## Exploring Functional Transcriptomics using Gene Regulatory Networks in Maize



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

<sup>1</sup> Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108

<sup>2</sup> Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108 <sup>3</sup> Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

## UC San Diego



## University of Minnesota

## 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**).

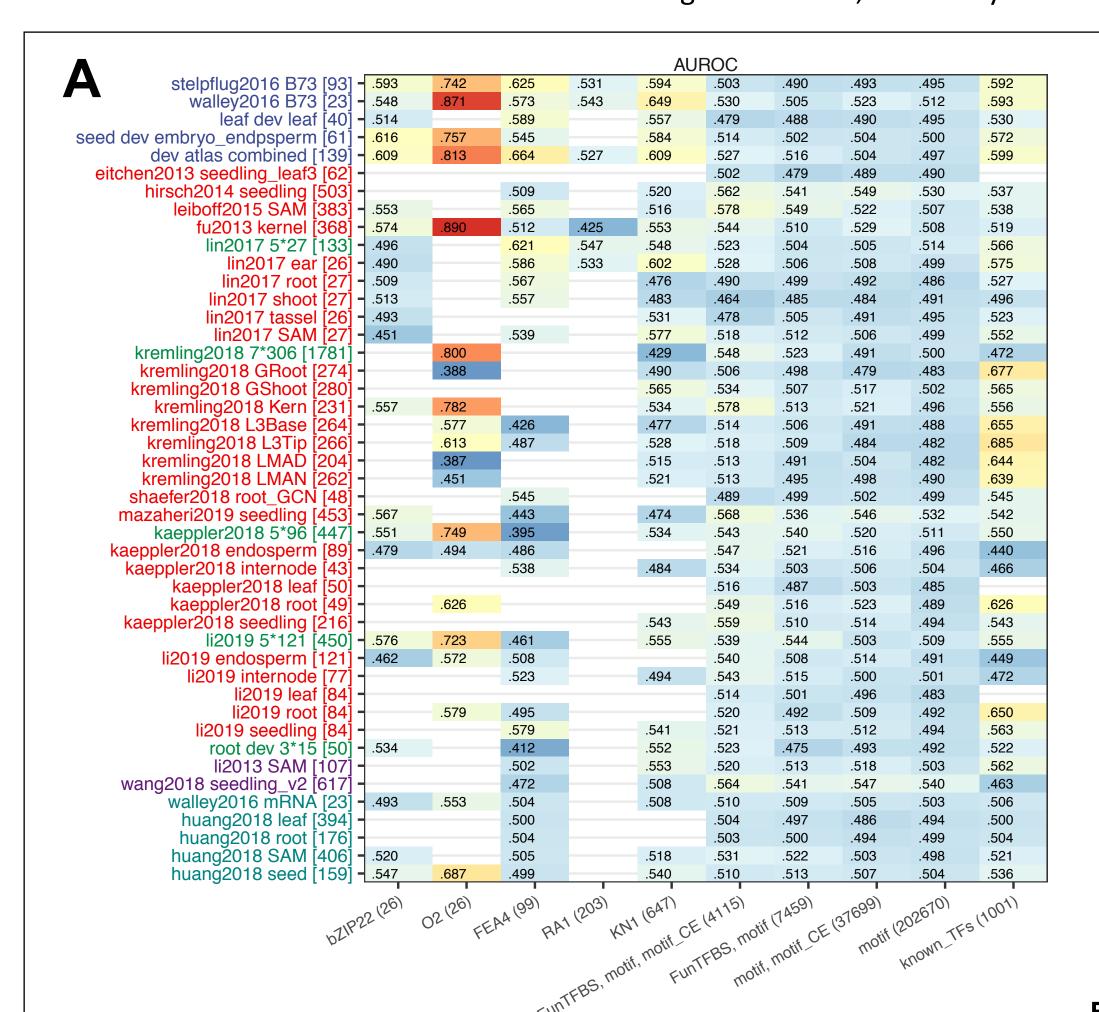


Figure A. Area under receiver-operating characteristic curve (AUROC) for each GRN using 5 maize TFs with known targets as ground truth. Numbers in each cell show the actual AUROC or AUPR values with white cells indicating missing data (the TF being filtered from the expression matrix due to invariable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

