

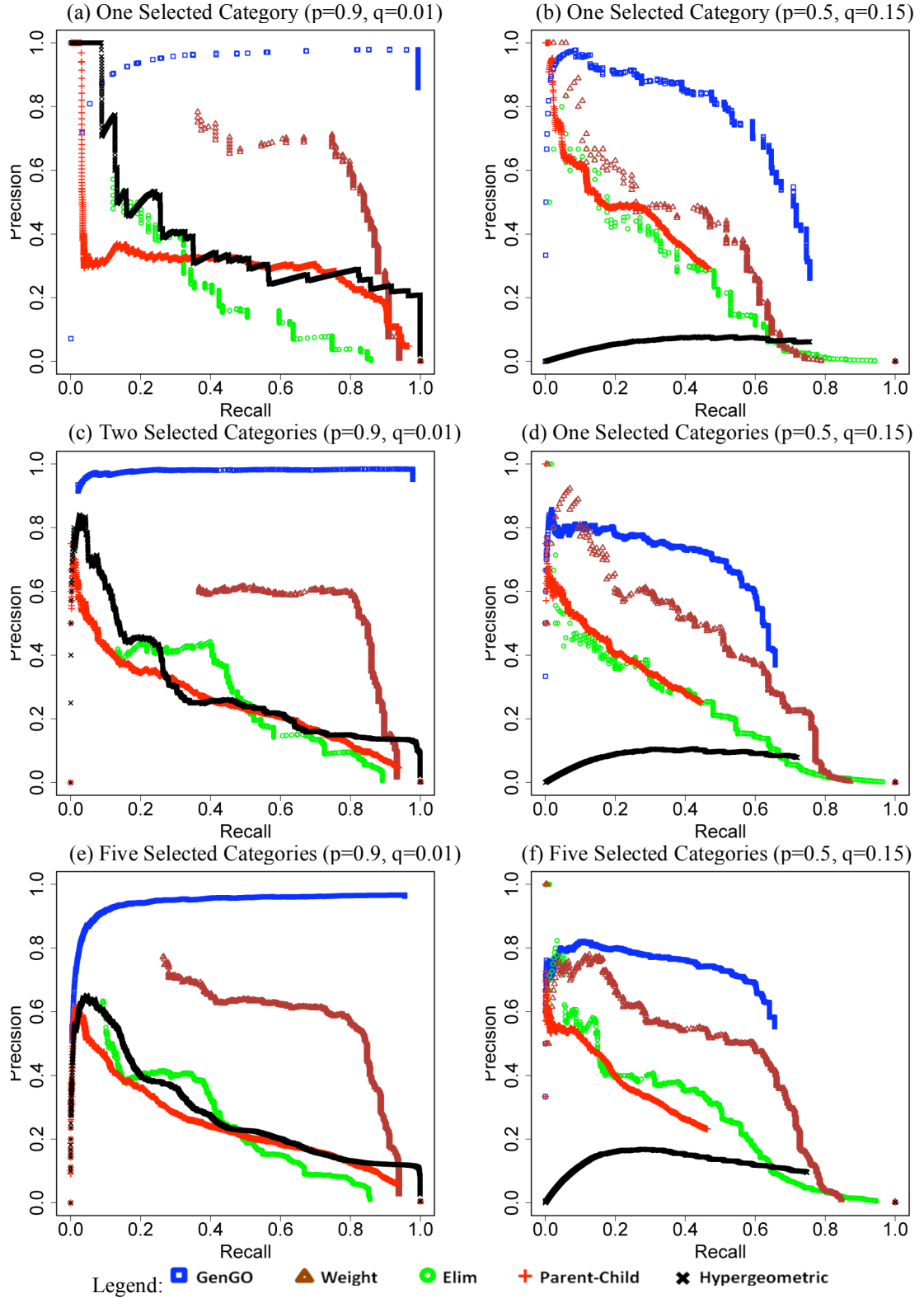
Supplementary Figures for:

