Exploring Functional Transcriptomics using Gene Regulatory Networks in Maize

Peng Zhou[1], Erika Magnusson[1], Zhi Li[2], Sarah N. Anderson[1], Peter Crisp[1], Jaclyn Noshay[1], Peter Hermanson[1], Candice N. Hirsch[2], Steven P. Briggs[3], Nathan M. Springer[1]

[1] Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108
[2] Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108
[3] Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

UNIVERSITY OF MINNESOTA

UC San Diego

Highlights

- 43 GRNs of different types (cross-tissue, cross-genotype, tissue-and-genotype) were constructed using public maize transcriptome datasets (**Table 1**);
- Most GRNs receive support from well-characterized transcription factors (TFs) and conserved TF binding sites (**Figure A**), and exhibit significant enrichment of biologically relevant interactions (Gene Ontology, CornCyc pathways, **Figures B and C**);
- Using the paired B73&Mo17 developmental expression dataset (**Figure D**), we show that the presence/absence of expression - rather than the relative expression levels - of a TF, tend to significantly affect target gene expression;
- Further support of GRNs comes from the BiomAP RNA-Seq dataset (**Figure E**), where we show a comprehensive developmental atlas network and several large-size tissue-specific networks having the highest predictive power across different test settings;
- GRNs are also supported by previous eQTL studies in maize (**Figure F**). Combining TF-target predictions with previously identified *trans*-eQTL hotspot information, we are able to pinpoint the *trans*-acting factors underlying 126 eQTL hotspots, clarifying regulatory relationship for a total of 238 TFs involved in a variety of biological pathways (**Figure G**).

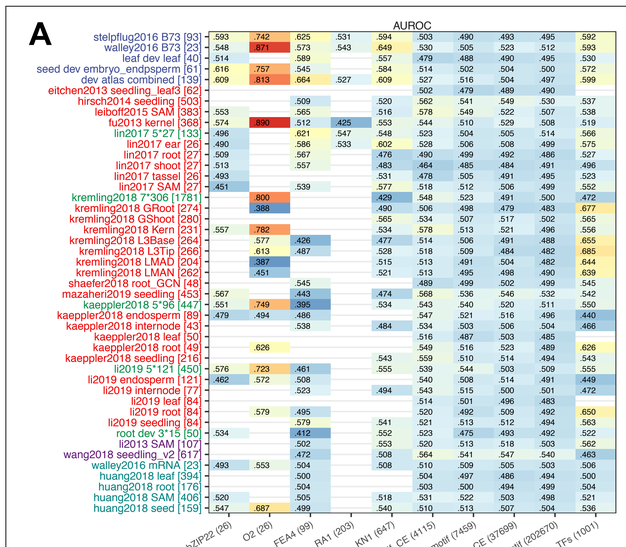


Figure A. Area under receiver-operating characteristic curve (AUROC) for each GRN using 5 maize TFs with known targets as ground truth. Numbers in each cell show the actual AUROC or AUPR values with white cells indicating missing data (the TF being filtered from the expression matrix due to invariable expression).

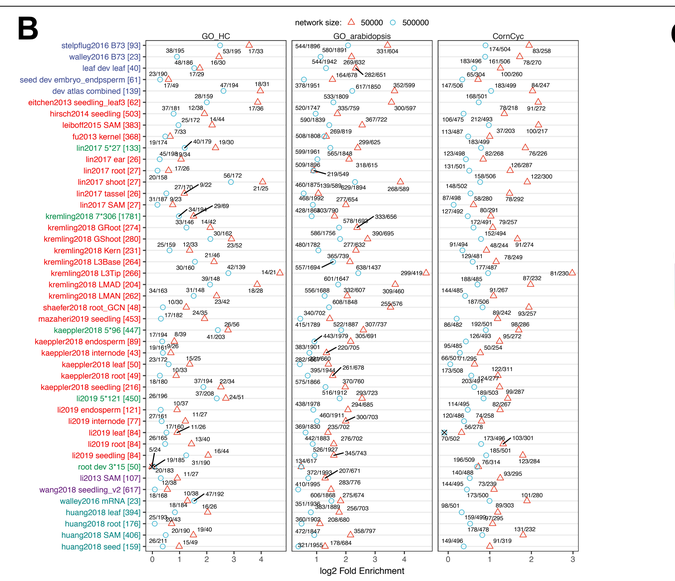
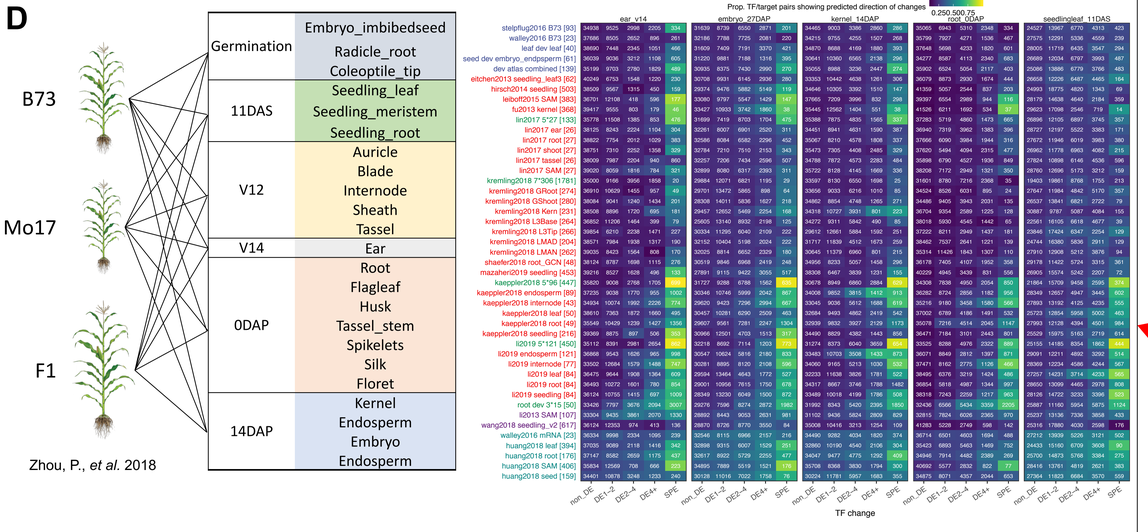


