

## Additional Data File 1

Detailed text profiles for clusters {B, C, D, E, F, G, H, J, and K} in

Eisen M, Spellman P, Brown P, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, 95:14863-14868.

For each cluster, the second column contains the top 10 scoring keywords extracted from MEDLINE abstracts that are associated in SGD to the genes of that cluster (listed in column 1). The last column contains the top 10 keywords (provided that many are available) derived from the GO annotations for the gene cluster under consideration.

In clusters B and C we observe the complementary nature of MEDLINE information. Terms extracted from MEDLINE such as 'spindle pole body', 'cytokinesis' and 'anaphase promotor complex' are more reminiscent of the overall function of cluster B than the most of the less-specific GO-extracted terms such as 'cytoskeleton', 'protein binding' and 'bud neck'. One exception is the GO-term 'microtubuli nucleation', which contributes to the biogenesis of the microtubule cytoskeleton. For cluster C, the difference is even more pronounced as the GO term 'degrad' is the only one referring (vaguely) to the asserted proteasome function of this cluster. For the remaining clusters, terms extracted from MEDLINE and GO refer to the same biological function. Hence, term-based analysis of literature can be valuable to complement annotations in public databases.

SGD yeast gene names	MEDLINE-based profiles	GO-annotation -based profiles
<b>CLUSTER B (spindle pole assembly and function)</b>		
stu2	spindl_pole_bodi	cytoskeleton
dhs1	cyclin	cyclin
bnr1	bud	axial_bud
spc42	mitosi	microtubul_nucleat
cnm67	cytokinesi	protein_bind
clb4	septin	g2
cdc10	spindl	bud_neck
cdc3	cell_cycl	division
clb3	anaphas_promot_complex	degrad
apc4	anaphas	g1_transition
cdc16		
<b>CLUSTER C (proteasome)</b>		
rpn11	proteasom	degrad
ufd1	26s_proteasom	endopeptidas

rpn9	regulatori	cytoplasm
rpt1	degrad	nuclear
rpt6	ubiquitin	
pre4	particl	
rpn6	19s	
rpt4	proteolysi	
rpn7	core	
rpn3	similar	
pup2		
scl1		
pre5		
pre9		
pre1		
pre2		
pre3		
pre10		
pup1		
pre6		
pre7		
rpn10		
rpt3		
prt5		
rpn12		
rpn5		
rpn8		
<b>CLUSTER D (mRNA splicing)</b>		
pop6	mitochondri	mrna_splice
caf16	rna	nuclear
mrs1	splice	transport
smd3	mrna	mrna_process
taf40	interact	rrna_process
pet309	nuclear	snrnp_recycl
pzf1	transcript	gtpase
ysh1	u6	pre_mrna_splice_factor
cbp1	u4	mitochondri_genom_mainten
prp24	matur	integr_membran_protein
std1		mitochondri
mgm1		
prp19		
sls1		
<b>CLUSTER E (glycolysis)</b>		
tpt1	glyceraldehyd_3_phosphat_dehydrogenas	gluconeogenesi
gpm1	glycolyt	glycolysi
pgk1	pyruv_decarboxylas	cytoplasm

tdh3	glucos	glyceraldehyd_3_phosphat_dehyd rogenas
tdh2	enzym	phosphoryl
eno2	ethanol	transketolas
tdh1	carbon	hexokinas
fba1	glycolysi	phosphoglycer_kinas
tkl1	ferment	phosphoglycer_mutas
pdc5	pyruv_kinas	pyruv_kinas
pdc6		
pdc1		
cdc19		
hvk2		
tye7		
pfk1		
acs2		
<b>CLUSTER F (mitochondrial ribosome)</b>		
dbi56	mitochondri_ribosom	mitochondri
mss51	mitochondri	biosynthesi
msf1	mrp	mitochondri_matrix
mrps5	protein_biosynthesi	protein_biosynthesi
msm1	commun	peptidyltransferas
cox10	cytoplasm	ribosom
msk1	translat	respir
mef1	individu	mrna_process
none	ribosom	mitochondri_intermembran_space
mrpl35	similar	cytochrom_c_oxidasi_biogenesi
mrp7		
none		
cox17		
mrp17		
mrpl6		
mrpl8		
rml2		
mrpl13		
mrpl9		
pet123		
ppa2		
img2		
<b>CLUSTER G (ATP synthesis)</b>		
sds22	synthas	atpas
cox4	atp	transport
cox13	mitochondri	cytochrom_c_oxidasi
cox8	assembli	mitochondri
atp1	cytochrom_c_oxidasi	mitochondri_inner_membran
atp5	oxid_phosphoryl	nuclear

atp2	atpas	ubiquinol_cytochrom_c_reductas
cox12	f	
atp4	growth	
atp17	aerob	
atp14		
atp3		
atp16		
atp7		
qcr8		
<b>CLUSTER H (chromatin structure)</b>		
hht1	histon	nuclear
hht2	h4	disassembli
hhf2	h2a	chromatin_assembli
hhf1	h3	dna_bind
htb1	h2b	
hta2	nucleosom	
hta1	chromatin	
htb2	core	
	tail	
	transcript	
<b>CLUSTER J (DNA replication)</b>		
cdc54	mcm	atp_depend_dna_helicas
mcm3	dna_replic	dna_replic_initi
mcm2	replic	cell_cycl
cdc47	initi	nuclear
dbf2	origin	
	cell_cycl	
	phase	
	mainten	
	mitosi	
	chromatin	
<b>CLUSTER K (tricarboxylic acid cycle and respiration)</b>		
por1	cytochrom	mitochondri_inner_membran
none	succin_dehydrogenas	ubiquinol_cytochrom_c_reductas
sdh1	mitochondri	ubiquinon
ndi1	carbon	succin_dehydrogenas_complex
qcr7	respiratori	succin
cox6	oxid	cytochrom_c_oxidas
rip1	reductas	malic_enzym
cox15	respir	mitochondri_outer_membran
cor1	cytochrom_bc	cytochrom_c_oxidas_biogenesi

sdh3	assembli	acetyl_coa_hydrolas
sdh4		
sdh2		
mdh1		
qcr6		
cyt1		
ach1		