PortEco overview
PortEco search
    Search for information about a particular gene
Gene expression
    What genes are induced/repressed in some set of experiments?
    What pathways are upregulated?
    What genes have expression patterns similar to my favorite gene?
ChIP data
Chemical Genomics Phenotypic Landscape
        Strain vs Condition
        Strain vs Strain
        Condition vs Condition
Annotation with Students (CACAO)
    What is GO?
    Browsing GO

PortEco overview

PortEco is a portal for *E. coli* research (K-12 strains and their phage and mobile elements) that aims to:

1. facilitate access to *E. coli* information that is distributed over the web
2. make *E. coli* genomics data (currently gene expression data and chemical genomics data) easy to access, search and analyze.
3. enable the community to add information to the knowledgebase via EcoliWiki
4. provide community features such as a calendar, colleague search, blog entries mentioning *E. coli*; and educational materials

Access to *E. coli* information
PortEco search: aggregated search results for 14 different web resources
EcoliHouse: a data warehouse with information from multiple data resources, including EcoCyc and EcoGene

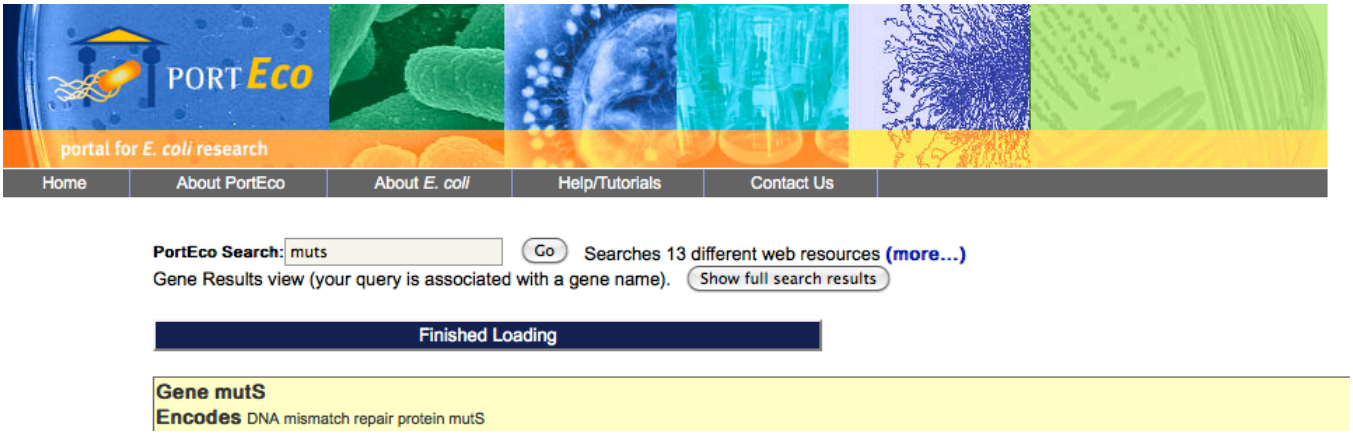
Genomics data
Gene expression data and analysis tools at http://expression.porteco.org
Chemical genomics data at http://ecoliwiki.net/tools/chemgen/

PortEco search

PortEco Search (search box at http://porteco.org/, or page at http://porteco.org/AjaxSearch.jsp) launches searches of 14 different web resources for *E. coli* information, and organizes the results onto a single results page. The page loads each result as it comes back from the

respective web resource: the progress bar on the page keeps track of how many resources have returned results.

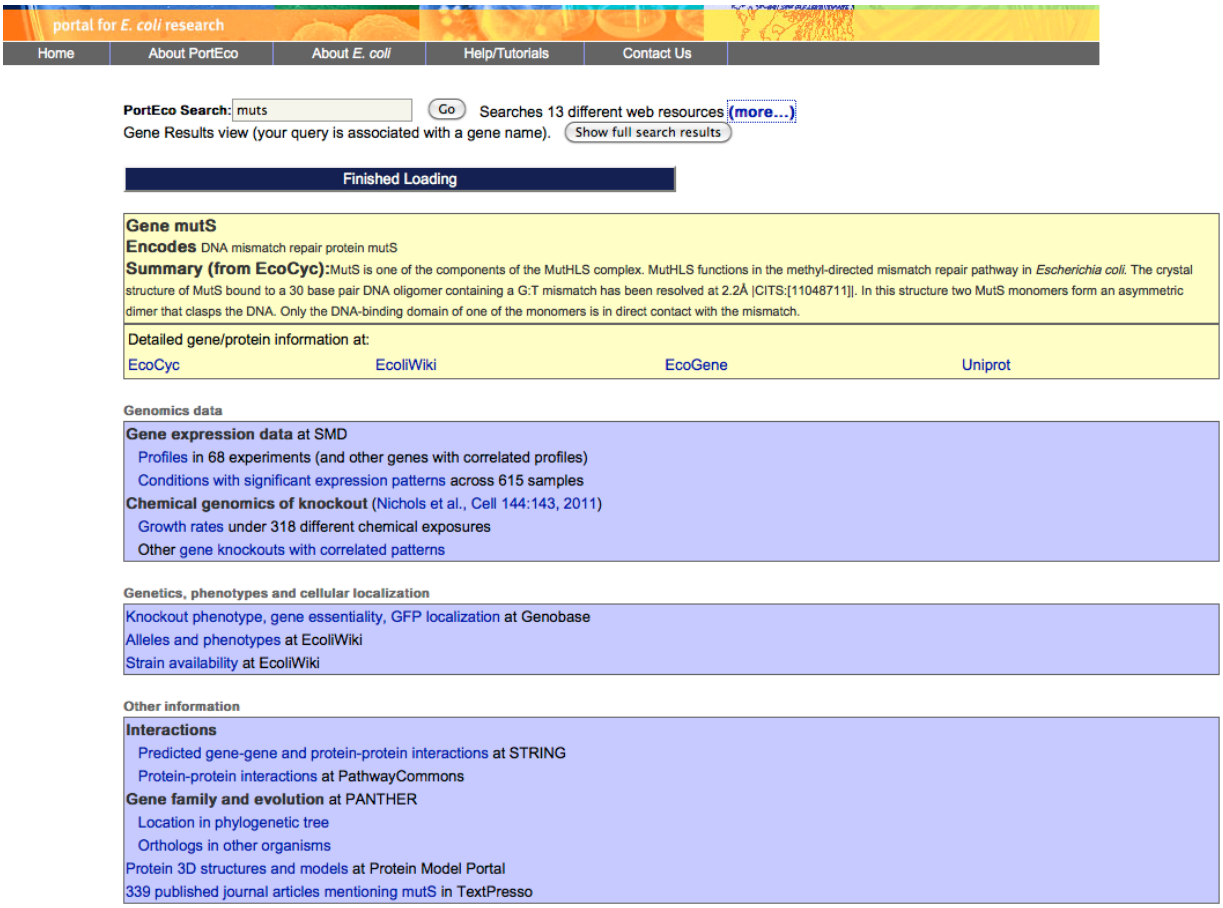
If an *E. coli* gene name is entered in the search box, the results are formatted into the "Gene Results view." The full results display can be shown by clicking on the "Show full search results" button:



The resources searched are listed at http://porteco.org/help/general.jsp.

Search for information about a particular gene

Try entering a gene name in the search box, and press the **Search** or **Go** button. An example is shown below for the search string "muts" (note that the search is not case-sensitive):



Search for general information

You can enter other search terms as well. For example, you can search for

- a person in the EcoliWiki colleague pages (e.g. “paul thomas”, see results page),
- a meeting in our PortEco Calendar (e.g. “phage meeting”, see results page),
- a general biological search term (e.g. “kinase”, see results page).

In this view, you can jump to different sections of the page by clicking on the name of the section in blue, in the yellow summary box:

For example, clicking on “Pathways and Interactions” will scroll down to the section of the page with results of searches at websites that have information about pathways and molecular interactions in *E. coli*, that contain the search term.

Gene expression

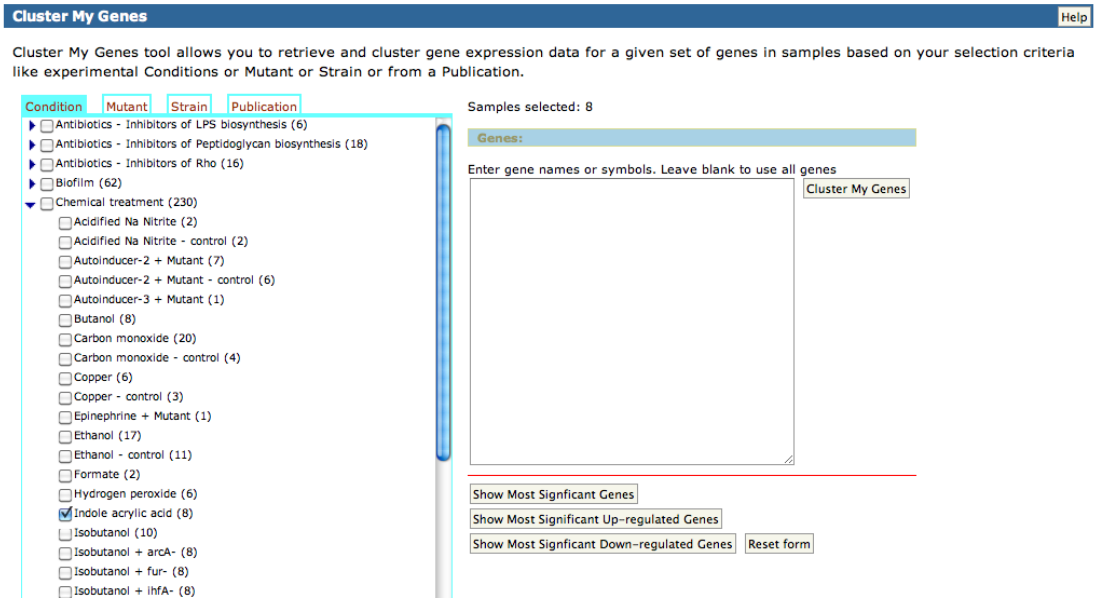
expression.porteco.org offers a variety of tools to let you explore expression data for *E. coli*. We have been curating expression data available from GEO and ArrayExpress to allow comparisons across different studies. Although we have loaded over 1,000 arrays from 75 publications, this is only about 35% of the available studies for *E. coli*.

