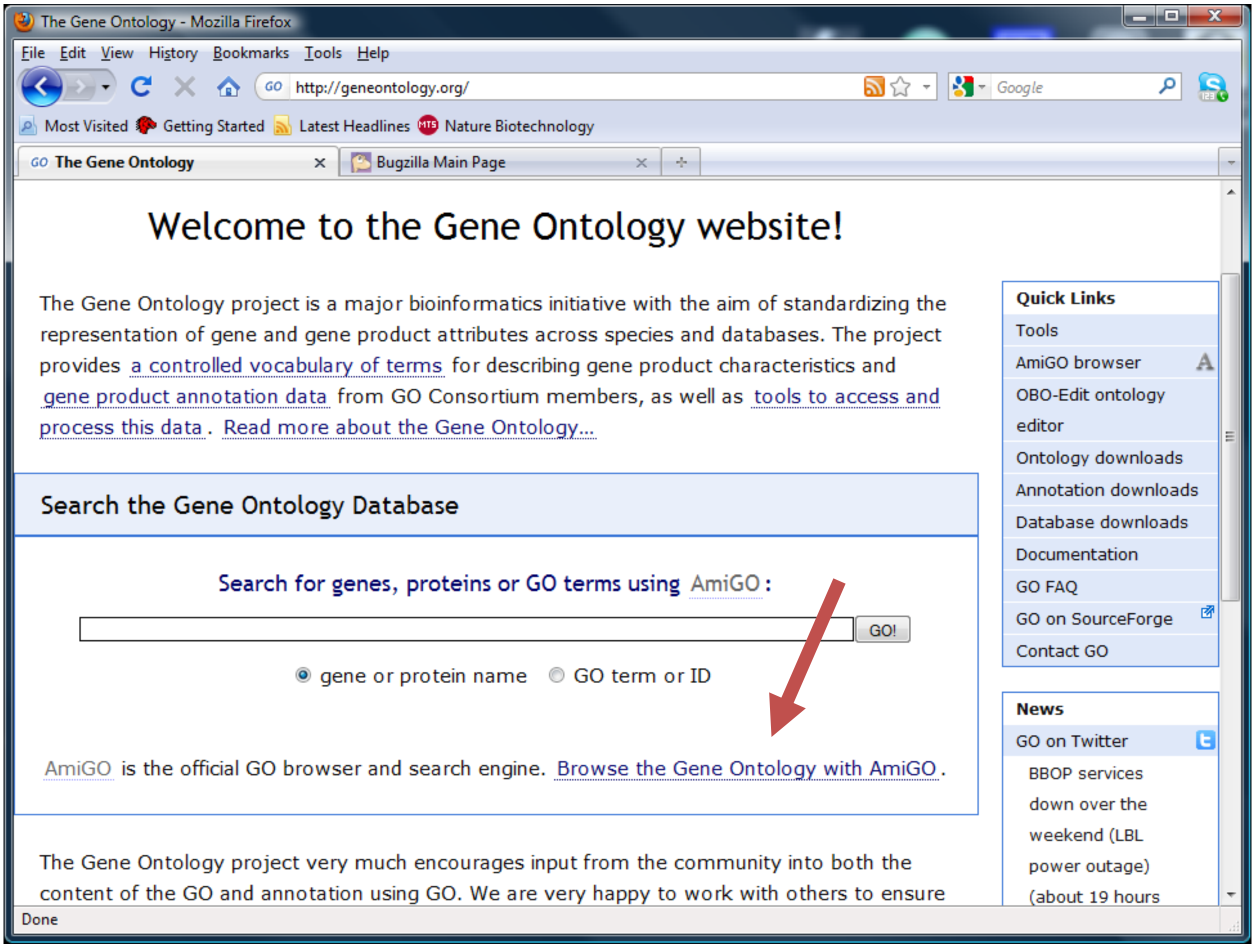
GO, DAVID and iRefScape revisited.

Ian Donaldson

MBV-INF 4410/9410

This exercise will revisit a number of applications introduced in the class. You will learn how to browse the Gene Ontology, select a set of genes that are annotated with some specific GO term and then you will analyze this gene list using DAVID and iRefScape.

Go to http://geneontology.org/



Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. Read more about the Gene Ontology...

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

● gene or protein name ○ GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

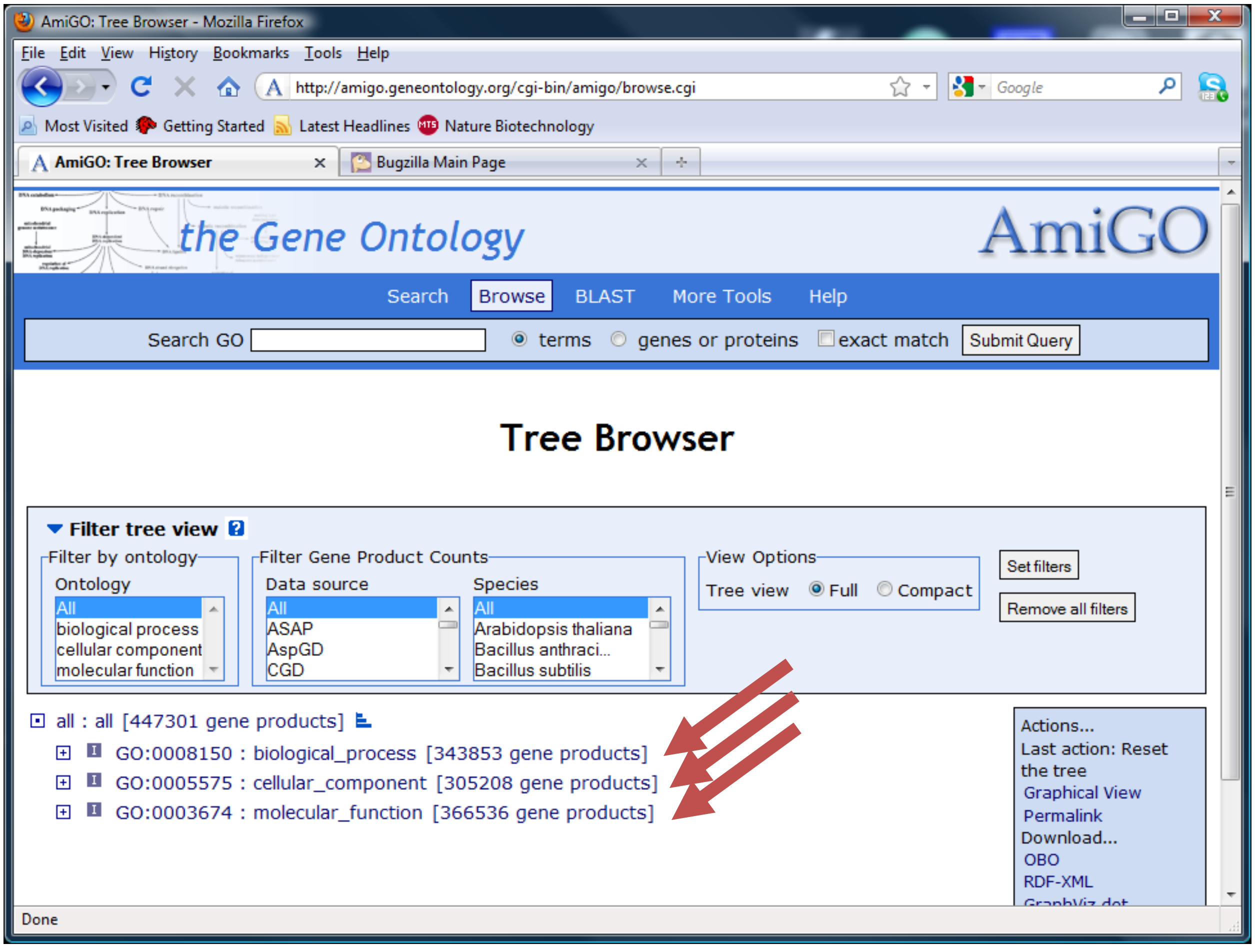
The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure

Quick Links
- Tools
- AmiGO browser
- OBO-Edit ontology editor
- Ontology downloads
- Annotation downloads
- Database downloads
- Documentation
- GO FAQ
- GO on SourceForge
- Contact GO

News
- GO on Twitter
- BBOP services down over the weekend (LBL power outage) (about 19 hours

Click on “Browse the Gene Ontology with AmiGO”.

Spend some time browsing through each of the three GO Ontologies. Look for terms that you are familiar with and see how they relate to terms above (parent terms) and below (child terms) in the GO.



Search GO ● terms ○ genes or proteins □ exact match

Tree Browser

Filter tree view

Filter by ontology
Ontology: All, biological process, cellular component, molecular function

Filter Gene Product Counts
Data source: All, ASAP, AspGD, CGD
Species: All, Arabidopsis thaliana, Bacillus anthraci..., Bacillus subtilis

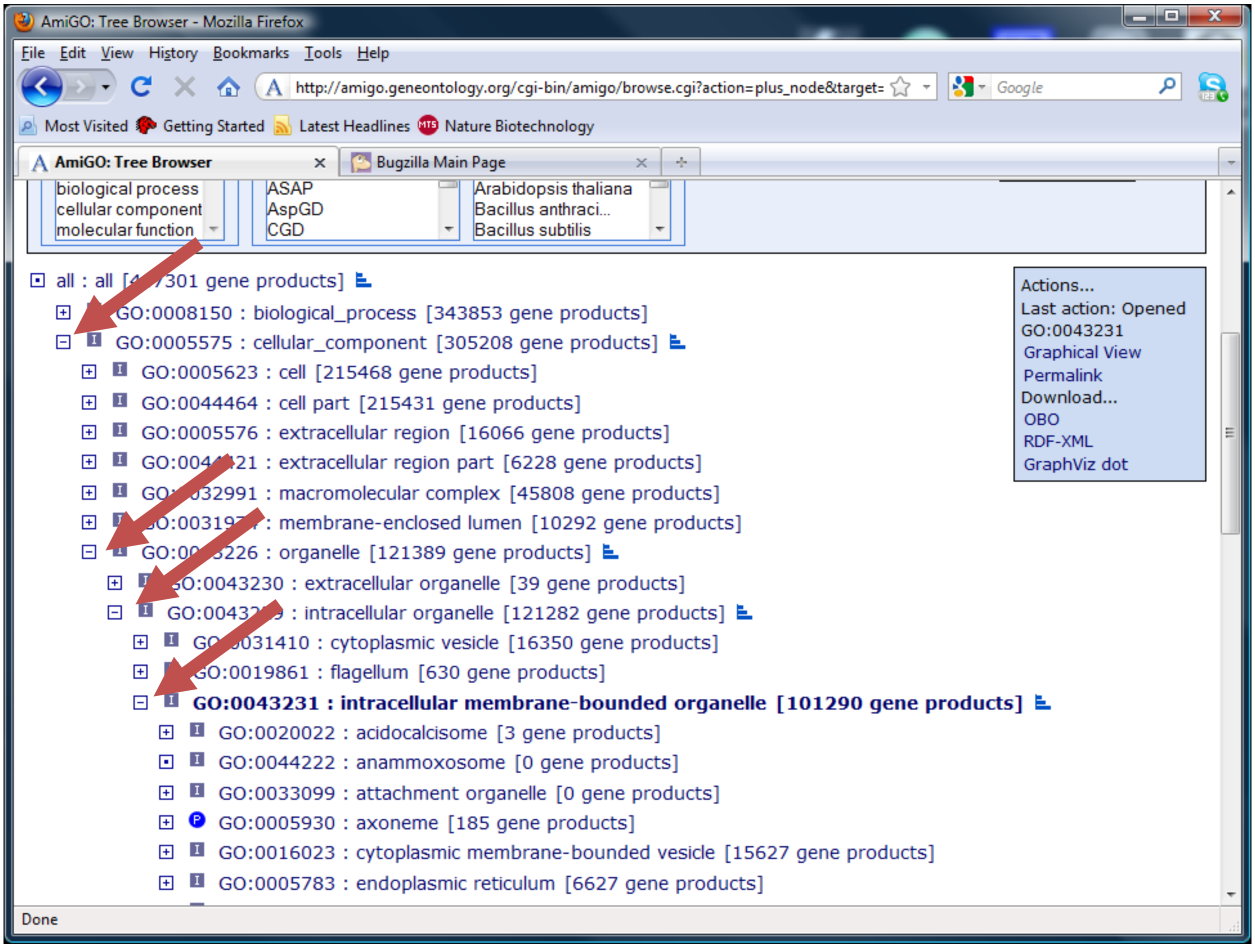
View Options
Tree view ● Full ○ Compact

all : all [447301 gene products]
GO:0008150 : biological_process [343853 gene products]
GO:0005575 : cellular_component [305208 gene products]
GO:0003674 : molecular_function [366536 gene products]

Actions...
Last action: Reset the tree
Graphical View
Permalink
Download...
OBO
RDF-XML
GraphViz dot

Try to make your way to the term that describes “nucleus” without directly searching for it. Instead, start by expanding the cellular_component “root node” (GO:0005755) by clicking on the + beside it. Look for the next closest thing to “nucleus” in the expanded list and then click on that. If you want to look at the definition of any given term, just click on it.

