UC San Diego

UC San Diego Previously Published Works

Title

Integration of omic networks in a developmental atlas of maize

Permalink

https://escholarship.org/uc/item/6vm6h10k

Journal

Science, 353(6301)

ISSN

0036-8075

Authors

Walley, Justin W
Sartor, Ryan C
Shen, Zhouxin
et al.

Publication Date

2016-08-19

DOI

10.1126/science.aag1125

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2018 February 12.

Published in final edited form as:

Science. 2016 August 19; 353(6301): 814–818. doi:10.1126/science.aag1125.

Integration of omic networks in a developmental atlas of maize

Justin W. Walley[1,2,*], **Ryan C. Sartor**[1,*], **Zhouxin Shen**[1], **Robert J. Schmitz**[3,4,†], **Kevin J. Wu**[1], **Mark A. Urich**[3,4], **Joseph R. Nery**[4], **Laurie G. Smith**[1], **James C. Schnable**[5], **Joseph R. Ecker**[3,4,6], and **Steven P. Briggs**[1,‡]

[1]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

[2]Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA

[3]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[4]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[5]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

[6]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract

Coexpression networks and gene regulatory networks (GRNs) are emerging as important tools for predicting functional roles of individual genes at a system-wide scale. To enable network reconstructions, we built a large-scale gene expression atlas composed of 62,547 messenger RNAs (mRNAs), 17,862 nonmodified proteins, and 6227 phosphoproteins harboring 31,595 phosphorylation sites quantified across maize development. Networks in which nodes are genes connected on the basis of highly correlated expression patterns of mRNAs were very different from networks that were based on coexpression of proteins. Roughly 85% of highly interconnected hubs were not conserved in expression between RNA and protein networks. However, networks from either data type were enriched in similar ontological categories and were effective in predicting known regulatory relationships. Integration of mRNA, protein, and phosphoprotein data sets greatly improved the predictive power of GRNs.

Predicting the functional roles of individual genes at a system-wide scale is a complex challenge in biology. Transcriptome data have been used to generate genome-wide gene regulatory networks (GRNs) (1–4) and coexpression networks (5–7), the design of which was based on the presumption that mRNA measurements are a proxy for protein abundance measurements. However, genome-wide correlations between the levels of proteins and mRNAs are weakly positive (8–15), which indicates that cellular networks built solely on transcriptome data may be enhanced by integration with proteomics data. We generated an integrated developmental atlas of the transcriptome, proteome, and phosphoproteome of the

Permissions: Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl

‡Corresponding author: sbriggs@ucsd.edu.

*These authors contributed equally to this work

†Present address: Department of Genetics, Davison Life Sciences, 120 East Green Street Athens GA 30602 USA

Supplementary Materials: www.sciencemag.org/content/353/6301/814/suppl/DC1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

model organism *Zea mays* (maize) and then used these three different cellular descriptions to generate transcriptome- and proteome-based networks.

We profiled 23 tissues spanning vegetative and reproductive stages of maize development to generate a comprehensive and integrated gene expression atlas. Specifically, transcriptome profiling by mRNA sequencing (mRNA-seq) (three biological replicates, 23 tissues) was carried out on a subset of the samples used for proteome profiling (three to seven biological replicates, 33 tissues) by electrospray ionization tandem mass spectrometry (14, 16–19) (tables S1 to S3). We assessed reproducibility of the biological replicates by calculating Pearson correlations and found an average of 0.9, 0.84, and 0.7 for the transcriptome, proteome, and phosphoproteome data sets, respectively (table S4). Transcripts were observed from 62,547 genes. Proteins and phosphoproteins were observed from 16,946 and 5587 genes, respectively. The RNA-seq data were bimodal, as reported for mouse and human (20, 21), with nearly all proteins and phosphoproteins arising from the 34,455 transcripts in the high-abundance population (right peak), with an average FPKM (fragments per kilobase of exon per million fragments mapped) greater than 1 (Fig. 1A). Proteins were observed from 46% of these transcripts (right peak). To determine whether coverage of the transcriptome by the proteome was constrained by the diversity of tissues sampled, we generated proteomics data from an additional 10 tissue types yielding proteins from a total of 18,522 genes (proteins from 17,862 genes and phosphoproteins from 6185 genes), but this only increased coverage of the high-abundance transcriptome to 48%.

