

# Chapter 13

## Exploring Protein Function Using the *Saccharomyces* Genome Database

Edith D. Wong

### Abstract

Elucidating the function of individual proteins will help to create a comprehensive picture of cell biology, as well as shed light on human disease mechanisms, possible treatments, and cures. Due to its compact genome, and extensive history of experimentation and annotation, the budding yeast *Saccharomyces cerevisiae* is an ideal model organism in which to determine protein function. This information can then be leveraged to infer functions of human homologs. Despite the large amount of research and biological data about *S. cerevisiae*, many proteins' functions remain unknown. Here, we explore ways to use the *Saccharomyces* Genome Database (SGD; <http://www.yeastgenome.org>) to predict the function of proteins and gain insight into their roles in various cellular processes.

**Keywords** *Saccharomyces cerevisiae*, GO Term Finder, Protein function, Gene Ontology, YeastMine

---

### 1 Introduction

The *Saccharomyces* Genome Database (SGD; <http://www.yeastgenome.org>) is the premier website for researchers interested in curated data about the budding yeast, *Saccharomyces cerevisiae* [1]. SGD provides comprehensive, curated biological information regarding protein function, cellular location, interactions, and expression, as well as information on mutant phenotypes for *S. cerevisiae* along with search and analysis tools to explore these data, thereby enabling the discovery of functional relationships between sequence and gene products in fungi and higher eukaryotic organisms. The tools provided by SGD leverage this expertly curated information. For predicting protein function, users can access the wealth of curated information available on SGD's locus summary pages, search and retrieve information using YeastMine, or analyze gene function using Gene Ontology (GO) enrichment tools.

For researchers interested in a single, specific protein, the Locus pages at SGD contain function summaries, gene ontology

annotations, protein-protein interaction information, as well as other curated, published experimental data to help predict a protein's role in cellular processes. Additionally, more specific protein information, such as predicted domains, can be found on Protein pages. This information can shed light on a protein's function, and help direct experimental research aimed at determining a protein's function.

By leveraging these curated data, users can also predict the functions of similar proteins using the GO Term Finder enrichment tool at SGD. The GO project (<http://geneontology.org/>) is a collaborative initiative involving many model organism databases, and was initiated by SGD, FlyBase (*Drosophila*), and Mouse Genome Informatics (MGI) to provide consistent descriptions of gene products across various organisms and databases [2]. GO has developed and maintains three structured ontologies to describe the biological processes, molecular functions, and cellular locations of gene products. GO Term Finder performs enrichment analysis by taking lists of genes as input, then looking for overrepresented ontology terms based on the associated annotations compared to all other annotations [3]. This enables the identification of biological processes, molecular functions, and cellular locations that the gene products in the input list may have in common.

Finally, YeastMine is a powerful search tool that can be queried to identify lists of proteins based on similarity in function, or other shared characteristics, to gain insight into biological processes [4]. YeastMine contains all curated information stored in the SGD database and more. Pregenerated queries can be employed by users interested in a specific process to find novel players, or to identify common functions within a group of proteins to predict function or involvement in a specific cellular process.

Individually or together, the various tools at SGD can provide insight into the cellular processes in which proteins are involved, thereby supporting the development of a fully defined model of the eukaryotic cell.

---

## 2 Materials

Using any of the tools mentioned above requires an internet-connected computer with a web-browser to connect to the *Saccharomyces* Genome Database (<http://www.yeastgenome.org>).

---

## 3 Methods

The data in SGD are updated on a daily basis as more information is curated. Therefore, results for these methods and case studies outlined below may be slightly different than described.

