FILE S2

GO Annotation Correlation: Scripts to correlate screened genes with gene ontology (GO) categories were written in Python 2.4 plus the scipy module. The GO term hierarchy structure was derived from the file http://www.geneontology.org/ontology/gene_ontology_edit.obo (revision 5.179, dated Feb 22 2007), which for each GO code gives names, definitions, top-level categories ('biological_process', 'molecular_function' or 'cellular_component') and parent and child GO codes. This tree-like structure was parsed into a Python dictionary (mapping) data structure. *S. cerevisiae* gene information, including GO terms (from the Ontology_term attribute) for each gene, was derived from the GFF3 formatted feature file on SGD: ftp://genome-ftp.stanford.edu/pub/yeast/chromosomal_feature/saccharomyces_cerevisiae.gff (version dated Feb 20 2007). Each GO term in the Python data structure was further decorated with the names of all *S. cerevisiae* genes annotated with that GO term. Those GO terms not used in *S. cerevisiae* annotation were subsequently erased from the mapping. The following terms were defined: the total number of *S. cerevisiae* genes considered (N) and the number of genes highlighted in the screen (n). For each screen the following calculations were performed for each GO term in the data mapping: the total number of *S. cerevisiae* genes assigned to that GO term (D), and that measure as a proportion of the total (D/N); the total number of those genes found in the screen (k), and that measure as a proportion of the total number of genes in the screen (k/n); the ratio k/D; and the probability (p-value) of randomly achieving k hits from a sample size of n when there are a total of D possible hits in the full population size N. Scipy's 'exp' and 'special.gammaln' functions were used to efficiently calculate the hyper-geometric distribution for the determination of this p-value. The same calculations were performed a second time but counting all child GO terms (and associated genes) as part of each parent GO term. The data were exported to Microsoft Excel where GO terms could be ranked by p-value (by either set of calculations), thereby highlighting which GO terms were statistically over-represented within the subpopulation of screened genes.