Annotation consistency using annotation intersections

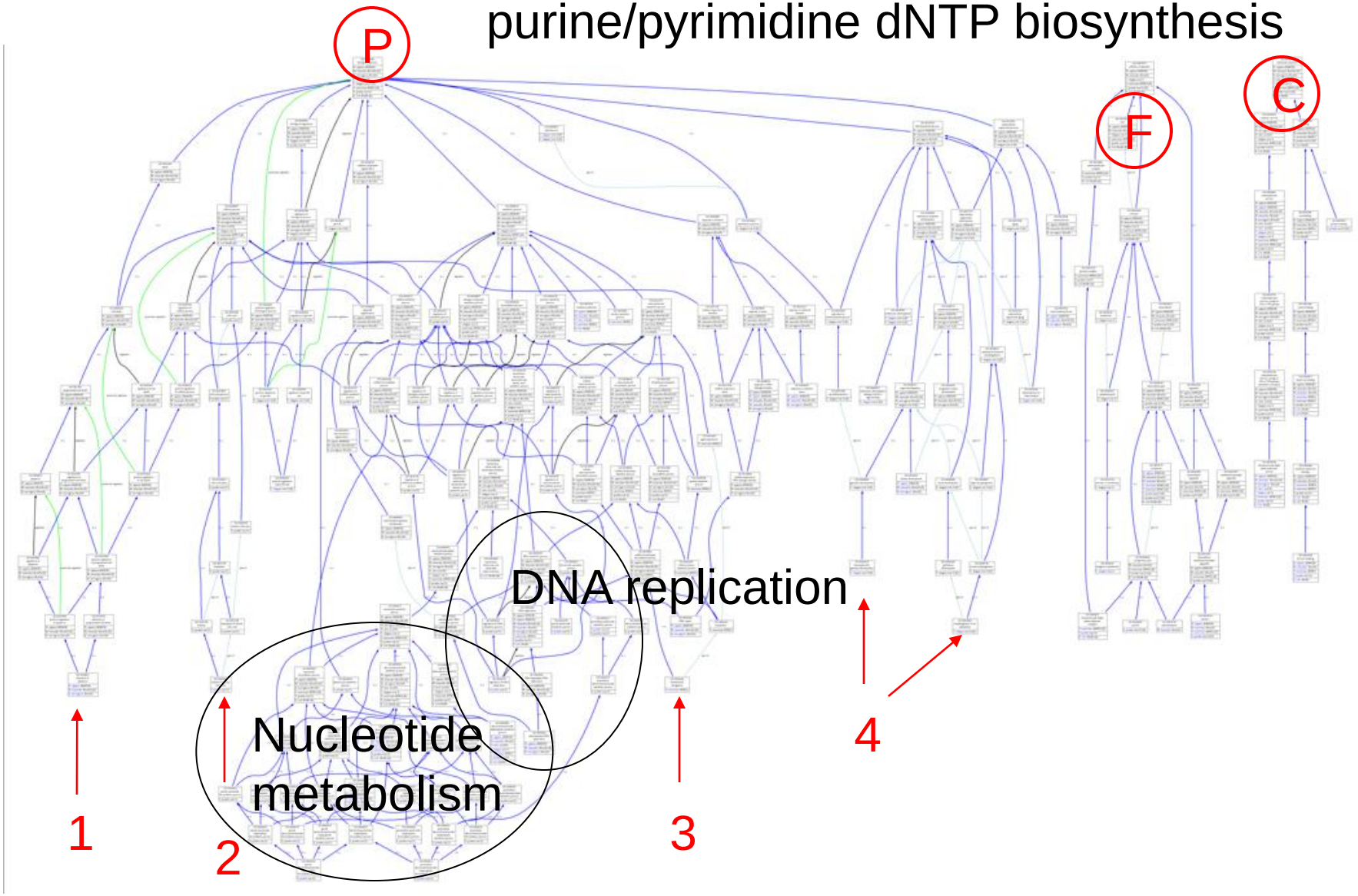
Valerie Wood

PomBase

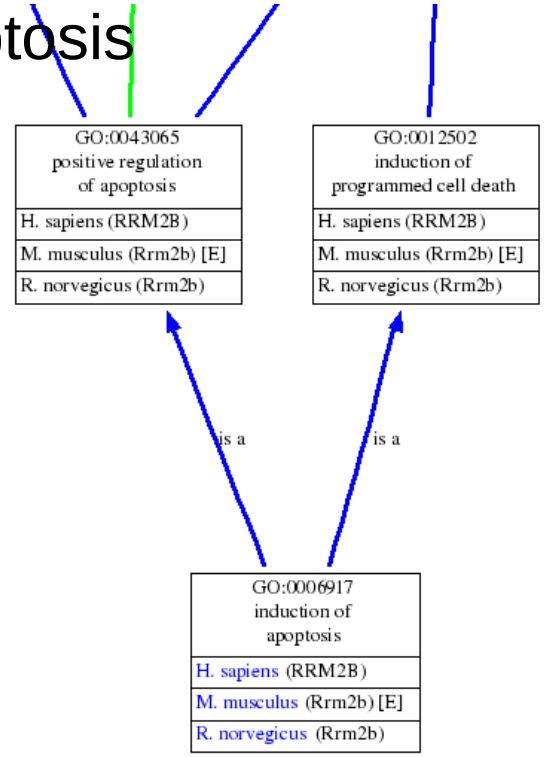
- Reference genome annotation in PAINT identifies annotation ‘outliers’ on a gene-by-gene basis
- Some ‘outliers’ are biologically valid

  (new biology, multiple functions)
- Others are, curator errors, mapping errors, ontology problems, experimental inconsistencies/legacy data

