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
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.

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...
How to leverage data rich SGD!

Curated Data

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

- 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
...or analyze your own gene list
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-bisphophate 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

#### Summary
- Imidazolylglycerol-phosphate dehydratase involved in histidine biosynthesis

#### GO Slim Terms
- cellular amino acid metabolic process

#### Biologically Curated

**Date Last Reviewed:** 2007-02-21

### Biological Process

<table>
<thead>
<tr>
<th>Gene Ontology Term</th>
<th>Qualifier</th>
<th>Evidence</th>
<th>Source</th>
<th>Assigned On</th>
<th>Annotation Extension</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>histidine biosynthetic process</td>
<td>IMP:2</td>
<td>SGD</td>
<td></td>
<td>2009-10-02</td>
<td></td>
<td>Fink GR (1966) PMID:14139041</td>
</tr>
<tr>
<td>histidine biosynthetic process</td>
<td>IMP:2</td>
<td>SGD</td>
<td></td>
<td>2007-02-15</td>
<td></td>
<td>Struhl K and Davis RW (1977) PMID:1941150</td>
</tr>
</tbody>
</table>

Showing 1 to 2 of 2 entries 10 records per page

**Download (txt)**

### Molecular Function

<table>
<thead>
<tr>
<th>Gene Ontology Term</th>
<th>Qualifier</th>
<th>Evidence</th>
<th>Source</th>
<th>Assigned On</th>
<th>Annotation Extension</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>imidazolylglycerol-phosphate dehydratase activity</td>
<td>IMP:2</td>
<td>SGD</td>
<td></td>
<td>2009-10-02</td>
<td></td>
<td>Fink GR (1966) PMID:14139041</td>
</tr>
<tr>
<td>imidazolylglycerol-phosphate dehydratase activity</td>
<td>IMP:2</td>
<td>SGD</td>
<td></td>
<td>2007-02-15</td>
<td></td>
<td>Struhl K and Davis RW (1977) PMID:1941150</td>
</tr>
</tbody>
</table>

Showing 1 to 2 of 2 entries 10 records per page

**Download (txt)**

### Cellular Component

<table>
<thead>
<tr>
<th>Gene Ontology Term</th>
<th>Qualifier</th>
<th>Evidence</th>
<th>Source</th>
<th>Assigned On</th>
<th>Annotation Extension</th>
<th>Reference</th>
</tr>
</thead>
</table>

Showing 1 to 1 of 1 entries 10 records per page

**Download (txt)**
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.
### SGD Gene Ontology Slim Mapper

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

<table>
<thead>
<tr>
<th>Enter Gene/ORF names (separated by a return or a space)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAL002W</td>
</tr>
<tr>
<td>YAL011W</td>
</tr>
<tr>
<td>YAL021C</td>
</tr>
<tr>
<td>YAL024C</td>
</tr>
<tr>
<td>YAL046C</td>
</tr>
</tbody>
</table>

### Step 2: Choose GO Slim Terms 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

-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

### Step 3: Refine your list of GO Slim Terms

<table>
<thead>
<tr>
<th>Select ALL Terms from Yeast GO-Slim: Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA recombination</td>
</tr>
<tr>
<td>DNA repair</td>
</tr>
</tbody>
</table>

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"

Search Reset

- 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.
## GO Slim Mapper: Results

