

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

About SGD

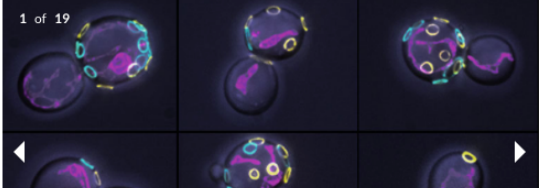
- Started by David Botstein in the early 90s based on the need to organize info on the genes of the budding yeast, *S. cerevisiae*.
- Mike Cherry P.I. of the project since 1992. On May 8th, 1994 the first external web hits were recorded
- Public, open, non-profit academic project funded by the NIH (NHGRI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

SGD

[About](#) [Blog](#) [Download](#) [Help](#) [YeastMine](#)    

 **Saccharomyces**
GENOME DATABASE

Analyze ▾ Sequence ▾ Function ▾ Literature ▾ Community ▾



1 of 19


Division rates of individual cells measured using TrackScar (yellow & cyan).
Image courtesy of C. Maxwell and P. Magwene, Duke University.
dx.doi.org/10.1111/mec.13955


About SGD


The *Saccharomyces* Genome Database (SGD) provides comprehensive integrated biological information for the budding yeast *Saccharomyces cerevisiae* along with search and analysis tools to explore these data, enabling the discovery of functional relationships between sequence and gene products in fungi and higher organisms.


Try this?

Meetings

[Yeast Genetics & Genomics Course](#) 
July 25, 2017 - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

[28th International Conference on Yeast Genetics and Molecular Biology \(ICYGMB\)](#) 
August 27, 2017 - Prague, Czech Republic
Deadline for late registration and abstract submission: July 14, 2017

[British Yeast Group Meeting – The Versatility of Yeasts](#) 
September 11, 2017 - University of Kent, Canterbury, UK

[Yeast Genetics Meeting](#) 
August 22, 2018 - Stanford University, Palo Alto, California



[View all meetings](#)

New & Noteworthy

[How Histones Use FACT\(s\) to Find Their Way](#)
07/05/2017
Some people (like me) have no sense of direction. Send me to the store and who knows where I'll end up! Tools like maps, a GPS system, and my iPhone all help to make sure I get to where I need to be. And seat belts, airbags and working brakes keep me safe while I am getting there. Histones are similar. These proteins, which help to organize and run our DNA, can get lost without a variety... [Read...](#)

[Yeast's Skynet Against Salt](#)
06/27/2017
In the Terminator franchise, the U.S. creates an artificial intelligence (AI)-based defense system called Skynet to, among other things, react more quickly to threats than any general or politician could. What starts out as an interesting idea almost dooms mankind to extinction once Skynet becomes conscious and decides to eliminate its greatest threat—humans. Our friend *Saccharomyces cerevisiae* has its own version of Skynet for when it is "attacked" by too many salt ions. No, the system... [Read...](#)

SGD About Blog Help Terms of Use
© Stanford University, Stanford, CA 94305.
























  **Stanford**
University

How to leverage data rich SGD!

Curated Data

Literature Curated Data

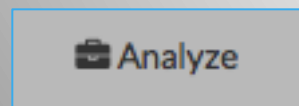
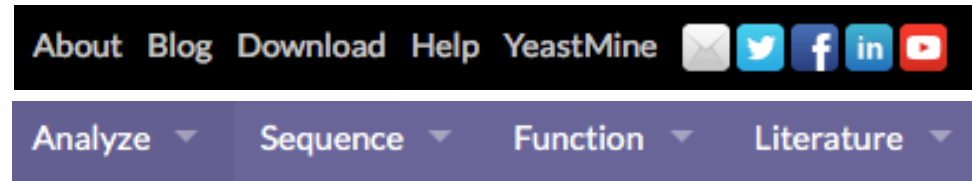
Information curated or otherwise assembled by the SGD staff.

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



GO Term Finder

Find common GO annotations between genes.

GO Slim Mapper

Sort genes into broad categories.

SPELL

View expression data.

YeastMine


Conduct advanced analysis.

...or analyze your own gene list

1 Upload list of identifiers

2 Verify identifier matches

List analysis



Create a new list

Select the type of list to create and either enter in a list of identifiers or upload identifiers from a file. A search will be performed for all the identifiers in your list.

- Separate identifiers by a **comma, space, tab, new line** or **semi-colon**.
- Qualify any identifiers that contain whitespace with double quotes like so: "even skipped".

Select Type:

for Organism:

Type/Paste in identifiers [\(click to see an example\)](#)

or Upload identifiers from a .txt file... no file selected

☐ **Match on case**

Presentation outline

- GO Slim Mapper
- GO Term Finder
- YeastMine

The Gene Ontology (GO) Project

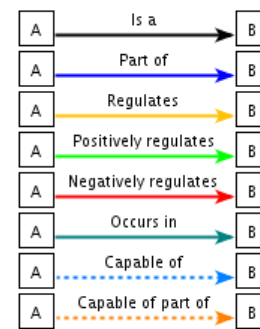
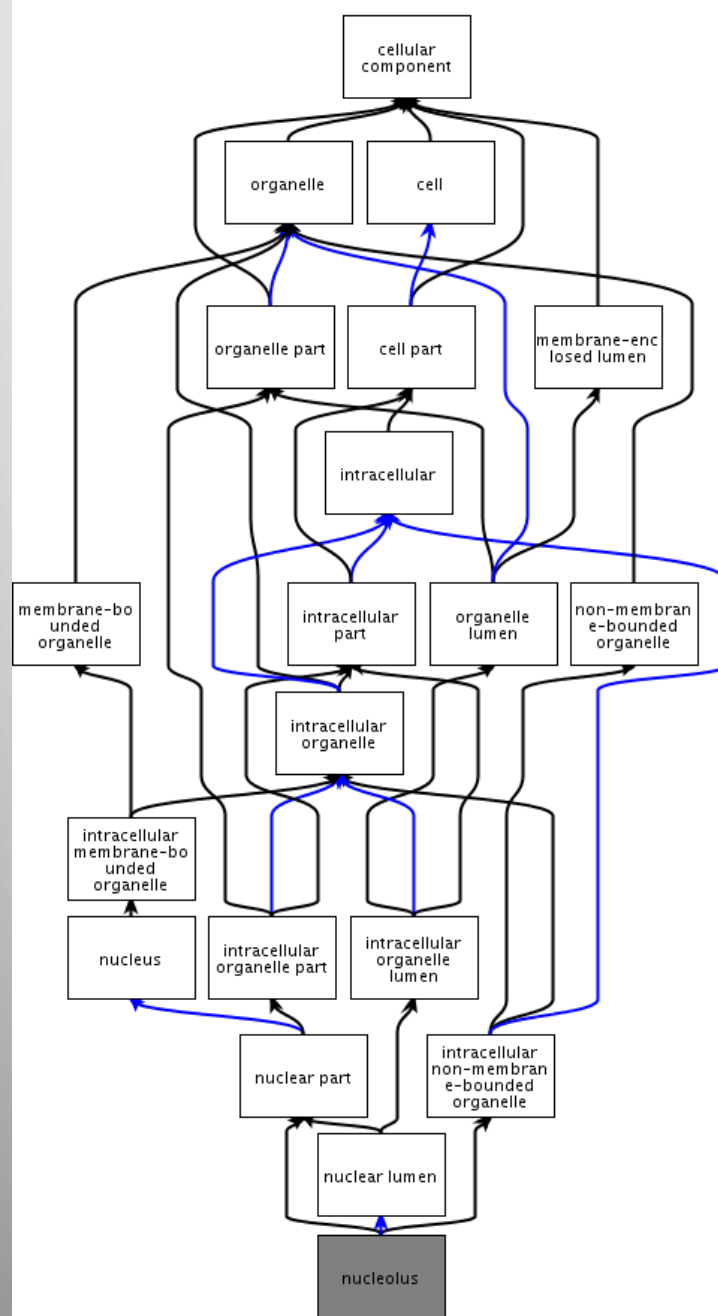
- A collaboration among MODs to improve queries within and across databases.
- Developed three structured ontologies to describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.
- The use of GO terms by collaborating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

