

GO, DAVID and ORA.

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.

Go to <http://geneontology.org/>

The Gene Ontology - Mozilla Firefox

File Edit View History Bookmarks Tools Help

GO http://geneontology.org/

Most Visited Getting Started Latest Headlines Nature Biotechnology

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

Done

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.

AmiGO: Tree Browser - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://amigo.geneontology.org/cgi-bin/amigo/browse.cgi

Most Visited Getting Started Latest Headlines Nature Biotechnology

AmiGO: Tree Browser Bugzilla Main Page

the Gene Ontology AmiGO

Search Browse BLAST More Tools Help

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

Done

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:0005575) 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 can't 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.

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=

Most Visited Getting Started Latest Headlines Nature Biotechnology

AmiGO: Tree Browser Bugzilla Main Page

biological process
cellular component
molecular function

ASAP
AspGD
CGD

Arabidopsis thaliana
Bacillus anthraci...
Bacillus subtilis

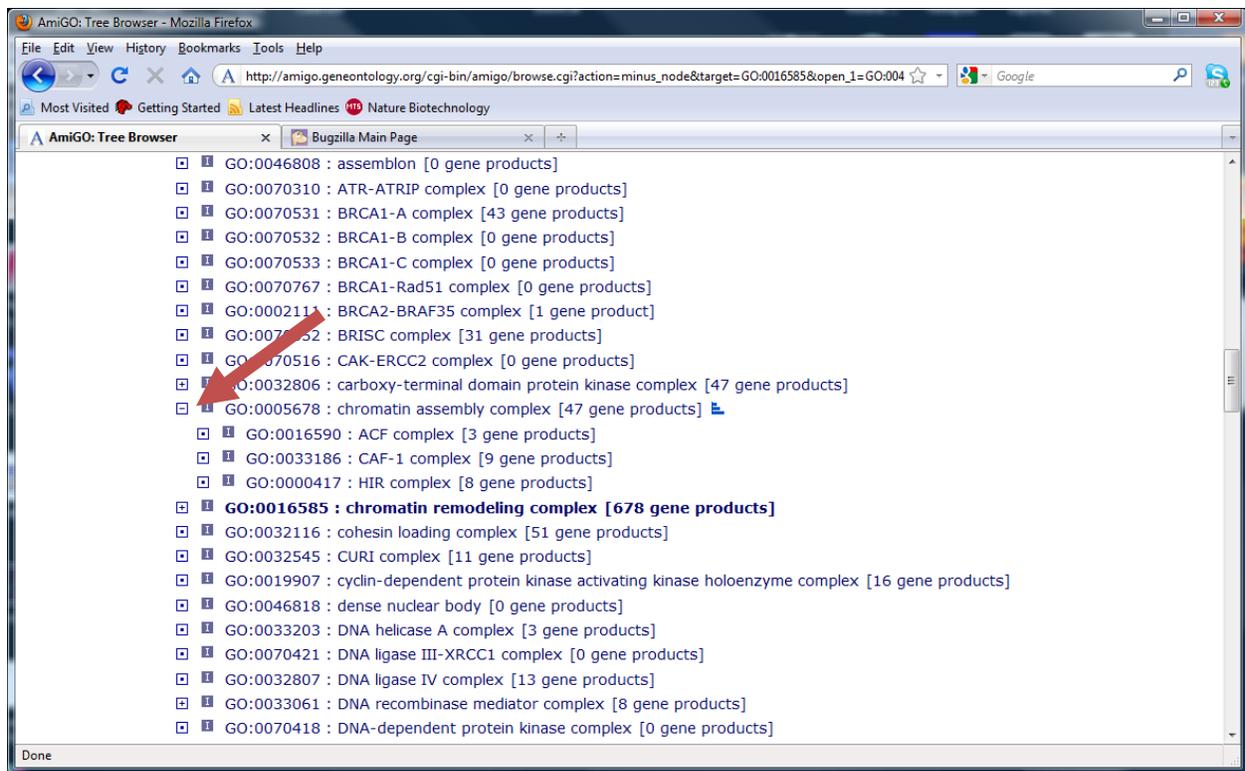
all : all [407301 gene products]

- 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:0031977 : membrane-enclosed lumen [10292 gene products]
 - GO:003226 : organelle [121389 gene products]
 - GO:0043230 : extracellular organelle [39 gene products]
 - GO:0043231 : 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]