The results of analyses we will demo in this workshop are likely to change as additional datasets are added, and as we incorporate data from RNA-seq studies.

What genes are induced/repressed in some set of experiments?

We will find what genes have significant expression changes when subjected to one of the chemical treatments studied in the set of transcriptome experiments in the database.

- Choose Cluster My Genes.
- Click Chemical Treatments.
- Select Indole Acrylic Acid.
- Click the **Most Significant Genes** button



You should get a list like this:

Individual conditions: Indole acrylic acid

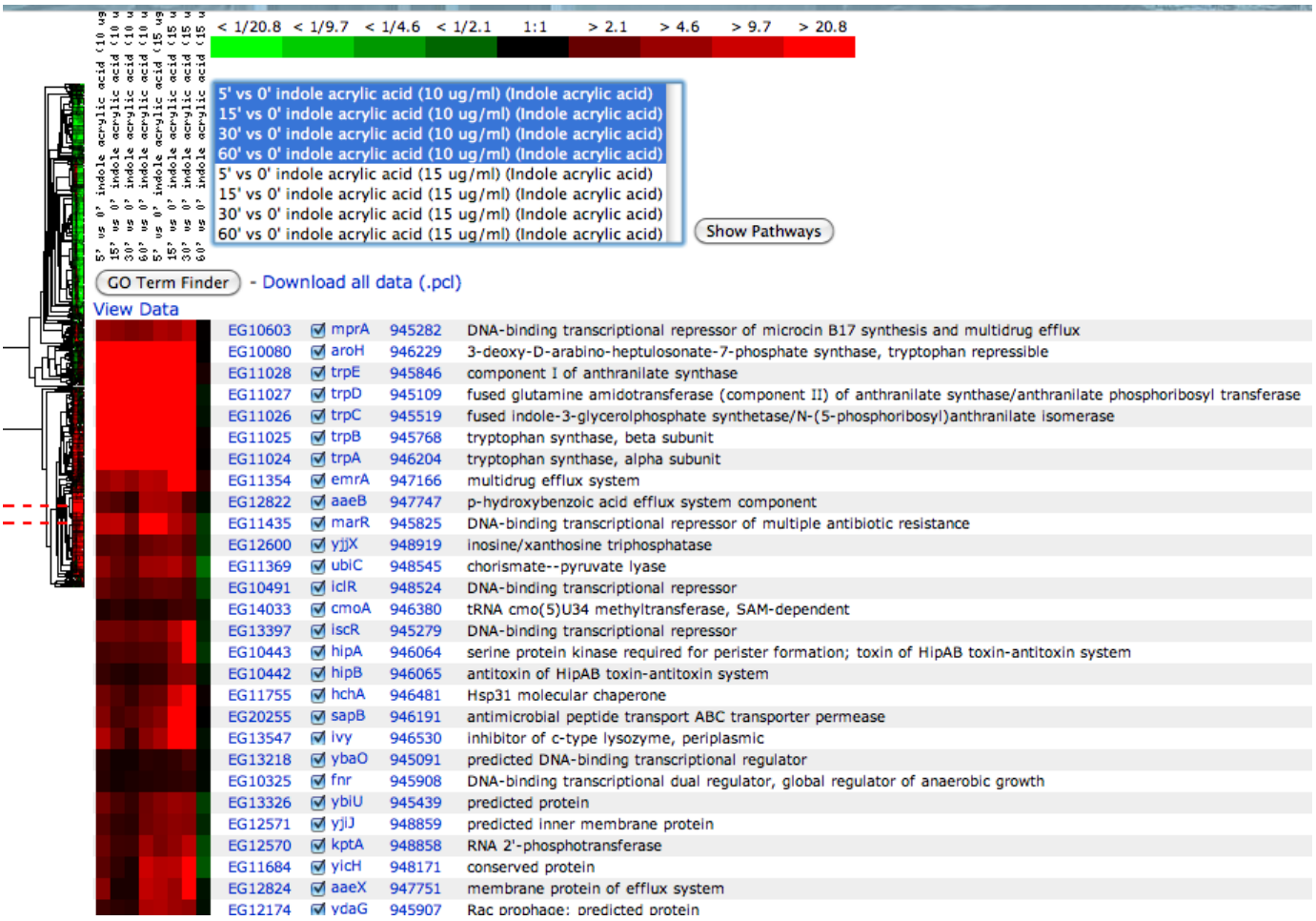
Gene	Median Z Score (Absolute Value)	Number of Samples
trpE	7.067	8
yjiC	6.679	1
trpD	6.288	8
trpC	5.311	8
mtr	5.115	8
puuB	5.024	8
aceK	4.855	6
eamB	4.670	8
gatB	4.623	6
trpB	4.610	8
puuE	4.585	8
oppC	4.411	8
pstS	4.377	8
trpA	4.289	8
fliK	4.288	7
yfjZ	4.280	1
ycjV	4.267	1
cysC	4.251	2
gatC	4.209	5
fliI	4.205	6
frdB	4.136	8
sucB	4.040	7
nuoH	4.032	5
argF	3.943	8

Indole Acrylic Acid is an inhibitor of tryptophan synthetase, and is known to induce expression of the trp operon. It makes sense that the trpE, trpD, trpC, trpB, and trpA all have significant changes in gene expression. To see a “heat map” of the data, click on **Go To Gene Profiles** at the bottom of the page to get this:

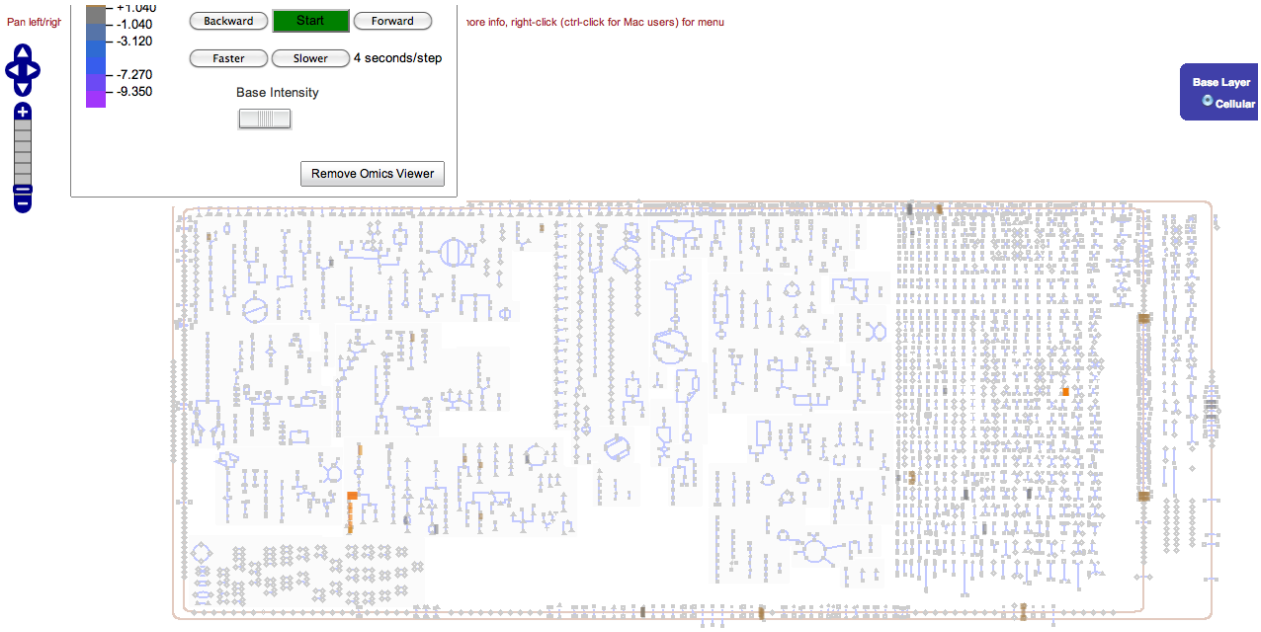
The small, almost unreadable graph shows a thumbnail of the clustering of the *E. coli* genes with significant expressions changes in the 8 selected samples. The red dotted lines show the area that is expanded below. You can move the selection by clicking.

What pathways are upregulated?

Click on the patch of bright red in the heat map to select the genes that are strongly upregulated. The select a subset of the samples.