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](#)

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](#)

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 ¹

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
YAL047C

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

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>

GENERIC GENE ONTOLOGY (GO) TERM FINDER

Welcome to the **GOTERMFINDER**, a tool for finding significant GO terms shared among a list of genes from your organism of choice, helping you discover what they may have in common.

This implementation, developed at the Lewis-Sigler Institute at Princeton, depends on the **GO-TermFinder** software written by Gavin Sherlock and Shuai Weng at Stanford University and the GO-View module written by Shuai Weng. It is made publicly available through the **GMOD project**. For more information, please see [Boyle et al. Bioinformatics \(2004\)](#).

New There is a new tool available with much the same functionality as GOTermFinder, only it is much more efficient. Although the backend has been tested extensively, please consider **LAGO** to be in beta at this time. Drop us a note to tell us what you think.

A previous job exists. If desired, you may [view the results](#) or [load the form options](#).

Required Basic Input Options [Help](#)

1. Enter a list of genes, one per line. [SGD sample gene list](#)

VAR007C
YAR019C
YBL031W
YBL063W
YBL097W
YBR170C
YCL029C
YCR089W
YDL003W
YDL008W

[CLEAR]

OR

Upload a file containing lists of genes: [Choose File](#) no file selected [CLEAR]
(10 MB max, plain text, one per line, gz ok)

This version of the Generic GO Term Finder has a batch mode to process multiple gene lists in parallel. To use this feature, upload an archive in tar, tar.gz, tgz, or zip format. Please refer to the [help document](#) for instructions.

Batch processing is done on one of our grid systems. Processing time will depend on the grid's load.

For long running jobs, it is highly recommended that an email address be provided so that notification may be sent when results are ready. If you do not do this, you will have to keep your browser window open to monitor the progress.
A notification email address is required for batch jobs.

Please enter a notification email address:
And please enter it again for confirmation:

2. Choose 1 of the 3 ontology aspects: ☐ Process ☒ Function ☐ Component

3. Choose annotation: [SGD - S. cerevisiae \(Yeast\)](#) (or upload your own in the advanced options)
Example identifiers: YJL166W S000003702 COR5 QC88
[Larger sample list](#)

4. Plain text will be produced. Choose additional output format(s): ☒ HTML table ☒ GO tree view images

Optional Advanced Input Options [Help](#)

Enter the number of products estimated for your organism (e.g. roughly 7000 for *Saccharomyces cerevisiae*):

OR provide a list of genes for the background population [Choose File](#) no file selected [CLEAR]
(100 MB max, plain text, one per line, gz ok)

Enter p-value cutoff for significant shared GO terms search (e.g. 0.01 is the default p-value cutoff)

☒ Bonferroni correction for p-values?

☒ Calculate false discovery rate (FDR)?

☒ Follow regulation links? (regulates, positively_regulates, negatively_regulates)

Enter URL for your organism (e.g. <http://db.yeastgenome.org/cgi-bin/SGD/locus.p?locus=> is the default url for *Saccharomyces cerevisiae*)

For batch mode, enter the extension for files containing gene lists (examples: list, txt)

Upload a custom gene association file [Choose File](#) no file selected [CLEAR]
(100 MB max, GAF format, gz ok)

Select **evidence codes** ☐ EXP ☐ IDA ☐ IMP ☐ IGI ☐ IEP ☐ ISS ☐ ISO ☐ ISA ☐ ISM ☐ IGC ☐ RCA ☐ TAS ☐ NAS ☐ IC ☐ ND to exclude: ☐ IEA Inferred from Electronic Annotation associations are included by default. Select this checkbox to exclude them.

Extra headers to include in text result files: ☐ Query parameters ☐ Duplicated, discarded, ambiguous, and unknown identifiers

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, YKL052C, YJL019W, YEL061C, YOR372C, YCL029C, YBL031W, YIL149C, YOR372C, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YAR007C, YDL139C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YIL144W, YJR088W, 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, YJR088W, YGL075C, 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, YJR088W, YGL075C, 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, YOR372C, YCL029C, YBL031W, YDR130C, YKR088C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YIL144W, YJR088W, YGL075C, 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, YKL052C, YJL019W, YEL061C, YOR372C, YCL029C, YBL031W, YIL149C, YOR372C, YCL029C, YPL124W, YNL188W, YBL031W, YDR130C, YKR088C, YAR007C, YDL139C, YGR098C, YIL031W, YFR027W, YHR127W, YOR373W, YBR170C, YGR140W, YMR043W, YBL063W, YDL028C, YGR113W, YML064C, YKR054C, YIL144W, YJR088W, 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, YJR088W, 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, YJR088W, 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

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.

The screenshot displays the YeastMine web application interface. At the top, the SGD YeastMine logo is on the left, and the text "Search and retrieve S. cerevisiae data with YeastMine, populated by SGD and powered by InterMine." is on the right. Below this, a navigation bar includes links for Home, Templates, Lists, QueryBuilder, Tools, Regions, Data Sources, API, and MyMine. A search bar on the right of the navigation bar contains the text "e.g. act1" and a "GO" button. The main content area is divided into three columns. The first column, titled "Search", contains a magnifying glass icon and a text input field with "e.g. act1" and a "SEARCH" button. The second column, titled "Analyse", contains a document icon and a text input field with "S. cerevisiae" and an "ANALYSE" button. The third column, titled "Welcome Back!", contains a video icon and a "TAKE A TOUR" button. Below these columns is a horizontal menu with tabs for GENOME, PROTEINS, FUNCTION, PHENOTYPES, INTERACTIONS, REGULATION, HOMOLOGY, EXPRESSION, and LITERATURE. The "GENOME" tab is selected, and a "Read more" link is visible. Below the menu, a "Query for genome:" section lists various search options with arrows indicating the type of query. A "popular templates" banner is visible in the bottom right corner of the main content area.

SGD YeastMine Search and retrieve S. cerevisiae data with YeastMine, populated by SGD and powered by InterMine.