Figure B. Enrichment of co-annotated GO/CornCyc terms in co-regulated network targets. For each network either the top 50,000 edges (red triangle) or 500,000 edges (blue circle) was taken to assess enrichment of GO/CornCyc annotation. Log2 fold enrichment is calculated as the observed number of shared GO/CornCyc terms (by targets regulated by a common TF) divided by the expected number of shared annotation terms (determined by permutation).

Figure C. Hierarchical clustering of 140 Gene Ontology (Uniprot.Plants) terms using log2 fold enrichment in different GRNs. Only GO terms enriched in at least 5 out of the 43 networks were used for clustering. Insignificant enrichment (p-value < 0.05) was treated as missing data and shown as white space in the figure.



Zhou, P., *et al.* 2018

Figure D. (Left) Experimental design of the B73 & Mo17 paired developmental atlas study (3 genotypes X 23 tissues) and (right) proportion of differentially expressed targets regulated by TFs showing different DE levels. Each TF-target pair is classified according to the DE level of the TF ("non_DE", "DE1-2", "DE2-4", "DE4+" or "SPE") in each network. The proportion of TF-target pairs with the target also showing DE was then determined for each category. Numbers in each cell stands for the total number of TF-target pairs falling in that category.

Table 1. GRNs built in this study.

net_type	study	note	sample_size	reference
tissue	stelpflug2016	B73	93	Stelpflug et al. 2016
	walley2016	B73	23	Walley et al. 2016
	leaf dev	leaf	40	Leaf Dev
	seed dev	embryo_endpsperm	61	Seed Dev
	dev atlas	combined	139	Dev Atlas
genotype	eitchen2013	seedling_leaf3	62	Eitchen et al. 2013
	hirsch2014	seedling	503	Hirsch et al. 2014
	leibotf2015	SAM	383	Leiboff et al. 2015
	fu2013	kernel	368	Fu et al. 2013
	lin2017	ear	26	Lin et al. 2017
		root	27	
		shoot	27	
		tassel	26	
		SAM	27	
	kremling2018	GRoot	274	Kremling et al. 2018
		GShoot	280	
		Kern	231	
		L3Base	264	
		L3Tip	266	
		LMAD	204	
		LMAN	262	
	shaefer2018	root_GCN	48	Shaefer et al. 2018
	mazaheri2019	seedling	453	Mazaheri et al. 2019
	kaeppler2018	endosperm	89	Kaeppler et al. 2018
		internode	43	
		leaf	50	
		root	49	
		seedling	216	
	li2019	endosperm	121	Li et al. 2019
		internode	77	
		leaf	84	
		root	84	
		seedling	84	
tissue*genotype	lin2017	5*27	133	Lin et al. 2017
	kremling2018	7*306	1781	Kremling et al. 2018
	kaeppler2018	5*96	447	Kaeppler et al. 2018
	li2019	5*121	450	Li et al. 2019
	root dev	3*15	50	Root Dev
ril	li2013	SAM	107	Li et al. 2013
	wang2018	seedling_v2	617	Wang et al. 2018
liftover	walley2016	mRNA	23	Walley et al. 2016
	huang2018	leaf	394	Huang et al. 2018
		root	176	
		SAM	406	
		seed	159	

