. doi:10.1111/ppl.13558

880 31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
881 RNA-seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-
882 0550-8

883 32. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the
884 agricultural community, 2017 update. Nucleic Acids Research. 2017;45: W122–W129.
885 doi:10.1093/nar/gkx382

886

887 ## Supplemental Table and Figures

888 *Table S1. Summary of enrichment results for the ethylene dataset [18]. The top three GO*
889 *annotations with the lowest FDRs were reported. TFs with targets enriched logFC ≥1 relative*
890 *to the genome were reported.*

PaLD group	Number of transcripts	Enriched GO annotations	TFs with enriched targets
1	86	Response to chemical	None
2	128	Carbohydrate or metal ion transport; oxidation-reduction process	None
3	106	Cellular homeostasis; oxidation-reduction process; photoperiodism	ABF2, bZIP(22, 26, 28), LOB, CAMTA1
4	64	**Ethylene-activated signaling pathway**; negative regulation of signal transduction; tissue development	ABF2, ASIL2, AT1G01250, AT3G11280, AT4G18450, AT4G31060, AT5G18450, bHLH31, bZIP(1, 22, 28, 44, 57), CRF4, DDF1, EIN3, ERF(2, 3, 19 38), ESE1, GAL3, IDD11, JKD, LBD23, LOB13, PUCHI, RAP(2.1, 2.6, 2.9, 2.12), TG, TGA10, TINY, WRKY31
5	106	Monocarboxylic acid biosynthetic process; oxidation-reduction process; response to nematode	CAMTA5, MYB3R-4, MYB121
6	132	**Hormone-mediated signaling pathway**	ABF2, AT5G18450, bZIP26, DOF4.2, EIN3
7	148	Carboxylic acid transport; cell wall polysaccharide biosynthetic process; cellular glucan metabolic process	ESE1, TCP7, TCP15, TINY
8	82	None	ERF10
9	105	Cellular response to nutrient levels; **ethylene-activated signaling pathway**; protein ubiquitination	bZIP(2, 44, 48), EIN3, GATA-4

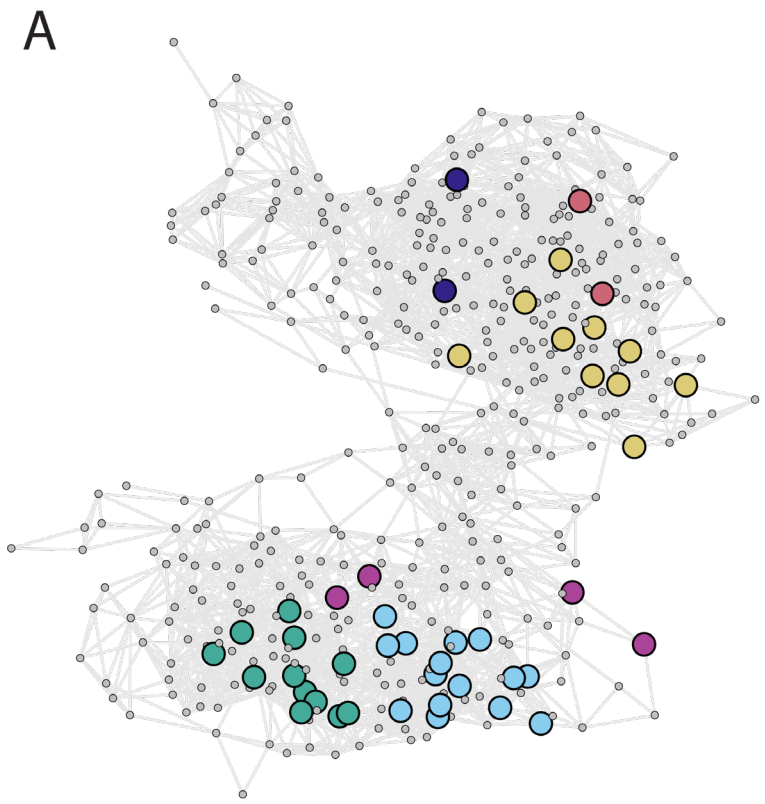


Fig. S1. Overlay of gold standard genes on the PaLD network. The colors correspond to the PaLD groups defined for the ACC network. The PaLD algorithm placed the gold standard genes into a small number of groups. Thus, the gold standard genes have similar response curves based on our defined distance function (Eq. (6)).