A Probabilistic Generative Model for GO Enrichment Analysis

Contents

Supplementary Figures 1-2 (larger set comparison for yeast and human)	2
Supplementary Figure 3 (Elim cell cycle categories)	5
Supplementary Figures 4-8 (amino acid starvation categories)	6
Supplementary Figures 9-13 (Swi6 targets categories)	11
Supplementary Figures 14-18 (E2F1 targets categories)	16

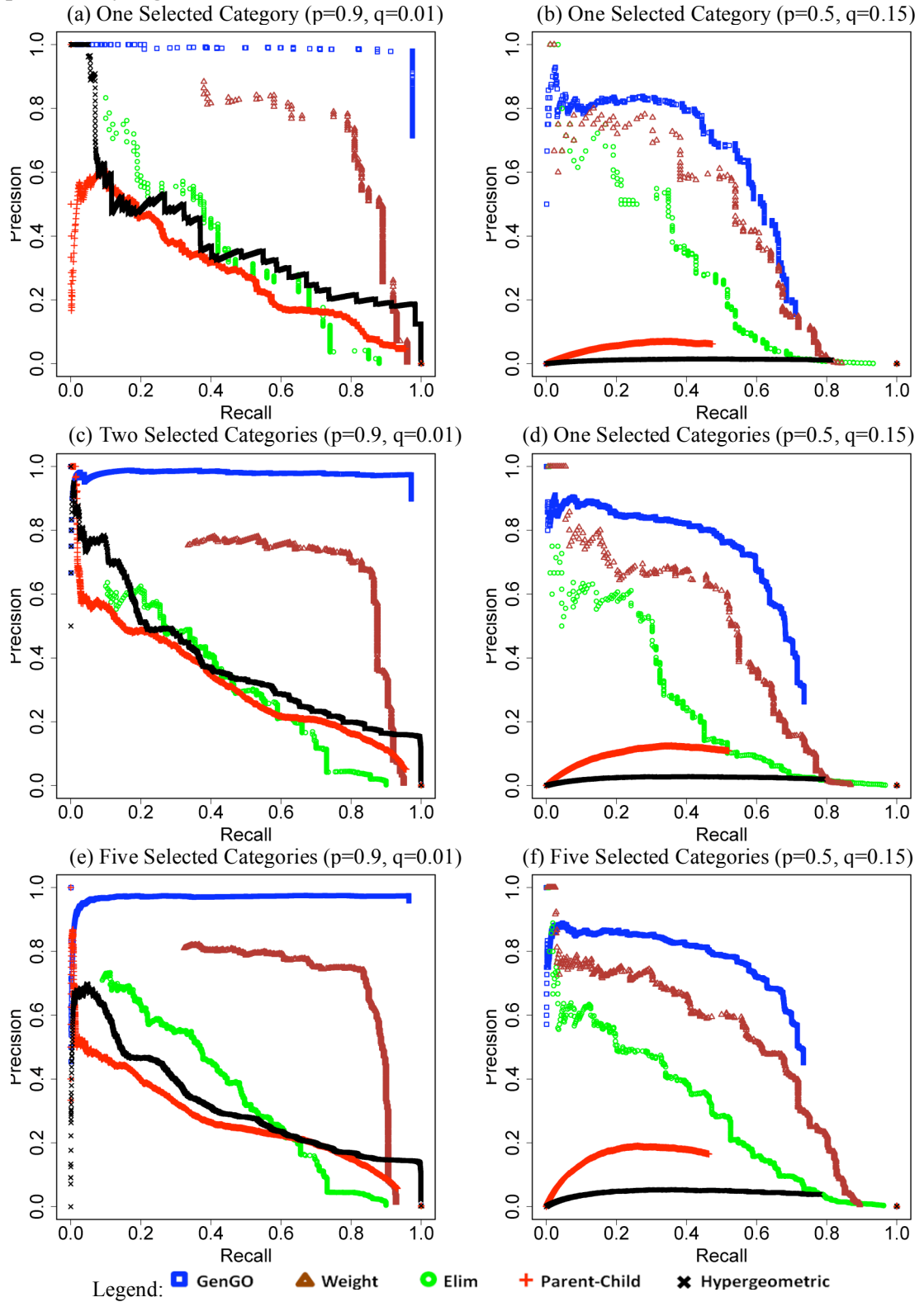
Supplementary Figure 1



Supplementary Figure 1. Comparison using GO for yeast. Performance comparison of five methods on data generated using the yeast GO database. The results for GenGO,