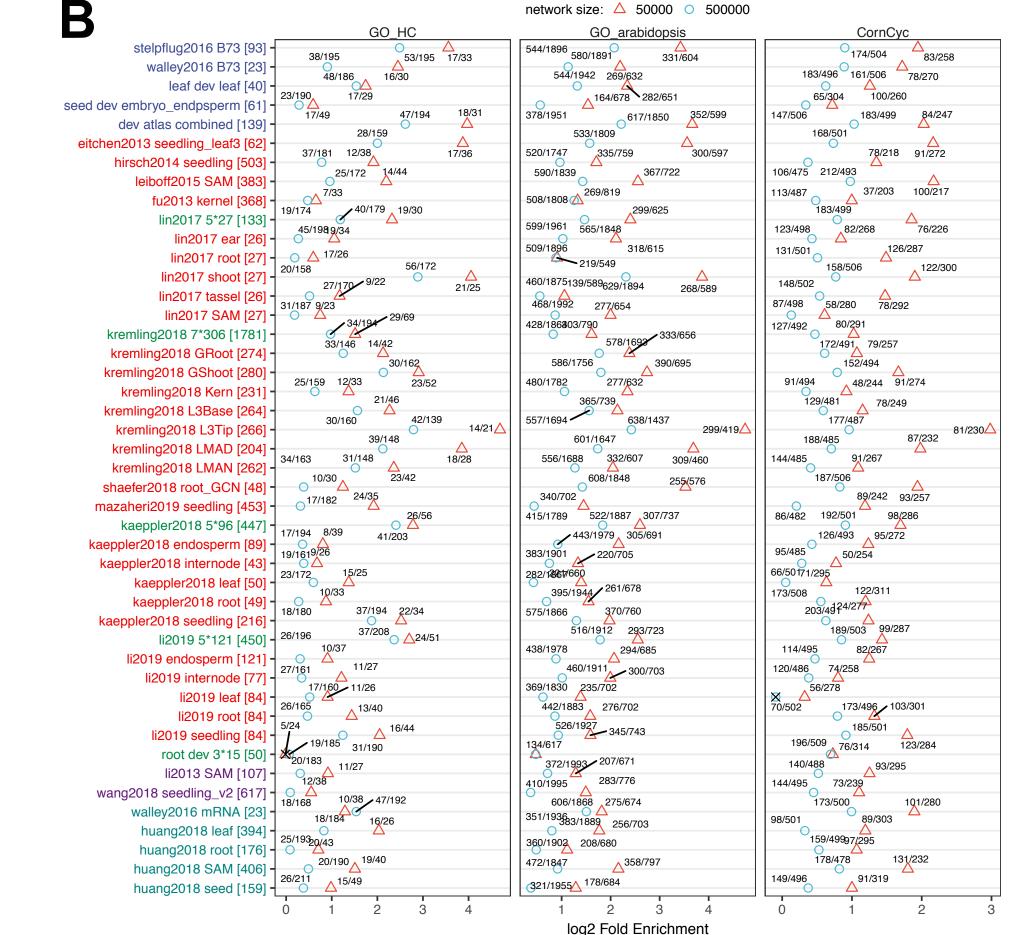


Figure B. Enrichment of co-annotated GO/CornCyc terms in co-regulated network targets. For each network either the top 50,000 edges (red triangle) or 500,000 edges (blue circle) was taken to assess enrichment of GO/CornCyc annotation. Log2 fold enrichment is calculated as the observed number of shared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