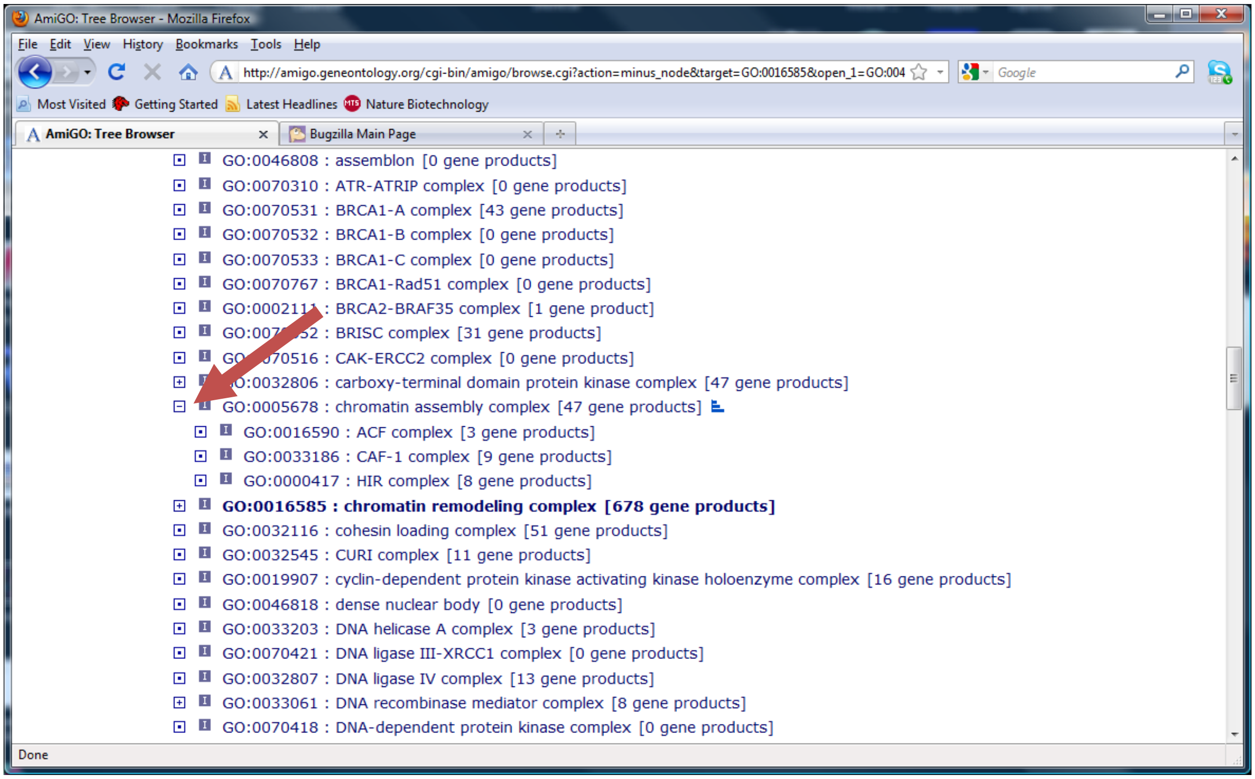
If you cant find the term entry for “nucleus”, use the next page as a hint or search for nucleus in the “Search GO” box at the top of the interface.



```
all : all [4[illegible]7301 gene products]
  GO:0008150 : biological_process [343853 gene products]
  GO:0005575 : cellular_component [305208 gene products]
    GO:0005623 : cell [215468 gene products]
    GO:0044464 : cell part [215431 gene products]
    GO:0005576 : extracellular region [16066 gene products]
    GO:0044421 : extracellular region part [6228 gene products]
    GO:0032991 : macromolecular complex [45808 gene products]
    GO:0031974 : membrane-enclosed lumen [10292 gene products]
    GO:0043226 : organelle [121389 gene products]
      GO:0043230 : extracellular organelle [39 gene products]
      GO:0043229 : intracellular organelle [121282 gene products]
        GO:0031410 : cytoplasmic vesicle [16350 gene products]
        GO:0019861 : flagellum [630 gene products]
        GO:0043231 : intracellular membrane-bounded organelle [101290 gene products]
          GO:0020022 : acidocalcisome [3 gene products]
          GO:0044222 : anammoxosome [0 gene products]
          GO:0033099 : attachment organelle [0 gene products]
          GO:0005930 : axoneme [185 gene products]
          GO:0016023 : cytoplasmic membrane-bounded vesicle [15627 gene products]
          GO:0005783 : endoplasmic reticulum [6627 gene products]
```

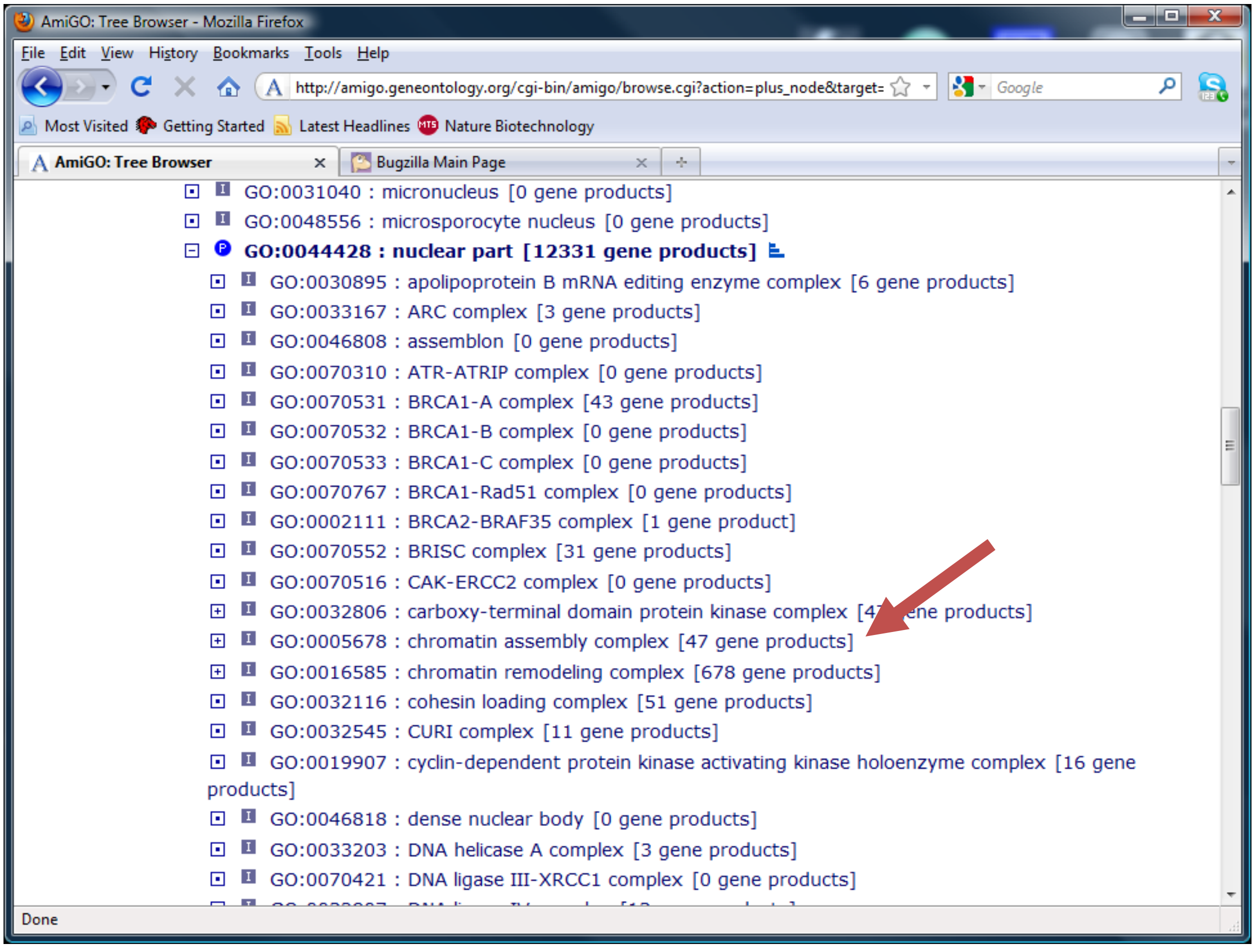
Actions...
Last action: Opened GO:0043231
Graphical View
Permalink
Download...
OBO
RDF-XML
GraphViz dot

Once you make it to the nucleus, keep navigating down the tree to chromatin assembly complex (GO:0005678). This term is actually used to refer to a number of complexes (try expanding the term). What are these complexes? Note that all of them are “leaf” nodes (you cant expand them any further).


AmiGO: Tree Browser - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://amigo.geneontology.org/cgi-bin/amigo/browse.cgi?action=minus_node&target=GO:0016585&open_1=GO:004
Google
Most Visited Getting Started Latest Headlines Nature Biotechnology
AmiGO: Tree Browser
Bugzilla Main Page
Done

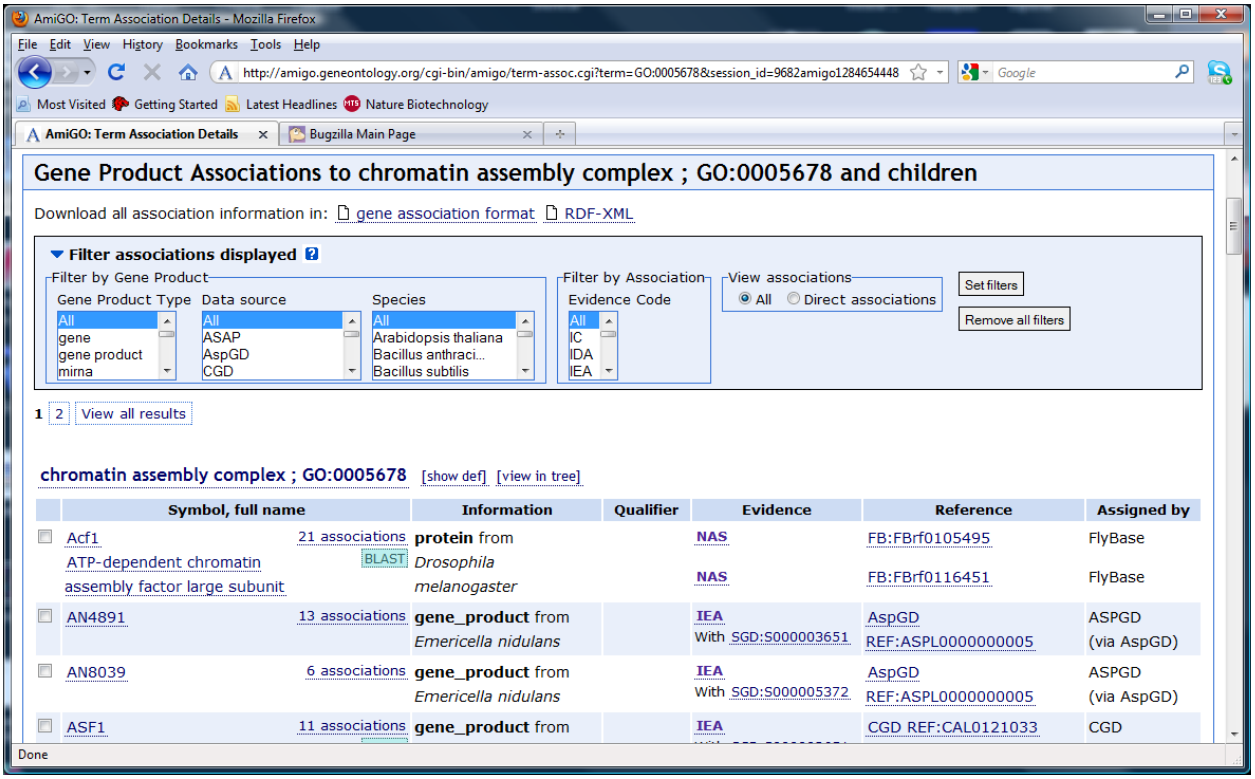
```
⊡ GO:0046808 : assemblon [0 gene products]
⊡ GO:0070310 : ATR-ATRIP complex [0 gene products]
⊡ GO:0070531 : BRCA1-A complex [43 gene products]
⊡ GO:0070532 : BRCA1-B complex [0 gene products]
⊡ GO:0070533 : BRCA1-C complex [0 gene products]
⊡ GO:0070767 : BRCA1-Rad51 complex [0 gene products]
⊡ GO:0002111 : BRCA2-BRAF35 complex [1 gene product]
⊡ GO:0070552 : BRISC complex [31 gene products]
⊡ GO:0070516 : CAK-ERCC2 complex [0 gene products]
⊞ GO:0032806 : carboxy-terminal domain protein kinase complex [47 gene products]
⊟ GO:0005678 : chromatin assembly complex [47 gene products]
    ⊡ GO:0016590 : ACF complex [3 gene products]
    ⊡ GO:0033186 : CAF-1 complex [9 gene products]
    ⊡ GO:0000417 : HIR complex [8 gene products]
⊞ GO:0016585 : chromatin remodeling complex [678 gene products]
⊡ GO:0032116 : cohesin loading complex [51 gene products]
⊡ GO:0032545 : CURI complex [11 gene products]
⊡ GO:0019907 : cyclin-dependent protein kinase activating kinase holoenzyme complex [16 gene products]
⊡ GO:0046818 : dense nuclear body [0 gene products]
⊡ GO:0033203 : DNA helicase A complex [3 gene products]
⊡ GO:0070421 : DNA ligase III-XRCC1 complex [0 gene products]
⊡ GO:0032807 : DNA ligase IV complex [13 gene products]
⊡ GO:0033061 : DNA recombinase mediator complex [8 gene products]
⊡ GO:0070418 : DNA-dependent protein kinase complex [0 gene products]
```

Now go back and click on the 47 gene products that correspond to the “chromatin assembly complex”.


AmiGO: Tree Browser - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://amigo.geneontology.org/cgi-bin/amigo/browse.cgi?action=plus_node&target=
Google
Most Visited
Getting Started
Latest Headlines
Nature Biotechnology
AmiGO: Tree Browser
Bugzilla Main Page
Done

GO:0031040 : micronucleus [0 gene products]
GO:0048556 : microsporocyte nucleus [0 gene products]
GO:0044428 : nuclear part [12331 gene products]
GO:0030895 : apolipoprotein B mRNA editing enzyme complex [6 gene products]
GO:0033167 : ARC complex [3 gene products]
GO:0046808 : assemblon [0 gene products]
GO:0070310 : ATR-ATRIP complex [0 gene products]
GO:0070531 : BRCA1-A complex [43 gene products]
GO:0070532 : BRCA1-B complex [0 gene products]
GO:0070533 : BRCA1-C complex [0 gene products]
GO:0070767 : BRCA1-Rad51 complex [0 gene products]
GO:0002111 : BRCA2-BRAF35 complex [1 gene product]
GO:0070552 : BRISC complex [31 gene products]
GO:0070516 : CAK-ERCC2 complex [0 gene products]
GO:0032806 : carboxy-terminal domain protein kinase complex [47 gene products]
GO:0005678 : chromatin assembly complex [47 gene products]
GO:0016585 : chromatin remodeling complex [678 gene products]
GO:0032116 : cohesin loading complex [51 gene products]
GO:0032545 : CURI complex [11 gene products]
GO:0019907 : cyclin-dependent protein kinase activating kinase holoenzyme complex [16 gene products]
GO:0046818 : dense nuclear body [0 gene products]
GO:0033203 : DNA helicase A complex [3 gene products]
GO:0070421 : DNA ligase III-XRCC1 complex [0 gene products]

Click on the “47 gene products” to view them.