Ribonucleotide reductase rrm2b



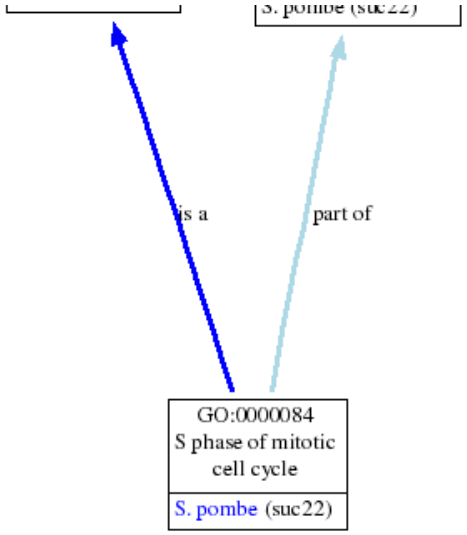
1. Apoptosis

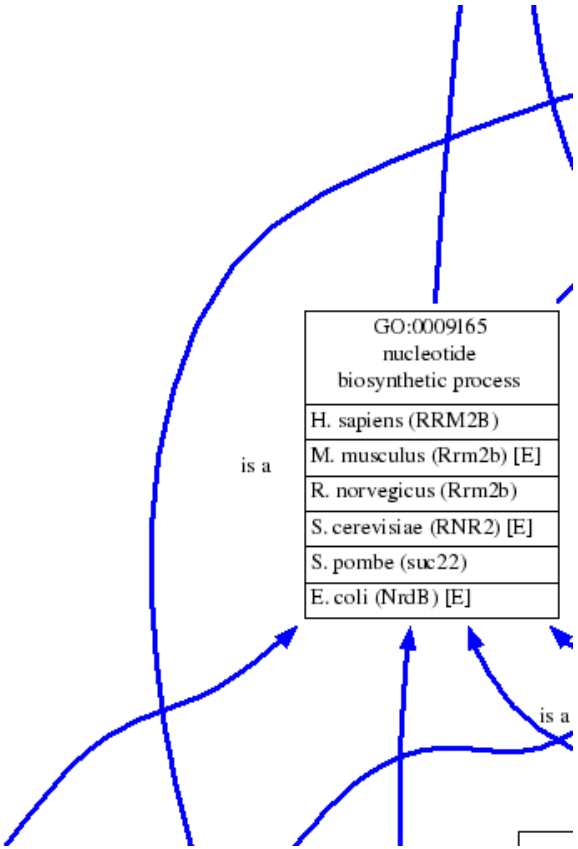


~~Apoptosis is downstream RNR, nucleotide depletion~~
positively “regulates apoptosis” via P53 pathway (valid annotation)

Other downstream effects in this paper (unnannotated) are renal failure, organ failure, growth retardation,early mortality, these are a consequence of apoptosis, but not
biological processes mechanistically controlled by this gene product

2. S-phase of mitotic cell cycle☐

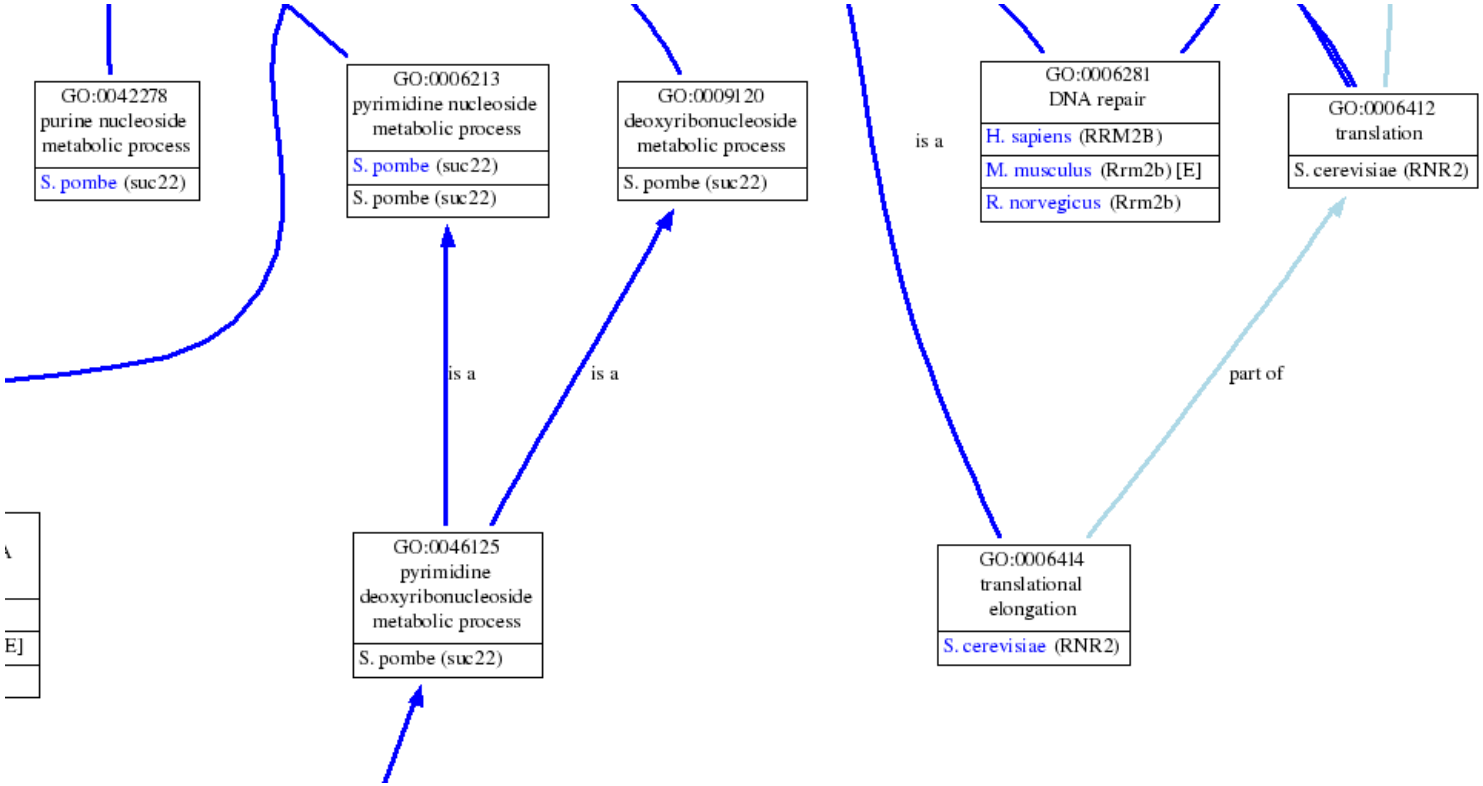




IC from replication

Recommendation, remove ?