Click **Show Pathways** to send data for the selected genes to the EcoCyc Omics viewer



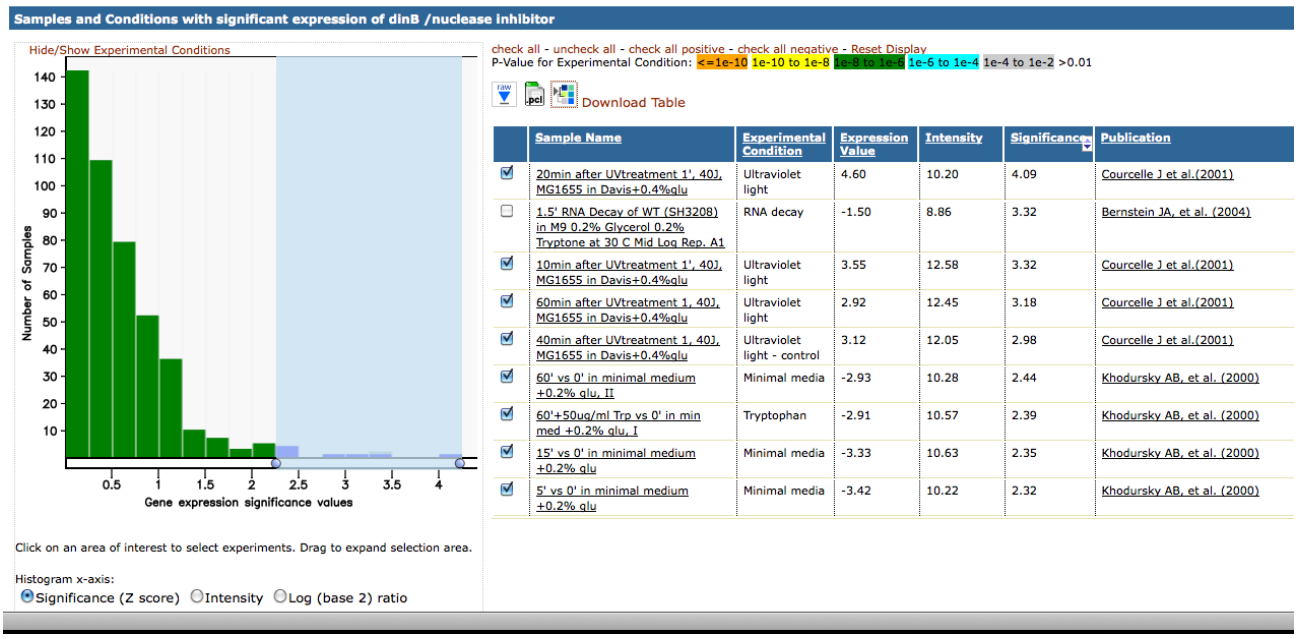
The trp biosynthetic pathway, along with a few other reactions, is highlighted. You can pan and zoom on the map to see more detail.

What genes have expression patterns similar to my favorite gene?

Genes that act in the same biological process often have similar patterns of gene expression. However, the converse is often not true: similar patterns of gene expression do not mean common biological functions. A major reason for this is that in most transcriptome studies, the majority of genes do not change their expression. Thus, many genes of unrelated function will cluster together by virtue of not doing anything.

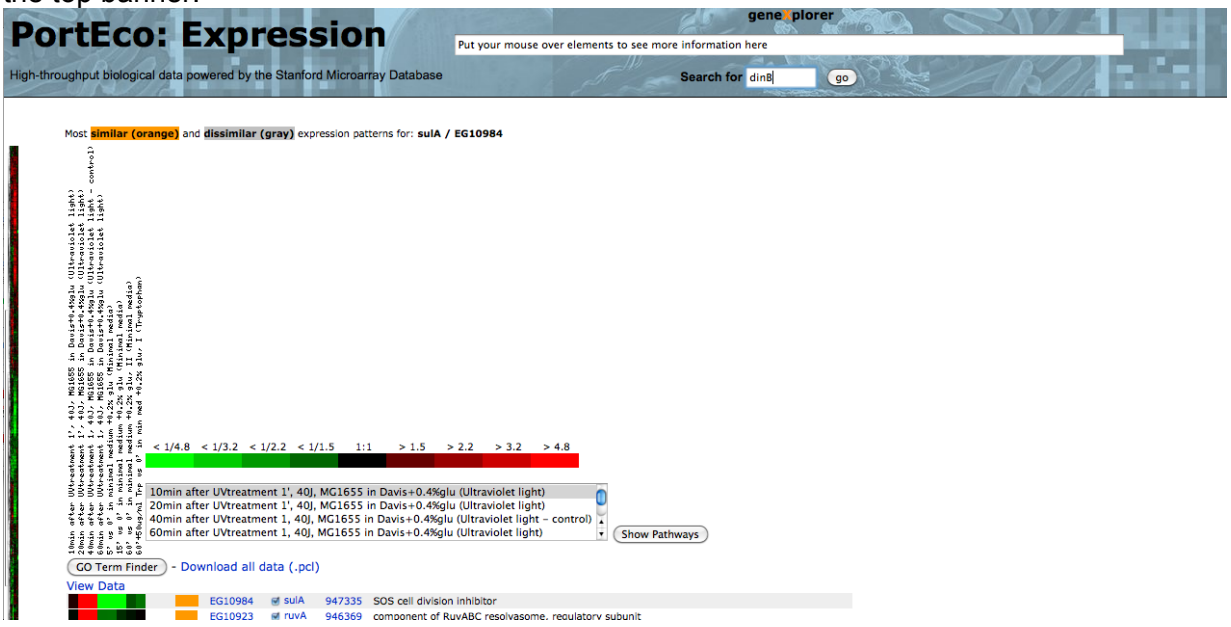
To avoid this, we would like to see what genes have common expression with our favorite genes under conditions where the expression of our favorite gene is doing something interesting.

Type dinB in search box for Samples and Conditions.

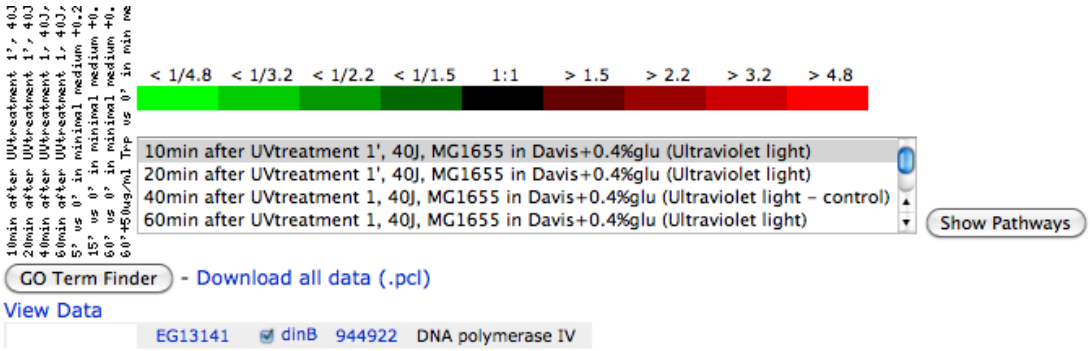


The histogram shows the distribution of significance scores for *dinB* across the studies available at expression.porteco.org. The adjustable light blue area selects studies from a range of values. These are listed on the right. To view a cluster across these studies, click on the cluster icon next to the Download Table link.

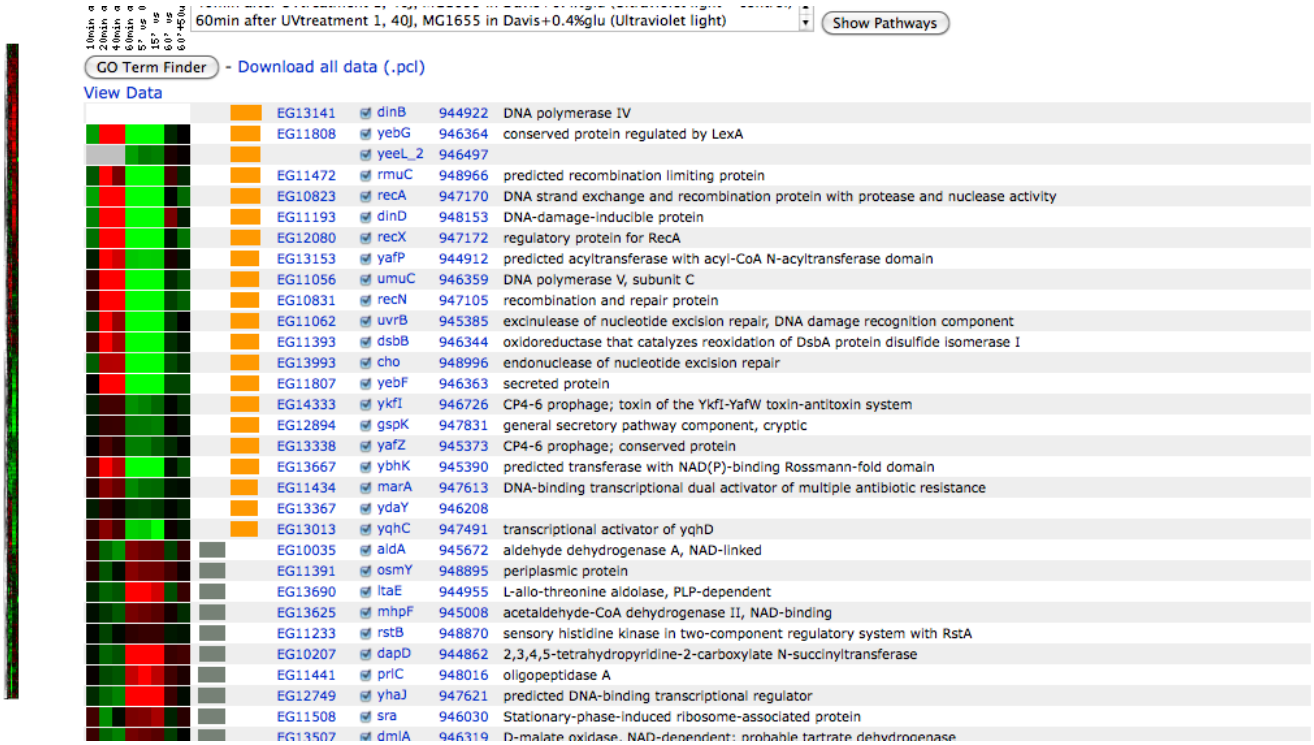
Wait for the clustering heatmap to come up, and enter *dinB* in the search box that will appear in the top banner.



