

Model Organism Databases

Saccharomyces Genome Database

Mike Cherry

Department of Genetics

Stanford University School of Medicine

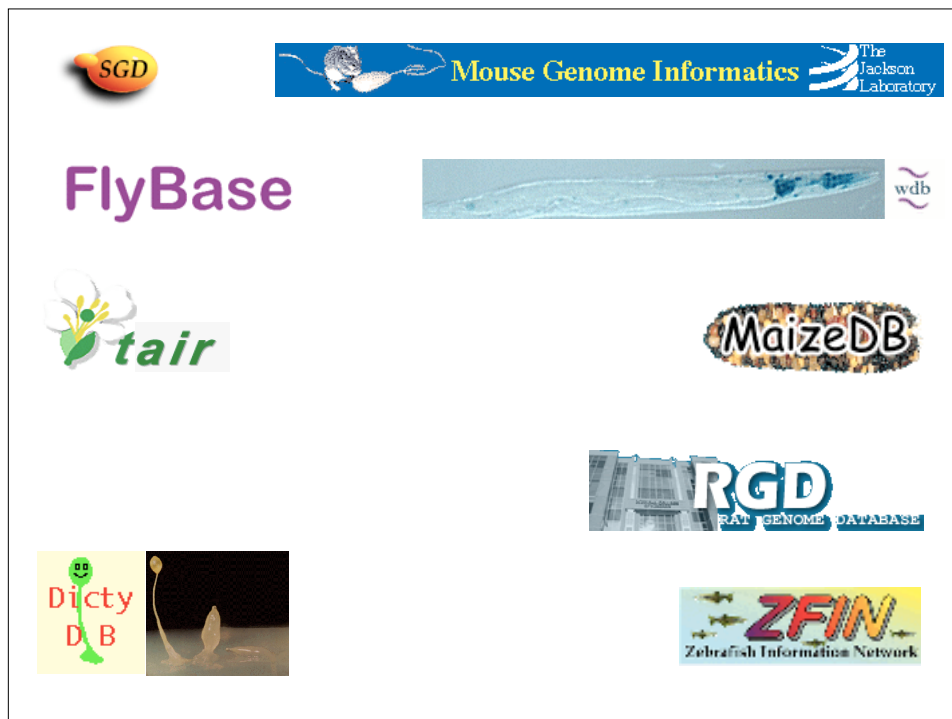
- Model Organism Databases
 - a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- Microarray Results
- Gene Ontology

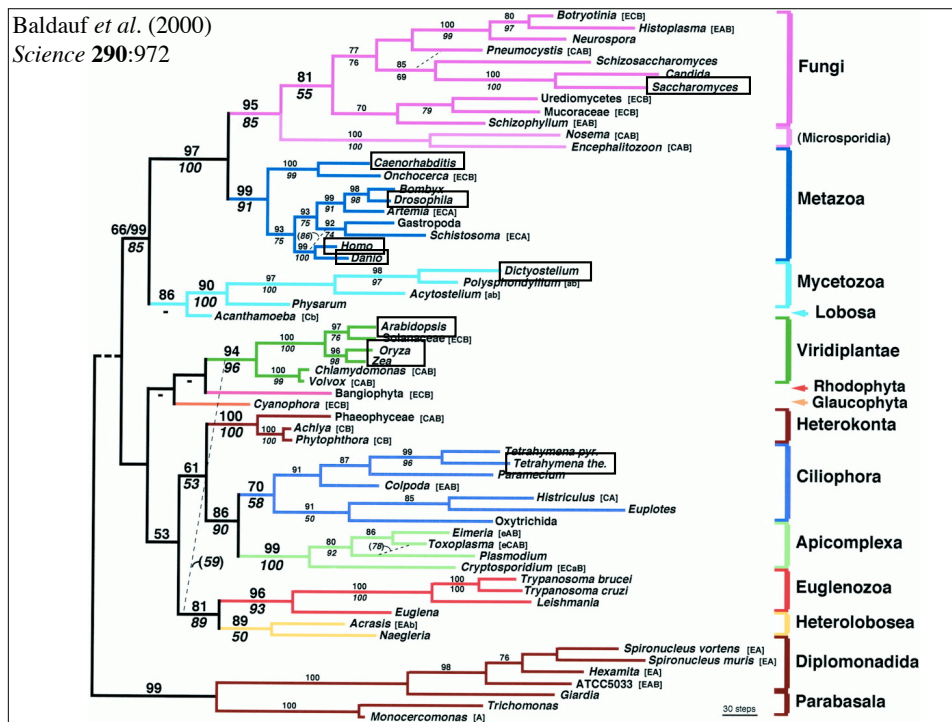
- **Model Organism Databases**
 - a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- Microarray Results
- Gene Ontology

Genome Databases

- Organism specific resource
- Typically includes any of the following
 - genetic and physical maps
 - DNA & protein sequences
 - Gene nomenclature
 - Summary of the literature
- Includes community specific data

- Common Entry Points
 - gene or ORF symbols (database IDs)
 - sequence (BLAST search)
 - full gene names
 - map location(s)
 - text search
 - hyperlinks from another resource
- Environment
 - Relational database
 - Perl CGI





- GoldenPath genome.ucsc.edu
- GDB www.gdb.org
- OMIM ncbi.nlm.nih.gov/Omim
- LocusLink ncbi.nlm.nih.gov/LocusLink
- GeneCards bioinformatics.weizmann.ac.il/cards
- MGD www.informatics.jax.org
- BDGP www.fruitfly.org
- FlyBase flybase.bio.indiana.edu
- TAIR www.arabidopsis.org
- MaizeDB www.agron.missouri.edu
- WormBase www.wormbase.org
- RGD rgd.mcg.edu
- SGD genome-www.stanford.edu
- ZFIN zfin.org

Generic Model Organism Database

www.gmod.org

- MODs working together to share software modules
- Modules built to shared
- Modules input/output using XML
 - DAS - www.biodas.org
 - GFF - www.sanger.ac.uk/Software/formats/GFF
- Open source, www.gmod.org
- BioPerl - bio.perl.org

- Model Organism Databases
 - a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- Microarray Results
- Gene Ontology

Curator?

- Museum
 - Conservator
 - Keeper
- Journal
 - Editor
- Educator
- Ontologist

Curation

- Information Acquisition
- Standardization
 - Nomenclature
 - Controlled Vocab.
- Connections
 - Collaborations for data or software
 - links & hyperlinks
- Quality Control
 - Thoughtful design, constant checks
 - design for what is known
- User Support
- Presentation
 - User interface
 - Needs

Trained professionals

- Ph.D. or postdoc in labs
- General scientific interests
- Work at a computer screen, and think
- Interested in the details
- Identify conflicting or ambiguous information
- Find missing details from the literature as a result of exploration

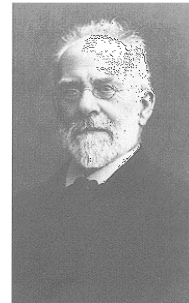
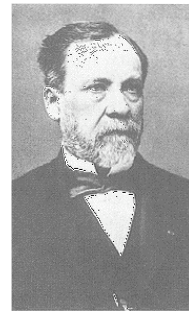
- Model Organism Databases
 - a.k.a. Genome Databases
- Curation
- **SGD**
- Sequence Comparison
- Microarray Results
- Gene Ontology

