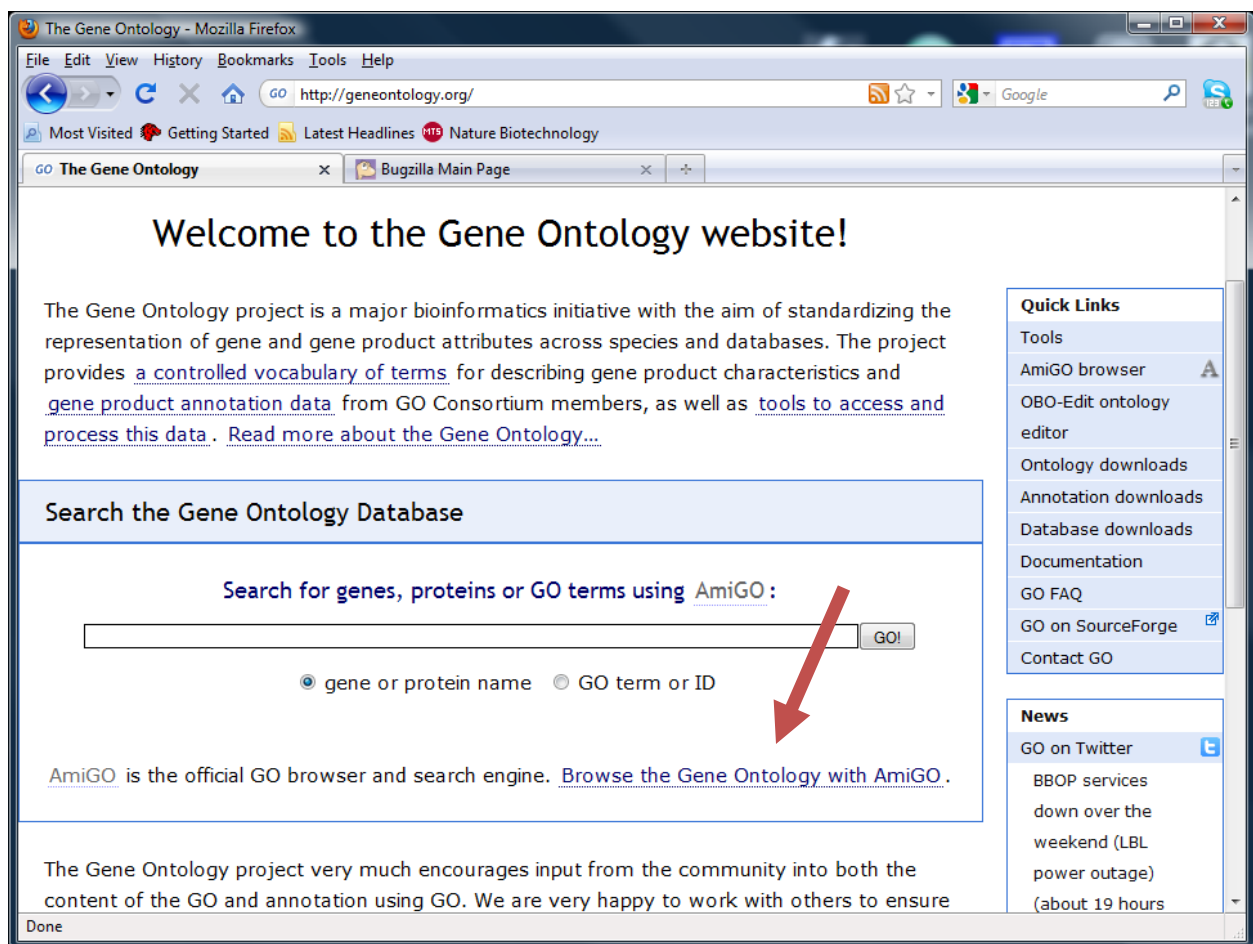# GO, DAVID and ORA.