Clicking **GO** or pressing return will give you a view like this, showing only *dinB*:

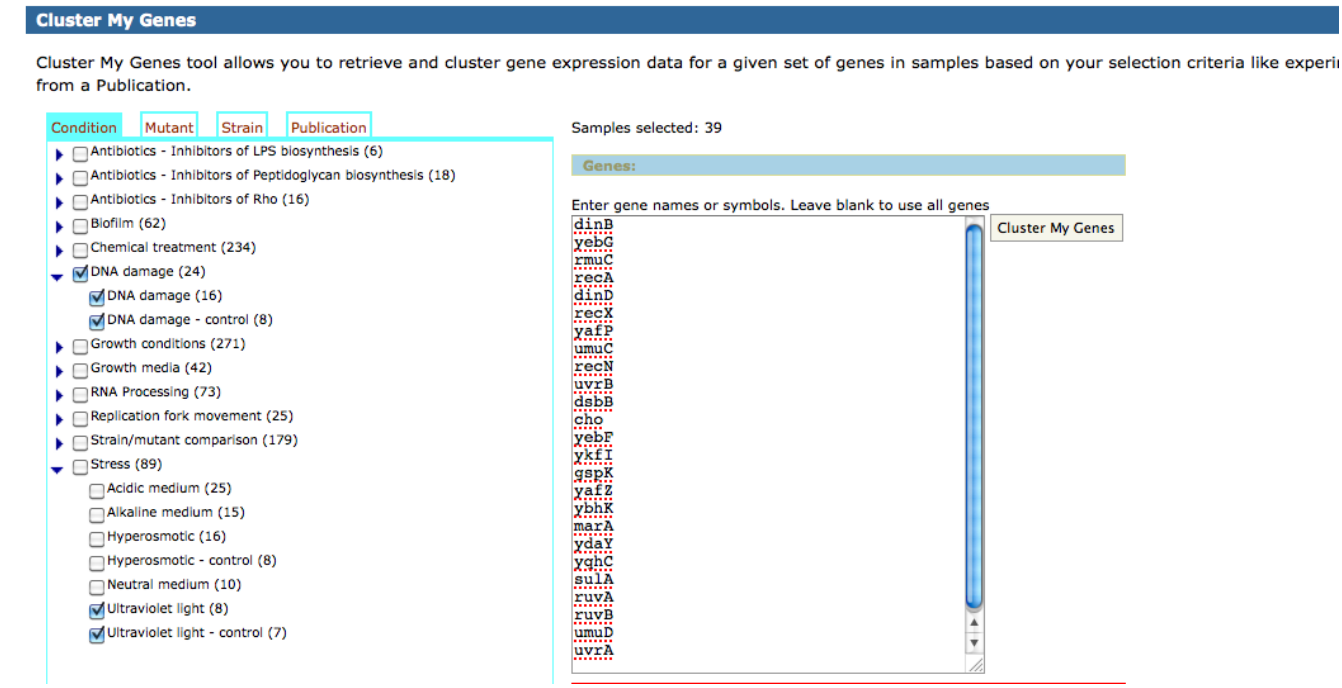


Sometimes there is no heat map to the left of the gene information - this is a bug. Mouse over that area and you will see that there is a link. You may see a tooltip popup that says “View this gene’s profile”. Click there.

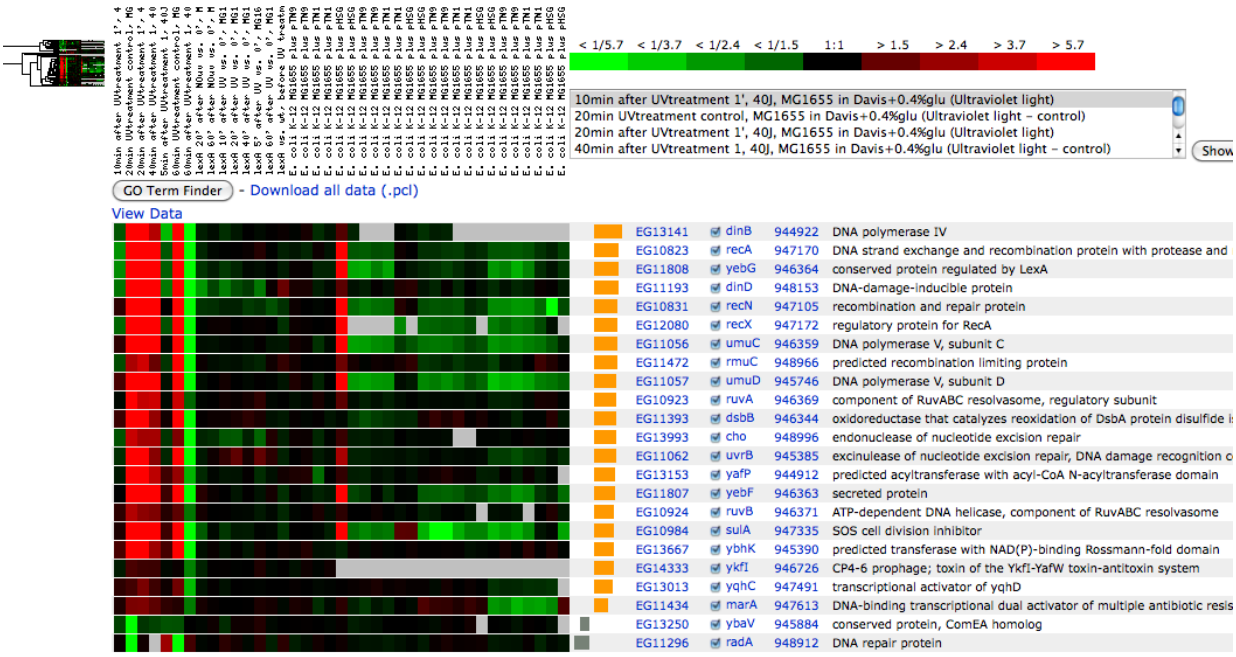


The orange bars indicate genes that cluster with *dinB*, while the grey bars indicate anticorrelation. The many of the genes we see are known members of the SOS response regulon, so this is an expected result. However some are genes of unknown function.

Interestingly, some other genes known to be in the SOS response are not clustering with *dinB* in these studies. Let’s compare some expression patterns over several studies.



Here, I've selected other studies that are annotated as involving DNA damage, and I've focused the clustering on the genes that clustered with dinB in the previous analysis, plus some other genes I think should have been in this cluster. For example, *sulA* is often used as a reporter for the SOS response. UmuD is in a complex with UmuC. The *umuC* gene clustered with *dinB*, so let's see what's going on with *umuD*. Click **Cluster My Genes**, and find *dinB* again.

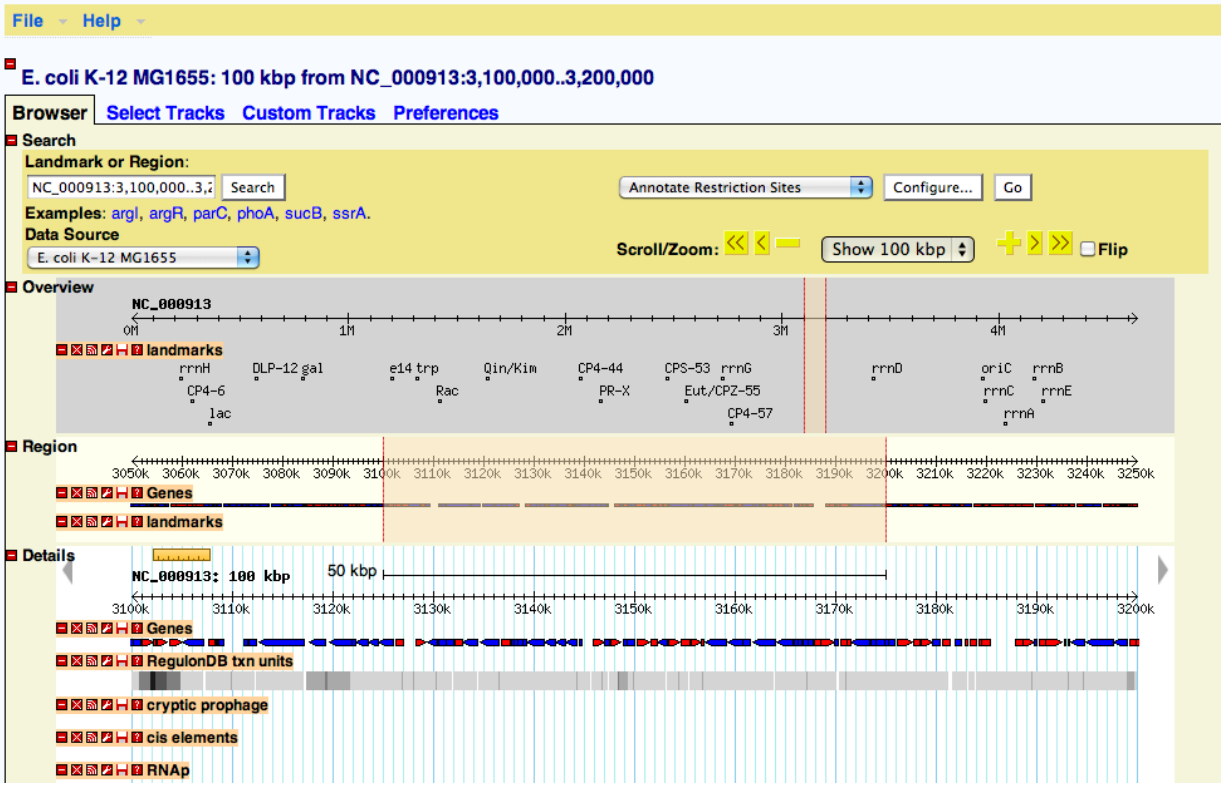


Note a few things. First, Cluster My Genes picked up some genes that were not on our list; this is because it's looking for matches not just in the gene name but also in the rest of the gene information. Second, in the larger comparisons, the other known SOS genes look similar to