Data Updated on: Jun-12-2017

Contact Us Video Tutorials Help Log in

Home Templates Lists QueryBuilder Tools Regions Data Sources API MyMine

Search: e.g. act1 GO

Search

Search YeastMine. Enter name, identifier or keyword for genes, proteins, ontology terms, authors, abstract etc. (e.g. rad54, Act1p, DNA binding, Betel D).

e.g. act1

SEARCH

Analyse

Enter a list of identifiers.

Gene S. cerevisiae

e.g. rad51; rad52; rad53; ddc1; rad55; rad57; spo11; dmc1; rad17; rad9; rad24; msh1; msh5; mre11; xrs2; ndt80; tid1; ssb1; pre3; acrl; doa3; rad54; ssf1

advanced

ANALYSE

Welcome Back!

See how YeastMine works from our video tour..

TAKE A TOUR

GENOME PROTEINS FUNCTION PHENOTYPES INTERACTIONS REGULATION HOMOLOGY EXPRESSION LITERATURE

[Read more](#)

Query for genome:

- Gene ➔ Flanking features within a specific distance
- Chromosomal Region ➔ All genes
- Gene ➔ Protein Sequence
- Feature Type ➔ Features of a selected feature Type
- All genes of a selected Feature Type ➔ Genes with introns
- Gene ➔ Non-Fungal and S. cerevisiae Homologs
- Gene ➔ Chromosomal location
- Gene ➔ Transcripts

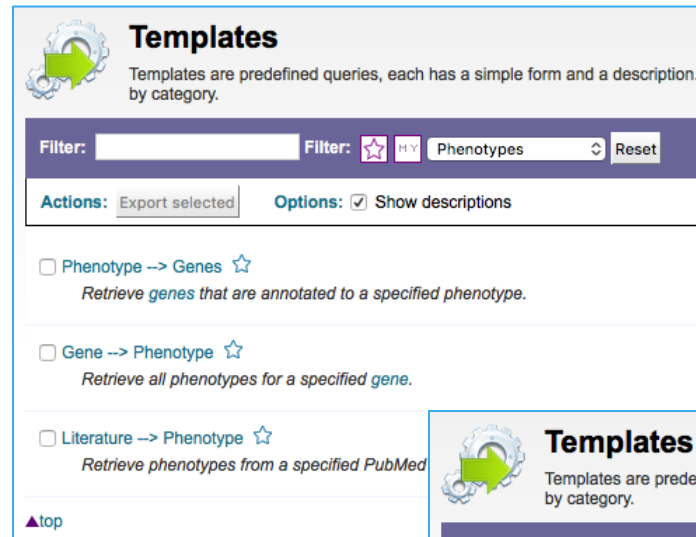
» [More queries](#)

popular templates

Basic features

Templates are predefined queries. Filter by category:

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



Templates
Templates are predefined queries, each has a simple form and a description. by category.

Filter: Filter: ☆ MY Phenotypes

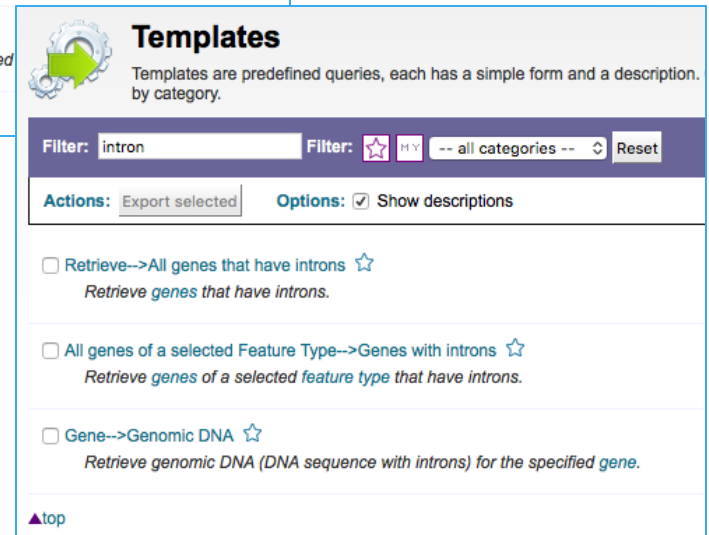
Actions: Options: ☒ Show descriptions

- ☐ Phenotype --> Genes ☆
Retrieve *genes* that are annotated to a specified *phenotype*.
- ☐ Gene --> Phenotype ☆
Retrieve all *phenotypes* for a specified *gene*.
- ☐ Literature --> Phenotype ☆
Retrieve *phenotypes* from a specified *PubMed*

▲top

or keyword:

- “intron”
- “sequence”
- “UTR”



Templates
Templates are predefined queries, each has a simple form and a description. by category.

Filter: Filter: ☆ MY -- all categories --

Actions: Options: ☒ Show descriptions


- ☐ Retrieve-->All genes that have introns ☆
Retrieve *genes* that have *introns*.
- ☐ All genes of a selected Feature Type-->Genes with introns ☆
Retrieve *genes* of a selected *feature type* that have *introns*.
- ☐ Gene-->Genomic DNA ☆
Retrieve *genomic DNA* (DNA sequence with *introns*) for the specified *gene*.

▲top


Template results page

 Manage Columns

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

 Save as List ▾

Save items, such as genes in list

 Export

Download results in different formats

Rows per page: 25



Navigation aids



Column sort



Remove column



Toggle column visibility



Filter by values in column



View column summary

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.)



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).

Search Selected Features within Genomic Region(s)

Select feature types to be searched within a specified genomic region (genomic regions can also be uploaded from a file).

[Genome coordinates help](#)

1. Select Organism: *genome build: not available*

2. ☒ Select Feature Types:

<input checked="" type="checkbox"/> Blocked Reading Frame [?]	<input checked="" type="checkbox"/> ORF [?]	<input checked="" type="checkbox"/> snRNA Gene [?]
<input checked="" type="checkbox"/> Gene [?]	<input checked="" type="checkbox"/> Origin Of Replication [?]	<input checked="" type="checkbox"/> Telomerase RNA Gene [?]
<input checked="" type="checkbox"/> Intein Encoding Region [?]	<input checked="" type="checkbox"/> Plus 1 Translational Frameshift [?]	<input checked="" type="checkbox"/> Transposable Element Gene [?]
<input checked="" type="checkbox"/> Matrix Attachment Site [?]	<input checked="" type="checkbox"/> Pseudogene [?]	<input checked="" type="checkbox"/> tRNA Gene [?]
<input checked="" type="checkbox"/> mRNA [?]	<input checked="" type="checkbox"/> rRNA Gene [?]	
<input checked="" type="checkbox"/> ncRNA Gene [?]	<input checked="" type="checkbox"/> snoRNA Gene [?]	


