Putting Genetic Interactions in Context through Global Modular Decomposition

Jeremy Bellay[1], Gowtham Atluri[1], Tina L. Sing[2], Kiana Toufighi[5], Michael Costanzo[3,4], Philippe Souza Moraes Ribeiro[1], Gaurav Pandey[6], Joshua Baller[1], Benjamin VanderSluis[1], Magali Michaut[3,4], Sangjo Han[3], Philip Kim[3], Grant W. Brown[2], Brenda J. Andrews[3,4], Charles Boone[3,4], Vipin Kumar[1], Chad L. Myers[1*]

1. Department of Computer Science & Engineering, University of Minnesota-Twin Cities, 200 Union St., Minneapolis MN, U.S.A. 55455

2. Department of Biochemistry, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Toronto ON, Canada M5S 3E1

3. Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Toronto ON, Canada M5S 3E1

4. Department of Molecular Genetics, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Toronto ON, Canada M5S 3E1

5. Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003 Barcelona, Spain

6. Plant & Molecular Biology Department, 461 Koshland Hall, University of California, Berkeley, CA 94720-3102

***** To whom correspondence should be addressed.

Supplemental Materials

Figures

Figure S1: Module Hypothesis
Here we present a small portion of the SGA interaction matrix that has been hierarchically clustered on both the query and array sides. Large negative and positive blocks immediately become apparent. We call out two blocks as specific examples of of the postive-within and negative-between pathway models. The set of genes *RAD51, RAD52, RAD54, RAD55, RAD57* are involved in DNA repair through recombination, while *REV1, REV3* and *REV7* are also involved in DNA repair but through translesion synthesis. We see that the RAD genes form a positive clique among themselves (within pathway) but form a negative biclique with the REV genes (between pathway).

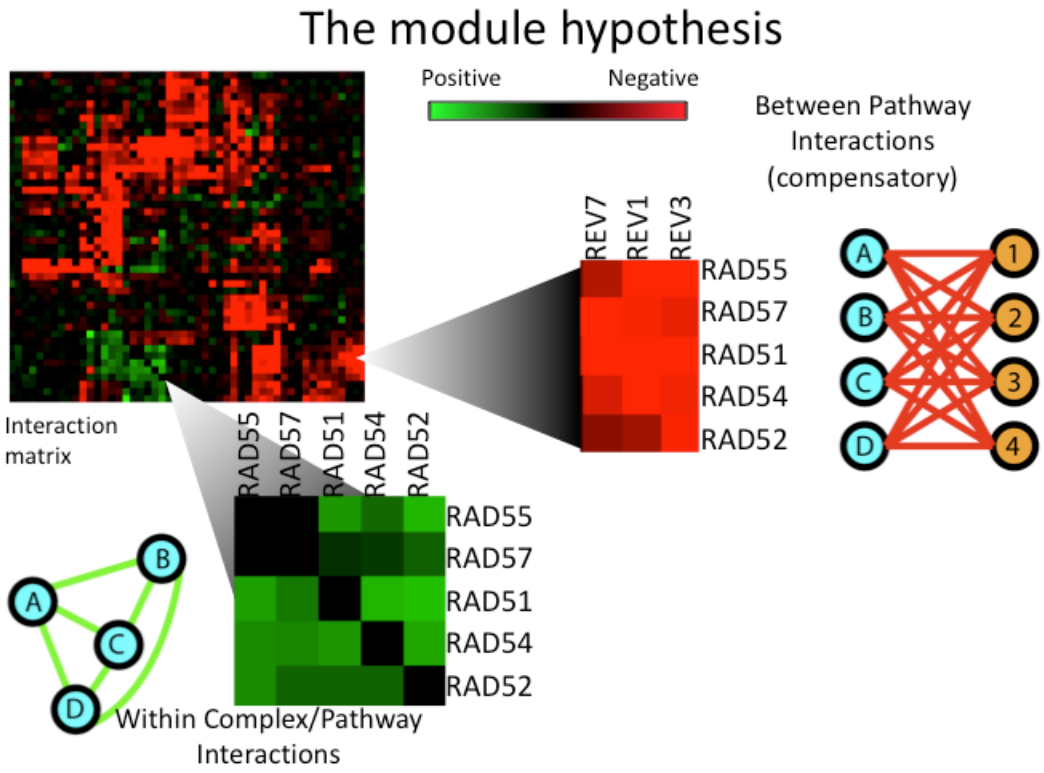


Figure S2: Network Randomization
The network was randomized by randomly selecting two interacting gene pairs (for example, Q1 and A1, and Q2 and A2) and switching the array genes as in the following figure. Thus, the connectivity of the network was randomized, but the degree (number of interactions) of each gene remained the same. If the random interaction pairs had overlapping query or array genes a new set of interaction pairs was selected. Maintaining the correct degree distribution is essential in cases where the distribution is extremely skewed, as is the case in most biological networks.