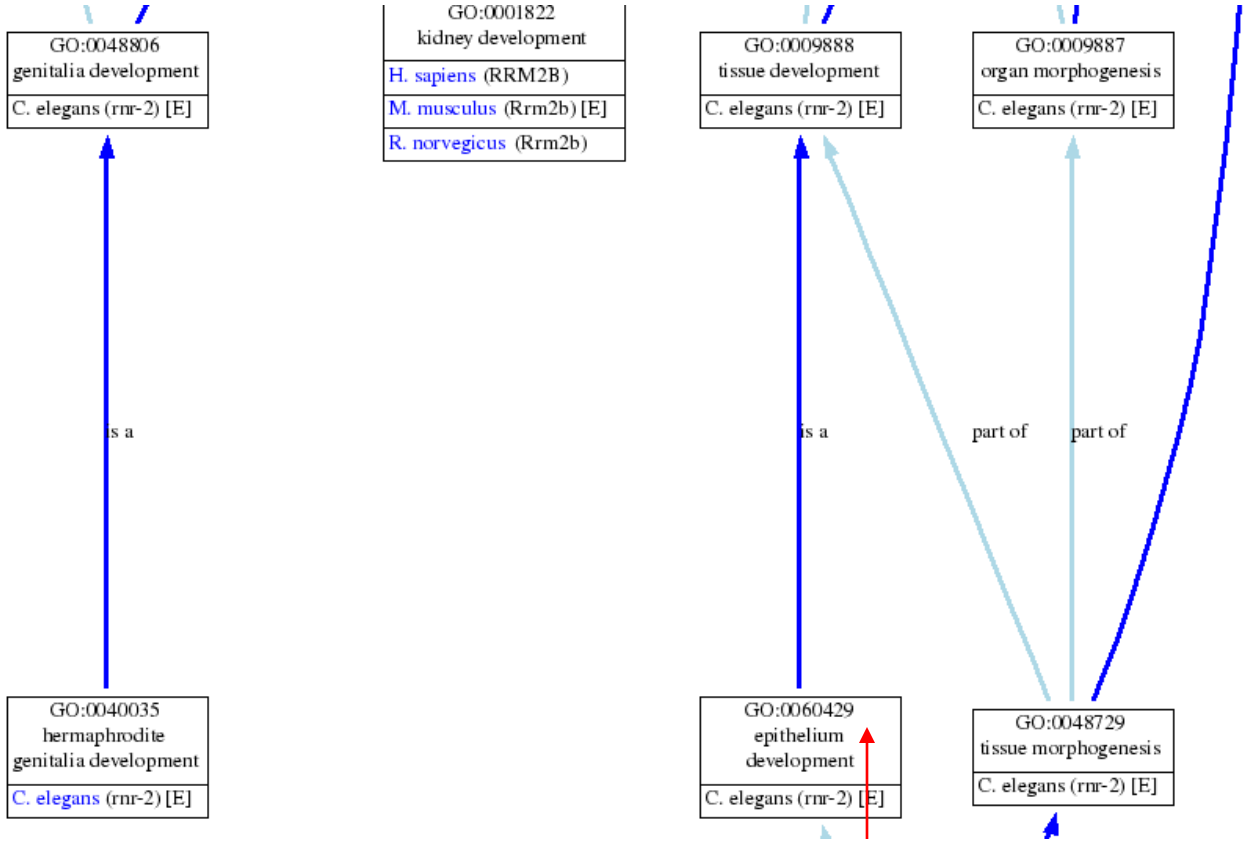
3. Translational elongation



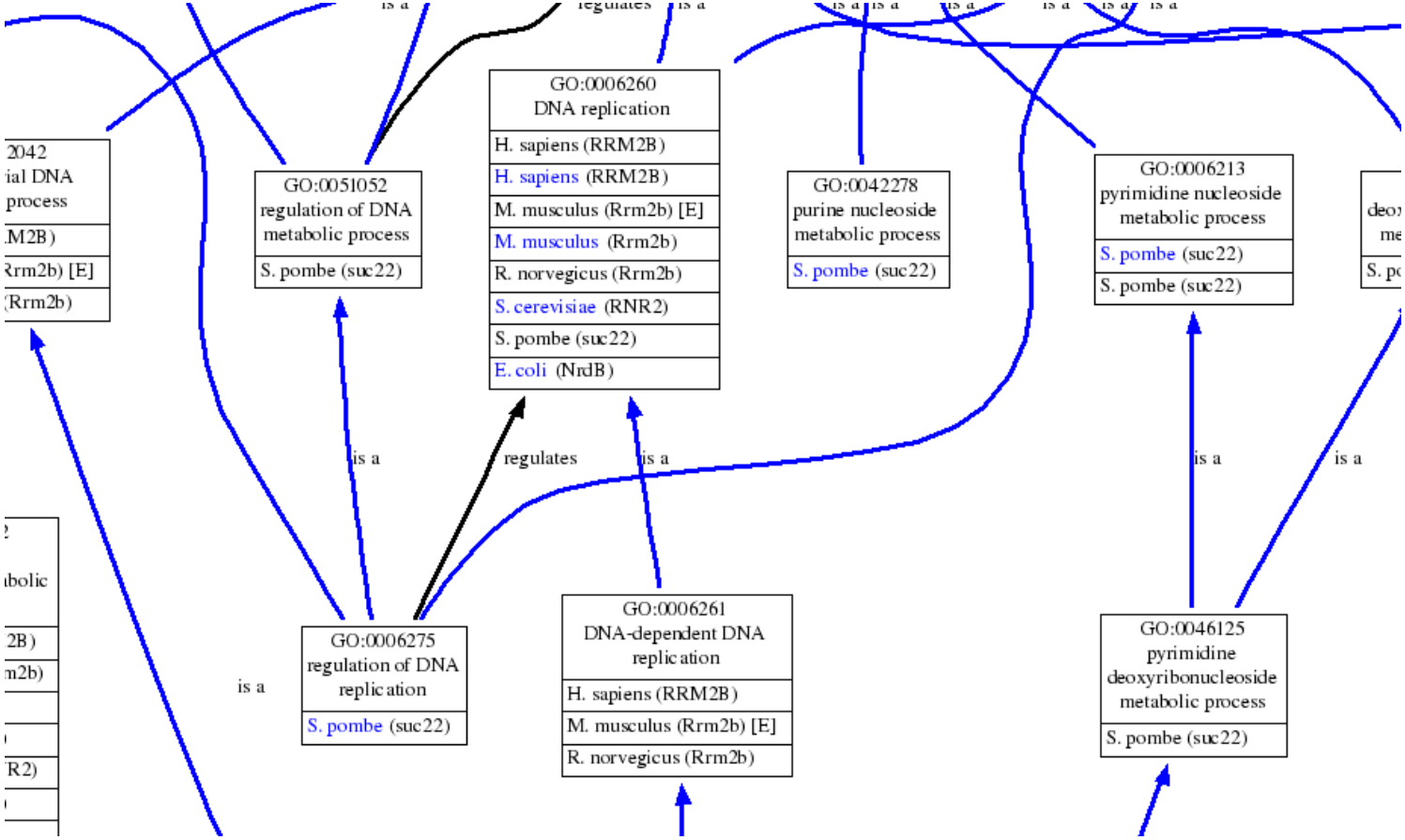
translational elongation	RCA: Reviewed Computational Analysis *Last updated 2009-08-06*	**Huttenhower C and Troyanskaya OG (2** integrating high-throughput datasets SGD Paper

Recommend mapping removed

4. Developmental processes☐



If these annotation are only a consequence of inability of cells to copy DNA and divide then probable the processes shouldn’t be annotated?



DNA replication is downstream (DSP) of nucleotide biosynthesis

Recommendation, “regulation of”?