There are a variety of possible technical and biological explanations for why we detect proteins from less than half of the high-abundance transcript-producing genes and why we do not observe corresponding mRNA for 245 quantified proteins. Previously, we found evidence for multiple mechanisms that may explain the detection of proteins but not mRNA. These mechanisms include (i) differential stability of mRNA and proteins; (ii) transport of proteins between tissues; and (iii) diurnal, out-of-phase accumulation of mRNAs and cognate proteins (14). The heightened sensitivity of transcriptomics relative to proteomics likely provides a partial explanation for why we detect proteins corresponding to less than half of the transcript-producing genes. Additionally, we observed a greater percentage of proteins arising from the annotated filtered gene set, which consists of 39,656 high-confidence gene models that exclude transposons, pseudogenes, and other low-confidence members present in the working gene set (Fig. 1B). Furthermore, a higher proportion of proteins than transcripts arise from genes annotated as protein coding (Fig. 1C), which suggests that transcripts from many genes may not produce proteins. Genes conserved at syntenic orthologous locations between maize and sorghum exhibited a unimodal, high-expression pattern, in contrast to genes in nonsyntenic locales (Fig. 1A). Considering all genes that expressed mRNAs, syntenic genes were nine times more likely than nonsyntenic genes to express proteins (Fig. 1D). To show that this observation is not due to the higher average transcript expression level of syntenic genes, we examined a range of transcript abundance cutoffs and obtained similar results, even when looking at the highest-abundance syntenic and nonsyntenic transcripts (fig. S1). A greater frequency of protein expression is a possible mechanistic explanation for the eightfold enrichment of genes responsible for visible mutant phenotypes among syntenically conserved genes in maize (22).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We next examined how genes and biological processes change throughout development. Initially, we focused on transcription factors (TFs), as they are key regulators of development, growth, and cell fate. Of the 2732 annotated TFs and transcriptional co-regulators, we detected 2627 as mRNA (23 tissues), 1026 as protein (33 tissues), and 559 as phosphoprotein (33 tissues). We used hierarchical clustering to identify 712 (mRNA), 469 (protein), and 419 (phosphoprotein) TFs that exhibited tissue-specific enrichment (figs. S2 to S4 and table S5). We also examined expression trends at the TF family level. First, we used traditional overrepresentation analysis to identify TF families whose members are detected in a given tissue at a greater frequency than chance (figs. S5A, S6A, and S7A). To augment the overrepresentation analysis, we also examined TF family-level expression profiles by quantifying the total amount of each TF family's mRNA, protein, and/or phosphoprotein present in given tissue (figs. S5B, S6B, and S7B). Taken together, these data describe the spatiotemporal expression pattern of individual TFs and TF families across development.

We expanded our analyses to examine the patterns of all gene types across maize development. We used the weighted gene coexpression network analysis (WGCNA) R package (23) to group similarly expressed genes—detected as mRNA (23 tissues), protein (33 tissues), or phosphoprotein (33 tissues) in at least four tissues—into modules (clusters). This approach enabled us to group 31,447 mRNAs, 13,175 proteins, and 4267 phosphoproteins into coexpression modules (fig. S8 and table S6). We next plotted the eigengene profile for each module in order to assign the tissue(s) in which each module is highly expressed (figs. S9 to S12). We observed that 36 well-characterized genes required for maize development—including the homeobox TFs Knotted1 [KN1, Maize Genetics and Genomics Database (MGGD) accession number GRMZM2G017087] (24) and Rough Sheath 1 (RS1, MGGD accession number GRMZM2G028041) (25), as well as the transcriptional co-repressor Ramosa1 Enhancer Locus2 (REL2, MGGD accession number GRMZM2G042992) (26) (table S6)—are present in mRNA, protein, and phosphoprotein modules that correspond to dividing and meristematic tissues. The phosphorylation pattern of these proteins is similar to their mRNA profile and occurs in tissues known to have altered developmental phenotypes in mutant plants, which suggests that phosphorylation of these proteins might positively regulate their function. Finally, we determined overrepresentation of MapMan functional categories in each module (table S6). As expected, we found that genes involved in photosynthetic light reactions have mRNA and protein that are enriched predominantly in the mature leaf. We did not detect an enrichment of light-reaction phosphoproteins in the mature leaf module, which suggests that phosphorylation is not a major regulator of the light reactions (fig. S11 and table S6).