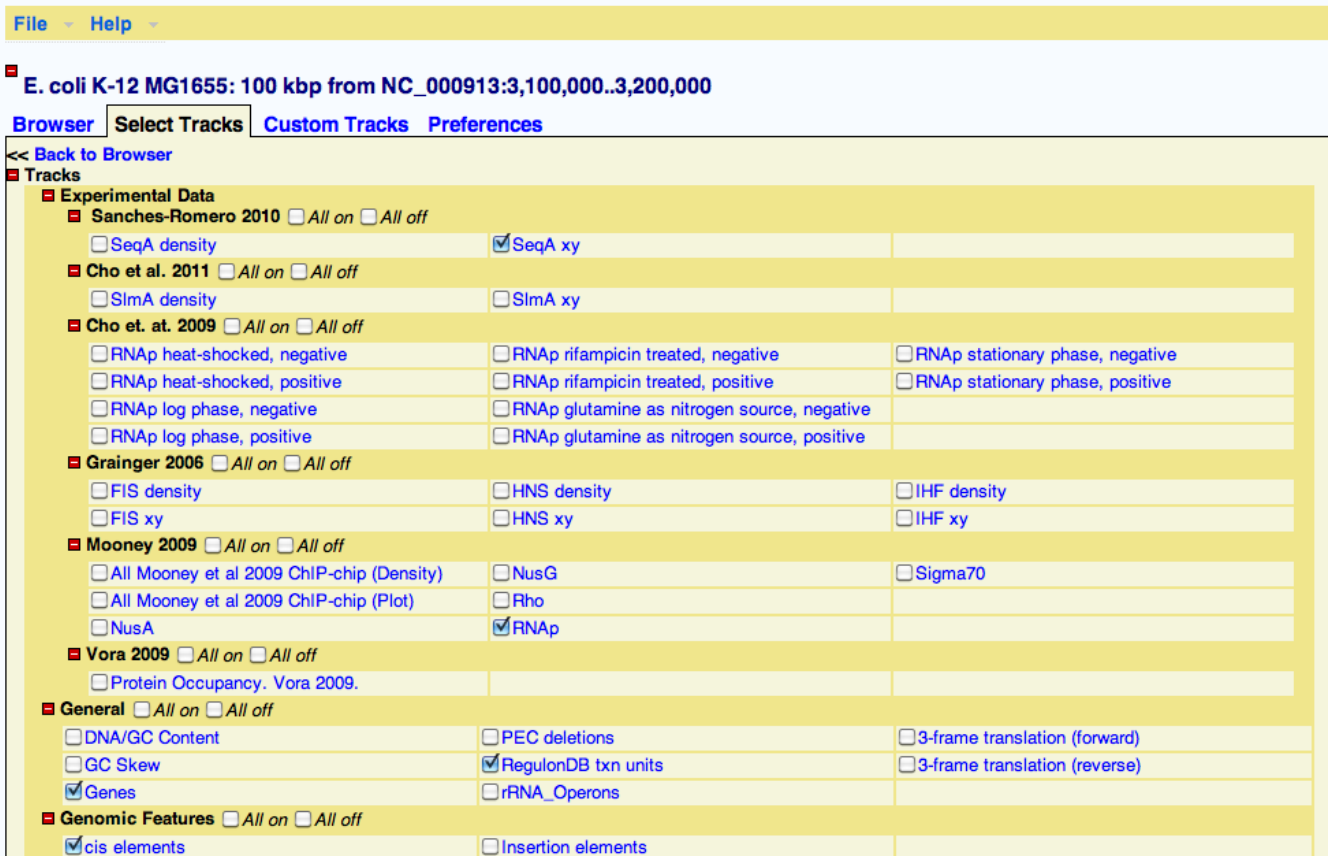
dinB. This second point illustrates how transcriptome analysis, and, more generally, cluster analysis does not have a unique “right answer”.

ChIP data

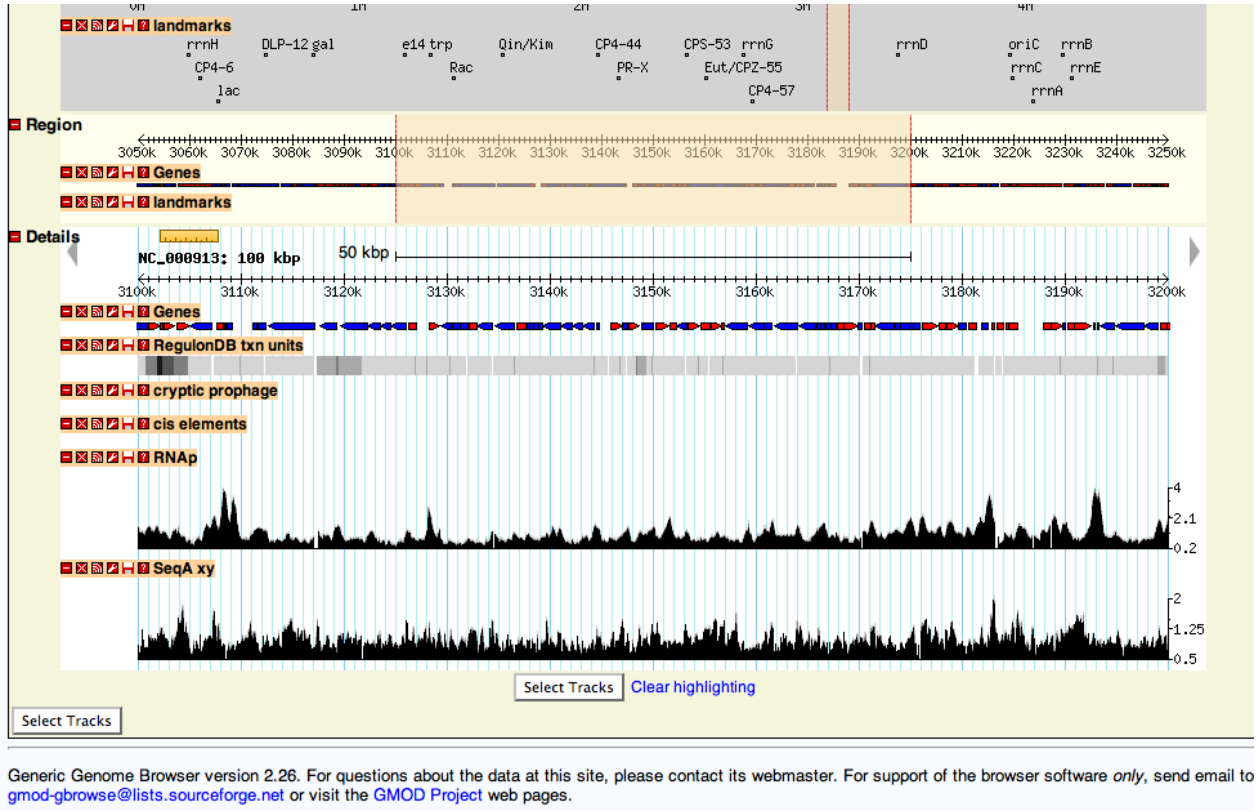
http://ecoliwiki.net/gbrowse will redirect you to the EcoliWiki genome browser.



