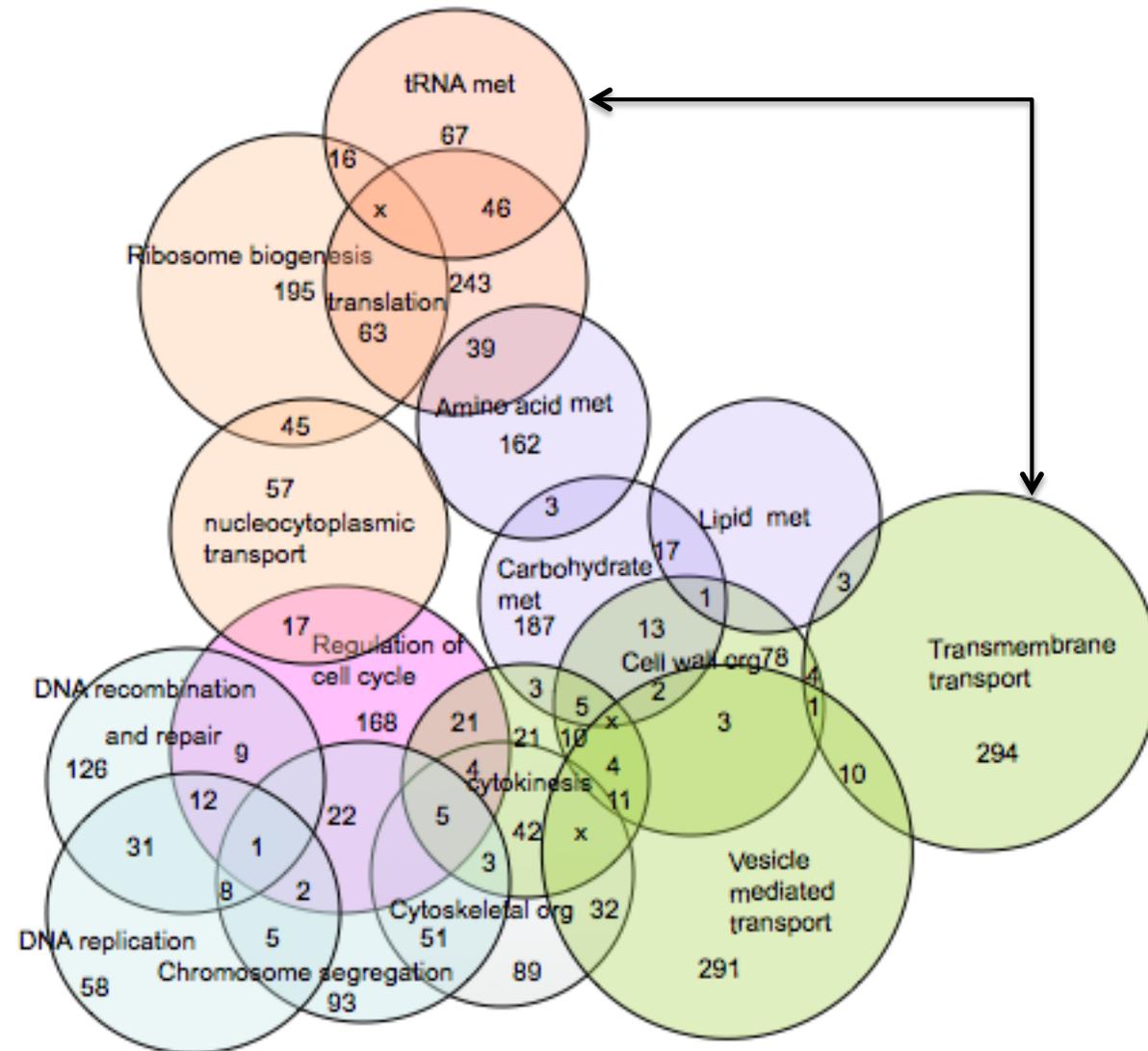


# **Using biological process co- annotation for ontology and annotation quality control**



Biological processes are accomplished by the coordinated action of sets of gene products. However, some processes are rarely connected to each other because they functionally, temporally or spatially distant.

## Matrix View

### Matrix Inputs

Axis 1:

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

Axis 2:

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

Axis 3 (optional):

Select species

S. pombe

Select relations

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

Submit Query

You are using an **AmiGO Labs** prototype. See [here](#) for more information.

# Matrix

Information about Matrix 

## Input GO IDS

GO:0043473 GO:0009987 GO:0022008

<http://amigo2.berkeleybop.org/cgi-bin/amigo2/matrix>

(caveat, only pombe filter currently available)

## Taxon

- All
- pombe

Selection

# Step 1

	DNA repair GO:0006281	tRNA metabolic process GO:0006399	translation GO:0006412	cellular amino acid metabolic process GO:0006520	vitamin metabolic process GO:0006766	ribosome biogenesis GO:0042254	transmembrane transport GO:0055085	nitrogen cycle metabolic process GO:0071941
DNA repair GO:0006281	197	0	4	1	0	4	0	0
tRNA metabolic process GO:0006399	0	154	46	37	1	32	0	0
translation GO:0006412	4	46	631	41	0	67	1	0
cellular amino acid metabolic process GO:0006520	1	37	41	234	14	0	0	11
vitamin metabolic process GO:0006766	0	1	0	14	41	0	0	0
ribosome biogenesis GO:0042254	4	32	67	0	0	327	0	0
transmembrane transport GO:0055085	0	0	1	0	0	0	304	1
nitrogen cycle metabolic process GO:0071941	0	0	0	11	0	0	1	17

**Step 1** Co-annotated terms (annotation intersections) were identified for all pairs of "high level process" (slim) GO terms  
 Can see that many intersections are zero (cerise) and some are low (red)

# Step 2

## **Step 2 :**

**Create rules for “zero” intersects  
(based on known biology)**

**GO:0006520,GO:0006310, NO OVERLAP**

**GO:0006520,GO:0006260, NO OVERLAP**

**GO:0006520,GO:0030437, NO OVERLAP**



"x is not usually co-annotated with y".