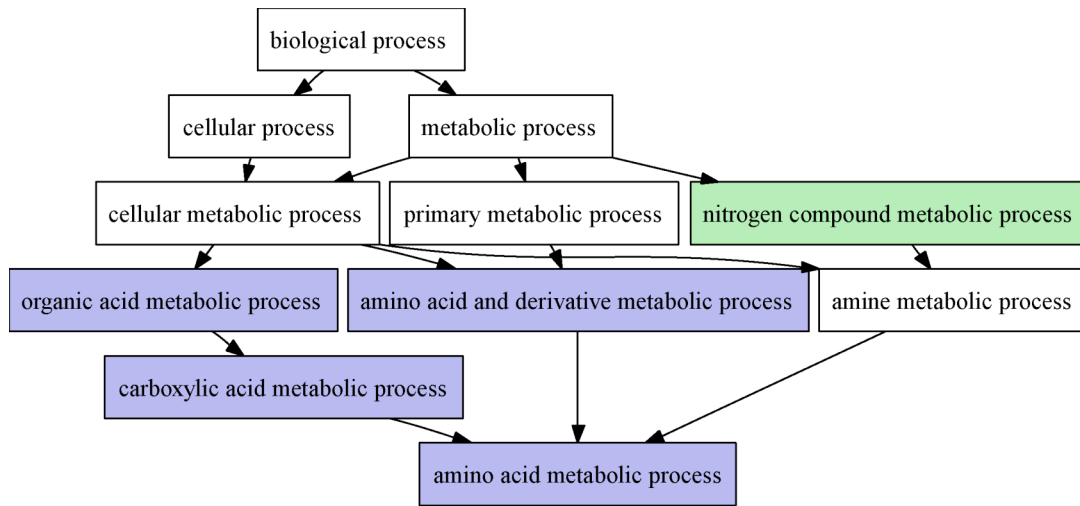
Parent-Child, and Hypergeometric methods were based on 1,000 random gene sets, while the results for Weight and Elim were based on 100 random gene sets. We ran the latter two methods on fewer sets because they are much more time consuming than the other methods. As in Figure 2 and 3 of the main text, we use p to represent the fraction of genes that are identified from an active GO category (true positive rate for a category, see Methods) and q to represent the fraction genes that are selected but do not belong to any active category. (a) Selecting one category with $p = 0.9$, $q=0.01$ (b) Selecting one category with $p = 0.5$, $q=0.15$. (c) and (d): same as (a) and (b) but using two categories. (e) and (f): same with five categories. Note that even when the noise is substantial (using 50% of genes in selected categories and 15% of all other genes, second column) GenGO is still able to accurately recover most of the correct categories.

Supplementary Figure 2

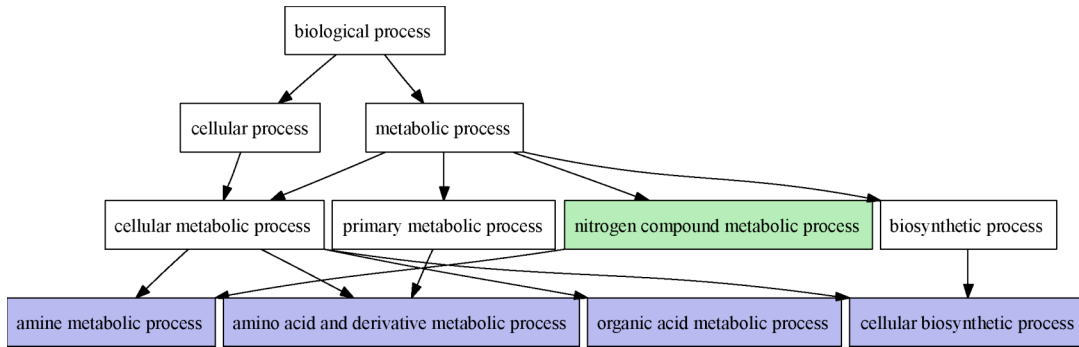


Supplementary Figure 2. Performance comparison of five methods on data generated using human GO database. (a-f) same as in Supplementary Figure 1 for human GO data.

