

Co-expression analysis of metabolic pathways in Arabidopsis

Ann Loraine(1,2), Xiangqin Cui (1), Hairong Wei(1), Grier Page (1), Staffan Persson(3)

(1) Department of Biostatistics & (2) Genetics, University of Alabama at Birmingham

(3) Plant Biology Department, Carnegie Institute

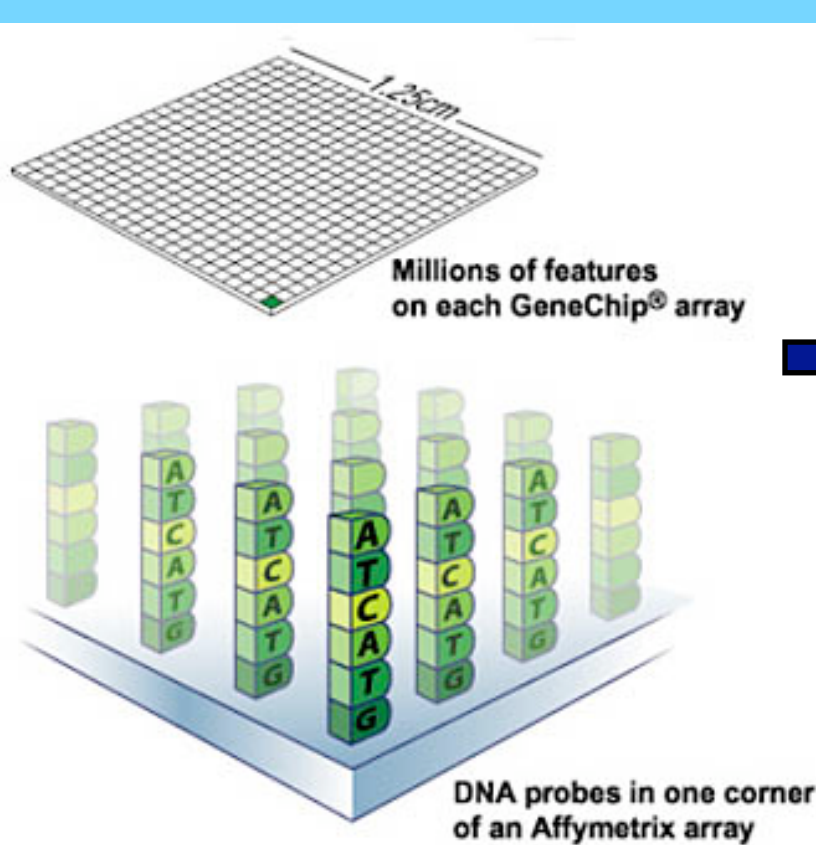
Abstract

We used linear regression to define co-expression relationships for genes encoding metabolic enzymes in Arabidopsis. Using a simple network-analysis algorithm based on intersection of co-expression (ICE), we identified and prioritized candidate pathway members, transcriptional regulators, and cross-pathway control points for 140 metabolic pathways.

Methods

We used microarray expression data from 486 quality-screened Arabidopsis ATH1 array hybridizations from the Nottingham Arabidopsis Stock Center AffyWatch service.