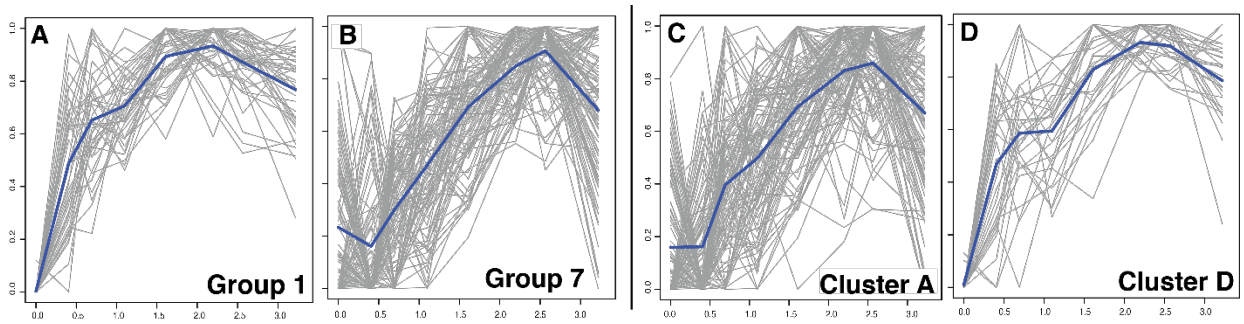


Fig. S2. provides a plot of the transformed data (prior to fitting of the cubic polynomials) for the transcripts in groups 1 and 7, alongside those for clusters A and D, with similar temporal response patterns. PaLD groups 1 and 7 are over-represented in genes involved in regulation of ethylene signaling and signal transducer activity, respectively. While cluster D is enriched in functional annotation (e.g., regulation of ethylene signaling), the large cluster A, consisting of nearly a third of the transcripts considered in Table 2, has no enrichment in function. This illustrates further the ability of the approach to more effectively group together transcripts based on common biological function. Corresponding plots for the remaining groups and clusters are included in Figs. 9 and 10.

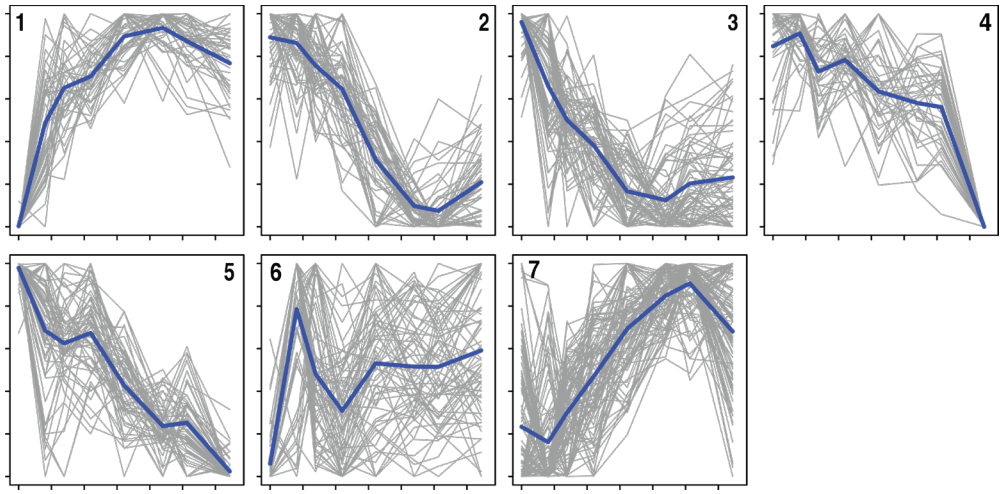


Fig. S3. The transformed abundance levels before fitting the cubic polynomials for all transcripts in PaLD groups 1–7. Indicated in blue is the average at the respective time points. Note how the curves in each group share fundamental graphical properties concerning relative maxima and minima and intervals of increase/decrease.

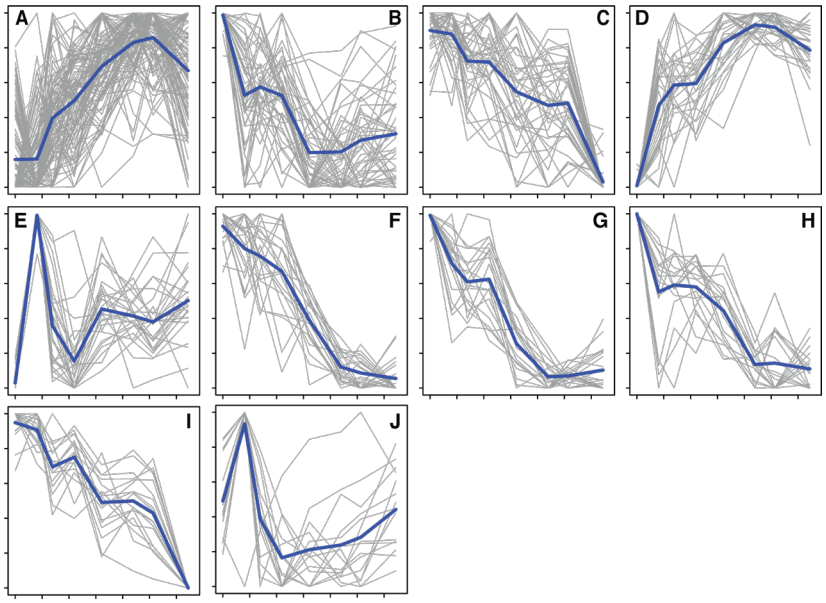


Fig. S4. The transformed abundance levels before fitting the cubic polynomials for all transcripts in the k-means clusters from Harkey et al. 2018. Indicated in blue is the average at the respective time points. The curve in each cluster share important graphical properties.