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

<table>
<thead>
<tr>
<th>GO-Slim term</th>
<th>Cluster frequency</th>
<th>Genome frequency</th>
<th>Genes annotated to the term</th>
</tr>
</thead>
<tbody>
<tr>
<td>mitotic cell cycle</td>
<td>76 out of 438 genes, 17.4%</td>
<td>319 of 6433 genes, 5%</td>
<td>ACT1, BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUD30, BUD31, BUR2, CCR4, CDC10, CDC11, CDC25, CDC28, CDC53, CDC55, CDH1, CHL4, CIC1, CIN8, CKS1, CLA1, CLA4, CLB2, CSM1, CTF18, CTF19, CTF4, CTF8, CTR9, DCK1, ELM1, HTF2, HIR3, HO1, HOC2, IPL1, IRC3, JNN1, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRCl, NDD1, NUP53, PAF1, PHO85, PIN4, PSE1, REF2, RM11, SGO1, SHE1, SHS1, SIT4, SMC2, SPC72, STE20, STN1, STU2, TOF1, TOPI, TOP3, TUB2, ULP2, YBP2, YRB1</td>
</tr>
<tr>
<td>organelle fission</td>
<td>64 out of 438 genes, 14.6%</td>
<td>302 of 6433 genes, 4.7%</td>
<td>BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CDC10, CDC28, CDC55, CDH1, CHL4, CIC1, CIN8, CLA4, CLB2, CSM1, CSM3, CST9, CTF18, CTF19, CTF4, CTF8, DBF2, DCC1, HTF2, HO1, HO2, HOS2, IPL1, IRC3, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRCl, NUP53, PAP2, PSE1, RM11, SGO1, SHE1, SHS1, SLZ1, SMC2, SPC72, STE20, STN1, STU2, TOF1, TOPI, TOP3, TUB2, ULP2, YBP2</td>
</tr>
<tr>
<td>transcription from RNA polymerase II promoter</td>
<td>62 out of 438 genes, 14.2%</td>
<td>476 of 6433 genes, 7.4%</td>
<td>ARPA4, BUD30, BUR2, CBF1, CCR4, CDC28, CDC36, CEG1, CKS1, CLA4, CTK1, CTR9, ESA1, GCN5, GCR1, GM13, GMP2, HCM1, HIR1, HIR3, HPC2, HPR1, HSF1, HTZ1, IMP2, INO80, LDB7, MFT1, MOT3, NDD1, NPL6, NUP48, OPP1, PAF1, PFD1, PHO80, PHO85, POB3, PO2, PSE1, REF2, REG1, RP89, RPN4, RSC2, SNF4, SNF8, SSN3, STE20, STH1, SUM1, SUD2, TAF14, THP2, TOPI, TRA1, VPS25, VPS36, YAP1, YKE2, YSH1</td>
</tr>
<tr>
<td>chromosome segregation</td>
<td>61 out of 438 genes, 13.9%</td>
<td>203 of 6433 genes, 3.2%</td>
<td>BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CBF1, CDC28, CDC55, CDH1, CHL4, CIC1, CIN8, CLA4, CLB2, CSM1, CSM3, CST9, CTF18, CTF19, CTF4, CTF8, DAM1, DCC1, HHT2, HO1, HO2, IES6, INO80, IPL1, IRC3, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCK1, MCM21, MIF2, MPS1, MPS3, MRCl, NDC80, RM11, RSC2, SGO1, SHE1, SHP1, SL15, SMC2, SMC6, SPC72, STN1, STU1, TOF1, TOPI, TOP3, TUB2, ULP2</td>
</tr>
<tr>
<td>cytoskeleton organization</td>
<td>58 out of 438 genes, 13.2%</td>
<td>238 of 6433 genes, 3.7%</td>
<td>ACT1, ARC18, ARC35, ARC40, ARP3, AVO1, BBP1, BEM2, BIK1, BIM1, BIM7, BUR1, CDC10, CDC28, CDC31, CDC37, CDH1, CIK1, CIN8, CLA4, CLB2, CMD1, COF1, CRN1, DAM1, ELM1, HCM1, HO1, IPL1, IRC3, JNN1, KAR9, KIP3, LAS17, LIA1, MDM20, MIF2, MPS1, MPS2, MPM3, NAT3, NDC80, PPF1, REF2, RHQ2, SAC6, SD5, SHE1, SHE4, SHS1, SIT4, SPC29, SPC72, STV2, STU4, STU2, TOR2, YBP2</td>
</tr>
<tr>
<td>regulation of organelle organization</td>
<td>52 out of 438 genes, 11.9%</td>
<td>283 of 6433 genes, 4.4%</td>
<td>ARC40, ARP3, BDF1, BFA1, BIK1, BIM1, BUB1, BUB2, BUB3, BUR2, CDC28, CDC55, CDH1, CHL4, CIK1, CLA4, CLB2, CRN1, CTF19, CTR9, DAM1, HHT2, HIR3, HO1, HO2, HOS2, IPL1, KAR9, KIP3, LAS17, LTE1, MAD1, MAD2, MAD3, MCM21, MDM20, MPS1, NAT3, NUP53, PAF1, PSE1, REF2, RM11, SAR1, SEC12, SGO1, SIR4, SMC2, SPC72, SRV2, STE20, ULP2</td>
</tr>
<tr>
<td>regulation of cell cycle</td>
<td>49 out of 438 genes, 11.2%</td>
<td>233 of 6433 genes, 3.6%</td>
<td>BFA1, BIM1, BUB1, BUB2, BUB3, CCR4, CDC25, CDC28, CDC36, CDC37, CDC55, CDH1, CHL4, CIK1, CKS1, CLA4, CLB2, CSM3, CTF19, DAM1, DBF2, ESA1, HHT2, HO1, HO2, IPL1, IRC3, KAR9, KIP3, LTE1, MAD1, MAD2, MAD3, MCM21, MDM20, MPS1, MRCl, NUP53, PHO80, PIN4, PSE1, REF2, SGO1, SL15, SMC2, SPC72, STE20, TOF1, TOR2, ULP2</td>
</tr>
</tbody>
</table>