DNA microarrays



expression vectors

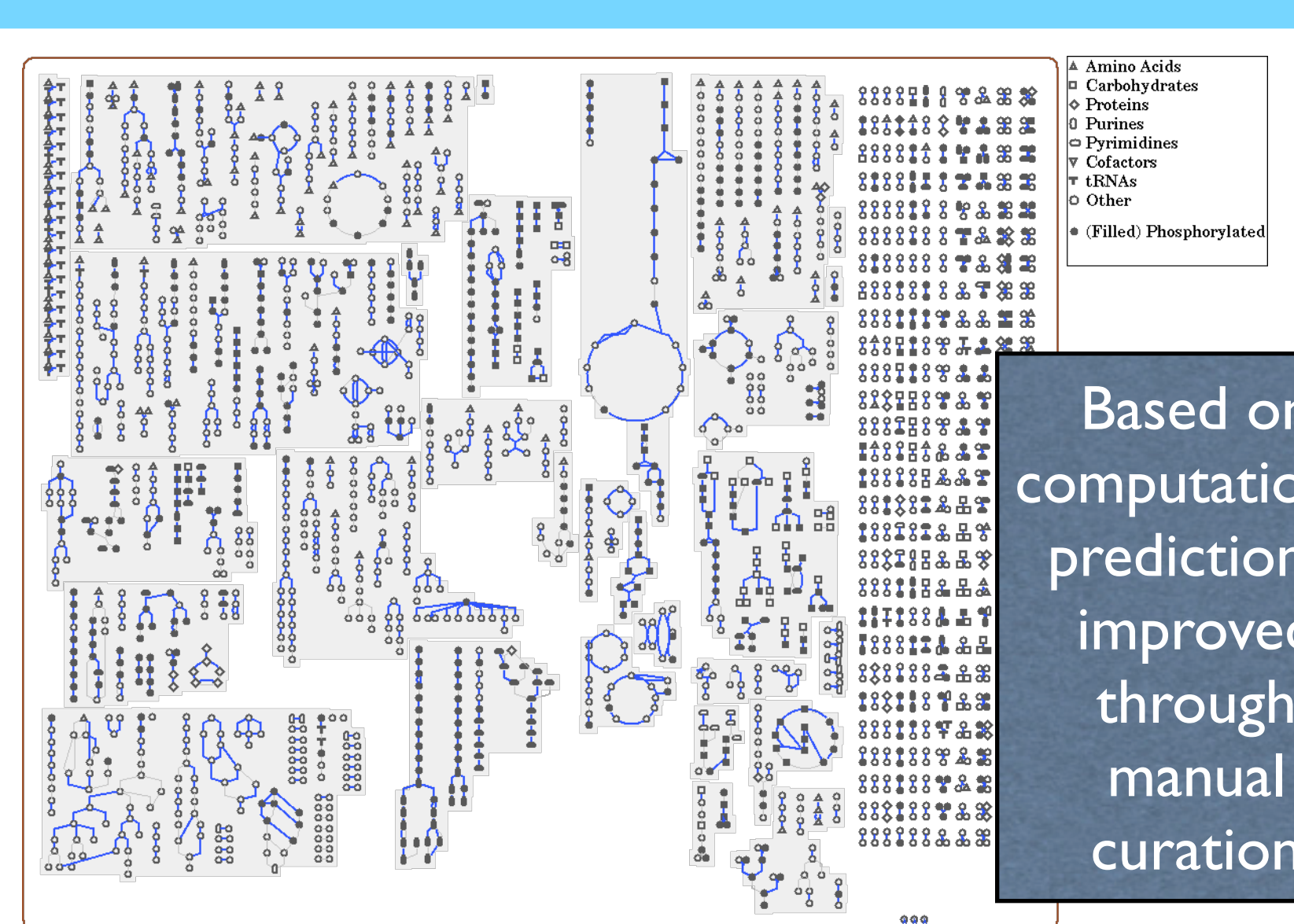
$$\begin{aligned}
 *g_0 &\rightarrow (e_1, e_2, e_3, \dots, e_M) \\
 *g_1 &\rightarrow (e_1, e_2, e_3, \dots, e_M) \\
 *g_2 &\rightarrow (e_1, e_2, e_3, \dots, e_M) \\
 &\dots \\
 *g_N &\rightarrow (e_1, e_2, e_3, \dots, e_M)
 \end{aligned}$$

N genes,
M experiments

Data source:
Nottingham Arabidopsis
Stock Center AffyWatch

The ATH1 expression microarray from Affymetrix contains 1,330 probe sets (out of ~22,000) that measure expression of genes encoding known and predicted metabolic enzymes from the AraCyc database.