# Step 3

GO:0006520,GO:0006310, NO OVERLAP  
GO:0006520,GO:0006260, NO OVERLAP  
GO:0006520,GO:0030437, NO OVERLAP  
**GO:0006520,GO:0006281, INTERSECT HAS GO:0006521 OR GO:0006338**  
**GO:0006520,GO:0007155, INTERSECT HAS GO:0006521**  
**GO:0006520,GO:0070882, INTERSECT HAS GO:0004360 OR GO:0004067**

Intersections where annotations were sparse were inspected for spurious annotations, which were either corrected, or, if the annotations were validated the rules were extended to allow these exceptions.

For example the intersect between GO:0006520 (amino acid metabolic process) and GO:0006399 (tRNA metabolic process) includes gene products annotated to GO:0043039 (tRNA aminoacylation) OR GO:0031071 (cysteine desulphurase), which are involved in both processes

More examples here

:<https://sourceforge.net/apps/trac/pombase/wiki/MatrixProject>

# Step 4

	DNA repair GO:0006281	tRNA metabolic process GO:0006399	translation GO:0006412	cellular amino acid metabolic process GO:0006520	vitamin metabolic process GO:0006766	ribosome biogenesis GO:0042254	transmembrane transport GO:0055085	nitrogen cycle metabolic process GO:0071941
DNA repair GO:0006281 GO:0006281	<u>197</u>	0	<u>4</u>	<u>1</u>	0	<u>4</u>	0	0
tRNA metabolic process GO:0006399 GO:0006399	0	<u>154</u>	<u>46</u>	<u>37</u>	<u>1</u>	<u>32</u>	0	0
translation GO:0006412 GO:0006412	<u>4</u>	<u>46</u>	<u>631</u>	<u>41</u>	0	<u>67</u>	<u>1</u>	0
cellular amino acid metabolic process GO:0006520 GO:0006520	<u>1</u>	<u>37</u>	<u>41</u>	<u>234</u>	<u>14</u>	0	0	<u>11</u>
vitamin metabolic process GO:0006766 GO:0006766	0	<u>1</u>	0	<u>14</u>	<u>41</u>	0	0	0
ribosome biogenesis GO:0042254 GO:0042254	<u>4</u>	<u>32</u>	<u>67</u>	0	0	<u>327</u>	0	0
transmembrane transport GO:0055085 GO:0055085	0	0	<u>1</u>	0	0	0	<u>304</u>	<u>1</u>
nitrogen cycle metabolic process GO:0071941 GO:0071941	0	0	0	<u>11</u>	0	0	<u>1</u>	<u>17</u>

Identify annotations which break rules.

### Step 1 :

Identify intersections between pairs of GO terms, using taxonomically diverse set of model species

GO term	Species 1	Species 2	Intersection	Species 1	Species 2	Intersection	Species 1	Species 2	Intersection
GO:0006520	1	1	1	1	1	1	1	1	1
GO:0006310	1	1	1	1	1	1	1	1	1
GO:0006260	1	1	1	1	1	1	1	1	1
GO:0030437	1	1	1	1	1	1	1	1	1



### Step 2 :

Create rules for "zero" intersects (based on known biology)

GO:0006520,GO:0006310, NO OVERLAP  
 GO:0006520,GO:0006260, NO OVERLAP  
 GO:0006520,GO:0030437, NO OVERLAP



### Step 3 :

Query annotation outliers and Fix annotations in contributing MODS

Gene	GO term	Annotation
XXXX		

# Add eggs

### Step 4 :

Periodically identify annotations which break rules across annotation corpus

GO term	Species 1	Species 2	Intersection	Species 1	Species 2	Intersection	Species 1	Species 2	Intersection
GO:0006520	1	1	1	1	1	1	1	1	1
GO:0006310	1	1	1	1	1	1	1	1	1
GO:0006260	1	1	1	1	1	1	1	1	1
GO:0030437	1	1	1	1	1	1	1	1	1

# ANNOTATION QC CYCLE

OR extend rules by allowing specific intersects

GO:0006520,GO:0006310, NO OVERLAP  
 GO:0006520,GO:0006260, NO OVERLAP  
 GO:0006520,GO:0030437, NO OVERLAP  
 GO:0006520,GO:0006281, INTERSECT HAS GO:0006521 OR GO:0006338  
 GO:0006520,GO:0007155, INTERSECT HAS GO:0006521  
 GO:0006520,GO:0070882, INTERSECT HAS GO:0004360 OR GO:0004067