3. Type/Paste in genomic regions in ☒ base coordinate [?] ☐ interbase coordinate [?]
([click to see an example](#))▼

chrIII:1356..20455
chrIV:11331..18001
chrVI:9856..100010

or Upload genomic regions from a .txt file...

no file selected

4. Extend your regions at both sides:



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

SGD YeastMine Search and retrieve *S. cerevisiae* data with YeastMine, populated by SGD and powered by InterMine. Data Updated on: Jul-4-2016

Contact Us Video Tutorials Help Log in

Home Templates Lists QueryBuilder Tools Regions Data Sources API MyMine

Search: e.g. act1 GO

Search
Search YeastMine. Enter name, identifier or keyword for genes, proteins, ontology terms, authors, abstract etc. (e.g. *rad54*, *Act1p*, DNA binding, Betel D).
e.g. act1
SEARCH

Analyse
Enter a list of identifiers.
Gene
S. cerevisiae
e.g. *rad51*; *rad52*; *rad53*; *ddc1*; *rad55*; *rad57*; *spo11*; *dmc1*; *rad17*; *rad9*; *rad24*; *msh1*; *msh5*; *mre11*; *xrs2*; *ndt80*; *tid1*; *ssb1*; *pre3*; *acr1*; *doa3*; *rad54*; *ssb1*
advanced
ANALYSE

Welcome Back!
See how YeastMine works from our video tour..
TAKE A TOUR

GO Term name [and children of this term] ➡ All genes ☆
Retrieve all genes that are annotated to the specified GO term and children of that specified GO Term. Wild card queries (such as *ascospore*) are supported. Only manually curated and high-throughput GO annotations are included.

GO Term > Name - Show genes annotated with GO term (and any children of this GO term):
= mitochondrial ribosome
Show Results Edit Query

web service URL Perl | Python | Ruby | Java [help] export XML

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

GO Term name [and children of this term] → All genes ☆
Retrieve all genes that are annotated to the specified GO term and children of that specified GO Term. Wild card queries (such as "ascospore") are supported. Only manually curated and high-throughput GO annotations are included.

Manage Columns Manage Filters Manage Relationships Save as List Generate Python code Export

Showing 1 to 25 of 110 rows Rows per page: 25 page 1

Gene Primary DBID	Gene Systematic Name	Gene Standard Name	Gene Parents	GO Annotation Term - Name	GO Annotation Term - Identifier	Code Annot Type	Gene Qualifier	GO Annotation Namespace	Publications PubMed ID	Publications Citation
S000000134	YBL038W	MRPL16	mitochondrial ribosome	GO:0005761	mitochondrial large ribosomal subunit	manually curated	Verified	cellular_component	7478995	Pen C and Mason TL (1995) Identification of the yeast nuclear gene for the mitochondrial homologue of bacterial ribosomal protein L16. Nucleic Acids Res 23(16):3673-7
S000000134	YBL038W	MRPL16	mitochondrial ribosome	GO:0005761	mitochondrial large ribosomal subunit	manually curated	Verified	cellular_component	9151978	Kitakawa M, et al. (1997) Identification and characterization of the genes for mitochondrial ribosomal proteins of <i>Saccharomyces cerevisiae</i> . Eur J Biochem 245(2):449-56

Create a new List of 90 Genes

List Name

mito_ribosome

Optional attributes

List Description

List of genes annotated to mito ribosome or child terms

NO TAGS Add a new tag add

Close Create List

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

Gene → Interaction ☆

Retrieve all interactions for a specified gene.

Gene

LOOKUP: act1

constrain to be IN saved Gene list mito_ribosome

Show Results

web service URL Perl Python Ruby Java [help] export XML

Create a new List of 1,062 Genes

List Name

mito_ribo_interactors

Optional attributes

List Description

List of genes which interacts with mito ribo genes

NO TAGS Add a new tag add

Close Create List

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

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

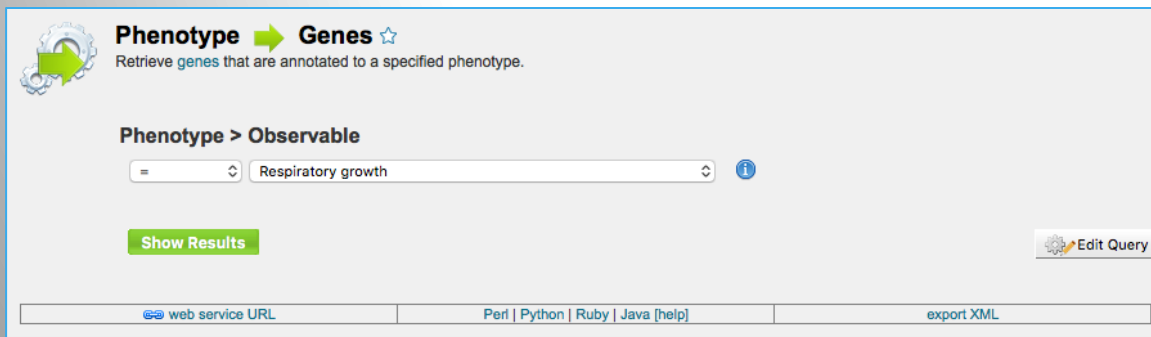
The screenshot shows the 'Lists' section of the SGD website. It includes a search bar, a filter dropdown, and a 'Reset' button. Below the search bar, there are action buttons for 'Union', 'Intersect', 'Subtract', and 'Asymmetric Difference', along with 'Copy' and 'Delete' icons. The 'Options' section is checked for 'Show descriptions'. Two lists are displayed: 'mito_ribo_interactors' with 1062 genes and a description 'List of genes which interacts with mito ribo genes', and 'Uncharacterized_ORFs' with 668 genes. The intersection of these two lists is highlighted in blue.

List Name	Gene Count	Description
mito_ribo_interactors	1062 Genes	List of genes which interacts with mito ribo genes
Uncharacterized_ORFs	668 Genes	

32 genes are uncharacterized

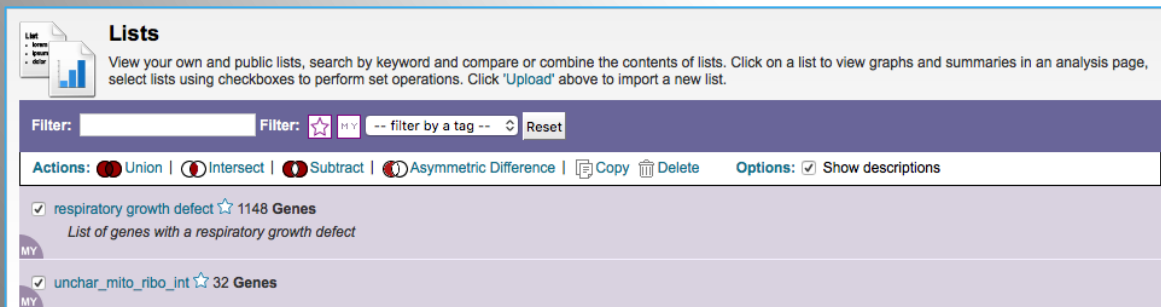
32 uncharacterized ORFs interact genetically or physically with known mitochondrial ribosomal proteins.