AraCyc database of Metabolic Pathways



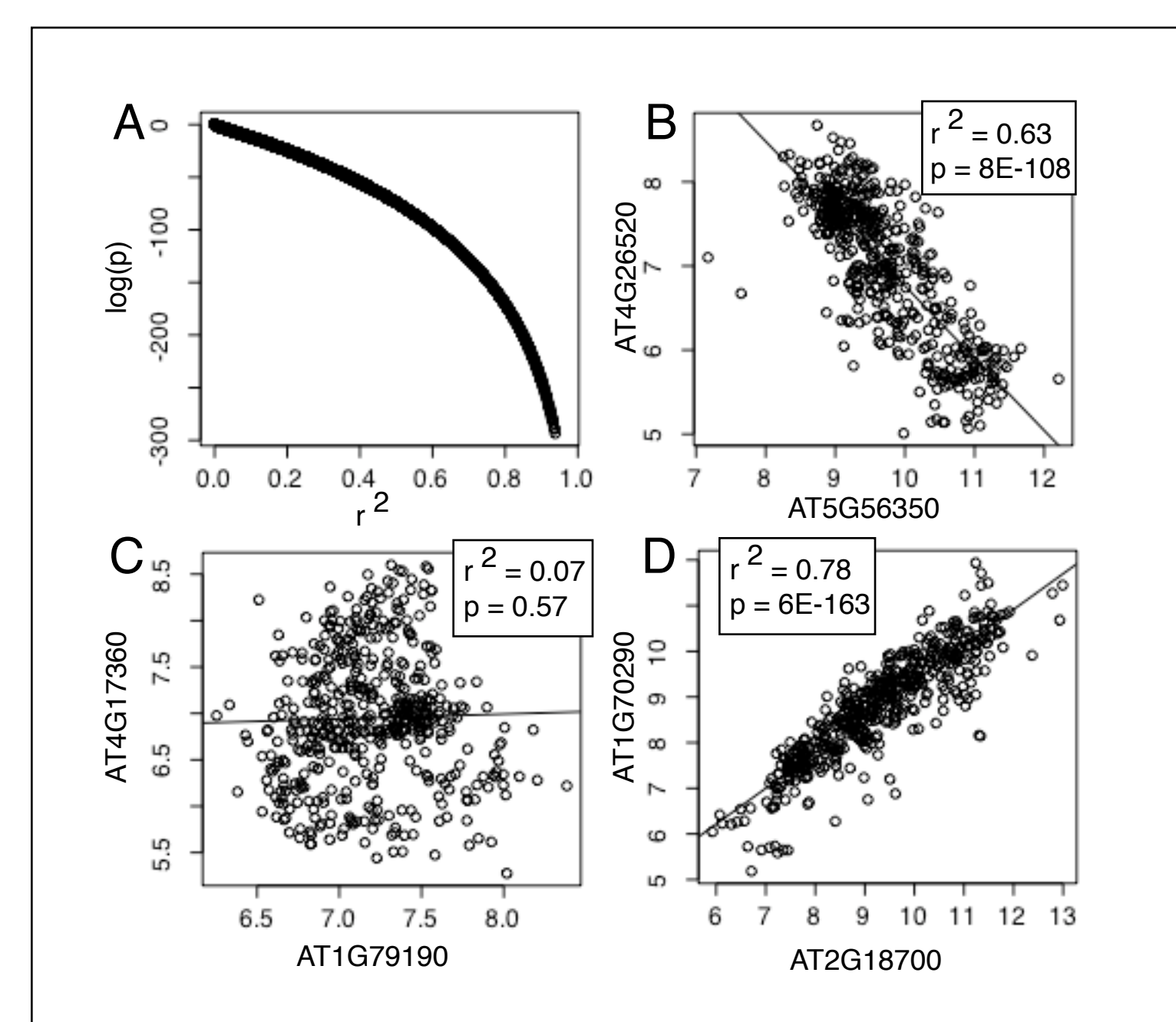
Based on computational predictions, improved through manual curation

Many pathway steps lack annotation, while others are annotated with multiple gene products.

Results

We used linear regression to identify highly-coexpressed gene pairs and define a co-expression network for metabolism.