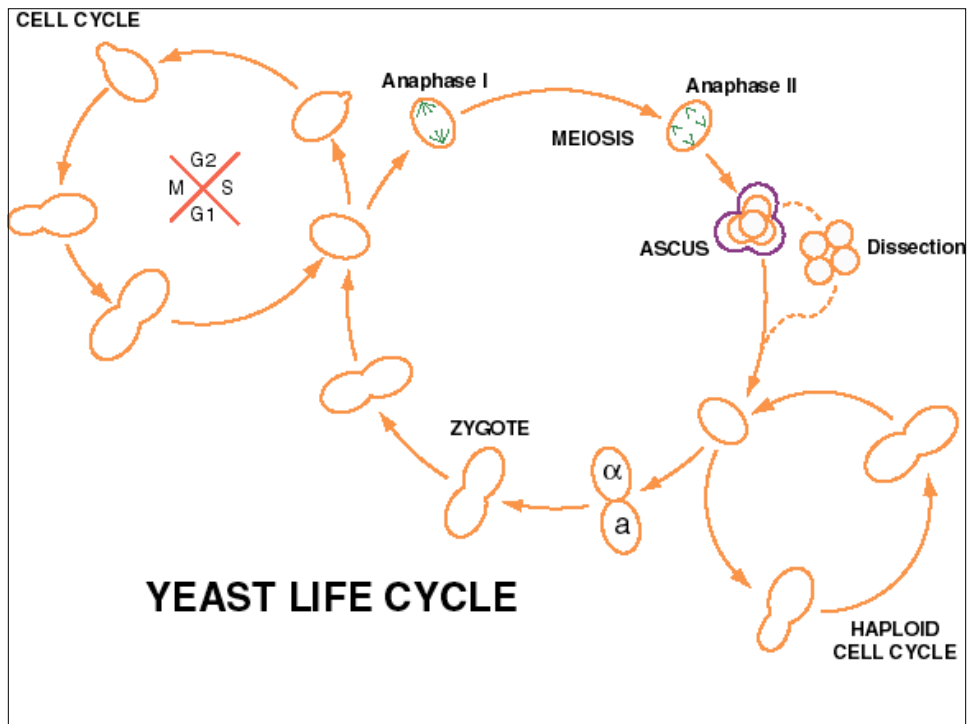
Ancient Brewers



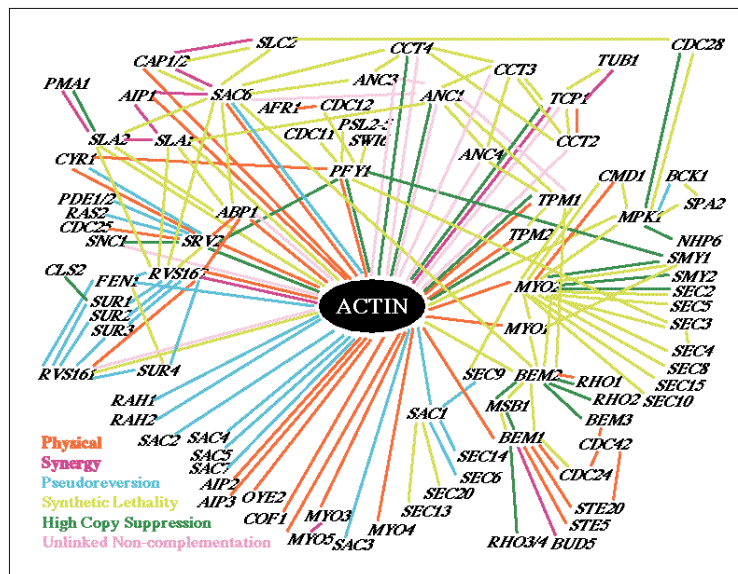
The Role of Yeast in Brewing

- 1866 - Louis Pasteur
 - Yeast was responsible for alcoholic fermentation.
- 1883 - Emil Christian Hansen
 - Developed pure culture technique
 - Isolated pure cultures of brewing yeasts





Physical & Genetics Interactions of Actin



CDC28/YBR160W - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Location: <http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?locus=cdc28>

SGD **CDC28/YBR160W** [Help](#)

[Search SGD](#) | [Gene/Seq Resources](#) | [Help](#) | [Gene Registry](#) | [Maps](#)
[BLAST](#) | [FASTA](#) | [PatMatch](#) | [Sacch3D](#) | [Primers](#) | [SGD Home](#)

CDC28 BASIC INFORMATION

Standard Name	<i>CDC28</i>
Alias	<i>CDK1, HSL5, SRM5</i>
Systematic Name	YBR160W
Feature Type	ORF
GO Molecular Function	<ul style="list-style-type: none">cyclin-dependent protein kinase
GO Biological Process	<ul style="list-style-type: none">cell cycle controlmitotic G1/S transitionmitotic G2/M transitionprotein phosphorylation
GO Cellular Component	<ul style="list-style-type: none">cytoplasm
Description	protein kinase catalytic subunit
Gene Product	protein kinase catalytic subunit
Phenotype	arrests at G(sub)1/S phase transition
Position	ChrII: coordinates 560036 to 560932 Genetic position: 90 Old format Sequence details

CDC28 RESOURCES

Click on map for expanded view

Literature
Gene_Info View

Retrieve Sequences
DNA (w/ introns) Retrieve

Sequence Analysis Tools
BLAST Analyze

Maps and Displays
Chr. Features Map View

Comparison Resources
Worm Homologs View

Document: Done

CDC28/YBR160W - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Location: <http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?locus=cdc28>

Alias	<i>CDK1, HSL5, SRM5</i>
Systematic Name	YBR160W
Feature Type	ORF
GO Molecular Function	<ul style="list-style-type: none">cyclin-dependent protein kinase
GO Biological Process	<ul style="list-style-type: none">cell cycle controlmitotic G1/S transitionmitotic G2/M transitionprotein phosphorylation
GO Cellular Component	<ul style="list-style-type: none">cytoplasm
Description	protein kinase catalytic subunit
Gene Product	protein kinase catalytic subunit
Phenotype	arrests at G(sub)1/S phase transition
Position	ChrII: coordinates 560036 to 560932 Genetic position: 90 Old format Sequence details
External Links	MIPS YFP SwissProt Entrez Protein Entrez Neighbors PIR-DE PIR-JP PIR-US Kyoto NiceZyme
Primary SGDID	S0000364

ADDITIONAL INFORMATION for *CDC28*

Locus History	Global Gene Hunter	Function
Mapping Data	Protein Info & Composition	Gene/Seq

CDC28 RESOURCES

Click on map for expanded view

Literature
Gene_Info View

Retrieve Sequences
DNA (w/ introns) Retrieve

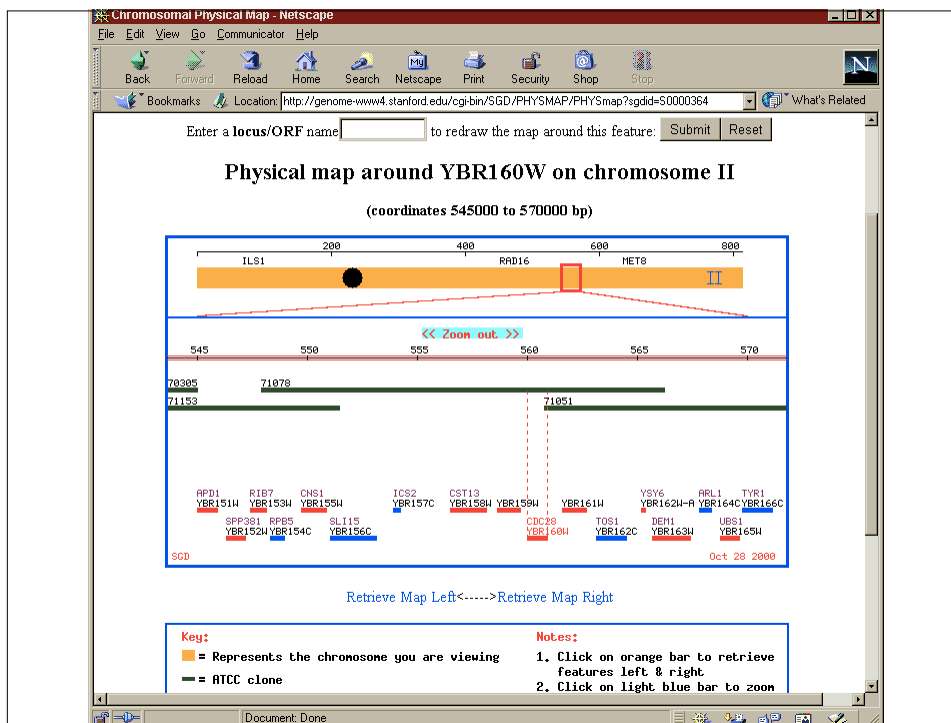
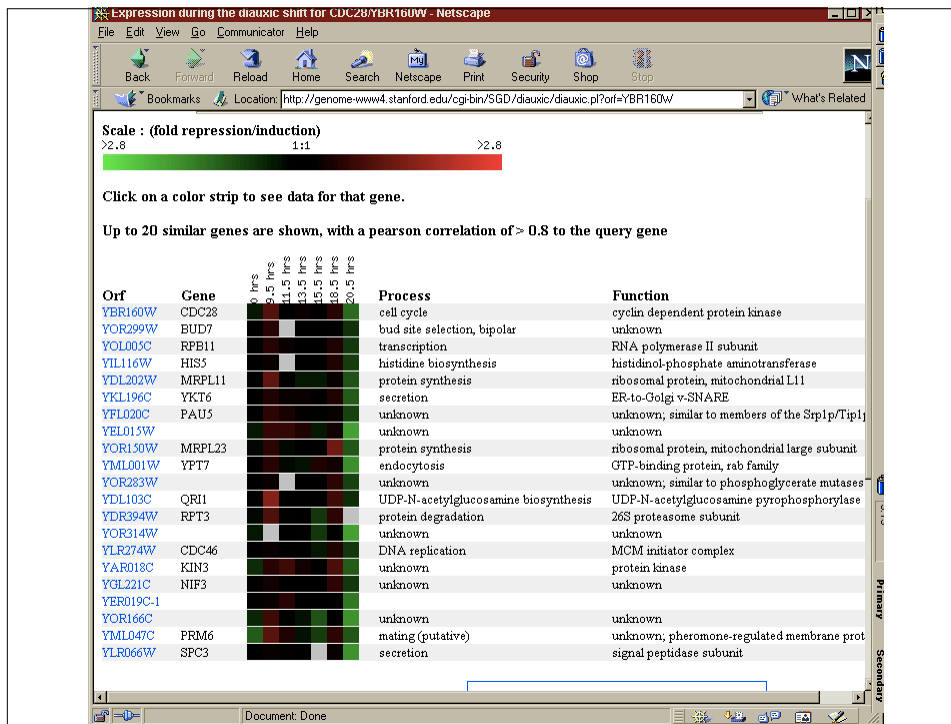
Sequence Analysis Tools
BLAST Analyze