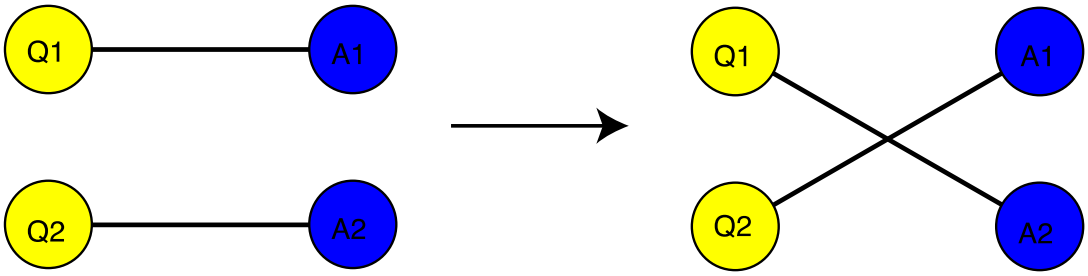


Figure S3: XMOD Flow Chart
The following diagram describes the process of discovering biclusters with XMOD. First, apriori is run to discover blocks of sizes bigger or equal to 6 query genes X 3 array genes or 3 query X 7 array for negative interactions and bigger than 3 query X 3 array for positive interactions. These groups are then compared with a distribution of blocks found on randomized graphs. Any block whose score (measure of how likely the block might arise from just the degree distribution) is a above a .0001 p-value with respect to the randomized distribution of scores is removed. To produce condensed blocks, smaller blocks that have 40% overlap or more by area with larger blocks are discarded.

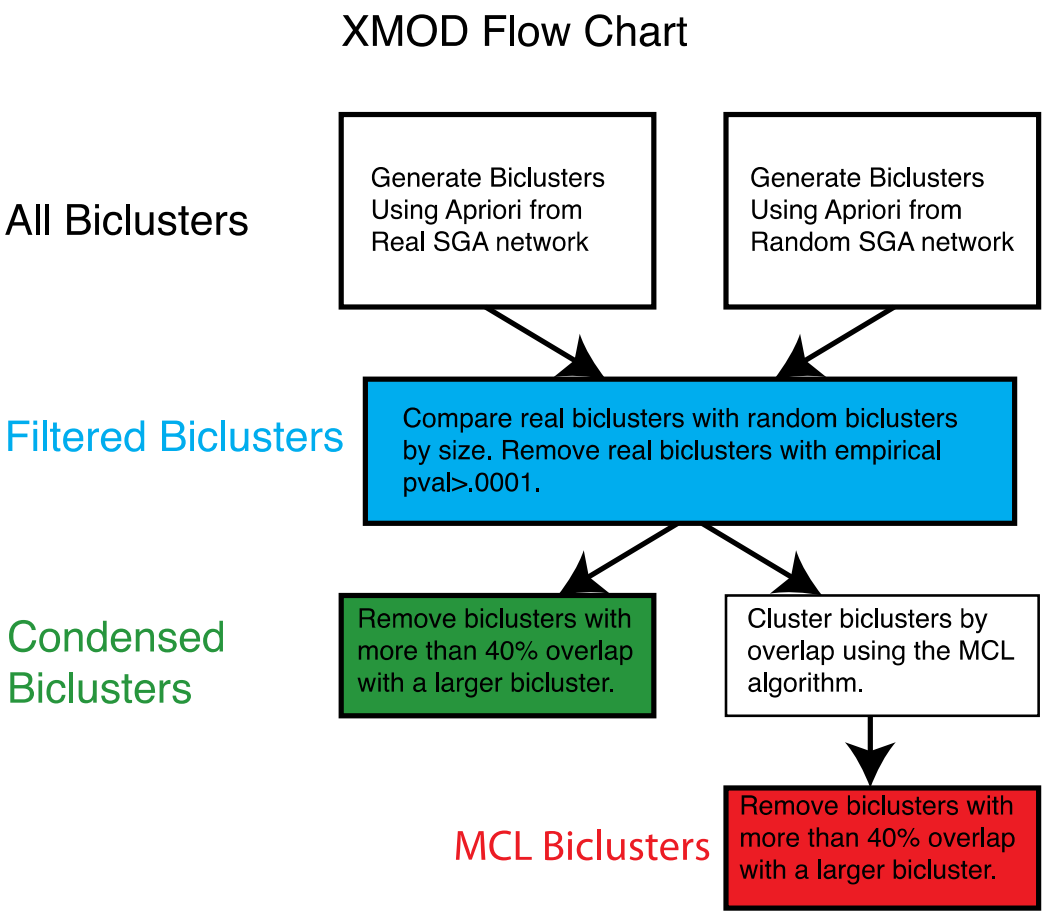


Figure S4: Filtering Positive Biclusters
The number of biclusters discovered for the real and randomized interaction network at various p-value cutoffs. The randomization and p-value assessments were performed separately on the positive and negative interaction networks.