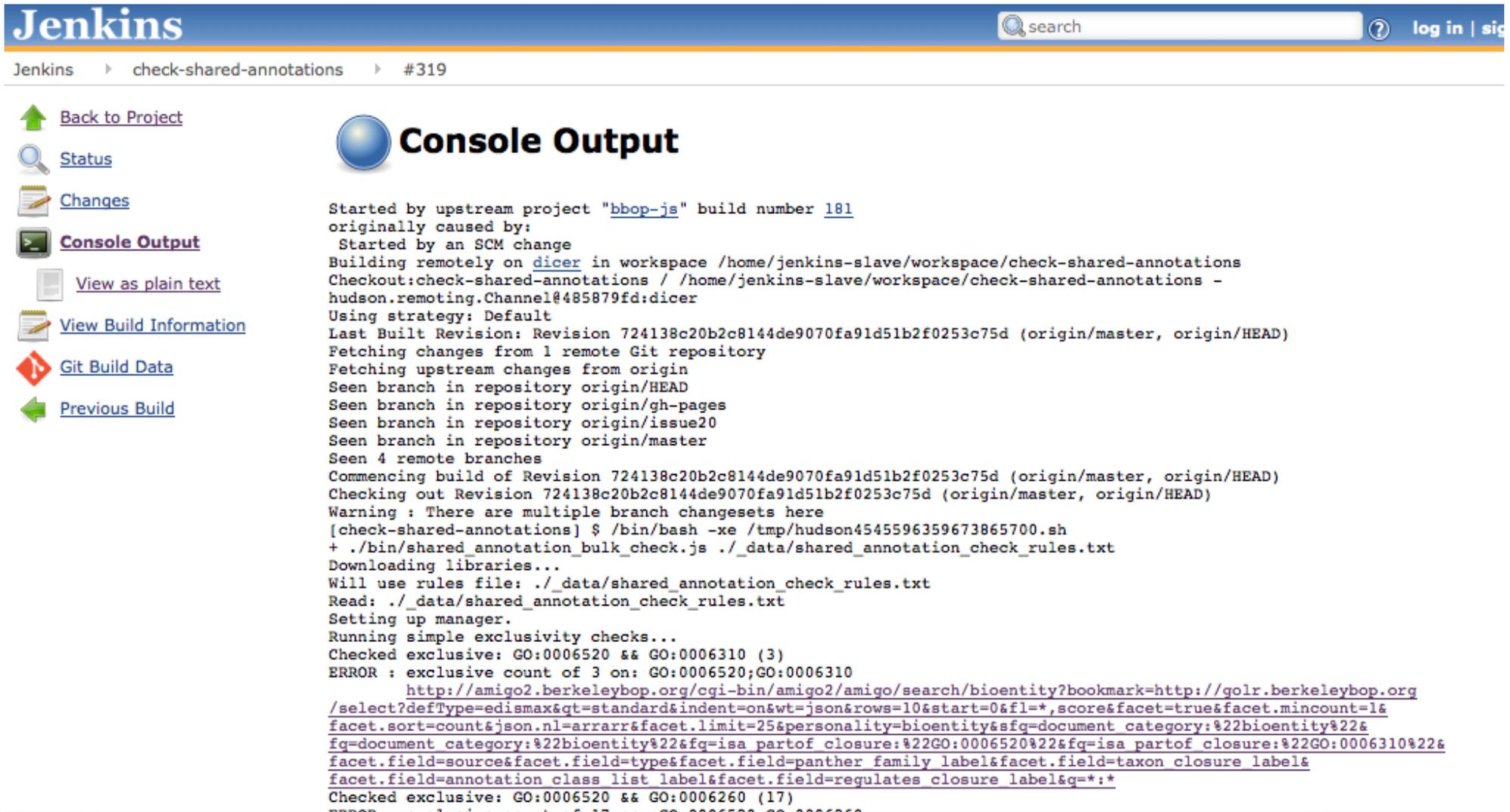
Fix annotations....  
 Extend rules....  
 Repeat....

# Rules so far

for intersections with:

- GO:0006520 cellular amino acid metabolic process
- GO:0006399 tRNA metabolic process
- GO:0042254 ribosome biogenesis
- GO:0006766 vitamin metabolic process
- GO:0007346 regulation of mitotic cell cycle
- GO:0006281 DNA repair
- GO:0006260 DNA replication

http://build.berkeleybop.org/job/check-shared-annotations/lastBuild/console



The screenshot shows the Jenkins web interface. At the top, the Jenkins logo is on the left, and a search bar and 'log in | sig' link are on the right. Below the header, the breadcrumb 'Jenkins > check-shared-annotations > #319' is visible. On the left sidebar, there are navigation links: 'Back to Project', 'Status', 'Changes', 'Console Output' (highlighted), 'View as plain text', 'View Build Information', 'Git Build Data', and 'Previous Build'. The main content area is titled 'Console Output' and contains the following text:

```
Started by upstream project "bbop-js" build number 181
originally caused by:
  Started by an SCM change
Building remotely on dicer in workspace /home/jenkins-slave/workspace/check-shared-annotations
Checkout:check-shared-annotations / /home/jenkins-slave/workspace/check-shared-annotations -
hudson.remoting.Channel@485879fd:dicer
Using strategy: Default
Last Built Revision: Revision 724138c20b2c8144de9070fa91d51b2f0253c75d (origin/master, origin/HEAD)
Fetching changes from 1 remote Git repository
Fetching upstream changes from origin
Seen branch in repository origin/HEAD
Seen branch in repository origin/gh-pages
Seen branch in repository origin/issue20
Seen branch in repository origin/master
Seen 4 remote branches
Commencing build of Revision 724138c20b2c8144de9070fa91d51b2f0253c75d (origin/master, origin/HEAD)
Checking out Revision 724138c20b2c8144de9070fa91d51b2f0253c75d (origin/master, origin/HEAD)
Warning : There are multiple branch changesets here
[check-shared-annotations] $ /bin/bash -xe /tmp/hudson4545596359673865700.sh
+ ./bin/shared_annotation_bulk_check.js ./_data/shared_annotation_check_rules.txt
Downloading libraries...
Will use rules file: ./_data/shared_annotation_check_rules.txt
Read: ./_data/shared_annotation_check_rules.txt
Setting up manager.
Running simple exclusivity checks...
Checked exclusive: GO:0006520 && GO:0006310 (3)
ERROR : exclusive count of 3 on: GO:0006520;GO:0006310
http://amigo2.berkeleybop.org/cgi-bin/amigo2/amigo/search/bioentity?bookmark=http://golr.berkeleybop.org/select?defType=edismax&qt=standard&indent=on&wt=json&rows=10&start=0&fl=\*,score&facet=true&facet.mincount=1&facet.sort=count&json.nl=arrarr&facet.limit=25&personality=bioentity&sfg=document\_category:%22bioentity%22&fq=document\_category:%22bioentity%22&fq=isa partof closure:%22GO:0006520%22&fq=isa partof closure:%22GO:0006310%22&facet.field=source&facet.field=type&facet.field=panther\_family\_label&facet.field=taxon\_closure\_label&facet.field=annotation\_class\_list\_label&facet.field=regulates\_closure\_label&q=\*\*
Checked exclusive: GO:0006520 && GO:0006260 (17)
```