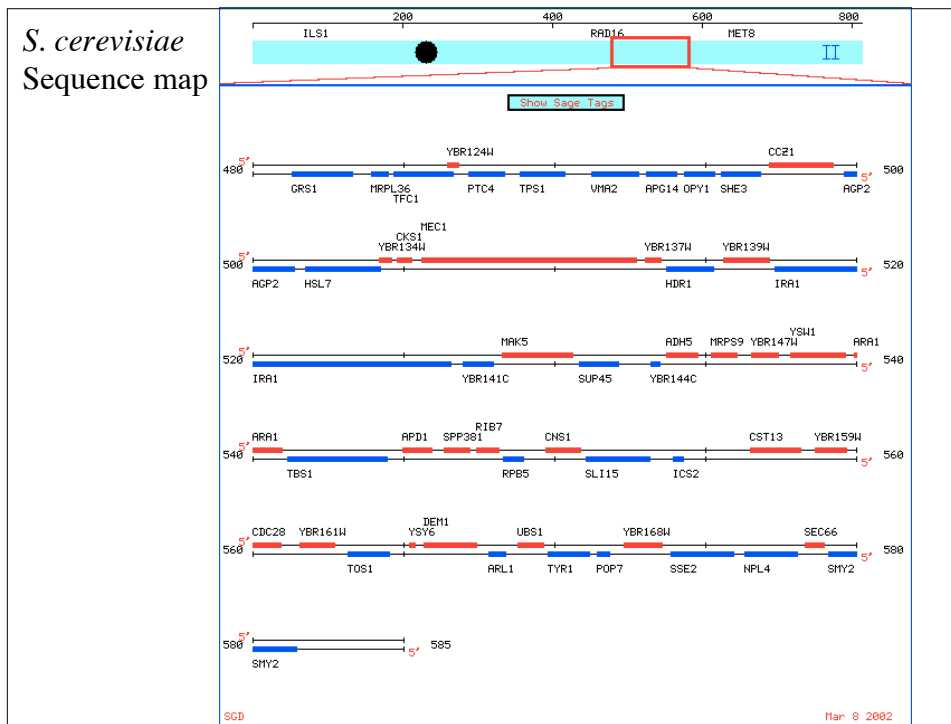
Maps and Displays
Chr. Features Map View

Comparison Resources
Worm Homologs View

Functional Analysis
Diauxic Shift View
Stanford Cell Cycle
Diauxic Shift
Sporulation Time Series
Glucose limitation
Portal PathCalling
YGAC Triples
UCLA Function Assignment

Document: Done





TEC1 Literature Guide [Help](#)

Search SGD: [Full Search](#) | [Gene/Seq Resources](#) | [Help](#) | [Gene Registry](#) | [Maps](#) | [BLAST](#) | [FASTA](#) | [PatMatch](#) | [Sacch3D](#) | [Primers](#) | [SGD Home](#)

TEC1 LITERATURE TOPICS (formerly Gene Info)

- Genetics/Cell Biology**
- Function/Process
- Genetic Interactions
- Mutants/Phenotypes
- Regulatory Role
- Nucleic Acid Information**
- DNA/RNA Sequence Features
- Mapping
- Transcription
- Protein Information**
- Protein-Nucleic Acid Interactions
- Protein-protein Interactions
- Substrates/Ligands/Cofactors
- Protein Sequence Features
- Related Genes/Proteins**
- Non-Yeast Related Genes/Proteins
- Yeast Related Genes/Proteins
- Research Aids**
- Alias
- Other Features
- Strains/Constructs
- Techniques and Reagents
- Genome-wide Analysis
- Curated Literature**

[TEC1 Locus Info](#)

TEC1 - Curated References

Reference	Other Genes Addressed
Liu H (2001) Transcriptional control of dimorphism in <i>Candida albicans</i> . <i>Curr Opin Microbiol</i> 4(6):728-35 SGD Curated Paper PubMed Access Full Text Access Full Text Access Full Text	CDC42 CPR1 NRG1 RIM101 TUP1
La Valle R and Wittenberg C (2001) A Role for the Swe1 Checkpoint Kinase During Filamentous Growth of <i>Saccharomyces cerevisiae</i> . <i>Genetics</i> 158(2):549-62 SGD Curated Paper PubMed Access Full Text	FLOS HSL1 HSL7 SWE1
Schweizer A, et al. (2000) The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in <i>Candida albicans</i> . <i>Mol Microbiol</i> 38(3):435-45 SGD Curated Paper PubMed Access Full Text	MUC1 SAP4
Lorenz MC, et al. (2000) Characterization of alcohol-induced filamentous growth in <i>Saccharomyces cerevisiae</i> . <i>Mol Biol Cell</i> 11(1):183-99 SGD Curated Paper PubMed Access Full Text	BEM1 BEM4 BUD8 CHD1 FIG1 GPA2 GPR1 HEM1 MEP2 MRP21 MSM1 STE11 STE12 STE20 MORE
Rupp S, et al. (1999) MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of	FLOS FUS3 HO IME1 KSS1 MUC1 STE12

Summary: Topoisomerases catalyze the interconversion between topological states of DNA by breaking and rejoining DNA strands. These changes in DNA topology are required during several cellular processes such as replication, transcription, recombination, and chromosome condensation (1). There are three classes of topoisomerases that are distinguished by substrate (IA, IB, II). Type I topoisomerases cleave one DNA strand, while Type II enzymes cleave a pair of complementary DNA strands (2). The type IB topoisomerases relax both positively and negatively supercoiled DNA, TOP1 encodes the type IB enzyme in yeast (3, 4) Type IA topoisomerase, encoded by TOP3 in yeast, relaxes only negatively supercoiled DNA, and yeast topoisomerase II is encoded by the TOP2 gene (2).









Topoisomerases are highly conserved; yeast Top1p shares 57% identity with human *Top1* (5). The *Top1* protein, like other type IB topoisomerases, relaxes supercoiled DNA by forming a DNA-enzyme complex and transiently cleaving one strand via a nucleophilic attack that results in a covalent linkage with the 3' end of the cleaved strand. The 5' end can then rotate freely (1).

Top1p is the target of the antitumor drug camptothecin (6). Camptothecin increases the half-life of the enzyme-DNA complex, which results in double-stranded DNA breaks during DNA replication (7). Specific amino acid substitutions in Top1p have the same effect as the drug (8). Suppressors of these mutations were identified that reduced the enzyme's affinity for DNA (9).

Created: 1999-11-09 KARA

Modified: 1999-11-09 KARA

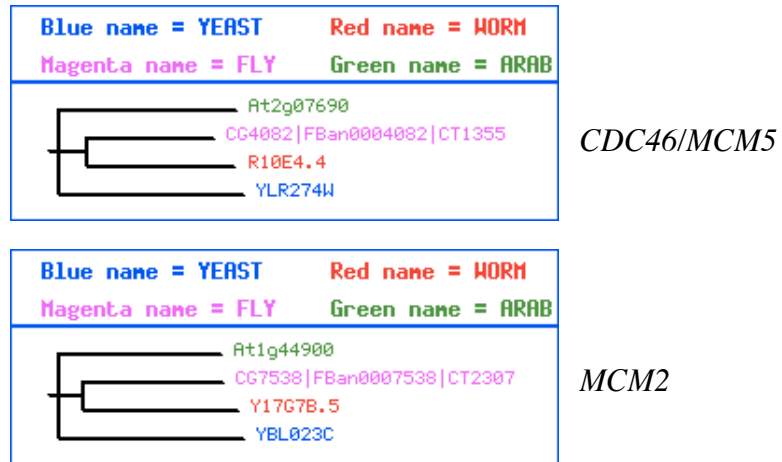
Summary References:

- 1) **Berger JM (1998)** Structure of DNA topoisomerases. *Biochim Biophys Acta* 1400(1-3):3-18
 
- 2) **Wang JC (1996)** DNA topoisomerases. *Annu Rev Biochem* 65():635-92
 
- 3) **Thrash C, et al. (1985)** Cloning, characterization, and sequence of the yeast DNA topoisomerase I gene. *Proc Natl Acad Sci U S A* 82(13):4374-8
 