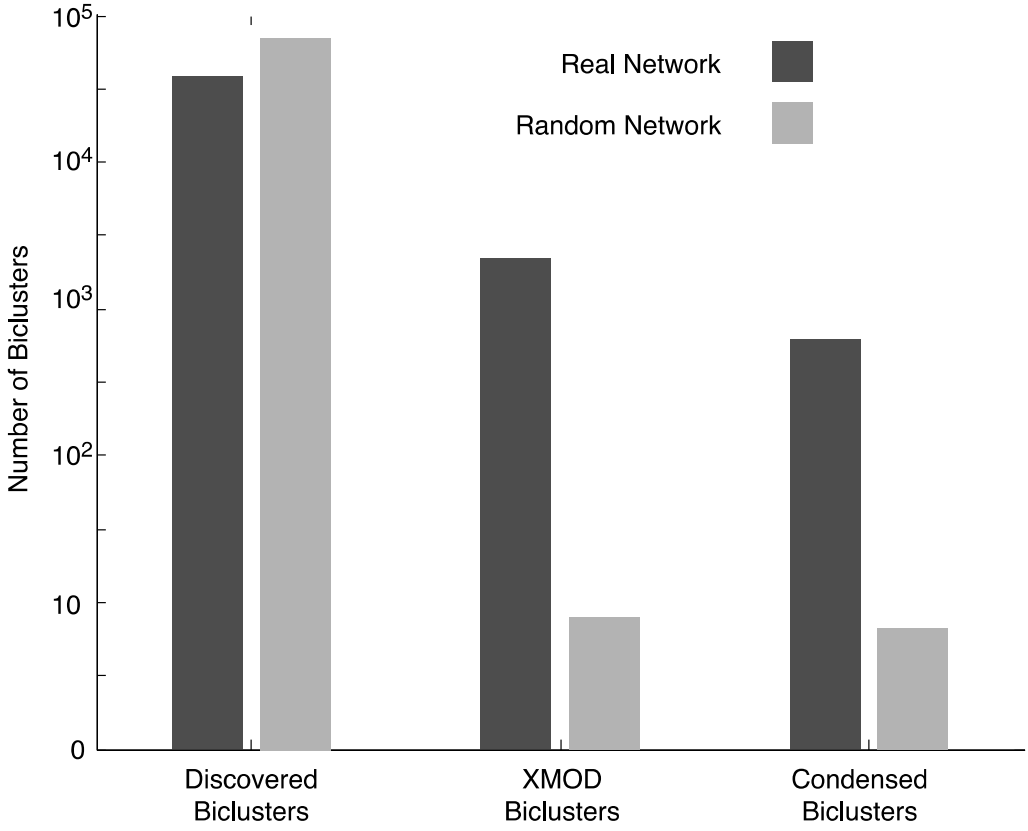


Figure S5: Size distribution of biclusters

Here we plot the size distribution of positive and negative biclusters. Each circle is proportional in size to the number of biclusters with the given number of query and array genes.