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



The screenshot shows the 'Phenotype > Genes' interface. At the top, it says 'Phenotype > Genes' with a green arrow icon and a star icon. Below this is the instruction 'Retrieve genes that are annotated to a specified phenotype.' The main section is titled 'Phenotype > Observable'. It features a dropdown menu with 'Respiratory growth' selected. To the right of the dropdown is an information icon. Below the dropdown is a green 'Show Results' button. To the right of the button is an 'Edit Query' button with a gear icon. At the bottom, there is a row of links: 'web service URL', 'Perl | Python | Ruby | Java [help]', and 'export XML'.

- create list of genes that confer a respiratory phenotype



The screenshot shows the 'Lists' interface. At the top, it says 'Lists' with a document icon. Below this is the instruction 'View your own and public lists, search by keyword and compare or combine the contents of lists. Click on a list to view graphs and summaries in an analysis page, select lists using checkboxes to perform set operations. Click 'Upload' above to import a new list.' The main section has a search bar with 'Filter:' and a 'Reset' button. Below the search bar is a row of 'Actions' buttons: 'Union', 'Intersect', 'Subtract', and 'Asymmetric Difference'. To the right of these buttons are 'Copy' and 'Delete' buttons. To the right of the 'Actions' buttons is an 'Options' section with a checkbox for 'Show descriptions'. Below the 'Options' section is a list of two items: 'respiratory growth defect' with 1148 Genes and 'unchar_mito_ribo_int' with 32 Genes. Each item has a checkbox and a star icon.

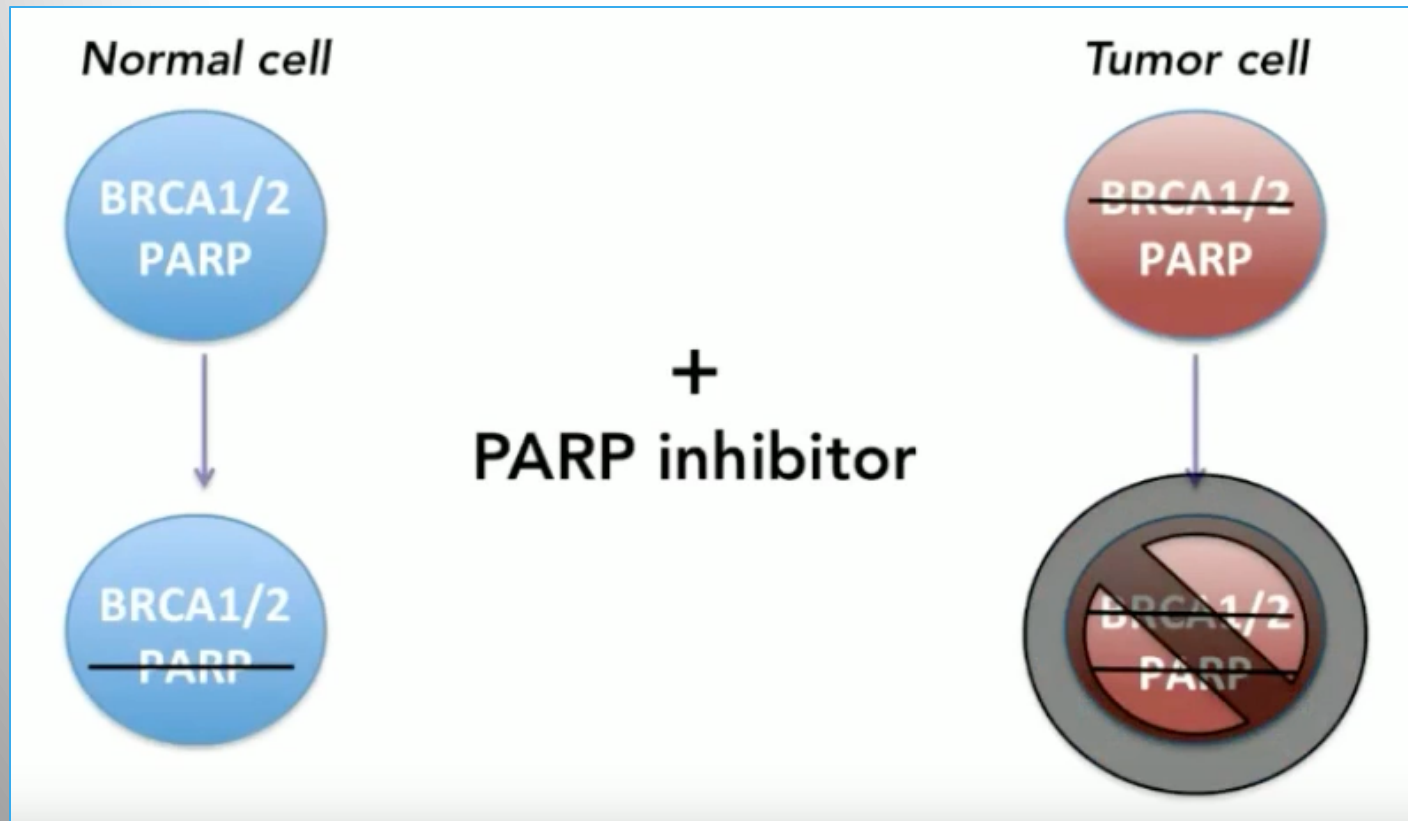
- find the intersection with the list of 32 uncharacterized ORFs

