Data analysis: YeastMine, GO tools, and use cases

SGD: www.yeastgenome.org

YeastMine: http://www.yeastmine.yeastgenome.org

Email: sgd-helpdesk@lists.stanford.edu

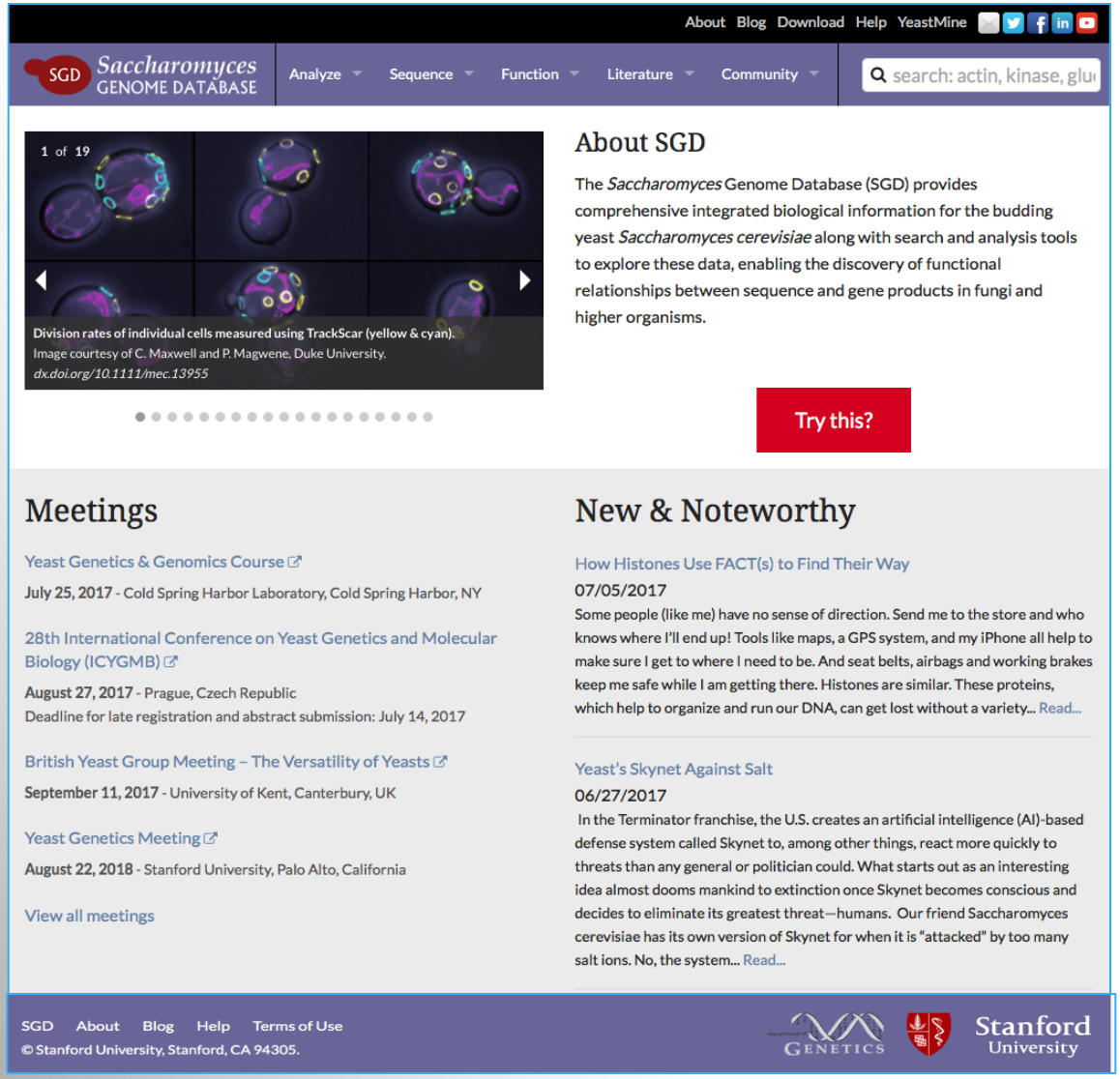
Rob Nash
Senior Biocuration Scientist
rnash@stanford.edu

Saccharomyces GENOME DATABASE

About SGD

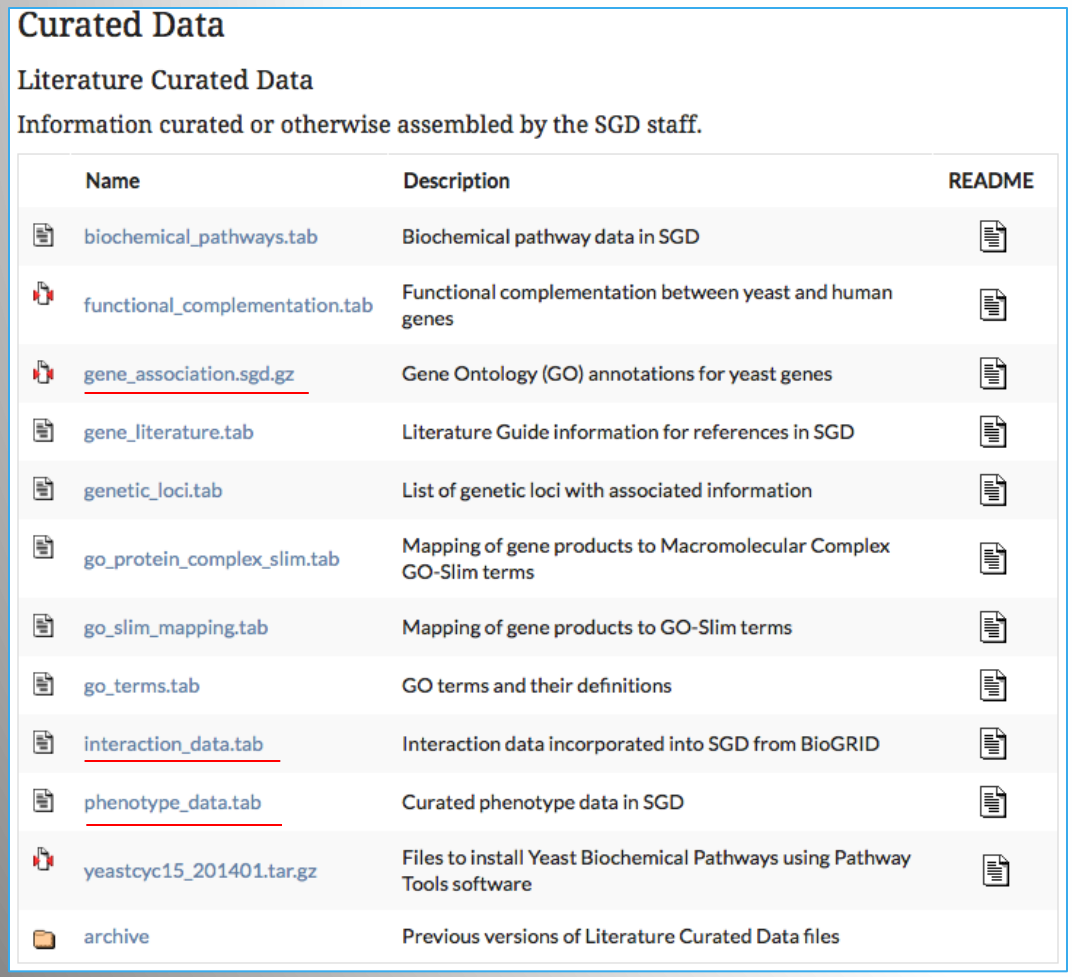
- Started by David Botstein in the early 90s based on the need to organize info on the genes of the budding yeast, *S. cerevisiae*.
- Mike Cherry P.I. of the project since 1992. On May 8th, 1994 the first external web hits were recorded
- Public, open, non-profit academic project funded by the NIH (NHGRI U41 grant)
- Completing the transition to a new flexible, expandable schema (postgres) with everything in the cloud

SGD



SGD Saccharomyces GENOME DATABASE

How to leverage data rich SGD!



Curated Data

Literature Curated Data

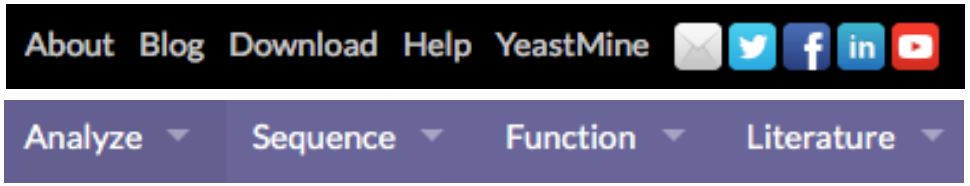
Information curated or otherwise assembled by the SGD staff.

Name	Description	README
biochemical_pathways.tab	Biochemical pathway data in SGD	
functional_complementation.tab	Functional complementation between yeast and human genes	
gene_association.sgd.gz	Gene Ontology (GO) annotations for yeast genes	
gene_literature.tab	Literature Guide information for references in SGD	
genetic_loci.tab	List of genetic loci with associated information	
go_protein_complex_slim.tab	Mapping of gene products to Macromolecular Complex GO-Slim terms	
go_slim_mapping.tab	Mapping of gene products to GO-Slim terms	
go_terms.tab	GO terms and their definitions	
interaction_data.tab	Interaction data incorporated into SGD from BioGRID	
phenotype_data.tab	Curated phenotype data in SGD	
yeastcyc15_201401.tar.gz	Files to install Yeast Biochemical Pathways using Pathway Tools software	
archive	Previous versions of Literature Curated Data files	

- 111K GO annotations (manual, HTP and computational)
- 146K phenotype annotations (manual and HTP)
- 342K physical (130K) and genetic (212K) interactions

Analysis entry points

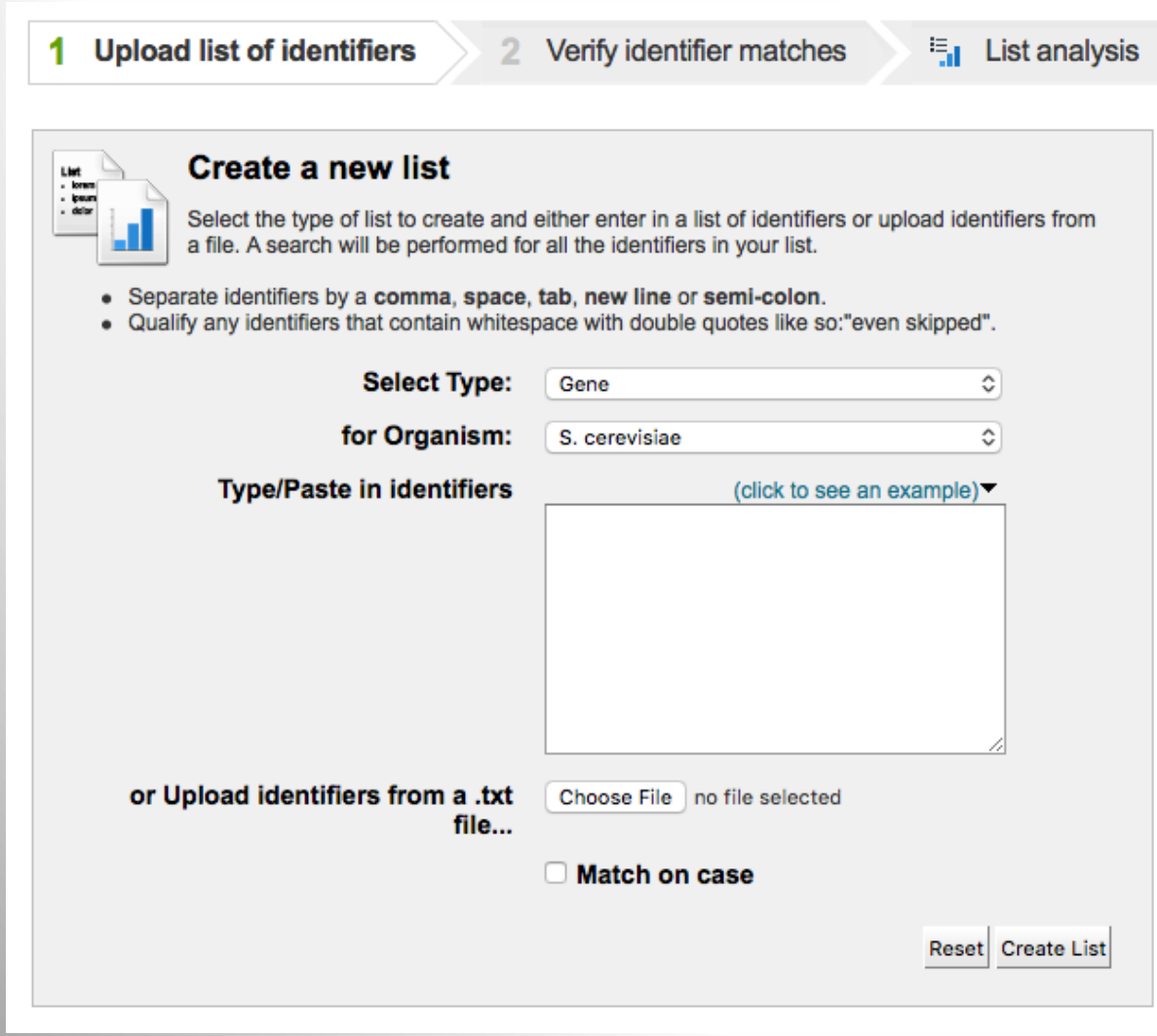
- SGD home page



- GO term pages (+/- child terms)
- Phenotype (observable-qualifier or obs. Pages)
- Interaction pages (Phys., Gen, Intersection, All)
- Domain pages (protein tab)
- E.C. number pages