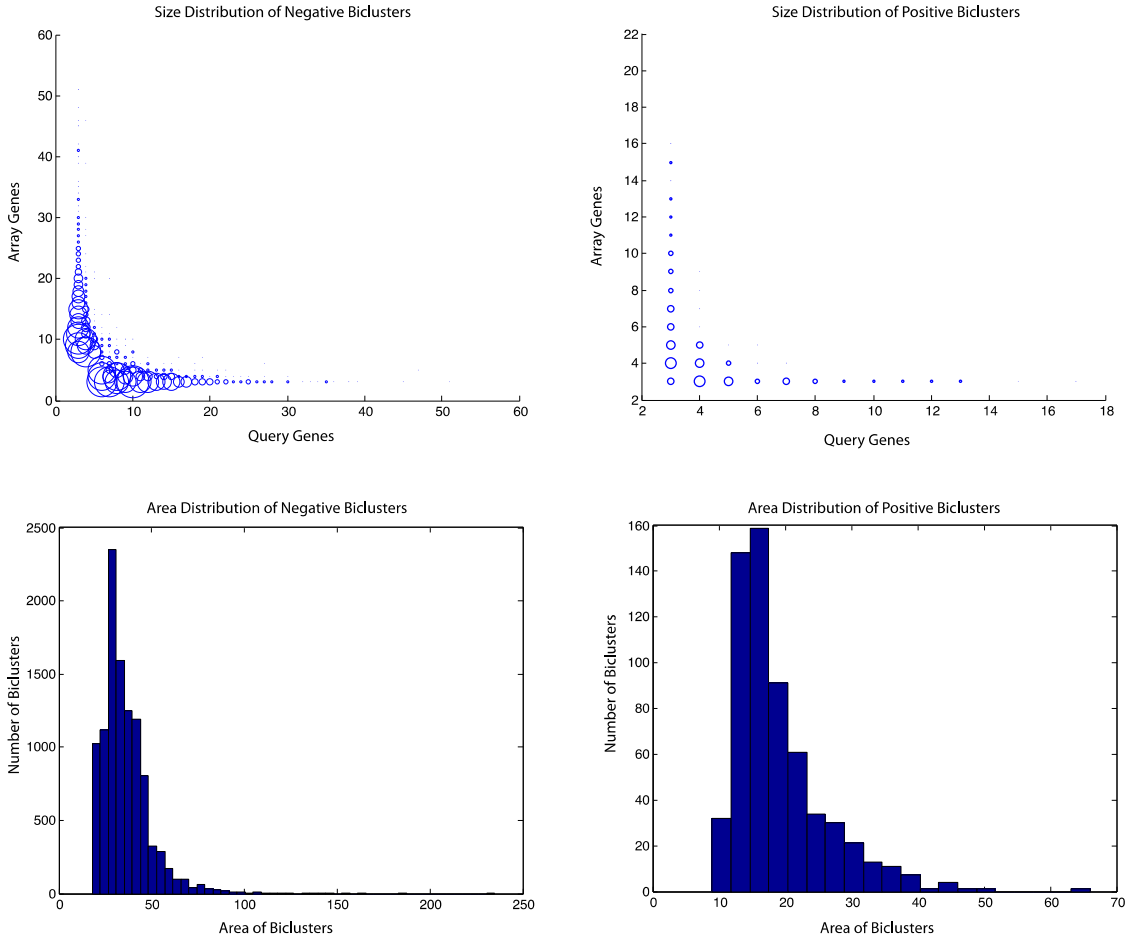


Figure S6: Functional coherence as measured by MEFIT while controlling for overlap on all methods.
We compared XMOD to other biclustering methods where we allowed only a 20% overlap by area among biclusters in every method. XMOD again performs well when compared to these other methods under this condition.

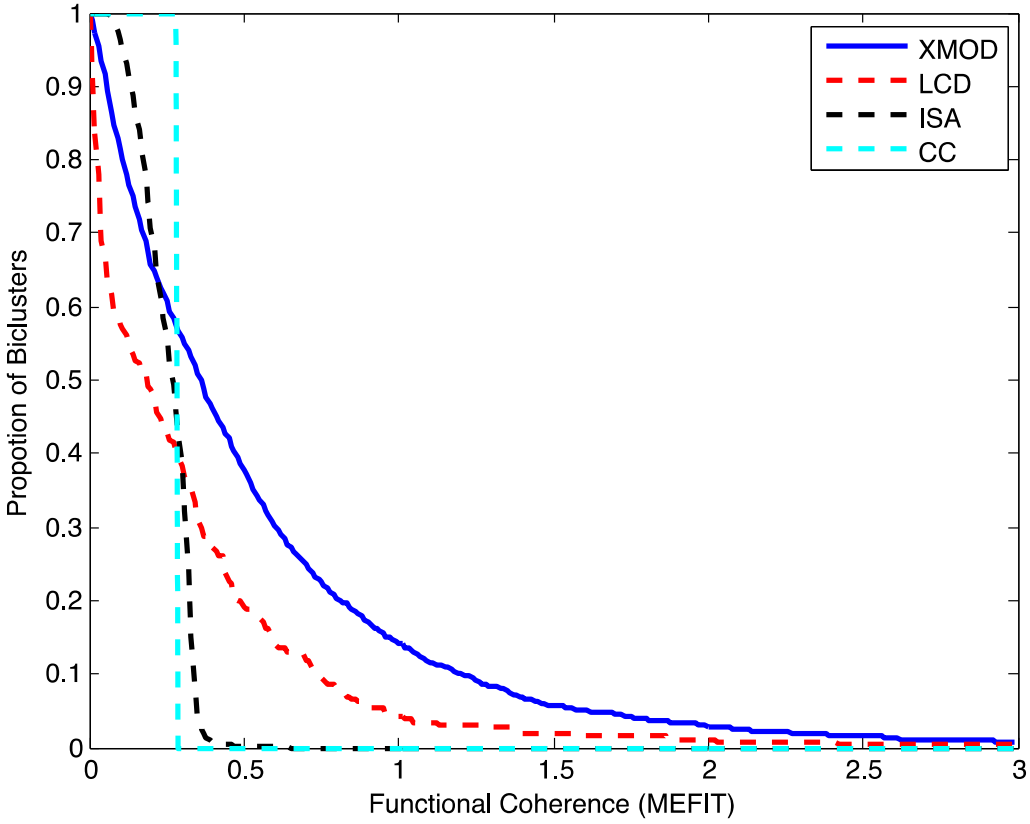


Figure S7: Functional coherence as measured by semantic similarity
We also compared XMOD to other biclustering techniques using semantic similarity on GO process terms as described in (Lin 1998). Again, XMOD out performs other methods. We opted to use the MEFIT functional network for our primary evaluation due to the bias of GO categories towards genetic interaction hubs.