### 3.1 Summary Pages

1. Open the URL, [www.yeastgenome.org](http://www.yeastgenome.org), in any modern web browser (e.g., Chrome, Firefox, or Safari).
2. Enter a gene name (e.g., NOG2) or systematic name (e.g., YNR053C) into the search box on the top right-hand side of the page and hit return.
3. The resulting NOG2/YNR053C Summary page has many sections that can be used to gain insight into the protein's potential function. The top Overview section contains a Description that summarizes curated information. The Gene Ontology (GO) section contains GO terms used to annotate the gene product, and also a brief summary of its function, cellular location, and biological processes in which it is involved (Fig. 1a). GO terms are linked to a term-specific page that displays the term's definition as well as a list of other co-annotated gene products. Proteins with similar functions or that are involved in the same processes have closely related annotations, thus GO annotations of characterized proteins can be leveraged to predict the functions of uncharacterized proteins.
4. Based on the premise that proteins acting in the same pathway or involved in the same process may share mutant phenotypes, the data in the Phenotype section can also provide insight into a protein's function. The Phenotype section contains a summary of the curated experimental phenotype data, as well as the individual mutant phenotype annotations (Fig. 1b).
5. Protein function can also be predicted based on the proteins that interact with it (*see Note 1*). The Interaction section of this page summarizes the methods and number of physical and genetic interactions that have been identified for this locus (Fig. 1c). Selecting the "Details" link on the right-hand side of any section will provide access to more detailed information, such as a list of the interacting genes and proteins for the "Interaction Details" link. By accessing a list of interacting proteins, available via the "Analyze" link on the details page, the GO Enrichment tool described below can be used to predict protein function.
6. For additional insight into a protein's function, click on "Protein Details" on the top right link of the "Protein" section of the page, or the "Protein" tab at the top of the page. The "Protein" page contains the following types of protein-specific information: cellular abundance, predicted domains, protein sequence, posttranslational modifications, and physical properties. The predicted domains can help predict a protein's function. For example, Nog2p, a putative GTPase, has predicted GTP-binding protein domains in multiple locations (Fig. 2).

**A.**

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

**NOG2 / YNR053C Overview**

**Standard Name:** [NOG2](#)<sup>1</sup>

**Systematic Name:** YNR053C

**SGD ID:** S000005336

**Aliases:** [NUG2](#)<sup>2</sup>

**Feature Type:** ORF, Verified

**Description:** Putative GTPase; associates with pre-60S ribosomal subunits in the nucleolus and is required for their nuclear export and maturation; recruited by ribosomal proteins L17, L35, and L37 to assembling ribosomes after 275B pre-rRNA is generated, immediately preceding removal of ITS2<sup>1,3</sup>

**Name Description:** [Nucleolar G-protein](#)<sup>1</sup>

**B.**

NOG2 / YNR053C	Phenotype
----------------	-----------

**Phenotype** [Phenotype Details](#)

**Summary:** Essential gene in S288C; depletion results in slow growth; in Sigma1278b null mutation causes defects in filamentous growth

**Classical Genetics**

**null:**

- inviable

**Large-scale Survey**

**null:**

- competitive fitness: decreased
- filamentous growth: decreased
- haploinsufficient
- invasive growth: absent
- inviable

**conditional:**

- protein/peptide accumulation: decreased

**reduction of function:**

- competitive fitness: increased

**C.**

NOG2 / YNR053C	Interaction
----------------	-------------

**Interaction** [Interaction Details](#)

186 total interactions for 82 unique genes

**Physical Interactions**

- Affinity Capture-MS: 140
- Affinity Capture-RNA: 6
- Affinity Capture-Western: 22
- Co-purification: 3
- Two-hybrid: 10

**Genetic Interactions**

- Dosage Rescue: 1
- Negative Genetic: 1
- Positive Genetic: 1
- Synthetic Growth Defect: 1
- Synthetic Lethality: 1

**Fig. 1** Summary of information on a locus summary page. (a) *NOG2*/*YNR053C* Overview section has a description of known protein function information. (b) Phenotype information can help predict a protein's function. (c) Interaction data can be used to predict protein function and process involvement

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

**NOG2 / YNR053C Protein** Protein Help

**Aliases:** NUG2<sup>2</sup>  
**Protein Product:** putative GTPase NOG2  
**Feature Type:** ORF, Verified  
**Description:** Putative GTPase; associates with pre-60S ribosomal subunits in the nucleolus and is required for their nuclear export and maturation; recruited by ribosomal proteins L17, L35, and L37 to assembling ribosomes after 275B pre-rRNA is generated, immediately preceding removal of ITS2<sup>1,3</sup>