Nucleic Acids Res. 2007;35(4):1187-97. Epub 2007 Jan 30.

Caf1 regulates translocation of ribonucleotide reductase by releasing nucleoplasmic Spd1-Suc22 assembly.

Takahashi S, Kontani K, Araki Y, Katada T.

Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, University of Tokyo, Tokyo 113-0033, Japan.

Abstract

Appropriate supply of deoxyribonucleotides by the ribonucleotide reductase (RNR) complex is essential for DNA replication and repair. One recent model for the RNR activation in Schizosaccharomyces pombe is translocation of the regulatory subunit Suc22 from the nucleoplasm to the cytoplasm. The RNR inhibitory protein Spd1, which retains Suc22 in the nucleoplasm, is rapidly degraded upon DNA-replication stress, resulting in release of Suc22 to form the active RNR complex in the cytoplasm. Here, we show that Caf1, a component of the Ccr4-Not complex, is responsible for resistance of the replication stress and control of the Suc22 translocation. Caf1 is required not only for the stress-induced

Cell cycle, DNA damage and heat shock regulate suc22+ expression in fission yeast.

Harris P, Kersey PJ, McInerny CJ, Fantes PA.

Institute of Cell and Molecular Biology, University of Edinburgh, UK.

Abstract

The suc22+ gene of Schizosaccharomyces pombe encodes the small subunit of ribonucleotide reductase. Two transcripts that hybridise to suc22+ have previously been described: a constitutive transcript of 1.5 kb, and a transcript of approximately 1.9 kb that is induced when DNA replication is blocked by hydroxyurea. In this paper we show that both transcripts derive from the suc22+ gene, are polyadenylated, and have transcription initiation sites separated by approximately 550 nucleotides. The absence of translation

Suggest that downstream process “regulation of DNA replication” *should* be annotated in this instance because there is a mechanism for the regulation of DNA replication involving these genes.

So ribonucleotide reductase is

Involved in nucleotide metabolic process (and children)
(specifically dNTP biosynthesis)

Regulates apoptosis (a downstream process, via a signaling pathway)

Regulates DNA replication (a downstream process, via a signaling pathway)

Has other downstream effects/phenotypes (organ morphogeneisis, tissue morphogenesis, impaired development, but via inability of cells to divide rather than an encoded regulatory mechanism)

- Time consuming way to check annotation consistency
- Unsustainable, changing ontology annotations and mappings
- Need a way to identify these issues globally
- Consistent rules based on known biology
- Flexible to include new biology

Annotation Intersections

- Identify slim term i.e nucleotide metabolic process
- What do all of the genes annotated to this term have in common with other processes?

You are using AmiGO Labs

Matrix View

Matrix Inputs

Axis 1:

```
GO:0005975
GO:0009117
GO:0006807
GO:0006790
GO:0032502
```

Axis 2:

```
GO:0009117
```

Axis 3 (optional):

Select species

S. pombe

Select relations

- (•) include all relations
- ( ) exclude regulates (waiting for database change...)

Submit Query

Fission yeast nucleotide met ∩ other GO slim processes

Results

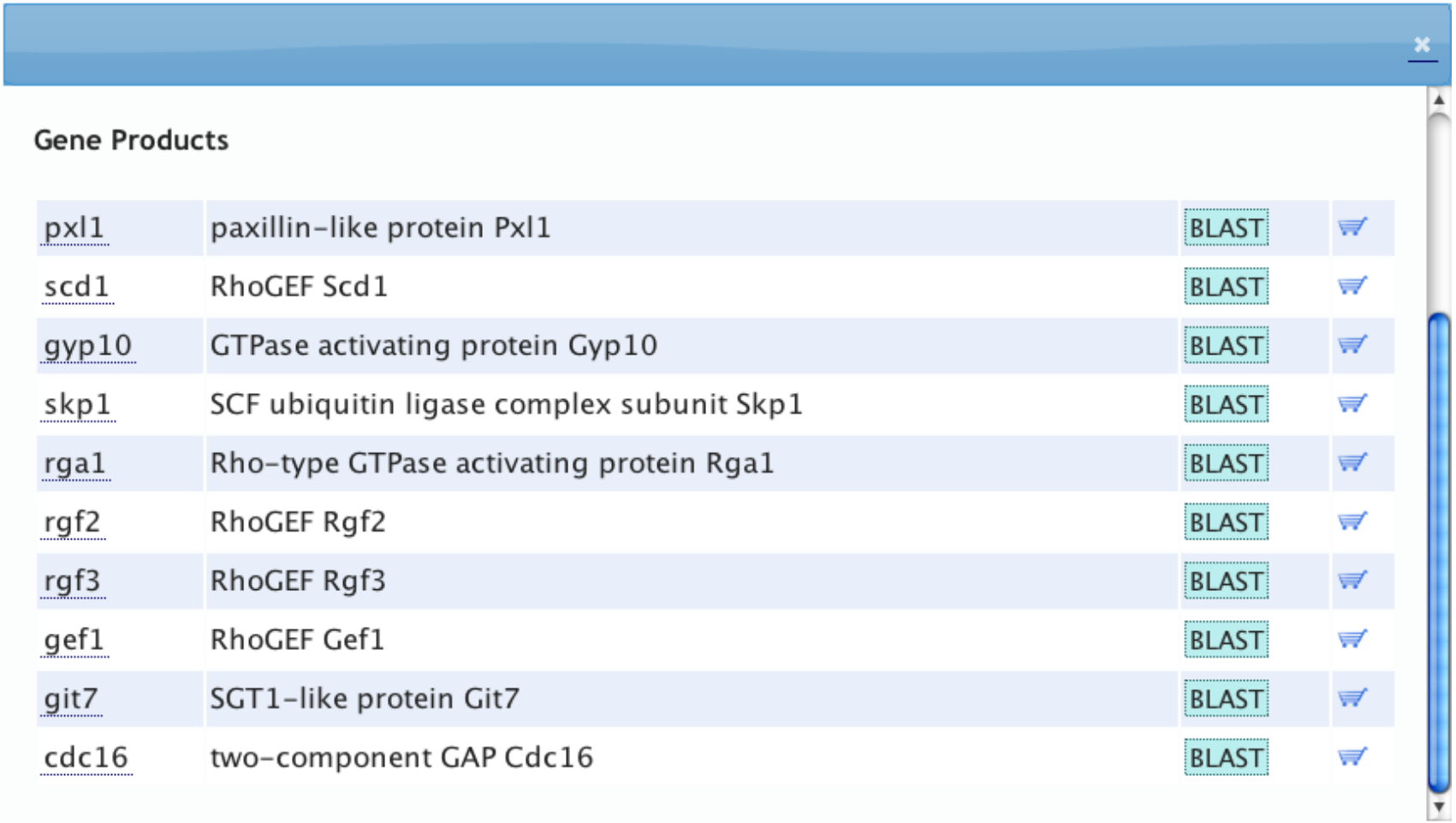
	nuclear mRNA splicing, via spliceosome GO:0000398	conjugation GO:0000746	cytokinesis GO:0000910	carbohydrate metabolic process GO:0005975	DNA metabolic process GO:0006259	DNA-dependent DNA replication GO:0006261	tRNA metabolic process GO:0006399	translation GO:0006412	protein folding GO:0006457	protein modification process GO:0006464	proteolysis GO:0006508	cellular amino acid metabolic process GO:0006520	lipid metabolic process GO:0006629	vitamin metabolic process GO:0006766	sulfur metabolic process GO:0006790	nitrogen compound metabolic process GO:0006807	nucleotide metabolic process GO:0009117
nucleotide metabolic process GO:0009117 GO:0009117	0	6	10	18	2	2	0	1	0	2	1	7	2	11	0	167	167