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/

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.
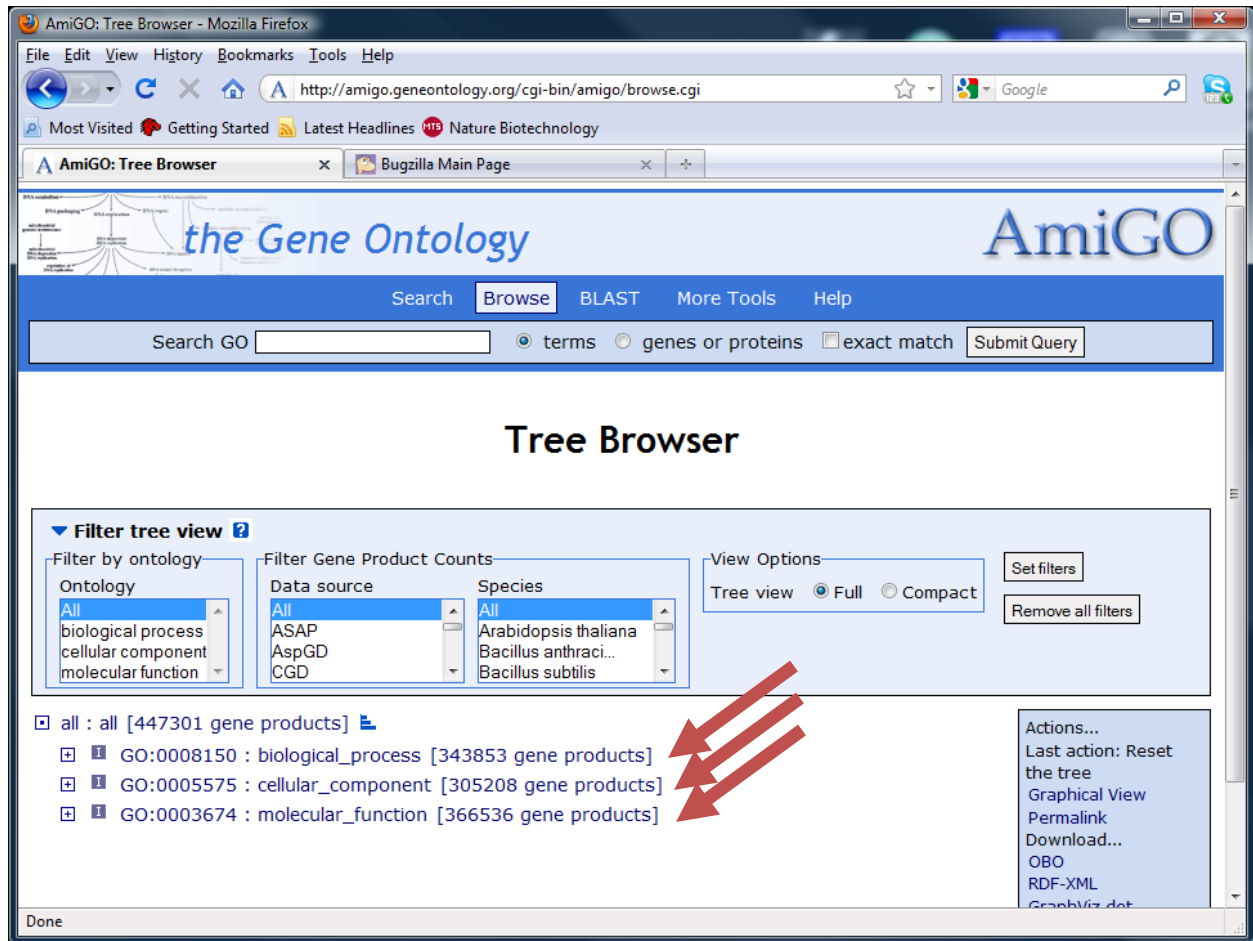

Try to make your way to the term that describes "nucleus" without directly searching for it. Instead, start by expanding the cellular_component "root node" (GO:0005755) by clicking on the + beside it. Look for the next closest thing to "nucleus" in the expanded list and then click on that. If you want to look at the definition of any given term, just click on it.