**Experimental Data**

Experiment	Result	Reference
protein abundance	67700 molecules/cell	Ghaemmaghami S, et al. (2003)
protein abundance	553 molecules/cell	Chong YT, et al. (2015)
protein abundance	1097 molecules/cell	Newman JR, et al. (2006)
protein abundance	6261 molecules/cell	Kulak NA, et al. (2014)

[Download \(.txt\)](#)

**Domains and Classification** 11 entries for 10 domains

Protein Coordinates	Accession ID	Description	Source	No. of Genes with Domain
25-484	<a href="#">PTHR11089</a>	gtp-binding protein-related	PANTHER	5
25-484	<a href="#">PTHR11089:SF9</a>	nucleolar gtp-binding protein 2	PANTHER	1
41-171	<a href="#">PF08153</a>	NGP1NT (NUC091) domain	Pfam	1
154-175	<a href="#">Coil</a>		-	1275
208-389	<a href="#">SSF52540</a>		SUPERFAMILY	330
211-373	<a href="#">G3DSA:3.40.50.300</a>		Gene3D	323
212-373	<a href="#">PSS1721</a>	Circularly permuted (CP)-type guanine nucleotide-binding (G) domain profile.	Prosite	5
318-381	<a href="#">PF01926</a>	50S ribosome-binding GTPase	Pfam	13
318-338	<a href="#">PRO0326</a>	GTP1/OBG GTP-binding protein family signature	PRINTS	8
364-379	<a href="#">PRO0326</a>	GTP1/OBG GTP-binding protein family signature	PRINTS	8

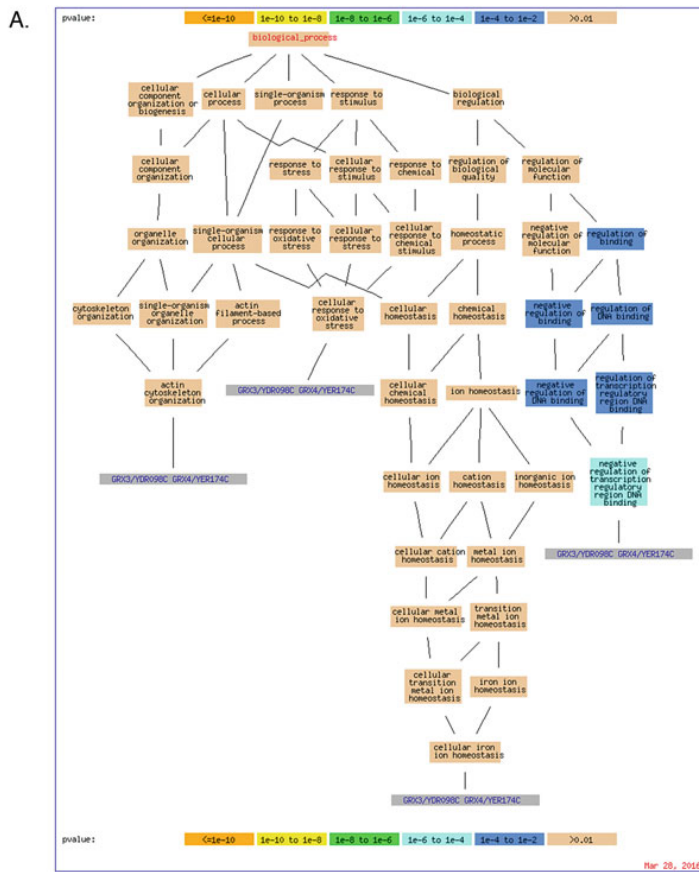
Showing 1 to 10 of 11 entries  records per page

**Fig. 2** NOG2 protein page. Protein abundance and domains on the protein page can be used to predict protein function

### 3.2 Gene Ontology Enrichment Tool

1. Open the URL [www.yeastgenome.org](http://www.yeastgenome.org) in any modern web browser (e.g., Chrome, Firefox, or Safari).
2. Enter a gene name in the search box as described above, such as YAL044W-A.