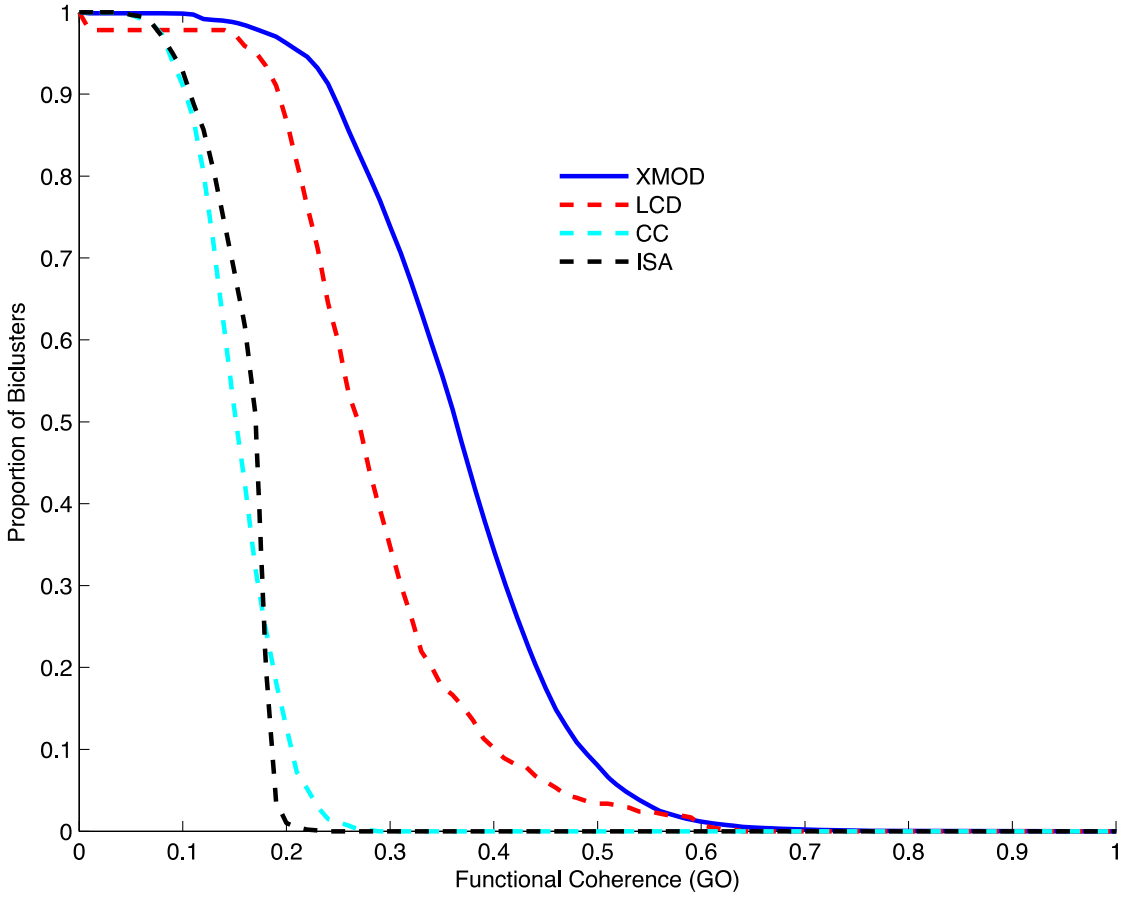


Figure S8: Robustness of XMOD to missing values
To test the robustness of XMOD to missing values, we randomly changed measured values to missing values for 1, 2, 5, and 10 percent of the measured values. We plot the proportion of interactions that were covered by XMOD filtered biclusters after introducing random missing values as a proportion of the interactions covered for the real, intact network.