Biological networks can be constructed based on many different types of data and serve to elucidate the structure underlying complex systems. Typically, transcript profiling data are used to generate various types of gene expression networks. However, we observed a weakly positive correlation between mRNA and protein levels in our data set (supplementary text, figs. S13 to S17, and table S7), in agreement with research done in a range of organisms (8–15). Although the modest correlation between mRNA and protein levels is well documented, a major outstanding question is whether transcriptome-based networks predict the same relationships as proteome-based networks. Given our extensive developmental gene expression atlas, we addressed this question by generating two different types of networks:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

coexpression networks and GRNs. We first generated coexpression networks (table S8), which are undirected networks where nodes are genes connected on the basis of highly correlated expression patterns (Fig. 2A) (5–7). For these network reconstructions, we used 10,979 genes that were detected as both transcripts and proteins in at least 5 of the 23 developmental gene expression atlas tissues in which we profiled both mRNA and protein. Pairwise mRNA-to-mRNA and protein-to-protein coexpression networks were built with Spearman correlations using WGCNA (fig. S18 and table S8). The biweight midcorrelation yielded similar results (figs. S19 and S20). To directly compare the mRNA- and protein-based coexpression networks and compile a high-confidence coexpression data set, each network was constrained to include only edges with a correlation score >0.75 (top 1 million edges), which is a frequently used correlation threshold for coexpression networks (table S8). As a measure of similarity, we calculated edge conservation by dividing the set intersect by the union (known as the Jaccard index) and reported this as a percentage. We found that 122,029 of the combined 2 million edges (6.1%) were conserved in both networks (Fig. 2B). Though this edge overlap is greater than the 0.8% expected by chance (P value = 0), the majority of relationships between genes were specific to each network, even when we expanded the network size to 10 million edges (fig. S20).

To examine whether the lack of edge overlap was due to experimental noise, we used single biological replicates (three mRNA and three protein networks) to create six new co-expression networks. Pairwise comparisons revealed a similar low level of edge conservation (5%) between the mRNA and protein coexpression networks. However, 46% of mRNA-to-mRNA edges and 36% of protein-to-protein edges were conserved between replicate coexpression networks (fig. S21). These data suggest that biological phenomena underpin the observed lack of edge conservation between transcriptome- and proteome-derived coexpression networks.

A key feature of scale-free networks is a small number of highly interconnected hubs. Because hubs are more likely than nonhubs to be required for network integrity and organism survival, the identification of so-called “hub genes” is of interest (23,27–30). We therefore determined the highly interconnected hub genes in each coexpression network, which we categorized as nodes in the top 10th percentile for most edges (Fig. 2C and fig. S22A). When we compared the hub genes from each network, we found that the majority (85%) were not shared between the mRNA and protein coexpression networks (Fig. 2C and fig. S22).

Groups of coexpressed genes (modules) were derived from the two networks. Each module was examined for over- or underrepresentation of MapMan categories (table S9). The majority of modules from each network (mRNA: 17 of 19; protein: 18 of 25) showed significant enrichment for one or more categories (adjusted P value < 0.05). Overall, we observed similar enrichment of categories between the two coexpression networks (fig. S23). Whereas the overall degree of enrichment was very similar for most categories in both coexpression networks, the actual genes that accounted for the significantly enriched categories were mostly specific to one network (35% protein-specific, 27% mRNA-specific, and 38% shared) (Fig. 3). Taken together, these results demonstrate that transcript- and protein-based coexpression networks yield differing predictions of gene relatedness and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function. Presumably, the discrepancy between transcriptome and proteome coexpression networks arises from the limited correlation between mRNA and protein abundance, which has been attributed to a range of factors that include differing stabilities of mRNA and protein, translational control, and protein movement from the tissue of synthesis (8, 14, 31).