1 2 3 4

Tentatively, dependent on the intersections in other organisms, we may be able to say nucleotide metabolism does not overlap with splicing, tRNA metabolism, sulfur metabolism.

Can also look at the specific types of annotations in the intersects. How doe nucleotide metabolism intersect with these processes?

1. Fission yeast nucleotide met ∩ cytokinesis



Gene Products

pxl1	paxillin-like protein Pxl1	BLAST	
scd1	RhoGEF Scd1	BLAST	
gyp10	GTPase activating protein Gyp10	BLAST	
skp1	SCF ubiquitin ligase complex subunit Skp1	BLAST	
rga1	Rho-type GTPase activating protein Rga1	BLAST	
rgf2	RhoGEF Rgf2	BLAST	
rgf3	RhoGEF Rgf3	BLAST	
gef1	RhoGEF Gef1	BLAST	
git7	SGT1-like protein Git7	BLAST	
cdc16	two-component GAP Cdc16	BLAST	

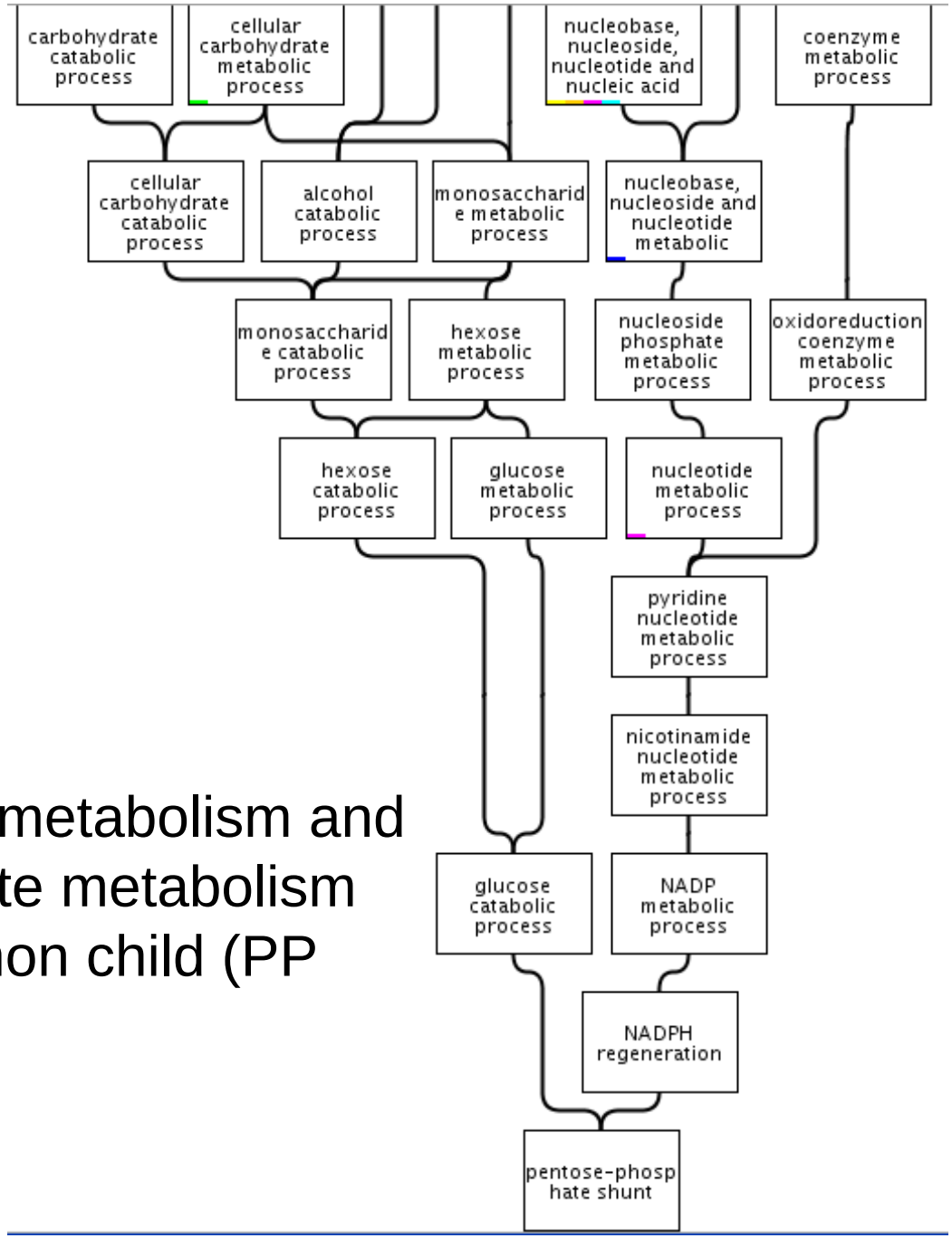
All signalling pathway upstream of cytokinesis
i.e. regulation of cytokinesis,
all are “GTP metabolism” a part of child of nucleotide metabolism

2. Fission yeast nucleotide met ∩ carbohydrate met

hxk1	hexokinase 1	PP-shunt□□
gpd2	glycerol-3-phosphate dehydrogenase Gpd2	Glycerol met /NADH oxidation
pgi1	glucose-6-phosphate isomerase	PP-shunt□□
adh1	alcohol dehydrogenase Adh1	
tal1	transaldolase	PP-shunt
zwf1	glucose-6-phosphate 1-dehydrogenase	PP-shunt□□
SPBC660.16	phosphogluconate dehydrogenase, decarboxylating	PP-shunt□□
gut2	glycerol-3-phosphate dehydrogenase Gut2	Glycerol met /NADH oxidation
SPCC794.01c	glucose-6-phosphate 1-dehydrogenase	PP-shunt□□
rga5	Rho-type GTPase activating protein Rga5	Upstream signalling
SPCC16C4.10	6-phosphogluconolactonase	PP-shunt□□
gpd1	glycerol-3-phosphate dehydrogenase Gpd1	Glycerol met /NADH oxidation
SPAC31G5.05c	ribulose phosphate 3-epimerase	PP-shunt
SPAC3C7.13c	glucose-6-phosphate 1-dehydrogenase	PP-shunt□□

glycerol met is a part of child of carbohydrate met
NADH oxidation is a part of child of nucleotide met