SGD Saccharomyces GENOME DATABASE

…or analyze your own gene list



Saccharomyces GENOME DATABASE

Presentation outline

- GO Slim Mapper
- GO Term Finder
- YeastMine

The Gene Ontology (GO) Project

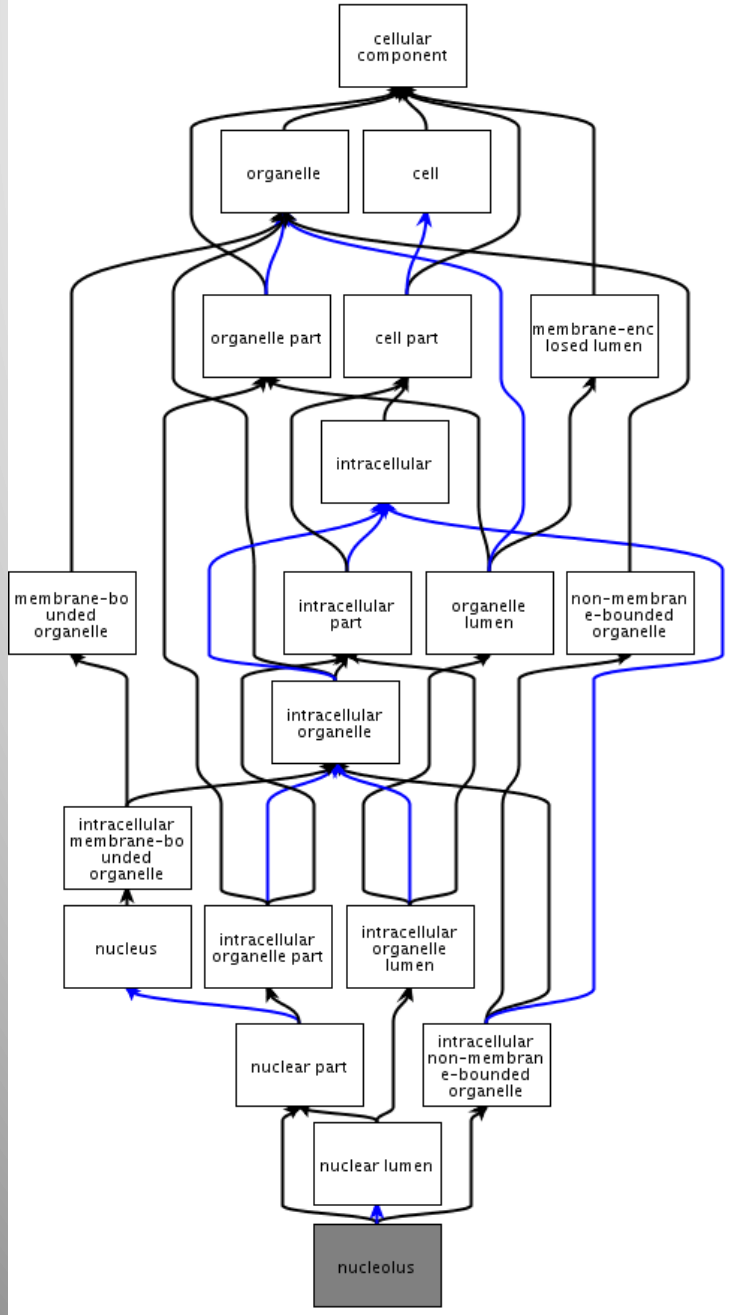
- A collaboration among MODs to improve queries within and across databases.
- Developed three structured ontologies to describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.
- The use of GO terms by collaborating databases facilitates uniform queries across all of them.

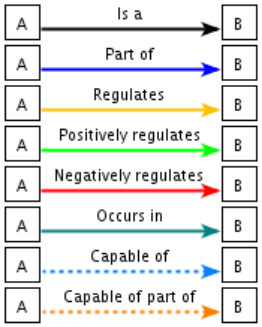
GO, a set of three independent structured, controlled vocabularies for describing the molecular function, biological process, and cellular component of gene products

Molecular function: the tasks performed by individual gene products, for example, *fructose-bisphosphate aldolase activity* or *protein serine/threonine kinase activity*.

Biological process: the broad biological goals, such as *mitosis* or *DNA replication*, that are accomplished by ordered assemblies of molecular functions.

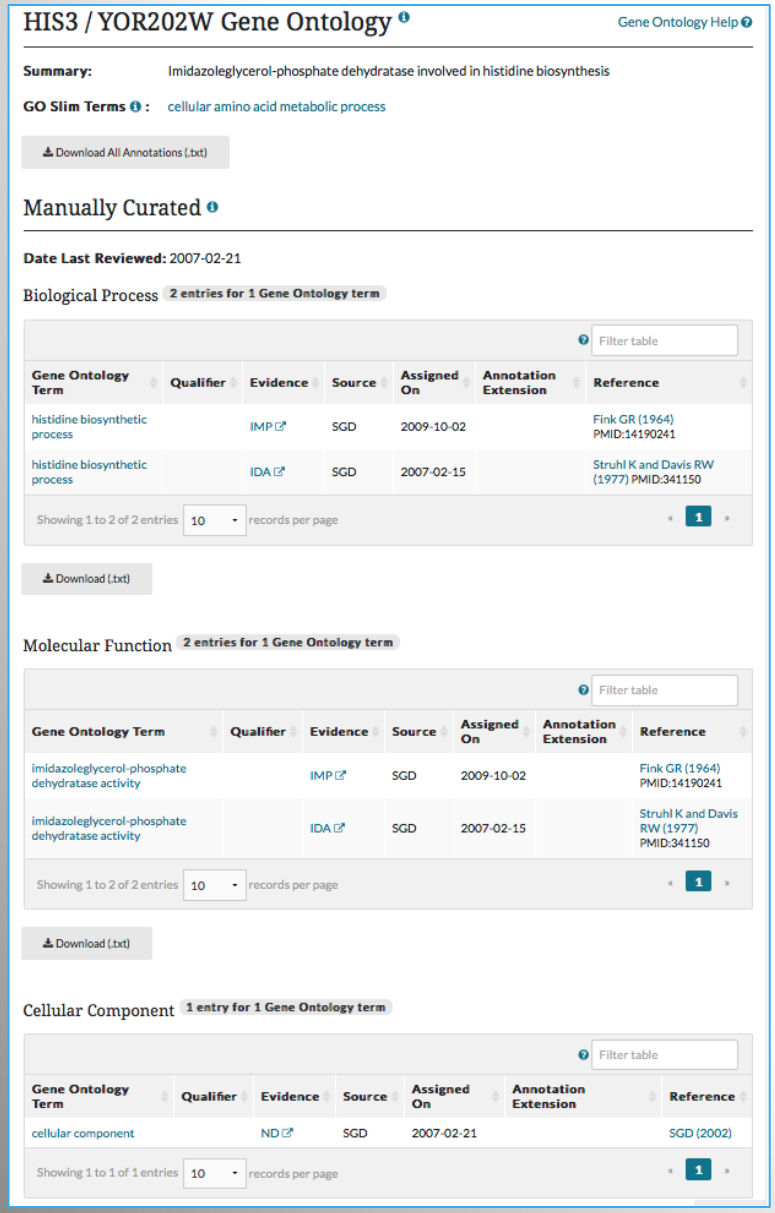
Cellular component: subcellular structures, locations, and macromolecular complexes, such as *nucleus, cellular bud tip,* and *origin recognition complex.*





Graph of ancestry for nucleolus in the cellular component aspect

GO Annotation Details



HIS3 / YOR202W Gene Ontology

Gene Ontology Help

Summary: Imidazoleglycerol-phosphate dehydratase involved in histidine biosynthesis

GO Slim Terms : cellular amino acid metabolic process

Download All Annotations (.txt)

Manually Curated

Date Last Reviewed: 2007-02-21

Biological Process — 2 entries for 1 Gene Ontology term

Filter table

Gene Ontology Term	Qualifier	Evidence	Source	Assigned On	Annotation Extension	Reference
histidine biosynthetic process		IMP	SGD	2009-10-02		Fink GR (1964) PMID:14190241
histidine biosynthetic process		IDA	SGD	2007-02-15		Struhl K and Davis RW (1977) PMID:341150

Showing 1 to 2 of 2 entries 10 records per page « 1 »

Download (.txt)

Molecular Function — 2 entries for 1 Gene Ontology term

Filter table

Gene Ontology Term	Qualifier	Evidence	Source	Assigned On	Annotation Extension	Reference
imidazoleglycerol-phosphate dehydratase activity		IMP	SGD	2009-10-02		Fink GR (1964) PMID:14190241
imidazoleglycerol-phosphate dehydratase activity		IDA	SGD	2007-02-15		Struhl K and Davis RW (1977) PMID:341150

Showing 1 to 2 of 2 entries 10 records per page « 1 »

Download (.txt)

Cellular Component — 1 entry for 1 Gene Ontology term

Filter table

Gene Ontology Term	Qualifier	Evidence	Source	Assigned On	Annotation Extension	Reference
cellular component		ND	SGD	2007-02-21		SGD (2002)

Showing 1 to 1 of 1 entries 10 records per page « 1 »

⟵ **GO Summary**

⟵ **Biological Process**

⟵ **Molecular Function**

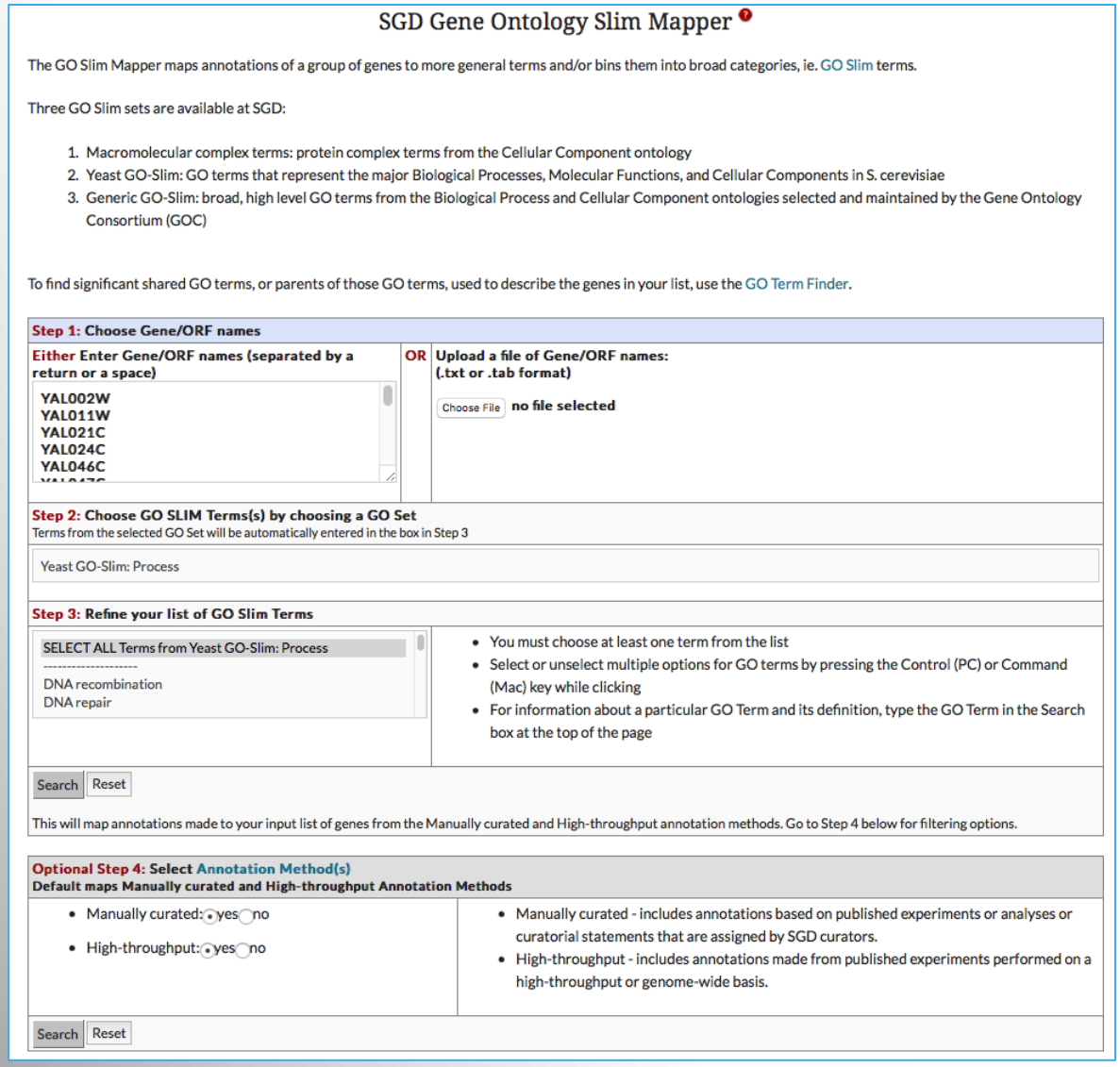
⟵ **Cellular Component**

GO Slim Mapper

Definition: Maps annotations of a group of genes to more general terms; binning them into broad categories

Scenario: You complete a screen looking for mutants with altered sensitivity to a drug and want to know based on the mutants identified what process might be affected.

GO Slim Mapper



SGD Gene Ontology Slim Mapper

The GO Slim Mapper maps annotations of a group of genes to more general terms and/or bins them into broad categories, ie. GO Slim terms.

Three GO Slim sets are available at SGD:

1. Macromolecular complex terms: protein complex terms from the Cellular Component ontology
2. Yeast GO-Slim: GO terms that represent the major Biological Processes, Molecular Functions, and Cellular Components in S. cerevisiae
3. Generic GO-Slim: broad, high level GO terms from the Biological Process and Cellular Component ontologies selected and maintained by the Gene Ontology Consortium (GOC)

To find significant shared GO terms, or parents of those GO terms, used to describe the genes in your list, use the GO Term Finder.

Step 1: Choose Gene/ORF names

Either Enter Gene/ORF names (separated by a return or a space)

YAL002W
YAL011W
YAL021C
YAL024C
YAL046C

OR Upload a file of Gene/ORF names: (.txt or .tab format)

Choose File no file selected

Step 2: Choose GO SLIM Terms(s) by choosing a GO Set
Terms from the selected GO Set will be automatically entered in the box in Step 3

Yeast GO-Slim: Process

Step 3: Refine your list of GO Slim Terms

SELECT ALL Terms from Yeast GO-Slim: Process

DNA recombination
DNA repair

- You must choose at least one term from the list
- Select or unselect multiple options for GO terms by pressing the Control (PC) or Command (Mac) key while clicking
- For information about a particular GO Term and its definition, type the GO Term in the Search box at the top of the page

Search Reset

This will map annotations made to your input list of genes from the Manually curated and High-throughput annotation methods. Go to Step 4 below for filtering options.

Optional Step 4: Select Annotation Method(s)
Default maps Manually curated and High-throughput Annotation Methods

- Manually curated: yes no
- High-throughput: yes no

- Manually curated - includes annotations based on published experiments or analyses or curatorial statements that are assigned by SGD curators.
- High-throughput - includes annotations made from published experiments performed on a high-throughput or genome-wide basis.

Search Reset

Saccharomyces GENOME DATABASE

GO Slim Mapper: Results

SGD Gene Ontology Slim Mapper

This page displays genes from your query that are annotated directly or indirectly (via a parent:child relationship) to the GO Slim terms of your choice.