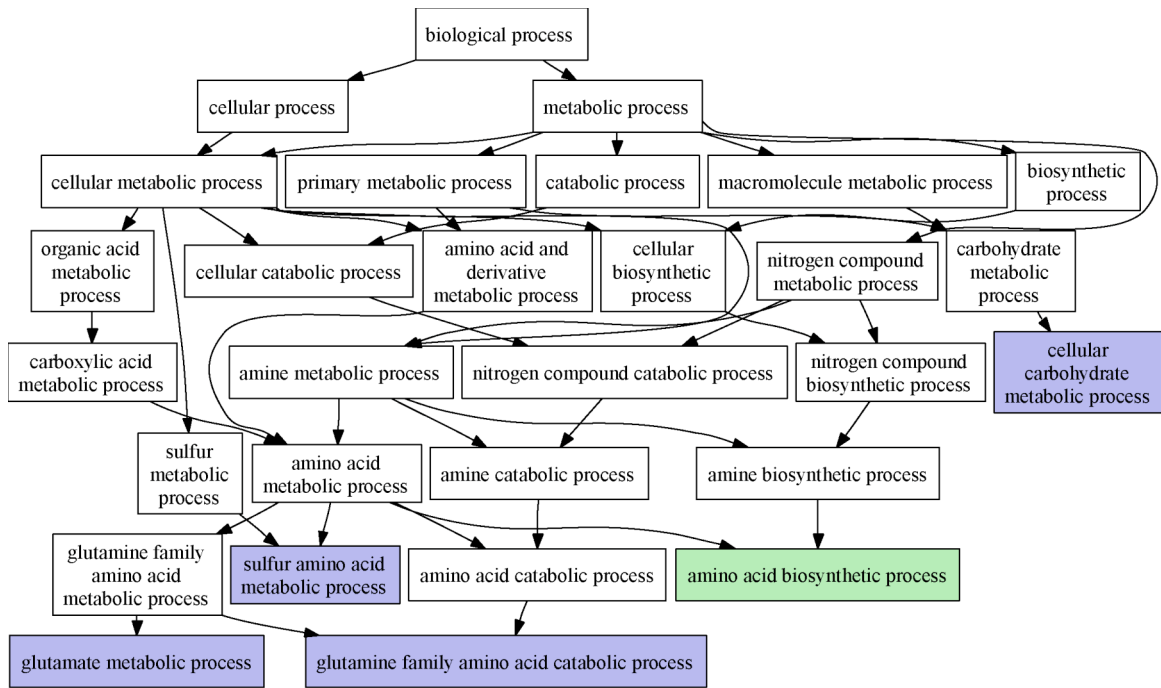
Supplementary Figure 4. Top 5 categories identified by the hypergeometric method for yeast genes induced in amino acid starvation.



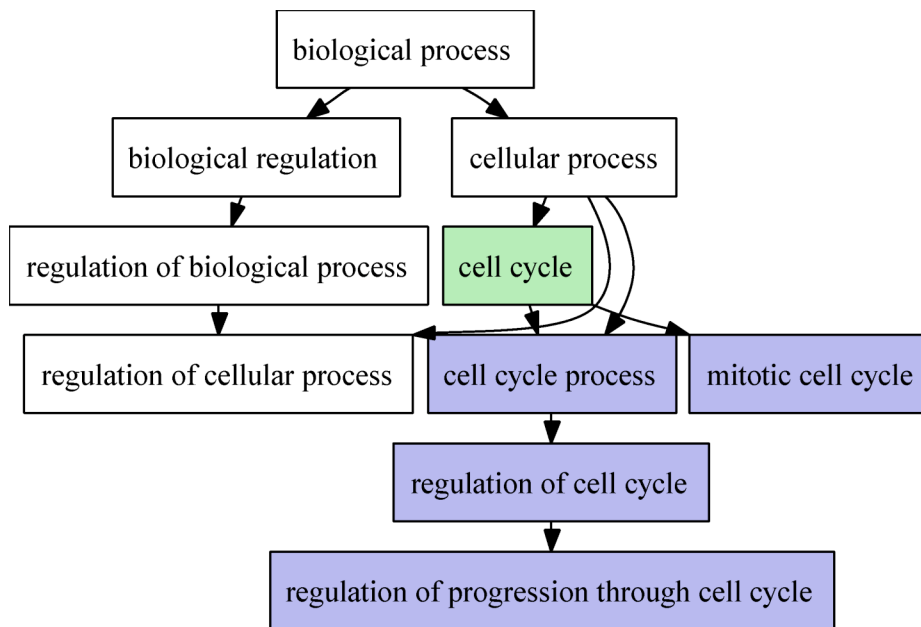
Supplementary Figure 5. Top 5 categories identified by the Parent-Child method for yeast genes induced in amino acid starvation.