http://build.berkeleybop.org/job/check-shared-annotations/lastBuild/console

```
facet.field=annotation class list label&facet.field=regulates closure label&q=**
Checked exclusion: GO:0042254 && GO:0006091 && !(GO:0023052) (108)
ERROR : bad co-annotations for: GO:0042254;GO:0006091;GO:0023052
http://amigo2.berkeleybop.org/cgi-bin/amigo2/amigo/search/bioentity?bookmark=http://golr.berkeleybop.org
/select?defType=edismax&qt=standard&indent=on&wt=json&rows=10&start=0&fl=*,score&facet=true&facet.mincount=1&facet.sort=count&
json.nl=arrarr&facet.limit=25&personality=bioentity&sfg=document category:%22bioentity%22&fg=document category:%22bioentity%22&
fg=isa partof closure:%22GO:0042254%22&fg=isa partof closure:%22GO:0006091%22&fg=-isa partof closure:%22GO:0023052%22&
facet.field=source&facet.field=type&facet.field=panther family label&facet.field=taxon closure label&
facet.field=annotation class list label&facet.field=regulates closure label&q=**
Checked exclusion: GO:0042254 && GO:0007126 && !(GO:0023052) (31)
ERROR : bad co-annotations for: GO:0042254;GO:0007126;GO:0023052
http://amigo2.berkeleybop.org/cgi-bin/amigo2/amigo/search/bioentity?bookmark=http://golr.berkeleybop.org
/select?defType=edismax&qt=standard&indent=on&wt=json&rows=10&start=0&fl=*,score&facet=true&facet.mincount=1&facet.sort=count&
json.nl=arrarr&facet.limit=25&personality=bioentity&sfg=document category:%22bioentity%22&fg=document category:%22bioentity%22&
fg=isa partof closure:%22GO:0042254%22&fg=isa partof closure:%22GO:0007126%22&fg=-isa partof closure:%22GO:0023052%22&
facet.field=source&facet.field=type&facet.field=panther family label&facet.field=taxon closure label&
facet.field=annotation class list label&facet.field=regulates closure label&q=**
Checked exclusion; trivially passed with no base overlap: GO:0055085 && GO:0000910 (0)
Checked exclusion: GO:0006766 && GO:0030163 && !(GO:0070647) (8)
ERROR : bad co-annotations for: GO:0006766;GO:0030163;GO:0070647
http://amigo2.berkeleybop.org/cgi-bin/amigo2/amigo/search/bioentity?bookmark=http://golr.berkeleybop.org
/select?defType=edismax&qt=standard&indent=on&wt=json&rows=10&start=0&fl=*,score&facet=true&facet.mincount=1&facet.sort=count&
json.nl=arrarr&facet.limit=25&personality=bioentity&sfg=document category:%22bioentity%22&fg=document category:%22bioentity%22&
fg=isa partof closure:%22GO:0006766%22&fg=isa partof closure:%22GO:0030163%22&fg=-isa partof closure:%22GO:0070647%22&
facet.field=source&facet.field=type&facet.field=panther family label&facet.field=taxon closure label&
facet.field=annotation class list label&facet.field=regulates closure label&q=**
Checked exclusion; trivially passed with no base overlap: GO:0006766 && GO:0070647 (0)
Checked exclusion: GO:0006281 && GO:0006913 && !(GO:0031080) (79)
ERROR : bad co-annotations for: GO:0006281;GO:0006913;GO:0031080
http://amigo2.berkeleybop.org/cgi-bin/amigo2/amigo/search/bioentity?bookmark=http://golr.berkeleybop.org
/select?defType=edismax&qt=standard&indent=on&wt=json&rows=10&start=0&fl=*,score&facet=true&facet.mincount=1&facet.sort=count&
json.nl=arrarr&facet.limit=25&personality=bioentity&sfg=document category:%22bioentity%22&fg=document category:%22bioentity%22&
fg=isa partof closure:%22GO:0006281%22&fg=isa partof closure:%22GO:0006913%22&fg=-isa partof closure:%22GO:0031080%22&
facet.field=source&facet.field=type&facet.field=panther family label&facet.field=taxon closure label&
facet.field=annotation class list label&facet.field=regulates closure label&q=**
Looked at 194 rules.
Completed with 145 broken rule(s).
Finished: SUCCESS
```



Next step?