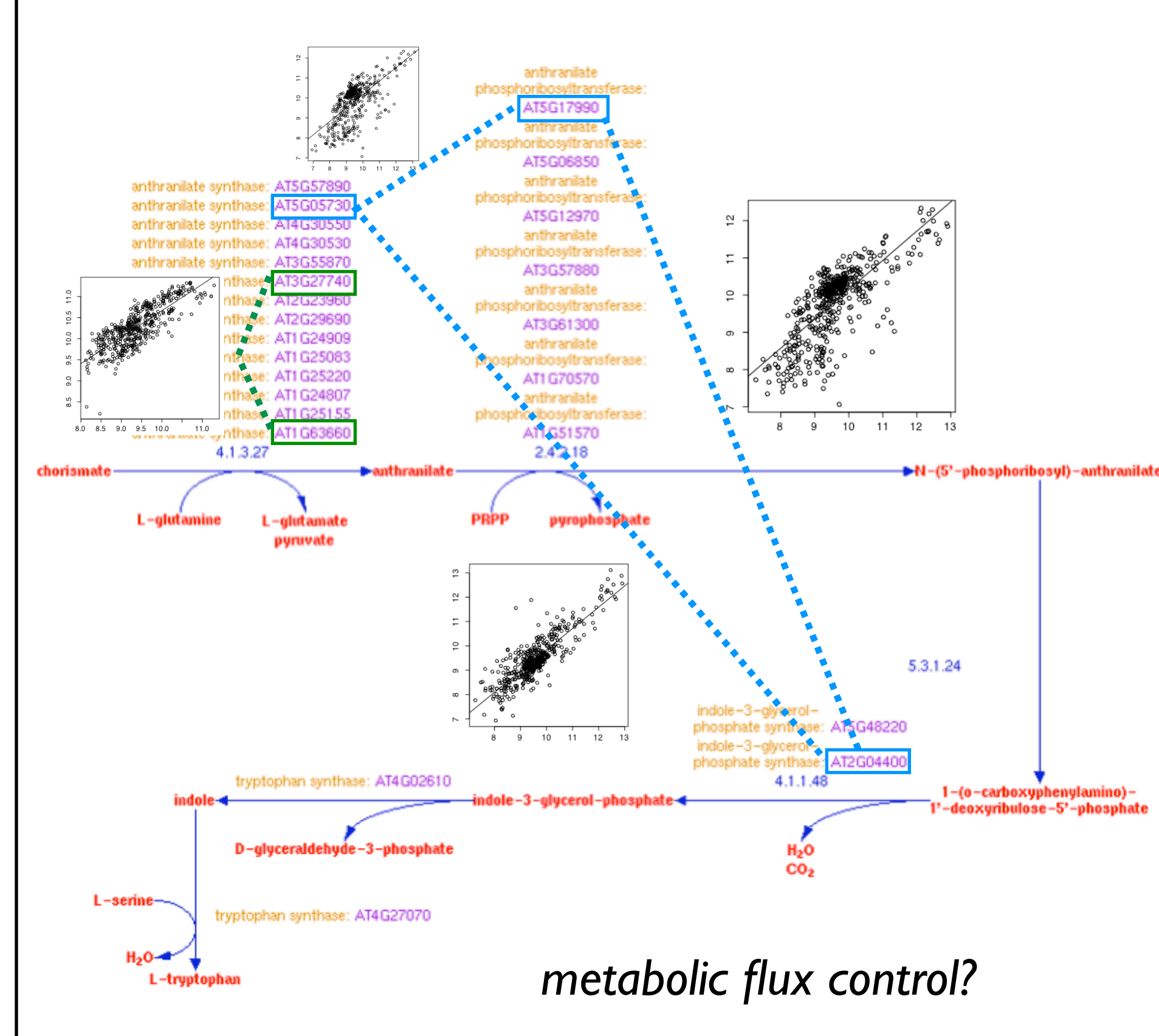
Co-expression as defined by simple linear regression



A. Regression p & r^2 values. Low p and large r^2 values indicate a greater degree of coordinate expression (co-expression). r^2 is the square of the correlation coefficient. **B-C.** Example plots and regression results for weakly (C) and strongly (B,D) co-expressed genes. B & D are examples of negative and positive co-expression.

Most pathways include cliques of genes that are highly-coexpressed with each other even across multiple sample types.

Co-expression cliques in the tryptophan biosynthesis pathway.



Results

Some pathways have unusually many co-expressed genes, more so than would be expected by chance.