Nucleotide metabolism and carbohydrate metabolism have common child (PP shunt)



3. Fission yeast nucleotide met ∩ DNA metabolism

View gene products as an association list

Terms

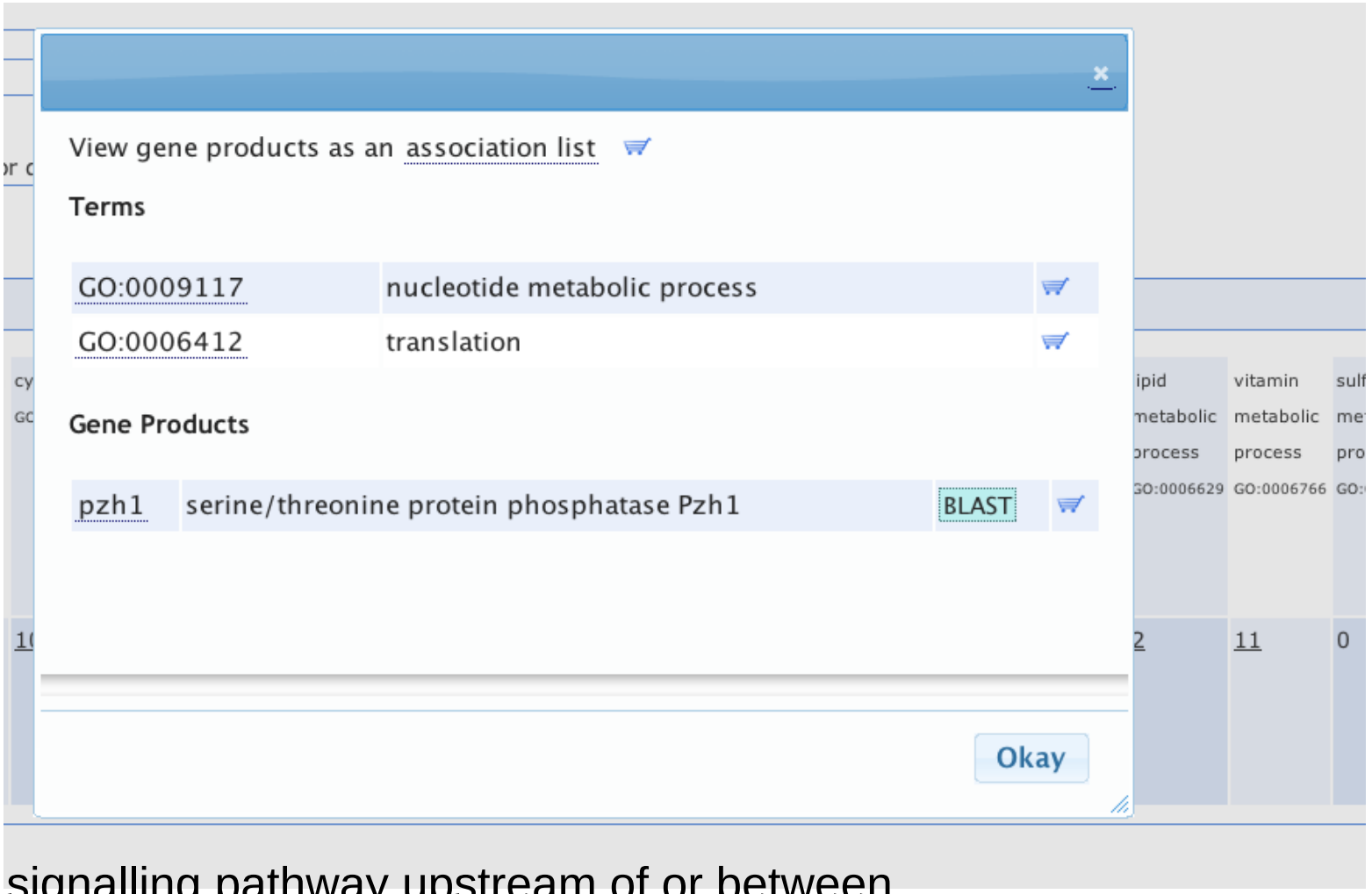
GO:0009117	nucleotide metabolic process
GO:0006259	DNA metabolic process

Gene Products

cdc22	ribonucleoside reductase large subunit Cdc22	BLAST
suc22	ribonucleotide reductase small subunit Suc22	BLAST

Known rule: Ribonucleotide reductase activity/complex is in the intersection between “nucleotide metabolism” and “regulation of DNA-dependent DNA replication”

4. Fission yeast nucleotide met ∩ translation



signalling pathway upstream of or between translation and nucleotide metabolism
(Q should these also be annotated to signal transduction?)

Rules Summary

1. Nucleotide met ∩ cytokinesis

Via common annotations to signalling processes(s) upstream of cytokinesis which are are a children of nucleotide metabolism (regulation of GTPase/GTP catabolic process)
In this case the signalling process is “upstream regulator of ” of cytokinesis but is a child of nucleotide met in the ontology

2 nucleotide met ∩ carbohydrate metabolism

Annotations to a child of both terms in the ontology (pentose phosphate shunt)
Annotations to a protein kinase which regulates is (upstream signalling pathway) for both processes
Annotations to child terms in both branches (glycogen metabolism and NADH regeneration)

3. Nucleotide met ∩ DNA metabolism

Via common annotations to
child of nucletide metabolism (dNTP biosynthesis)
child of DNA metabolism (regulation of DNA-dependent DNA-replication)
In this example dNTP biosynthesis *regulates* DNA replication

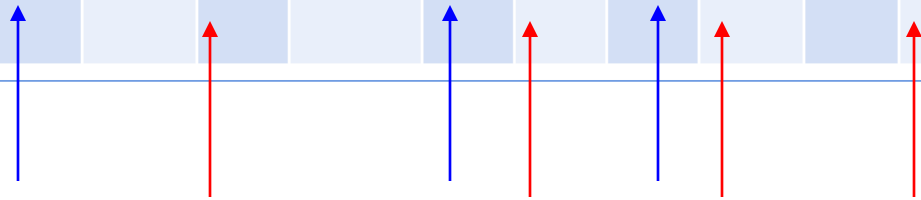
4. Nucleotide met ∩ translation

Via signalling processes upstream of nucleotide met and translation (protein kinase)