- 4) **Goto T and Wang JC (1985)** Cloning of yeast TOP1, the gene encoding DNA topoisomerase I, and construction of mutants defective in both DNA topoisomerase I and DNA topoisomerase II. *Proc Natl Acad Sci U S A* 82(21):7178-82
 

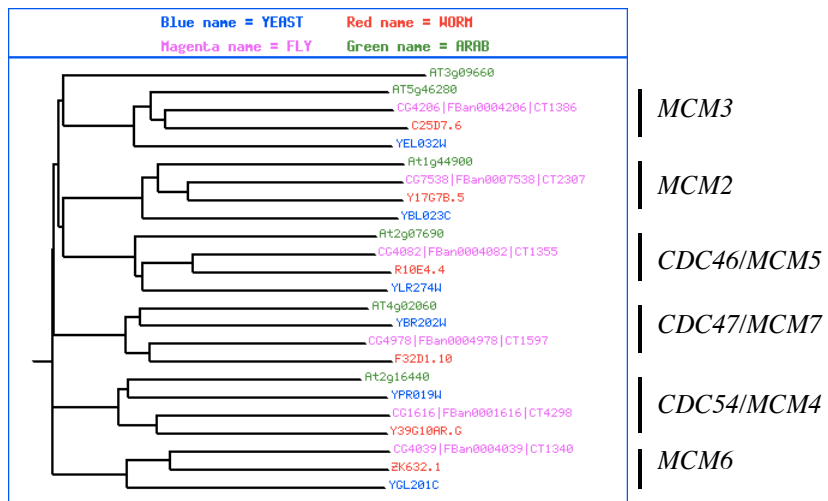
- Model Organism Databases
– a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- Microarray Results
- Gene Ontology

DNA Replication Initiation Complex: MCM Proteins

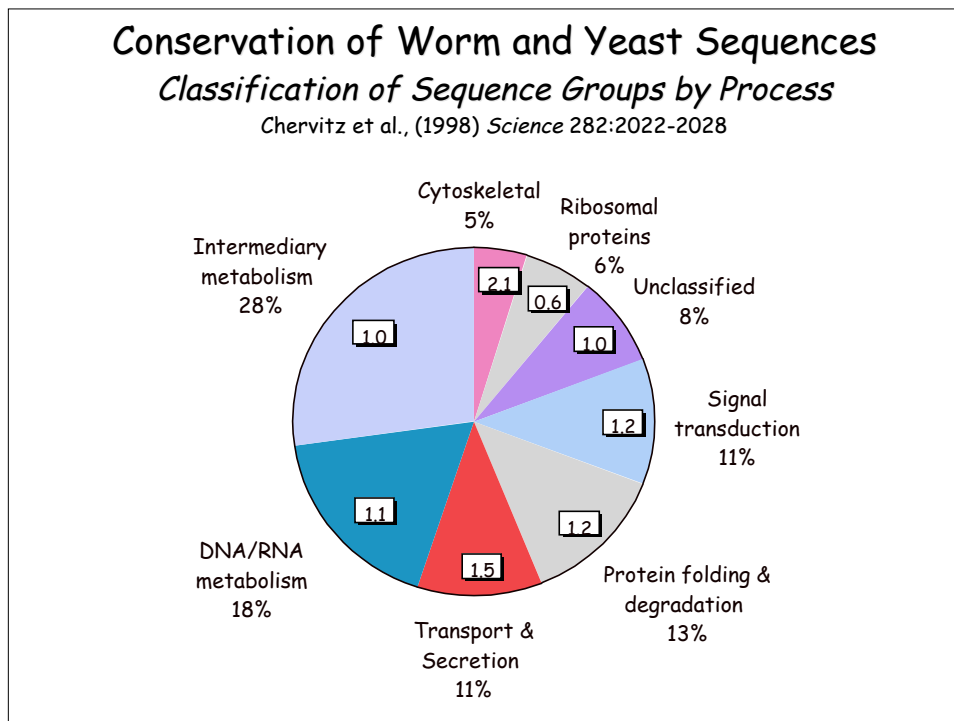
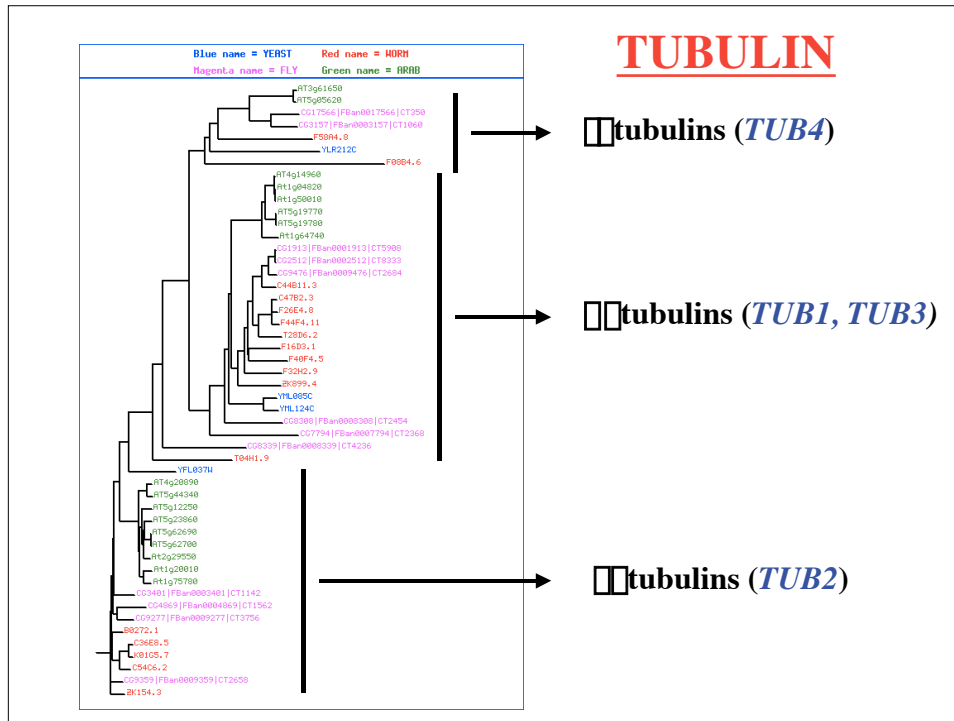
Results of Highest Stringency BLAST Search



DNA Replication Initiation Complex: MCM Proteins

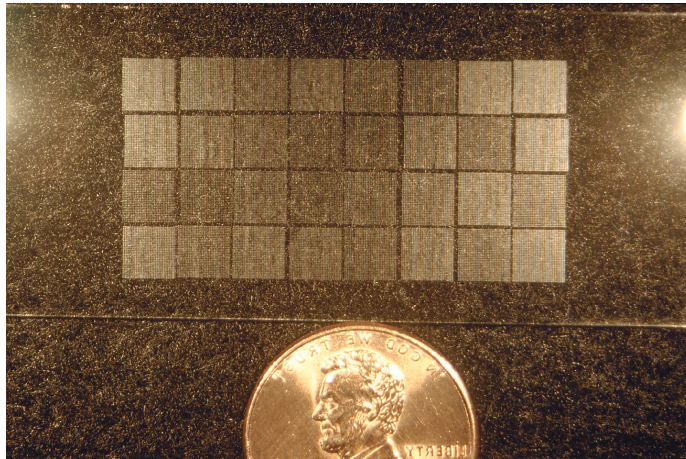


These proteins form a hexamer in the species that have been examined



- Model Organism Databases
 - a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- **Microarray Results**
- Gene Ontology

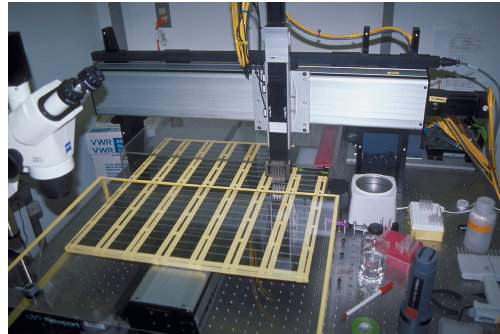
DNA Microarray



18,000 human cDNA clone inserts

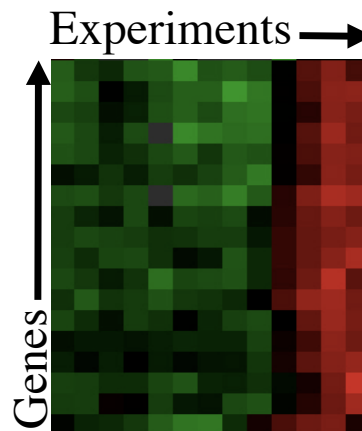
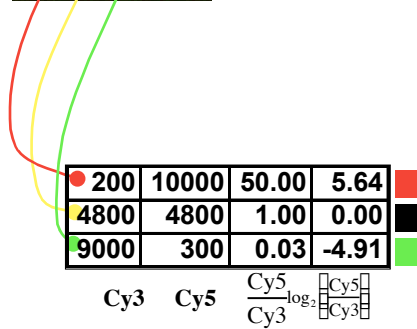
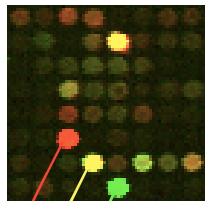
Microarray technology