	DNA recombination (GO:0006310)	DNA repair (GO:0006281)	DNA replication (GO:0006260)	ascospore formation (GO:0030437)	carbohydrate metabolic process (GO:0006975)	cell adhesion (GO:0007155)	cellular amino acid metabolic process (GO:0001668)	cell wall organization or biogenesis (GO:007155)	chromatin modification (GO:0016688)	chromatin organization (GO:0051276)	chromosome segregation (GO:0007059)	cofactor metabolic process (GO:0051186)	conjugation with cellular fusion (GO:0000747)	cytokinesis (GO:0000910)	cytoplasmic translation (GO:0002181)	cytoskeleton organization (GO:0007010)	establishment or maintenance of cell polarity (GO:0006629)	generation of precursor metabolites and energy	lipid metabolic process (GO:0006629)	mRNA metabolic process (GO:0016071)	meiotic nuclear division (GO:0007126)	mitochondrion organization (GO:0007005)	nitrogen cycle metabolic process (GO:0071941)	nucleobase-containing small molecule metabolism	nucleocytoplasmic transport (GO:0006913)	peroxisome organization (GO:0007091)	protein catabolic process (GO:0030163)	protein complex assembly (GO:0006461)	protein folding (GO:0006457)	protein glycosylation (GO:0006486)	protein maturation (GO:0051604)	protein modification by small protein conjugation	protein targeting (GO:0006605)	regulation of mitotic cell cycle (GO:0007346)	ribosome biogenesis (GO:0042254)	signaling (GO:0023052)	tRNA metabolic process (GO:0006399)	transcription, DNA-templated (GO:0006351)	transmembrane transport (GO:0055085)	vacuole organization (GO:0007033)	vesicle-mediated transport (GO:0016192)	vitamin metabolic process (GO:0006766)								
(GO:0006310)	112	63	51	3					5	42	20																																							
(GO:0006281)	63	205	57						44	85	15		2	2	1					3	54	7	25	1																										
(GO:0006260)	51	57	133						5	36	10																																							
(GO:0030437)	3				94	12		19	1	7	8	1	8	6	1	14	4	1	5		94																													
(GO:0005975)				12	223		6	69				20	2	8		1	31	28																																
(GO:0007155)						14						3																																						
(GO:0006520)				6			238	1	1	1	32				23			16	4	1	20	11	21					2		3											1		37				15			
(GO:0071554)				19	69		1	117	1	1		1	4	17		10	7		6		20			9				3	1	2	21	6	1	2	3			18			3	1	10							
(GO:0016688)	5	44	5	1			1	1	197	197	34	1		2		6	2			2	8			19	1		5	21	1	5	14		15	1	5	1	36									1				
(GO:0051276)	42	85	36	7			1	1	197	396	147				1	3	10	58	5		3	65	1	42	5		13	54	1	16	24	2	53	3	12	1	47									1				
(GO:0007059)	20	15	10	8			1		34	147	166				1	9	58	4			1	56		15	3		6	28		7	7	31	1	8		9											1			
(GO:0051186)				1	20		32	1	1	1	1	139						27	17			1		56				2																				27		
(GO:0000747)	8	2	3				4	3	1			65	4		7	1				2	1	10		7			6	1	1	9	3	5	1	27			2	4	2	2										
(GO:0000910)				6	8		17	2	10	9		4	101		47	14					10		8			3	9	2	1	2	1	18	25			1	1	16												
(GO:0002181)		2					23								445					4	1	2		17	7		2	1	4		4	3		6	61	1	21													
(GO:0007010)	5	2		14			10	6	58	58		7	47		216	46				3	1	37	4		27	1	1	4	46	15		4		2	29	3	38		1		1	27								
(GO:0007163)		1		4	1		7	2	5	4		1	14		46	58					5	2		13			2	13	3		2	1	11	16												6				
(GO:0006091)				1	31		16			27										115	11		1	8		30		4		1	2															29				
(GO:0006629)				5	28		4	6		17	2				3					11	211		5	4		2		4	1		11	4		10	2		22									1	5	20		
(GO:0016071)	1	3					2	3	1		1				4	1				238	3	1		17	3		2		4		4	6		5	31	4	19	11												
(GO:0007126)	54	34	20	94	14		1	20	8	65	56	1	10	10	1	37	5	1	5	3	250		22	2	1	10	10	1		14	10	13	19		15		12	3	3	20										
(GO:0007005)	3	7	6				20		1						2	4	2	8	4	1		283	30		1	2	39	9		13	50		3	1	27	3	46													
(GO:0071941)							11																16	2																										
(GO:0055086)	25	40	29	6	28		21	9	19	42	15	56	7	8	17	27	13	30	2	17	22	30	2	497	11	4	12	17	11	6	15	6	16	27	36	49	12	14	78	4	38	7								
(GO:0006913)	1	1	1	1			1	5	3						7	1				3	2		11	112			2	1	1	3	24	14	41	6																
(GO:0007031)																							4		22			1				3	14								6									
(GO:0030163)		6	2	5	1		3	5	13	6		6	3	2	4	2				4	2	10		12			193	8	7	1	188	63	19	59		6												2	24	
(GO:0006461)	3	12	2	5			1	21	54	28	2	1	9	1	46	13	4			1		10	39		17	2	1	8	177	20	9	4	11	18	2	5											4	10	6	
(GO:0006457)							2	2	1	1		1	2	4	15	3				4	1	9	1	11	1		7	20	113		8	1	10	5	1	3													1	
(GO:0006486)					63			21															6					1																						
(GO:0051604)	11	8	13	8			3	6	5	16	7	1	9	2	4	4	2	1	4	4	14	13		15	1		188	9	8		285	70	33	62	2	9	2	3	2	5	24	1								
(GO:0070647)	6	23	6	1				1	14	24	7				1	3								6	3	3	63	4	1	70	174	3	34	2	4	7	1	2	7											
(GO:0006605)						11	3		2			3			2	1	2	10							16	24	14	19	11	10		33	3	178	6	2	3											52	9	39
(GO:0007346)	11	37	26	2			3	15	53	31	1	5	18	6	29	11				2	5	19		27	14		59	18	5	62	34	6	273	6	65												1			
(GO:0042254)		4	1					1	3	1						61	3					31	3		36	41		2	1	2	2	2	6		328	4		29	5											
(GO:0023052)	1	5	3	11	1	1	18	5	12	8		27	25	1	38	16				22	4	15		49	6		6	5	3	1	9	3	65	4	330	1	6									10	51			
(GO:0006399)							37	1	1															12																										
(GO:0006351)	3	15	4	2				36	47	9																																								

