

Systematic Integration of GWAS with co-expression networks to detect causal genes for elemental accumulation in Zea mays and Arabidopsis thaliana using Camoco

Robert Schaefer, Jean Michel-Michno, Joseph Jeffers, Owen Hoekenga, Brian Dilkes, Ivan Baxter, Chad Myers

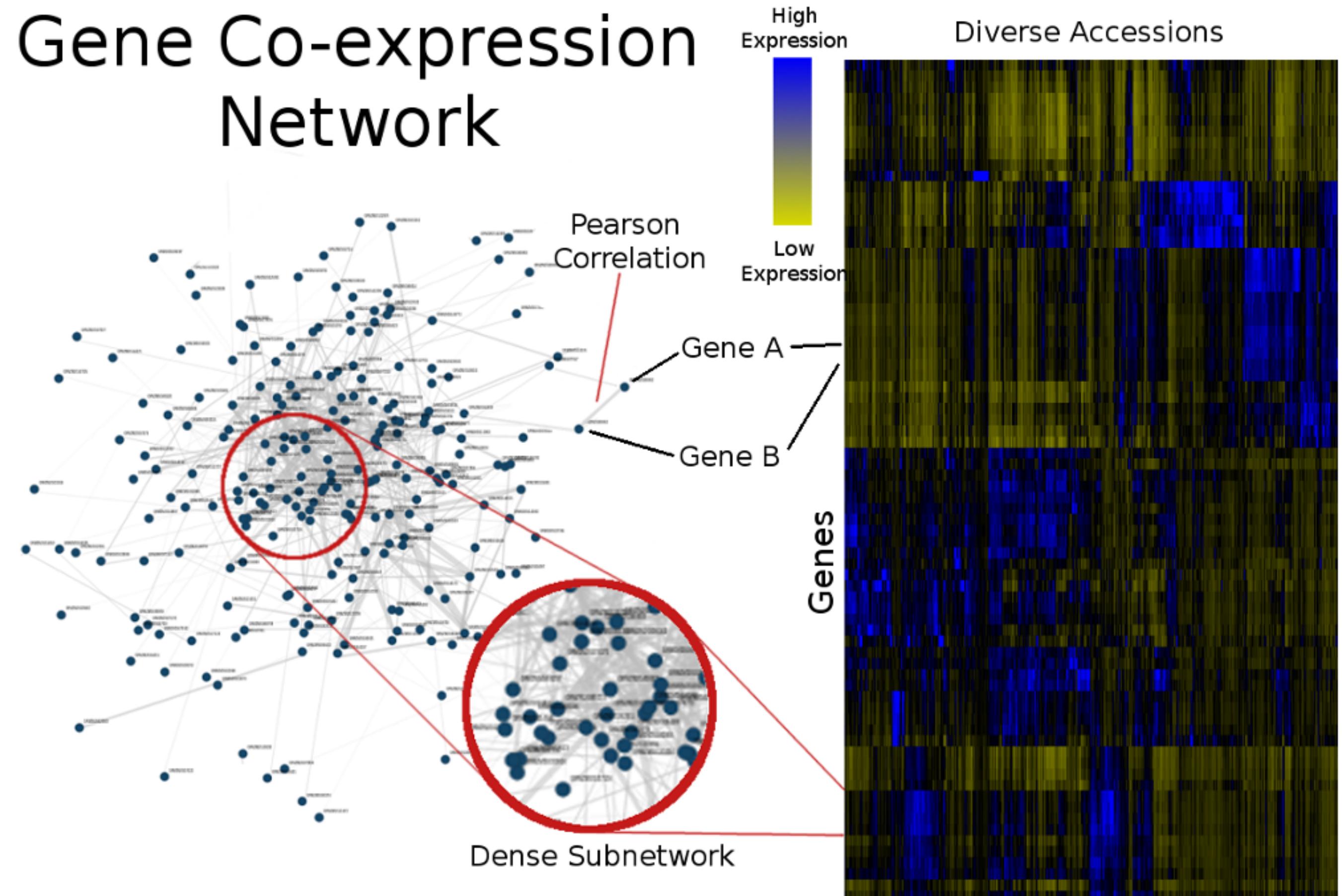
Biomedical Informatics and Computational Biology, University of Minnesota, Minneapolis, MN; Genetics Consultant, Ithaca, NY; Department of Biochemistry, Purdue University, West Layfayette, IN; Donald Danforth Plant Science Center, St Louis, MO; Department of Computer Science, University of Minnesota, Minneapolis, MN;