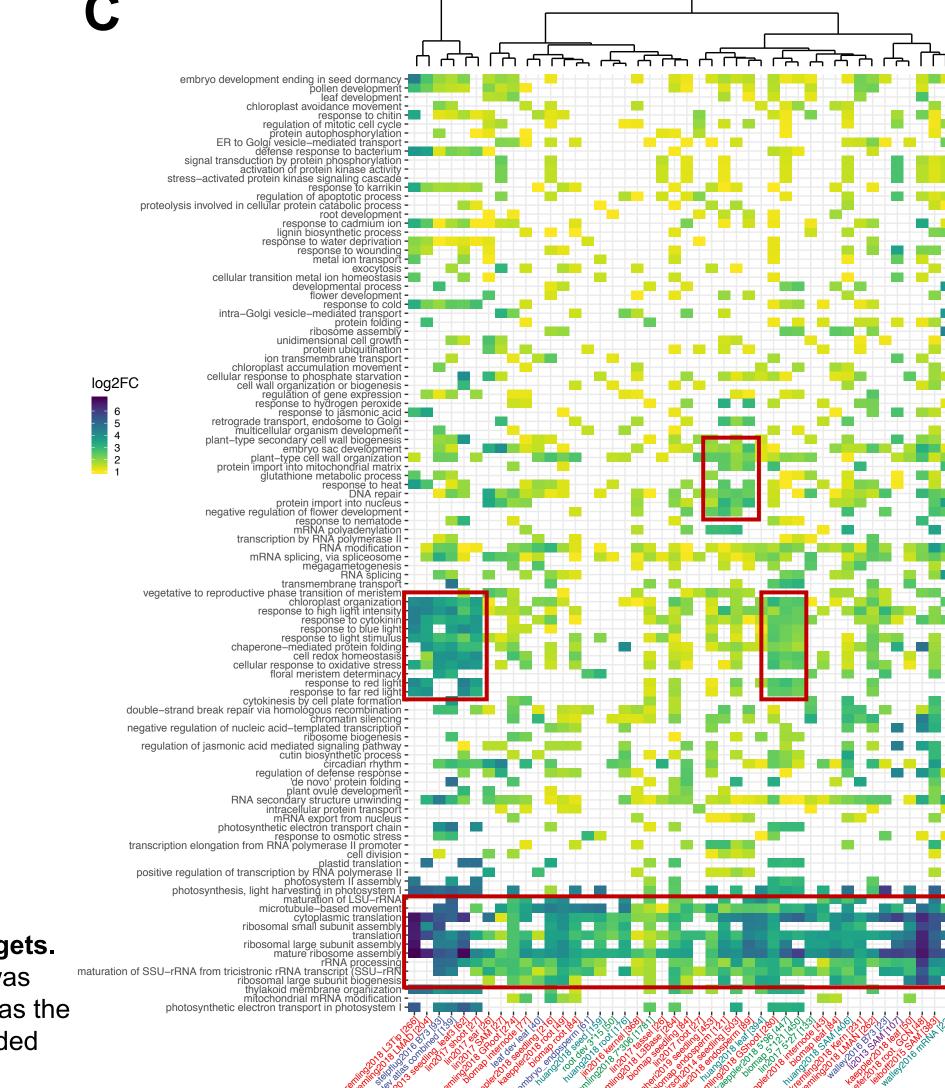


Figure C. Hierarchical clustering of 140 Gene Ontology (Uniprot.Plants) terms using log2 fold enrichment in different GRNs. Only GO terms enriched in at least 5 out of the 43 networks were used for clustering. Insignificant enrichment (p-value < 0.05) was treated as missing data and shown as white space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