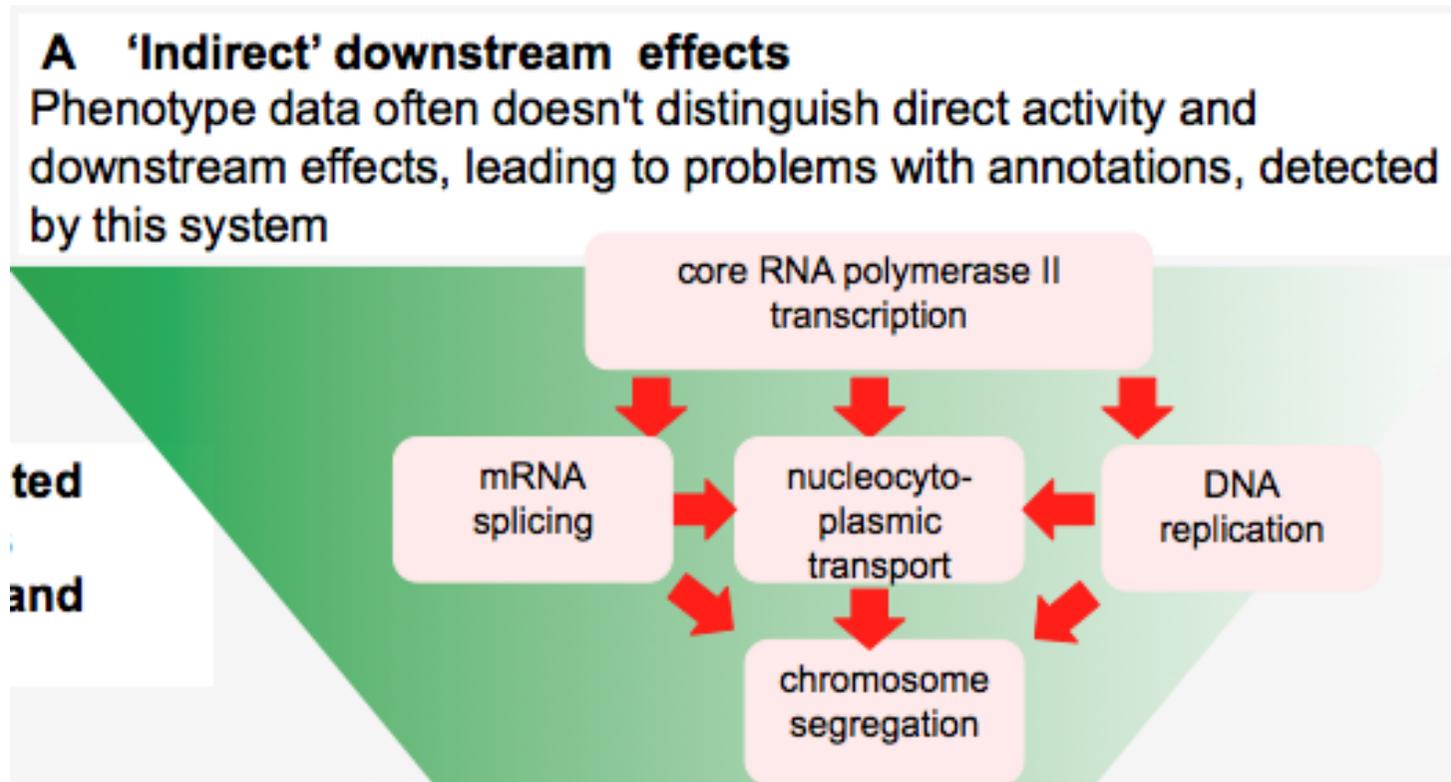
	DNA recombination (GO:0006310)	DNA repair (GO:0006281)	DNA replication (GO:0006260)	ascospore formation (GO:0030437)	carbohydrate metabolic process (GO:0005975)	cell adhesion (GO:0007155)	cellular amino acid metabolic process (GO:0005975)	cell wall organization or biogenesis (GO:00715E)	chromatin modification (GO:0016568)	chromosome organization (GO:0051276)	chromosome segregation (GO:0007059)	cofactor metabolic process (GO:0051186)	conjugation with cellular fusion (GO:0000747)	cytokinesis (GO:0000910)	cytoplasmic translation (GO:0002181)	cytoskeleton organization (GO:0007010)	establishment or maintenance of cell polarity (GO:0007163)	generation of precursor metabolites and energy (GO:0006091)	lipid metabolic process (GO:0006629)	mRNA metabolic process (GO:0016071)	meiotic nuclear division (GO:0007126)	mitochondrion organization (GO:0007005)	nitrogen cycle metabolic process (GO:0071941)	small molecule metabolic process (GO:0055086)	nucleocytoplasmic transport (GO:0006913)	peroxisome organization (GO:0007031)	protein catabolic process (GO:0030163)	protein complex assembly (GO:0006461)	protein folding (GO:0006457)	protein glycosylation (GO:0006486)	protein maturation (GO:0051604)	protein conjugation or removal (GO:0070647)	protein targeting (GO:0006605)	regulation of mitotic cell cycle (GO:0007346)	ribosome biogenesis (GO:0042254)	signaling (GO:0023052)	tRNA metabolic process (GO:0006399)	transcription, DNA-templated (GO:0006351)	transmembrane transport (GO:0055085)	vacuole organization (GO:0007033)	vesicle-mediated transport (GO:0016192)	vitamin metabolic process (GO:0006766)					
DNA recombination (GO:0006310)	28 28 20 3 21 277 138 593	4 62	98 2 21 20 95 130 65	39 2 46 98 11 3 184 190 19 337 19 237 2 275 27 8	15 45 18 20 16 315 943 428	8 1 83	124 85 48 125 27 130 31	89 42 22 43 93 10 74 7 90 136 31 169	28 48 15 130 153 5 1 365 51 100 50 29 100 138 1	28 76 45 153 1299 326 398 204 136 197 40 188 48 159	20 11 18 5 326 724 56 673 69 88 19 29 65 103	3 19 20 1 398 56 267 67 55 72 14 1768	21 23 16 355 294 273 67 910 57 112 35 37 134 444	277 858 316 51 136 69 55 57 689 893 409 57 18 167	136 216 943 100 197 88 72 112 689 100 253 66 49 33	593 584 428 50 40 19 14 35 409 205 398 9 22 229	3 8 29 139 29 37 57 66 9 306 23	37 40 21 15 59 1196 9	4 1 100 48 65 134 18 49 22 880 75	62 106 83 139 159 103 6 444 167 333 229 23 75 294	97 1 28	98 226 129 124 224 607 63 281 249 834 963 37 87 1028	2 59 8 85 223 236 12 341 72 183 131 40 68 870 1	21 11 41 48 283 24 301 36 52 73 21 471 2	90 47 35 129 279 216 447 403 98 116 15 153 26 43	95 175 58 27 78 26 17 86 172 324 59 13 27 8 18 75 50 43 64 6772 107 55	338 1124 324 136 207 26 28 400 367 185 1196 77 128 336 1	65 121 177 31 66 51 284 25 43 106 9 41 8 29 9 199 69 266 163 55 55 4978 15 703 39 78 106 103 182 11 455 101 1194 76 52 632 280 207 1027 40 141 18	1039 1473 768 89 181 235 990 356 423 1058 517 288 114 510 58 1061 427 1401 644 359 768 703 24 6090 701 130 414 1298 345 286 720 131 645 368 741 6493 325 616 8002 220 1800 214	39 101 53 42 117 62 29 63 108 284 45 21 31 99 46 230 151 64 73 332 68 39 8 701 3698 1 79 197 39 21 142 124 1277 258 655 1091 29 442 59 49 168	2 2 1 22 19 3 9 27 1 9 2 12 22 33 17 15 388 15 33 78	46 387 87 43 338 46 68 102 288 463 101 47 75 76 25 336 105 102 208 89 139 156 414 79 39 8701 620 132 113 7745 2819 331 549 22 905 7 216 98 72 486 8	98 283 195 93 352 267 272 60 484 297 398 252 44 391 7 1800 263 499 536 116 285 1032 13 129 197 224 620 1919 359 56 823 303 564 462 136 1745 34 670 1001 51 891 34	11 9 23 10 84 18 37 52 32 59 1 5 33 14 16 132 23 34 49 53 25 182 14 345 39 5 132 359 5332 40 213 39 188 50 27 326 10 59 149 11 192 2	3 44 2 3021 96 16 474 47 66 4 5 20 32 31 56 53 310 18 4 11 286 21 7 113 56 40 1021 132 56 38 9 1 239 24 104 32 139 1	184 418 229 74 410 327 271 284 351 596 141 71 129 141 54 461 130 171 465 144 215 455 5 720 142 56 775 823 213 132 3999 3190 609 701 33 2073 33 400 262 176 818 14	190 575 95 7 119 50 39 62 653 858 127 18 24 74 36 205 23 21 68 167 142 101 1 131 124 70 2819 303 39 56 3933 7715 163 476 44 1280 79 389 88 21 209	19 60 20 90 158 67 75 88 71 181 18 53 50 34 4 209 82 99 475 141 123 1194 8 645 1277 845 331 954 188 38 609 163 6000 160 164 1233 29 281 1021 242 1013 9	337 672 407 136 127 149 49 114 438 1001 592 15 69 402 40 70 235 48 153 192 452 76 368 259 1 549 462 50 9 701 476 160 3090 90 1511 22 552 63 31 149 1	19 44 65 31 91 28 54 25 59 141 35 82 26 31 165 66 87 127 73 424 40 52 741 605 22 136 27 1 33 44 164 90 6000 128 386 280 20 13 72 8	237 469 155 169 954 2337 340 891 750 1139 251 112 372 796 11 1600 900 435 200 380 470 632 24 643 1081 152 905 1745 326 239 2070 1286 1233 1812 128 6661 61 4004 3239 255 3710 46	2 30 11 11 2 2038 10 18 37 11 27 8 37 31 10 12 264 7 280 325 29 7 34 10 33 79 29 22 386 61 5091 94 3 24 13	279 909 393 77 529 305 295 253 233 2700 179 222 32 80 313 112 200 701 799 411 207 33 616 442 13 216 670 59 24 400 389 281 592 280 4691 94 989 78 17 255 16	27 25 12 40 409 180 143 218 27 36 14 123 56 33 2 109 84 1299 605 26 58 1027 7 8002 59 611 98 1001 149 104 262 88 1021 63 20 3039 3 78 9899 103 565 30	12 81 111 11 7 72 37 51 13 10 55 76 111 70 46 174 2 85 40 220 49 50 72 51 11 32 176 21 242 31 13 255 17 103 655 602 14	8 74 19 181 396 530 130 266 73 164 63 61 51 506 3 1025 463 116 693 80 289 141 8 1025 168 118 486 891 192 139 818 209 1013 149 72 3710 24 255 565 602 18196 14	2 3 122 10 505 9 3 3 1 1002 4 1 33 197 2 4 18 214 8 34 2 1 14 9 1 8 46 13 16 30 14 2416