---

[SGD Saccharomyces Genome Database]
GO Term Finder

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

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

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

Advantages:

• process multiple gene lists in parallel
• handles longer gene lists
• large number of available organisms
### 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.
Basic features

Templates are predefined queries. Filter by category:

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

or keyword:

- “intron”
- “sequence”
- “UTR”
Template results page

- 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

- Column sort
- Remove column
- Toggle column visibility
- Filter by values in column
- View column summary

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

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

4. Create a list of 1,062 interacting genes/proteins.
Are any of the interacting genes/proteins uncharacterized?

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

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

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

- create list of genes that confer a respiratory phenotype
- find the intersection with the list of 32 uncharacterized ORFs
Three uncharacterized ORFs exhibit genetic or physical interactions with known mt ribosomal proteins AND block respiratory growth when mutated

<table>
<thead>
<tr>
<th>Systematic name</th>
<th>Gene name</th>
<th>Name Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YBL095W</td>
<td>MRX3</td>
<td>Mitochondrial organization of gene expression</td>
<td>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</td>
</tr>
<tr>
<td>YDL157C</td>
<td></td>
<td></td>
<td>Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies</td>
</tr>
<tr>
<td>YPR109W</td>
<td></td>
<td></td>
<td>Predicted membrane protein; SWAT-GFP and mCherry fusion proteins localize to the endoplasmic reticulum; diploid deletion strain has high budding index</td>
</tr>
</tbody>
</table>
Predicting chemotherapy targets

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

Synthetic lethality: loss of two non-essential genes = inviability
Step 1: Create human gene list
Step 2: Find yeast homologs & save yeast genes
Step 3: ID synthetic lethal interactors

Run query and filter by interaction detection methods to obtain just synthetic lethals. Save as "List3: Synthetic Lethal Interactors"
Step 4: ID human homologs of SL interactors

Run query with SL interactors and then save list of human homologs as “List4: Human homologs of yeast SL Interactors”
YeastMine Scenario: Predicting potential chemotherapy targets

<table>
<thead>
<tr>
<th>S. cerevisiae</th>
<th>H. sapiens</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEC1</td>
<td>ATR</td>
</tr>
<tr>
<td>YBR136W</td>
<td>POL3</td>
</tr>
<tr>
<td>Mitosis Entry</td>
<td>POLymerase</td>
</tr>
<tr>
<td>Checkpoint</td>
<td></td>
</tr>
<tr>
<td>YDL102W</td>
<td>POLD1</td>
</tr>
</tbody>
</table>

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!
Explore a gene PRP8

1. Identify PRP8 interactors
2. Use OMIM to ID yeast orthologs of human genes involved in retinitis pigmentosa
3. Intersect the two lists to identify PRP8 interactors with orthologs involved in RP
1. Select template “Gene -> Interaction”, enter “PRP8” and show results

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

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

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

6. Create a second list of yeast orthologs of human genes associated with RP
7. Now intersect the two lists (PRP8 interactors and RP orthologs)

<table>
<thead>
<tr>
<th>Gene Primary DBID</th>
<th>Gene Systematic Name</th>
<th>Gene Organism . Short Name</th>
<th>Gene Standard Name</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>S000000259</td>
<td>YBR255C</td>
<td>S. cerevisiae</td>
<td>PRP8</td>
<td>Pre-mRNA Processing</td>
</tr>
<tr>
<td>S000000974</td>
<td>YER172C</td>
<td>S. cerevisiae</td>
<td>BRR2</td>
<td>Bad Response to Refrigeration</td>
</tr>
<tr>
<td>S000002881</td>
<td>YDR473C</td>
<td>S. cerevisiae</td>
<td>PRP3</td>
<td>Pre-mRNA Processing</td>
</tr>
<tr>
<td>S000006382</td>
<td>YPR178W</td>
<td>S. cerevisiae</td>
<td>PRP4</td>
<td>Pre-mRNA Processing</td>
</tr>
</tbody>
</table>

In fact, there is evidence that these 4 proteins are associated with Prp8p, as part of the U4/U6-U5 tri-snRNP spliceosome complex!
Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8

Sina Mozaffari-Jovin, Traudt Wandersleben, Karine F. Santos, Cindy L. Will, Reinhard Lührmann, Markus C. Wahl

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 Smu114 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 (Brr2H) with tandem helicase cassettes (18) in complex with hPrp8ab1 at 3.6 Å resolution (fig. S1 and table S1) (19). hPrp8ab1 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

80

5 JULY 2013  VOL 341  SCIENCE  www.sciencemag.org
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)
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