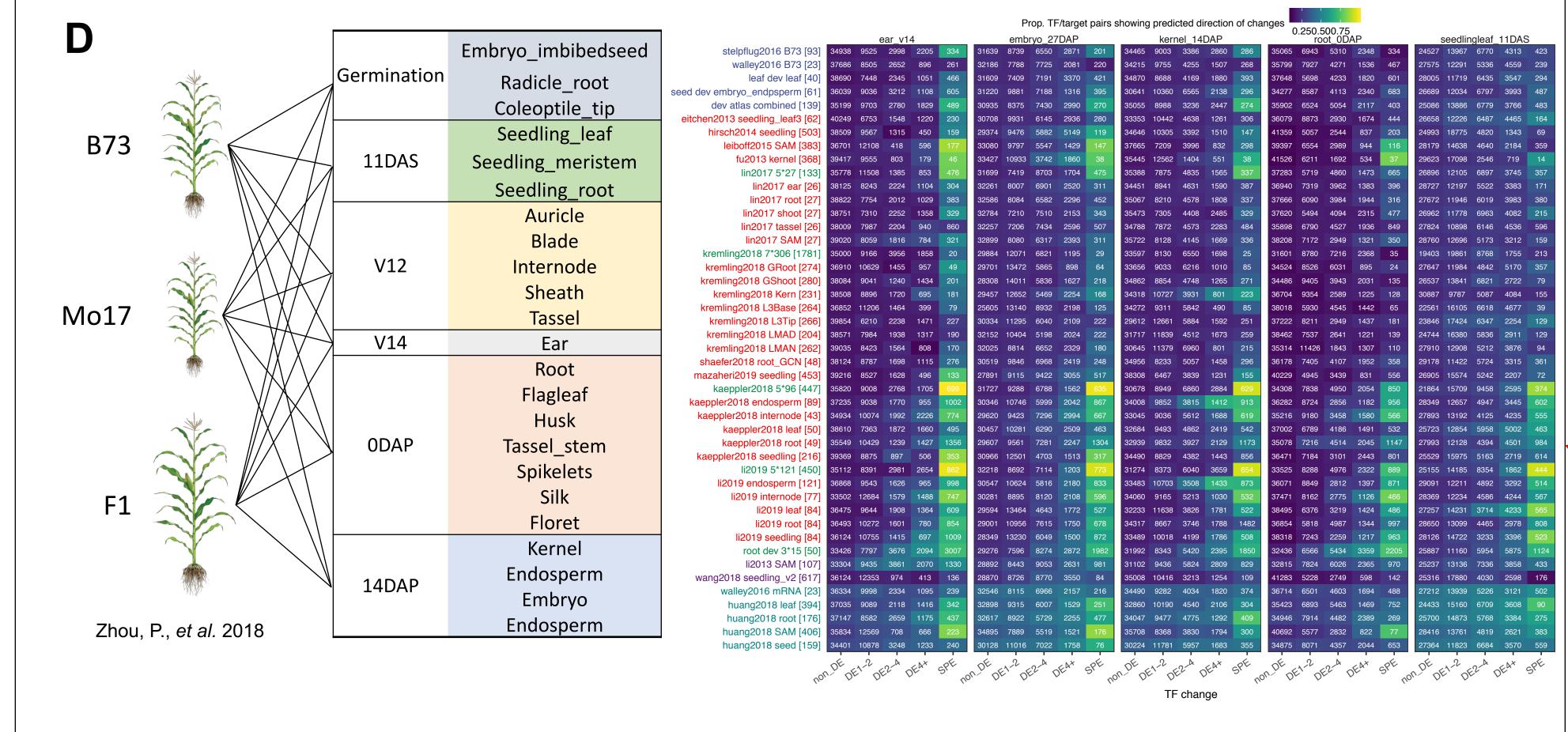


Figure D. (Left) Experimental design of the B73 & Mo17 paired developmental atlas study (3 genotypes X 23 tissues) and (right) proportion of differentially expressed targets regulated by TFs showing different DE levels. Each TF-target pair is classified according to the DE level of the TF ("non\_DE", "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.