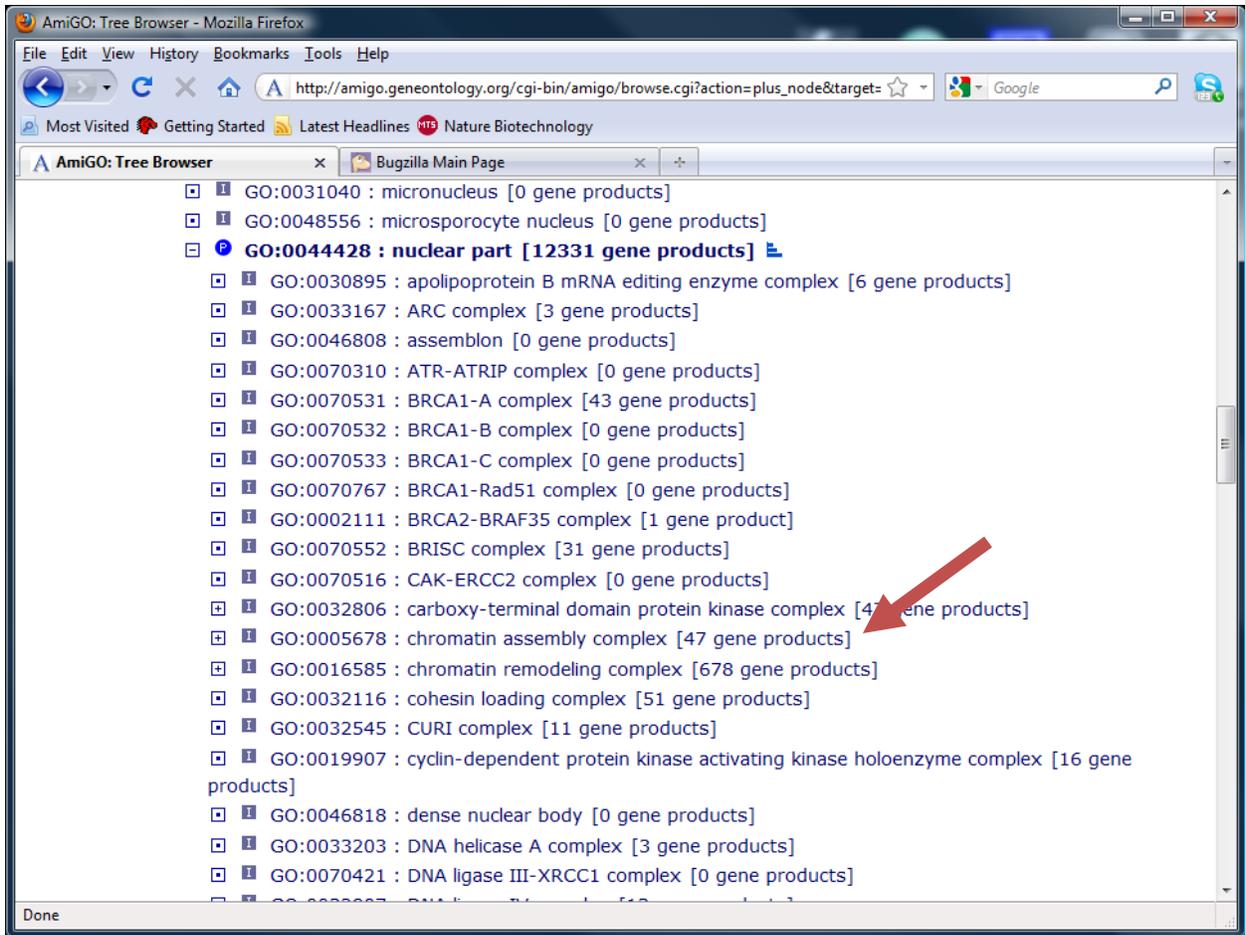
Actions...
Last action: Opened
GO:0043231
Graphical View
Permalink
Download...
OBO
RDF-XML
GraphViz dot

Done

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).



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



Click on the "47 gene products" to view them.