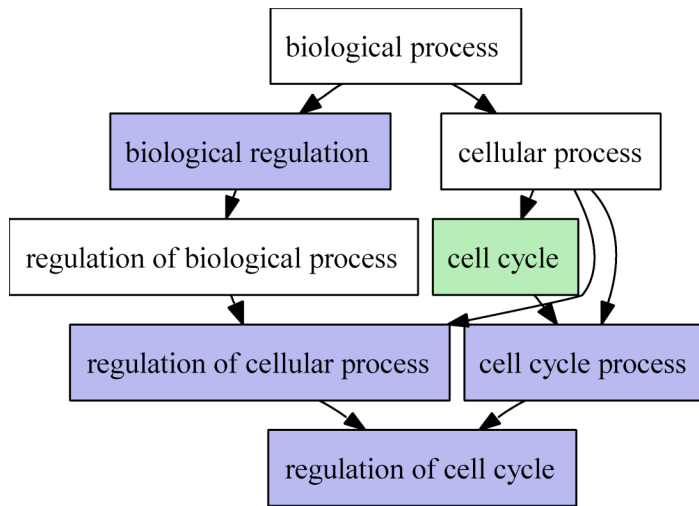
Supplementary Figure 7. Top 5 categories identified by the Weight method for yeast genes induced in amino acid starvation.