Most highly-coexpressed pathways

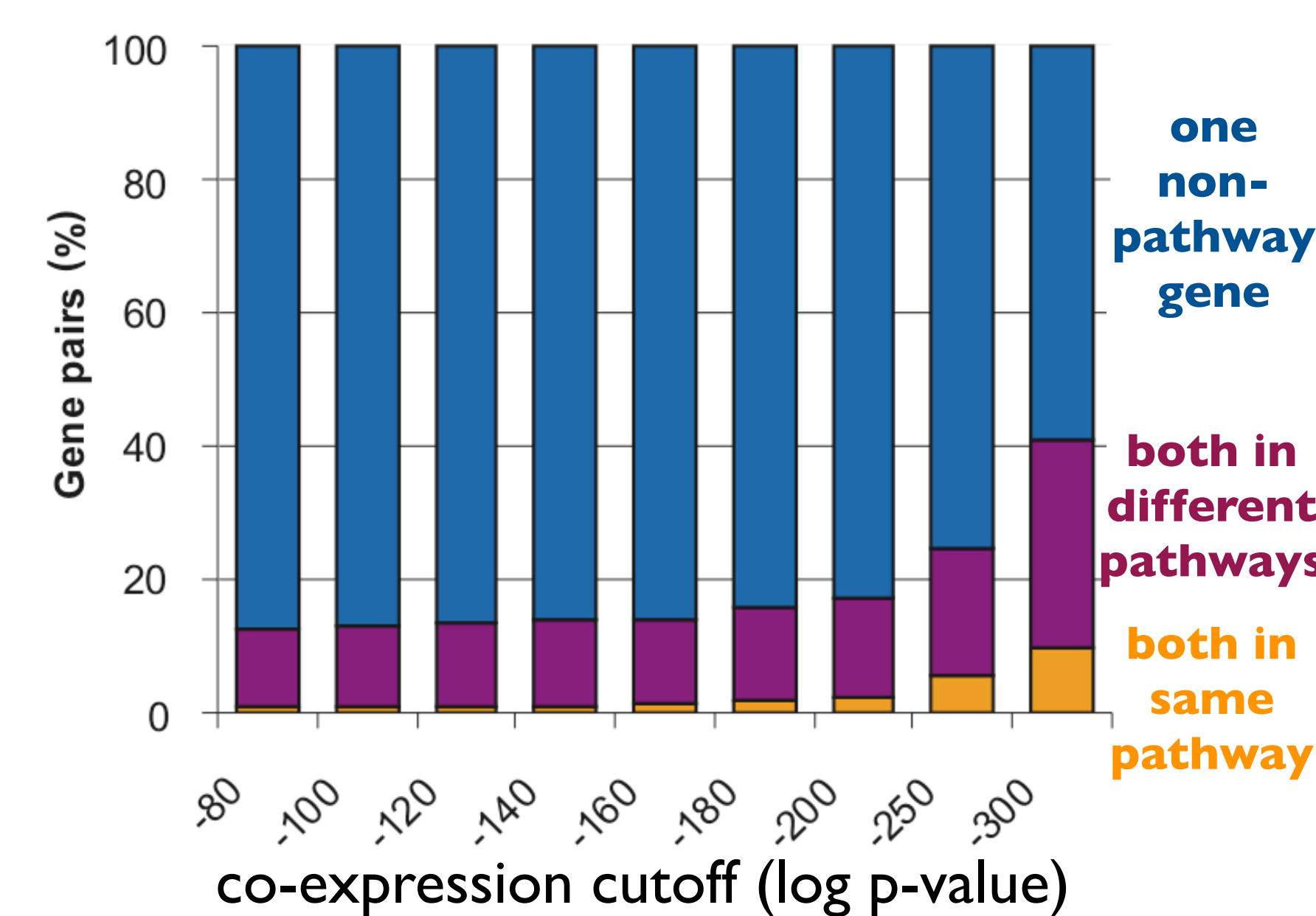
Biosynthesis
chlorophyll
carotenoids
purine nucleotides
arginine
lysine
fatty acids (initial steps)

**Generation of Precursor
Metabolites & energy**
glycolysis I
TCA cycle
photorespiration
acetyl-CoA assimilation
aerobic respiration
photosynthesis, light reaction

**Degradation/Utilization/
Assimilation**
Calvin cycle
glycine degradation

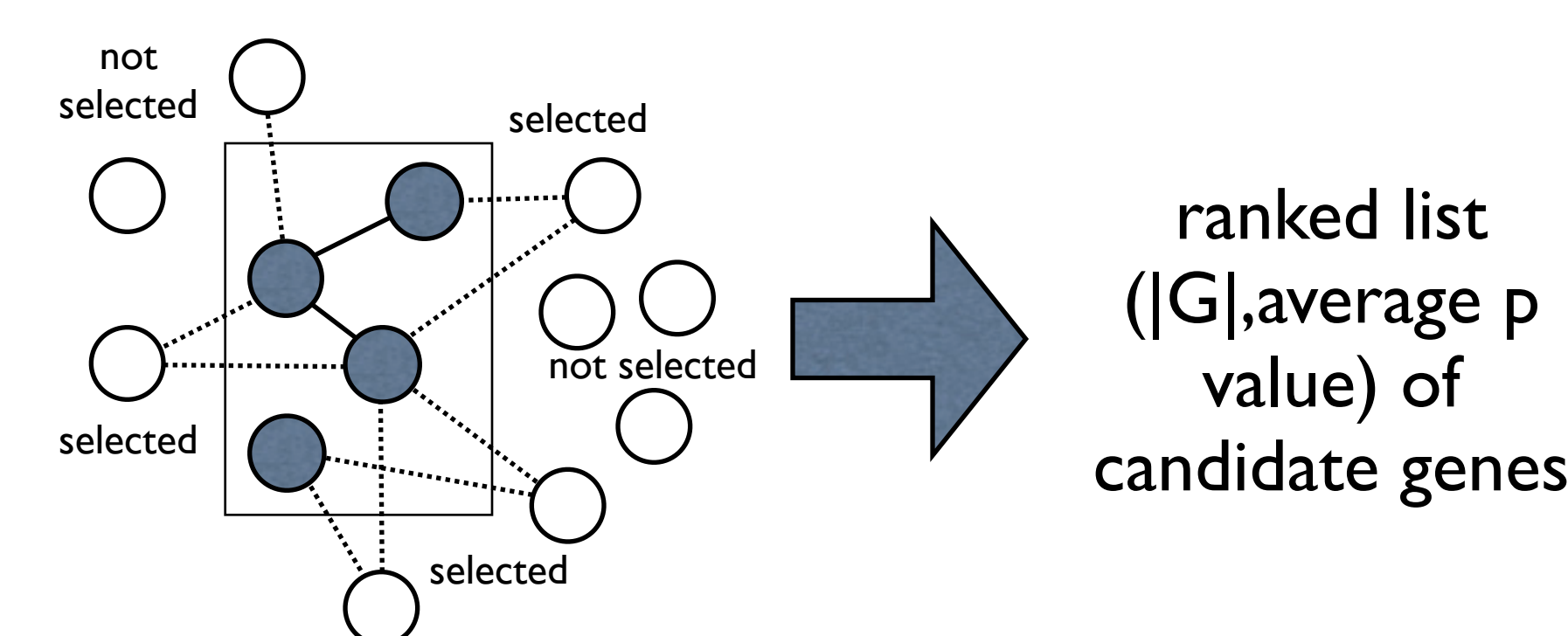
Empirical p -value < 0.001, based on 10,000 random samples