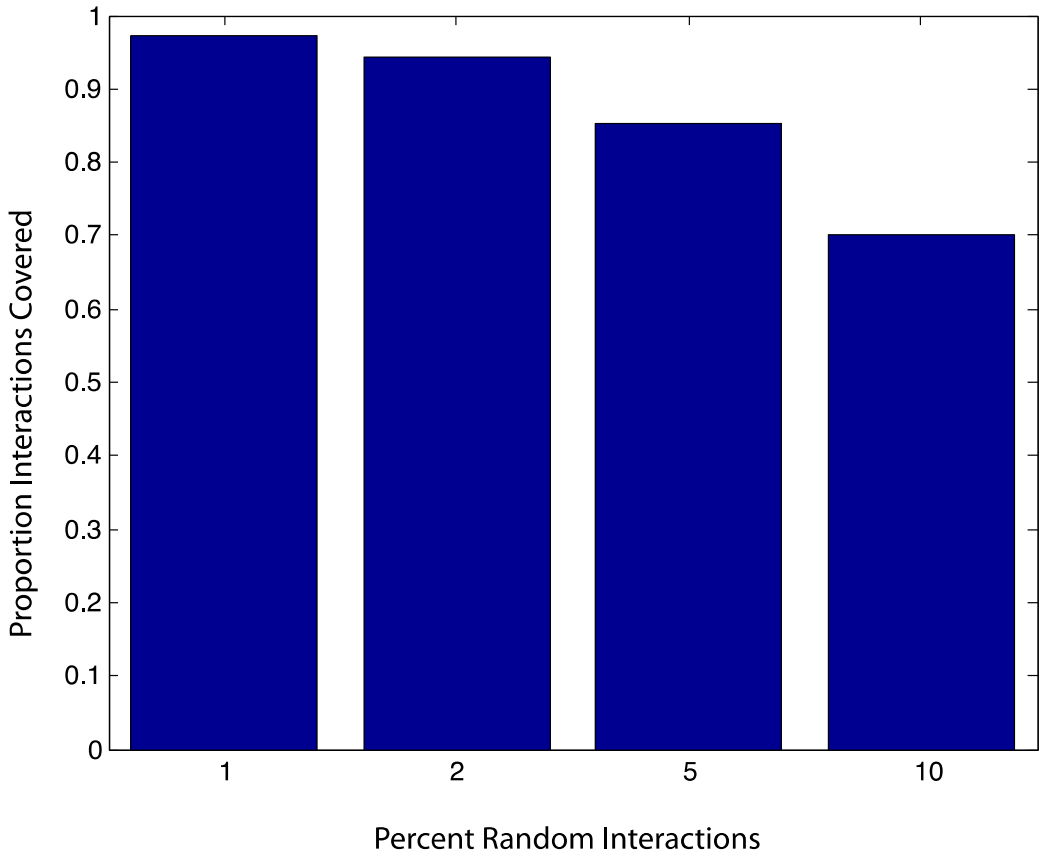


Figure S9: Validation of module derived functions
We investigated the functions associated with module membership by using the bioPIXIE functional network (Myers et al. 2005). In short, if gene A was associated with process X, we considered the mean bioPIXIE score between gene A and the genes associated with X (not including X or other genes in the given module). We then gave the association between A and X a score by considering the proportion of screened genes that have a lower association with term X than does gene A. For example, if gene A is better associated with term X than any other gene, the association of A and X would have a score of 1.

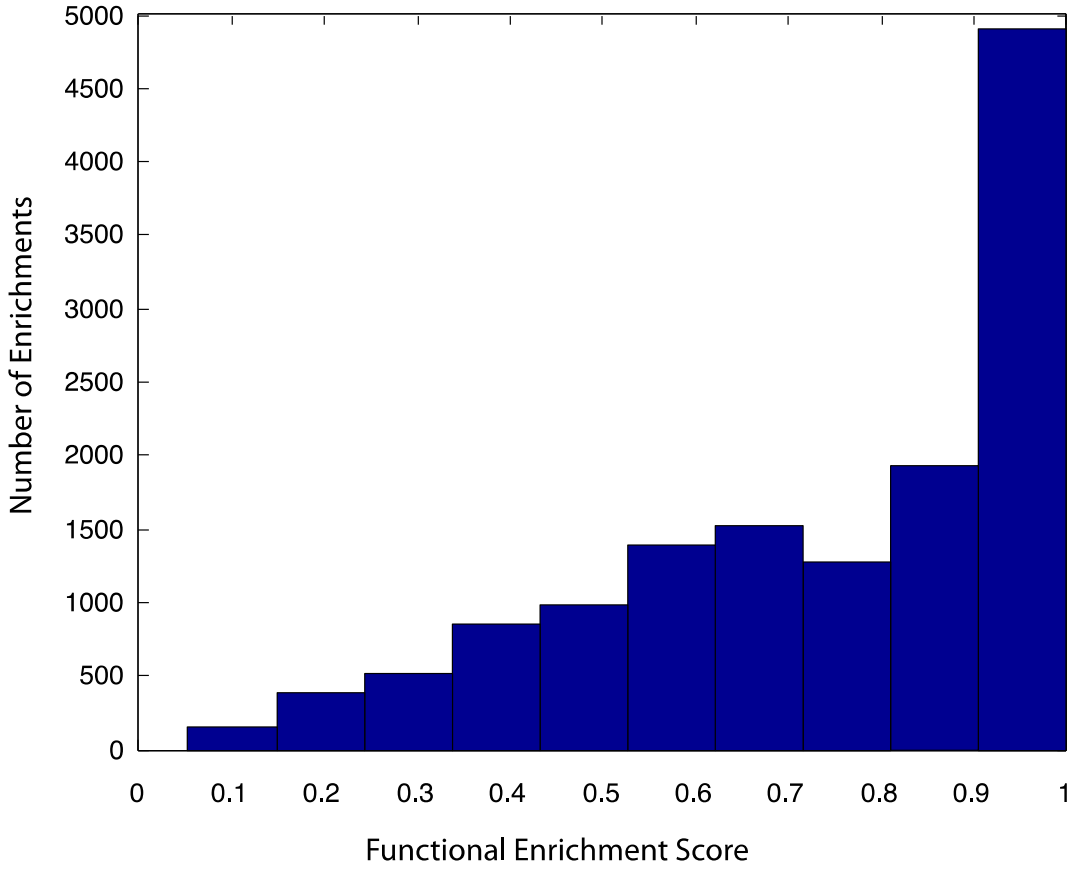


Figure S10: Proportion of q-cliques and bicliques on deletion biclusters
We found proportion of q-cliques and bicliques on biclusters that only depend on deletion alleles (bigger that 3x3 without TS or DAMP alleles). 38% of positive biclusters are q-cliques, while 13% of negative biclusters are q-cliques.