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; Evidence Code: All, IC, IDA, IEA

View associations: All Direct associations

1 2 [View all results](#)

chromatin assembly complex ; GO:0005678 [\[show def\]](#) [\[view in tree\]](#)

Symbol, full name	Information	Qualifier	Evidence	Reference	Assigned by
<input type="checkbox"/> Acf1 ATP-dependent chromatin assembly factor large subunit	21 associations BLAST protein from <i>Drosophila melanogaster</i>		NAS	FB:FBrf0105495	FlyBase
<input type="checkbox"/> AN4891	13 associations gene_product from <i>Emericella nidulans</i>		IEA With SGD:5000003651	AspGD REF:ASPL0000000005	ASPGD (via AspGD)
<input type="checkbox"/> AN8039	6 associations gene_product from <i>Emericella nidulans</i>		IEA With SGD:5000005372	AspGD REF:ASPL0000000005	ASPGD (via AspGD)
<input type="checkbox"/> 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

http://amigo.geneontology.org/cgi-bin/amigo/term-assoc.cgi?gptype=all&speciesdb=all&taxid=9606&evcode=all&t

Search GO terms genes or proteins exact match Submit Query

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 sulfur...
gene product	AspGD	Haloarcula marism...
mirna	CGD	Homo sapiens

Filter by Association

Evidence Code

All Direct associations

Set filters Remove all filters

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

Symbol, full name	Information	Qualifier	Evidence	Reference	Assigned by
<input type="checkbox"/> CHAF1A	view associations protein from <i>Homo sapiens</i>		TAS	PMID:7600578	Proteome Inc.

Done

Mozilla Firefox

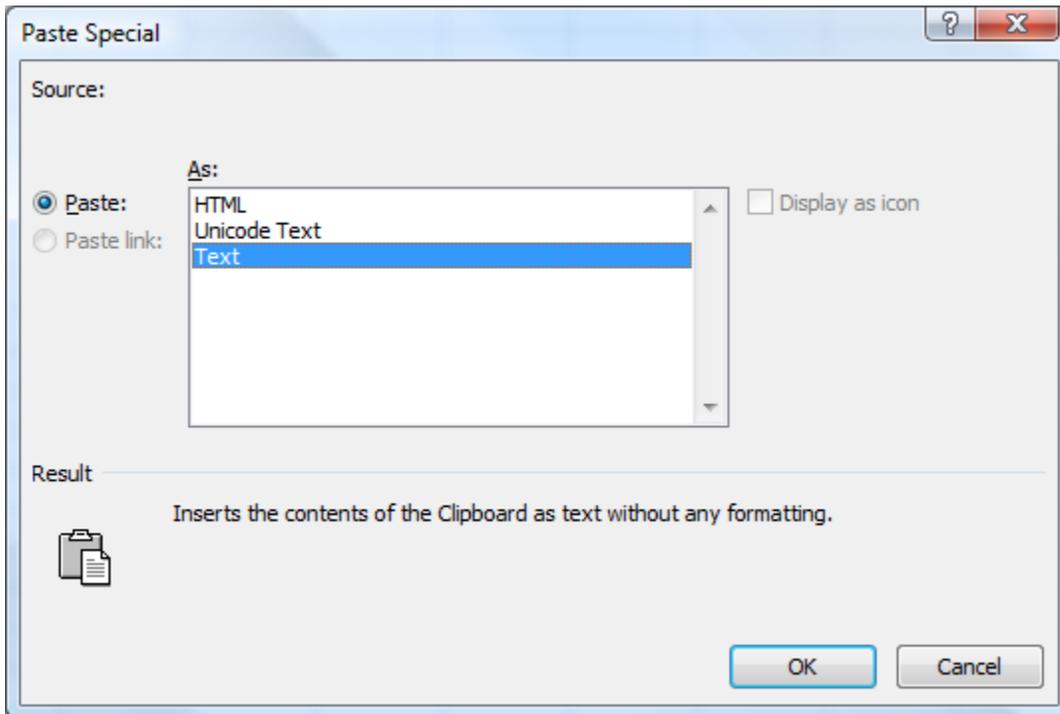
http://amigo.geneontology.org/cgi-bin/amigo/term-assoc.cgi?term=GO:0005678&format=go_assoc&session_id=968:

http://amigo.ge...migo1284654448 x Bugzilla Main Page

UniProtKB	Q13112	CHAF1B	GO:0005678	PMID:7600578	TAS	C	Chromatin assembly factor 1 subunit B	CAF1A CAF1B
UniProtKB	P55209	NAP1L1	GO:0005678	PMID:8297347	TAS	C	Nucleosome assembly protein 1-like 1	IFI000238
UniProtKB	Q99733	NAP1L4	GO:0005678	PMID:9325046	TAS	C	Nucleosome assembly protein 1-like 4	IFI009414
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 B2RE62
UniProtKB	Q99457	NAP1L3	GO:0005678	PMID:8976385	TAS	C	Nucleosome assembly protein 1-like 3	B2RCMO B2RCMP
UniProtKB	Q9NRL2	BAZ1A	GO:0016590	PMID:10662543	TAS	C	Bromodomain adjacent to zinc finger domain prote	