- Robotically spotted PCR products
 - cDNA clone inserts (human, mouse, Arabidopsis)
 - ORFs amplified from genomic DNA (yeast, worm, *E.coli*)



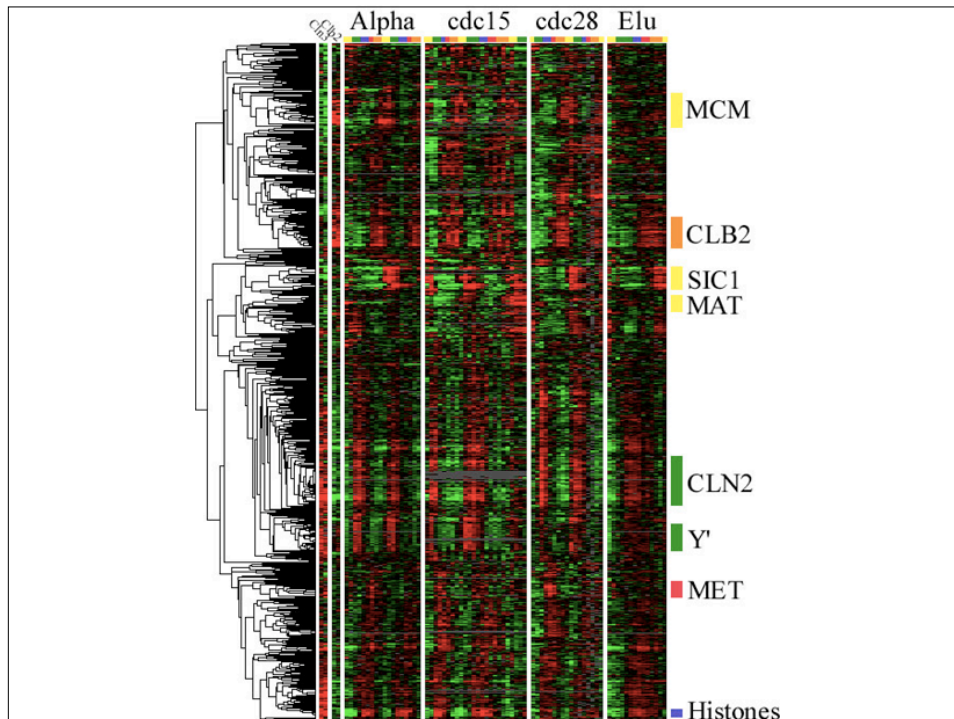
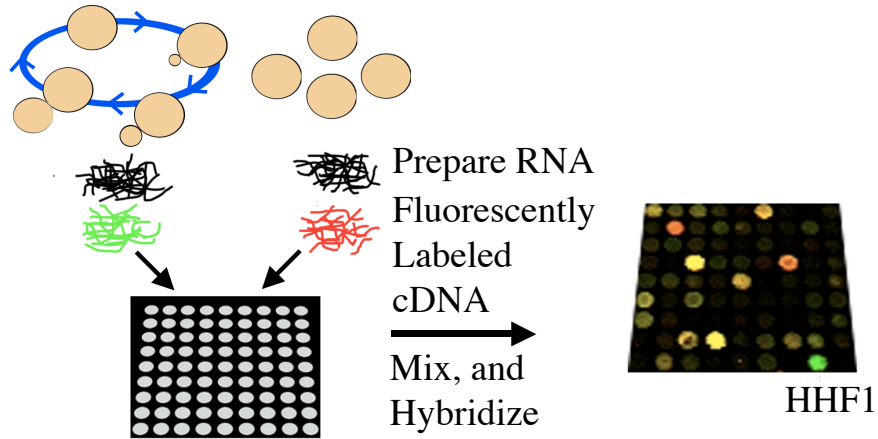
URL: <http://cmgm.stanford.edu/pbrown>

Extracting Data



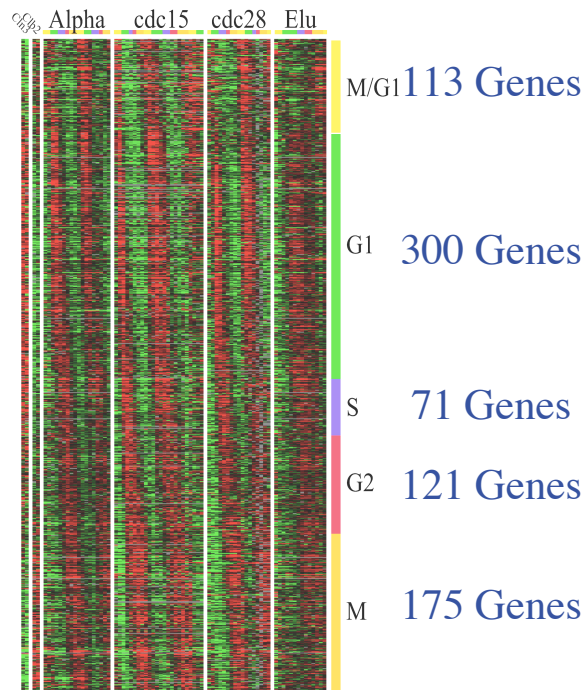
The Microarray Procedure

Asynchronous Synchronized



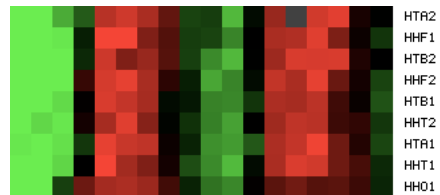
'Phasogram'

800
periodically
expressed
genes sorted
by phase:

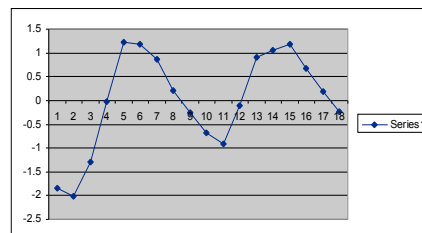


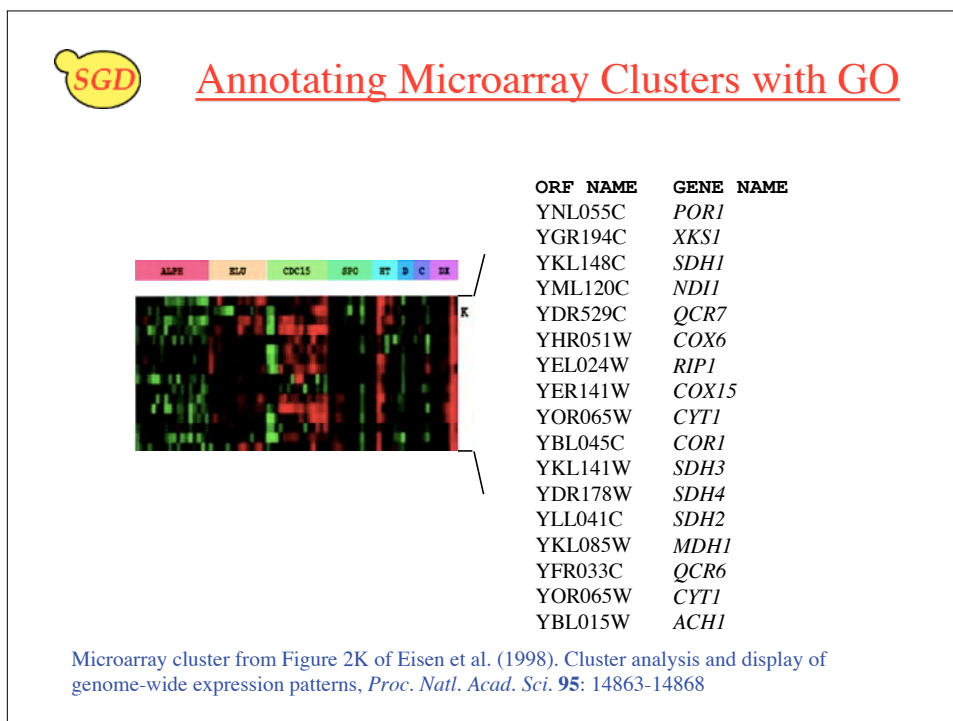
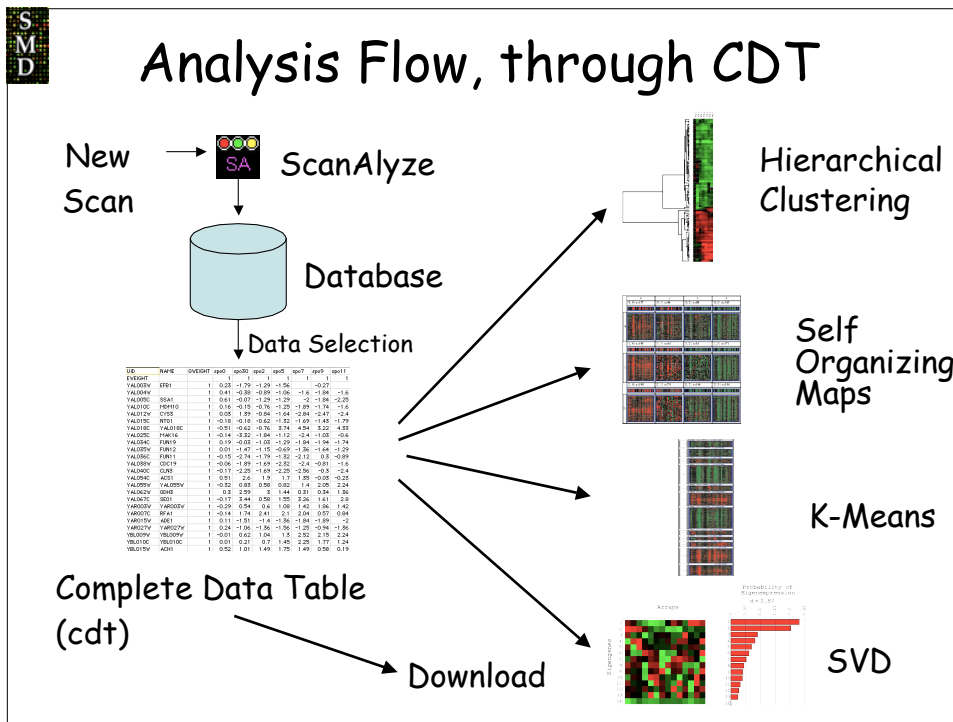
Identifying correlated gene expression