Select Tracks to see available ChIP and other data types



Select one or more tracks and return to the browser



Gbrowse allows you to upload your own private tracks to compare with our tracks. For example, try this file:

http://ecoliwiki.net/files/phage2011/pec_nonessential.gff

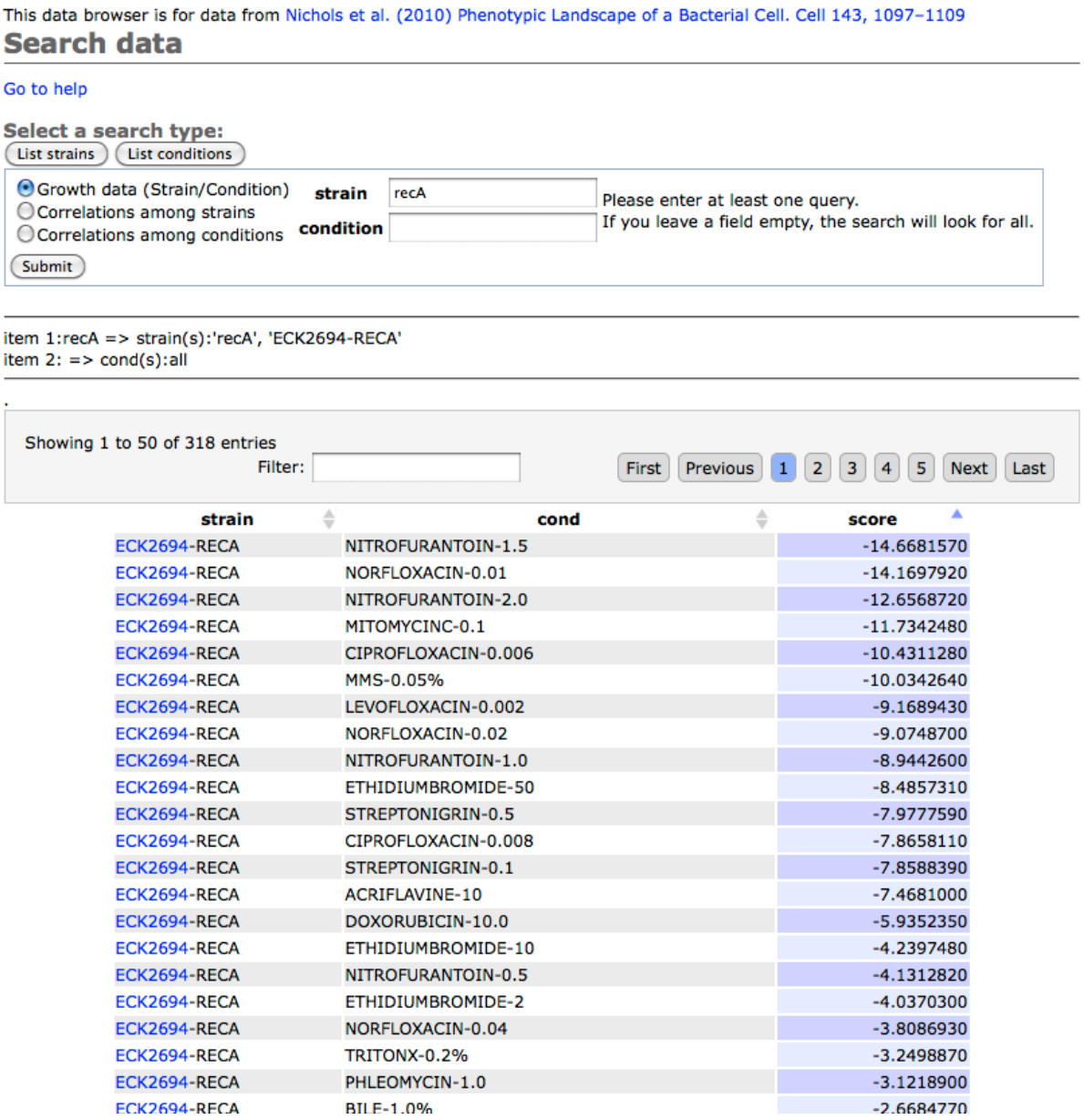
Chemical Genomics Phenotypic Landscape

Carol Gross' keynote will describe some of the work her lab based on high-throughput phenotyping (Nichols et al. (2010) Phenotypic Landscape of a Bacterial Cell. Cell 143, 1097–1109). We have a tool to allow you to browse the data at: http://ecoliwiki.net/tools/chemgen/

There are three kinds of searches you can do.

Strain vs Condition

This allows you to search for the behavior of a knockout mutant on one of the 324 conditions. Enter *recA* and sort the score to show the most negative first.



strain	cond	score
ECK2694-RECA	NITROFURANTOIN-1.5	-14.6681570
ECK2694-RECA	NORFLOXACIN-0.01	-14.1697920
ECK2694-RECA	NITROFURANTOIN-2.0	-12.6568720
ECK2694-RECA	MITOMYCINC-0.1	-11.7342480
ECK2694-RECA	CIPROFLOXACIN-0.006	-10.4311280
ECK2694-RECA	MMS-0.05%	-10.0342640
ECK2694-RECA	LEVOFLOXACIN-0.002	-9.1689430
ECK2694-RECA	NORFLOXACIN-0.02	-9.0748700
ECK2694-RECA	NITROFURANTOIN-1.0	-8.9442600
ECK2694-RECA	ETHIDIUMBROMIDE-50	-8.4857310
ECK2694-RECA	STREPTONIGRIN-0.5	-7.9777590
ECK2694-RECA	CIPROFLOXACIN-0.008	-7.8658110
ECK2694-RECA	STREPTONIGRIN-0.1	-7.8588390
ECK2694-RECA	ACRIFLAVINE-10	-7.4681000
ECK2694-RECA	DOXORUBICIN-10.0	-5.9352350
ECK2694-RECA	ETHIDIUMBROMIDE-10	-4.2397480
ECK2694-RECA	NITROFURANTOIN-0.5	-4.1312820
ECK2694-RECA	ETHIDIUMBROMIDE-2	-4.0370300
ECK2694-RECA	NORFLOXACIN-0.04	-3.8086930
ECK2694-RECA	TRITONX-0.2%	-3.2498870
ECK2694-RECA	PHLEOMYCIN-1.0	-3.1218900
ECK2694-RECA	BILE-1.0%	-2.6684770

Fitness is based on colony size doubly normalized for the sizes of all colonies on the plate and the size of the specific strain under other conditions. Scores are not directly correlated to doubling times or growth rates: they're a statistical measure of how far this sample is from the average behavior of all strains on this condition and this strain on all conditions. Positive scores mean better than average fitness, while negative scores mean greater than average sensitivity.

In the example, we can see that *recA* is more sensitive to nitrofurantoin, norfloxacin, cipro, mitomycin etc. These all make physiological sense, as they lead to DNA damage.

Strain vs Strain

We can look at what other genes have similar patterns of increased or decreased fitness by changing to a strain vs strain comparison. Leave *recA* in strain 1 and leave the other blank.

Search data

Go to help

Select a search type:

List strains | List conditions

Growth data (Strain/Condition)
Correlations among strains
Correlations among conditions

strain 1 recA

strain 2

Please enter at least one query.
If you leave a field empty, the search will look for all.

Submit

item 1:recA => strain(s):'recA', 'ECK2694-RECA'
item 2: => strain2(s):all

.

Showing 1 to 50 of 3979 entries
Filter:

First | Previous | 1 | 2 | 3 | 4 | 5 | Next | Last

strain	strain2	correlation_coefficient
ECK2694-RECA	ECK2694-RECA	1.000000000
ECK2694-RECA	ECK1862-RUVA	0.588771000
ECK2694-RECA	ECK2818-RECC	0.544787000
ECK2694-RECA	ECK1864-RUVC	0.541649000
ECK2694-RECA	ECK3642-RECG	0.538428000
ECK2694-RECA	ECK2816-RECB	0.489492000
ECK2694-RECA	ECK4050-UVRA	0.483626000
ECK2694-RECA	ECK2612-RECN	0.438514000
ECK2694-RECA	ECK0768-UVRB	0.412939000
ECK2694-RECA	ECK0621-LIPA	0.407332000
ECK2694-RECA	ECK3808-UVRD	0.401450000
ECK2694-RECA	ECK3806-XERC	0.393383000
ECK2694-RECA	ECK1310-YCJS	0.360886000
ECK2694-RECA	ECK1912-UVRC	0.323210000
ECK2694-RECA	ECK2270-NUON	0.323139000
ECK2694-RECA	ECK4017-PGI	0.322410000
ECK2694-RECA	ECK0654-UBIF	0.319181000
ECK2694-RECA	ECK2889-XERD	0.311637000
ECK2694-RECA	ECK0388-RDGC	0.301641000
ECK2694-RECA	ECK3085-EXUR	0.296974000
ECK2694-RECA	ECK2281-NUOB	0.296758000
ECK2694-RECA	ECK2282-NUOA	0.294667000
ECK2694-RECA	ECK0398-ACPH	0.292268000

The score here is a correlation coefficient for all the phenotypes of each pair of strains. Based on Fig 3 of Nichols et al, the P-values for these correlation coefficients are