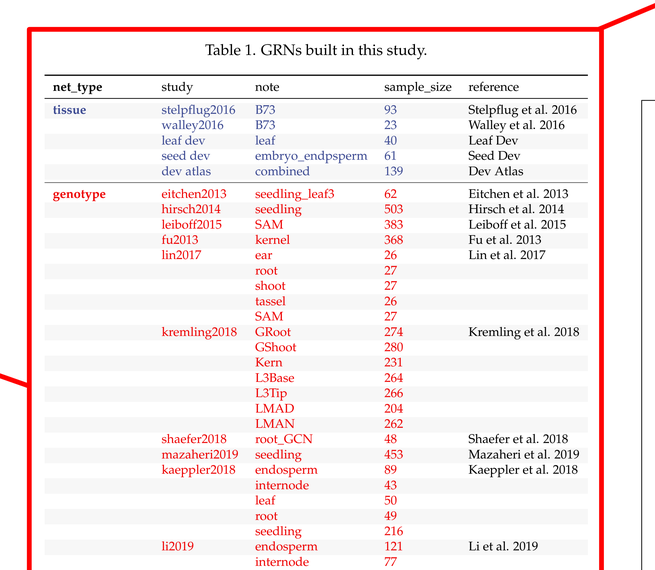
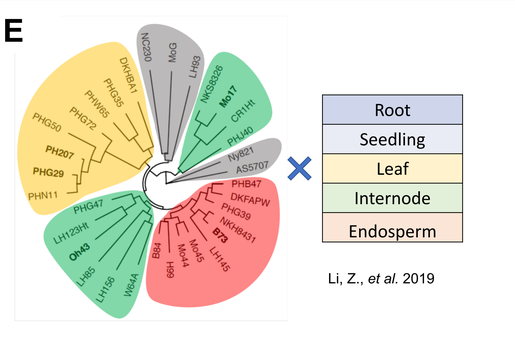


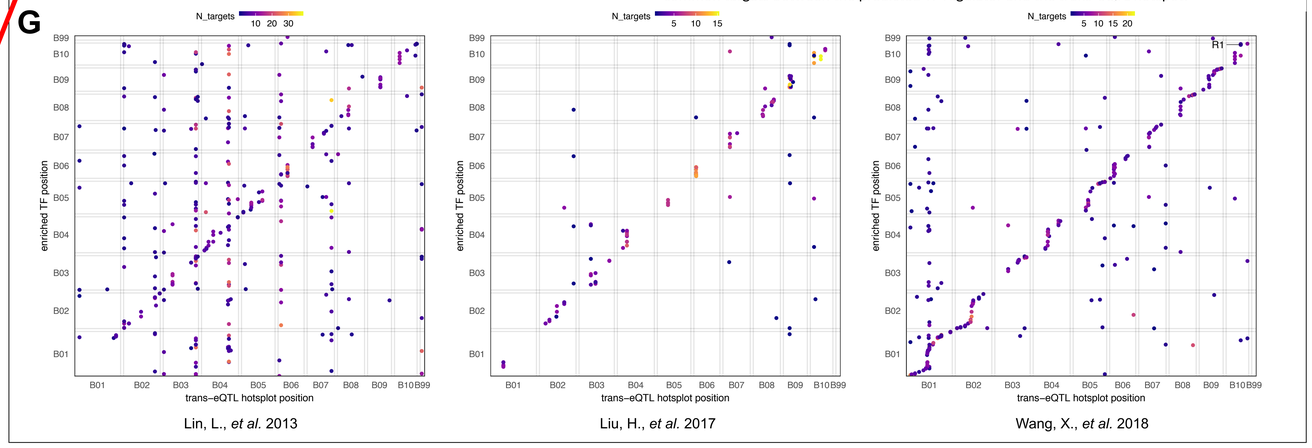
Figure F. Enrichment of co-regulated targets between known *trans*-eQTL hotspots (identified in previous studies) and GRN-predicted TF-target associations. For each network either the top 50,000 edges (red triangle) or 500,000 edges (blue circle) was taken to assess enrichment of co-regulation. Log2 fold enrichment is determined by the same permutation approach described in Figure B.



Li, Z., *et al.* 2019

Figure E. . (Left) Experimental design of the BiomAP study (33 inbred lines X 5 tissues) and (right) Enrichment in significantly correlated TF-target pairs based on the BiomAP dataset. Only the top 50k TF-target pairs from each GRN was taken. Color in each cell represents the fold enrichment of significantly correlated (target expression first ordered by TF level, followed by a Mann-Whitney U test between genotypes where TF shows CPM<0.1 and those with TF showing CPM>1, P-value < 0.05) TF-target proportions relative to permutation average (TF-target pairs randomly shuffled). Texts in each cell represents the total number of testable TF-target pairs, the observed proportion of significant pairs, and the fold enrichment over expected (permutated) proportion of pairs showing significance.

Figure G. Identification of acting transcription factors underlying *trans*-eQTL hotspots in previous eQTL studies. For each *trans*-eQTL hotspot identified in three previous eQTL studies, the physical genomic locations were extracted and converted to AGPv4 coordinates, and the regulated targets (i.e., eTraits) for each *trans*-eQTL hotpot were also converted to AGPv4 gene IDs. Each dot stands for a significant co-regulation between a *trans*-eQTLhotspot (x-axis) and a TF in GRN (y-axis) with the level of significance determined in Figure F. Color of dot represents the number of common targets between the predicted TF regulator and the *trans*-eQTL hotspot.



Lin, L., *et al.* 2013 | Liu, H., *et al.* 2017 | Wang, X., *et al.* 2018