All organisms

190	298	165	DNA recombination (GO:0006310)
			DNA repair (GO:0006281)
			DNA replication (GO:0006260)
341	749	167	asospore formation (GO:0030437)
		47	carbohydrate metabolic process (GO:0005975)
637	896	272	cell adhesion (GO:0007155)
		90	cellular amino acid metabolic process (GO:0001001)
			cell wall organization or biogenesis (GO:0071551)
			chromatin modification (GO:0016568)
			chromosome organization (GO:0051276)
			chromosome segregation (GO:0007059)
			cofactor metabolic process (GO:0051186)
			conjugation with cellular fusion (GO:0000747)
185		2	cytokinesis (GO:0000910)
		1074	cytoplasmic translation (GO:0002181)
340	133	553	cytoskeleton organization (GO:0007010)
		171	establishment or maintenance of cell polarity (GO:0007010)
365	190		generation of precursor metabolites and energy (GO:0006913)
		15	lipid metabolic process (GO:0006629)
			mRNA metabolic process (GO:0016071)
			meiotic nuclear division (GO:0007126)
			mitochondrion organization (GO:0007005)
621	208	9	nitrogen cycle metabolic process (GO:0071941)
			nucleobase-containing small molecule metabolism (GO:0006913)
		278	nucleocytoplasmic transport (GO:0006913)
575		50	peroxisome organization (GO:0007031)
			protein catabolic process (GO:0030163)
			protein complex assembly (GO:0006461)
			protein folding (GO:0006457)
114	608		protein glycosylation (GO:0006486)
433			protein maturation (GO:0051604)
228	554		protein modification by small protein conjugation (GO:0006605)
100	4401	39	protein targeting (GO:0006605)
			regulation of mitotic cell cycle (GO:0007346)
			ribosome biogenesis (GO:0042254)
			signaling (GO:0023052)
		3432	tRNA metabolic process (GO:0006399)
336		68	transcription, DNA-templated (GO:0006351)
		1037	transmembrane transport (GO:0055085)
		32	vacuole organization (GO:0007033)
			vesicle-mediated transport (GO:0016192)
			vitamin metabolic process (GO:0006766)
		1682	embryo development (GO:0009790)

Embryo development

# Classes of errors in intersections



# B Automated mapping errors

## ▾ GO Biological Process

[Ontology Graph](#)

[Contact Curators](#)

Term ID	Term Name	Evidence	With/From	Reference
<a href="#">GO:0006355</a>	regulation of transcription, DNA-dependent	IEA	KW-0805 <b>X</b>	<a href="#">GO_REF:0000004</a>
<a href="#">GO:0051083</a>	'de novo' cotranslational protein folding	ISO	<a href="#">SGD:S000001236</a>	<a href="#">GO_REF:0000024</a>