AmiGO: Term Association Details - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://amigo.geneontology.org/cgi-bin/amigo/term-assoc.cgi?term=GO:0005678&session_id=9682amigo1284654448
Google
Most Visited Getting Started Latest Headlines Nature Biotechnology
AmiGO: Term Association Details
Bugzilla Main Page
Done

Gene Product Associations to chromatin assembly complex ; GO:0005678 and children

Download all association information in: gene association format RDF-XML

Filter associations displayed

Filter by Gene Product

Gene Product Type: All, gene, gene product, mirna

Data source: All, ASAP, AspGD, CGD

Species: All, Arabidopsis thaliana, Bacillus anthraci..., Bacillus subtilis

Filter by Association

Evidence Code: All, IC, IDA, IEA

View associations: ● All ○ Direct associations

Set filters

Remove all filters

1 2 View all results

chromatin assembly complex ; GO:0005678 [show def] [view in tree]

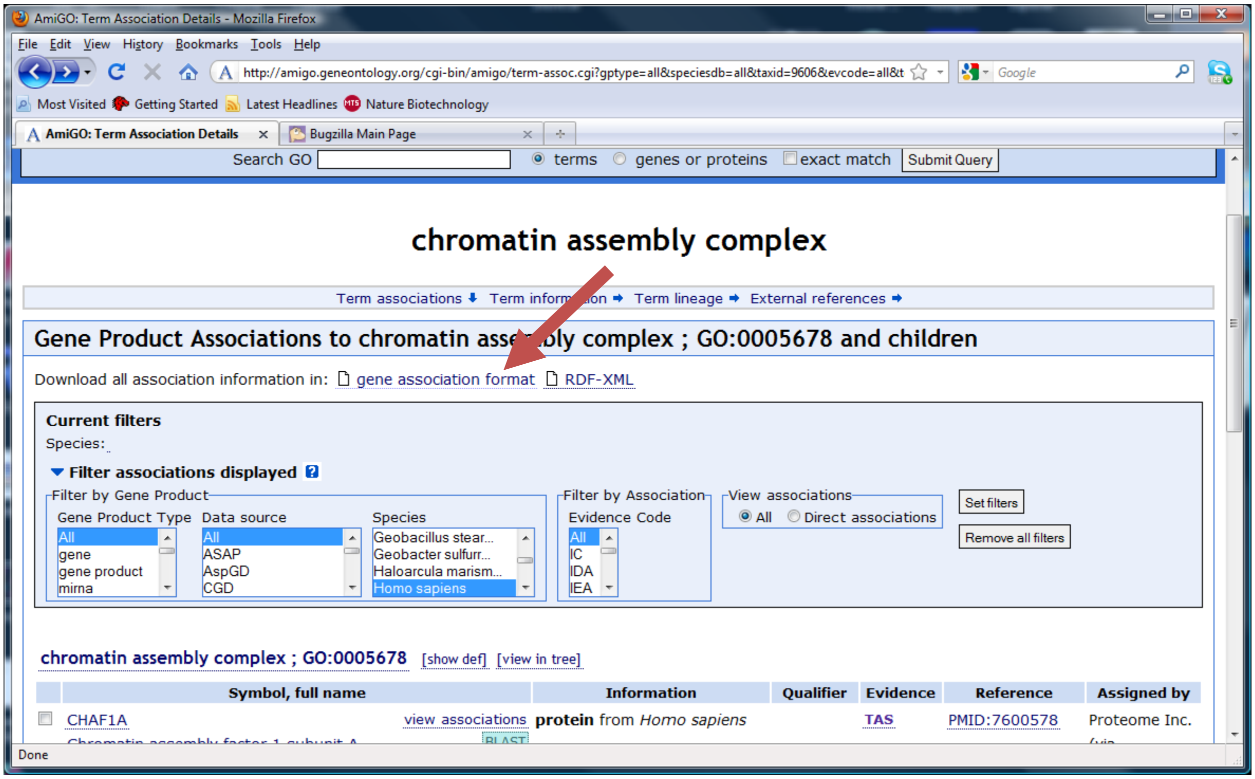
	Symbol, full name		Information	Qualifier	Evidence	Reference	Assigned by
□	Acf1 ATP-dependent chromatin assembly factor large subunit	21 associations BLAST	**protein** from *Drosophila melanogaster*		NAS NAS	FB:FBrf0105495 FB:FBrf0116451	FlyBase FlyBase
□	AN4891	13 associations	**gene_product** from *Emericella nidulans*		IEA With SGD:S000003651	AspGD REF:ASPL0000000005	ASPGD (via AspGD)
□	AN8039	6 associations	**gene_product** from *Emericella nidulans*		IEA With SGD:S000005372	AspGD REF:ASPL0000000005	ASPGD (via AspGD)
□	ASF1	11 associations	**gene_product** from		IEA	CGD REF:CAL0121033	CGD

Spend some time browsing the links from this page (or just hovering over them). Scroll down. Genes annotated with each of the three leaf terms are grouped separately on the page.

What kinds of evidence are provided for these assignments. What kinds might you be cautious of (less likely to believe) or more likely to believe? Who assigns these GO terms to the genes?

Try filtering the list using the filters at the top. Make selections from the menus (Gene Product Type, Data source, Species and Evidence code and then click “Set filters”.

Next, try to export your results for all genes from all data sources from human (Homo sapiens) that have any evidence code. Set filters and then click on “Download all association information in gene association format”.



AmiGO: Term Association Details - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://amigo.geneontology.org/cgi-bin/amigo/term-assoc.cgi?gptype=all&speciesdb=all&taxid=9606&evcode=all&t
Google
Most Visited
Getting Started
Latest Headlines
Nature Biotechnology
AmiGO: Term Association Details
Bugzilla Main Page
Search GO
terms
genes or proteins
exact match
Submit Query
Set filters
Remove all filters
Done

chromatin assembly complex

Term associations ⬇ Term information ➡ Term lineage ➡ External references ➡

Gene Product Associations to chromatin assembly complex ; GO:0005678 and children

Download all association information in: gene association format RDF-XML

Current filters

Species:

Filter associations displayed

Filter by Gene Product

Gene Product Type	Data source	Species
All	All	Geobacillus stear...
gene	ASAP	Geobacter sulfurr...
gene product	AspGD	Haloarcula marism...
mirna	CGD	Homo sapiens

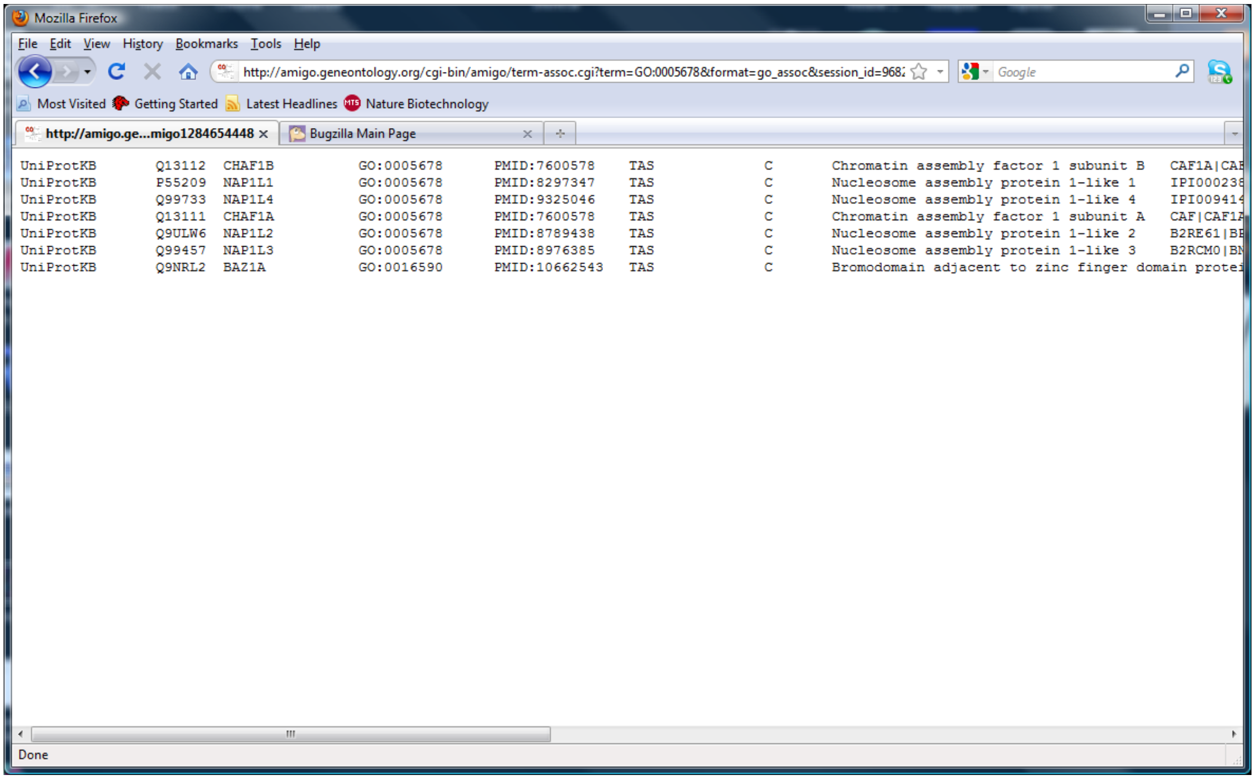
Filter by Association

Evidence Code: All, IC, IDA, IEA

View associations: ● All ○ Direct associations

chromatin assembly complex ; GO:0005678 [show def] [view in tree]

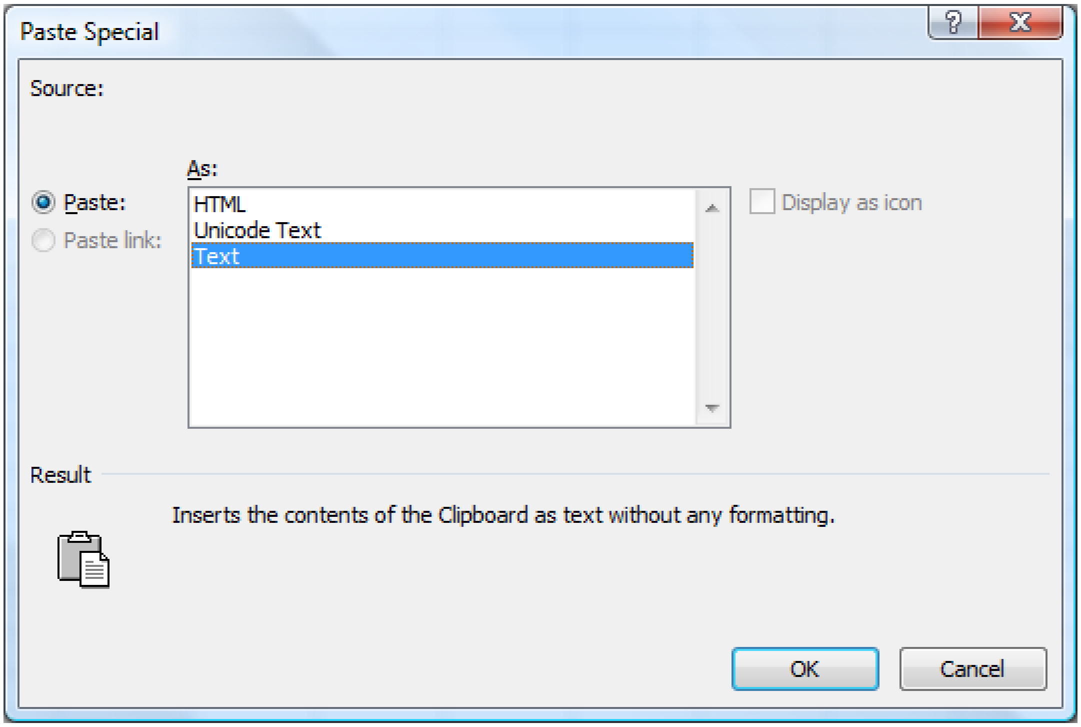
	Symbol, full name		Information	Qualifier	Evidence	Reference	Assigned by
□	CHAF1A Chromatin assembly factor 1 subunit A	view associations BLAST	**protein** from *Homo sapiens*		TAS	PMID:7600578	Proteome Inc. (via



Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://amigo.geneontology.org/cgi-bin/amigo/term-assoc.cgi?term=GO:0005678&format=go_assoc&session_id=968
Google
Most Visited
Getting Started
Latest Headlines
Nature Biotechnology
http://amigo.ge...migo1284654448
Bugzilla Main Page
Done

```
UniProtKB       Q13112  CHAF1B          GO:0005678      PMID:7600578    TAS             C       Chromatin assembly factor 1 subunit B   CAF1A|CAE
UniProtKB       P55209  NAP1L1          GO:0005678      PMID:8297347    TAS             C       Nucleosome assembly protein 1-like 1    IPI00023€
UniProtKB       Q99733  NAP1L4          GO:0005678      PMID:9325046    TAS             C       Nucleosome assembly protein 1-like 4    IPI009414
UniProtKB       Q13111  CHAF1A          GO:0005678      PMID:7600578    TAS             C       Chromatin assembly factor 1 subunit A   CAF|CAF1A
UniProtKB       Q9ULW6  NAP1L2          GO:0005678      PMID:8789438    TAS             C       Nucleosome assembly protein 1-like 2    B2RE61|BE
UniProtKB       Q99457  NAP1L3          GO:0005678      PMID:8976385    TAS             C       Nucleosome assembly protein 1-like 3    B2RCM0|BN
UniProtKB       Q9NRL2  BAZ1A           GO:0016590      PMID:10662543   TAS             C       Bromodomain adjacent to zinc finger domain protei
```

Copy and paste this to an Excel spread sheet (or similar). Right-click, paste special, as text.