However, most co-expression partners of pathway genes are not in the same pathway. This is true regardless of how stringently we define co-expression.



The basic assumption of our study is that co-expression suggests related function, which can then be tested experimentally or evaluated using other data sources.

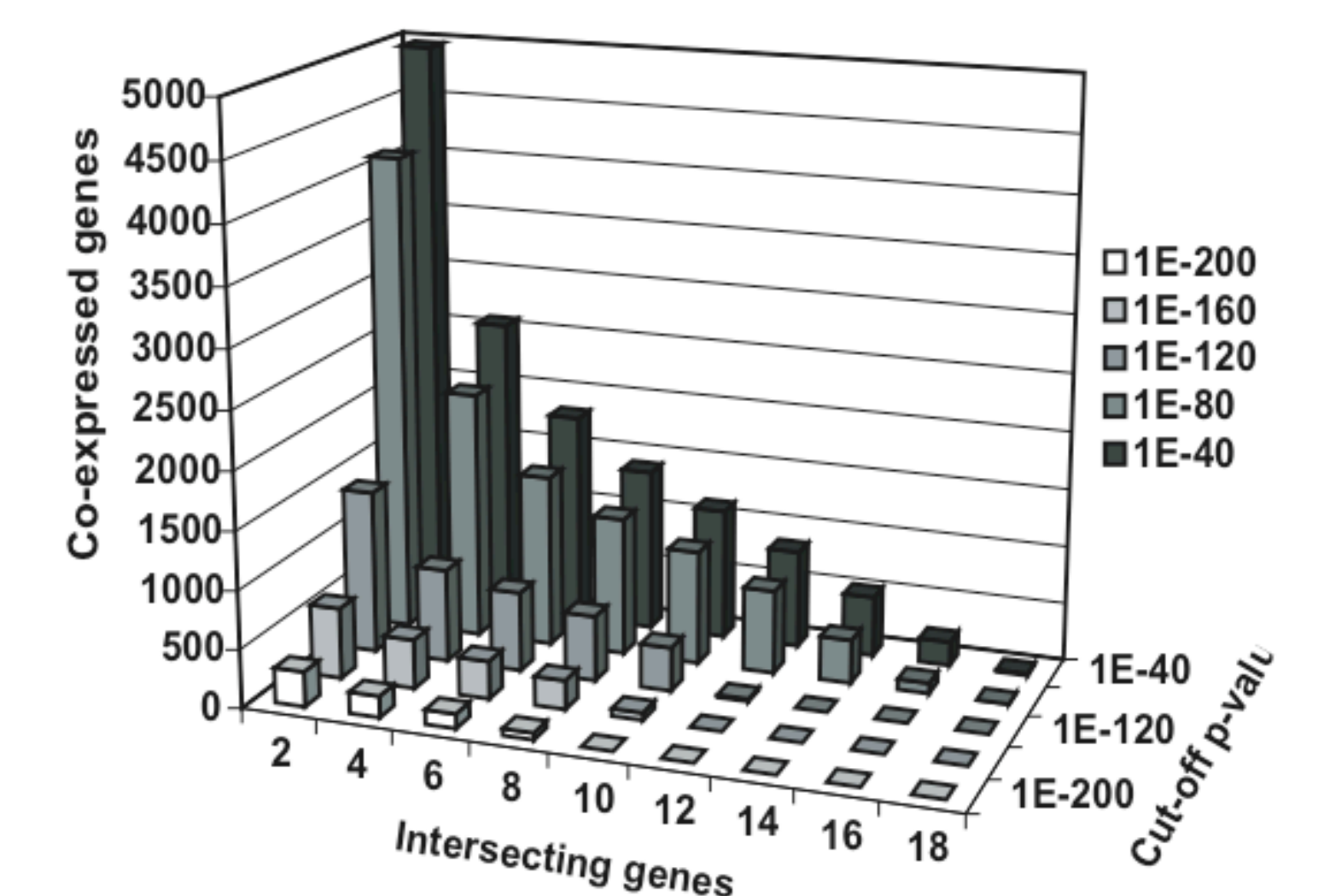
To identify genes co-expressed with a pathway (rather than a single pathway member) we simply select genes connected with multiple pathway members.