Results for the mapping of 438 genes to the Yeast GO-Slim Process

GO-Slim term	Cluster frequency	Genome frequency	Genes annotated to the term
mitotic cell cycle | AmiGO	76 out of 438 genes, 17.4%	319 of 6433 genes, 5%	ACT1, BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUD30, BUD31, BUR2, CCR4, CDC10, CDC11, CDC25, CDC28, CDC53, CDC55, CDH1, CHL4, CIK1, CIN8, CKS1, CLA4, CLB2, CSM3, CTF18, CTF19, CTF4, CTF8, CTR9, DCC1, ELM1, HHT2, HIR3, HOF1, HPC2, IPL1, IRC15, JNM1, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRC1, NDD1, NUP53, PAF1, PHO85, PIN4, PSE1, REF2, RMI1, SGO1, SHE1, SHS1, SIT4, SMC2, SPC72, STE20, STH1, STU2, TOF1, TOP1, TOP3, TUB2, TUB3, ULP2, YBP2, YRB1
organelle fission | AmiGO	64 out of 438 genes, 14.6%	302 of 6433 genes, 4.7%	BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CDC10, CDC28, CDC55, CDH1, CHL4, CIK1, CIN8, CLA4, CLB2, CSM1, CSM3, CST9, CTF18, CTF19, CTF4, CTF8, DBF2, DCC1, HHT2, HOP1, HOP2, HOS2, IPL1, IRC15, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRC1, NUP53, PAP2, PSE1, RMI1, SAR1, SGO1, SHE1, SHS1, SLZ1, SMC2, SPC72, STE20, STH1, STU2, TOF1, TOP1, TOP3, TUB2, TUB3, ULP2, YBP2
transcription from RNA polymerase II promoter | AmiGO	62 out of 438 genes, 14.2%	476 of 6433 genes, 7.4%	ARP4, BUD30, BUR2, CBF1, CCR4, CDC28, CDC36, CEG1, CKS1, CLA4, CTK1, CTR9, ESA1, GCN5, GCR1, GIM3, GIM5, GPB2, HCM1, HIR1, HIR3, HPC2, HPR1, HSF1, HTZ1, IMP2', INO80, LDB7, MFT1, MOT3, NDD1, NPL6, NUP84, OPI1, PAF1, PFD1, PHO80, PHO85, POB3, POP2, REB1, REF2, REG1, RPB9, RPN4, RSC2, SNF4, SNF8, SSN3, STE20, STH1, SUM1, SWD2, TAF14, THP2, TOP1, TRA1, VPS25, VPS36, YAP1, YKE2, YSH1
chromosome segregation | AmiGO	61 out of 438 genes, 13.9%	203 of 6433 genes, 3.2%	BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CBF1, CDC28, CDC55, CDH1, CHL4, CIK1, CIN8, CLB2, CSM1, CSM3, CST9, CTF18, CTF19, CTF4, CTF8, DAM1, DCC1, HHT2, HOP1, HOP2, IES6, INO80, IPL1, IRC15, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRC1, NDC80, RMI1, RSC2, SGO1, SHE1, SHP1, SLI15, SMC2, SMC6, SPC72, STH1, STS1, TOF1, TOP1, TOP3, TUB2, TUB3, ULP2
cytoskeleton organization | AmiGO	58 out of 438 genes, 13.2%	238 of 6433 genes, 3.7%	ACT1, ARC18, ARC35, ARC40, ARP3, AVO1, BBP1, BEM2, BIK1, BIM1, BIT61, CDC10, CDC28, CDC31, CDC37, CDH1, CIK1, CIN8, CLA4, CLB2, CMD1, COF1, CRN1, DAM1, ELM1, HCM1, HOF1, IPL1, IRC15, JNM1, KAR9, KIP3, LAS17, LIA1, MDM20, MIF2, MPS1, MPS2, MPS3, NAT3, NDC80, PFD1, REF2, RHO2, SAC6, SCD5, SHE1, SHE4, SHS1, SIT4, SPC29, SPC72, SRV2, STH1, STU1, STU2, TOR2, YBP2
regulation of organelle organization | AmiGO	52 out of 438 genes, 11.9%	283 of 6433 genes, 4.4%	ARC40, ARP3, BDF1, BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CDC28, CDC55, CDH1, CHL4, CIK1, CLA4, CLB2, CRN1, CTF19, CTR9, DAM1, HHT2, HIR3, HOF1, HOP1, HOS2, IPL1, KAR9, KIP3, LAS17, LTE1, MAD1, MAD2, MAD3, MCM21, MDM20, MPS1, NAT3, NUP53, PAF1, PSE1, REF2, RIM21, SAR1, SEC12, SGO1, SIR4, SMC2, SPC72, SRV2, STE20, ULP2
regulation of cell cycle | AmiGO	49 out of 438 genes, 11.2%	233 of 6433 genes, 3.6%	BFA1, BIM1, BUB1, BUB2, BUB3, CCR4, CDC25, CDC28, CDC36, CDC37, CDC55, CDH1, CHL4, CIK1, CKS1, CLA4, CLB2, CSM3, CTF19, DAM1, DBF2, ESA1, HHT2, HOP1, HOS2, IPL1, IRC15, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCM21, MPS1, MRC1, NUP53, PHO80, PIN4, PSE1, REF2, SGO1, SLI15, SMC2, SPC72, STE20, TOF1, TOR2, ULP2

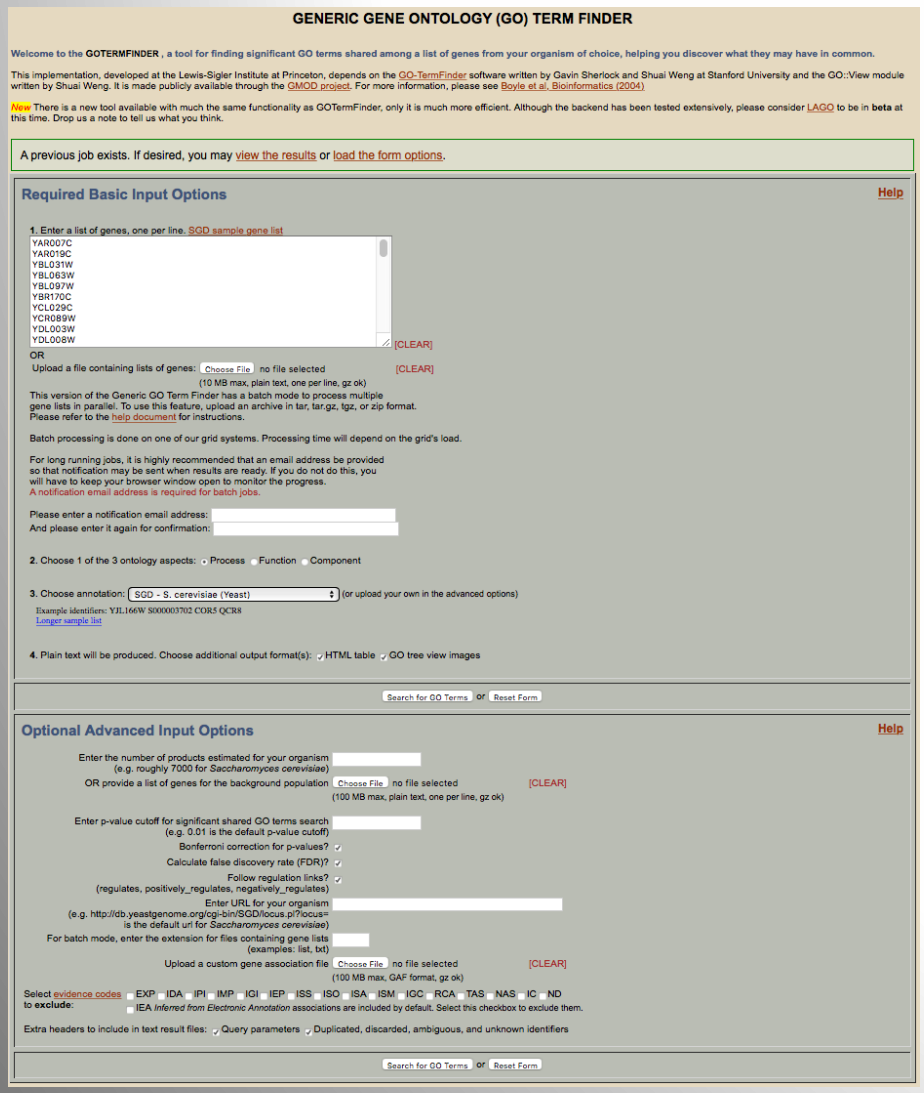
GO Term Finder

Definition: Searches for significant shared GO terms or parents of these terms, to help discover what a set of genes may have in common.

Scenario: You complete a screen looking for mutants with possible spindle defects and want to know whether you are on the right track.

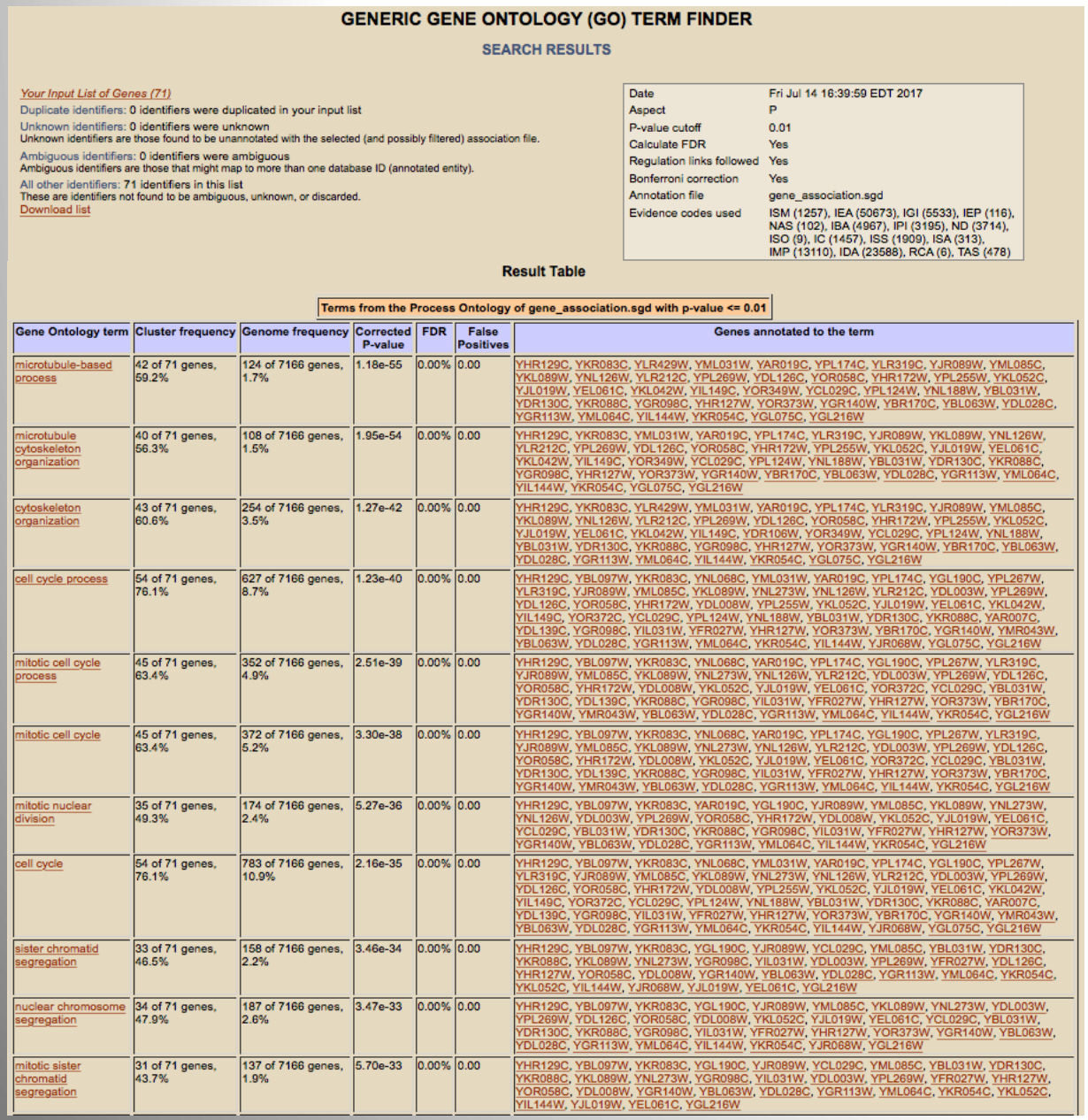
Batch GO Term Finder

http://go.princeton.edu/cgi-bin/GOTermFinder



Advantages:

- process multiple gene lists in parallel
- handles longer gene lists
- large number of available organisms



GENERIC GENE ONTOLOGY (GO) TERM FINDER

SEARCH RESULTS

Your Input List of Genes (71)

Duplicate identifiers: 0 identifiers were duplicated in your input list

Unknown identifiers: 0 identifiers were unknown
Unknown identifiers are those found to be unannotated with the selected (and possibly filtered) association file.

Ambiguous identifiers: 0 identifiers were ambiguous
Ambiguous identifiers are those that might map to more than one database ID (annotated entity).

All other identifiers: 71 identifiers in this list
These are identifiers not found to be ambiguous, unknown, or discarded.