To further explore the regulatory patterns of gene expression across maize development, we generated GRNs, which are directed networks of TFs and their target genes (Fig. 4A) (1). Unsupervised GRNs were created using GENIE3, which takes advantage of the random forest machine learning algorithm and was the top-performing method in the DREAM4 and -5 GRN reconstruction challenges (32, 33). Three independent GRNs were generated from the 23 tissues in which we profiled both mRNA and protein. To construct these networks, we varied whether the TFs (termed “regulators”) were quantified as mRNAs (2200 TFs), proteins (545 TFs), or phosphopeptides (441 TFs) and used a common set of 41,021 quantified mRNAs (termed “target genes”) (table S10). We evaluated the GRNs by using published data for two classical maize TFs, the homeobox TF KN1 and the bZIP TF Opaque2 (O2). These TFs were chosen as benchmarks because they have been the subject of high-quality RNA-seq and chromatin immunoprecipitation (ChIP)–seq studies in both wild-type and null mutant backgrounds, and they represent two distinct types of TFs with key developmental roles (24, 34). Target genes are bound by their TF in a ChIP-seq assay, and their mRNA levels change when their TF is knocked out. Using the published direct targets of KN1 and O2, we generated receiver operating characteristic (ROC) and precision-versus-recall curves, which are two methods commonly used to evaluate the power of a predictive model (35). These curves showed that the overall qualities of all three GRNs were similar (fig. S24). However, when we looked at the top 500 scoring GENIE3 predictions for KN1 and O2 in each GRN, we observed a performance advantage for the two protein-based GRNs in accurately predicting target genes (Fig. 4B and fig. S25A). Specifically, the KN1 subnetworks accurately predicted 108 (mRNA), 129 (protein), and 125 (phosphopeptide) targets, with the O2 subnetworks performing similarly. Additionally, 44% (KN1) and 31% (O2) of all correctly predicted targets were specific to a single type of GRN (Fig. 4B and fig. S25A). These results indicated that predictions made by all three GRNs were largely complementary to each other.

We expanded our analyses to examine all TFs in the three GRNs. Again, we found that there was low edge conservation between the GRNs, with the vast majority of edges being present in a single GRN (fig. S26). Specifically, when considering one million edges, 93% were present in a single GRN (Fig. 4C). This amount increased to 96% for the 200,000 highest-confidence predictions, which we determined using KN1 precision data as the cutoff (fig. S25, B and C). This finding illustrates that the different accumulation patterns of mRNA, protein, and phosphorylation for a given TF (fig. S27) result in disparate GRN predictions.

The three preceding GRNs were constructed using different-sized sets of TF regulators, which complicated direct comparisons of networks constructed using TF abundance measurements at the mRNA or protein level. Therefore, we used 539 TFs quantified as both mRNAs and proteins to reconstruct GRNs. Evaluation of these GRNs using the KN1 and O2 data indicated quality and accuracy similar to those of the full-sized networks (fig. S28). We still observed a performance advantage for the protein GRN, as well as limited edge

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conservation between the mRNA- and protein-based GRNs, with only 6% of the top 200,000 edges being shared (figs. S28 and S29). We examined several possible features of the TF regulators to help further our understanding of the limited overlap in TF target predictions. The TFs connected by edges that were present only in the transcript GRN had lower and more variable protein abundance than the TFs connected by edges that were shared with or specific to the protein GRN (fig. S30, A to D). As expected, the mRNA-to-protein correlations were higher for targets of edges present in both GRNs (fig. S30E).

To further validate GRN predictions and test whether network relationships were consistent between different maize varieties, we took advantage of natural variation in regulator abundance arising from the natural genetic variation present in another inbred line, Mo17. Specifically, we compared mRNA and protein abundance in primary roots of Mo17 to B73. Whereas most TFs and target genes were expressed at similar levels in B73 and Mo17, we identified 149 (mRNA), 26 (protein), and 16 (phosphopeptide) regulatory TFs that were expressed at significantly different levels. We found, with high confidence, that for many of these differentially expressed TFs, their GRN predicted target groups were also significantly enriched for differentially expressed transcripts (figs. S31 to S33). Thus, elements of the GRN structure were preserved, and quantitative changes in regulator abundance levels are associated with altered network output and gene expression patterns. Additionally, these findings validated the GRN approaches used in this study and demonstrated the utility of applying this method to examine dynamics of gene regulation.