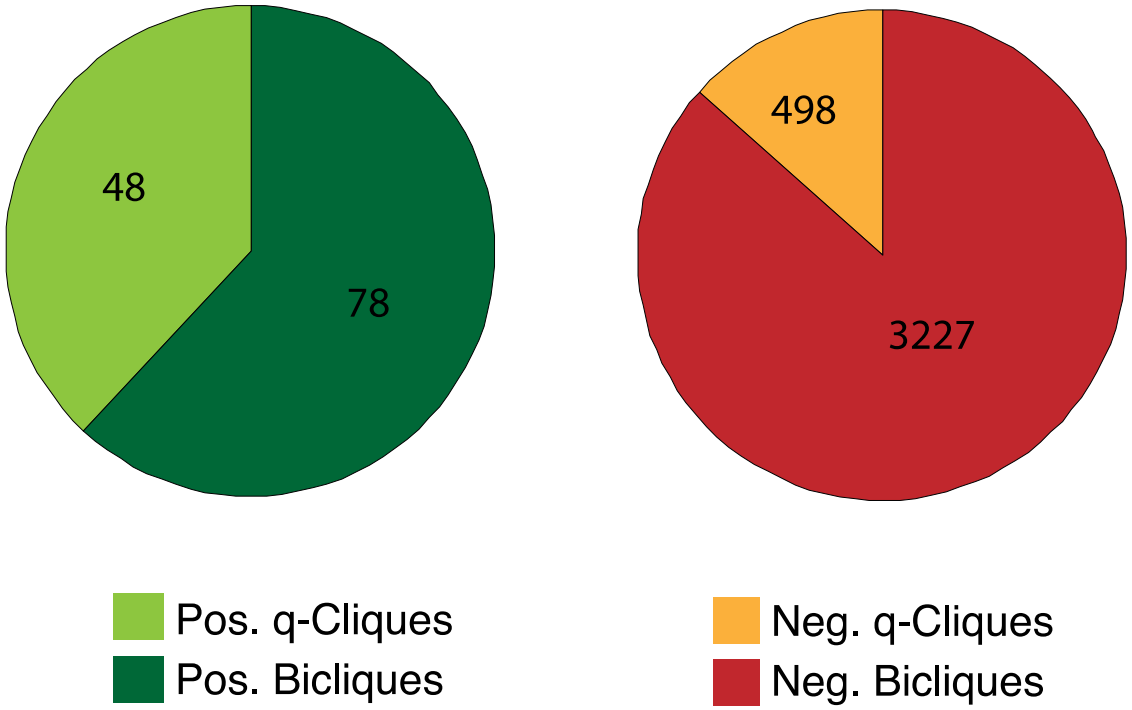


Figure S11: Coexpression enrichments on modules supported only by deletion alleles
We found the average enrichments for coexpression of modules who's support only depended on deletion mutants (bigger that 3x3 without TS or DAMP alleles). We find no striking differences in coexpression between q-cliques or bicliques.