Paste Special

Source:

As:

● Paste:

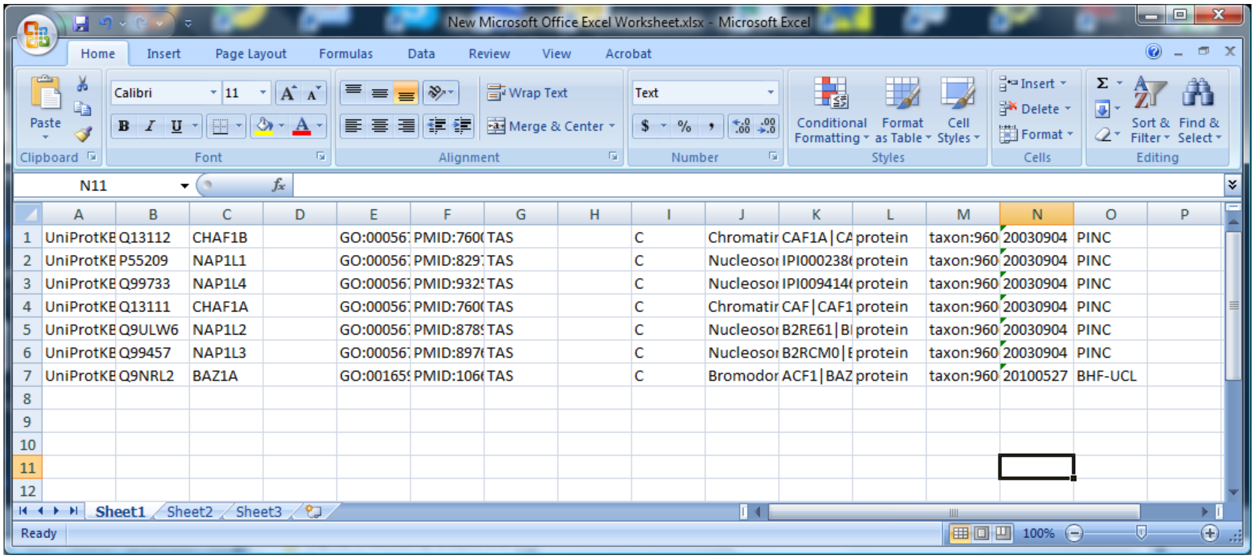
○ Paste link:

HTML
Unicode Text
Text

□ Display as icon

Result

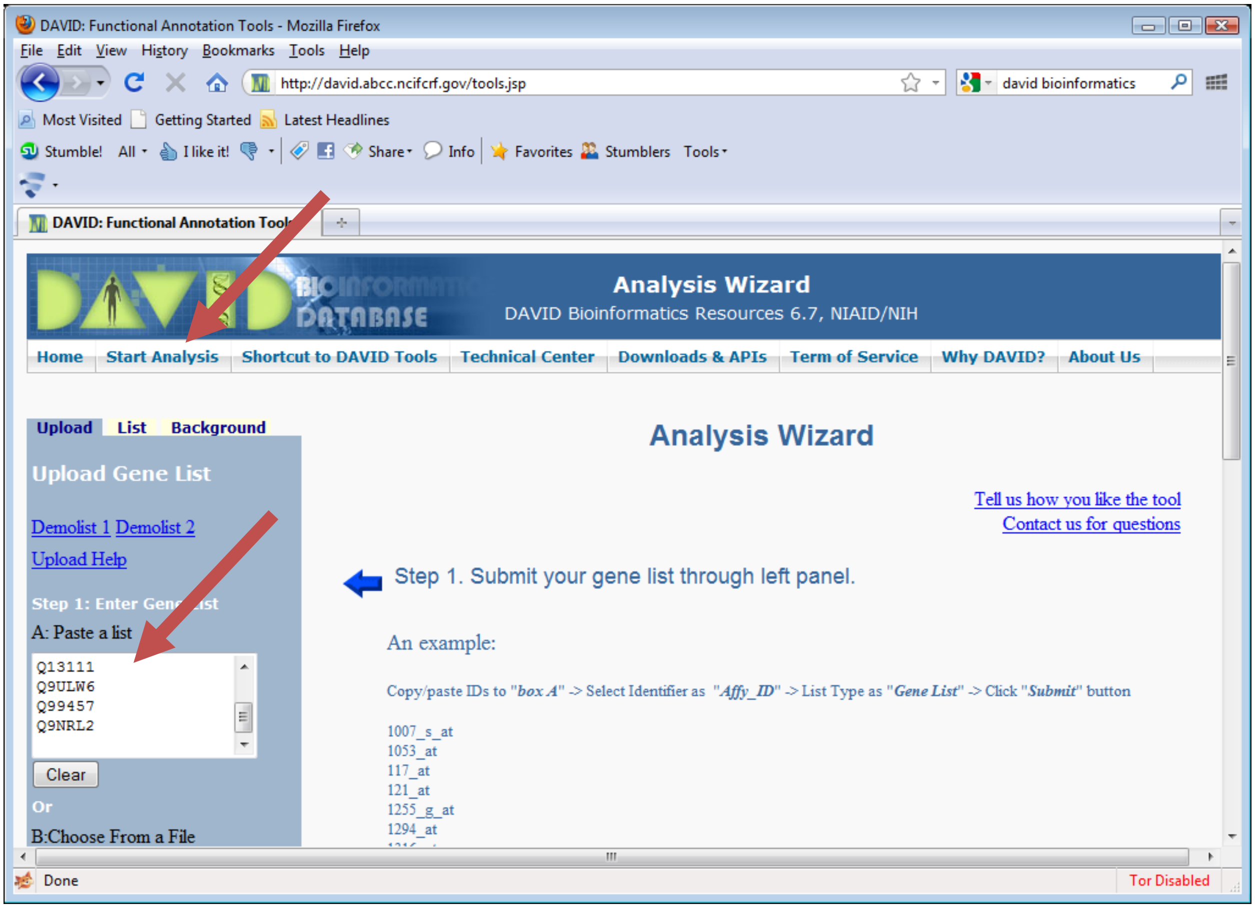
Inserts the contents of the Clipboard as text without any formatting.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	UniProtKE	Q13112	CHAF1B		GO:00056	PMID:760	TAS		C	Chromatir	CAF1A|CA	protein	taxon:960	20030904	PINC	
2	UniProtKE	P55209	NAP1L1		GO:00056	PMID:829	TAS		C	Nucleoso	IPI000238	protein	taxon:960	20030904	PINC	
3	UniProtKE	Q99733	NAP1L4		GO:00056	PMID:932	TAS		C	Nucleoso	IPI009414	protein	taxon:960	20030904	PINC	
4	UniProtKE	Q13111	CHAF1A		GO:00056	PMID:760	TAS		C	Chromatir	CAF|CAF1	protein	taxon:960	20030904	PINC	
5	UniProtKE	Q9ULW6	NAP1L2		GO:00056	PMID:878	TAS		C	Nucleoso	B2RE61|B	protein	taxon:960	20030904	PINC	
6	UniProtKE	Q99457	NAP1L3		GO:00056	PMID:897	TAS		C	Nucleoso	B2RCM0|E	protein	taxon:960	20030904	PINC	
7	UniProtKE	Q9NRL2	BAZ1A		GO:00165	PMID:106	TAS		C	Bromodor	ACF1|BAZ	protein	taxon:960	20100527	BHF-UCL	

Select and copy the UniProt accessions from column B.

Go to http://david.abcc.ncifcrf.gov/home.jsp , click on Start Analysis and then paste the list of accessions into the query box like this:


DAVID: Functional Annotation Tools - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Getting Started Latest Headlines
Stumble! All I like it! Share Info Favorites Stumblers Tools
DAVID: Functional Annotation Tools
BIOINFORMATICS DATABASE
Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us
Upload List Background
Demolist 1 Demolist 2
Upload Help
Clear
Tell us how you like the tool
Contact us for questions
Done
Tor Disabled

http://david.abcc.ncifcrf.gov/tools.jsp

david bioinformatics

Analysis Wizard

DAVID Bioinformatics Resources 6.7, NIAID/NIH

Analysis Wizard

Upload Gene List

Step 1: Enter Gene List

A: Paste a list

```
Q13111
Q9ULW6
Q99457
Q9NRL2
```

Or

B:Choose From a File

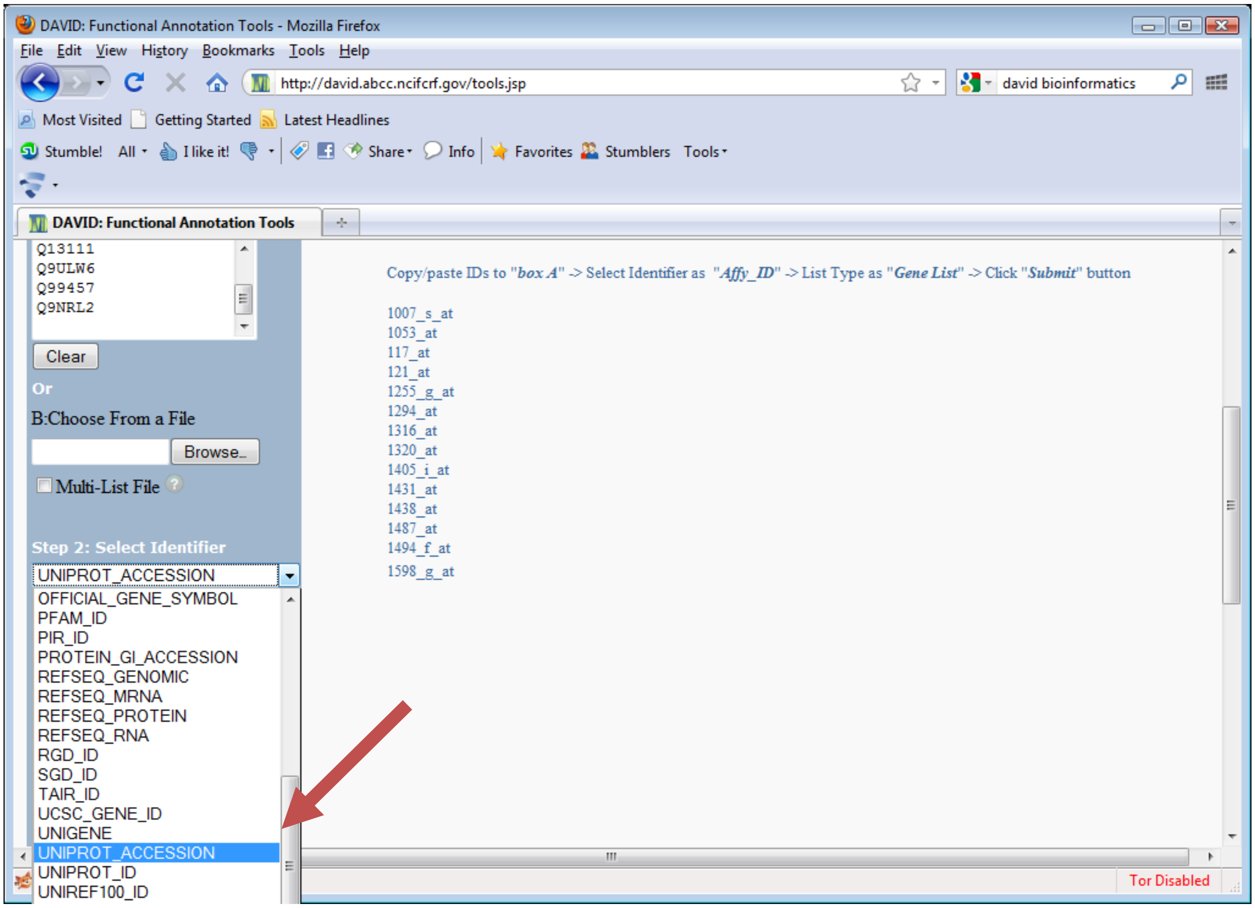
Step 1. Submit your gene list through left panel.

An example:

Copy/paste IDs to "*box A*" -> Select Identifier as "*Affy_ID*" -> List Type as "*Gene List*" -> Click "*Submit*" button

1007_s_at
1053_at
117_at
121_at
1255_g_at
1294_at

The tell DAVID that you have entered UniProt Accessions. Like this


DAVID: Functional Annotation Tools - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://david.abcc.ncifcrf.gov/tools.jsp
david bioinformatics
Most Visited Getting Started Latest Headlines
Stumble! All I like it! Share Info Favorites Stumblers Tools
DAVID: Functional Annotation Tools
Clear
Or
B:Choose From a File
Browse...
Multi-List File
OFFICIAL_GENE_SYMBOL
PFAM_ID
PIR_ID
PROTEIN_GI_ACCESSION
REFSEQ_GENOMIC
REFSEQ_MRNA
REFSEQ_PROTEIN
REFSEQ_RNA
RGD_ID
SGD_ID
TAIR_ID
UCSC_GENE_ID
UNIGENE
UNIPROT_ACCESSION
UNIPROT_ID
UNIREF100_ID
Tor Disabled

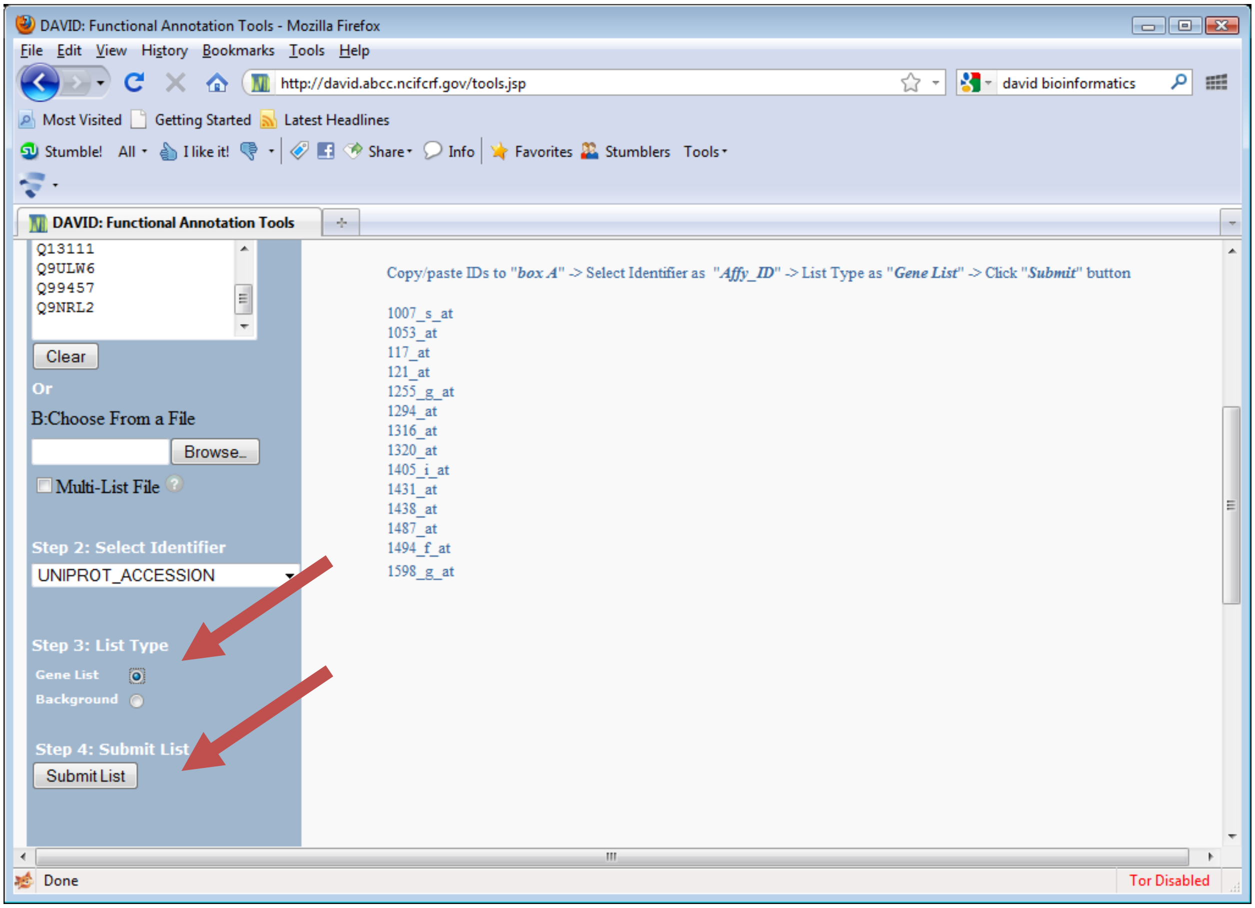
Q13111
Q9ULW6
Q99457
Q9NRL2

Step 2: Select Identifier: UNIPROT_ACCESSION

Copy/paste IDs to "*box A*" -> Select Identifier as "*Affy_ID*" -> List Type as "*Gene List*" -> Click "*Submit*" button