Done

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

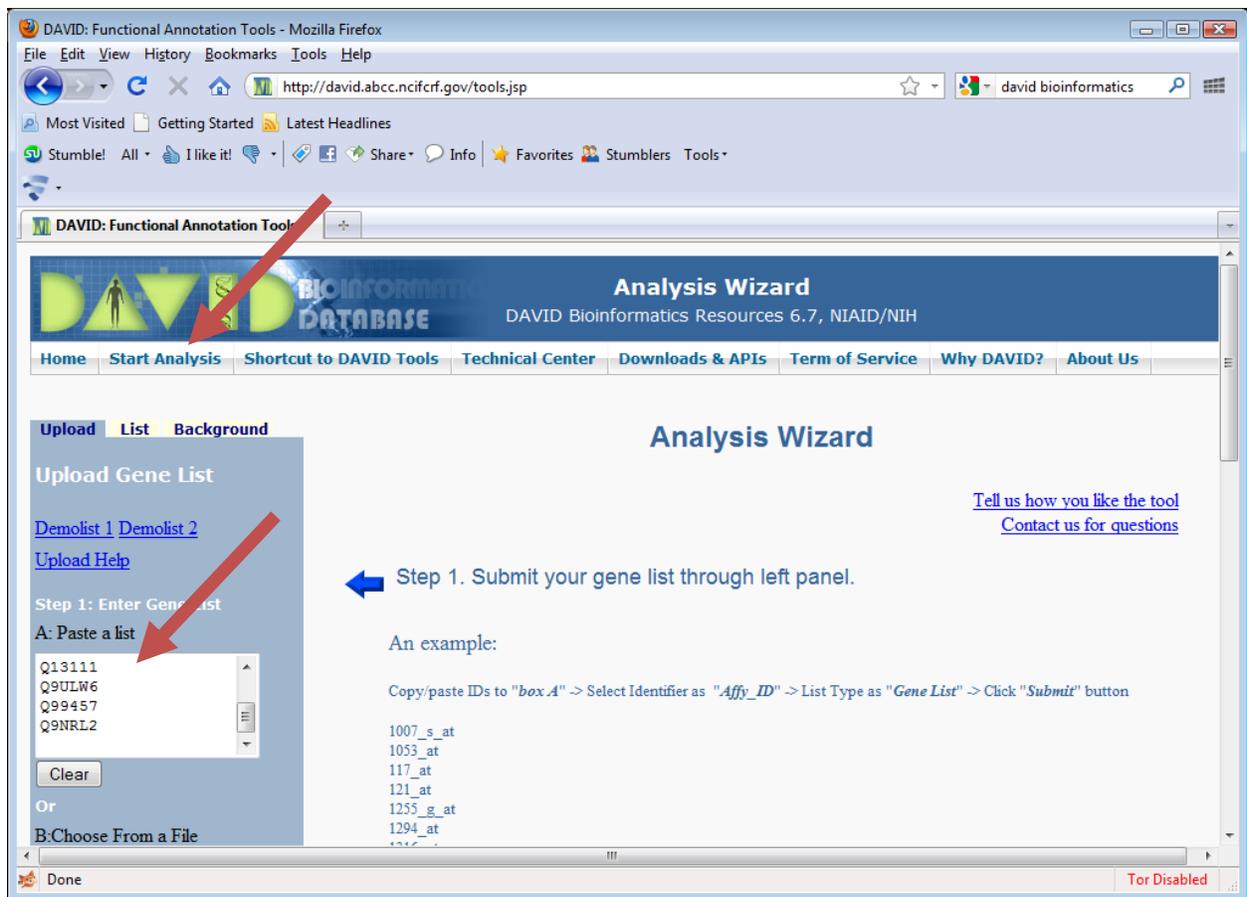


The screenshot shows a Microsoft Excel spreadsheet with the following data:

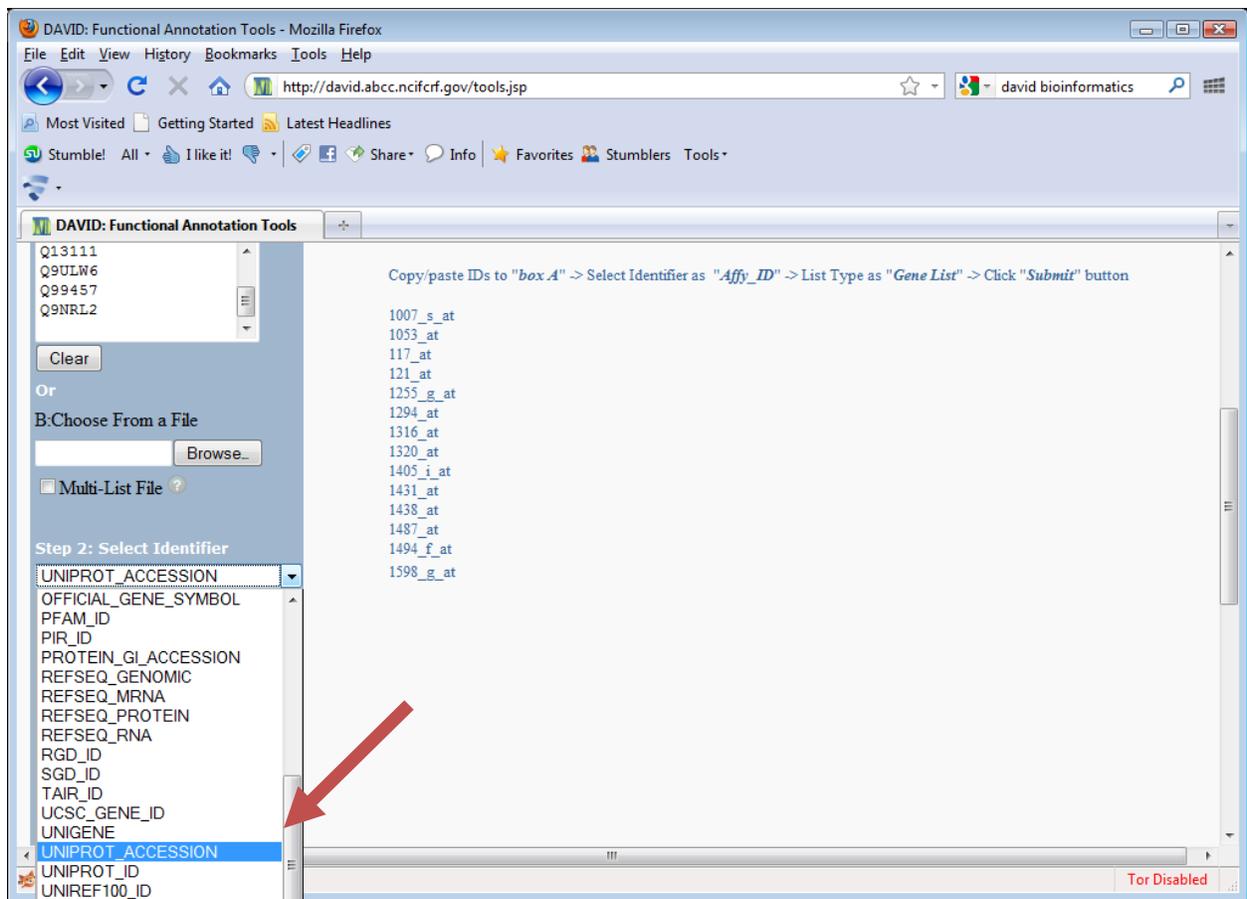
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	UniProtKE Q13112	CHAF1B		GO:00056; PMID:760; TAS					C	Chromatin CAF1A CA protein	taxon:960	20030904	PINC			
2	UniProtKE P55209	NAP1L1		GO:00056; PMID:829; TAS					C	Nucleosor IPI00238; protein	taxon:960	20030904	PINC			
3	UniProtKE Q99733	NAP1L4		GO:00056; PMID:932; TAS					C	Nucleosor IPI009414; protein	taxon:960	20030904	PINC			
4	UniProtKE Q13111	CHAF1A		GO:00056; PMID:760; TAS					C	Chromatin CAF CAF1 protein	taxon:960	20030904	PINC			
5	UniProtKE Q9ULW6	NAP1L2		GO:00056; PMID:878; TAS					C	Nucleosor B2RE61 Bi protein	taxon:960	20030904	PINC			
6	UniProtKE Q99457	NAP1L3		GO:00056; PMID:897; TAS					C	Nucleosor 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			
8																
9																
10																
11																
12																

Select and copy the UniProt accessions from column B.