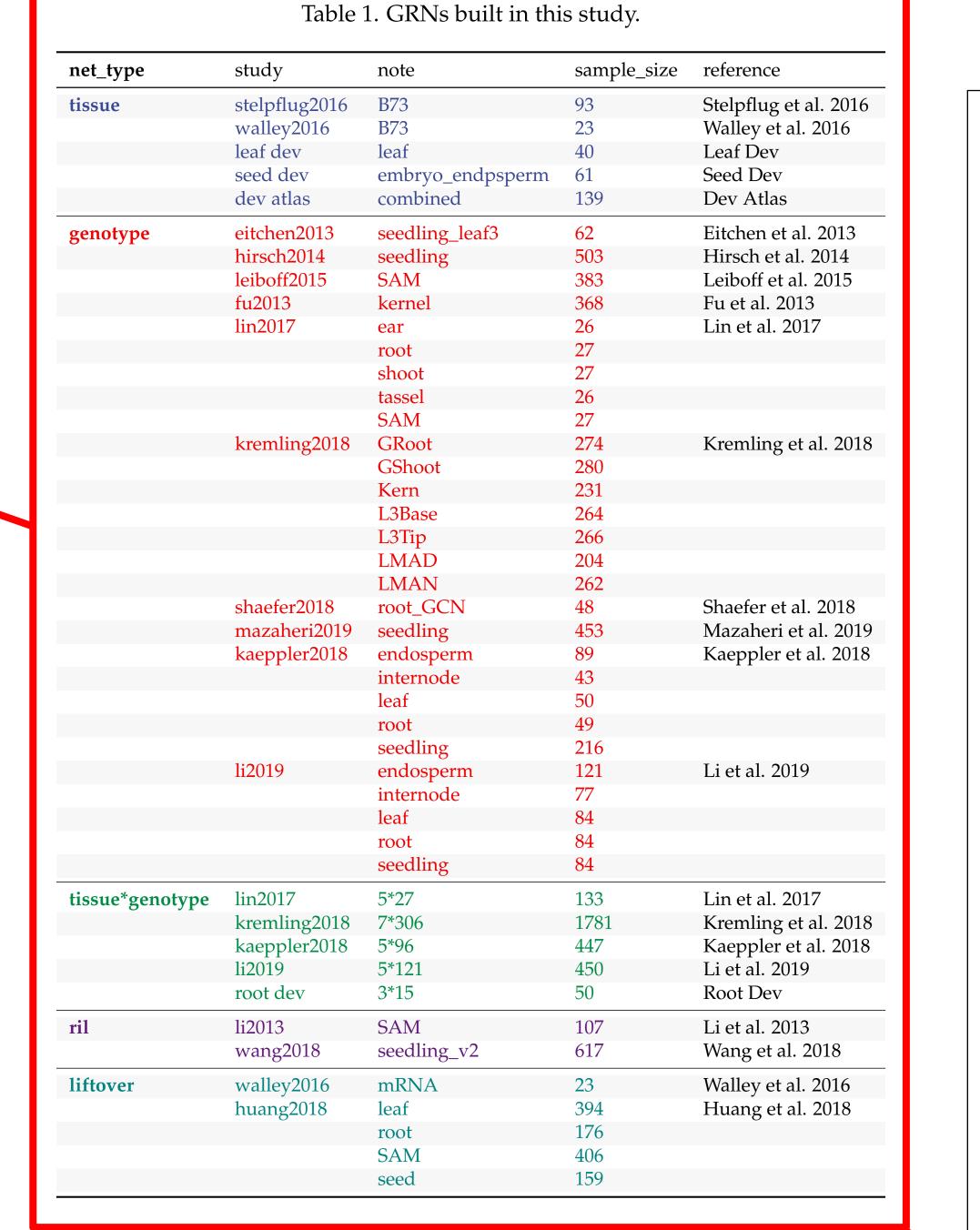


Figure F. Enrichment of co-regulated targets between known *trans*-eQTL hotspots (identified in previous studies) and GRN-predicted TF-target associations. For each network either the top 50,000 edges (red triangle) or 500,000 edges (blue circle) was taken to assess enrichment of co-regulation. Log2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

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.