3. As previously outlined, a locus-specific Summary page will appear. Click the “Interactions” tab at the top of the page to open the Interactions page, which lists proteins that interact with the query protein.
4. On the Interactions page, a table with the interacting proteins and corresponding experimental information is shown. This table can be sorted and filtered on any column, such as for specific interacting proteins of interest, in the “Interactor” column. Underneath this table, the “Analyze” button will feed this list of interactors into a selection of SGD’s analysis tools.
5. After clicking on the “Analyze” button, the resulting “Analysis” page provides links to four analysis tools auto-populated with the list of interacting proteins, a table listing the interactors with curated summaries of their functions, and a precomputed list of shared GO biological processes. GO Term Finder will look for GO molecular function, biological process, or cellular component terms shared by these gene products. GO Slim Mapper is similar to GO Term Finder, but will sort genes into broad categories. SPELL provides visualization of genes with similar expression patterns to the genes on the list. Finally, YeastMine is an advanced data mining and analysis tool that will be described in detail below. For general information about these tools, *see Note 2*.
6. If “GO Term Finder” is selected, the “Gene Ontology Term Finder” form will have the interactors preentered into the “Query Set.” The default settings for this tool are to compare these genes’ GO annotations to all those in the SGD database annotated using both manual and high-throughput methods, and will only return hits with a  $p$ -value of less than 0.01. These settings can be changed in the Optional steps: 3, 4, and 5. Additionally, as defined in the “Step 2” section of the page, the Process branch of the Gene Ontology is used for enrichment, by default. This can be changed to another branch, such as Function or Component (*see Note 3*).
7. The results for GO Term Finder are displayed in both graphic and tabular form (Fig. 3). The graphic view displays the GO terms in a hierarchical tree; color coded by the  $p$ -value, and also indicates the genes that are annotated to the specific term. The table view displays the GO terms to which the genes are directly annotated, as well as parent terms, along with cluster and background frequencies and  $p$ -values. The results can also be downloaded using the “Download Full Results” link at the bottom of the results table. Links to GO term pages, as well as to the GO consortium’s search and browse tool (AmiGO; <http://amigo.geneontology.org/amigo>), are provided within the table of results.



B.

Terms from the Process Ontology						
Gene Ontology term	Cluster frequency	Background frequency	P-value	FDR	Expected False Positives	Genes annotated to the term
negative regulation of transcription regulatory region DNA binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	2 out of 7165 background genes, 0.0%	8.42e-05	0.00%	0.00	GRX3/YDR098C, GRX4/YER174C
negative regulation of DNA binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	3 out of 7165 background genes, 0.0%	0.00025	0.00%	0.00	GRX3/YDR098C, GRX4/YER174C
negative regulation of binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	4 out of 7165 background genes, 0.1%	0.00051	0.00%	0.00	GRX3/YDR098C, GRX4/YER174C
regulation of transcription regulatory region DNA binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	5 out of 7165 background genes, 0.1%	0.00084	0.00%	0.00	GRX3/YDR098C, GRX4/YER174C
regulation of DNA binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	7 out of 7165 background genes, 0.1%	0.00177	0.00%	0.00	GRX3/YDR098C, GRX4/YER174C
regulation of binding   AmiGO <sup>®</sup>	2 out of 7 genes, 28.6%	12 out of 7165 background genes, 0.2%	0.00554	0.00%	0.02	GRX3/YDR098C, GRX4/YER174C

[Download Full Results](#)

**Fig. 3** GO Term Finder results. (a) Graphic view of GO term enrichment of proteins that interact with Yal044w-ap. (b) Table view of GO Term Finder results for Yal044w-ap interacting proteins