The experiments
are in columns
containing a unique
gene in each row



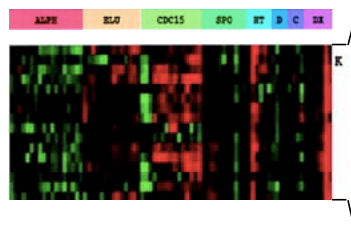
Build a series of
profiles and calculate
correlations for each
gene to each profile.







Annotating Microarray Clusters with GO



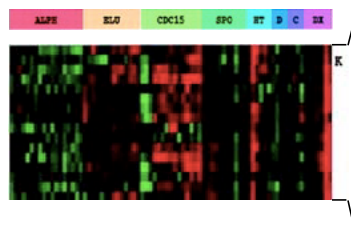
FUNCTION

voltage-gated ion channel
xylulokinase
succinate dehydrogenase
NADH dehydrogenase
ubiquinol--cytochrome-c reductase
cytochrome-c oxidase
Rieske Fe-S protein
cytochrome-c oxidase
cytochrome-c1
ubiquinol--cytochrome-c reductase
succinate dehydrogenase subunit
succinate dehydrogenase subunit
succinate dehydrogenase subunit
malate dehydrogenase
ubiquinol--cytochrome-c reductase
cytochrome-c1
acetyl-CoA hydrolase

Microarray cluster from Figure 2K of Eisen et al. (1998). Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci.* **95**: 14863-14868



Annotating Microarray Clusters with GO



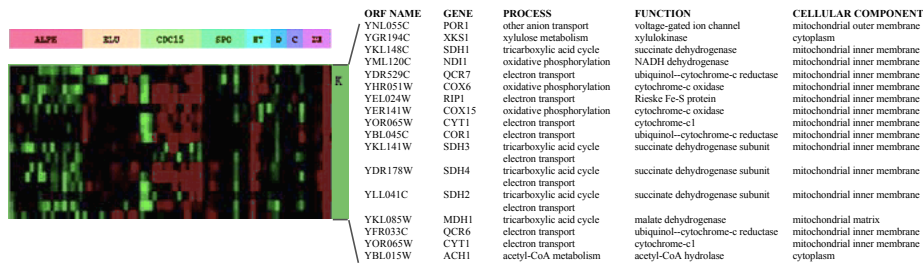
PROCESS

other anion transport
xylulose metabolism
tricarboxylic acid cycle
oxidative phosphorylation
electron transport
oxidative phosphorylation
electron transport
oxidative phosphorylation
electron transport
electron transport
tricarboxylic acid cycle & electron transport
tricarboxylic acid cycle & electron transport
tricarboxylic acid cycle & electron transport
tricarboxylic acid cycle
electron transport
electron transport
acetyl-CoA metabolism

Microarray cluster from Figure 2K of Eisen et al. (1998). Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci.* **95**: 14863-14868



Annotating Microarray Clusters with GO



Microarray data from Figure 2K of Eisen et al. (1998). Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci.* **95** (25): 14863-14868.



Annotating Microarray Clusters with GO

ORF NAME	GENE	PROCESS	FUNCTION	CELLULAR COMPONENT
YNL055C	POR1	other anion transport	voltage-gated ion channel	mitochondrial outer membrane
YGR194C	XKS1	xylulose metabolism	xylulokinase	cytoplasm
YKL148C	SDH1	tricarboxylic acid cycle	succinate dehydrogenase	mitochondrial inner membrane
YML120C	NDH1	oxidative phosphorylation	NADH dehydrogenase	mitochondrial inner membrane
YDR529C	QCR7	electron transport	ubiquinol-cytochrome-c reductase	mitochondrial inner membrane
YHR051W	COX6	oxidative phosphorylation	cytochrome-c oxidase	mitochondrial inner membrane
YEL024W	RIP1	electron transport	Rieske Fe-S protein	mitochondrial inner membrane
YER141W	COX15	oxidative phosphorylation	cytochrome-c oxidase	mitochondrial inner membrane
YOR065W	CYT1	electron transport	cytochrome-c1	mitochondrial inner membrane
YBL045C	COR1	electron transport	ubiquinol-cytochrome-c reductase	mitochondrial inner membrane
YKL141W	SDH3	tricarboxylic acid cycle	succinate dehydrogenase subunit	mitochondrial inner membrane
YDR178W	SDH4	tricarboxylic acid cycle	succinate dehydrogenase subunit	mitochondrial inner membrane
YLL041C	SDH2	tricarboxylic acid cycle	succinate dehydrogenase subunit	mitochondrial inner membrane
YKL085W	MDH1	tricarboxylic acid cycle	malate dehydrogenase	mitochondrial matrix
YFR033C	QCR6	electron transport	ubiquinol-cytochrome-c reductase	mitochondrial inner membrane
YOR065W	CYT1	electron transport	cytochrome-c1	mitochondrial inner membrane
YBL015W	ACH1	acetyl-CoA metabolism	acetyl-CoA hydrolase	cytoplasm

- Model Organism Databases
 - a.k.a. Genome Databases
- Curation
- SGD
- Sequence Comparison
- Microarray Results
- Gene Ontology

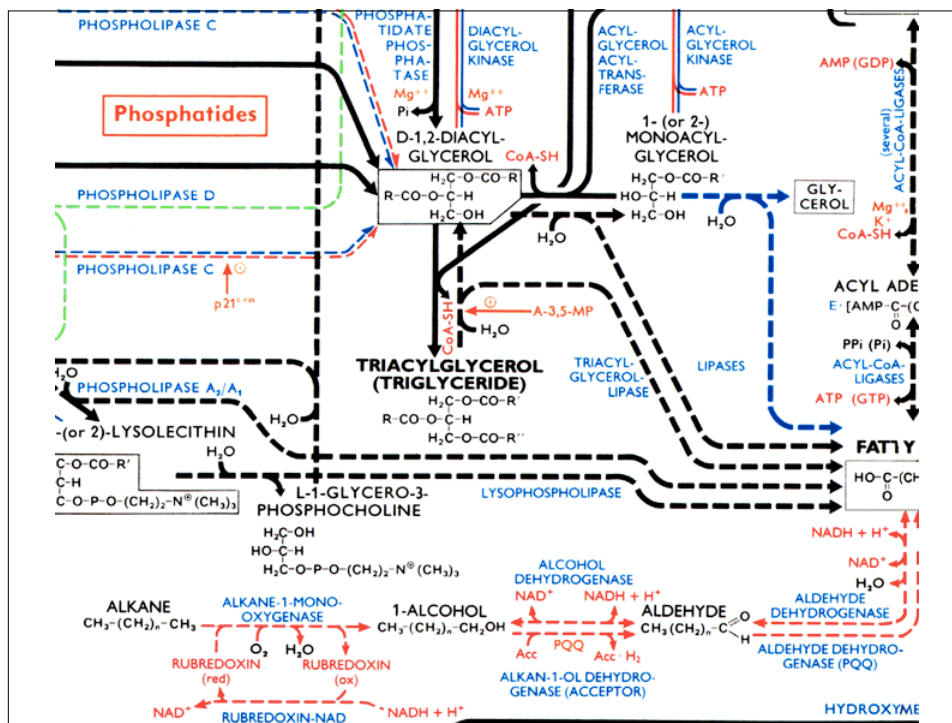
Knowledge of a protein's biological process in one organism can often be transferred to another organism.