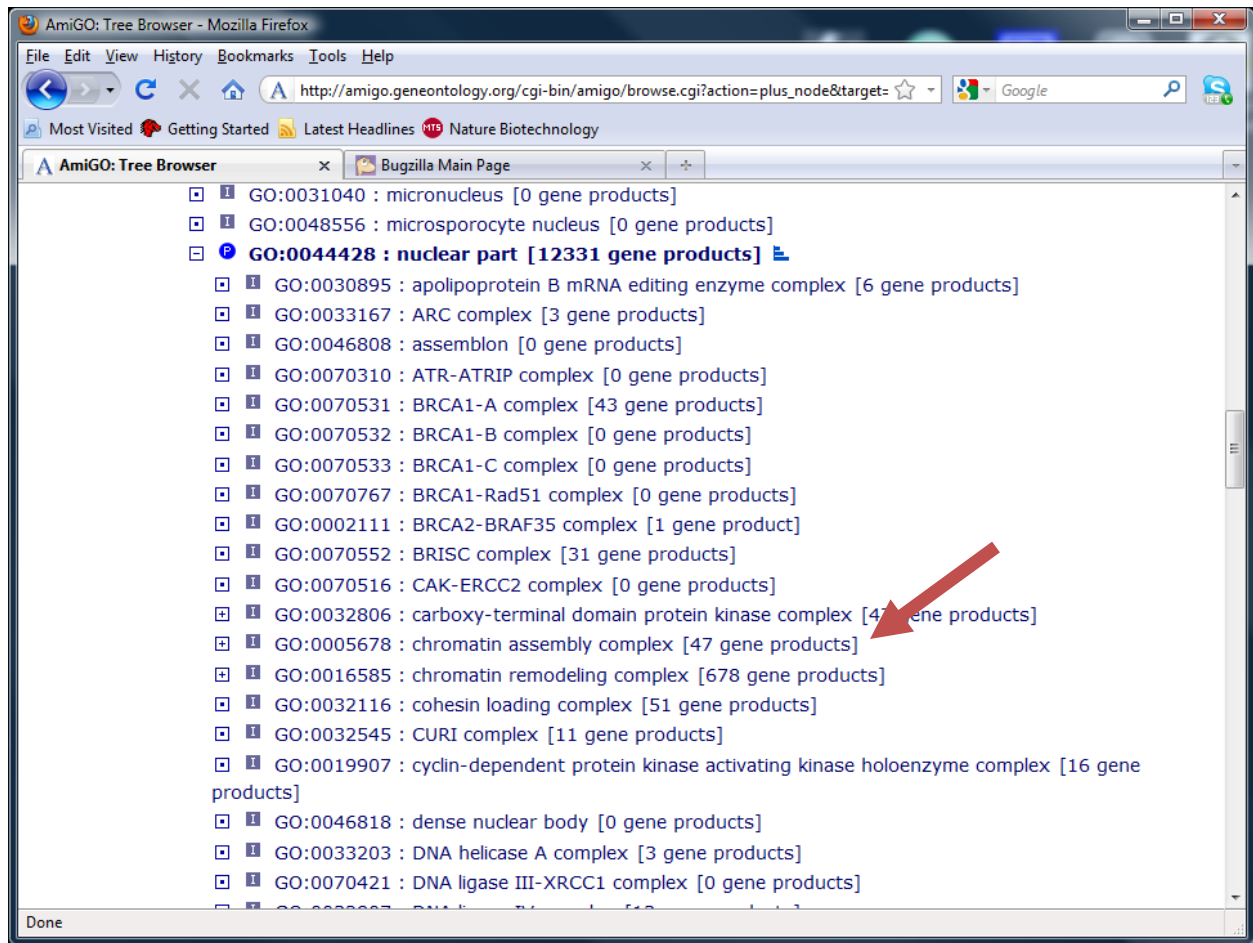
If you cant find the term entry for "nucleus", use the next page as a hint or search for nucleus in the "Search GO" box at the top of the interface.