1007_s_at
1053_at
117_at
121_at
1255_g_at
1294_at
1316_at
1320_at
1405_i_at
1431_at
1438_at
1487_at
1494_f_at
1598_g_at

Then select “Gene List” under List type and click on Submit list:



DAVID: Functional Annotation Tools - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://david.abcc.ncifcrf.gov/tools.jsp
david bioinformatics
Most Visited
Getting Started
Latest Headlines
Stumble!
All
I like it!
Share
Info
Favorites
Stumblers
Tools
DAVID: Functional Annotation Tools
Clear
Browse...
Submit List
Done
Tor Disabled

Q13111
Q9ULW6
Q99457
Q9NRL2

Or

B:Choose From a File

□ Multi-List File

Step 2: Select Identifier

UNIPROT_ACCESSION

Step 3: List Type

Gene List ●

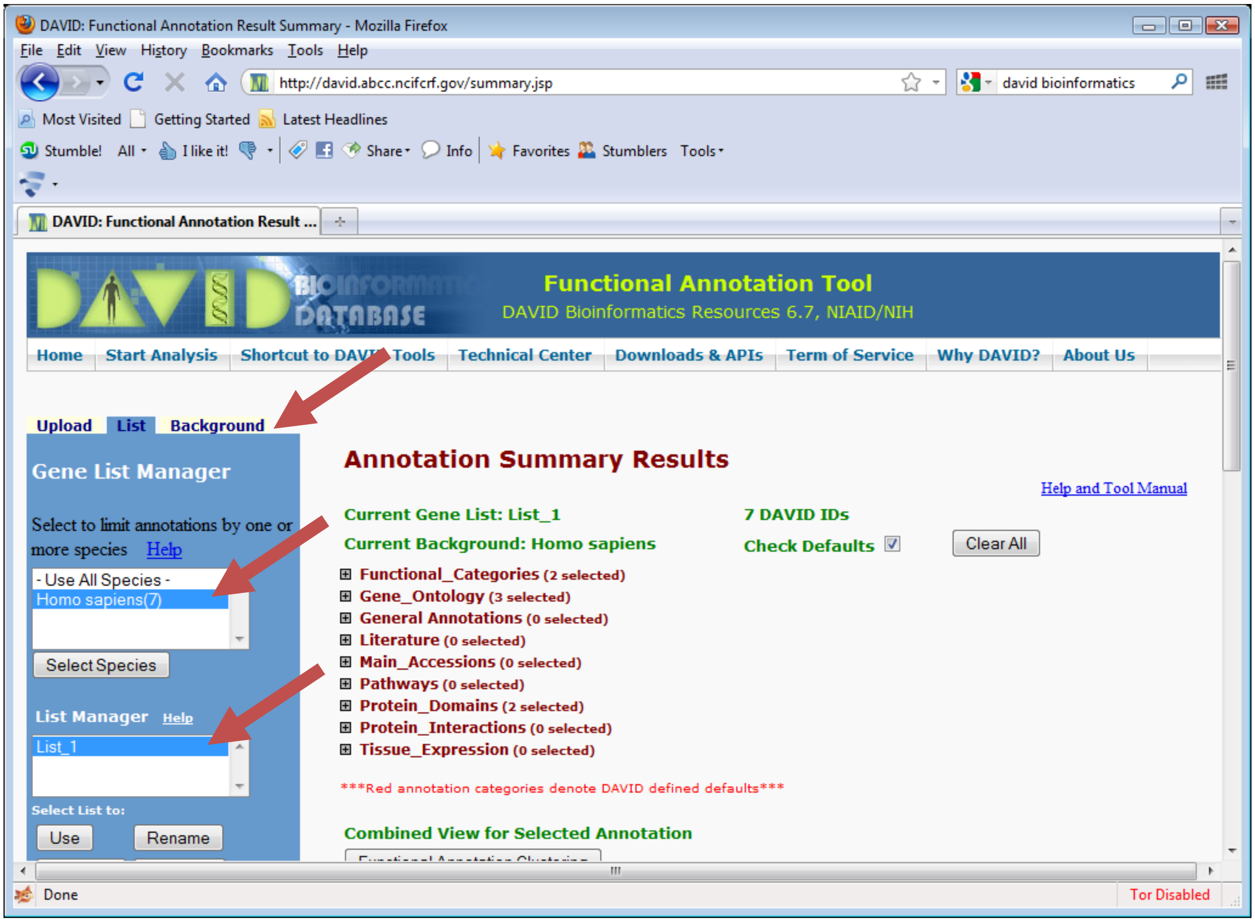
Background ○

Step 4: Submit List

Copy/paste IDs to "*box A*" -> Select Identifier as "*Affy_ID*" -> List Type as "*Gene List*" -> Click "*Submit*" button

1007_s_at
1053_at
117_at
121_at
1255_g_at
1294_at
1316_at
1320_at
1405_i_at
1431_at
1438_at
1487_at
1494_f_at
1598_g_at

Your list has been saved by DAVID as List_1 and it has automatically recognized “Homo sapiens” as the species from which the list is derived:


DAVID: Functional Annotation Result Summary - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://david.abcc.ncifcrf.gov/summary.jsp
david bioinformatics
Most Visited Getting Started Latest Headlines
Stumble! All I like it! Share Info Favorites Stumblers Tools
DAVID: Functional Annotation Result ...
DAVID BIOINFORMATICS DATABASE
Functional Annotation Tool
DAVID Bioinformatics Resources 6.7, NIAID/NIH
Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us
Upload List Background
Help and Tool Manual
Clear All
Select Species
Use Rename
Combined View for Selected Annotation
Functional Annotation Clustering
Done
Tor Disabled

Gene List Manager

Select to limit annotations by one or more species Help

- Use All Species -
Homo sapiens(7)

List Manager Help

List_1

Select List to:

Annotation Summary Results

Current Gene List: List_1 7 DAVID IDs

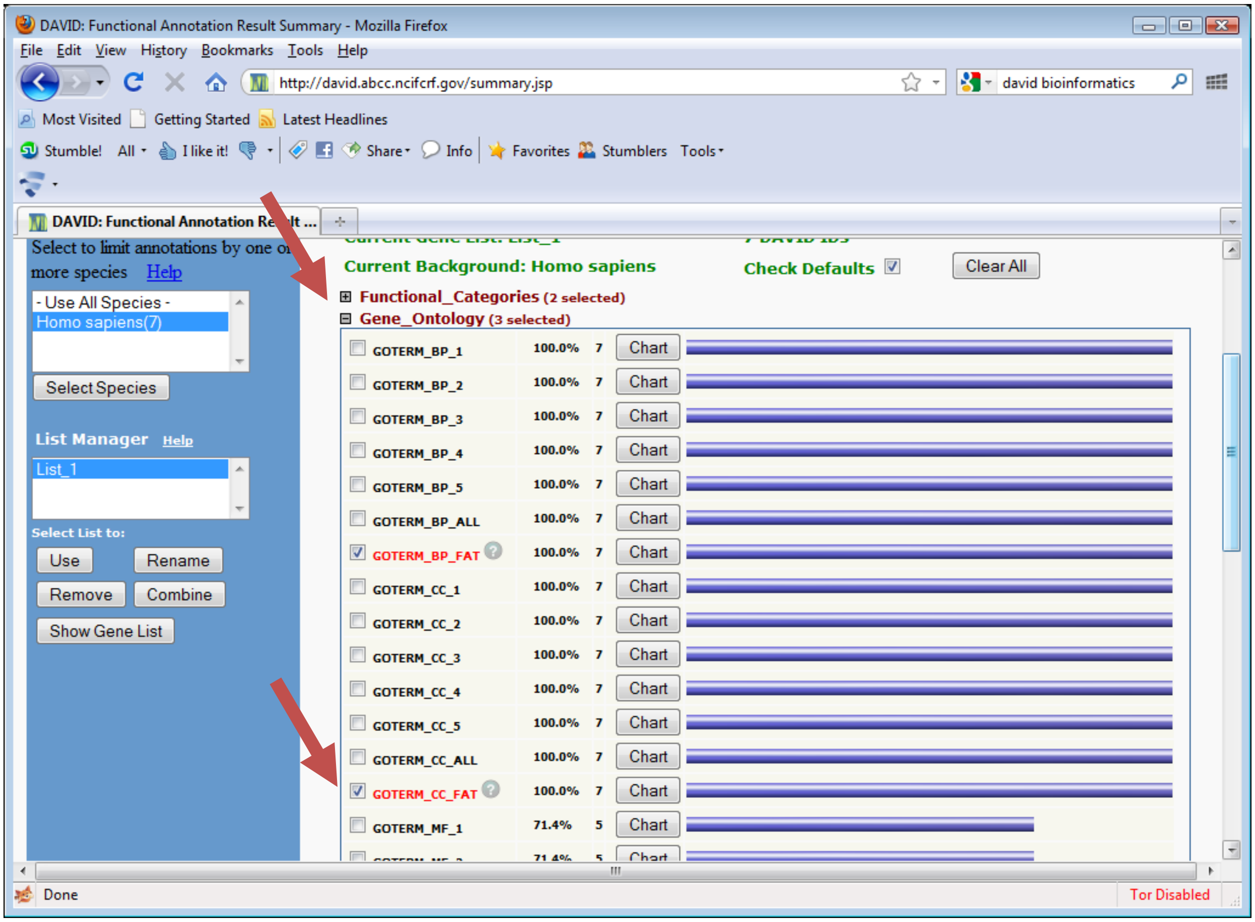
Current Background: Homo sapiens Check Defaults ☑

⊞ Functional_Categories (2 selected)
⊞ Gene_Ontology (3 selected)
⊞ General Annotations (0 selected)
⊞ Literature (0 selected)
⊞ Main_Accessions (0 selected)
⊞ Pathways (0 selected)
⊞ Protein_Domains (2 selected)
⊞ Protein_Interactions (0 selected)
⊞ Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

If you click on the “Background” tab, you will see that DAVID has also set “Homo sapiens” as the “background”. Its important that you check these settings and change them if necessary. Discuss why.

You can now look at categories that are over-represented in your list. For example, click on the + beside Gene_Ontology and scroll down to GOTERM_CC_FAT like this:


DAVID: Functional Annotation Result Summary - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://david.abcc.ncifcrf.gov/summary.jsp
david bioinformatics
Most Visited Getting Started Latest Headlines
Stumble! All I like it! Share Info Favorites Stumblers Tools
DAVID: Functional Annotation Re...lt ...
Select Species
Help
Use Rename
Remove Combine
Show Gene List
Clear All
Done
Tor Disabled

Select to limit annotations by one or more species Help

- Use All Species -
Homo sapiens(7)

List Manager Help

List_1

Select List to:

Current Background: Homo sapiens

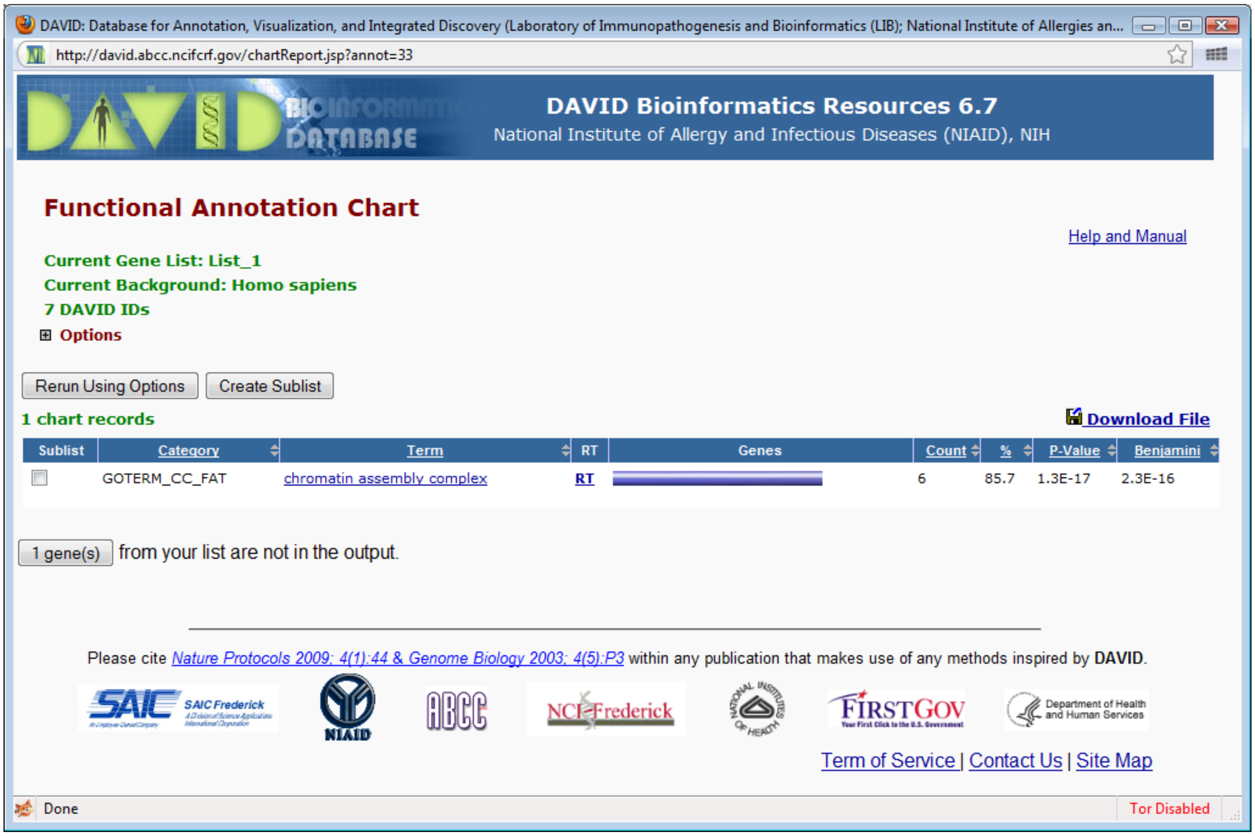
Check Defaults ☑

⊞ Functional_Categories (2 selected)