Genome wide association studies (GWAS) have become the de facto approach to identify agriculturally important genes. However, complex traits such as elemental accumulation are often associated with potentially hundreds of statistically significant SNPs, many of which are located outside known gene models. Identifying causal genes for complex traits quickly become unwieldy even using straightforward marker to gene mapping such as the closest flanking gene. Here, we present a computational framework called Camoco (Co-analysis of molecular components) that integrates loci identified by GWAS with co-expression networks to identify a focused set of candidate loci with functional coherence. This framework analyses the overlap between candidate loci generated from GWAS and the co-expression interactions that occur between them. Camoco implements a suite of functions that perform SNP-to-gene mapping, builds and functionally validates co-expression networks, and provides a robust bootstrapping model to evaluate the statistical significance of candidate loci.

Using Camoco, candidate genes were identified from GWAS measuring elemental accumulation in maize seeds as well as arabidopsis leaves, seeds and roots. Three maize and four arabidopsis co-expression networks were generated using diverse genotypic as well as tissue based gene expression data. On average, candidate gene lists identified by GWAS were reduced by two orders of magnitude by integrating co-expression network information, which produced a focused set of candidates with both strong associations with the phenotype as well as evidence for functional coherence in the co-expression network. Identifying candidates is an important step to resolve genes responsible for elemental accumulation and other agriculturally important traits.

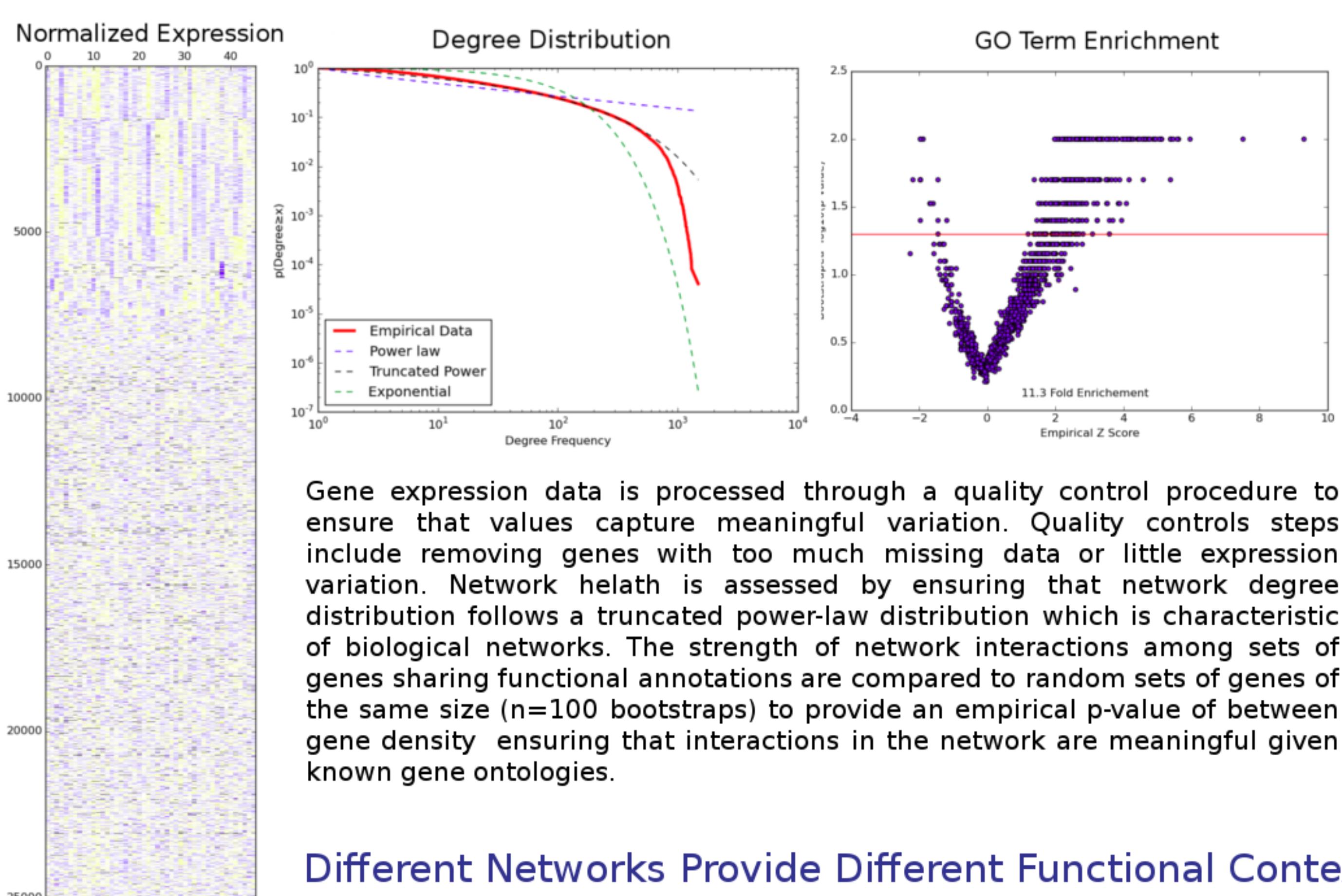
Building and Validating Gene Co-Expression Networks Using the Camoco Framework