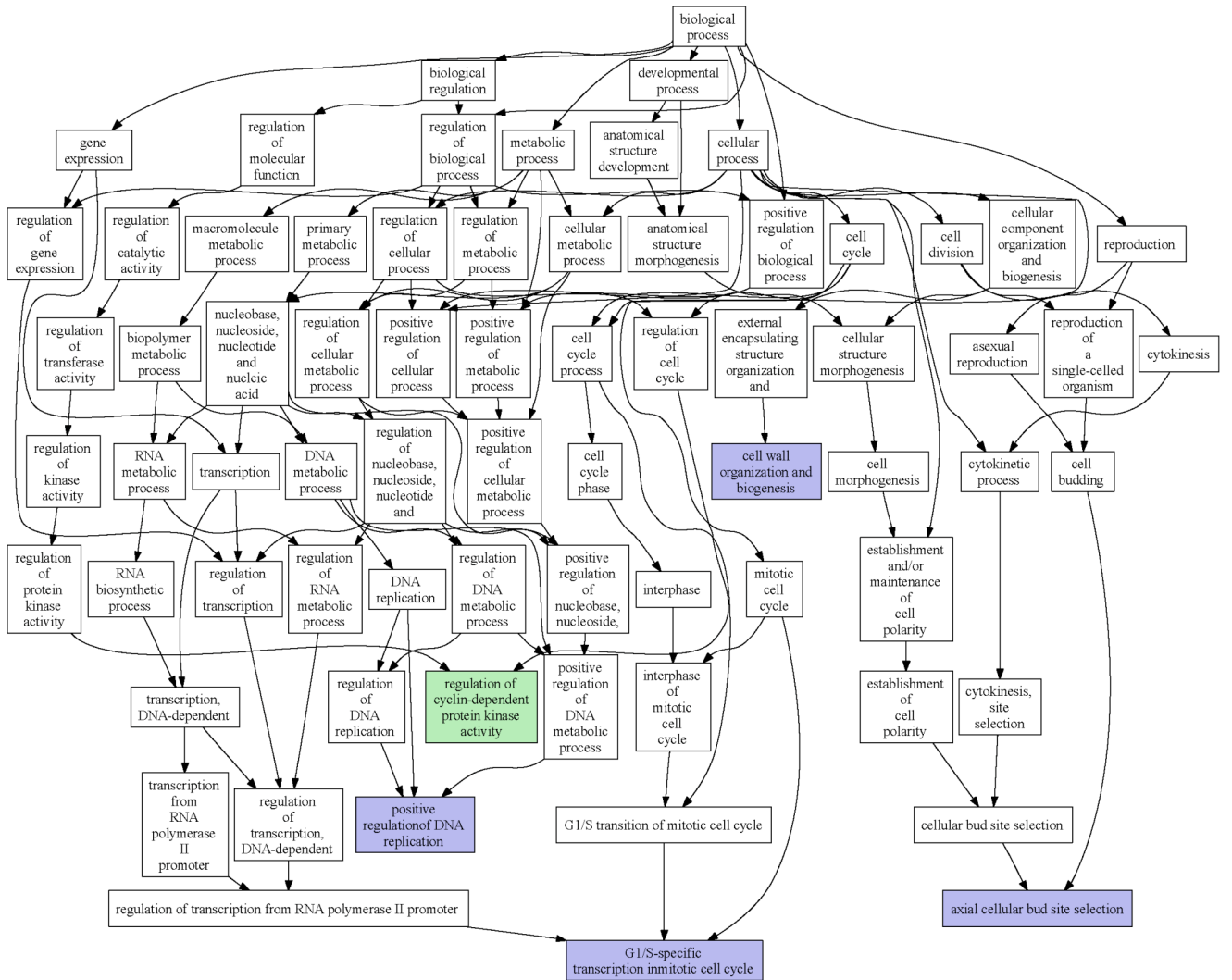
Supplementary Figure 9. Top 5 categories identified by the hypergeometric method for yeast Swi6 targets.



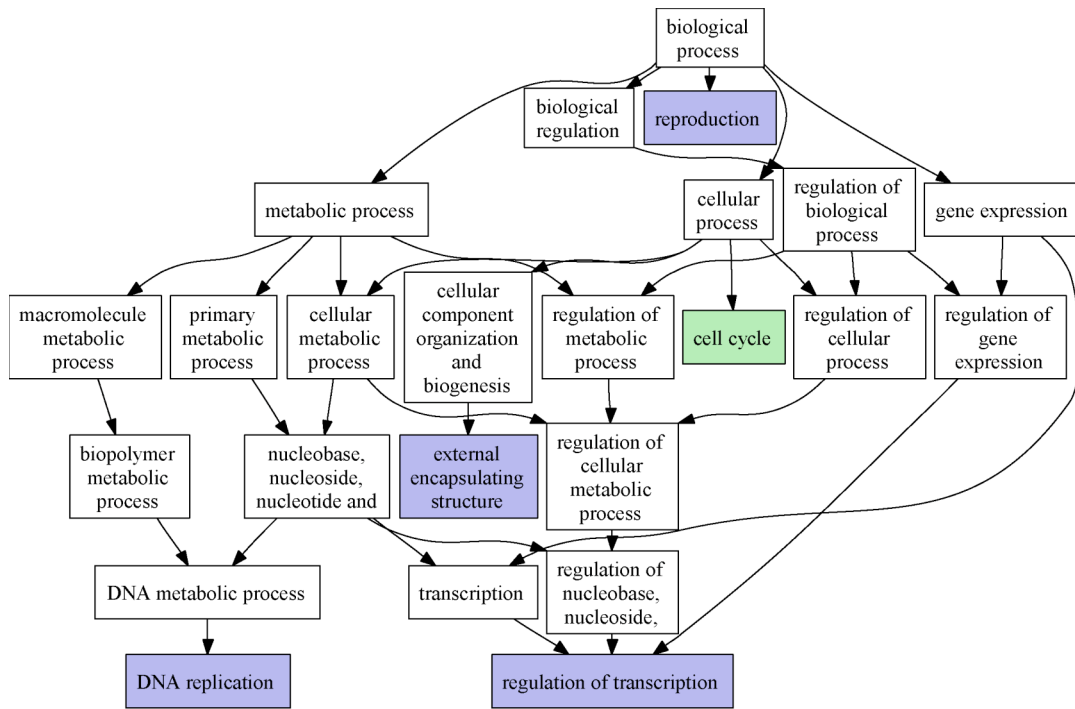
Supplementary Figure 10: Top 5 categories identified by the Parent-Child method for yeast Swi6 targets.