After analyzing separate mRNA- and protein-based GRNs, we considered integrating the data sets to determine whether the resulting single GRN would have improved inference over the individual GRNs. Specifically, we constructed four additional GRNs, each consisting of combinations of TF regulators quantified as mRNA, protein, and/or phosphopeptides (table S10). Details of how the combined mRNA, protein, and phosphopeptide GRNs were made are described in the supplementary materials. We examined the performance of the resulting networks using the validation set of KN1 and O2 published targets (Fig. 4 and fig. S24). All GRNs reconstructed with combinations of TF regulators performed better than single-input GRNs. This finding demonstrates that integrating readouts of gene expression quantified at different levels results in improved GRN inference. Our use of TF mRNA levels to infer TF activity had provided good GRN predictive power. The area under the ROC curve (AUC) was 0.657, compared with 0.500 for random predictions. When the mRNA measurements were combined with protein abundance and phosphorylation levels to infer TF activity, the AUC increased to 0.717. Thus, if an investigator wished to use network predictions with a false-positive rate of 20%, the mRNA-only network would predict 40% of the true positives, compared with 50% for the combined network (Fig. 4D and fig. S24A). Likewise, examination of Fig. 4E and fig. S24B reveals that if an investigator wished to use network predictions with a precision of 0.021 (which is three times higher than expected at random), then 16% of the true positives would be recalled from the mRNA-only network versus 41% for the combined network.

By quantitatively measuring mRNAs, proteins, and phosphoproteins in parallel in a tissue-specific manner, we discovered unexpected relationships among these cellular readouts across maize development. In particular, our comparison of transcriptome- to proteome-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

based dendrograms and coexpression networks showed little overlap at the gene level, even though the samples were classified similarly and had similar ontological enrichments. The discovery that most protein-expressing genes are conserved and syntenic also was unexpected. The coexpression networks and GRNs provide a conceptual framework for future detailed studies in a model organism that is central to food security and bioenergy. Our findings highlight the importance of studying gene regulation at multiple levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank V. Walbot for tassel and anther tissues, F. Hochholdinger for root tissues, and J. Fowler for pollen tissues. This work was supported by the NSF (grant 0924023 to S.P.B. and grants MCB-0929402 and MCB1122246 to J.R.E.), an NIH National Research Service Award Postdoctoral Fellowship (F32GM096707 to J.W.W.), and the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation (grant GBMF3034 to J.R.E.). Sequence data can be downloaded from the National Center for Biotechnology Information's Sequence Read Archive (accession number GSE50191). Raw mass spectra have been deposited at the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository (accession numbers MSV000079290 and MSV000079303). Normalized expression data have been integrated into the genome browser and individual gene model pages at the central Maize Genetics and Genomics Database (www.maizegdb.org/).

References and Notes

1. Krouk G, Lingeman J, Colon AM, Coruzzi G, Shasha D. Genome Biol. 2013; 14:123. [PubMed: 23805876]
2. Gardner TS, Faith JJ. Phys Life Rev. 2005; 2:65–88. [PubMed: 20416858]
3. Bar-Joseph Z, et al. Nat Biotechnol. 2003; 21:1337–1342. [PubMed: 14555958]
4. De Smet R, Marchal K. Nat Rev Microbiol. 2010; 8:717–729. [PubMed: 20805835]
5. van Noort V, Snel B, Huynen MA. Trends Genet. 2003; 19:238–242. [PubMed: 12711213]
6. Stuart JM, Segal E, Koller D, Kim SK. Science. 2003; 302:249–255. [PubMed: 12934013]
7. Horvath S, Dong J. PLOS Comput Biol. 2008; 4:e1000117. [PubMed: 18704157]
8. Schwanhäusser B, et al. Nature. 2011; 473:337–342. [PubMed: 21593866]
9. Vogel C, et al. Mol Syst Biol. 2010; 6:400. [PubMed: 20739923]
10. Ghaemmaghami S, et al. Nature. 2003; 425:737–741. [PubMed: 14562106]
11. Baerenfaller K, et al. Science. 2008; 320:938–941. [PubMed: 18436743]
12. Ghazalpour A, et al. PLOS Genet. 2011; 7:e1001393. [PubMed: 21695224]
13. Ponnala L, Wang Y, Sun Q, van Wijk KJ. Plant J. 2014; 78:424–440. [PubMed: 24547885]
14. Walley JW, et al. Proc Natl Acad Sci USA. 2013; 110:E4808–E4817. [PubMed: 24248366]
15. Washburn MP, et al. Proc Natl Acad Sci USA. 2003; 100:3107–3112. [PubMed: 12626741]
16. Qiao H, et al. Science. 2012; 338:390–393. [PubMed: 22936567]
17. Chang KN, et al. eLife. 2013; 2:e00675. [PubMed: 23795294]
18. Facette MR, Shen Z, Björnsdóttir FR, Briggs SP, Smith LG. Plant Cell. 2013; 25:2798–2812. [PubMed: 23933881]
19. Marcon C, et al. Plant Physiol. 2015; 168:233–246. [PubMed: 25780097]
20. Hebenstreit D, et al. Mol Syst Biol. 2011; 7:497. [PubMed: 21654674]
21. Nagaraj N, et al. Mol Syst Biol. 2011; 7:548. [PubMed: 22068331]
22. Schnable JC, Freeling M. PLOS ONE. 2011; 6:e17855. [PubMed: 21423772]
23. Langfelder P, Horvath S. BMC Bioinformatics. 2008; 9:559. [PubMed: 19114008]
24. Bolduc N, et al. Genes Dev. 2012; 26:1685–1690. [PubMed: 22855831]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