Gene Co-expression Network



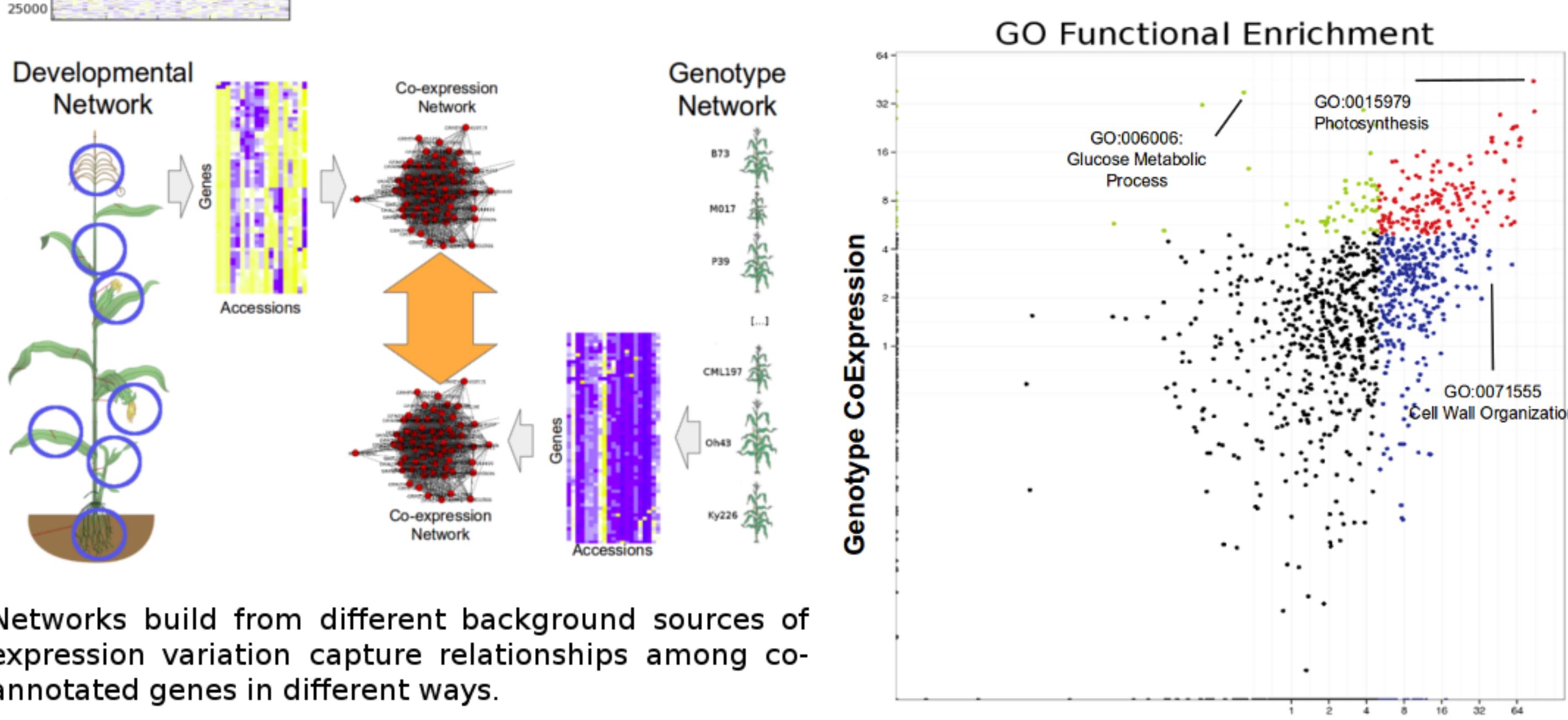
Camoco provides robust tools to build co-expression networks from diverse sources of gene expression such as RNA-Seq or from the large amount of publicly available microarray data.