### 3.3 YeastMine

1. YeastMine can be accessed in many ways. To directly access YeastMine, open the URL <http://yeastmine.yeastgenome.org/yeastmine/begin.do> in a browser window. Alternatively, as described above, a pregenerated list of genes can be fed directly into YeastMine by selecting the “YeastMine” link on the “Analysis” page, which can be accessed via the “Analyze” button below any table of genes on SGD pages. Finally, below the search box on the top-right-hand side of every SGD page, YeastMine is accessible via the “Batch Analysis” or “Advanced Search” links.
2. Users can create and analyze their own lists of genes, or choose from one of a number of premade lists provided by SGD, which include, but are not limited to, the following: human genes with yeast homologs, all uncharacterized ORFs in *S. cerevisiae*, all curated macromolecular complexes, and all ORFs verified by experimental data (Verified\_ORFs). These collections are accessible by the “Lists” link in the purple toolbar that runs across at the top of the page, or on the YeastMine front page (Fig. 4b).
3. Pregenerated queries are grouped into different categories, depending on the types of annotated results that are analyzed (Fig. 4c). The “Function” templates will return lists of genes that are annotated to particular GO terms, or pathways. Conversely, for a given list of genes, other Function queries will return all the pathways or GO terms that are annotated within the starting list. The “Proteins” templates will return lists of genes based on criteria such as molecular weight, sequence, posttranslational modifications, domains, or cellular abundance.
4. Returned lists of results can be further screened by looking for similarities within other curated data, such as interactions and phenotypes, as outlined below in Subheading 4.
5. In addition, lists can be saved by creating a “MyMine” account, using the “Log In” link found on the top right-hand side of the header (Fig. 4a).

---

## 4 Case Studies

### 4.1 Predicting the Function of an Uncharacterized Protein with GO Term Finder

In this example, GO Term Finder will be used to predict the putative function of YAL044W-A, a protein of unknown function. As outlined in Subheading 3, after entering YAL044W-A in the search box, a page summarizing all curated information about YAL044W-A will be returned. YAL044W-A is noted to be similar to a *Schizosaccharomyces pombe* putative DNA repair protein, but no other functional information is known. However, it has been shown to physically interact with seven unique proteins (Fig. 5).



The screenshot shows the YeastMine website interface. At the top, there is a navigation bar with links for Home, Templates, Lists, QueryBuilder, Tools, Regions, Data Sources, API, MyMine, and a Log in button. Below the navigation bar is a search bar with the text "Search: e.g. act1". The main content area is divided into three columns: Search, Analyze, and Welcome Back. The Search column has a search box and a "SEARCH" button. The Analyze column has a "Enter a list of identifiers" section with a text input field and an "ANALYSE" button. The Welcome Back column has a "TAKE A TOUR" button. Below the main content area is a horizontal menu with tabs for GENOME, PROTEINS, FUNCTION, PHENOTYPES, INTERACTIONS, REGULATION, HOMOLOGY, EXPRESSION, and LITERATURE. The "FUNCTION" tab is selected. Below the menu is a "Query for genome:" section with a list of pre-generated queries. At the bottom is a "Lists" section with a list of pre-generated lists. Annotations with dashed arrows point from text boxes on the right to specific elements: "A. Create accounts to save lists" points to the "Log in" button; "B. Analyze lists of genes" points to the "Analyze" section; and "C. Pregenerated queries for analysis" points to the "Query for genome:" and "Lists" sections.

**Fig. 4** Analysis tools that are available on the YeastMine home page

Researchers can gain insight into the putative biological function of the uncharacterized gene, YAL044W-A, by leveraging the curated information about these interacting proteins' functions with GO Term Finder to determine GO terms that have been annotated to this input set.

First, to identify known proteins that interact with YAL044W-A, click either on the "Interactions" tab at the top of the page, or the "Interactions Details" link in the "Interactions" section of the page. In the "Annotations" section, a list of seven proteins that interact with Yal044w-ap, along with experimental details and references, are listed in a table (Fig. 5b). An "Interaction Network" below the "Annotations" section is a graphic display of the interactions of these proteins. In this example, Grx3p/Ydr098cp, Grx4p/Yer174cp, Mgs1p/Ynl218wp, Mpa43p/Ynl249cp,

A.

Summary Sequence Protein Gene Ontology Phenotype Interactions Regulation Expression Literature

**YAL044W-A**

**Reserved Name:** BOL1<sup>2</sup> *This name is reserved with SGD according to the Gene Naming Guidelines agreed upon by the yeast community.*

**Systematic Name:** YAL044W-A

**SGD ID:** S000007586

**Feature Type:** ORF, Uncharacterized

**Description:** Putative protein of unknown function; similar to *S. pombe* uvi31 which is a putative DNA repair protein<sup>1</sup>

**Name Description:** BolA-like protein<sup>2</sup>

B.