25. Schneeberger RG, Becraft PW, Hake S, Freeling M. Genes Dev. 1995; 9:2292–2304. [PubMed: 7557382]
26. Gallavotti A, et al. Development. 2010; 137:2849–2856. [PubMed: 20699296]
27. Mukhtar MS, et al. Science. 2011; 333:596–601. [PubMed: 21798943]
28. Albert R, Jeong H, Barabasi AL. Nature. 2000; 406:378–382. [PubMed: 10935628]
29. Calderwood MA, et al. Proc Natl Acad Sci USA. 2007; 104:7606–7611. [PubMed: 17446270]
30. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Nature. 2001; 411:41–42. [PubMed: 11333967]
31. Baerenfaller K, et al. Mol Syst Biol. 2012; 8:606. [PubMed: 22929616]
32. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. PLOS ONE. 2010; 5:e12776. [PubMed: 20927193]
33. Marbach D, et al. Nat Methods. 2012; 9:796–804. [PubMed: 22796662]
34. Li C, et al. Plant Cell. 2015 10.1105/tpc.114.134858.
35. Schrynemackers M, Küffner R, Geurts P. Front Genet. 2013; 4:262. [PubMed: 24348517]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

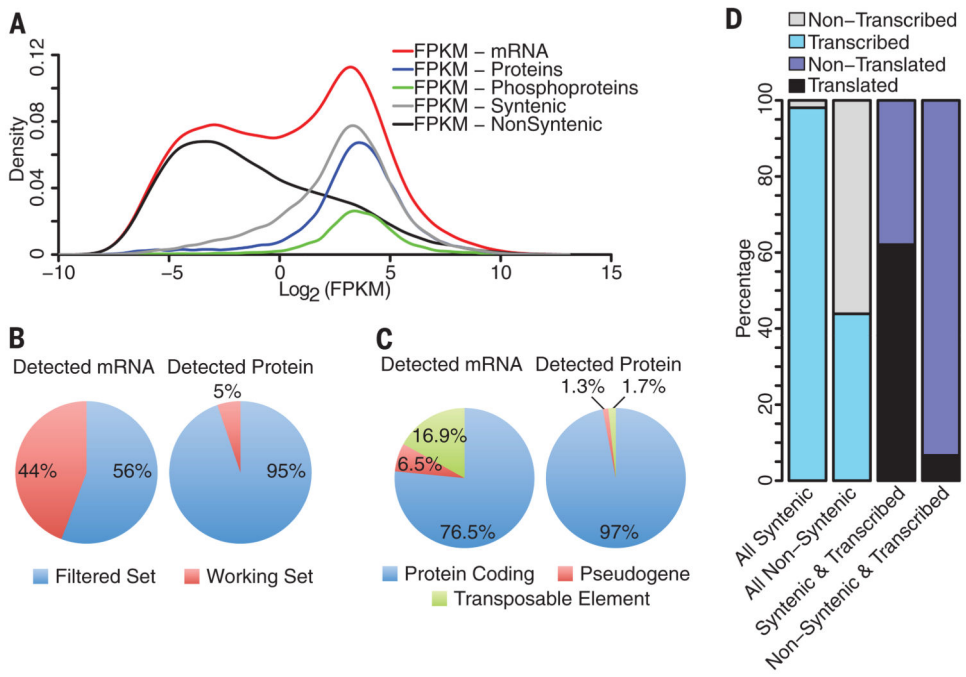

Fig. 1. Comparison of transcriptome and proteome data sets

(**A**) FPKM distribution of mRNA abundance (red). FPKM values of transcripts corresponding to quantified proteins (blue), phosphopeptides (green), syntenic genes conserved between maize and sorghum (gray), and nonsyntenic