Supplementary Tables for:

A Probabilistic Generative Model for GO Enrichment Analysis

Contents

Supplementary Table 1: Analysis of random gene sets.

Random Genes	Classic	Parent-Child	Elim	Weight	GenGO
1%	69%	100%	67%	64%	100%
5%	0%	83%	74%	71%	98%
10%	0%	7%	51%	44%	100%

Supplementary Table 1: 1%, 5%, and10% of all human genes were randomly selected as a test set, and the five algorithms were run to identify significant categories. Categories were only selected if they achieved a p-value < 0.001 following Bonferroni correction for multiple hypothesis testing. The procedure is repeated 100 times, and the percentages of sets ***without*** any significant GO categories are listed in the table. As can be seen, while GenGO correctly determined that there were no significant categories in more than 98% of tests, other methods identified much more erroneous categories in these experiments.

Supplementary Table 2: Additional analysis of random gene sets.

Random Genes	Hypergeometric	Parent-Child	Elim	Weight	GenGO
0.1%	87%	100%	23%	19%	98%
0.2%	13%	90%	22%	23%	98%
0.5%	0%	42%	2%	11%	89%

Supplementray Table 2: 0.1%, 0.2%, and 0.5% genes from each level-2 categories ("biological_process" being at level 1) were selected into a test set, and the five algorithms were run to identify significant categories. The procedure is repeated 100 times, and the percentages of sets without any significant GO categories are listed in the table. Categories were only selected if they achieved a p-value < 0.001 following Bonferroni correction for multiple hypothesis testing.

Supplementary Table 3: GO Analysis of yeast cell cycle genes in different phases

	Classic	Parent-Child	Elim	Weight	GenGO
G1	DNA replication	DNA metabolic process	mitotic sister chromatid cohesion	DNA strand elongation during DNA replica...	DNA replication
	DNA-dependent DNA replication	cell cycle	lagging strand elongation	mitotic sister chromatid cohesion	mitotic sister chromatid cohesion
	DNA metabolic process	cell cycle process	microtubule nucleation	DNA repair	microtubule nucleation
	DNA strand elongation during DNA replication	DNA replication	mismatch repair	microtubule nucleation	telomere maintenance via recombination
	DNA strand elongation	response to endogenous stimulus	leading strand elongation	DNA replication	septin cytoskeleton organization and biogenesis
S	sulfur metabolic process	sulfur metabolic process	sulfate assimilation	chromatin assembly or disassembly	sulfur metabolic process
	sulfur amino acid metabolic process		chromatin assembly or disassembly	sulfur amino acid metabolic process	chromatin assembly or disassembly
	sulfur amino acid biosynthetic process		methionine biosynthetic process	sulfate assimilation	microtubule-based process
	sulfur compound biosynthetic process		microtubule nucleation	microtubule nucleation	
	sulfur utilization		mitotic spindle organization and biogene...	mitotic spindle organization and biogene...	
S/G2			nuclear migration, microtubule-mediated	nuclear migration, microtubule-mediated	nuclear migration, microtubule-mediated
			methionine metabolic process	methionine metabolic process	methionine metabolic process
			negative regulation of microtubule depol...	response to xenobiotic stimulus	amine transport
			amino acid biosynthetic process	organelle inheritance	polysaccharide biosynthetic process
			amino acid transport	axial bud site selection	axial cellular bud site selection
G2/M	cation transport	ion transport	cation transport	iron ion transport	cation transport
	ion transport		DNA unwinding during	DNA unwinding during	DNA unwinding during

			replication	replication	replication
	metal ion transport		siderophore-iron transport	arginine catabolic process	amino acid catabolic process
	iron ion transport		arginine catabolic process	nuclear division	polyamine transport
	transition metal ion transport		nuclear division	ATP transport	G1-specific transcription in mitotic cell cycle
M-G1	response to pheromone during conjugation with cellular fusion	multi-organism process	hexose transport	response to pheromone during conjugation...	response to pheromone during conjugation with cellular fusion
	response to pheromone	carbohydrate transport	pheromone-dependent signal transduction ...	hexose transport	monosaccharide transport
	conjugation with cellular fusion	response to pheromone	agglutination during conjugation with ce...	N-terminal protein lipidation	protein myristoylation
	conjugation		N-terminal protein lipidation	N-terminal protein myristoylation	pre-replicative complex formation
	sexual reproduction		N-terminal protein myristoylation	pre-replicative complex formation	telomere maintenance via recombination

Supplementary Table 2: Top five GO categories identified by different methods from yeast cell cycle genes whose expression peak in each cell cycle phase (Spellman et al. 1998).

Supplementary Table 4: Categories for amino acid starvation

Classic	Parent-Child	Elim	Weight	GenGO
nitrogen compound metabolic process	nitrogen compound metabolic process	arginine biosynthetic process	amino acid biosynthetic process	amino acid biosynthetic process
carboxylic acid metabolic process	organic acid metabolic process	glutamate biosynthetic process	glutamate metabolic process	sulfur metabolic process
organic acid metabolic process	amino acid and derivative metabolic process	sulfate assimilation	sulfur amino acid metabolic process	amino acid catabolic process
amino acid metabolic process	amine metabolic process	transposition, RNA-mediated	main pathways of carbohydrate metabolic process	purine base metabolic process
amino acid and derivative metabolic process	cellular biosynthetic process	methionine biosynthetic process	glutamine family amino acid catabolic process	monosaccharide catabolic process

Supplementary Table 3: Top five GO categories identified by different methods from the list of yeast genes induced following amino acid starvation (Gasch et al. 2000).

Supporting Table 5: Categories for Swi6 targets identified by ChIP-chip experiments.

Hypergeometric	Parent-Child	Elim	Weight	GenGO
cell cycle	cell cycle	regulation of cyclin-dependent protein kinase activity	regulation of cyclin-dependent protein kinase activity	cell cycle
mitotic cell cycle	cell cycle process	G1/S-specific transcription in mitotic cell cycle	interphase of mitotic cell cycle	external encapsulating structure organization and biogenesis
regulation of progression through cell cycle	biological regulation	cell wall organization and biogenesis	regulation of progression through mitotic cell cycle	DNA replication
regulation of cell cycle	regulation of cellular process	axial bud site selection	axial bud site selection	reproduction
cell cycle process	regulation of cell cycle	positive regulation of DNA replication	cell wall organization and biogenesis	regulation of transcription

Supplementary Table 4: Top five GO categories identified from the list of yeast Swi6 targets determined by ChIP-chip (Harbison et al. 2004).

Supporting Table 6: Categories for Human E2F1 targets identified by ChIP-chip experiments

Hypergeometric	Parent-Child	Elim	Weight	GenGO
DNA metabolic process	cell cycle process	cell division	DNA replication	DNA replication
cell cycle process	cell cycle	DNA replication	mitosis	Double-strand break repair
cell cycle	DNA metabolic process	DNA replication initiation	cell division	mitotic checkpoint
DNA replication	response to endogenous stimulus	mitosis	regulation of progression through cell cycle	mitotic syster chromatid segregation
cell cycle phase	regulation of cell cycle	regulation of cyclin-dependent protein kinase activity	DNA repair	G2/M transition of mitotic cell cycle

Supplementary Table 5: Top five GO categories identified from the list of human E2F1 targets determined by ChIP-chip (Ren et al. 2002).