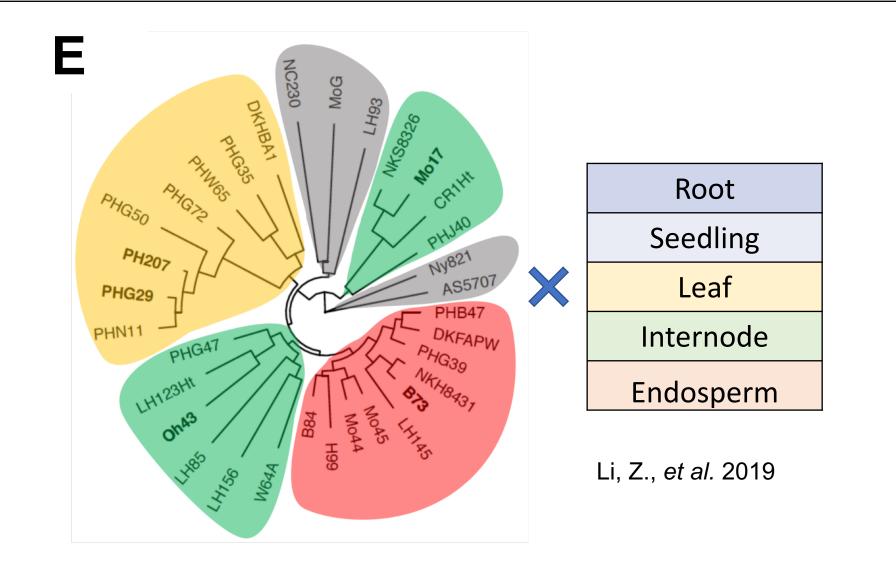


Figure E. . (Left) Experimental design of the BiomAP study (33 inbred lines X 5 tissues) and (right) Enrichment in significantly correlated TF-target pairs based on the BiomAP dataset. Only the top 50k TF-target pairs from each GRN was taken. Color in each cell represents the fold enrichment of significantly 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.