Co-expression Network Quality Control: Assessing overall Network Health and Enrichment for Known Functional Terms



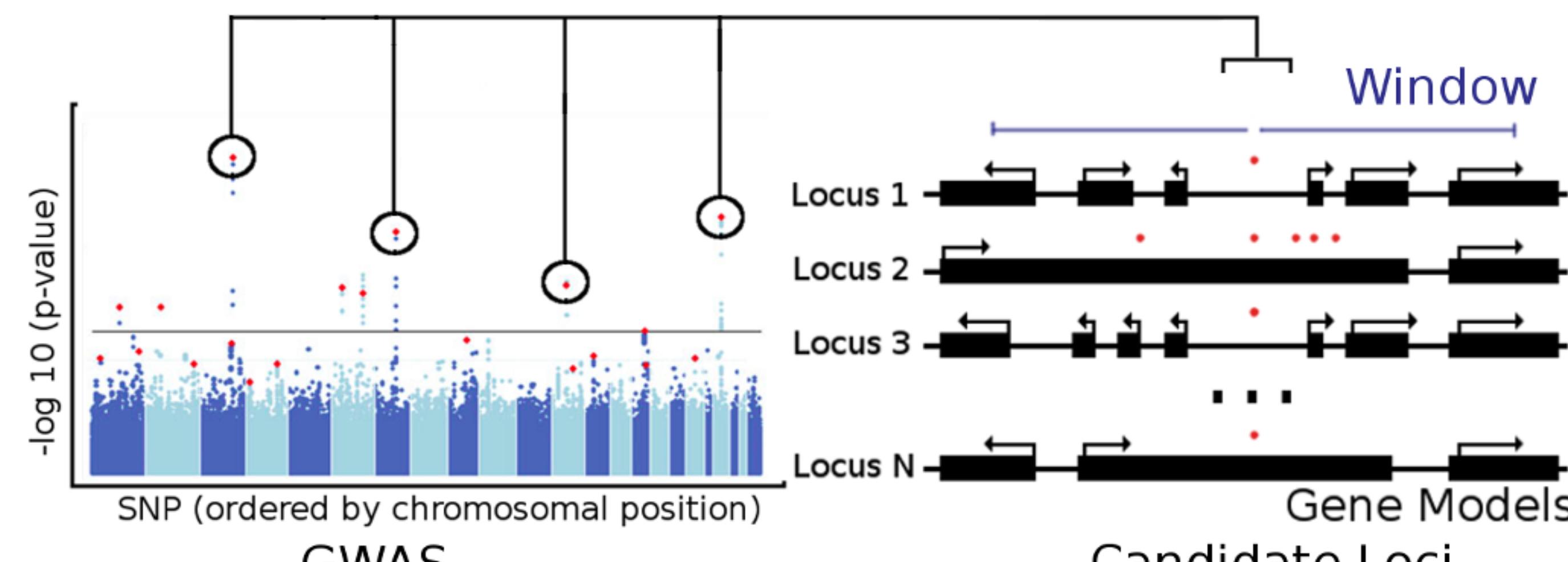
Gene expression data is processed through a quality control procedure to ensure that values capture meaningful variation. Quality controls steps include removing genes with too much missing data or little expression variation. Network health is assessed by ensuring that network degree distribution follows a truncated power-law distribution which is characteristic of biological networks. The strength of network interactions among sets of genes sharing functional annotations are compared to random sets of genes of the same size ($n=100$ bootstraps) to provide an empirical p-value of between gene density ensuring that interactions in the network are meaningful given known gene ontologies.