⊟ Gene_Ontology (3 selected)

☐	GOTERM_BP_1	100.0%	7	Chart
☐	GOTERM_BP_2	100.0%	7	Chart
☐	GOTERM_BP_3	100.0%	7	Chart
☐	GOTERM_BP_4	100.0%	7	Chart
☐	GOTERM_BP_5	100.0%	7	Chart
☐	GOTERM_BP_ALL	100.0%	7	Chart
☑	GOTERM_BP_FAT	100.0%	7	Chart
☐	GOTERM_CC_1	100.0%	7	Chart
☐	GOTERM_CC_2	100.0%	7	Chart
☐	GOTERM_CC_3	100.0%	7	Chart
☐	GOTERM_CC_4	100.0%	7	Chart
☐	GOTERM_CC_5	100.0%	7	Chart
☐	GOTERM_CC_ALL	100.0%	7	Chart
☑	GOTERM_CC_FAT	100.0%	7	Chart
☐	GOTERM_MF_1	71.4%	5	Chart
☐	GOTERM_MF_2	71.4%	5	Chart

If you then click on the “Chart” button, you see a new window open like this:


DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunopathogenesis and Bioinformatics (LIB); National Institute of Allergies an...
http://david.abcc.ncifcrf.gov/chartReport.jsp?annot=33
DAVID BIOINFORMATICS DATABASE
DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH
Help and Manual
Options
Rerun Using Options
Create Sublist
Download File
1 gene(s)
SAIC SAIC Frederick
NIAID
ABCC
NCI-Frederick
FIRSTGOV
Department of Health and Human Services
Term of Service | Contact Us | Site Map
Done
Tor Disabled

Functional Annotation Chart

Current Gene List: List_1
Current Background: Homo sapiens
7 DAVID IDs

1 chart records

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
□	GOTERM_CC_FAT	chromatin assembly complex	RT		6	85.7	1.3E-17	2.3E-16

1 gene(s) from your list are not in the output.

Please cite *Nature Protocols 2009; 4(1):44 & Genome Biology 2003; 4(5):P3* within any publication that makes use of any methods inspired by **DAVID**.

6 out of the 7 genes in your list (85.7%) are annotated with the GO term for chromatin assembly complex. The probability of randomly choosing 7 genes from the human genome where 6 of them all have this associated GO term is 1.3E.17. When you correct for multiple hypothesis testing (Benjamini), the probability is 2.3E-16 (still quite surprising – i.e. we would suspect that whatever "process" was used to pick out these 7 genes was not unrelated to this annotation). And of course, we know this to be the case.

Click on the "chromatin assembly complex" to see details about the term that is overrepresented.

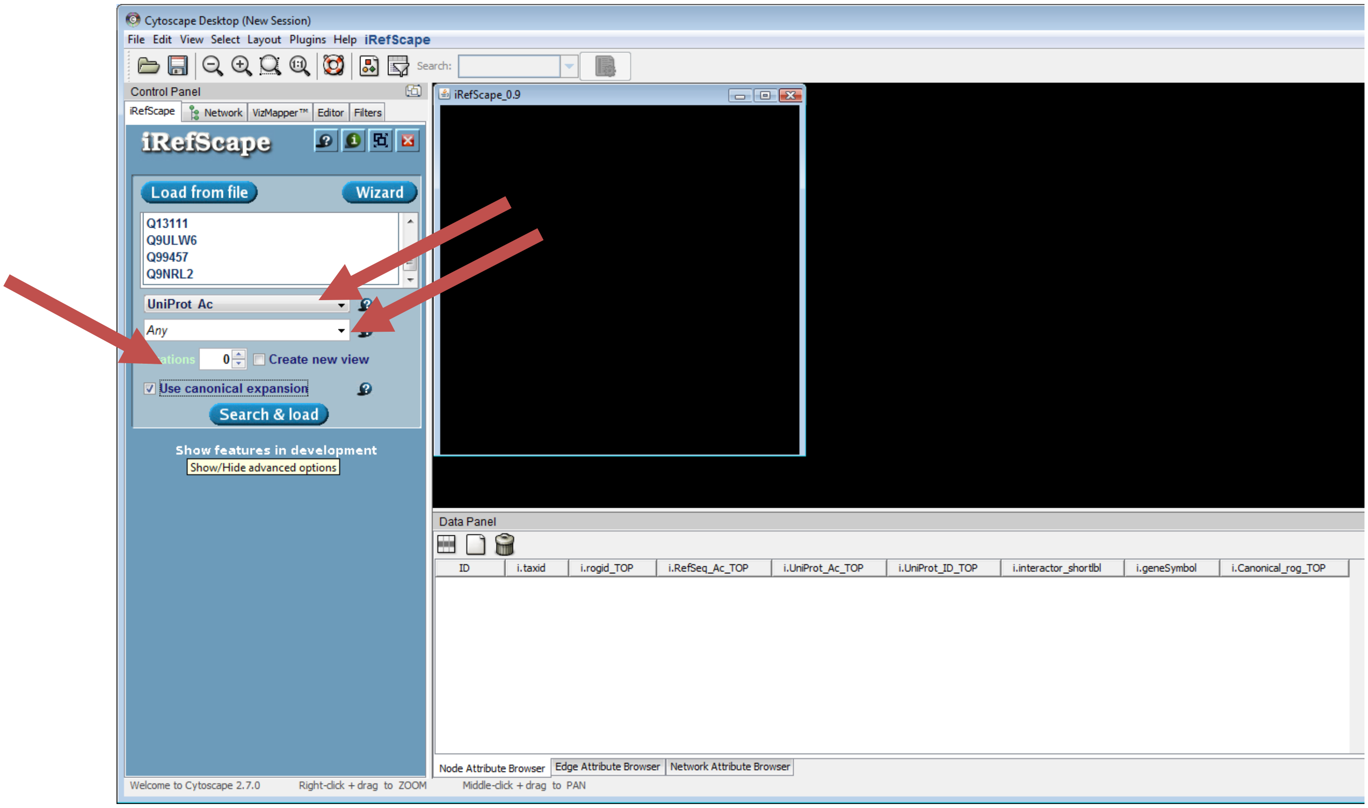
Click on the blue bar underneath "Genes" to see the list of genes that had this annotation.

Go back to the "Annotation summary results" and explore whether other categories of annotation were over represented in this list.

There is a lot of other material you can explore on this site. Make a note of the Nature Protocols tutorial on use of DAVID for later.
http://www.nature.com/nprot/journal/v4/n1/pdf/nprot.2008.211.pdf

Next, we will explore known interactions between the genes in our list.

Start Cytoscape and the iRefScape plugin from Plugins menu?iRefScape 0.9. Then copy and paste the identifiers from column B of the Excel spreadsheet to the query box. Like this…



Q13111
Q9ULW6
Q99457
Q9NRL2

UniProt Ac

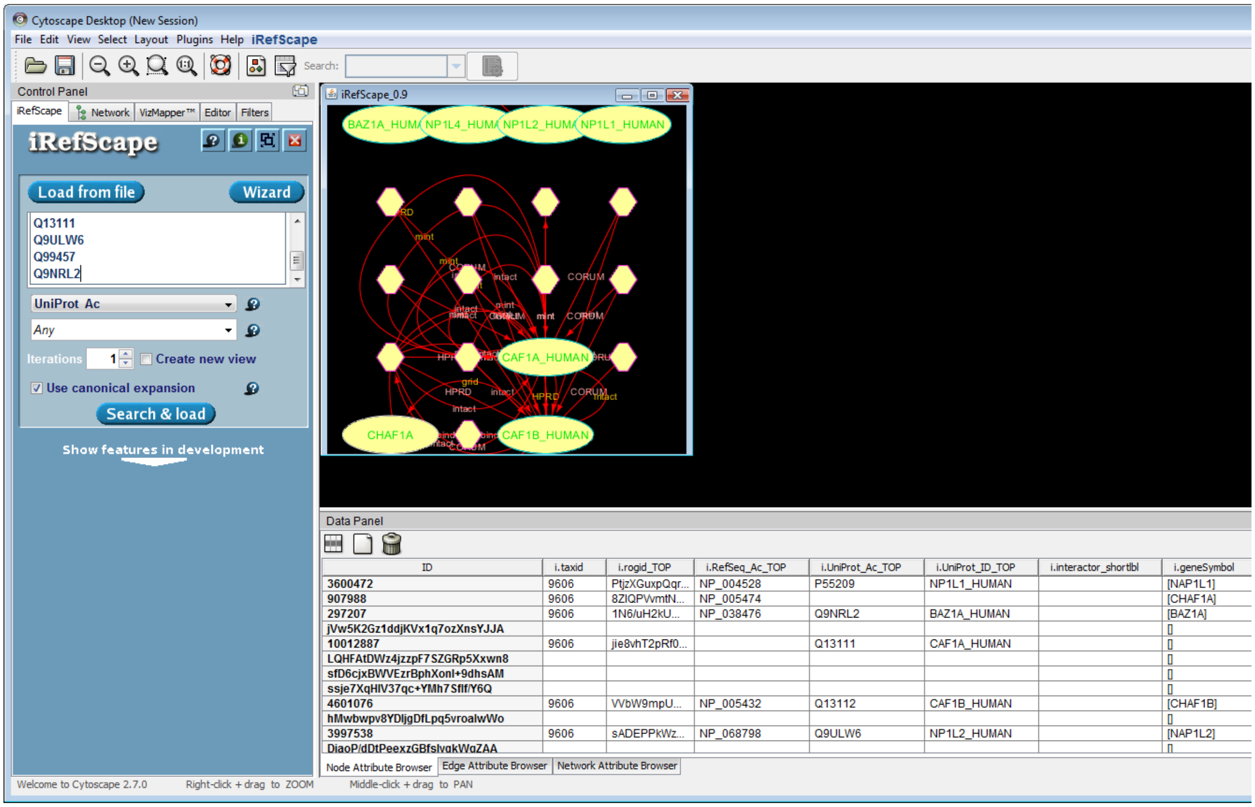
Any

ations 0 ☐ Create new view

☑ Use canonical expansion

Make sure that the search type is set to UniProt Ac, the taxon is set to Any (or Homo Sapiens) and that “Use canonical expansion” is selected and that iterations are set to 0 (see red arrows above). Iterations of zero will only return interactions that occur between proteins in our query list.   Then click on search and load.

The initial results will look like this



Q13111
Q9ULW6
Q99457
Q9NRL2

UniProt Ac

Any

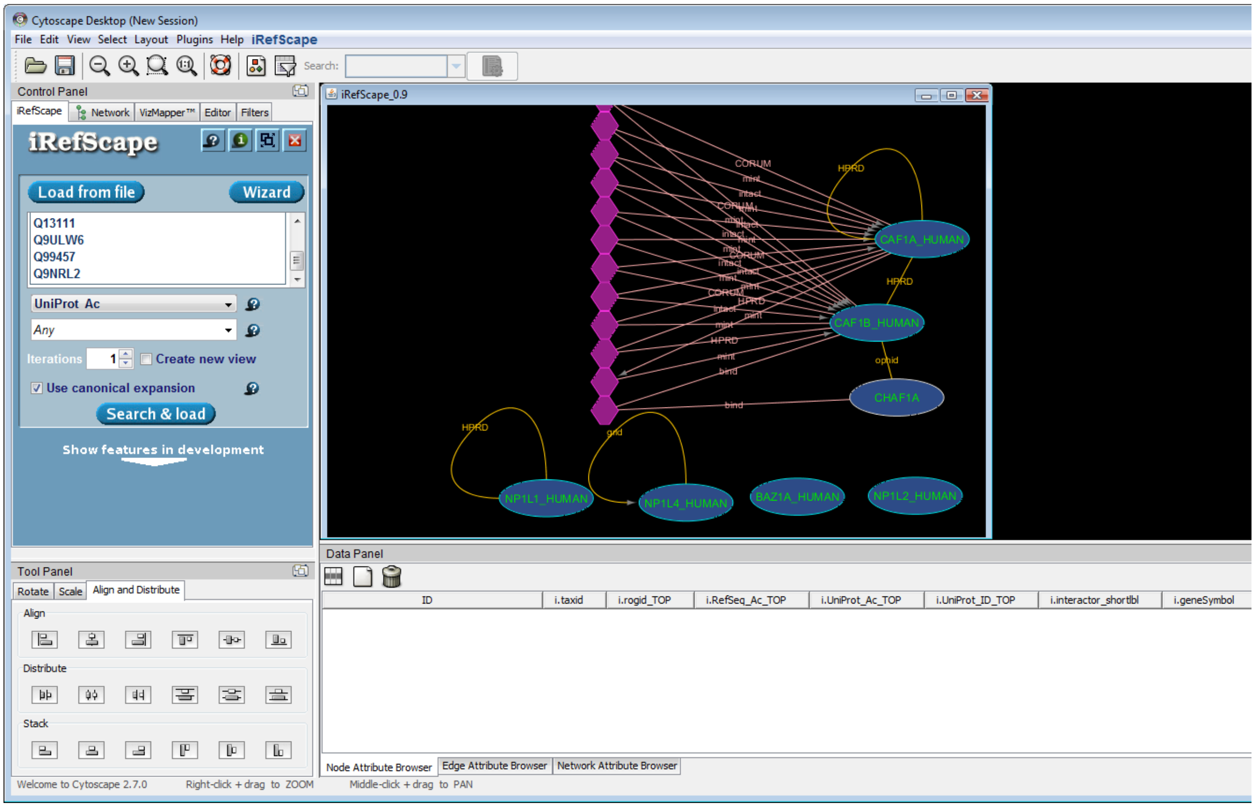
Iterations 1 ☐ Create new view

☑ Use canonical expansion

ID	i.taxid	i.rogid_TOP	i.RefSeq_Ac_TOP	i.UniProt_Ac_TOP	i.UniProt_ID_TOP	i.interactor_shortlbl	i.geneSymbol
3600472	9606	PtjzXGuxpQqr...	NP_004528	P55209	NP1L1_HUMAN		[NAP1L1]
907988	9606	8ZIQPVvmtN...	NP_005474				[CHAF1A]
297207	9606	1N6/uH2kU...	NP_038476	Q9NRL2	BAZ1A_HUMAN		[BAZ1A]
jVw5K2Gz1ddjKVx1q7ozXnsYJJA							[]
10012887	9606	jie8vhT2pRf0...		Q13111	CAF1A_HUMAN		[]
LQHFAtDWz4jzzpF7SZGRp5Xxwn8							[]
sfD6cjxBWVEzrBphXonl+9dhsAM							[]
ssje7XqHlV37qc+YMh7Sflf/Y6Q							[]
4601076	9606	VVbW9mpU...	NP_005432	Q13112	CAF1B_HUMAN		[CHAF1B]
hMwbwpv8YDljgDfLpq5vroalwWo							[]
3997538	9606	sADEPPkWz...	NP_068798	Q9ULW6	NP1L2_HUMAN		[NAP1L2]
DiaoP/dDtPeexzGBfslvgkWgZAA							[]

And you should be able to clean them up (rearrange them) to look like this.