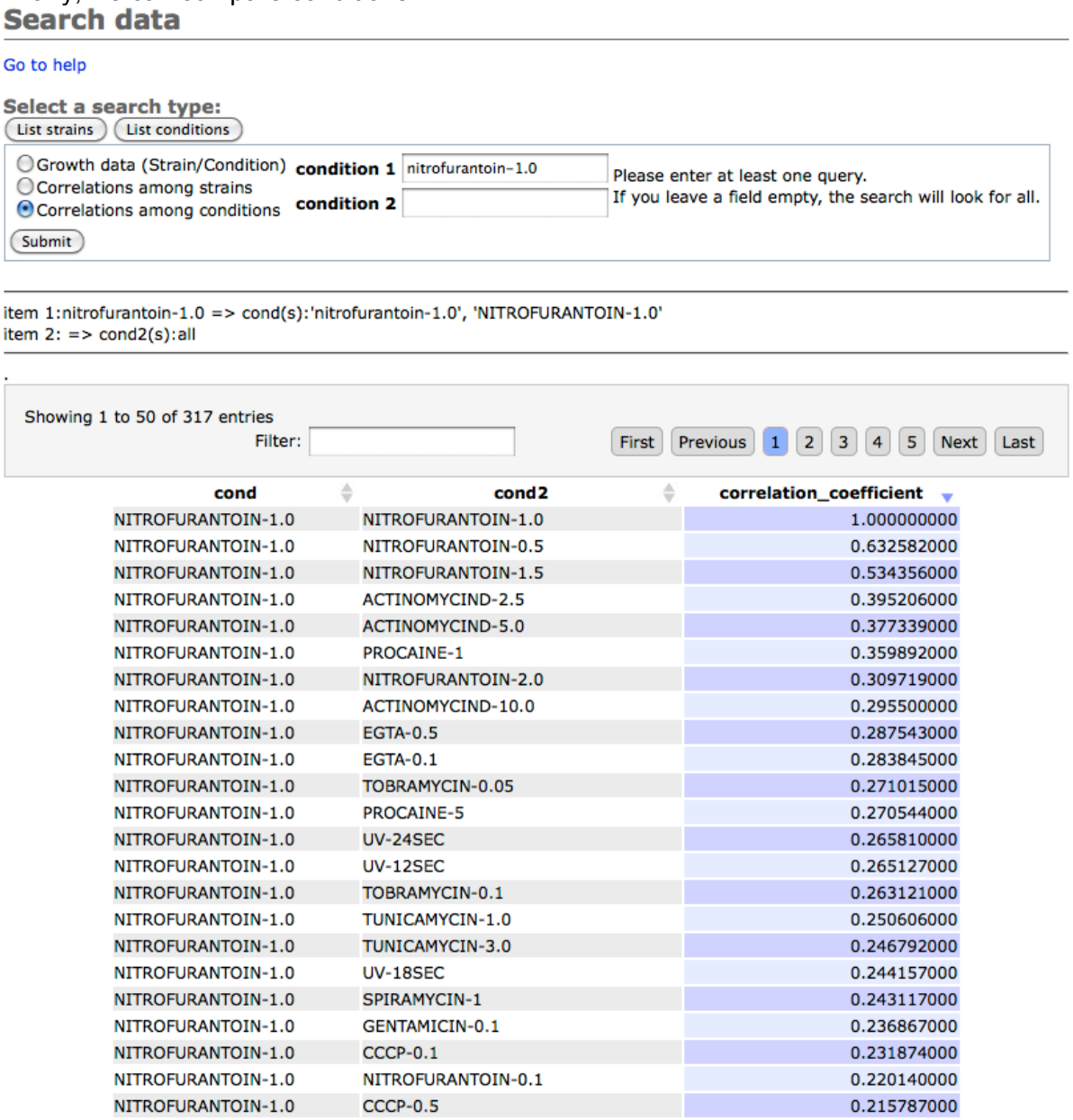
correlation coefficient	P-value
0.4	10^{-14}
0.5	10^{-22}
0.6	10^{-34}
0.7	10^{-50}

0.8	10^{-75}
0.9	10^{-120}

The fitness effects of mutations in a particular condition are often due to complex indirect effects. This means looking at the strain-condition scores is often not informative in terms of the biological function of a gene. However, strain-strain correlation will pick up cases where both the direct and indirect effects are similar, making the strain-strain comparisons more informative.

Condition vs Condition

Similarly, we can compare conditions



cond	cond2	correlation_coefficient
NITROFURANTOIN-1.0	NITROFURANTOIN-1.0	1.000000000
NITROFURANTOIN-1.0	NITROFURANTOIN-0.5	0.632582000
NITROFURANTOIN-1.0	NITROFURANTOIN-1.5	0.534356000
NITROFURANTOIN-1.0	ACTINOMYCIND-2.5	0.395206000
NITROFURANTOIN-1.0	ACTINOMYCIND-5.0	0.377339000
NITROFURANTOIN-1.0	PROCAINE-1	0.359892000
NITROFURANTOIN-1.0	NITROFURANTOIN-2.0	0.309719000
NITROFURANTOIN-1.0	ACTINOMYCIND-10.0	0.295500000
NITROFURANTOIN-1.0	EGTA-0.5	0.287543000
NITROFURANTOIN-1.0	EGTA-0.1	0.283845000
NITROFURANTOIN-1.0	TOBRAMYCIN-0.05	0.271015000
NITROFURANTOIN-1.0	PROCAINE-5	0.270544000
NITROFURANTOIN-1.0	UV-24SEC	0.265810000
NITROFURANTOIN-1.0	UV-12SEC	0.265127000
NITROFURANTOIN-1.0	TOBRAMYCIN-0.1	0.263121000
NITROFURANTOIN-1.0	TUNICAMYCIN-1.0	0.250606000
NITROFURANTOIN-1.0	TUNICAMYCIN-3.0	0.246792000
NITROFURANTOIN-1.0	UV-18SEC	0.244157000
NITROFURANTOIN-1.0	SPIRAMYCIN-1	0.243117000
NITROFURANTOIN-1.0	GENTAMICIN-0.1	0.236867000
NITROFURANTOIN-1.0	CCCP-0.1	0.231874000
NITROFURANTOIN-1.0	NITROFURANTOIN-0.1	0.220140000
NITROFURANTOIN-1.0	CCCP-0.5	0.215787000

Annotation with Students (CACAO)

Our ability to mine data about *E. coli,* or any other organism, is limited by how well we capture what is in the literature. To speed up curation of the literature, we have created a way for people to get teaching credit for having students do curation using the Gene Ontology (GO).

Brenley McIntosh will present a poster with more detail about this activity, which we call Community Assessment of Community Annotation with Ontologies. Today we're just going to show you:

- What GO is
- How to browse in GONUTS
- How to create an editable page

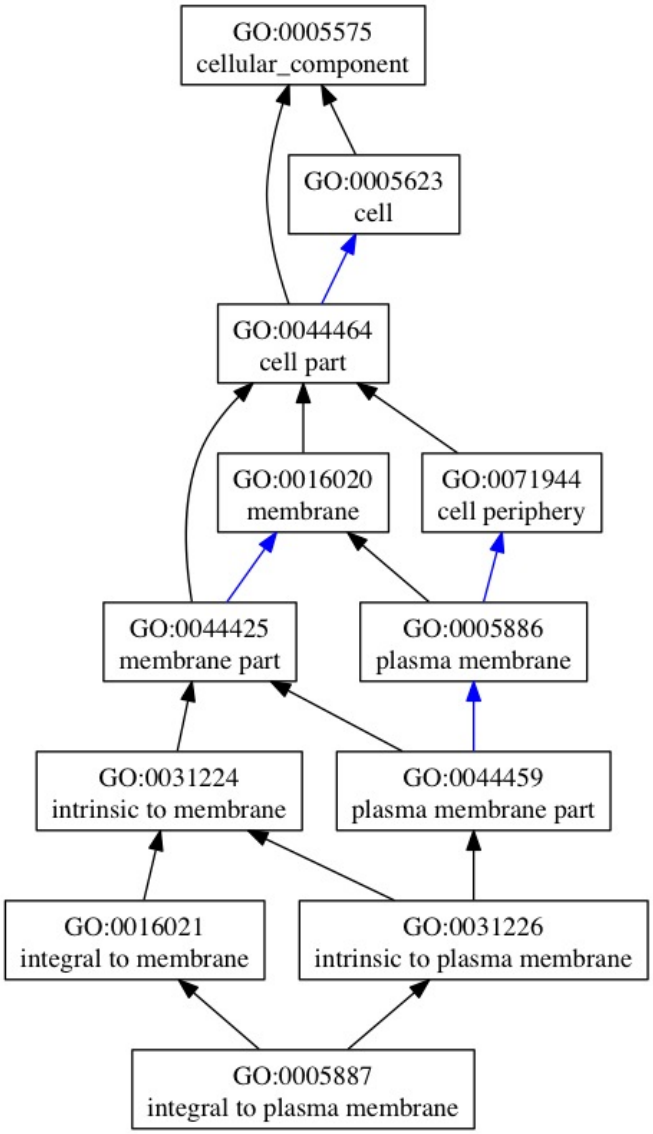
What is GO?

Ontologies can be thought of as vocabularies for types or classifications combined with defined relationships between those types.

Example:

phosphofructokinase activity **is_a** carbohydrate kinase activity

Relationships between types in ontologies can be complex directed acyclic graphs (DAG). Ontologies can be thought of as concept maps for different areas of knowledge. Below is a DAG for the GO term integral to plasma membrane.



Although it is counterintuitive, the DAG is usually drawn with the **root node** at the top. The terms at the ends of the branches are often referred to as **leaf nodes**. As you move from the root toward the leaves, the terms become more specific. The more specific terms below a term are also called **child terms**, while the less specific relatives are **parent terms**.