Different Networks Provide Different Functional Contexts

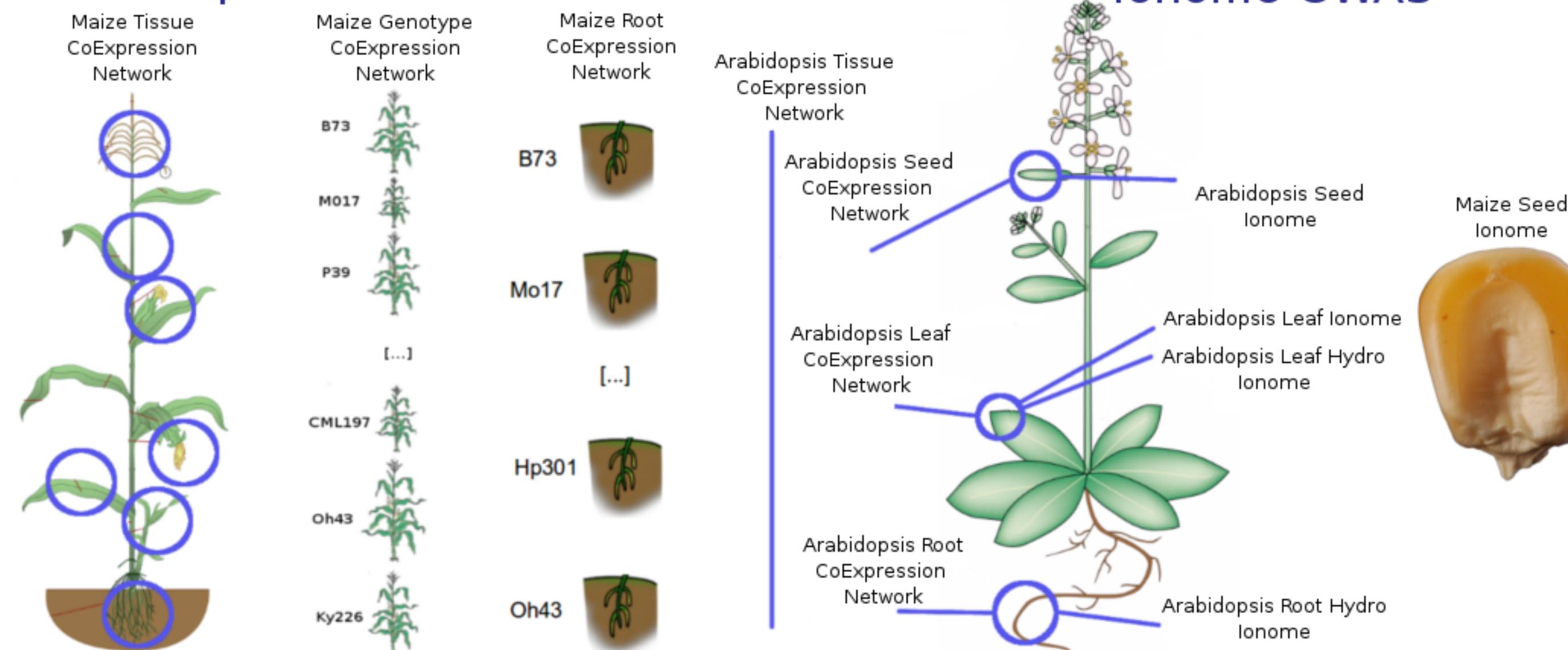


Networks build from different background sources of expression variation capture relationships among co-annotated genes in different ways.