Hint use Layout/align and distribute to align and stack the purple hexagons (nodes that represent complexes) and then iRefScape/ViewTools/Toggle selected multi-edges to hide multi-edges (representing multiple experiments that support the same protein-protein interaction).



From this view it should be apparent that

1) 4 of the nodes don't interact with any of the other proteins in the list. Although 2 of them are self interacting (loops)
2) 2 of the proteins (CAF1A and CAF1B) appear to be co-members of multiple complexes that are documented by multiple databases (these complexes are represented by the purple hexagons). By definition, these hexagons represent complex records with 3 or more proteins.
3) These same two proteins are reported to interact with one another by one database (HPRD). You will only se this interaction if you are using the version of iRefScape that contains HPRD data (you have to ask for this version).
4) If you want to explore the evidence for any interaction, click on the edge and look at details in the edge attribute browser. For example, the i.PMID feature lists the publication where the evidence for the interaction was found.
5) If you want to explore the other members of each of the complexes, click on the complex node ("pseudo-node") and then select iRefScape/Search tools/ retrieve interactions for selected nodes.

Go back to the Gene Ontology pages. Should more of these proteins have been involved in a single complex?

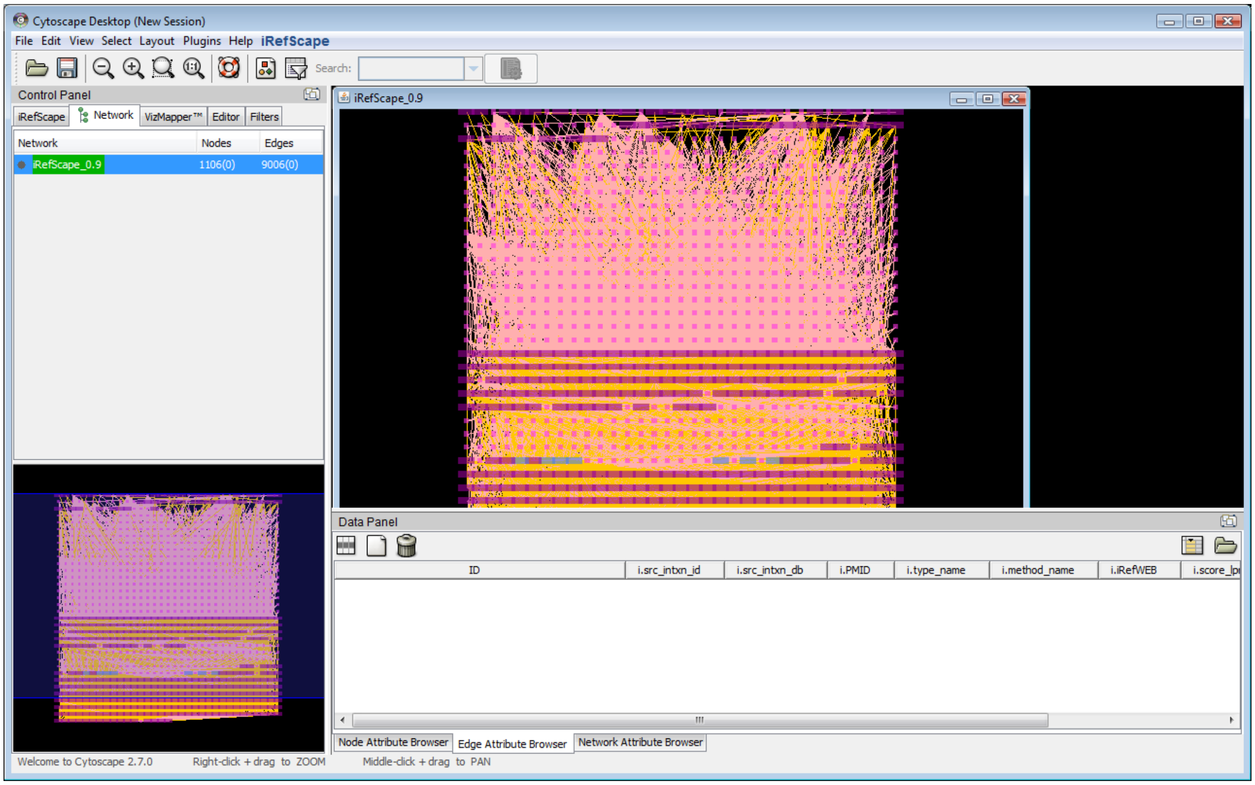
It is left as an exercise to see if you would get different results by querying iRefScape with proteins from different organisms (say Drosophila) that are annotated as belonging to the "chromatin assembly complex".

Next, we will move on to a technique that will identify proteins related to our initial query. We want to find proteins that interact with 2 or more of the proteins in our starting list (i.e. things that are associated with the "chromatin assembly complex".

Select all nodes in the current view.

Retrieve their neighbours using iRefIndex menu/Search tools/Retrieve interactions for selected nodes.

You will see (a ridiculously large network – 1106 nodes and 9006 edges) like this:

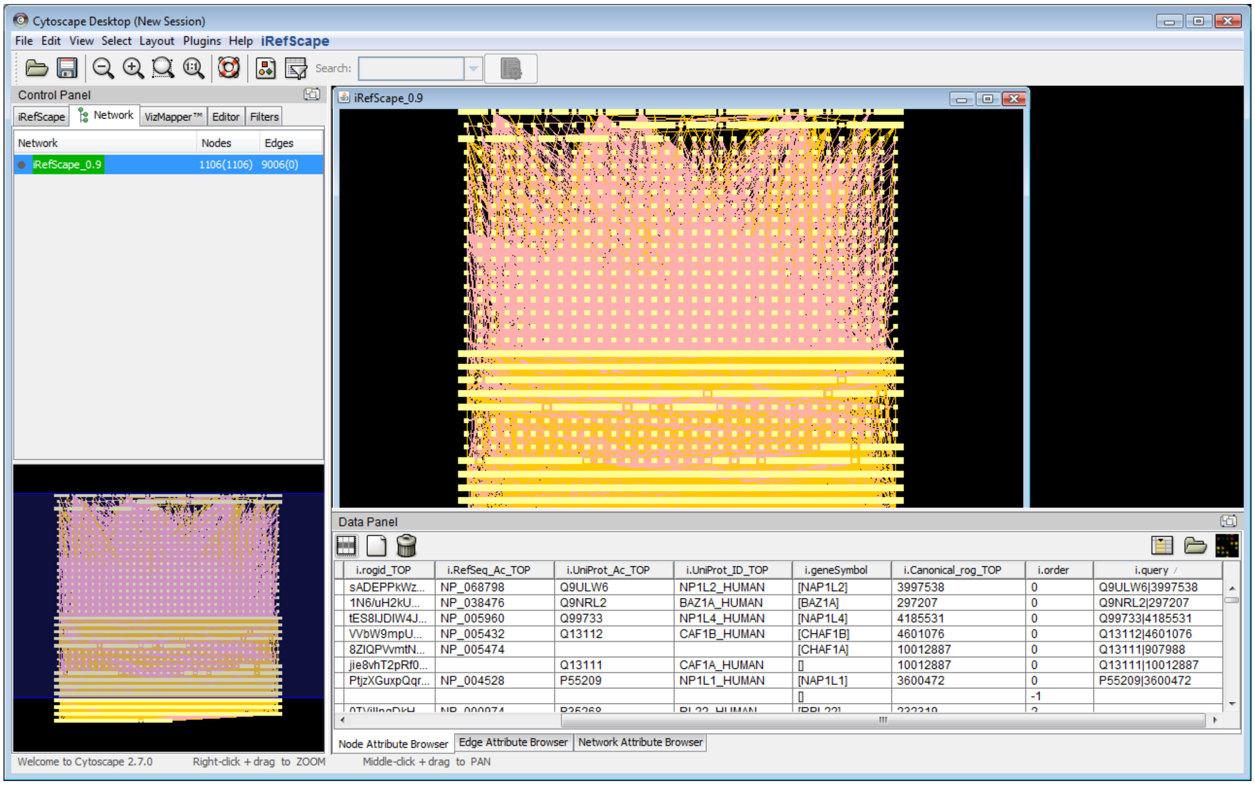


Network	Nodes	Edges
iRefScape_0.9	1106(0)	9006(0)

Welcome to Cytoscape 2.7.0 Right-click + drag to ZOOM Middle-click + drag to PAN

Select all nodes using control-A.

In the node attribute browser, select to view the i.query node feature and then sort on this feature by clicking on the i.query column heading. You should see all the nodes from your initial search at the top of the node attribute browser. Like this:

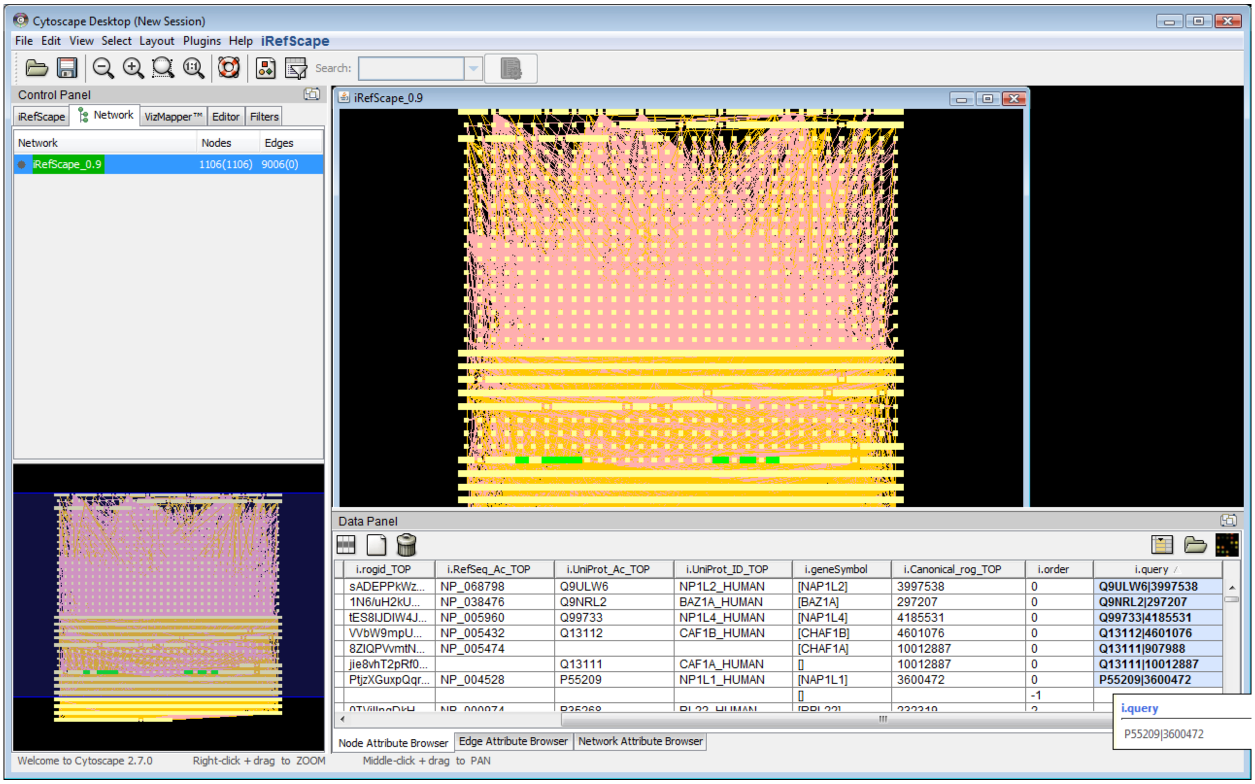

Cytoscape Desktop (New Session)
File Edit View Select Layout Plugins Help iRefScape
Search:
Control Panel
iRefScape
Network
VizMapper™
Editor
Filters
iRefScape_0.9
Data Panel
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser
Welcome to Cytoscape 2.7.0
Right-click + drag to ZOOM
Middle-click + drag to PAN

Network	Nodes	Edges
iRefScape_0.9	1106(1106)	9006(0)

i.rogid_TOP	i.RefSeq_Ac_TOP	i.UniProt_Ac_TOP	i.UniProt_ID_TOP	i.geneSymbol	i.Canonical_rog_TOP	i.order	i.query
sADEPPkWz...	NP_068798	Q9ULW6	NP1L2_HUMAN	[NAP1L2]	3997538	0	Q9ULW6|3997538
1N6/uH2kU...	NP_038476	Q9NRL2	BAZ1A_HUMAN	[BAZ1A]	297207	0	Q9NRL2|297207
tES8IJDIW4J...	NP_005960	Q99733	NP1L4_HUMAN	[NAP1L4]	4185531	0	Q99733|4185531
VVbW9mpU...	NP_005432	Q13112	CAF1B_HUMAN	[CHAF1B]	4601076	0	Q13112|4601076
8ZIQPVvmtN...	NP_005474			[CHAF1A]	10012887	0	Q13111|907988
jie8vhT2pRf0...		Q13111	CAF1A_HUMAN	[]	10012887	0	Q13111|10012887
PtjzXGuxpQqr...	NP_004528	P55209	NP1L1_HUMAN	[NAP1L1]	3600472	0	P55209|3600472
				[]		-1	
0TVillnqDkH	NP_000974	P35268	RL22_HUMAN	[RPL22]	222319	2	

Left-click and drag over these table entries to select them with the mosue.