## ▾ GO Cellular Component

[Ontology Graph](#)

[Contact Curators](#)

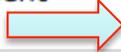
Term ID	Term Name	Evidence	With/From	Reference
<a href="#">GO:0005634</a>	nucleus	IEA	KW-0539	<a href="#">GO_REF:0000004</a>
<a href="#">GO:0005829</a>	cytosol	IDA		<a href="#">Matsuyama A et al. (2006)</a>
<a href="#">GO:0005854</a>	nascent polypeptide-associated complex	ISO	<a href="#">SGD:S000001236</a>	<a href="#">GO_REF:0000024</a>
<a href="#">GO:0042788</a>	polysomal ribosome	IDA		<a href="#">Andersen KM et al. (2007)</a>

More example

[https://sourceforge.net/tracker/?group\\_id=36855&atid=605890&source=navbar](https://sourceforge.net/tracker/?group_id=36855&atid=605890&source=navbar)

# C: Manual Annotation Errors

## Amino acid metabolism annotation updates

Database	gene	update	reason
PomBase?	fmt1	removed methionyl-tRNA aminoacylation 	missannotation should be methionyl-tRNA to N-formyl-methionyl-tRNA
PomBase?	SPAC823.09	waiting	
PomBase?	hst4	mitotic sister chromatid segregation	indirect
PomBase?	sla1	removed regulation of conjugation with cellular fusion	indirect
PomBase?	sla1	removed regulation of ascospore formation	indirect
SGD	MET5	removed fungal-type cell wall organization	indirect
SGD	ARG7	removed fungal-type cell wall organization	indirect
SGD	GAP1	ammonia assimilation cycle	indirect
SGD	FMT1	removed methionyl-tRNA aminoacylation 	error, should be methionyl-tRNA to N-formyl-methionyl-tRNA
SGD	ARG81	replaced arginine metabolic process with "regulation of arginine metabolic process"	upstream regulation
SGD	ARG81	removed negative regulation of calcium ion-dependent exocytosis 	error
SGD	CDC60	removed DNA replication initiation	indirect
SGD	MET18	removed methionine metabolic process	indirect
SGD	THR4	removed endocytosis	indirect
SGD	FPR1	replaced arginine metabolic process with "regulation of arginine metabolic process"	upstream regulation
SGD	CBF1	methionine biosynthetic process	indirect
MGI	Sod1	query nucleotide excision repair IMP 	error, incorrect interpretation

# D Ontology True Path Violations



More recent examples:

<http://tinyurl.com/e4rmh>

GO:0006606 protein import into nucleus

removed parent

GO:0044743 intracellular protein transmembrane import

GO:0042307 positive regulation of protein import into nucleus

remove parent

GO:0034762 regulation of transmembrane transport

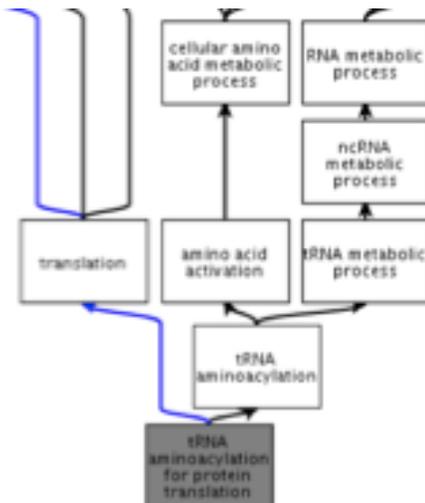
GO:0010559 regulation of glycoprotein biosynthetic process

remove parent

GO:0006417 regulation of translation

# E: Paradigm shifts and new biology

# Allowed Intersections



## 1. Multiple Parentage

Specific function/process is involved in more than one process simultaneously, Link between processes is encoded in the ontology

e.g. tRNA ligase is a child of tRNA metabolism, amino acid metabolism, translation

e.g. pentose phosphate shunt is a child of nucleotide metabolism, carbohydrate metabolism and coenzyme metabolism

## 2. Multifunctional Gene Product

gene product or pathway functions independently in the multiple processes

e.g. Noc3: intranuclear transport ; and DNA replication initiation

e.g. Urmylation pathway ADD

e.g. Tandem fusion ADD

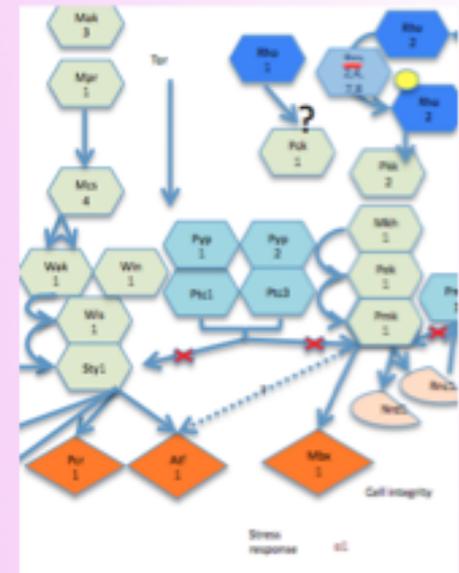
## 3. One process directly regulating another

In most cases are probably indirect downstream effects, but are sometimes regulatory.

e.g. Meiosis specific mRNA splicing

e.g. (eg of regulated transport controlling a process?) ADD

e.g. chromatin remodeling -> DNA replication? FING EG



## 4. Upstream signalling

regulating multiple diverse processes, via protein modification/transcription

# Future work

- Implement rules into GO QC
- Implement rules for intersects with many annotations to cover entire annotation space
- Create rules for within and between other ontology annotations (process/ component, component phenotype etc...)
- Rules can be used to create a set of “negative” annotations for \*\*\*\* (applications? I know some users need these for reasoning...)

# Acknowledgements

- Seth Carbon & Chris Mungall (BDGP)
- Antonia Lock and Midori Harris (PomBase)
- Maria Costanzo and ALL SGD curators
- David Hill, Harold Drabkin (MGI)
- InterPro (Everyone)
- GOA (Rachael Huntley, Emily Dimmer)
- UniPROT (Mark Feuermann Ivo Pedruzzi and others)