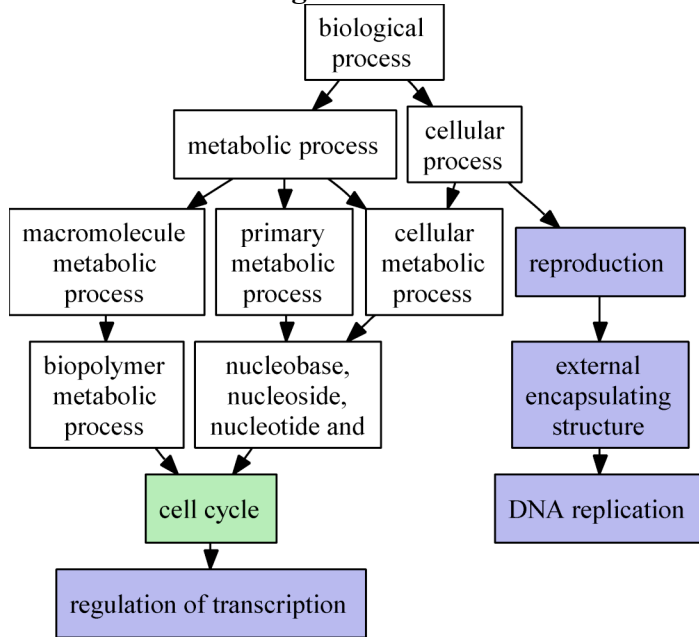
Supplementary Figure 11: Top 5 categories identified by the Elim method for yeast Swi6 targets.



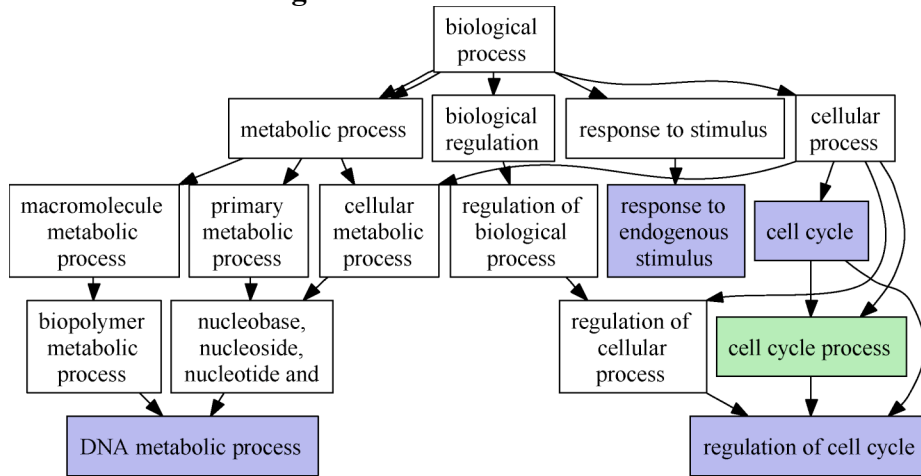
Supplementary Figure 13: Top 5 categories identified by GenGO for yeast Swi6 targets.



Supplementary Figure 14: Top 5 categories identified by hypergeometric method for Human E2F1 targets.



Supplementary Figure 15: Top 5 categories identified by the Parent-Child method for Human E2F1 targets.



Supplementary Figure 17: Top 5 categories identified by the Weight method for Human E2F1 targets.