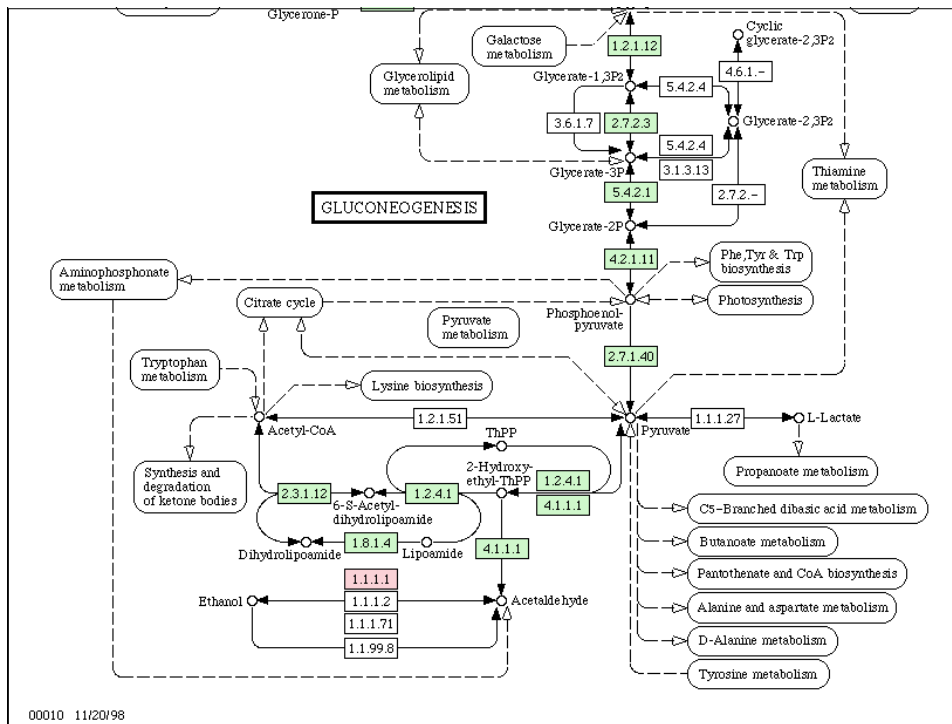
The goal of the *Gene Ontology Consortium* is to produce a dynamic, controlled vocabulary that can be applied to all eukaryotes (& prokaryotes) even as knowledge of gene and protein roles in cells is accumulating and changing.

The Categories of GO

Gene Ontology

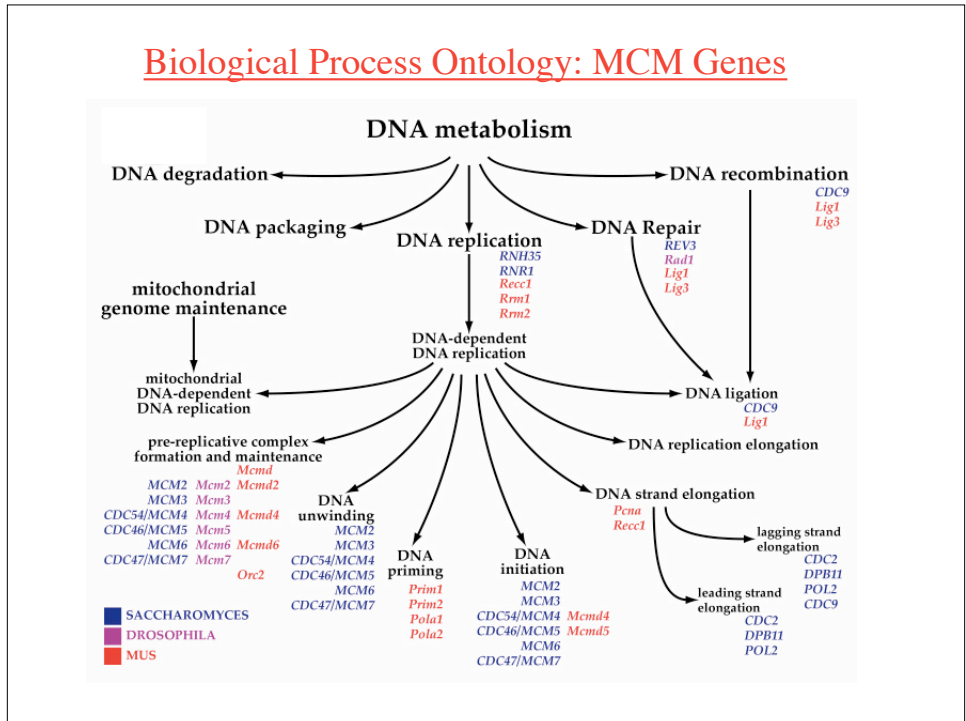
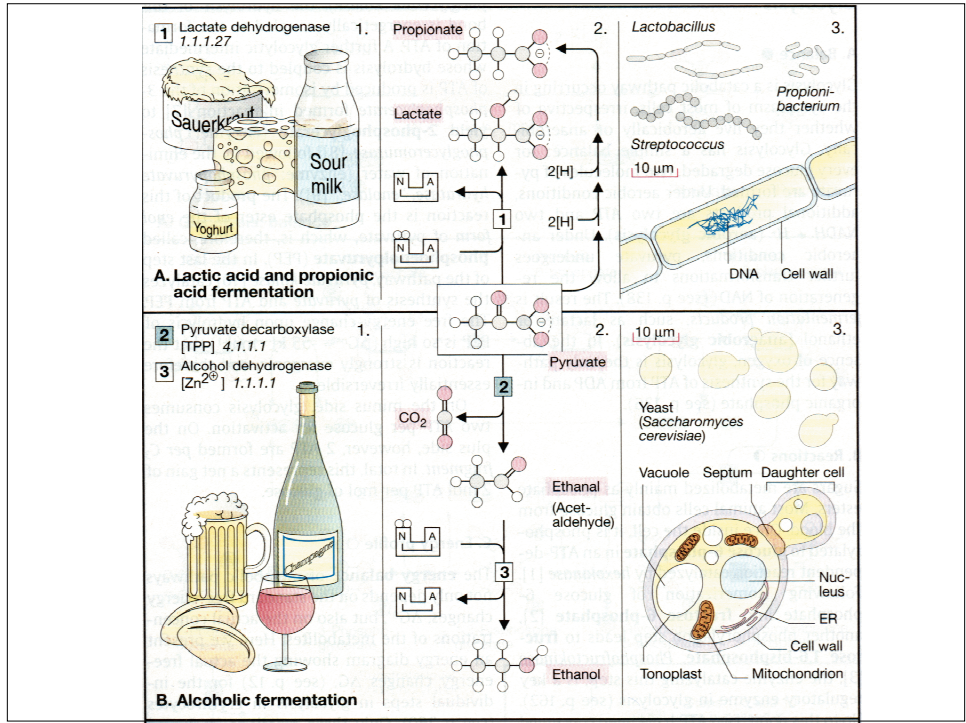
- *Biological Process* = goal or objective
(e.g. ion transport, budding, stress response)
- *Molecular Function* = elemental activity/task
(e.g. ATPase, polymerase, protein kinase)
- *Cellular Component* = location or complex
(e.g. membrane protein, ribosomal subunit)