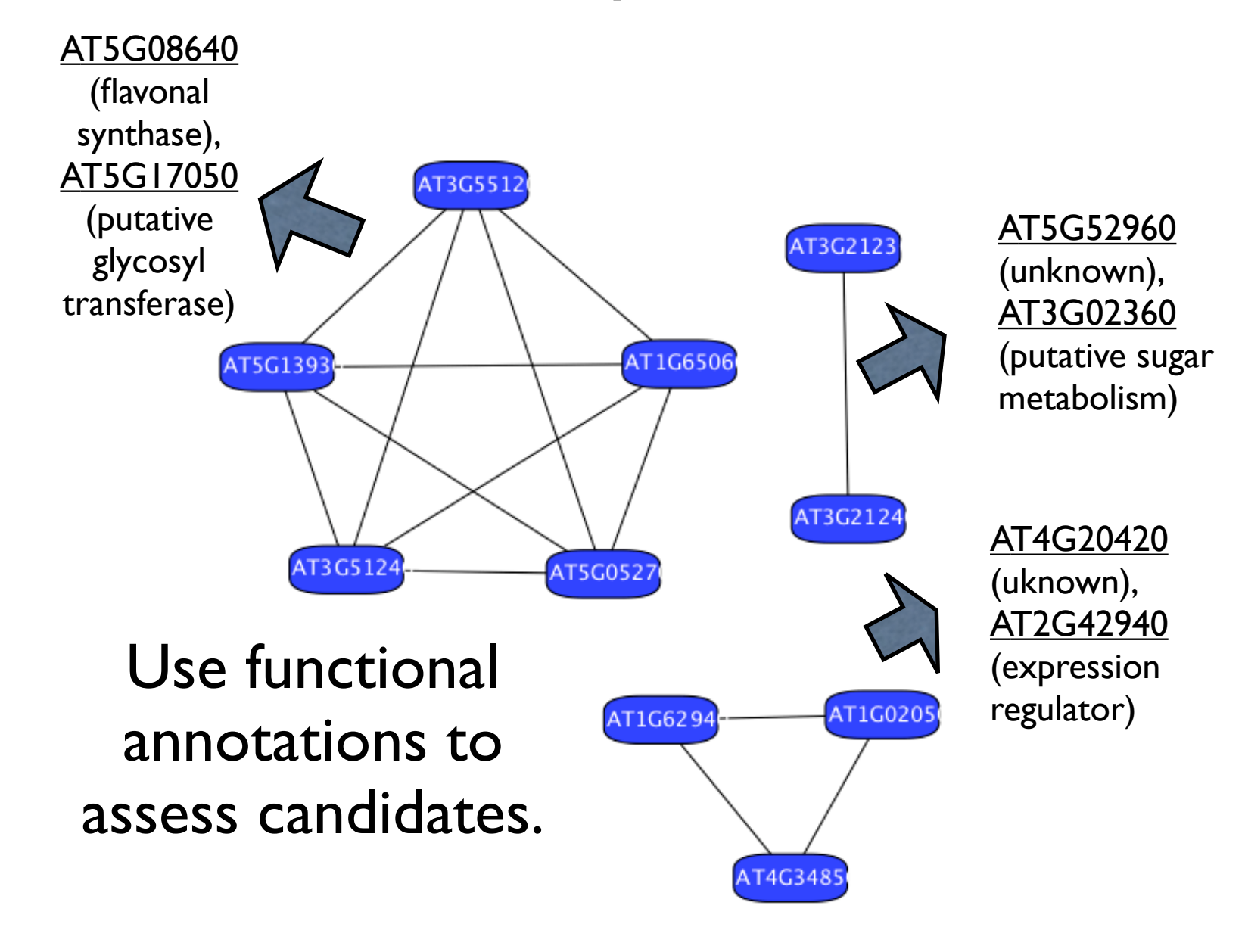
This is Intersection of Co-Expression (ICE). It finds co-expression partners common to two or more members (G) of a group of "bait" genes.