GO consists of three DAGs for three different aspects of gene function

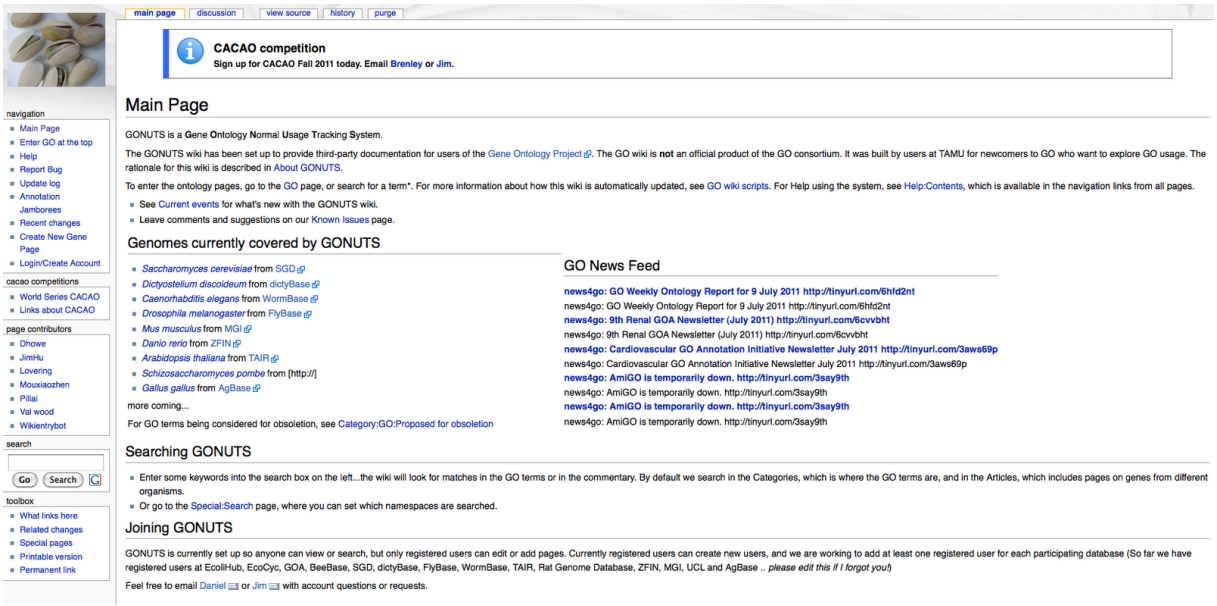
- Molecular function - what a gene product does on its own
- Biological process - pathways or processes the product participates in
- Cellular component - where it acts; compartments and complexes

Browsing GO in GONUTS

There are many tools to browse GO. The Gene Ontology Normal Usage Tracking System (GONUTS) at http://gowiki.tamu.edu has a couple of important features that make it different from other GO browsers.

- Users can add notes about terms
- Users can add annotations for anything in UniProt

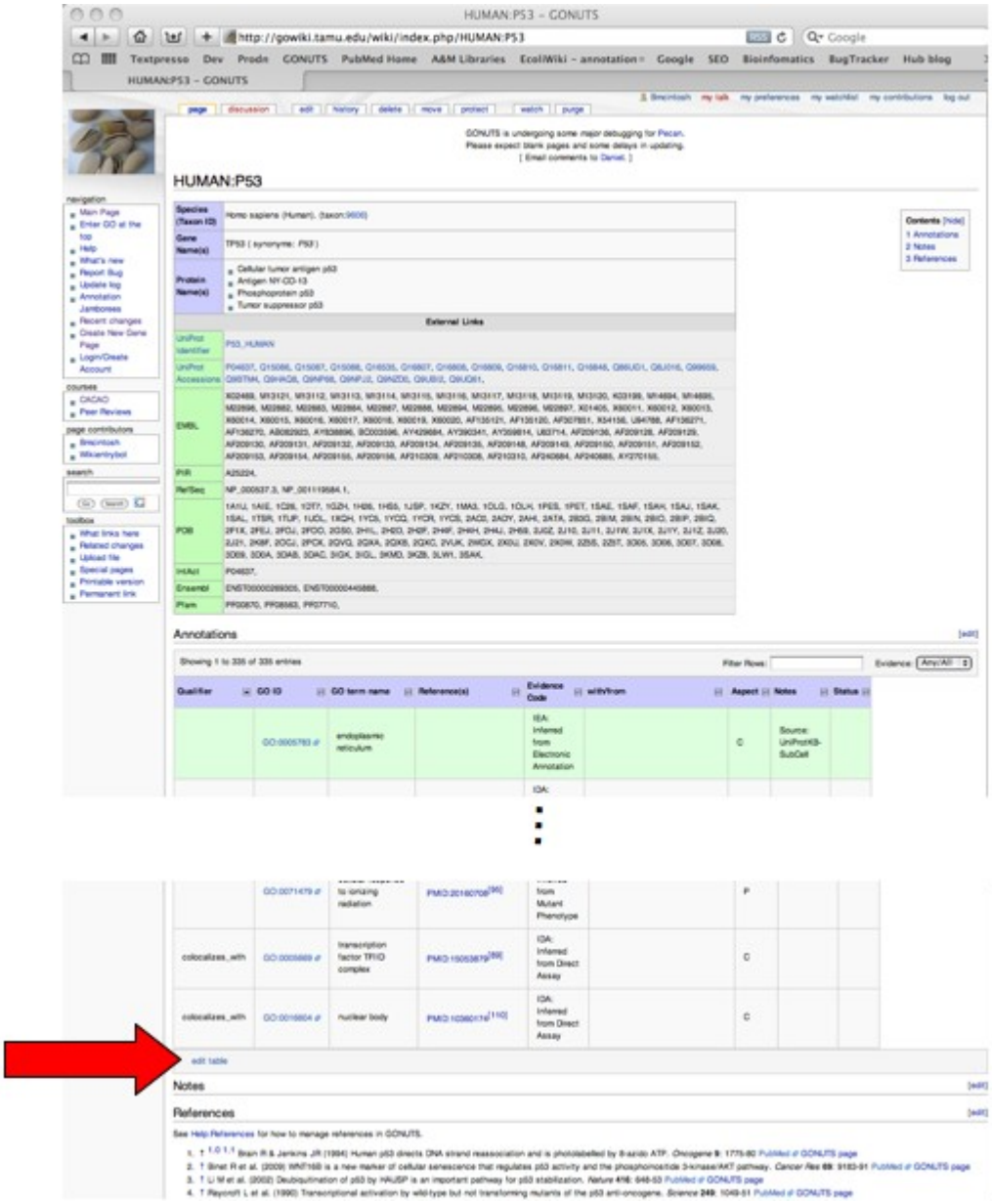
Go to GONUTS and log in. You should have received an email to set up your account. If not, we can create one for you.



In the search box, enter some words that sound like a function. If you match a GO term name, you'll go straight to a term page; otherwise you will get a list of possible matches. If you go to one of those, you can often find what you want by navigating up or down the DAG.

Creating gene pages

GoPageMaker – GONUTS
http://gowiki.tamu.edu/wiki/index.php/Special:GoPageMaker
Reader
Google
Textpresso Dev Prodn GONUTS PubMed Home A&M Libraries EcoliWiki – annotation Google SEO Bioinfomatics BugTracker
GoPageMaker – GONUTS
GO:0016021 ! integral to membra...
Bmcintosh my talk my preferences my watchlist my contributions log out
special page
page maker
translator
GONUTS is undergoing some major debugging for Pecan.
Please expect blank pages and some delays in updating.
[Email comments to Daniel.]
GoPageMaker
To create a new gene page, please select a database and enter a unique identifier such as an ID or an accession number.
It may take a few minutes to gather data from the primary sources, please be patient.
UniProt Id/Acc
Create
[edit]
The GONUTS gene page maker creates a gene page where you can add GO annotations for any gene that has a UniProt Identifier (ID/Entry or Accession) or a NCBI Identifier (GI Number, RefSeq Accession, GenPept Accession). The information generated by the GONUTS gene page maker is UniProt centric. If NCBI identifiers are used to create a gene page, they are mapped to the corresponding UniProt Accession using in-house tools adapted from the documentation listed here.
To create a gene page, use the pull-down menu and enter the appropriate protein identifier from the database of your choice. If the gene is from an organism which we recognize to be an existing model organism database (MOD), a link will be provided which would allow the user to access the community annotation or other input system implemented by the respective MOD. Once the GONUTS gene page maker creates a gene page, it would have a table containing GO Annotations (if any) as it exists for the UniProt Record. Annotators can also edit the table and add GO annotations for the gene.
navigation
Main Page
Enter GO at the top
Help
What's new
Report Bug
Update log
Annotation Jamborees
Recent changes
Create New Gene Page
Login/Create Account
courses
CACAO
Peer Reviews
search
Go
Search
toolbox


HUMAN:P53 - GONUTS
http://gowiki.tamu.edu/wiki/index.php/HUMAN:P53
Textpresso Dev Prodn GONUTS PubMed Home A&M Libraries EcoliWiki - annotation Google SEO Bioinfomatics BugTracker Hub blog
HUMAN:P53 - GONUTS
page discussion edit history delete move protect watch purge
GONUTS is undergoing some major debugging for Pecan.
Please expect blank pages and some delays in updating.
[Email comments to Daniel.]
HUMAN:P53
Species (Taxon ID)
Homo sapiens (Human), (taxon:9606)
Gene Name(s)
TP53 (synonyms: P53)
Protein Name(s)
Cellular tumor antigen p53
Antigen NY-CO-13
Phosphoprotein p53
Tumor suppressor p53
External Links
UniProt Identifier
P53_HUMAN
UniProt Accessions
EMBL
PIR
RefSeq
PDB
IHAct
Ensembl
Pfam
Contents [hide]
1 Annotations
2 Notes
3 References
Annotations
Showing 1 to 335 of 335 entries
Filter Rows:
Evidence:
Qualifier
GO ID
GO term name
Reference(s)
Evidence Code
with/from
Aspect
Notes
Status
endoplasmic reticulum
IEA: Inferred from Electronic Annotation
C
Source: UniProtKB-SubCell
IDA:
to ionizing radiation
from Mutant Phenotype
P
colocalizes_with
transcription factor TFIID complex
IDA: Inferred from Direct Assay
C
colocalizes_with
nuclear body
IDA: Inferred from Direct Assay
C
edit table
Notes
References
See Help:References for how to manage references in GONUTS.