Download list

Date	Fri Jul 14 16:39:59 EDT 2017
Aspect	P
P-value cutoff	0.01
Calculate FDR	Yes
Regulation links followed	Yes
Bonferroni correction	Yes
Annotation file	gene_association.sgd
Evidence codes used	ISM (1257), IEA (50673), IGI (5533), IEP (116), NAS (102), IBA (4967), IPI (3195), ND (3714), ISO (9), IC (1457), ISS (1909), ISA (313), IMP (13110), IDA (23588), RCA (6), TAS (478)

Result Table

Terms from the Process Ontology of gene_association.sgd with p-value <= 0.01

Gene Ontology term	Cluster frequency	Genome frequency	Corrected P-value	FDR	False Positives	Genes annotated to the term
microtubule-based process	42 of 71 genes, 59.2%	124 of 7166 genes, 1.7%	1.18e-55	0.00%	0.00	YHR129C, YKR083C, YLR429W, YML031W, YAR019C, YPL174C, YLR319C, YJR089W, YML085C, YKL089W, YNL126W, YLR212C, YPL269W, YDL126C, YOR058C, YHR172W, YPL255W, YKL052C, YJL019W, YEL061C, YKL042W, YIL149C, YOR349W, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YGR098C, YHR127W, YOR373W, YGR140W, YBR170C, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL075C, YGL216W
microtubule cytoskeleton organization	40 of 71 genes, 56.3%	108 of 7166 genes, 1.5%	1.95e-54	0.00%	0.00	YHR129C, YKR083C, YML031W, YAR019C, YPL174C, YLR319C, YJR089W, YKL089W, YNL126W, YLR212C, YPL269W, YDL126C, YOR058C, YHR172W, YPL255W, YKL052C, YJL019W, YEL061C, YKL042W, YIL149C, YOR349W, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YGR098C, YHR127W, YOR373W, YGR140W, YBR170C, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL075C, YGL216W
cytoskeleton organization	43 of 71 genes, 60.6%	254 of 7166 genes, 3.5%	1.27e-42	0.00%	0.00	YHR129C, YKR083C, YLR429W, YML031W, YAR019C, YPL174C, YLR319C, YJR089W, YML085C, YKL089W, YNL126W, YLR212C, YPL269W, YDL126C, YOR058C, YHR172W, YPL255W, YKL052C, YJL019W, YEL061C, YKL042W, YIL149C, YDR106W, YOR349W, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YGR098C, YHR127W, YOR373W, YGR140W, YBR170C, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL075C, YGL216W
cell cycle process	54 of 71 genes, 76.1%	627 of 7166 genes, 8.7%	1.23e-40	0.00%	0.00	YHR129C, YBL097W, YKR083C, YNL068C, YML031W, YAR019C, YPL174C, YGL190C, YPL267W, YLR319C, YJR089W, YML085C, YKL089W, YNL273W, YNL126W, YLR212C, YDL003W, YPL269W, YDL126C, YOR058C, YHR172W, YDL008W, YPL255W, YKL052C, YJL019W, YEL061C, YKL042W, YIL149C, YOR372C, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YAR007C, YDL139C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YIL144W, YJR068W, YGL075C, YGL216W
mitotic cell cycle process	45 of 71 genes, 63.4%	352 of 7166 genes, 4.9%	2.51e-39	0.00%	0.00	YHR129C, YBL097W, YKR083C, YNL068C, YAR019C, YPL174C, YGL190C, YPL267W, YLR319C, YJR089W, YML085C, YKL089W, YNL273W, YNL126W, YLR212C, YDL003W, YPL269W, YDL126C, YOR058C, YHR172W, YDL008W, YKL052C, YJL019W, YEL061C, YOR372C, YCL029C, YBL031W, YDR130C, YDL139C, YKR088C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL216W
mitotic cell cycle	45 of 71 genes, 63.4%	372 of 7166 genes, 5.2%	3.30e-38	0.00%	0.00	YHR129C, YBL097W, YKR083C, YNL068C, YAR019C, YPL174C, YGL190C, YPL267W, YLR319C, YJR089W, YML085C, YKL089W, YNL273W, YNL126W, YLR212C, YDL003W, YPL269W, YDL126C, YOR058C, YHR172W, YDL008W, YKL052C, YJL019W, YEL061C, YOR372C, YCL029C, YBL031W, YDR130C, YDL139C, YKR088C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL216W
mitotic nuclear division	35 of 71 genes, 49.3%	174 of 7166 genes, 2.4%	5.27e-36	0.00%	0.00	YHR129C, YBL097W, YKR083C, YAR019C, YGL190C, YJR089W, YML085C, YKL089W, YNL273W, YNL126W, YDL003W, YPL269W, YOR058C, YHR172W, YDL008W, YKL052C, YJL019W, YEL061C, YCL029C, YBL031W, YDR130C, YKR088C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YGR140W, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YGL216W
cell cycle	54 of 71 genes, 76.1%	783 of 7166 genes, 10.9%	2.16e-35	0.00%	0.00	YHR129C, YBL097W, YKR083C, YNL068C, YML031W, YAR019C, YPL174C, YGL190C, YPL267W, YLR319C, YJR089W, YML085C, YKL089W, YNL273W, YNL126W, YLR212C, YDL003W, YPL269W, YDL126C, YOR058C, YHR172W, YDL008W, YPL255W, YKL052C, YJL019W, YEL061C, YKL042W, YIL149C, YOR372C, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YAR007C, YDL139C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YIL144W, YJR068W, YGL075C, YGL216W
sister chromatid segregation	33 of 71 genes, 46.5%	158 of 7166 genes, 2.2%	3.46e-34	0.00%	0.00	YHR129C, YBL097W, YKR083C, YGL190C, YJR089W, YCL029C, YML085C, YBL031W, YDR130C, YKR088C, YKL089W, YNL273W, YGR098C, YIL031W, YDL003W, YPL269W, YFR027W, YDL126C, YHR127W, YOR058C, YDL008W, YGR140W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YKL052C, YIL144W, YJR068W, YJL019W, YEL061C, YGL216W
nuclear chromosome segregation	34 of 71 genes, 47.9%	187 of 7166 genes, 2.6%	3.47e-33	0.00%	0.00	YHR129C, YBL097W, YKR083C, YGL190C, YJR089W, YML085C, YKL089W, YNL273W, YDL003W, YPL269W, YDL126C, YOR058C, YDL008W, YKL052C, YJL019W, YEL061C, YCL029C, YBL031W, YDR130C, YKR088C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YGR140W, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YKR054C, YJR068W, YGL216W
mitotic sister chromatid segregation	31 of 71 genes, 43.7%	137 of 7166 genes, 1.9%	5.70e-33	0.00%	0.00	YHR129C, YBL097W, YKR083C, YGL190C, YJR089W, YCL029C, YML085C, YBL031W, YDR130C, YKR088C, YKL089W, YNL273W, YGR098C, YIL031W, YDL003W, YPL269W, YFR027W, YHR127W, YOR058C, YDL008W, YGR140W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YKL052C, YIL144W, YJL019W, YEL061C, YGL216W

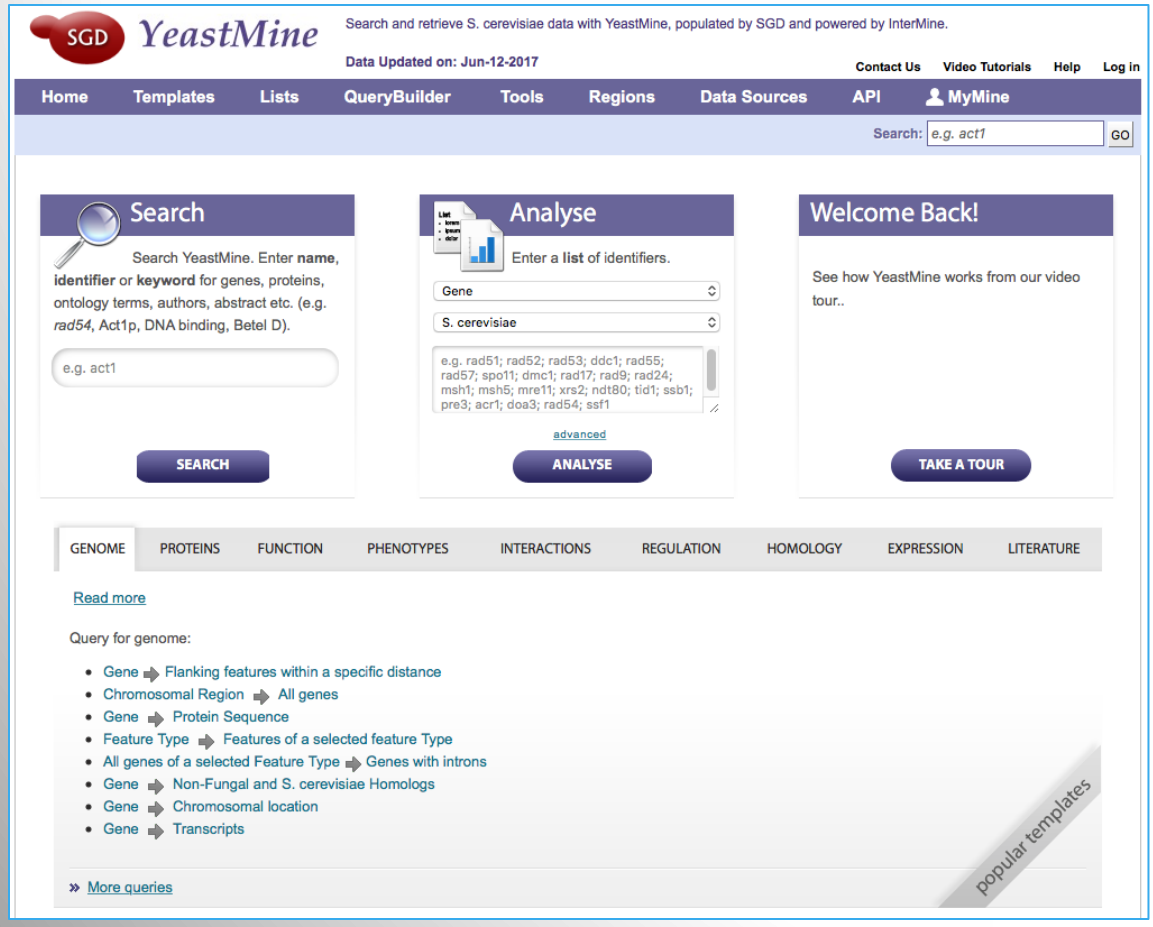
Batch GTF Results

- Ordered by statistical significance
- Save results as HTML, plain text, or as tab-delimited file
- GO tree view displayed based on annotated location

SGD Saccharomyces GENOME DATABASE

YeastMine

A multifaceted search and retrieval environment that provides access to diverse data types. Initiate searches, with a gene, or list of genes. Results can be combined for further analysis and saved or downloaded in customizable file formats.

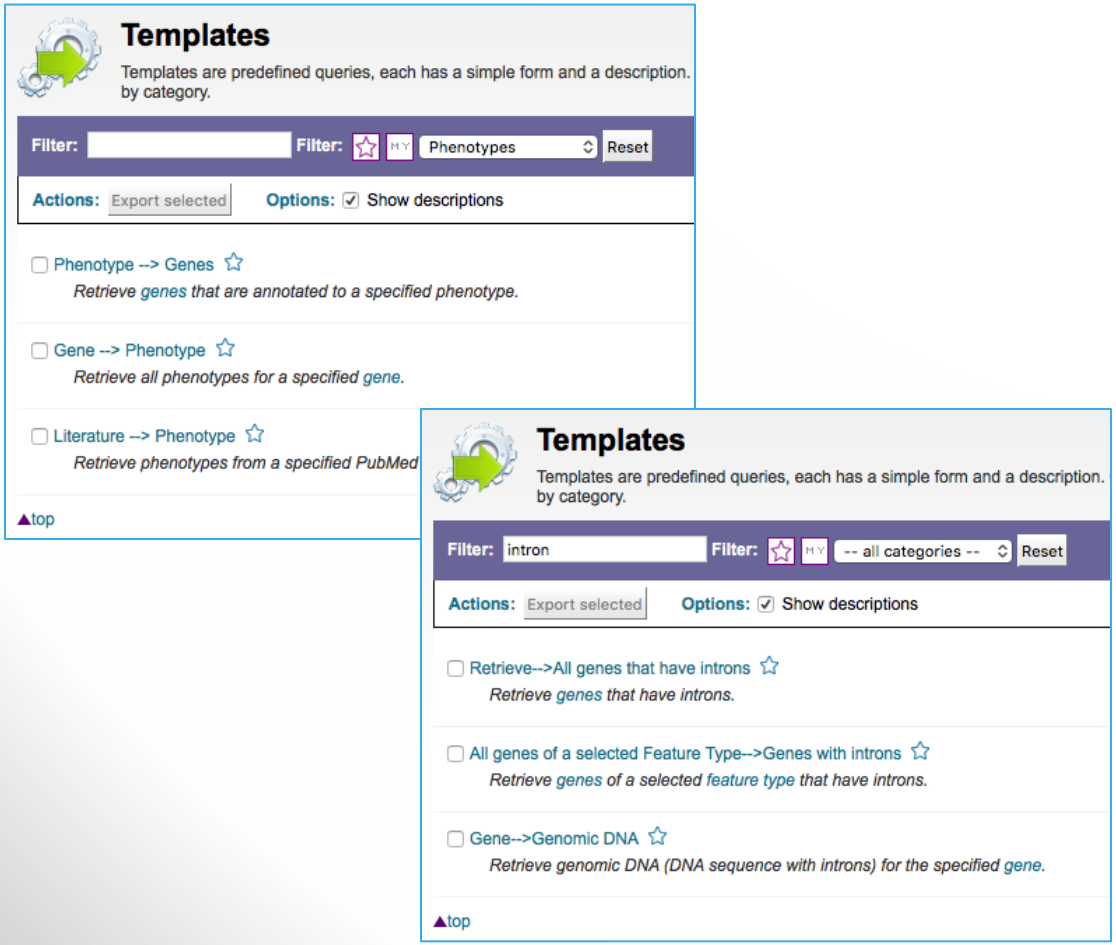


Saccharomyces GENOME DATABASE

Basic features

Templates are predefined queries. Filter by category:

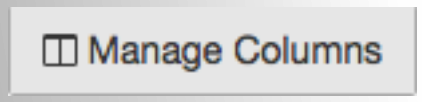
- Genome
- Proteins
- Function
- Phenotypes
- Interactions
- Literature
- Expression
- Regulation
- Homology



or keyword:

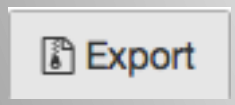
- “intron”
- “sequence”
- “UTR”

Template results page



Re-arrange, and/remove columns, change sort order

Save items, such as genes in list



Download results in different formats

Navigation aids

Column sort

Remove column

Toggle column visibility

Filter by values in column

View column summary

SGD Saccharomyces GENOME DATABASE

Lists and list operations

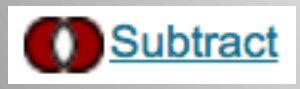
- List creation
  - Create, Save as List, Pick items from table
  - Add to List
  - Name and add description
  - Rename, share with MyMine

- List operations

Intersection (DNA replication AND DNA repair or genes on ChrIV, that are inviable when deleted)

Union (DNA replication and/or DNA repair, two interactions, etc.)