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

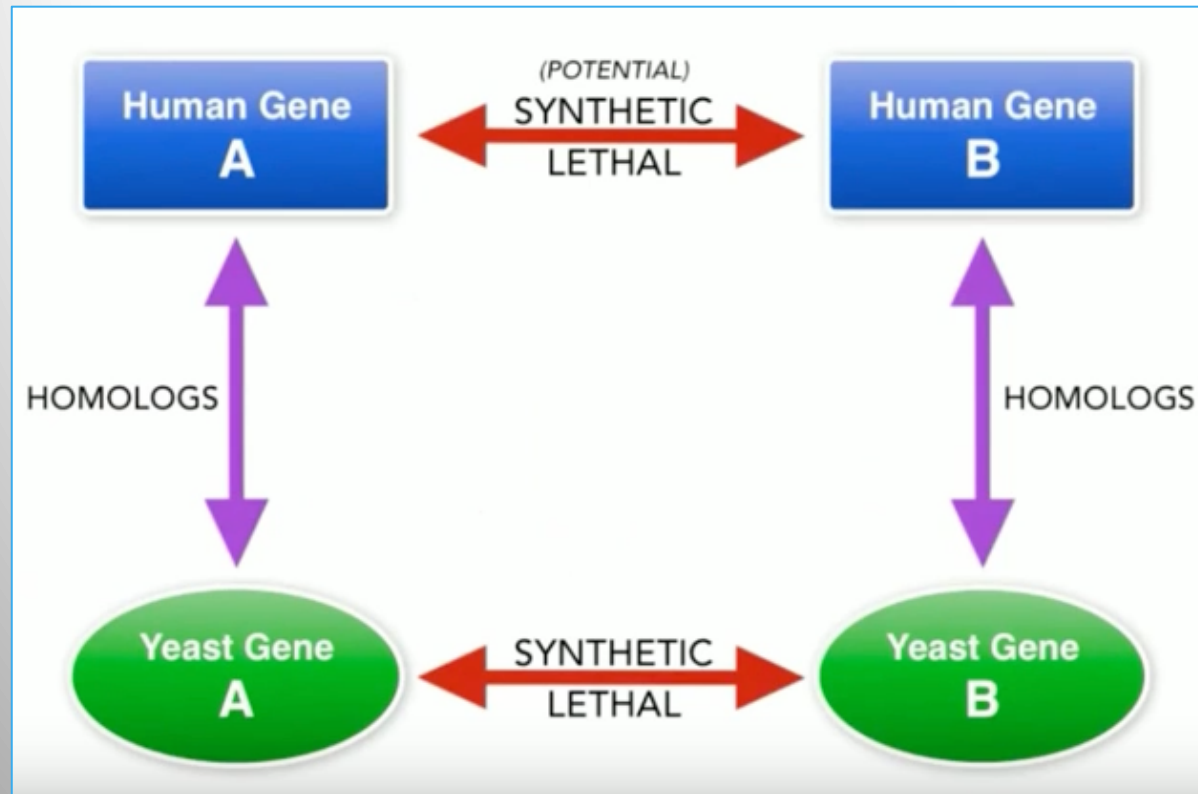
Predicting chemotherapy targets

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

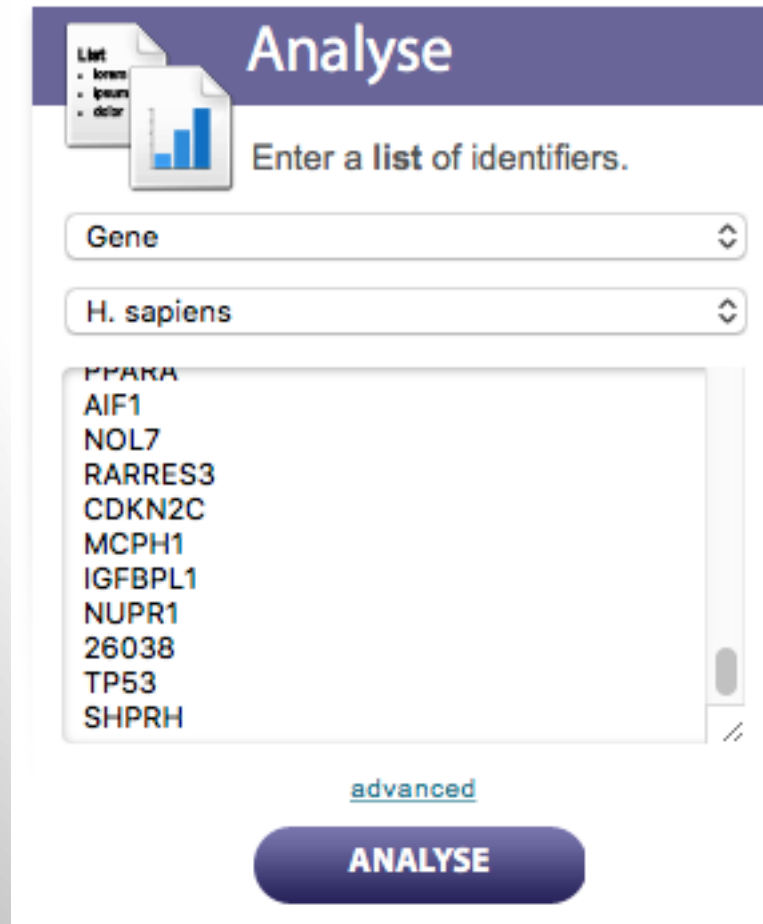


Predicting chemotherapy targets

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



Step 1: Create human gene list



The screenshot shows the 'Analyse' section of the SGD website. It features a purple header with the word 'Analyse' and a document icon. Below the header, there is a text input field labeled 'Gene' and a dropdown menu showing 'H. sapiens'. A large text area contains a list of gene identifiers: PPARG, AIF1, NOL7, RARRES3, CDKN2C, MCPH1, IGFBPL1, NUPR1, 26038, TP53, and SHPRH. At the bottom, there is a blue button labeled 'ANALYSE' and a link labeled 'advanced'.

Analyse

Enter a **list** of identifiers.

Gene


H. sapiens

PPARG
AIF1
NOL7
RARRES3
CDKN2C
MCPH1
IGFBPL1
NUPR1
26038
TP53
SHPRH

[advanced](#)

ANALYSE

Step 2: Find yeast homologs & save yeast genes



Human Gene -> Yeast Homolog(s) -> OMIM Disease Phenotype ☆

For a given human gene(s) retrieve associated OMIM disease phenotype(s) and yeast homolog(s).

Gene

LOOKUP:

☒ constrain to be saved Gene list

[web service URL](#) [Perl](#) [Python](#) [Ruby](#) [Java \[help\]](#) [export XML](#)

Create a new List of 93 Genes


List Name

Optional attributes

List Description


NO TAGS

Step 3: ID synthetic lethal interactors



**Gene** → **Interaction** ☆
Retrieve all interactions for a specified [gene](#).

Gene
LOOKUP:
☒ constrain to be saved Gene list
Show Results

[web service URL](#) [Perl](#) | [Python](#) | [Ruby](#) | [Java](#) [\[help\]](#)