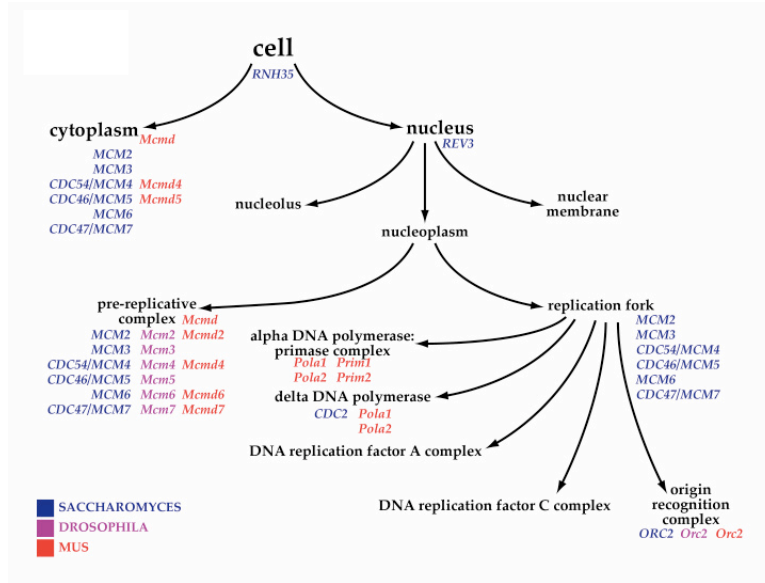


WormPD™ 56 Functional Categories

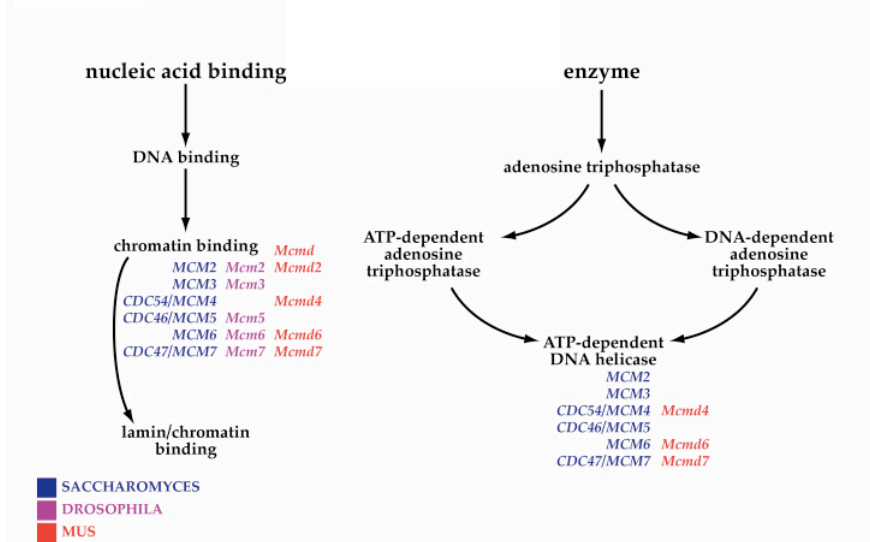
142	ATPase	6	Nuclear import/export protein
36	ATP-binding cassette	126	Nuclease [endo, exo, ribo, deoxyribo]
22	Activator	31	Other
85	Active transporter, primary	80	Other kinase
231	Active transporter, secondary	32	Other phosphatase
31	Adhesin/agglutinin	372	Oxidoreductase
4	Anchor Protein	194	Protease (other than proteasomal)
223	Channel [passive transporter]	41	Proteasome subunit
49	Chaperones	34	Protein conjugation factor
20	Complex assembly protein	381	Protein kinase
12	Conserved ATPase domain	106	Protein phosphatase
10	Cyclin	25	RNA polymerase subunit
68	DNA polymerase or subunit	214	RNA-binding protein
429	DNA-binding protein	10	Receptor (protein translocation)
19	Docking protein	1204	Receptor (signalling)
21	GTPase activating protein	28	Regulatory subunit
119	GTP-binding protein/GTPase	5	Ribosomal subunit
16	Guanine nucleotide exchange factor	237	Small molecule-binding protein
30	Heat shock protein	8	Spliceosomal subunit
70	Helicase	245	Structural protein
698	Hydrolase	9	Topoisomerase
32	Inhibitor or repressor	263	Transcription factor
55	Isomerase	614	Transferase
26	Ligand	44	Translation factor
85	Ligase	423	Transporter
99	Lyase	11136	Unknown
41	Major Facilitator Superfamily	2	Vesicle coat protein
51	Motor protein	30	tRNA synthetase

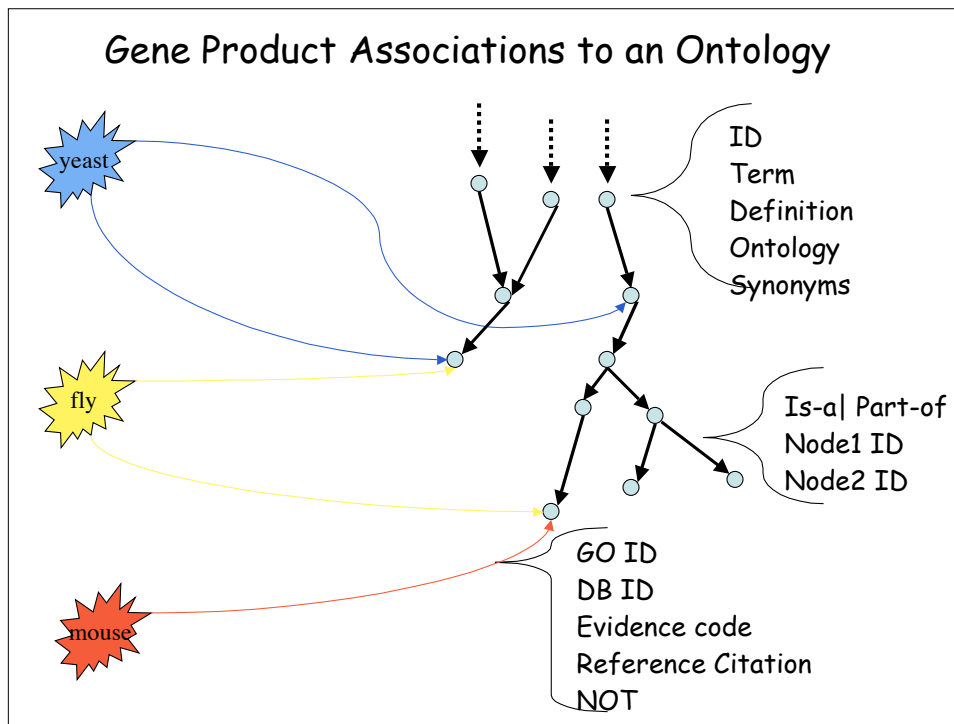


Cellular Location Ontology: MCM Genes



Molecular Function Ontology: MCM Genes





<http://genome-www.stanford.edu/Saccharomyces/>

The screenshot shows a Netscape browser window displaying the SGD (Saccharomyces Genome Database) page for the RPN11/YFR004W gene. The browser title is "RPN11/YFR004W - Netscape". The address bar shows the URL: <http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?sgdid=50001900>. The page content includes the SGD logo, a "Help" button, and navigation links: Search SGD, Gene/Seq Resources, Help, Gene Registry, Maps, BLAST, FASTA, PatMatch, Sacch3D, Primers, and SGD Home.

RPN11 BASIC INFORMATION

Standard Name	RPN11
Alias	MPR1
Systematic Name	YFR004W
Feature Type	ORF
Molecular Function	<ul style="list-style-type: none"> multicatalytic endopeptidase
Biological Process	<ul style="list-style-type: none"> ubiquitin-dependent protein degradation
Cellular Component	<ul style="list-style-type: none"> 19S proteasome regulatory particle
Description	Suppressor of mutant (ts on glycerol) tRNA gene deficient in the processing of its 3'-end; homologous to <i>S. pombe</i> PAD1 gene - global positive regulator of nuclear transcription and is involved in maintenance of chromatin structure

RPN11 RESOURCES

Click on map for expanded view

The genomic map shows the RPN11 gene structure with coordinates 151000 to 156000. The gene is flanked by NICO6 and YFR03C S801. The map shows the 5' and 3' ends of the gene.

- Literature**: Gene_Info View
- Retrieve Sequences**: DNA (w/ introns) Retrieve
- Sequence Analysis Tools**: BLASTP Analyze

GO: ubiquitin-dependent protein degradation

Search SGD | Gene/Seq Resources | Help | Gene Registry | Maps
BLAST | FASTA | PatMatch | Sacch3D | Primers | SGD Home

ubiquitin-dependent protein degradation (GO:0006511): (biological process ontology).

The following genes have been annotated to this term:

Locus	Reference(s)	Evidence
APC1	Zachariae, W. and K. Nasmyth (1999) Whose end is destruction: cell division and the anaphase-promoting complex. <i>Genes Dev</i> 13:2039-2058	TAS
APC2	Zachariae, W. and K. Nasmyth (1999) Whose end is destruction: cell division and the anaphase-promoting complex. <i>Genes Dev</i> 13:2039-2058	TAS
APC4	Zachariae, W. and K. Nasmyth (1999) Whose end is destruction: cell division and the anaphase-promoting complex. <i>Genes Dev</i> 13:2039-2058	TAS
APC5	Zachariae, W. and K. Nasmyth (1999) Whose end is destruction: cell division and the anaphase-promoting complex. <i>Genes Dev</i> 13:2039-2058	TAS
APC9	Zachariae, W. and K. Nasmyth (1999) Whose end is destruction: cell division	TAS

Gene Ontology provides ...

- Three structured controlled vocabularies
 - ID, Term, Definition, Synonyms
- Expert groups provide high quality associations for gene products
 - IDs, Evidence, Citations
- Software
 - Web site for open source distribution.
 - Database for vocabularies & associations
 - Browsers (Java applet and HTML)