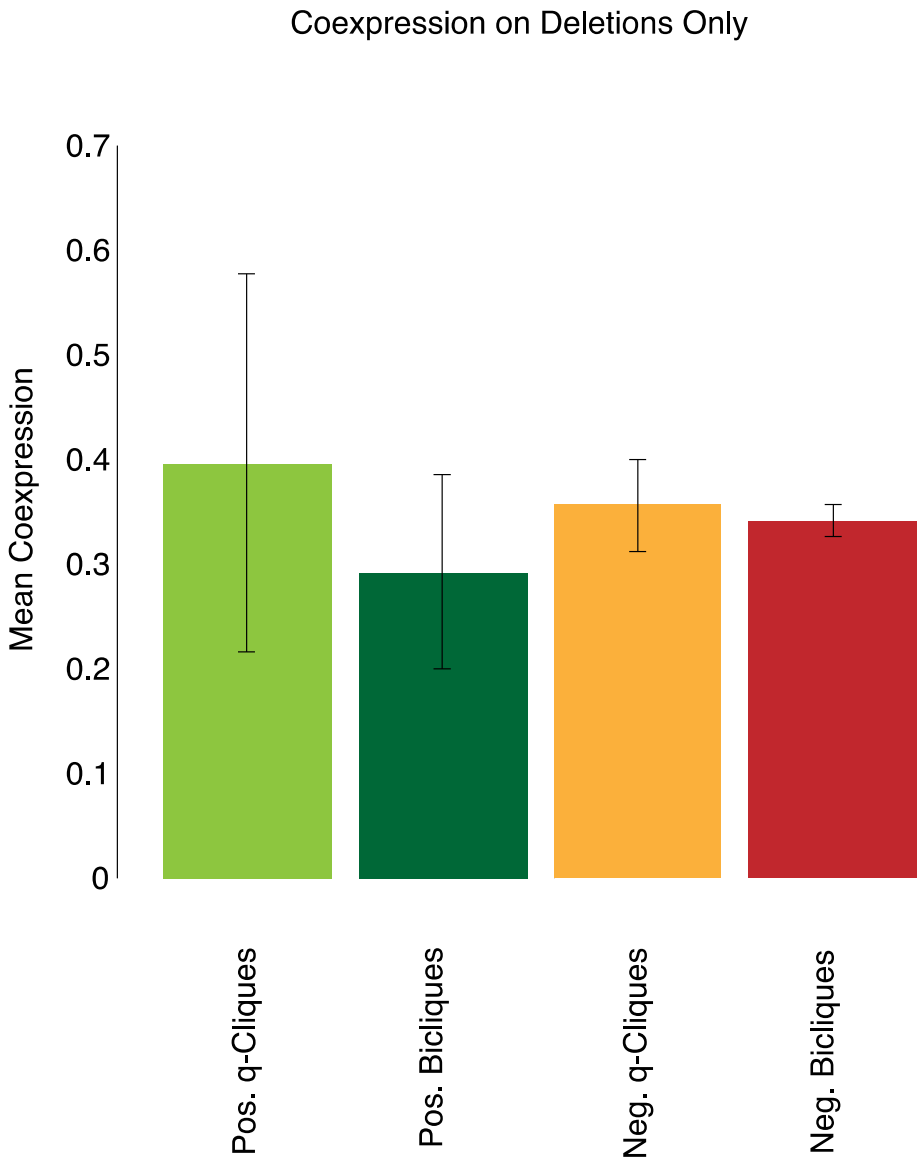
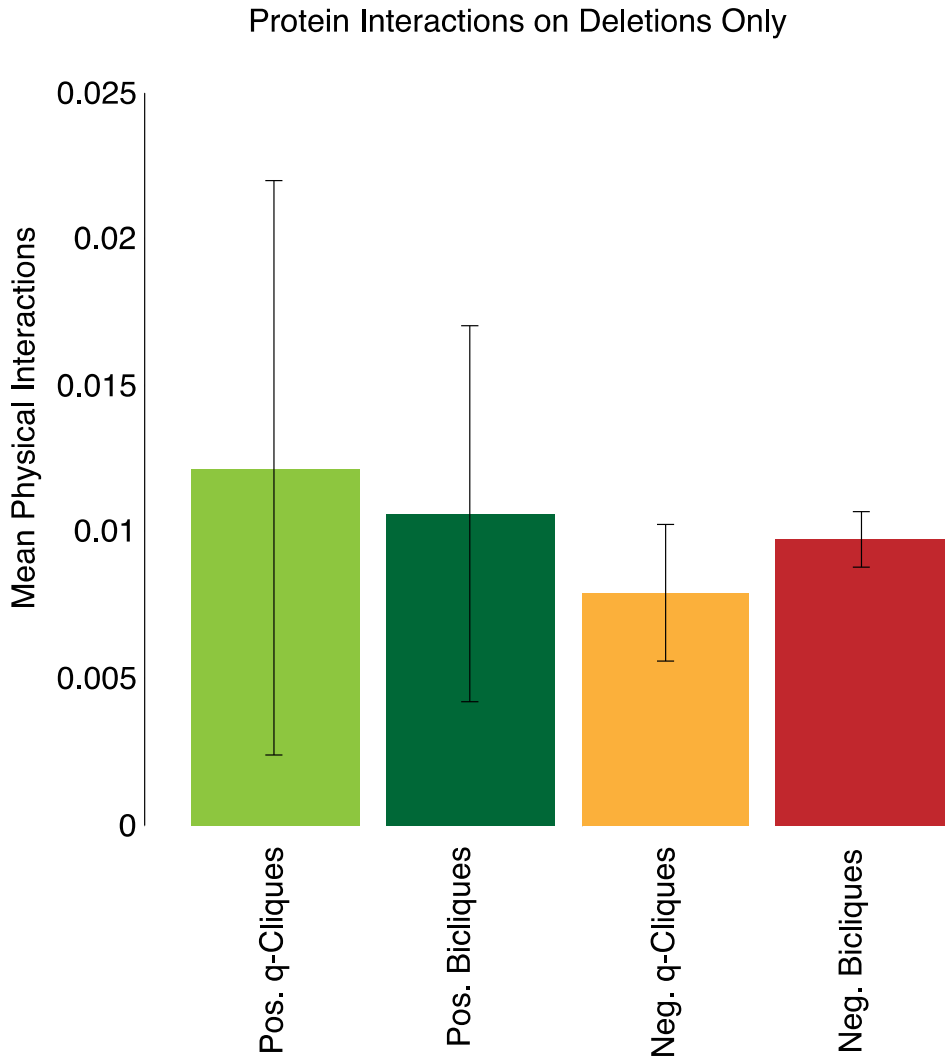


Figure S12: PPI enrichments on modules supported only by deletion alleles
We found the average enrichments for protein-protein interactions of modules who's support only depended on deletion alleles (bigger that 3x3 without TS or DAMP alleles). Here we find negative bicliques once again have a significantly higher mean PPI level than negative q-cliques ($p < 0.002$, Ranksum).



Supplemental Tables

Table S1: Enrichment of genes that appear on both sides of negative biclusters. See the attached file “neg_clique_enrich.txt”.

Supplemental References

Lin D. 1998. An Information-Theoretic Definition of Similarity. *IN PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING* 296--304. (Accessed February 18, 2011).

Myers CL, Robson D, Wible A, Hibbs MA, Chiriac C, Theesfeld CL, Dolinski K, and Troyanskaya OG. 2005. Discovery of biological networks from diverse functional genomic data. *Genome Biol* **6**: R114-R114.