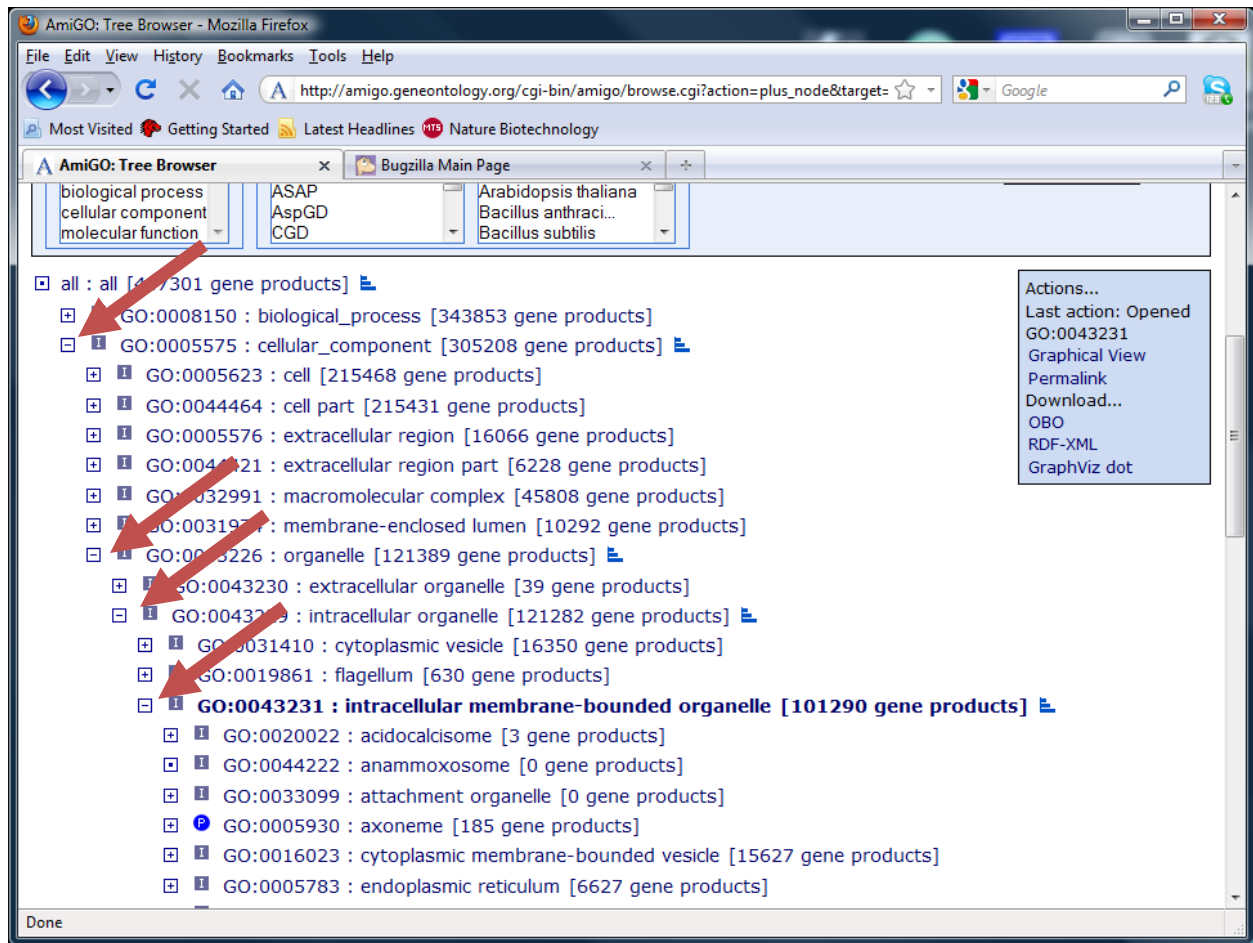
Once you make it to the nucleus, keep navigating down the tree to chromatin assembly complex (GO:0005678). This term is actually used to refer to a number of complexes (try expanding the term). What are these complexes? Note that all of them are "leaf" nodes (you cant expand them any further).