Mapping SNPs to Candidate Genes



Maize and Arabidopsis Co-expression Networks



Assessing GWAS/Network Overlap

WindowSize	50Kb	50Kb	50Kb	100Kb	100Kb	100Kb	500Kb	500Kb	500Kb	5
FlankLimit	1	2	5	1	2	5	1	2	5	
A127	0.6	0.76	0.95	0.83	0.69	0.96	0.98	0.98	0.98	0.98
A575	0.08	0.49	0.72	0	0.33	0.36	0.06	0.19	0.74	
B11	0.9	0.64	0.79	0.79	0.67	0.63	0.61	0.38	0.75	
Ca43	0.9	0.88	0.99	0.91	0.79	0.92	0.96	0.99	0.93	
Cd111	0.57	0.75	0.77	0.54	0.49	0.64	0.39	0.69	0.96	
Cu65	0.66	0.97	1	0.6	0.95	0.99	0.73	0.63	0.88	
Fe57	0.58	0.46	0.64	0.23	0.21	0.48	0.04	0.02	0.12	
K39	0.73	0.89	0.75	0.31	0.51	0.24	0.15	0.54	0.93	
Krb	0.28	0.31	0.36	0.35	0.46	0.29	0.69	0.37	0.68	
Mg25	0.43	0.66	0.56	0.44	0.83	0.94	0.56	0.8	0.85	
Mn55	0.14	0.39	0.65	0.22	0.62	0.54	0.1	0.46	0.66	
Mo698	0.73	0.83	0.66	0.18	0.14	0.39	0.04	0.28	0.91	
Na23	0.47	0.34	0.4	0.72	0.65	0.95	0.66	0.61	0.94	
Ni60	0.92	0.99	0.95	0.87	0.85	0.92	0.59	0.71	0.52	
P31	0.89	0.86	0.9	0.86	0.9	0.95	0.85	0.97	0.97	
Rb85	0.31	0.6	0.38	0.32	0.65	0.55	0.88	0.75	0.88	
S34	0.55	0.64	0.68	0.5	0.55	0.69	0.32	0.65	0.45	
Se62	0.68	0.79	0.65	0.25	0.52	0.57	0.97	0.8	1	
Sr88	0.83	0.94	0.96	0.9	0.98	0.79	0.69	0.91	0.91	
Zn66	0.05	0.31	0.4	0.17	0.47	0.63	0.15	0.52	0.45	

Assessing GWAS/Network Overlap

WindowSize	50Kb	50Kb	50Kb	100Kb	100Kb	100Kb	500Kb	500Kb	500Kb	5
FlankLimit	1	2	5	1	2	5	1	2	5	
A127	0.6	0.77	0.76	0.29	0.77	0.75	0.33	0.51	0.77	
A575	0.5	0.81	0.9	0.33	0.46	0.63	0.19	0.4	0.91	
B11	0.25	0.61	0.65	0.1	0.12	0.15	0.4	0.39	0.33	
Ca43	0.94	0.77	0.71	0.94	0.95	0.88	0.64	0.83	0.96	
Cd111	0.38	0.36	0.31	0.35	0.41	0.2	0.16	0.67	0.55	
Cu65	0.81	0.93	0.86	0.94	0.95	0.98	0.97	0.91	0.99	
Fe57	0.77	0.9	0.95	0.66	0.82	0.78	0.62	0.89	0.99	
K39	0.64	0.85	0.67	0.49	0.46	0.18	0.59	0.4	0.42	
Krb	0.53	0.83	0.77	0.54	0.7	0.36	0.68	0.65	0.64	
Mg25	0.39	0.5	0.51	0.21	0.5	0.61	0.63	0.82	0.96	
Mn55	0.81	0.98	0.94	0.82	0.97	0.92	0.9	0.97	0.89	
Mo698	0.65	0.36	0.24	0.18	0.26	0.56	0.43	1	1	
Na23	0.06	0.12	0.03	0.26	0.13	0.48	0.52	0.3	0.46	
Ni60	0.73	0.67	0.69	0.53	0.7	0.75	0.62	0.9	1	
P31	0.07	0.38	0.22	0.94	0.97	0.95	0.97	0.98	0.98	
Rb85	0.76	0.69	0.62	0.47	0.34	0.18	0.34	0.51	0.41	
S34	0.9	0.79	0.81	0.7	0.4	0.21	0.68	0.27	0.55	
Se62	0.92	0.95	0.88	0.57	0.85	0.87	0.84	0.96	0.79	
Sr88	0.02	0.42	0.18	0.47	0.42	0.34	0.34	0.24	0.15	
Zn66	0.25	0.5	0.03	0.43	0.66	0.59	0.38	0.5	0.51	

Maize Tissue

WindowSize	50Kb	50Kb	50Kb	100Kb	100Kb	100Kb	500Kb	500Kb	500Kb	5
FlankLimit	1	2	5	1	2	5	1	2	5	
A127	0.									