	Fold Enrich. in	Prop. Target Sig.	Corr. 1.5	2.0 2.5	3.0 3.5
stelpflug2016 B73 [93]	47400/0.25/2.57	46898/0.28/2.66	45187/0.33/2.33	43430/0.20/2.41	42976/0.19/2.24
walley2016 B73 [23]	49658/0.22/2.01	49200/0.23/2.07	48249/0.31/2.13	46857/0.13/1.54	46643/0.17/1.98
leaf dev leaf [40]	49974/0.28/2.35	49542/0.26/2.12	49011/0.31/1.99	47598/0.16/1.79	46579/0.16/1.81
seed dev embryo_endpsperm [61]	48321/0.18/1.76	48279/0.18/1.70	47305/0.23/1.64	46507/0.11/1.37	49112/0.22/2.28
dev atlas combined [139]	45667/0.27/2.98	44892/0.31/2.97	43110/0.37/2.71	41260/0.22/2.64	41308/0.23/2.61
eitchen2013 seedling_leaf3 [62]	48976/0.21/2.06	48730/0.20/1.96	47730/0.22/1.56	48009/0.25/2.53	45741/0.12/1.49
hirsch2014 seedling [503]	49052/0.62/3.01	48926/0.49/2.67	48012/0.44/2.36	47558/0.39/3.01	46830/0.35/2.62
leiboff2015 SAM [383]	48885/0.53/3.47	48585/0.54/3.55	47150/0.56/2.84	46132/0.32/3.12	44748/0.35/3.00
fu2013 kernel [368]	48555/0.50/2.21	48569/0.40/2.16	47582/0.43/1.84	46926/0.30/2.03	48326/0.47/2.42
lin2017 5*27 [133]	47067/0.25/2.61	46473/0.31/2.71	44355/0.36/2.36	42013/0.15/1.93	41517/0.16/1.93
lin2017 ear [26]	49372/0.23/2.14	48770/0.19/1.86	48904/0.25/1.89	47377/0.14/1.70	47758/0.15/1.78
lin2017 root [27]	49357/0.16/1.75	49302/0.22/2.20	47444/0.18/1.48	46976/0.12/1.45	44840/0.12/1.50
lin2017 shoot [27]	49536/0.17/1.75	49263/0.19/1.88	47969/0.22/1.75	47437/0.13/1.56	45433/0.11/1.42
lin2017 tassel [26]	47756/0.13/1.50	47403/0.13/1.39	46656/0.16/1.36	44985/0.10/1.29	44151/0.10/1.30
lin2017 SAM [27]	49564/0.21/1.94	49277/0.19/1.76	49135/0.22/1.60	48080/0.13/1.54	48188/0.15/1.77
kremling2018 7*306 [1781]	49148/0.23/1.64	49045/0.26/1.60	48310/0.32/1.50	46173/0.20/1.74	46609/0.16/1.43
kremling2018 GRoot [274]	47314/0.25/1.94	47306/0.44/2.30	45915/0.42/2.09	44824/0.19/1.79	43386/0.18/1.59
kremling2018 GShoot [280]	47873/0.31/2.51	47748/0.29/2.29	46397/0.42/2.31	46111/0.23/2.21	44736/0.22/2.12
kremling2018 Kern [231]	47464/0.20/1.85	46730/0.19/1.72	46064/0.23/1.65	45793/0.15/1.71	45327/0.34/2.96
kremling2018 L3Base [264]	48186/0.27/2.09	47771/0.29/2.00	47506/0.53/2.33	46083/0.19/1.91	45838/0.18/1.77
kremling2018 L3Tip [266]	42972/0.18/2.08	42159/0.14/1.67	40983/0.18/1.66	39888/0.21/2.54	39205/0.11/1.44
kremling2018 LMAD [204]	44703/0.16/1.79	44268/0.13/1.47	42479/0.16/1.43	41998/0.19/2.28	41354/0.11/1.35
kremling2018 LMAN [262]	45899/0.17/1.68	45714/0.14/1.41	44130/0.17/1.30	43500/0.17/1.93	42787/0.12/1.46
shaefer2018 root_GCN [48]	48911/0.24/2.34	48815/0.36/2.97	47421/0.26/1.93	45936/0.17/2.05	44899/0.16/1.90
mazaheri2019 seedling [453]	49344/0.57/3.14	49115/0.43/2.62	48428/0.34/2.16	48251/0.29/2.80	47040/0.27/2.26
root dev 3*15 [50]	49937/0.21/2.29	49973/0.28/2.69	48952/0.23/1.96	47891/0.15/1.76	47042/0.13/1.76
li2013 SAM [107]	49933/0.23/2.07	49462/0.21/1.81	49605/0.28/1.99	48899/0.15/1.69	48470/0.14/1.63
wang2018 seedling_v2 [617]	49958/0.54/2.82	49819/0.38/2.49	49789/0.45/2.37	49371/0.30/2.68	48770/0.28/2.28
walley2016 mRNA [23]	47943/0.22/1.95	47535/0.24/1.99	46527/0.29/1.97	45084/0.14/1.64	44834/0.18/1.79
huang2018 leaf [394]	48806/0.25/1.88	48669/0.22/1.60	48327/0.26/1.45	48036/0.19/1.92	46579/0.14/1.46
huang2018 root [176]	48816/0.24/2.06	48817/0.32/2.31	47369/0.27/1.64	47188/0.14/1.62	45294/0.15/1.59
huang2018 SAM [406]	49269/0.47/2.86	49129/0.49/2.96	48840/0.52/2.31	48027/0.25/2.45	47081/0.29/2.41
huang2018 seed [159]	47492/0.18/1.63	47255/0.18/1.52	46888/0.23/1.47	45084/0.12/1.29	48265/0.24/2.23
	Seedling	Root	Internode	Leaf	Endosperm