The right-click on one of the hi-lited entries and choose “Select from table”.

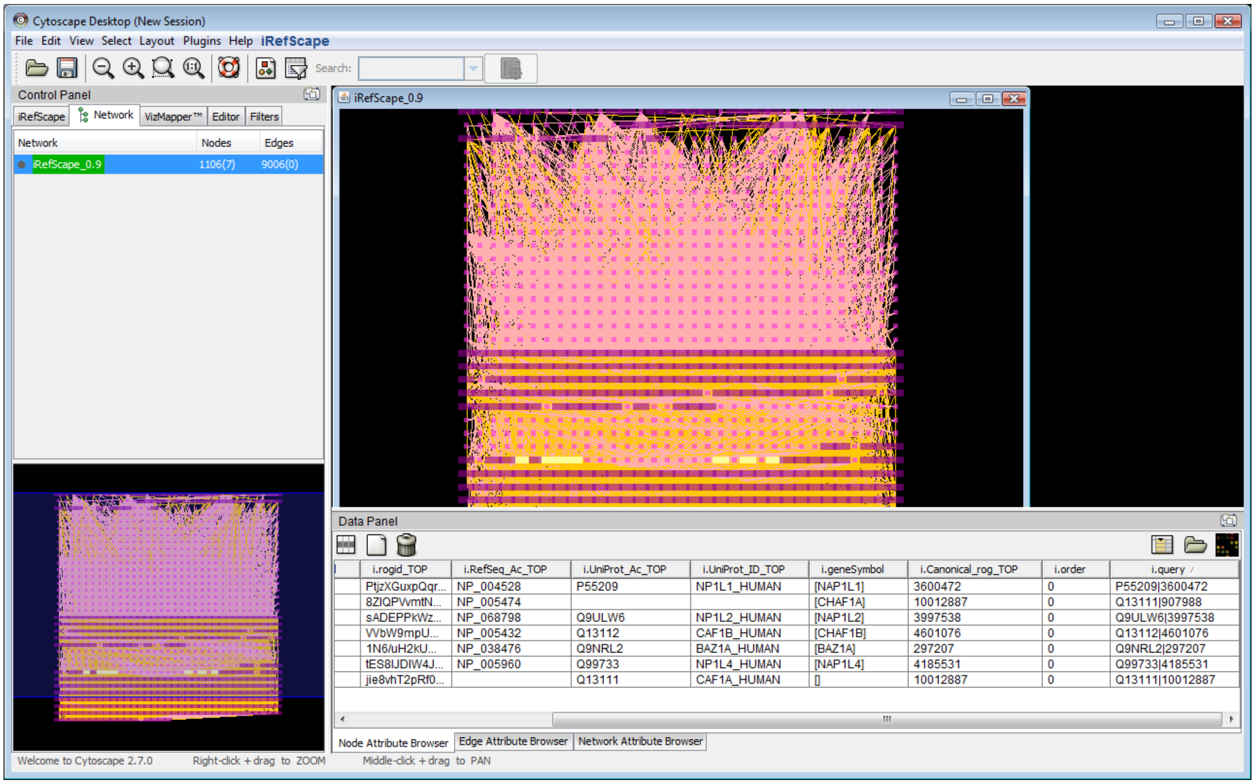

Cytoscape Desktop (New Session)
File Edit View Select Layout Plugins Help iRefScape
Search:
Control Panel
iRefScape
Network
VizMapper™
Editor
Filters
iRefScape_0.9
Data Panel
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser
Welcome to Cytoscape 2.7.0
Right-click + drag to ZOOM
Middle-click + drag to PAN

Network	Nodes	Edges
iRefScape_0.9	1106(1106)	9006(0)

i.rogid_TOP	i.RefSeq_Ac_TOP	i.UniProt_Ac_TOP	i.UniProt_ID_TOP	i.geneSymbol	i.Canonical_rog_TOP	i.order	i.query
sADEPPkWz...	NP_068798	Q9ULW6	NP1L2_HUMAN	[NAP1L2]	3997538	0	Q9ULW6|3997538
1N6/uH2kU...	NP_038476	Q9NRL2	BAZ1A_HUMAN	[BAZ1A]	297207	0	Q9NRL2|297207
tES8IJDIW4J...	NP_005960	Q99733	NP1L4_HUMAN	[NAP1L4]	4185531	0	Q99733|4185531
VVbW9mpU...	NP_005432	Q13112	CAF1B_HUMAN	[CHAF1B]	4601076	0	Q13112|4601076
8ZIQPVvmtN...	NP_005474			[CHAF1A]	10012887	0	Q13111|907988
jie8vhT2pRf0...		Q13111	CAF1A_HUMAN	[]	10012887	0	Q13111|10012887
PtjzXGuxpQqr...	NP_004528	P55209	NP1L1_HUMAN	[NAP1L1]	3600472	0	P55209|3600472
				[]		-1	
0TVillngDkH...	NP_000974	P35268	RL22_HUMAN	[RPL22]	232319	2	

i.query
P55209|3600472

After this operation, only the original nodes directly returned by the query will apeear in the node attribute browser.


Cytoscape Desktop (New Session)
File Edit View Select Layout Plugins Help iRefScape
Search:
Control Panel
iRefScape
Network
VizMapper™
Editor
Filters
Network
Nodes
Edges
iRefScape_0.9
1106(7)
9006(0)
iRefScape_0.9
Data Panel
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser
Welcome to Cytoscape 2.7.0
Right-click + drag to ZOOM
Middle-click + drag to PAN

i.rogid_TOP	i.RefSeq_Ac_TOP	i.UniProt_Ac_TOP	i.UniProt_ID_TOP	i.geneSymbol	i.Canonical_rog_TOP	i.order	i.query
PtjzXGuxpQqr...	NP_004528	P55209	NP1L1_HUMAN	[NAP1L1]	3600472	0	P55209|3600472
8ZIQPVvmtN...	NP_005474			[CHAF1A]	10012887	0	Q13111|907988
sADEPPkWz...	NP_068798	Q9ULW6	NP1L2_HUMAN	[NAP1L2]	3997538	0	Q9ULW6|3997538
VVbW9mpU...	NP_005432	Q13112	CAF1B_HUMAN	[CHAF1B]	4601076	0	Q13112|4601076
1N6/uH2kU...	NP_038476	Q9NRL2	BAZ1A_HUMAN	[BAZ1A]	297207	0	Q9NRL2|297207
tES8IJDIW4J...	NP_005960	Q99733	NP1L4_HUMAN	[NAP1L4]	4185531	0	Q99733|4185531
jie8vhT2pRf0...		Q13111	CAF1A_HUMAN	[]	10012887	0	Q13111|10012887

Then go to the iRefScape menu/View tools/Select between nodes.

Use the node attribute browser to select all nodes that are either

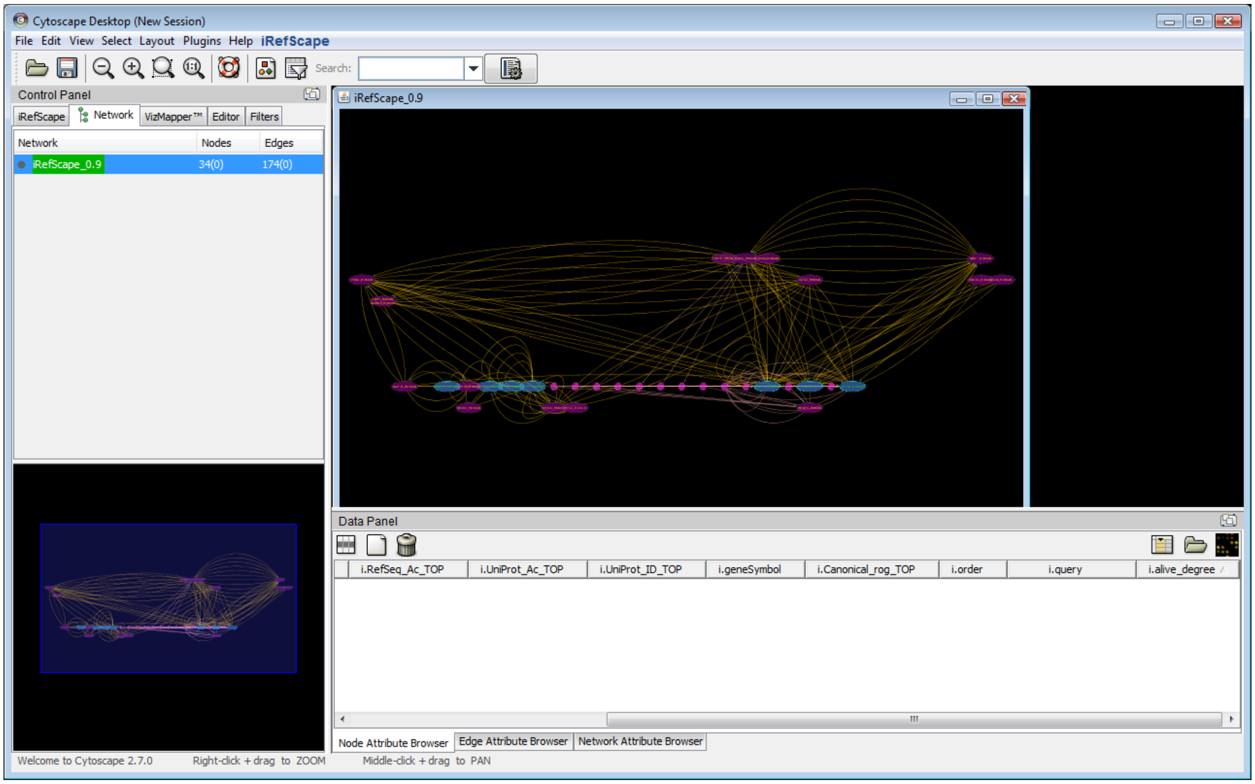
1) Returned by the original query (see i.query node feature) or
2) Have an i.alive_degree of 2 or more (see the i.alive_degree node feature)

This may require a live demo. Ask if you have problems.

Then select all other nodes and delete them. Use Select/Nodes/Invert node selection

And then hit delete.

You should see something like this.



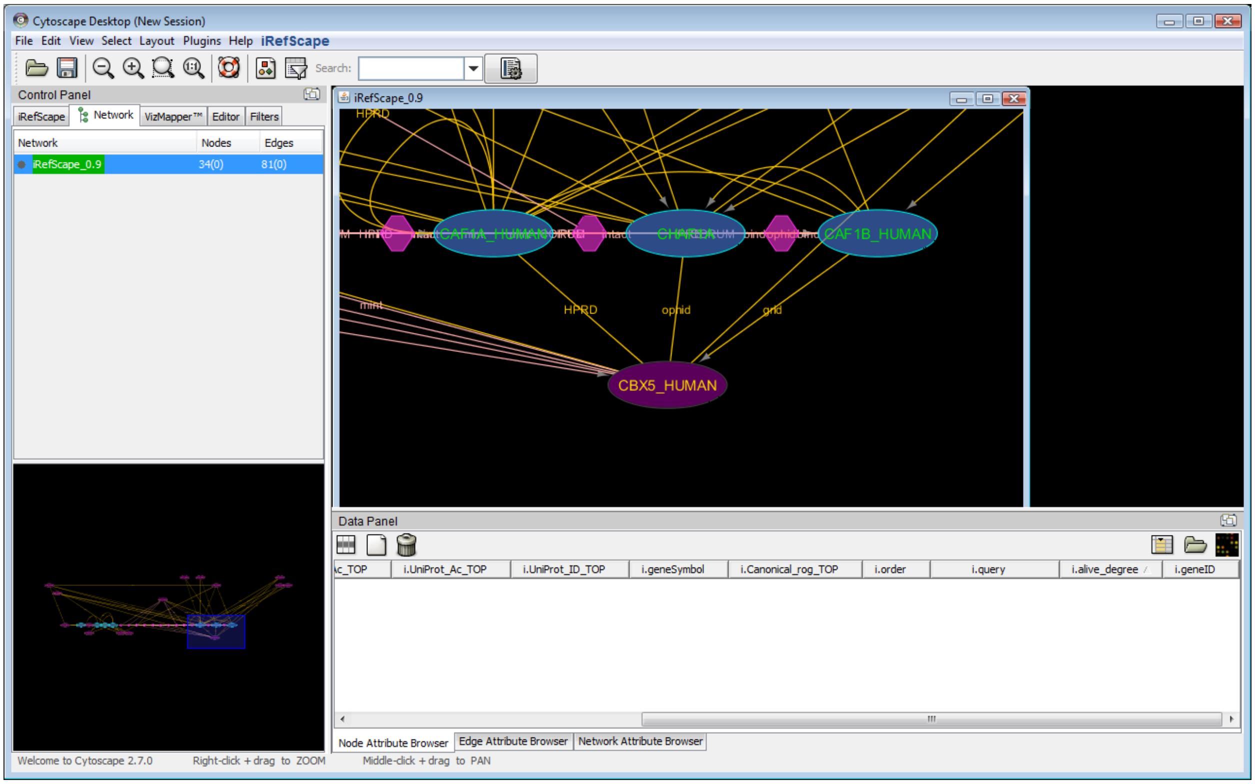
Network	Nodes	Edges
iRefScape_0.9	34(0)	174(0)

i.RefSeq_Ac_TOP	i.UniProt_Ac_TOP	i.UniProt_ID_TOP	i.geneSymbol	i.Canonical_rog_TOP	i.order	i.query	i.alive_degree

Welcome to Cytoscape 2.7.0 Right-click + drag to ZOOM Middle-click + drag to PAN

These new “between nodes” are connected to two or more nodes from your original search list (seed list). Guilt-by-association makes these nodes candidates to look at for things that may be related to the “chromatin assembly complex”. The higher the i.alive_degree, the better.

For example, CBX5_HUMAN interacts with CAF1A, CAF1B and CHAF1A from our original search.



A brief review of the Entrez Gene record for CBX5 (Entrez Gene ID 23468) shows that there is evidence to support this connection. It is left as an exercise to review the papers that support evidence of an interaction between CBX5 and each of the three original query proteins.

If you have time left, you can use the techniques described in this tutorial to investigate another GO term and its associated genes that are of specific interest to you.