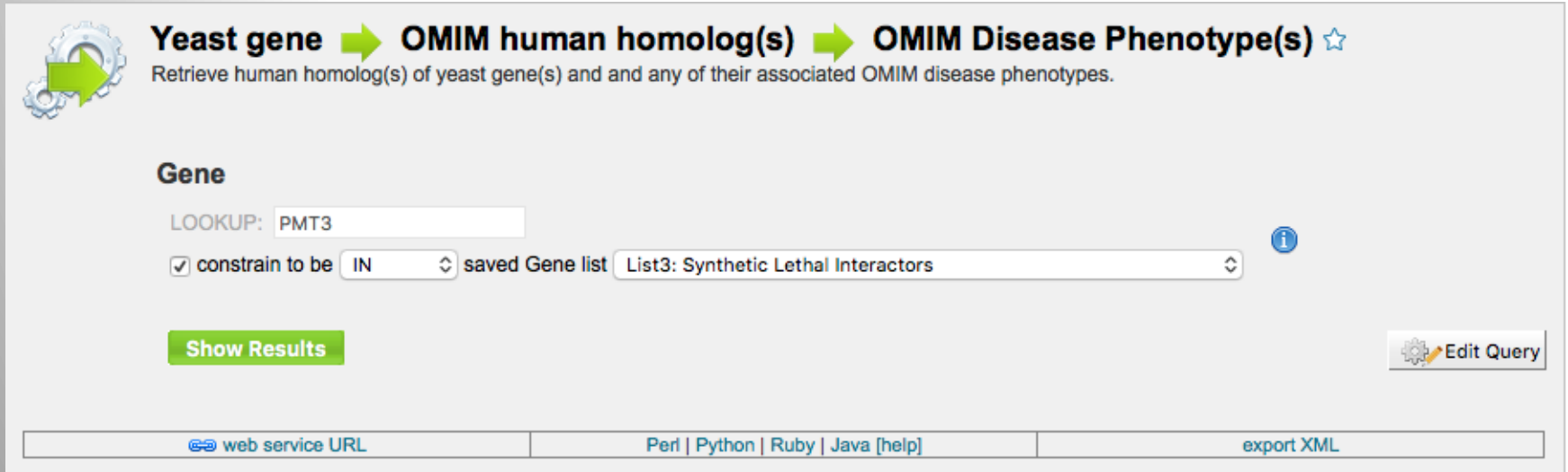
27 Interaction Term Identifiers

749 Items Selected
Filter values

	Interaction Term Identifier	Count
<input type="checkbox"/>	Negative Genetic	5,787
<input type="checkbox"/>	Affinity Capture-MS	2,474
<input type="checkbox"/>	Synthetic Growth Defect	1,528
<input type="checkbox"/>	Positive Genetic	1,258
<input type="checkbox"/>	Two-hybrid	983
<input type="checkbox"/>	Affinity Capture-Western	933
<input type="checkbox"/>	Phenotypic Enhancement	856
<input type="checkbox"/>	Biochemical Activity	751
<input checked="" type="checkbox"/>	Synthetic Lethality	749
<input type="checkbox"/>	Synthetic Rescue	570

Filter   [Download data](#)

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



The screenshot shows the SGD OMIM query interface. At the top, a header bar contains a gear icon with a green arrow, followed by the text "Yeast gene → OMIM human homolog(s) → OMIM Disease Phenotype(s) ☆". Below this, a subtitle reads "Retrieve human homolog(s) of yeast gene(s) and any of their associated OMIM disease phenotypes." The main form area is titled "Gene" and contains a "LOOKUP:" label followed by a text input field containing "PMT3". Below the input field is a checkbox labeled "constrain to be" followed by a dropdown menu showing "IN" and a "saved Gene list" dropdown menu showing "List3: Synthetic Lethal Interactors". A green "Show Results" button is located below the form. To the right of the form is an "Edit Query" button with a gear icon. At the bottom of the interface, there are three links: "web service URL", "Perl | Python | Ruby | Java [help]", and "export XML".

Yeast gene → OMIM human homolog(s) → OMIM Disease Phenotype(s) ☆
Retrieve human homolog(s) of yeast gene(s) and any of their associated OMIM disease phenotypes.

Gene

LOOKUP:

☒ constrain to be saved Gene list

[Show Results](#) [Edit Query](#)









[web service URL](#) [Perl](#) | [Python](#) | [Ruby](#) | [Java \[help\]](#) [export XML](#)

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

 MEC1	 YBR136W	 Mitosis Entry Checkpoint	 ATR
 POL3	 YDL102W	 POLymerase	 POLD1

H. sapiens

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!

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

Sandra Hocke¹, Yang Guo¹, Albert Job², Michael Orth³, Andreas Ziesch¹, Kirsten Lauber³, Enrico N. De Toni¹, Thomas M. Gress², Andreas Herbst¹, Burkhard Göke¹, Eike Gallmeier^{1,2}

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

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

³Department 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

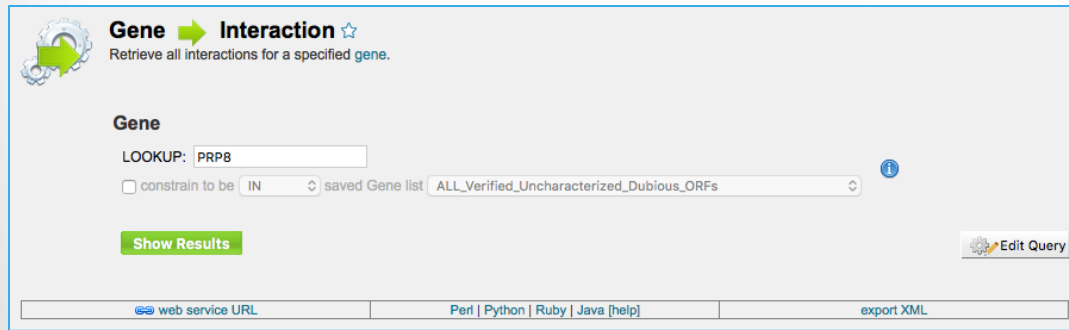
Feature Type: ORF, Verified

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



Gene → Interaction ☆
Retrieve all interactions for a specified [gene](#).

Gene

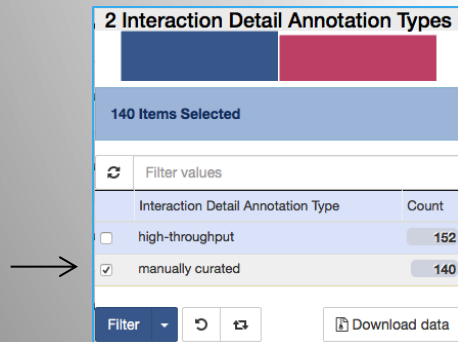
LOOKUP:

☐ constrain to be saved Gene list

[Show Results](#) [Edit Query](#)

[web service URL](#) [Perl](#) [Python](#) [Ruby](#) [Java \[help\]](#) [export XML](#)

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



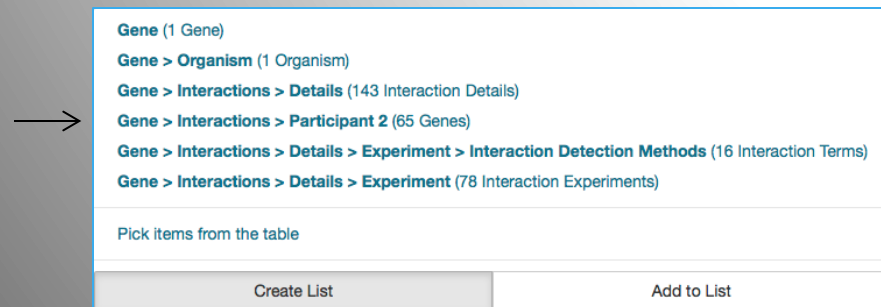
2 Interaction Detail Annotation Types

140 Items Selected

Filter values

Interaction Detail Annotation Type	Count
<input type="checkbox"/> high-throughput	152
<input checked="" type="checkbox"/> manually curated	140

[Filter](#) [Download data](#)



Gene (1 Gene)

Gene > Organism (1 Organism)

Gene > Interactions > Details (143 Interaction Details)

Gene > Interactions > Participant 2 (65 Genes)

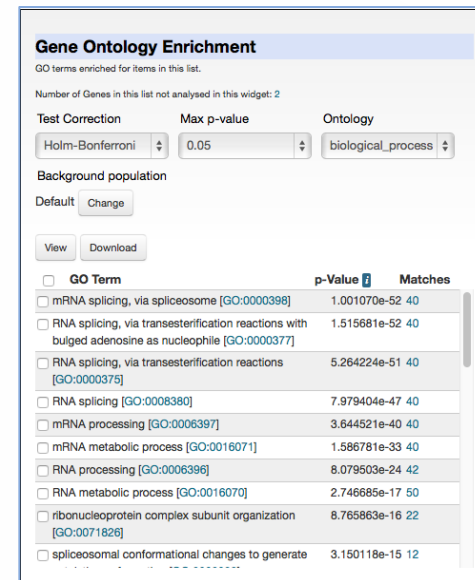
Gene > Interactions > Details > Experiment > Interaction Detection Methods (16 Interaction Terms)

Gene > Interactions > Details > Experiment (78 Interaction Experiments)

Pick items from the table

[Create List](#) [Add to List](#)

3. View enrichment



Gene Ontology Enrichment

GO terms enriched for items in this list.