genes (black) are shown. Data are the average expression from the 23 tissues profiled. (**B**) Percentage of quantified mRNA and proteins in the annotated filtered (high-confidence gene models) and working (all gene models) gene sets. (**C**) Breakdown of detected mRNA and proteins, based on annotations. (**D**) Percentages of all annotated genes that are transcribed and percentages of all transcribed genes that are translated, for both the syntenic and nonsyntenic gene sets.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

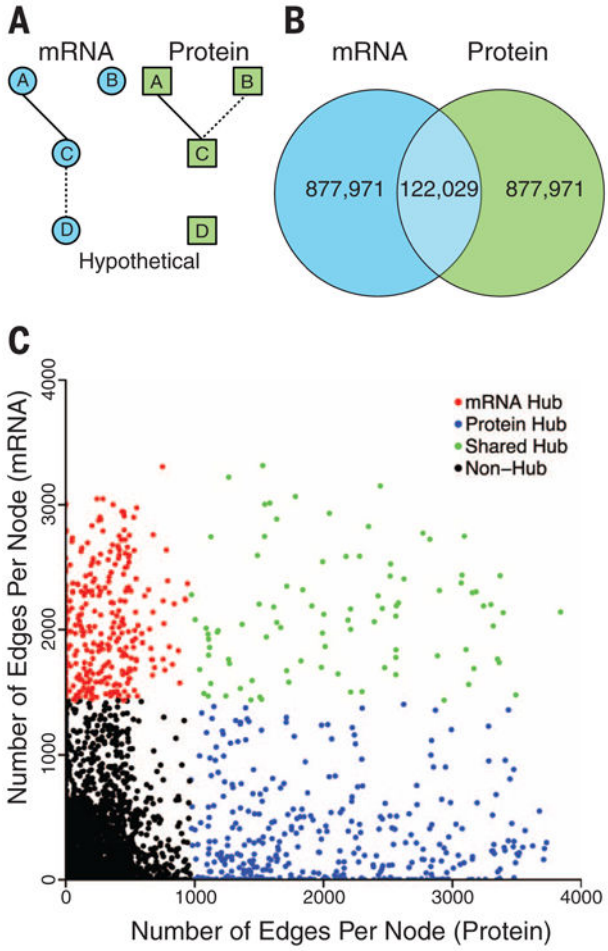

Fig. 2. Coexpression network analyses

(**A**) Hypothetical undirected coexpression subnetwork showing conserved (solid lines) and nonconserved (dotted lines) coexpression edges between mRNA and protein networks. (**B**) Venn diagram depicting edge conservation (solid lines in Fig. 2A) between the two coexpression networks. (**C**) Number of edges a given gene (node) has in the protein (*x* axis) and mRNA (*y* axis) coexpression networks. Nodes above the 90th percentile for the number of edges are considered hubs and are colored according to whether they are hubs in the protein (blue) or mRNA (red) network or both (green). Black dots represent non-hub nodes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