Results

Varying co-expression cutoffs yields varying numbers of candidates. For instance, using 1,330 genes from 209 pathways as "bait," we found 1,500 candidates for 140 pathways at a p value cutoff of $1E-120$.

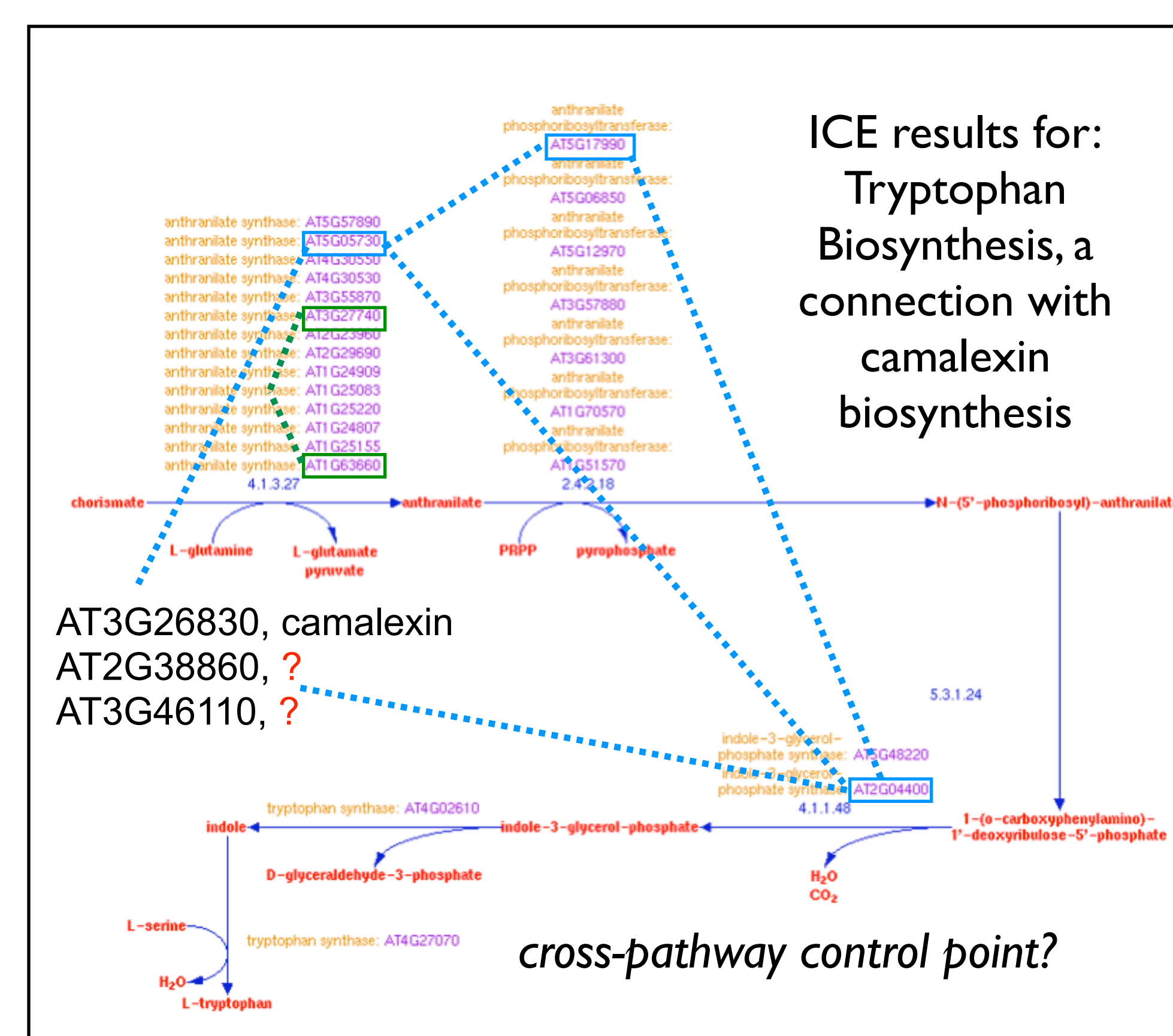


ICE results for: Flavonoid Biosynthesis



Use functional annotations to assess candidates.

ICE results for: Tryptophan Biosynthesis, a connection with camalexin biosynthesis



Conclusions

Co-expression analysis yields numerous, plausible candidates for functionally-related genes. Results are on-line at http://www.transvar.org/at_coexpress/analysis/web.