Budding yeast nucleotide met ∩ GO slim processes

Results

	nuclear mRNA splicing, via spliceosome GO:0000398	conjugation GO:0000746	cytokinesis GO:0000910	carbohydrate metabolic process GO:0005975	DNA metabolic process GO:0006259	DNA-dependent DNA replication GO:0006261	tRNA metabolic process GO:0006399	translation GO:0006412	protein folding GO:0006457	protein modification process GO:0006464	proteolysis GO:0006508	cellular amino acid metabolic process GO:0006520	lipid metabolic process GO:0006629	vitamin metabolic process GO:0006766	sulfur metabolic process GO:0006790	nitrogen compound metabolic process GO:0006807	nucleotide metabolic process GO:0009117
nucleotide metabolic process GO:0009117 GO:0009117	1	3	0	35	12	2	0	25	0	16	4	16	9	23	4	249	249



Zero intersections

Look more closely at some non zero intersections

Budding yeast nucleotide metabolism ∩ translation (25)

□

TPI1,PGI1,TKL1,NPP2,ALD6,SDH2,ADH2,5,PMA2,MIR1,
GYP6,ADE3,GND1,COX5A,AGE2,APA1,IMD2,3,4,URA2,7BNA4,GPD
2,RNR2,4

translational elongation	RCA: Reviewed Computational Analysis *Last updated 2009-08-06*	**Huttenhower C and Troyanskaya OG (2** integrating high-throughput datasets SGD Paper

Function prediction, from 2-Hybrid FP for highly expressed genes,

Recommend remove/suppress RCA to term

Budding yeast nucleotide metabolism ∩ DNA metabolism (12)

☐

TFP1	Subunit A of the eight-subunit V1 peripheral membrane domain of the vacuolar H+-ATPase	
RAD53	Protein kinase, required for cell-cycle arrest in response to DNA damage	
RNR4	Ribonucleotide-diphosphate reductase (RNR), small subunit	✓
TRX2	Cytoplasmic thioredoxin isoenzyme of the thioredoxin system which protects cells against oxidative and with Pbi2p, acts as a cofactor for Tsa1p, required for ER-Golgi transport and vacuole inheritance	✓
SML1	Ribonucleotide reductase inhibitor involved in regulating dNTP production	✓
RNR2	Ribonucleotide-diphosphate reductase (RNR), small subunit	✓
BNA2	Putative tryptophan 2,3-dioxygenase or indoleamine 2,3-dioxygenase, required for the de novo biosynt kynurenine	
RNR3	One of two large regulatory subunits of ribonucleotide-diphosphate reductase	✓
CDC21	Thymidylate synthase, required for de novo biosynthesis of pyrimidine deoxyribonucleotides	
YNK1	Nucleoside diphosphate kinase, catalyzes the transfer of gamma phosphates from nucleoside triphosph diphosphates by a mechanism that involves formation of an autophosphorylated enzyme intermediate	
TRX1	Cytoplasmic thioredoxin isoenzyme of the thioredoxin system which protects cells against oxidative and with Pbi2p, acts as a cofactor for Tsa1p, required for ER-Golgi transport and vacuole inheritance	✓
RNR1	One of two large regulatory subunits of ribonucleotide-diphosphate reductase	✓

TFP1 Subunit A of the eight-subunit V1 peripheral membrane domain of the vacuolar H+-ATPase

Biological Process

Manually curated

- cellular protein metabolic process (IDA)
- intron homing (TAS)
- vacuolar acidification (IMP)

intron homing	IEA: Inferred from Electronic Annotation with EBI:KW-0404 *Last updated 2010-04-20*	**GOA curators (2000)** Ge SGD Paper Access Full Text

Recommend replace TAS and delete KW mapping to intron homing

YNK1 Nucleoside diphosphate kinase, catalyzes the transfer of gamma phosphates from nucleoside triphosp diphosphates by a mechanism that involves formation of an autophosphorylated enzyme intermediate

DNA repair	IEA: Inferred from Electronic Annotation with EBI:KW-0234 *Last updated 2010-04-20*

Recommend delete KW mapping to DNA repair

CDC21 Thymidylate synthase, required for de novo biosynthesis of pyrimidine deoxyribonucleotides

DNA repair is a downstream process, Recommend “regulation of”, or removal”

BNA2 Putative tryptophan 2,3-dioxygenase or indoleamine 2,3-dioxygenase, required for the de novo biosy kynurenine

- de novo NAD biosynthetic process from tryptophan (IGI)
- telomere capping (IEP, IGI)

Is upregulated in response to “telomere capping” , Recommend removal”

TFP1	Subunit A of the eight-subunit V1 peripheral membrane domain of the vacuolar H+-ATPase
RAD53	Protein kinase, required for cell-cycle arrest in response to DNA damage
RNR4	Ribonucleotide-diphosphate reductase (RNR), small subunit
TRX2	Cytoplasmic thioredoxin isoenzyme of the thioredoxin system which protects cells against oxidative and with Pbi2p, acts as a cofactor for Tsa1p, required for ER-Golgi transport and vacuole inheritance
SML1	Ribonucleotide reductase inhibitor involved in regulating dNTP production
RNR2	Ribonucleotide-diphosphate reductase (RNR), small subunit
BNA2	Putative tryptophan 2,3-dioxygenase or indoleamine 2,3-dioxygenase, required for the de novo biosynt kynurenine
RNR3	One of two large regulatory subunits of ribonucleotide-diphosphate reductase
CDC21	Thymidylate synthase, required for de novo biosynthesis of pyrimidine deoxyribonucleotides
YNK1	Nucleoside diphosphate kinase, catalyzes the transfer of gamma phosphates from nucleoside triphosph diphosphates by a mechanism that involves formation of an autophosphorylated enzyme intermediate
TRX1	Cytoplasmic thioredoxin isoenzyme of the thioredoxin system which protects cells against oxidative and with Pbi2p, acts as a cofactor for Tsa1p, required for ER-Golgi transport and vacuole inheritance
RNR1	One of two large regulatory subunits of ribonucleotide-diphosphate reductase

Rules

ribonucleotide reductase activity/complex allowed

Thioredoxin system (TRX1/2) allowed

RAD53 protein kinase/upstream signaling?

Budding yeast nucleotide met ∩ protein modification

ATP1,PPA1,ADE1,SRP1,AAH1,HAM1,PRS4,GUK1,URA5

peptidyl-amino acid modification	RCA: Reviewed Computational Analysis *Last updated 2009-08-06*	**Huttenhower C and Troyanskaya OG (2009)** F integrating high-throughput datasets SGD Paper

Recommend remove/suppress RCA to term

SPF1,PMR1

In both cases that the glycosylation defect is a downstream effect of altered ionic concentrations in membrane organelles due to transporter defects. annotations to "protein amino acid glycosylation" were removed. Phenotype annotations were added.

