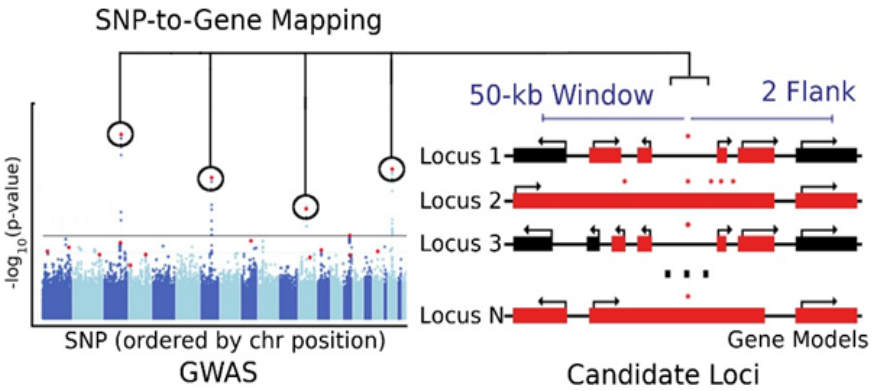
Plant Cell Advance Publication. Published on December 3, 2018, doi:10.1105/tpc.18.00908

IN BRIEF

Camoco: A Net for the Sea of Candidate Genes

Connecting phenotypic variance to causal genetic variants is hard. Forward and reverse genetic approaches allow us to build these connections only one at a time, and rely on chance and prior knowledge of appropriate targets, respectively. On the other hand, quantitative genetic approaches give us blurry snapshots of the whole structure of interacting genetic variants underlying a given phenotype, and bringing any one of those connections into sharp focus almost always takes significant effort. Quantitative trait loci identified by linkage mapping and genome-wide association studies (GWAS) often span large portions of the genome, with linkage disequilibrium (LD) limiting our ability to narrow them. Even when GWAS yield single nucleotide polymorphisms with apparently strong phenotypic associations, these are often segregating in LD with tens or more putative candidate genes or outside of annotated gene regions (see Figure; Wallace et al., 2014). Further, functional annotations for candidate genes are often incomplete, even for species of high economic importance like maize (Andorf et al., 2016). To move forward with functional analysis of every candidate gene that emerges from a quantitative genetics study is practically impossible—but how to prioritize?

Schaefer et al. (2018) introduce Camoco (Co-analysis of molecular components), a free, open-source computational framework that integrates gene co-expression networks with GWAS results (see Figure) to identify highly promising candidate genes. Co-expression networks quantify shared patterns of gene expression across biological contexts. As shared patterns of expression likely indicate shared regulation and/or function (Wolfe et al. 2005), Camoco ranks GWAS-derived candidate genes by their co-expression, prioritizing those with high similarity for further functional analysis. Schaefer et al. (2018) demonstrate the utility of this approach using GWAS of maize kernel elemental composition and three gene expression datasets, validate two candidate genes thus identified using mutants, and find that the gene expression dataset from genetically diverse individuals in biologically relevant tissue (roots) outperforms the two datasets from less diverse and less specific sources. While Schaefer and colleagues are not the first to bring together gene co-expression data and GWAS results, the implementation of this approach in Camoco is generalizable to any species with relevant gene expression data, even those with no other data on gene function.



The problem: typical GWAS identifies several SNPs (circled) highly associated with phenotypic variance linked to multiple candidate genes (red). On which should the researchers prioritize functional genetic analysis? *(Reprinted from Schaefer et al. [2018] Figure 1A.)*

Of course, this approach is not a cure-all for your quantitative genetics ails. It requires gene expression data across a broad sample of genotypes in the appropriate biological context(s). If these data do not already exist, they may be time- or cost-prohibitive to obtain. Relying on co-expression networks for functional information may also miss causal genes that are expressed at low levels or in short temporal windows or that are otherwise not well represented in the gene expression data, as well as causal genes that are not co-expressed and causal non-gene variants. This method may also identify false positive co-expression networks, driven by population structure rather that shared function or regulation. Still, when used with appropriate data and caution, Camoco could make bridging the gap between phenotype and genotype significantly easier.

Brook T. Moyers
Colorado State University
Fort Collins, CO
brook.moyers@gmail.com
ORCID: 0000-0003-0340-9488

REFERENCES

Andorf, C. M. et al. (2016). MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res. **44:** D1195–D1201.

Schaefer, R. J., Michno, J.-M., Jeffers, J., Hoekenga, O., Dilkers, B., Baxter, I., & Myers, C. L. (2018). Integrating co-expression networks with GWAS to prioritize causal genes in maize. Plant Cell. https://doi.org/10.1105/tpc.18.00299.

Wallace, J. G., Bradbury, P. J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E. S. (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. **10:** e1004845.

Wolfe, C. J., Kohane, I. S., & Butte, A. J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics **6:** 227.