sets of

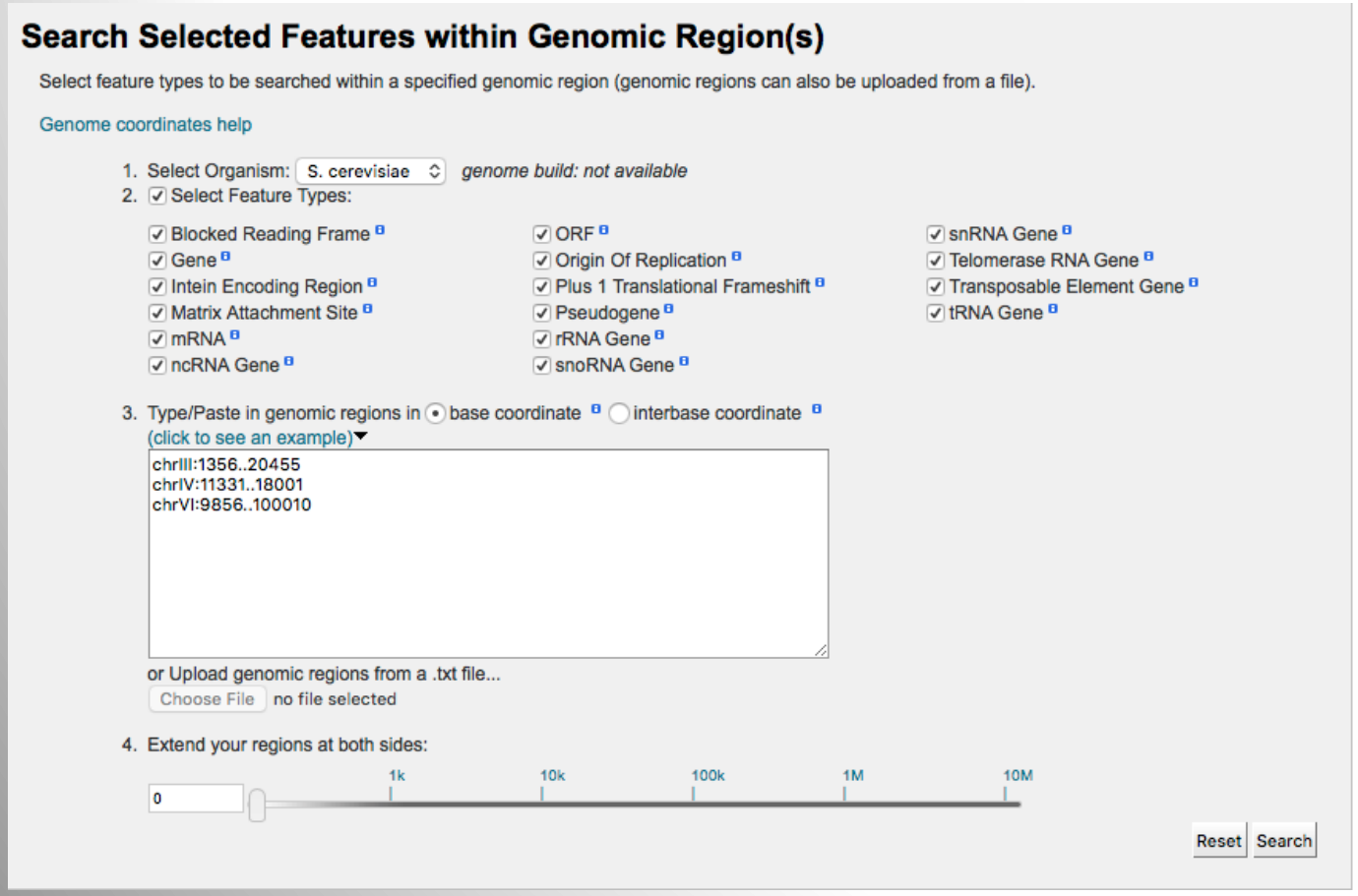


Subtract (DNA replication or DNA repair)

Asymmetric diff. (DNA replication minus repair; DNA repair minus replication)

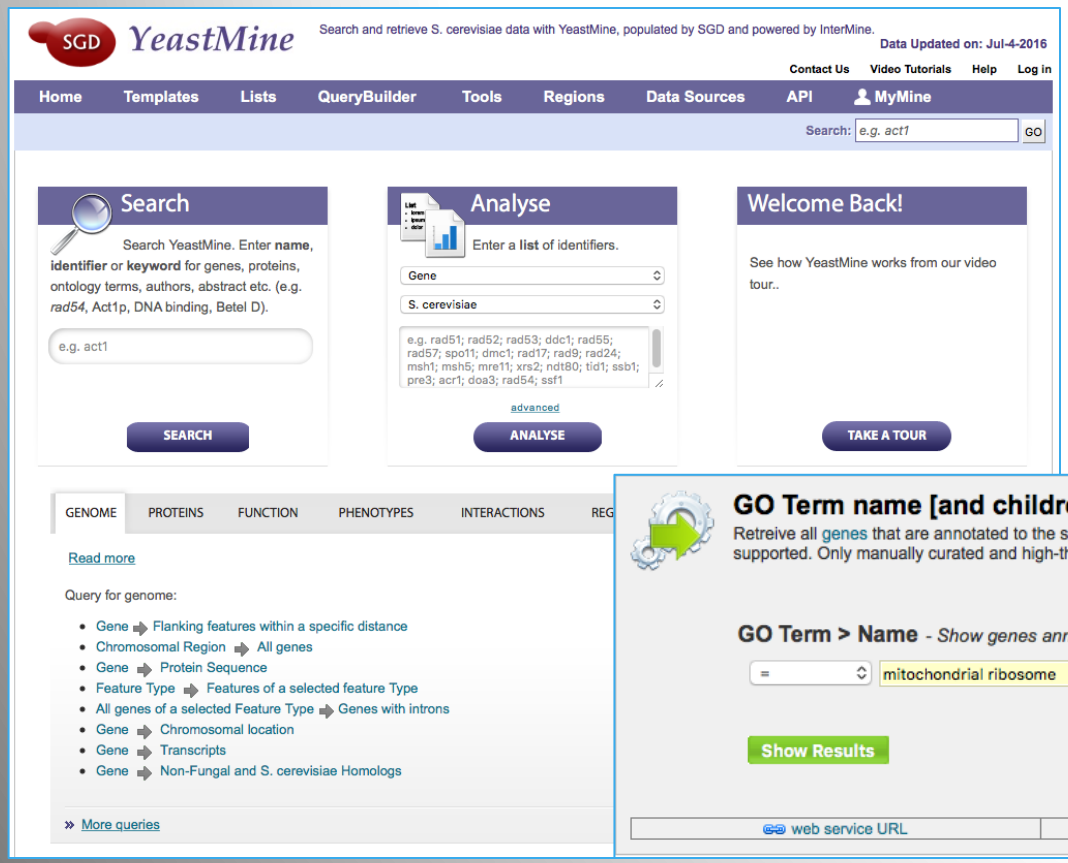
Regions tab

Select feature types to be searched within a specified genomic region (or upload from a file).



Use case: finding novel mitoribosomal proteins

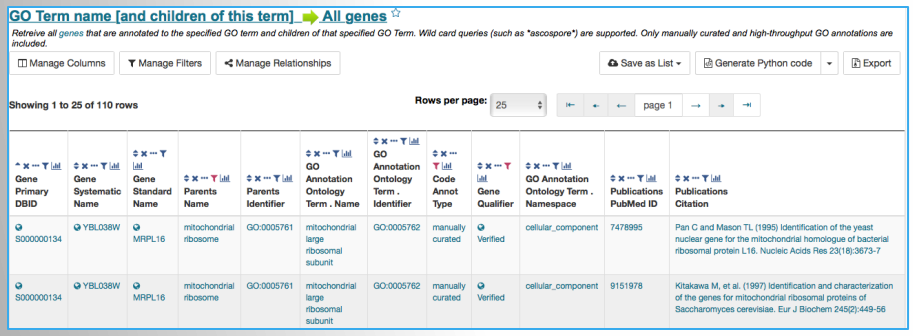
I'm interested in the mitochondrial ribosome. Does it have any as-yet-undiscovered subunits?

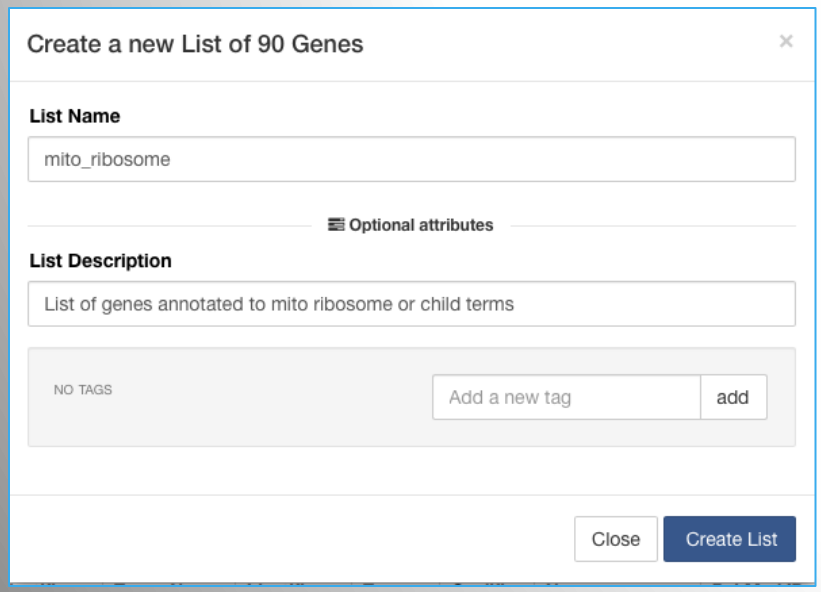


1. Find the known mitochondrial ribosomal proteins using YeastMine

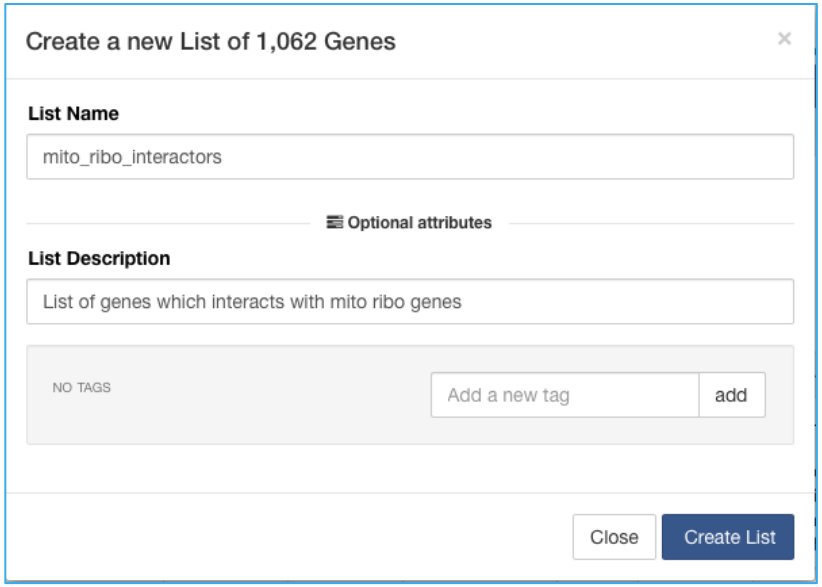
Saccharomyces GENOME DATABASE

2. Create a list of the results (90 genes)





3. Look for genes/proteins that interact with mt_ribosomal proteins

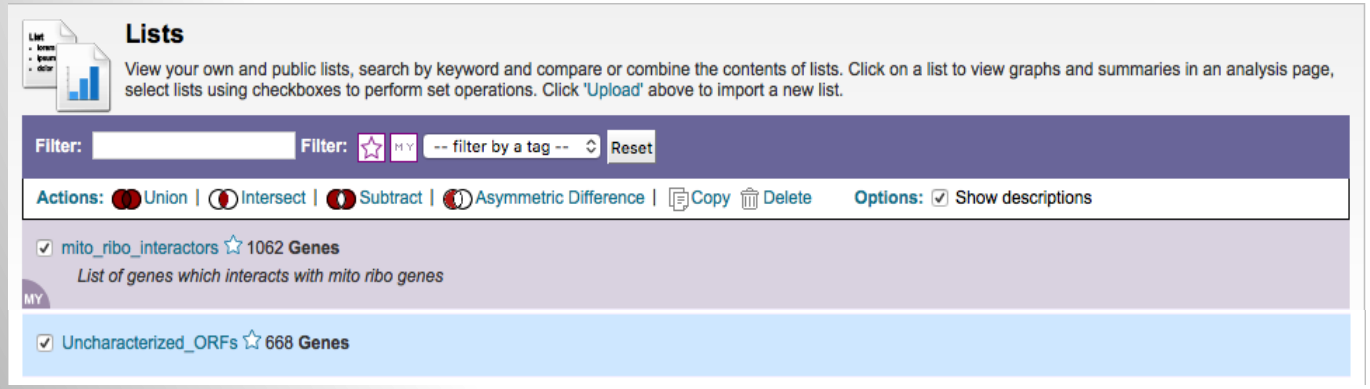


4. Create a list of 1,062 interacting genes/proteins.

SGD Saccharomyces GENOME DATABASE

Are any of the interacting genes/proteins uncharacterized?

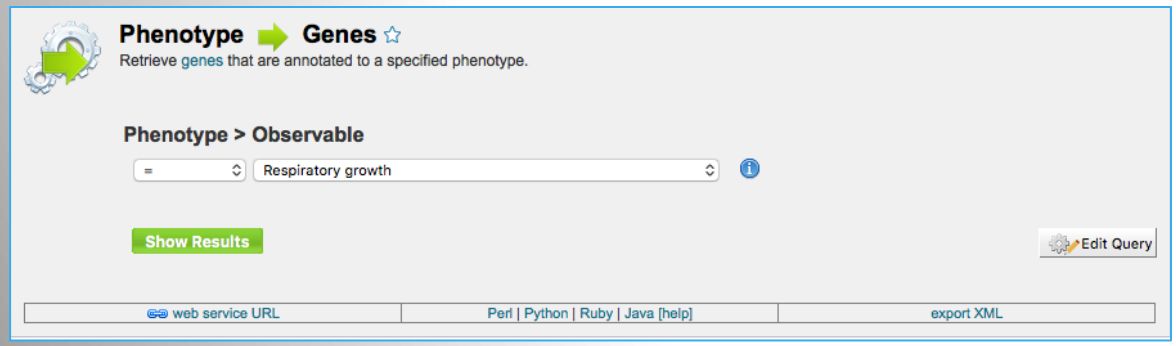
Determine the intersection between the pre-composed list of uncharacterized genes and the list of mitochondrial ribosome-interacting genes



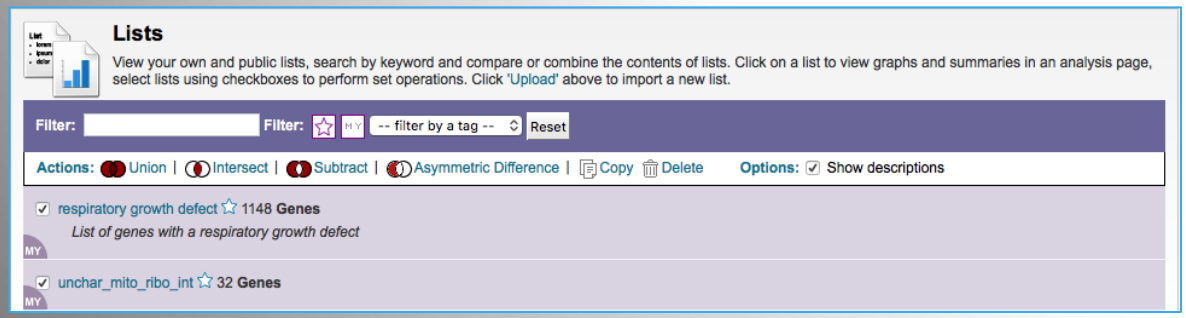
32 genes are uncharacterized

32 uncharacterized ORFs interact genetically or physically with known mitochondrial ribosomal proteins.