Number of Genes in this list not analysed in this widget: 2

Test Correction: Max p-value: Ontology:

Background population: [Change](#)

[View](#) [Download](#)

GO Term	p-Value	Matches
<input type="checkbox"/> mRNA splicing, via spliceosome [GO:000398]	1.001070e-52	40
<input type="checkbox"/> RNA splicing, via transesterification reactions with bulged adenosine as nucleophile [GO:0000377]	1.515681e-52	40
<input type="checkbox"/> RNA splicing, via transesterification reactions [GO:0000375]	5.264224e-51	40
<input type="checkbox"/> RNA splicing [GO:0008380]	7.979404e-47	40
<input type="checkbox"/> mRNA processing [GO:0006397]	3.644521e-40	40
<input type="checkbox"/> mRNA metabolic process [GO:0016071]	1.586781e-33	40
<input type="checkbox"/> RNA processing [GO:0006396]	8.079503e-24	42
<input type="checkbox"/> RNA metabolic process [GO:0016070]	2.746685e-17	50
<input type="checkbox"/> ribonucleoprotein complex subunit organization [GO:0071826]	8.765863e-16	22
<input type="checkbox"/> spliceosomal conformational changes to generate	3.150118e-15	12

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

Trail: Query
OMIM Disease Phenotype → **human gene(s)** → **yeast homolog(s)** ★
 Specify OMIM phenotype(s) (by keyword or name) and retrieve all associated human gene(s) and the yeast homologs of these gene(s).

1 to 25 of 64

Rows per page: 25 page 1

Disease Name	Primary ID	Cross References Identifier	Homologue Standard Name	Homologue Systematic Name
21 Disease Names				
RETINITIS PIGMENTOSA 10; RP10	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 59; RP59	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 13; RP13	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 46; RP46	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 67; RP67	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 70; RP70	557	613598	PZF1	YPR186C
HYPOPREBETALIPOPROTEINEMIA, ACANTHOCYTOSIS, RETINITIS PIGMENTOSA, AND PALLIDAL DEGENERATION	226	608830	ENV9	YOR246C
LEBER CONGENITAL AMAUROSIS 13; LCA13 RETINITIS PIGMENTOSA 53, INCLUDED; RP53, INCLUDED	226	608830	ENV9	YOR246C
RETINITIS PIGMENTOSA 11; RP11	226	608830	ENV9	YOR246C

61 Items Selected

Filter values

Filter

Download data

Trail: Query
OMIM Disease Phenotype → **human gene(s)** → **yeast homolog(s)** ★
 Specify OMIM phenotype(s) (by keyword or name) and retrieve all associated human gene(s) and the yeast homologs of these gene(s).

1 to 25 of 64

Save as List

→

Disease (20 Diseases)
 Disease > Genes (19 Genes)
 Disease > Genes > Cross References (19 Cross References)
 Disease > Genes > Homologues > Homologue (27 Genes)
 Disease > Genes > Homologues > Data Sets > Data Source (4 Data Sources)
 Disease > Genes > Homologues > Homologue > Organism (1 Organism)
 Disease > Genes > Organism (1 Organism)
 Disease > Genes > Cross References > Source (1 Data Source)

Pick items from the table

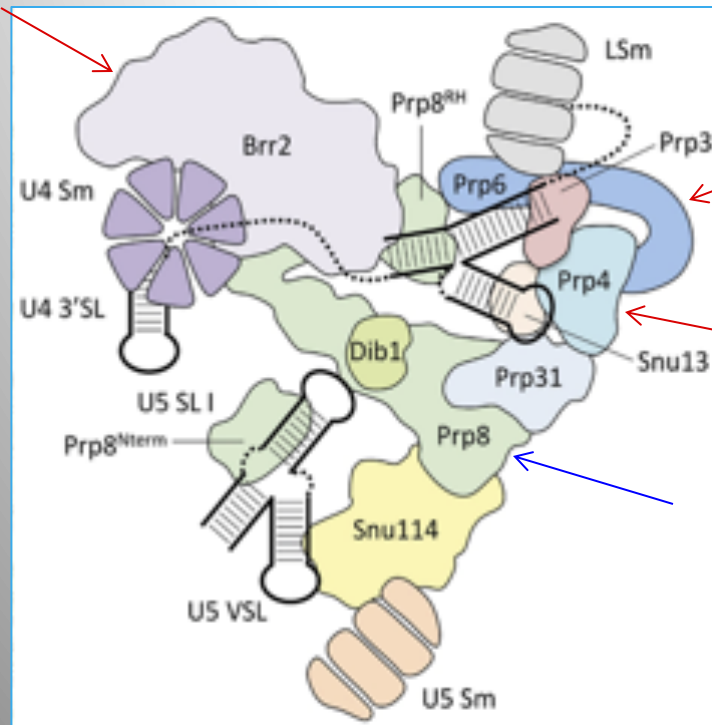
Create List Add to List

Disease Name	Primary ID	Cross References Identifier	Homologue Standard Name	Homologue Systematic Name
RETINITIS PIGMENTOSA 10; RP10	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 59; RP59	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 13; RP13	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 46; RP46	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 67; RP67	0594	607300	PRP8	YHR165C
RETINITIS PIGMENTOSA 70; RP70	557	613598	PZF1	YPR186C
HYPOPREBETALIPOPROTEINEMIA, ACANTHOCYTOSIS, RETINITIS PIGMENTOSA, AND PALLIDAL DEGENERATION	226	608830	ENV9	YOR246C
LEBER CONGENITAL AMAUROSIS 13; LCA13 RETINITIS PIGMENTOSA 53, INCLUDED; RP53, INCLUDED	226	608830	ENV9	YOR246C
RETINITIS PIGMENTOSA 11; RP11	226	608830	ENV9	YOR246C

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

¹Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. ²Laboratory 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



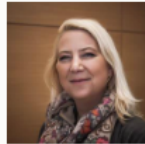
Gail Binkley, M.S.
Project Manager, SGD
Database Administrator
gail.binkley@stanford.edu



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
tj@stanford.edu



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



Olivia Lang
Biocuration Assistant
owlang@stanford.edu



**Kevin MacPherson,
M.S.**
Biocuration Assistant
kmacpher@stanford.edu



Stuart Miyasato
Senior Systems
Administrator
miyasato@stanford.edu



Rob Nash, Ph.D.
Senior Biocuration Scientist
mash@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