Budding yeast nucleotide met ∩ GO slim processes

Results

	nuclear mRNA splicing, via spliceosome GO:0000398	conjugation GO:0000746	cytokinesis GO:0000910	carbohydrate metabolic process GO:0005975	DNA metabolic process GO:0006259	DNA-dependent DNA replication GO:0006261	tRNA metabolic process GO:0006399	translation GO:0006412	protein folding GO:0006457	protein modification process GO:0006464	proteolysis GO:0006508	cellular amino acid metabolic process GO:0006520	lipid metabolic process GO:0006629	vitamin metabolic process GO:0006766	sulfur metabolic process GO:0006790	nitrogen compound metabolic process GO:0006807	nucleotide metabolic process GO:0009117
nucleotide metabolic process GO:0009117 GO:0009117	1	3	0	35	12	2	0	25	0	16	4	16	9	23	4	249	249

8 0 0

Zero intersections

Look more closely at some non zero intersections

Revise rules for sc?

1=annotation OK 1=annotation questionable (see tables below) 1=regulation ∩ 1=downsteam process ∩	Spombe, replication	cerevisiae,replication	dicty, replication	worm, replication	fly, replication	rat, replication	mouse, replication	human, replication	weed, replication	e.coli, replication
DNA-dept. replication	127	98	34	22	58	55	36	73	49	0?
splicing	-	1	-	-	1	-	-	–	–	
conjugation	2?	-	-	-	-	-	-	–	–	
cytokinesis	-	-	-	7?	-	1?	1	–	2	
carbohydrate met	-	3	-	-	-	-	-	–	–	
DNA met	127	98	34	22	58	55	36	73	49	
tRNA met	-	-	-	-	1	-	-	1?	–	
translation	2?	2?	-	-	3	-	-	2?	–	
protein folding	-	1	-	-	-	1	1	1	–	
protein modification	19	19	-	-	1	2	1	6	–	
proteolysis	5	5	-	-	2	-	-	–	–	
amino acid met	-	1	-	-	1~~?~~	-	-	1?	–	
lipid met	-	1	-	-	-	-	-	–	1?	
vitamin met	-	-	-	-	-	-	-	–	–	
sulphur met	-	1	-	-	-	-	-	–	–	
nitrogen compound met	127	98	34	22	58	55	36	73	49	
mitochondrion org	4	4	-	2	5	8	6	8	–	
cell cycle	68	58	5	7	28	17	12	30	31	
chromosome seg.	23	19	-	1	6	6	2	6	1	
meiosis	13	16	-	2	3	6	2	10	1	
cell polarity	-	-	-	-	-	-	-	–	–	
signal transduction	39	12	4	1	3	5	4	13	3	
nucleotide met	-	2	1	-	-	2?	2?	3?	–	
mRNA met	2-1	3	-	-	1~~?~~	-	-	1?	–	
actin cytoskeleton org	-	-	-	-	1	-	-	–	–	
developmental process	7?	10?	-	16?	14?	10?	6?	14?	21?	
ribosome biogenesis	-	2	-	-	-	-	-	–	–	
cofactor met.	-	1	-	4?	-	-	-	–	–	

Valid intersections are:

1. Annotation to a term which is a child of both terms

   I.e pentose phosphate shunt is a child of nucleotide metabolism and carbohydrate metabolism. tRNA acetylation is a child or tRNA metabolism and amino acid metabolism

2. Upstream or downstream effects

- Upstream signalling pathway regulating both processes (I.e a protein kinase regulating translation and nucleotide metabolism)
- One process acts upstream of another (i.e nucleotide metabolism upstream regulates DNA replication, apoptosis

3. Multifunctional gene products, for example NOC3 appears to function in both pathways of replication and rRNA processing (note this could be an example of 2...)

Note: It is possible that for 2 one or both annotations should always be “regulation of”

Invalid intersections are:

1. Annotation errors
2. Experimental errors or legacy experiemnts
3. Downstream effects which are not “mechanistically linked” to the process but are downstream consequences of when thing go wrong.

Advantages

- Provides a set of annotation rules for curators (annotation accuracy and consistency). Can identify rapidly annotation difference between organisms.
- Provides a check for curators to consider whether the annotation should be “regulation of” , or whether a downstream annotation (embryonic development) is a known consequence of a role in replication). Makes the scope of “regulation” more consistent
- Allows rapid detection of new annotation errors (from IEA mappings and RCA predictions)
- Improve the ontology, identifying missing parents, makes it easy to identfy changes which affect high level terms
- Can be used as quality checks for new annotation datasets and function predictions

Future

Establish rules for all high level processes

Establish more specific rules, taxon specific rules, function and component rules

Ultimately will be able to assess whether annotations are correct in the context of known biology, or whether they identify new previously unknown connections between divergent processes

Removal of experimental annotations, need more rigorous alerting for unsupported ISS annotations.

High level terms where annotations can consistently be transferred could be identified , I.e transcription, translation, replication, x metabolism (improve GO slims, easier to identify total “unknowns”)

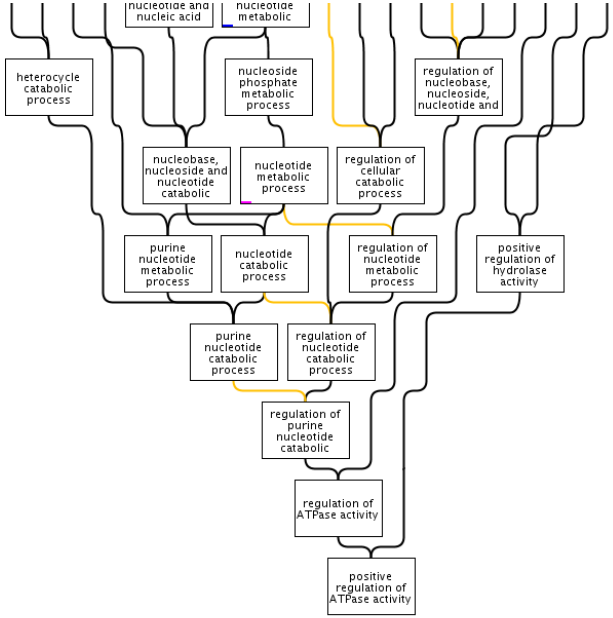
Spare slides

Budding yeast nucleotide met ∩ splicing

Biological Process

Manually curated

- maturation of SSU-rRNA (IGI)
- nuclear mRNA splicing, via spliceosome (IGI)
- positive regulation of ATPase activity (IDA, IMP)
- positive regulation of helicase activity (IDA, IMP)
- rRNA processing (IGI)



ickGO - http://www.ebi.ac.uk/QuickGO