Mutation of a mt ribosomal subunit would block respiratory growth. Do any of these 32 genes exhibit this mutant phenotype?



- create list of genes that confer a respiratory phenotype



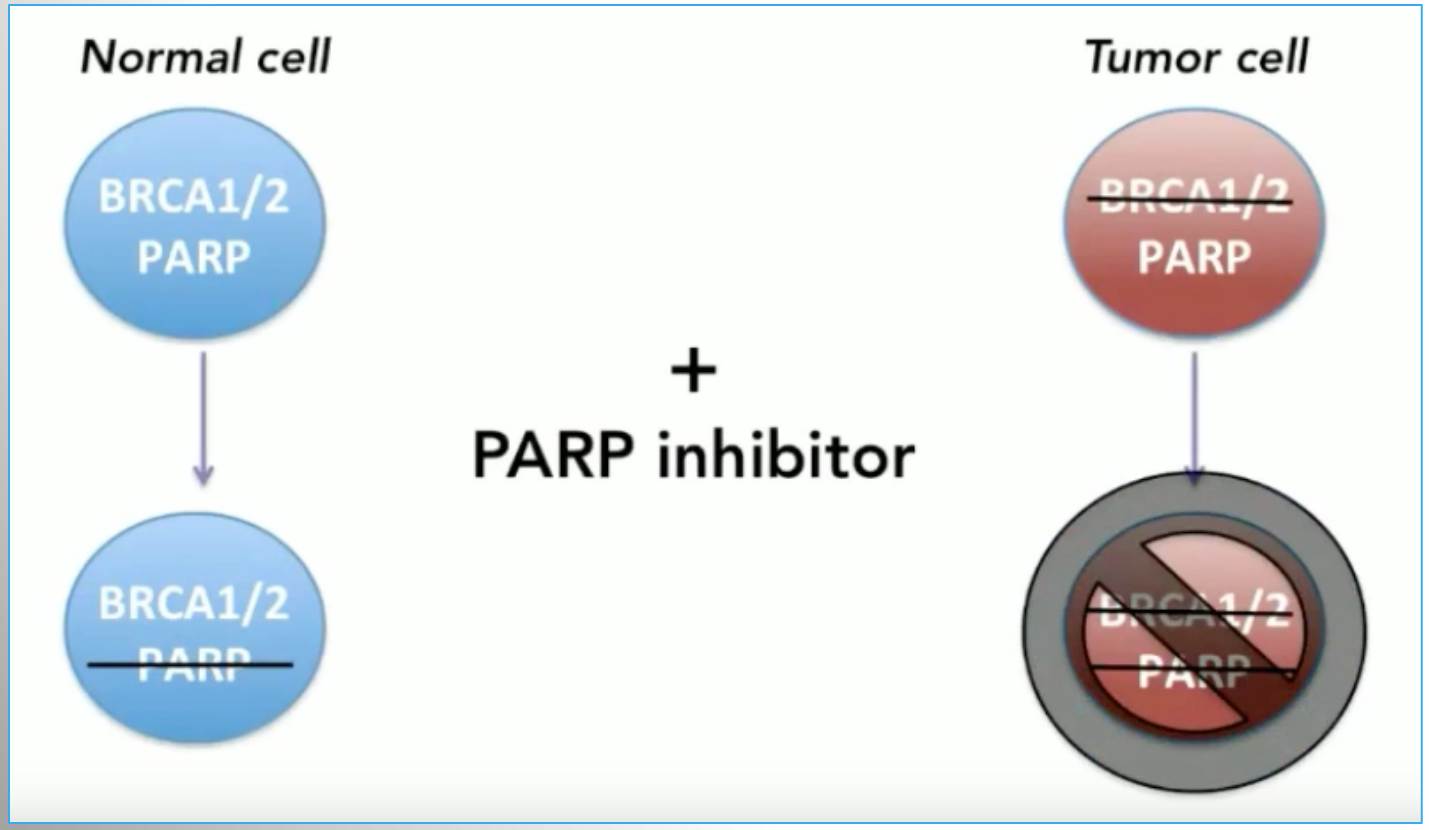
- find the intersection with the list of 32 uncharacterized ORFs

Three uncharacterized ORFs exhibit genetic or physical interactions with known mt ribosomal proteins AND block respiratory growth when mutated

Systematic name	Gene name	Name Description	Description
YBL095W	MRX3	Mitochondrial organization of gene expression	Protein that associates with mitochondrial ribosome; likely functions in cristae junction formation; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
YDL157C			Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
YPR109W			Predicted membrane protein; SWAT-GFP and mCherry fusion proteins localize to the endoplasmic reticulum; diploid deletion strain has high budding index

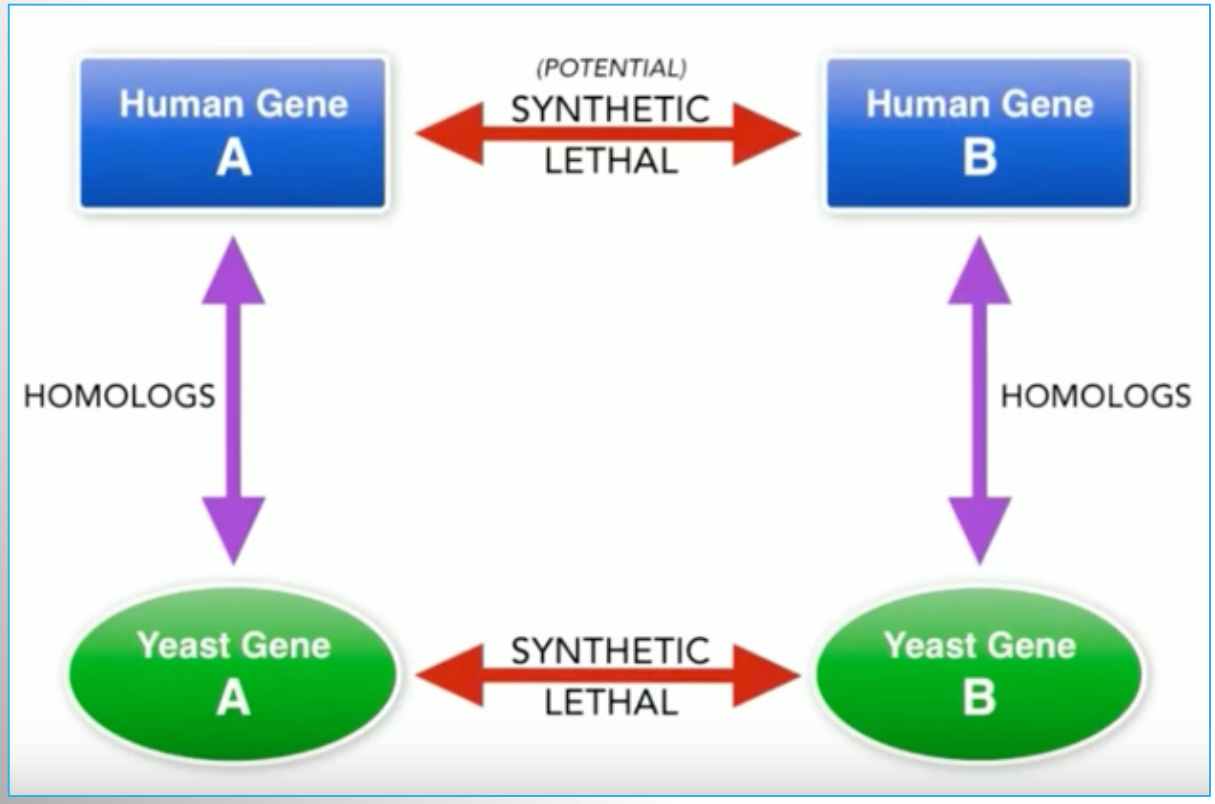
Predicting chemotherapy targets

Using yeast human homology data human to predict synthetic lethal interactions in the human genome that can be exploited for chemotherapy

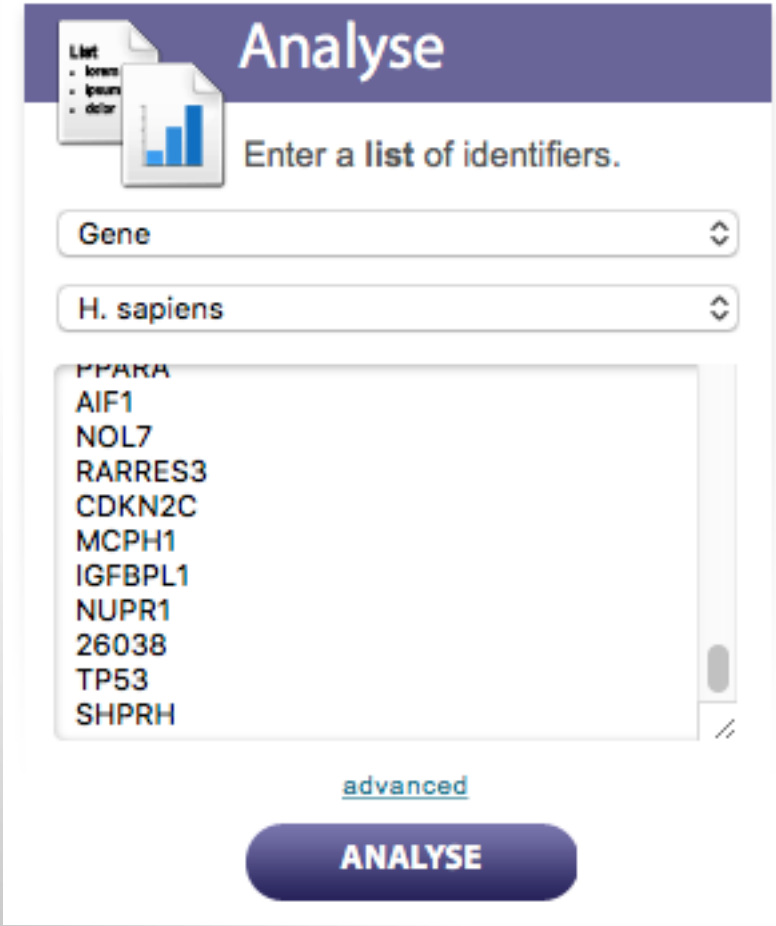


Predicting chemotherapy targets

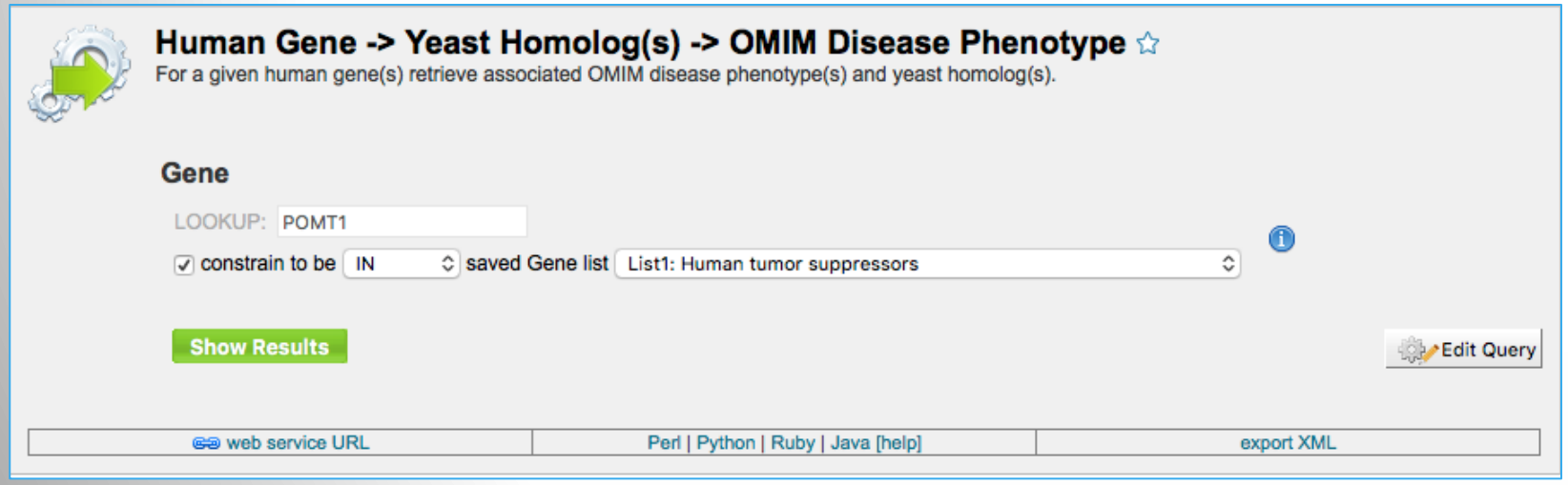
Synthetic lethality: loss of two non-essential genes = inviability