**YAL044W-A Interactions** <sup>1</sup> [Interaction Help](#)

Source: All physical and genetic interaction annotations listed in SGD are curated by [BioGRID](#).

Physical  
7

Analyze  
 Physical  Genetic  Intersection  All

**Annotations** <sup>1</sup> 7 entries for 7 genes

Interactor	Type	Assay	Annotation	Action	Modification	Phenotype	Reference
GRX3	Physical	Two-hybrid	high-throughput	Bait	No Modification		Yu H, et al. (2008) PMID:18719252
GRX4	Physical	Two-hybrid	high-throughput	Bait	No Modification		Yu H, et al. (2008) PMID:18719252
MGS1	Physical	Two-hybrid	high-throughput	Bait	No Modification		Yu H, et al. (2008) PMID:18719252
MPA43	Physical	Two-hybrid	high-throughput	Bait	No Modification		Yu H, et al. (2008) PMID:18719252
<sup>1</sup> NAB2	Physical	Affinity Capture-RNA	high-throughput	Hit	No Modification		Batisse J, et al. (2009) PMID:19840948
<sup>1</sup> SRO9	Physical	Affinity Capture-RNA	high-throughput	Hit	No Modification		Schenk L, et al. (2012) PMID:22271760
<sup>1</sup> SSB2	Physical	Affinity Capture-RNA	high-throughput	Hit	No Modification		Willmund F, et al. (2013) PMID:23332755

Showing 1 to 7 of 7 entries 10 records per page 1

[Download \(txt\)](#) [Analyze](#)

**Fig. 5** Curated information about YAL044W-A. (a) Description of YAL044W-A, an uncharacterized ORF. (b) List of proteins that physically interact with Yal044w-ap

Nab2p/Ygl122cp, Sro9p/Ycl037cp, and Ssb2p/Ynl209wp all have been shown to physically interact with Yal044w-ap.

Once identified, the annotations for these interactors can be used with the GO enrichment tool to determine YAL044W-A's putative function. Directly below the annotations table, select the "Analyze" button to pass these gene names to SGD's analysis tools. On this "Analyze" page, GO Term Finder, GO Slim Mapper, SPELL, and YeastMine are listed above the input list of genes. This list of interactors also includes the curated description of their functions, which could also shed some light on the function of the gene of interest. To find common GO terms to which these genes have been annotated, this list can be sent directly to GO Term Finder via the "GO Term Finder" link above the list of interactors.

In this case, the GO Term Finder form will be prefilled with GRX3/YDR098C, GRX4/YER174C, MGS1/YNL218W, MPA43/YNL249C, NAB2/YGL122C, SRO9/YCL037C, and SSB2/YNL209W. To predict the function of YAL044W-A, the researcher would enrich for "Process" annotations, which describe broad biological concepts (the default selection). In this case study, all default settings would be used. After clicking "Search," the GO Term Finder results are displayed both graphically and in tabular form (Fig. 3). Within the graphical display of the results, the hierarchical structure of GO can be observed (Fig. 3a). Based on the enrichment of GO terms in the input list of interactors, YAL044W-A is predicted to have a role in regulating transcription with GRX3 and GRX4, likely through negative regulation of DNA binding (<http://yeastgenome.org/go/GO:2000678/overview>). Additionally, the table displays the calculated frequency of GO terms in the cluster (input list) and the background list of genes, the entire yeast genome in this case. The  $p$ -value of  $8.42e-05$  and expected false positive of 0.00 indicate a high confidence in this GO term process enrichment (Fig. 3b). The enrichment of processes, such as this case study, can provide direction toward experimentally determining a protein's function.

#### **4.2 Identifying Novel Components of the Mitochondrial Ribosome**

In this example, YeastMine will be used to identify novel, potential subunits of the mitochondrial ribosome. In yeast, there are 91 genes annotated to the mitochondrial ribosome or a ribosomal subunit, less than half of what has been found for cytosolic ribosomes. Leveraging the wealth of curated information in SGD, as well as the power of YeastMine's search and analysis, a list of putative novel mitochondrial ribosomal subunits can be identified for future lab experiments. First, the researcher could identify all known mitochondrial ribosome proteins. To identify novel proteins, this list can be used to generate a second list of proteins that interact with the known mitochondrial ribosome proteins. Finally, restricting the list of interactors to uncharacterized proteins would reduce the list to a manageable target list of possible genes of interest.