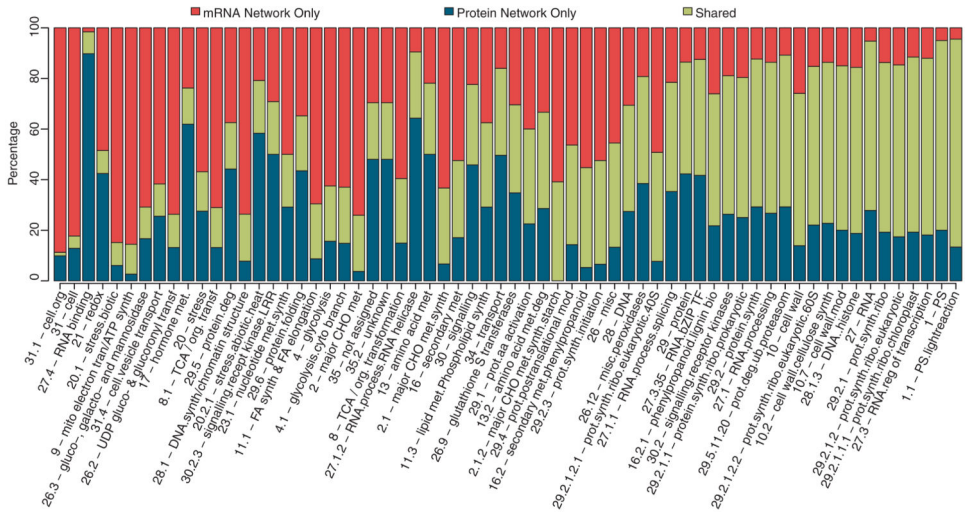

Fig. 3. Categorical enrichment analysis of coexpression modules

Coexpression modules were determined by WGCNA and functionally annotated using MapMan categories. Categories enriched (Benjamini-Hochberg adjusted *P* value ≤ 0.05) in one or more modules are represented by vertical bars and labeled with the bin number and name. For each category, the genes accounting for the enrichment were extracted separately from mRNA and protein modules. Only functional categories with at least 20 genes are shown. Colored bars represent the proportion of genes in each enriched category that are specific to one network (mRNA, red; protein, blue) or shared between the networks (green).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

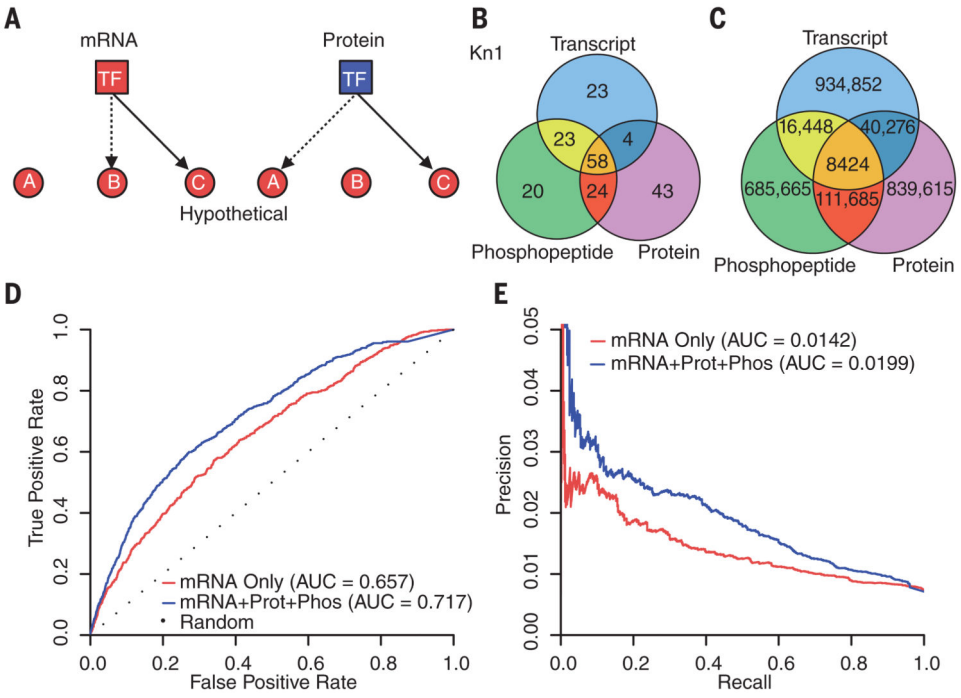

Fig. 4. Unsupervised GRN analyses

(**A**) Hypothetical GRN subnetwork depicting a TF regulator (square) and potential target genes (circle) quantified as mRNA (red) or protein (blue). GRN-specific and -conserved predictions are depicted by dotted and solid lines, respectively. (**B**) Overlap of the true-positive predictions from the top 500 true GRN predictions for KN1 quantified as mRNA, protein, or phosphopeptide. True KN1 targets were identified by Bolduc *et al.* (24). (**C**) Overlap of the top 1 million TF target predictions between the GRNs reconstructed using TF abundance quantified at the mRNA, protein, or phosphopeptide level. (**D**) ROC curves and (**E**) precision-recall curves generated using known Kn1 and O2 target genes for a mRNA-only GRN (red) and a fully integrated GRN built by combining mRNA, protein, and phospho-protein data into a single GRN (blue).