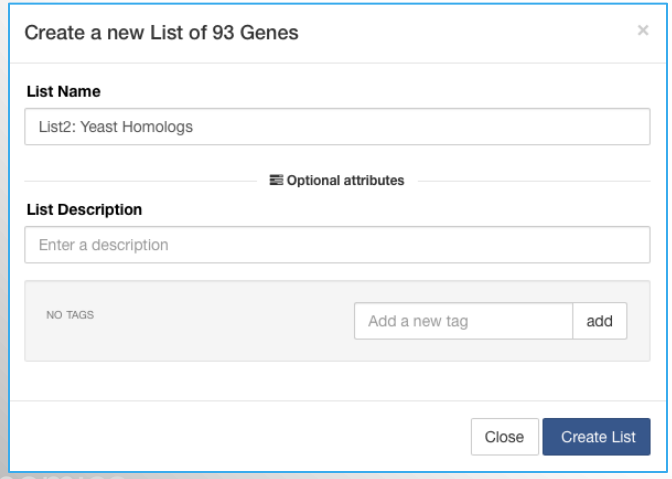


Step 1: Create human gene list



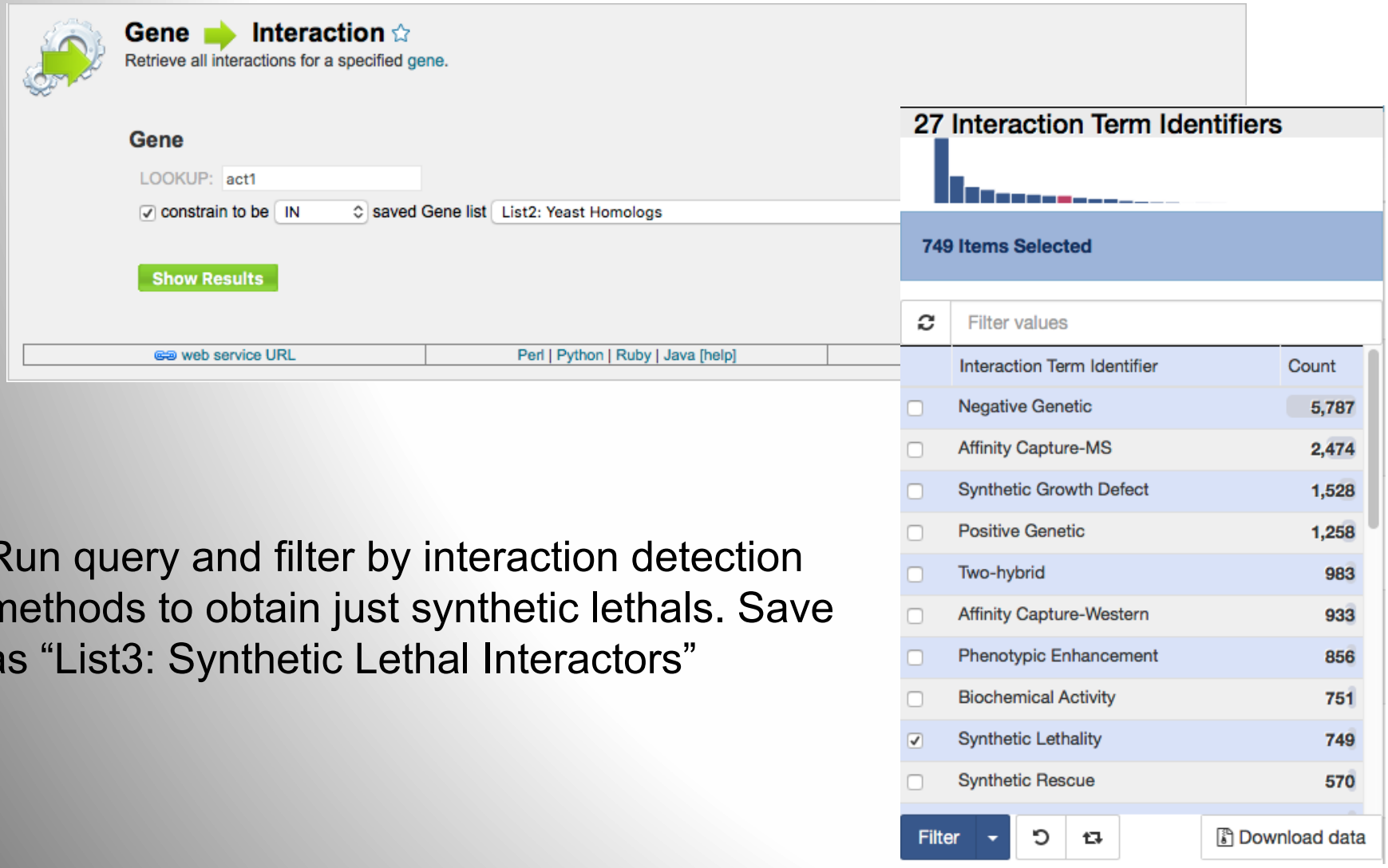
Step 2: Find yeast homologs & save yeast genes





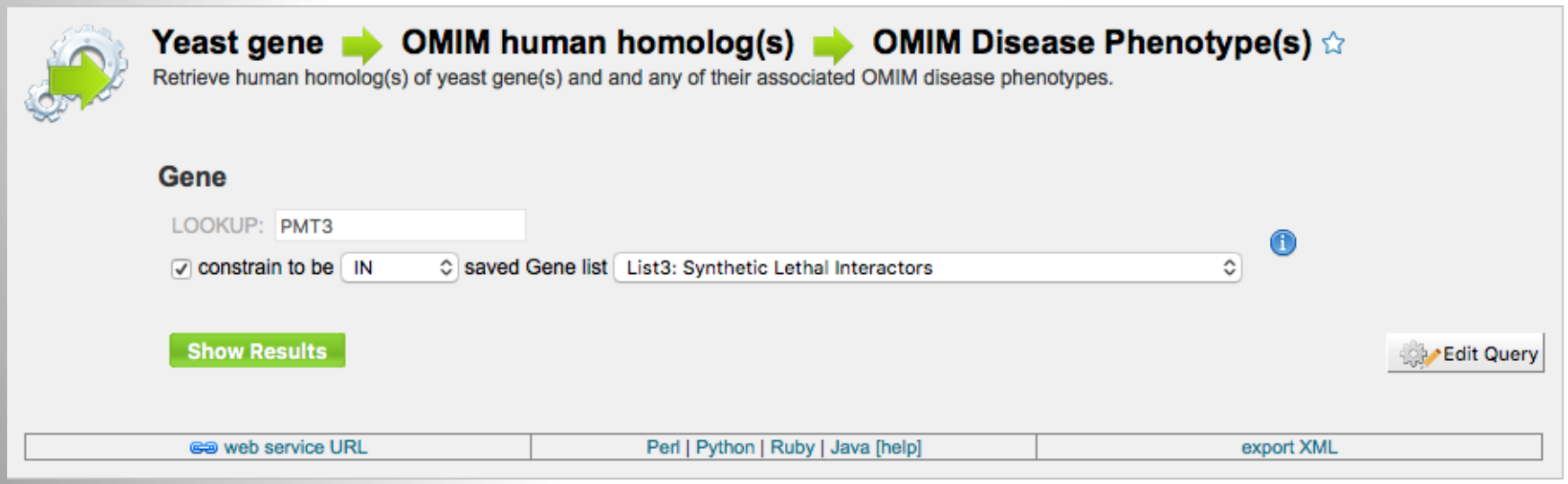
SGD Saccharomyces GENOME DATABASE

Step 3: ID synthetic lethal interactors



Run query and filter by interaction detection methods to obtain just synthetic lethals. Save as “List3: Synthetic Lethal Interactors”

Step 4: ID human homologs of SL interactors



Run query with SL interactors and then save list of human homologs as “List4: Human homologs of yeast SL Interactors”

YeastMine Scenario:

Predicting potential chemotherapy targets

S. cerevisiae			*H. sapiens*
MEC1	YBR136W	Mitosis Entry Checkpoint	ATR
POL3	YDL102W	POLymerase	POLD1

Human ATR and POLD1 potentially share a synthetic lethal interaction.

There is evidence that ATR-POLD1 have a SL interaction!

A synthetic lethal screen identifies ATR-inhibition as a novel therapeutic approach for POLD1-deficient cancers

Recent paper characterizes just such a synthetic lethal interaction, and POLD1 deficient cancers could be selectively killed by treatment with ATR inhibitors!

www.impactjournals.com/oncotarget/ Oncotarget, Vol. 7, No. 6

A synthetic lethal screen identifies ATR-inhibition as a novel therapeutic approach for POLD1-deficient cancers

Sandra Hocke[1], Yang Guo[1], Albert Job[2], Michael Orth[3], Andreas Ziesch[1], Kirsten Lauber[3], Enrico N. De Toni[1], Thomas M. Gress[2], Andreas Herbst[1], Burkhard Göke[1], Eike Gallmeier[1,2]

[1]Department of Medicine II, Ludwig-Maximilians-University of Munich, 81377 Munich, Germany

[2]Department of Gastroenterology, Endocrinology and Metabolism, University Hospital of Marburg, Philipps-University of Marburg, 35043 Marburg, Germany

[3]Department of Radiotherapy and Radiation Oncology, Ludwig-Maximilians-University of Munich, 81377 Munich, Germany

Correspondence to: *Eike Gallmeier,* ***e-mail:*** *Eike.Gallmeier@med.uni-marburg.de*

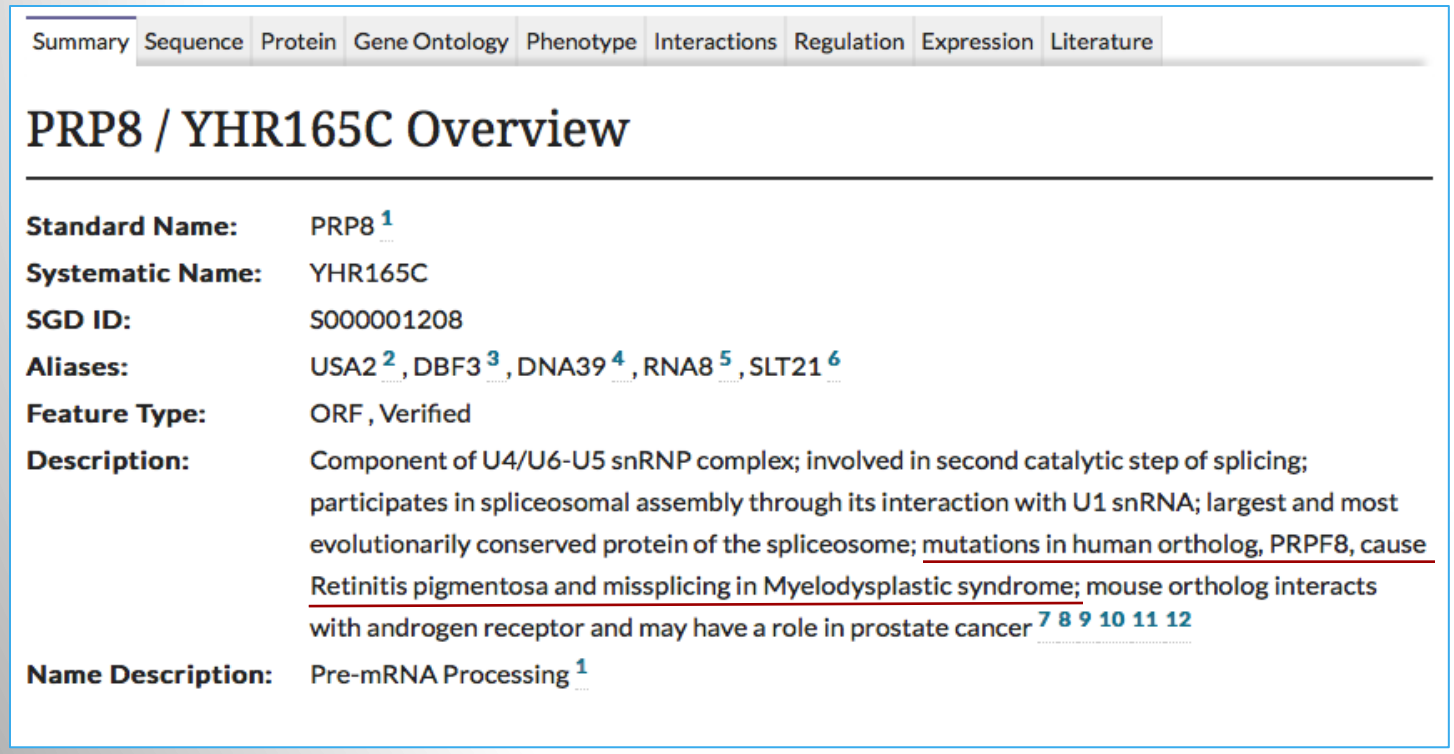
Keywords: *ATR, POLD1, synthetic lethality, DNA repair, targeted therapy*

Received: *August 04, 2015* ***Accepted:*** *January 01, 2016* ***Published:*** *January 09, 2016*

ABSTRACT

The phosphoinositide 3-kinase-related kinase ATR represents a central checkpoint regulator and mediator of DNA-repair. Its inhibition selectively eliminates certain subsets of cancer cells in various tumor types, but the underlying genetic determinants remain enigmatic. Here, we applied a synthetic lethal screen directed against 288 DNA-repair genes using the well-defined *ATR* knock-in model of DLD1 colorectal cancer cells to identify potential DNA-repair defects mediating these effects. We identified a set of DNA-repair proteins, whose knockdown selectively killed *ATR*-deficient cancer cells. From this set, we further investigated the profound synthetic lethal interaction between *ATR* and *POLD1*. *ATR*-dependent *POLD1* knockdown-induced cell killing was reproducible pharmacologically in POLD1-depleted DLD1 cells and a panel of other colorectal cancer cell lines by using chemical inhibitors of ATR or its major effector kinase CHK1. Mechanistically, POLD1 depletion in *ATR*-deficient cells caused caspase-dependent apoptosis without preceding cell cycle arrest and increased DNA-damage along with impaired DNA-repair. Our data could have clinical implications regarding tumor genotype-based cancer therapy, as inactivating *POLD1* mutations have recently been identified in small subsets of colorectal and endometrial cancers. *POLD1* deficiency might thus represent a predictive marker for treatment response towards ATR- or CHK1-inhibitors that are currently tested in clinical trials.

Explore a gene PRP8



Summary | Sequence | Protein | Gene Ontology | Phenotype | Interactions | Regulation | Expression | Literature

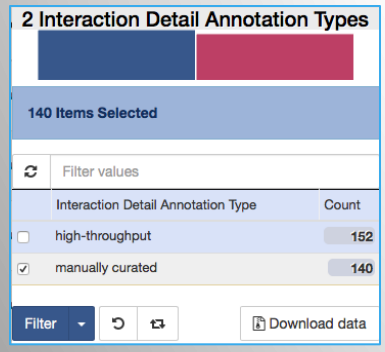
PRP8 / YHR165C Overview

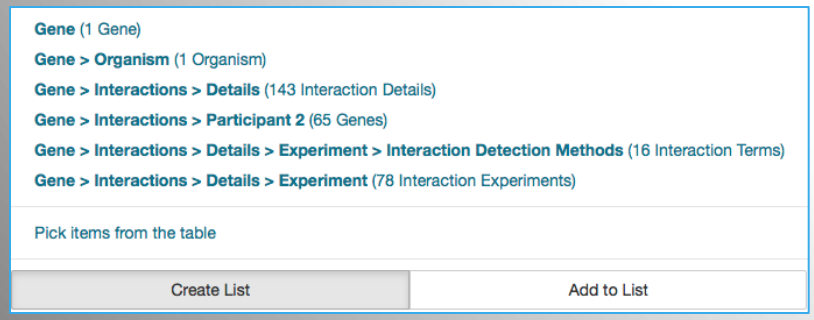
Standard Name:	PRP8 [1]
Systematic Name:	YHR165C
SGD ID:	S000001208
Aliases:	USA2 [2], DBF3 [3], DNA39 [4], RNA8 [5], SLT21 [6]
Feature Type:	ORF, Verified
Description:	Component of U4/U6-U5 snRNP complex; involved in second catalytic step of splicing; participates in spliceosomal assembly through its interaction with U1 snRNA; largest and most evolutionarily conserved protein of the spliceosome; mutations in human ortholog, PRPF8, cause Retinitis pigmentosa and missplicing in Myelodysplastic syndrome; mouse ortholog interacts with androgen receptor and may have a role in prostate cancer [7 8 9 10 11 12]
Name Description:	Pre-mRNA Processing [1]