To identify all the genes encoding subunits of the mitochondrial ribosome, researchers would use the “Function” template query called “GO Term name [and children of this term] → All genes” (Fig. 6b). To keep the search broad, the constraints “CONTAINS” “mito\* ribosom\*” would be used. This returns a list of 91 unique genes, annotated to GO terms such as “mitochondrial ribosome,” “mitochondrial large ribosomal subunit,” and “mitochondrial small ribosomal subunit.” This list should be saved and named using the “Save as List” button above the results table.

The second step would be used to identify all proteins that interact with this list of ribosomal subunits. To do this, the user would use the “Gene → Interaction” query found under the “Interaction” category in Templates. Rather than looking up the interactions for a single gene, the checkbox next to “constrain to be IN saved Gene list” would be selected, and then the list created and saved in the previous step would be selected from the dropdown. This would return 3478 interactions involving 1030 unique genes (listed as Participant 2 Standard Name) for the original list. Again, this list of interacting proteins would be named and saved using the “Save as List” button above the results table.

This list of proteins that interact with known mitochondrial ribosome proteins can be filtered against a pregenerated list of uncharacterized genes. To do this, the researcher would select “Lists,” found at the top of the page, next to “Templates.” Under lists, the “View” tab will display all the SGD pregenerated lists as well as any personal lists that the researcher created. First, the recently created list of proteins that interact with known mitochondrial ribosomal subunits and the “Uncharacterized\_ORFs” list of 672 genes are selected. Then, selecting “Intersect” under “Actions:” above all the lists will return a list of 34 genes that are uncharacterized and have been shown to interact with known mitochondrial ribosomal subunits. This list of 34 genes would be a good starting point for experimentally confirming which of these predicted mitochondrial ribosomal subunits were actual mitochondrial ribosomal subunits. A detailed video description of this case study can be found on the SGD YouTube channel, here: <https://www.youtube.com/watch?v=AoVfmaZmTK4>. To learn more about the many features of YeastMine, visit the SGD YouTube channel (<https://www.youtube.com/SaccharomycesGenomeDatabase>) and enter “YeastMine” in the search box at the top of the page.

---

## 5 Notes

SGD provides a number of tips and help for the resources mentioned above.

1. General help about curated protein function information at SGD can be found on the function help page: <http://www.yeastgenome.org/help/function-help>.

A.

B.

**Fig. 6** Function queries in YeastMine for predicting protein function. (a) A list of pregenerated queries available for examining protein function. (b) Using specific query to identify all known mitochondrial ribosome proteins

2. Specific help about the analysis tools, GO Term Finder, GO Slim Mapper, and YeastMine can be found at <http://www.yeastgenome.org/help/analyze/>.
3. Default settings are a good initial point for enrichment, with adjustments made to  $p$ -value cutoffs, ontology branches, and evidence codes made as necessary. If no enrichment for GO annotation terms within the input list is returned, the  $p$ -value can be increased in optional Step 5. This typically increases the number of terms that are enriched.
4. A number of help videos on SGD's YouTube channel are also useful starting points for using these tools: <https://www.youtube.com/SaccharomycesGenomeDatabase>.
5. For more specific questions, SGD curators can be contacted directly at [sgd-helpdesk@lists.stanford.edu](mailto:sgd-helpdesk@lists.stanford.edu).

---

## Acknowledgments

I would like to thank the staff at SGD for their help in putting together this article, particularly Rob Nash and Stacia Engel. I would also like to thank J. Michael Cherry for this opportunity and support. This work was funded by the National Human Genome Research Institute (NHGRI), US National Institutes of Health [5U41HG001315-18].

## References

1. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40(Database issue):D700–D705
2. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25–29
3. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G (2004) GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20(18):3710–3715
4. Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM (2012) YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database (Oxford)* 2012:201