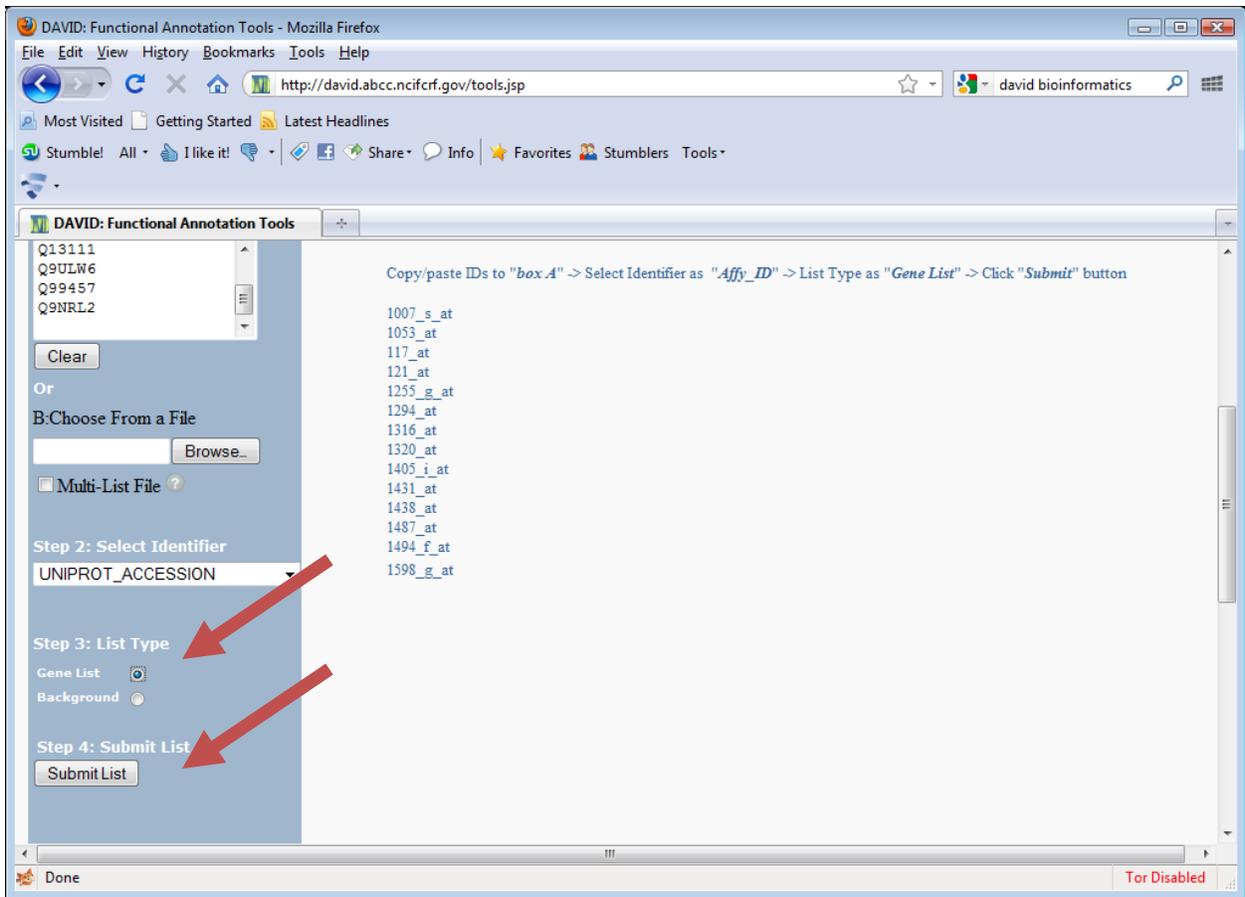
Go to <http://david.abcc.ncifcrf.gov/home.jsp> , click on Start Analysis and then paste the list of accessions into the query box like this:



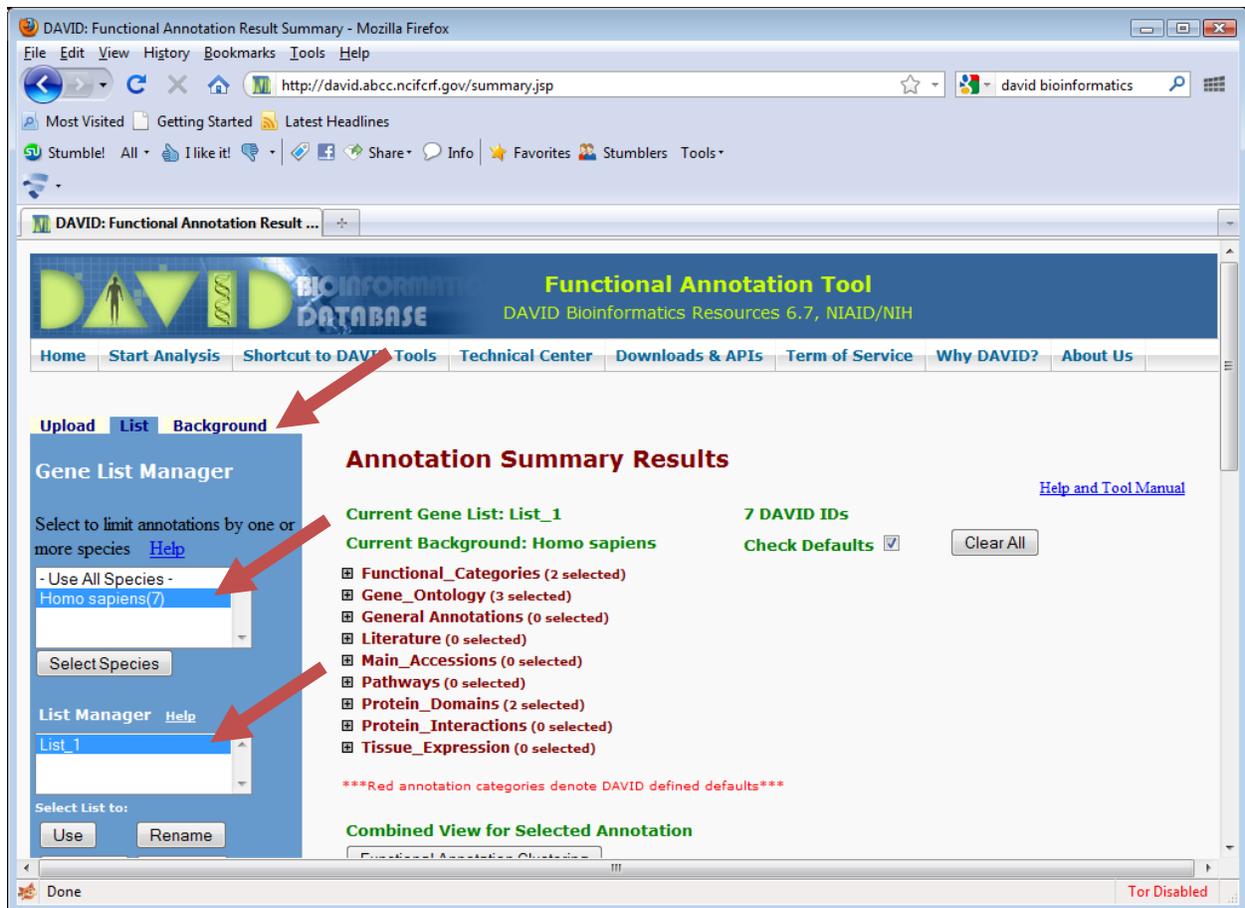
The tell DAVID that you have entered UniProt Accessions. Like this



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



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:



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

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

Current Background: Homo sapiens

Functional_Categories (2 selected)

Gene_Ontology (3 selected)

Category	Percentage	Count	Chart
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
<input checked="" type="checkbox"/> 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
<input checked="" type="checkbox"/> 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 and Infectious Diseases (NIAID), NIH)

http://david.abcc.ncifcrf.gov/chartReport.jsp?annot=33

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Chart

[Help and Manual](#)

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

Rerun Using Options Create Sublist

1 chart records [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	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.









[Term of Service](#) | [Contact Us](#) | [Site Map](#)

Done Tor Disabled

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>