1. Identify PRP8 interactors
2. Use OMIM to ID yeast orthologs of human genes involved in retinitis pigmentosa
3. Intersect the two lists to identify PRP8 interactors with orthologs involved in RP

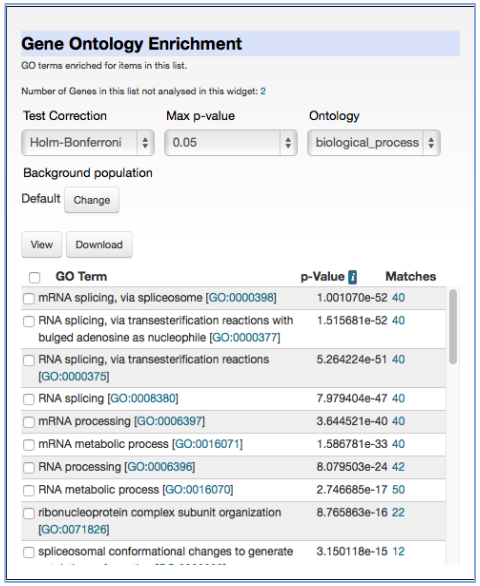
1. Select template “Gene -> Interaction”, enter “PRP8” and show results

2. Select manual annotations only by filtering and save list of interacting genes/proteins



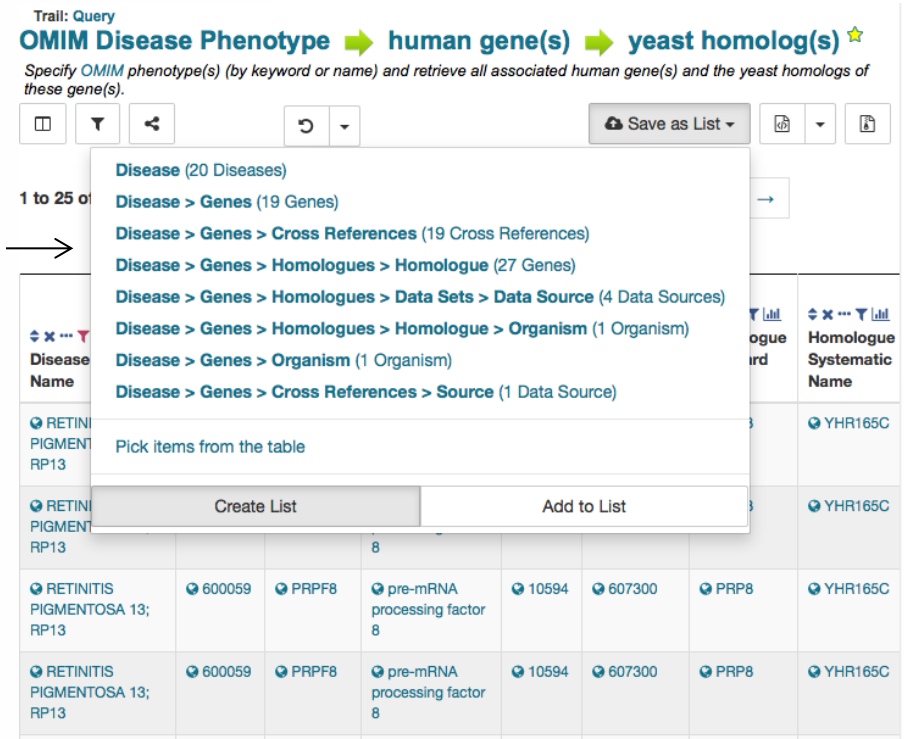


3. View enrichment



4. Go from human disease to genes to orthologs with “OMIM Disease Phenotype -> human gene(s) -> yeast homolog(s)” and enter “retinitis pigmentosa”

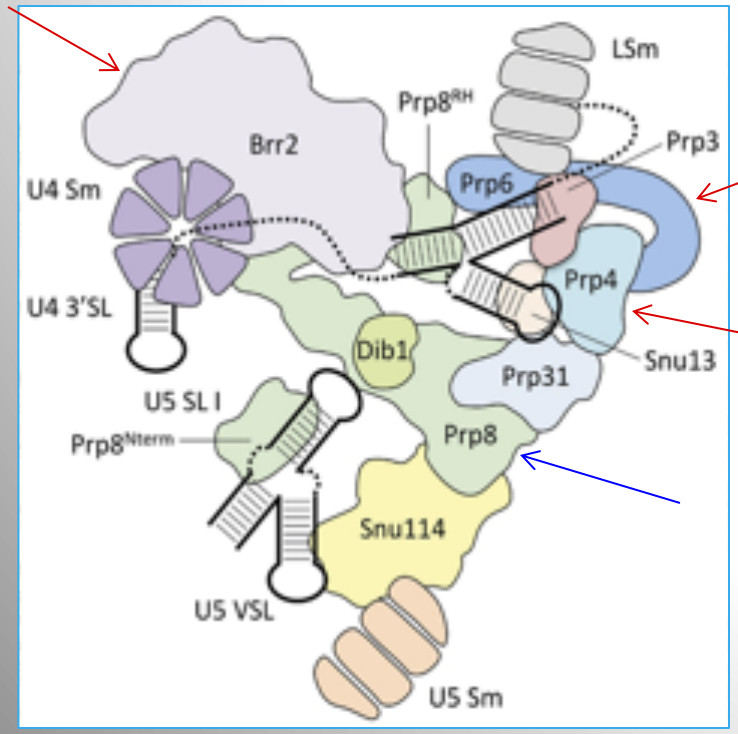
6. Create a second list of yeast orthologs of human genes associated with RP



5. Perform an inverse selection using column summary to remove “LEBER CONGENITAL …“

7. Now intersect the two lists (PRP8 interactors and RP orthologs)

Gene Primary DBID	Gene Systematic Name	Gene Organism . Short Name	Gene Standard Name	Gene Name
S000000259	YBR055C	S. cerevisiae	PRP6	Pre-mRNA Processing
S000000974	YER172C	S. cerevisiae	BRR2	Bad Response to Refrigeration
S000002881	YDR473C	S. cerevisiae	PRP3	Pre-mRNA Processing
S000006382	YPR178W	S. cerevisiae	PRP4	Pre-mRNA Processing



In fact, there is evidence that these 4 proteins are associated with Prp8p, as part of the U4/U6-U5 tri-snRNP spliceosome complex!

Prp8p inhibits Brr2p but not RP mutants

Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8

Sina Mozaffari-Jovin,[1]* Traudy Wandersleben,[2]* Karine F. Santos,[2]* Cindy L. Will,[1] Reinhard Lührmann,[1]† Markus C. Wahl[2]†

The Ski2-like RNA helicase Brr2 is a core component of the spliceosome that must be tightly regulated to ensure correct timing of spliceosome activation. Little is known about mechanisms of regulation of Ski2-like helicases by protein cofactors. Here we show by crystal structure and biochemical analyses that the Prp8 protein, a major regulator of the spliceosome, can insert its C-terminal tail into Brr2's RNA-binding tunnel, thereby intermittently blocking Brr2's RNA-binding, adenosine triphosphatase, and U4/U6 unwinding activities. Inefficient Brr2 repression is the only recognizable phenotype associated with certain retinitis pigmentosa–linked Prp8 mutations that map to its C-terminal tail. Our data show how a Ski2-like RNA helicase can be reversibly inhibited by a protein cofactor that directly competes with RNA substrate binding.

For each round of pre-mRNA splicing, a spliceosome is assembled, catalytically activated, and, after splicing catalysis, disassembled (*1*). During spliceosome activation, the U5 small nuclear ribonucleoprotein (snRNP) protein, Brr2, unwinds U4/U6 di-snRNAs, allowing U6 to base-pair with U2 and the 5′ splice site and a catalytically important U6 internal stem-loop to form (*2*–*4*). Additional requirements for Brr2 during splicing catalysis (*5*) and spliceosome disassembly (*6*) are independent of its adenosine triphosphatase (ATPase) and helicase activities (*5*, *7*), suggesting that after spliceosome activation, Brr2 must be repressed. Brr2 must also be silenced in the U4/U6-U5 tri-snRNP, where it encounters its U4/U6 substrate before association with the spliceosome. The U5 snRNP proteins Prp8 and Snu114 interact with Brr2 and modulate its activity (*6*, *8*, *9*). A C-terminal Jab1/MPN (Jab1) domain of Prp8 interacts directly with Brr2 (*10*–*13*), and many mutations leading to a severe form of retinitis pigmentosa (RP13) in humans (*14*, *15*) cluster in the C terminus of this domain (*16*, *17*).

We determined the crystal structure of a fragment of human (h) Brr2 comprising its helicase region ($Brr2^{HR}$) with tandem helicase cassettes (*18*) in complex with $hPrp8^{Jab1}$ at 3.6 Å resolution (fig. S1 and table S1) (*19*). $hPrp8^{Jab1}$ directly interacts with all six domains of the N-terminal hBrr2 cassette but does not contact the C-terminal cassette (Fig. 1A and fig. S2). One flank of

[1]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. [2]Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany.

*These authors contributed equally to this work.
†Corresponding author. E-mail: reinhard.luehrmann@mpi-bpc.mpg.de (R.L.); mwahl@zedat.fu-berlin.de (M.C.W.)

Constructing queries with QueryBuilder #1

User interested in the # of protein-coding and RNA genes on each one of the two strands of all yeast chromosomes?

Start at the tab QueryBuilder

- select data type: sequence feature
- constrain to type " ORF" or to be in premade list "ALL Verified Uncharacterized Dubious ORFs"
- constrain qualifier so not equal to dubious
- constrain status to be equal to active
- under "sequence feature" show: Primary DBID, Secondary ID, Standard name, SGD alias, Description, and Feature Type
- under "chromosome" show: Identifier and Organism – name
- under "chromosomal location" show: Strand

Queries

- Can use “save query” to name it for future use
- Can “show results” to run the query and reorder columns etc or after naming it can use action “Run”
- Saved query can be run later, edited or exported (shared for others such as a colleague)

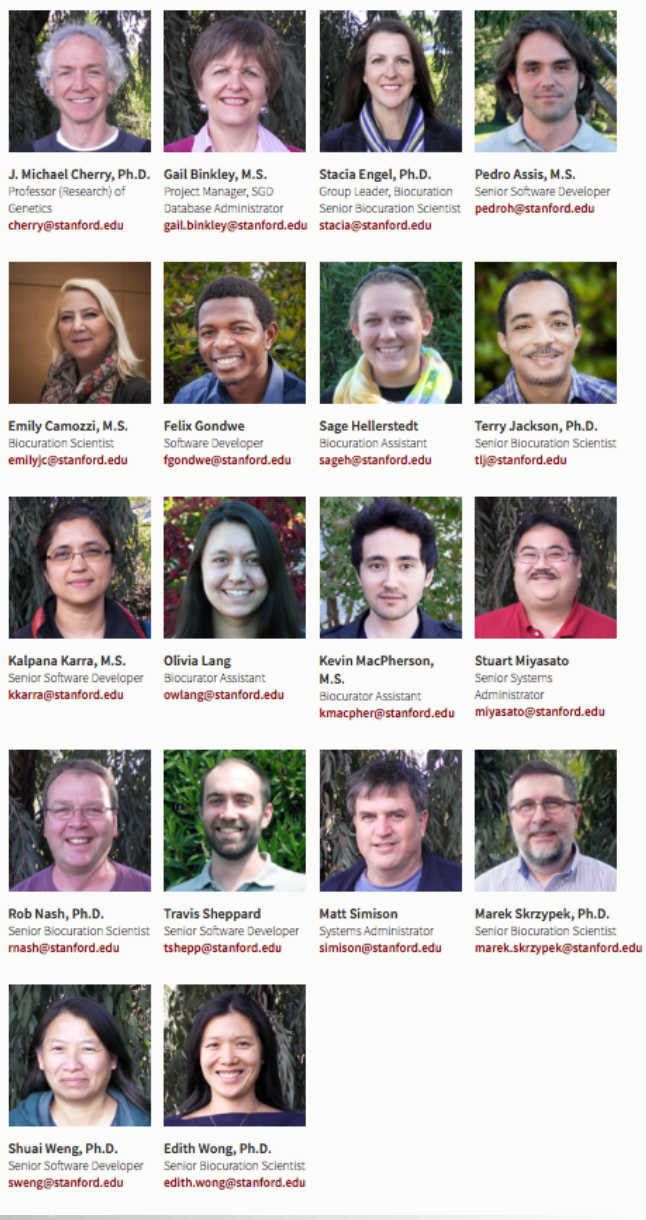
Constructing queries with QueryBuilder #2

How about the # of RNA genes on each one of the two strands of all yeast chromosomes?

Start at the tab Querybuilder

- select data type: sequence feature
  - constrain by feature type; since all RNA genes (rRNA_gene, tRNA_gene etc.) contain “RNA_gene” use “contains” and add text
  - no need to constrain qualifier for dubious as this is ORF specific
  - constrain status = active
  - under “sequence feature” show: primary DBID, Secondary ID, standard name, SGD alias, description, and feature type
- under “chromosome”, show identifier and under “organism” – name
- under “chromosomal location” show Strand

SGD Staff



J. Michael Cherry, Ph.D.
Professor (Research) of Genetics
cherry@stanford.edu

Gail Binkley, M.S.
Project Manager, SGD
Database Administrator
gail.binkley@stanford.edu

Stacia Engel, Ph.D.
Group Leader, Biocuration
Senior Biocuration Scientist
stacia@stanford.edu

Pedro Assis, M.S.
Senior Software Developer
pedroh@stanford.edu

Emily Camozzi, M.S.
Biocuration Scientist
emilyjc@stanford.edu

Felix Gondwe
Software Developer
fgondwe@stanford.edu

Sage Hellerstedt
Biocuration Assistant
sageh@stanford.edu

Terry Jackson, Ph.D.
Senior Biocuration Scientist
tlj@stanford.edu

Kalpana Karra, M.S.
Senior Software Developer
kkarra@stanford.edu

Olivia Lang
Biocurator Assistant
owlang@stanford.edu

Kevin MacPherson, M.S.
Biocurator Assistant
kmacpher@stanford.edu

Stuart Miyasato
Senior Systems Administrator
miyasato@stanford.edu

Rob Nash, Ph.D.
Senior Biocuration Scientist
rnash@stanford.edu

Travis Sheppard
Senior Software Developer
tshepp@stanford.edu

Matt Simison
Systems Administrator
simison@stanford.edu

Marek Skrzypek, Ph.D.
Senior Biocuration Scientist
marek.skrzypek@stanford.edu

Shuai Weng, Ph.D.
Senior Software Developer
sweng@stanford.edu

Edith Wong, Ph.D.
Senior Biocuration Scientist
edith.wong@stanford.edu

SGD Saccharomyces GENOME DATABASE