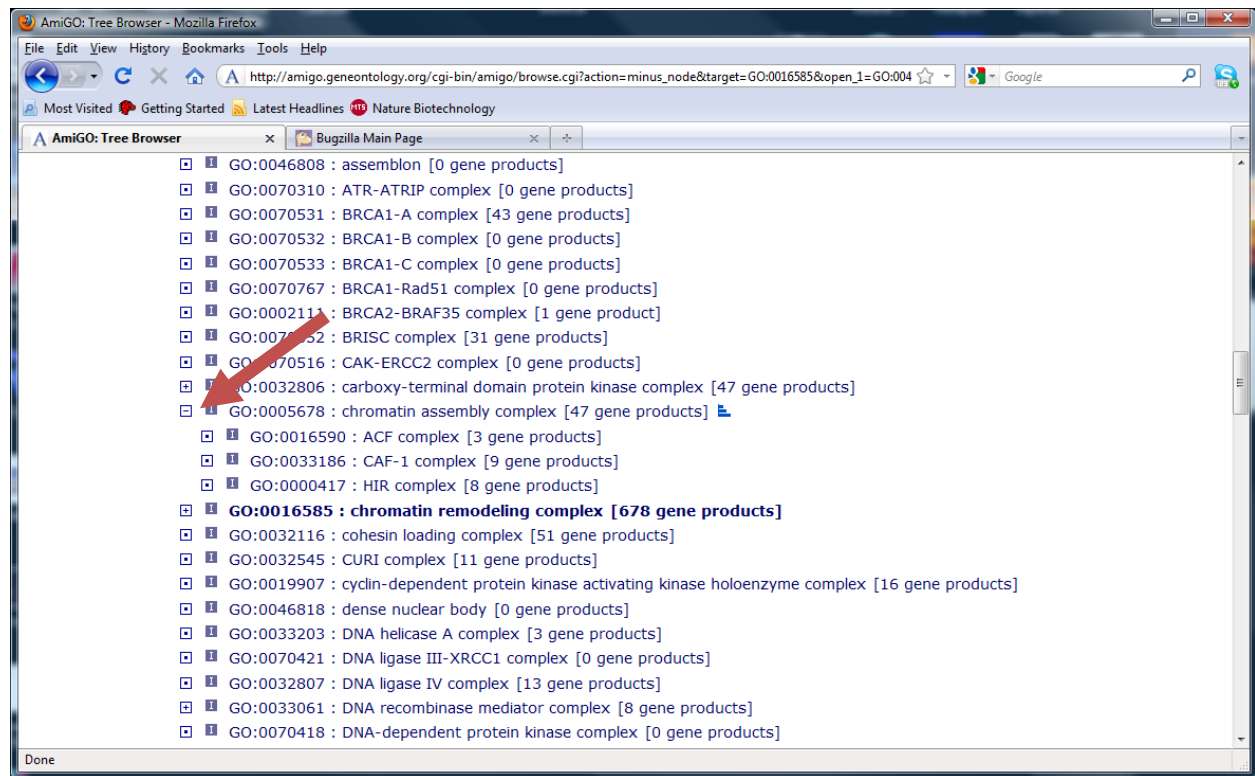
Now go back and click on the 47 gene products that correspond to the "chromatin assembly complex".

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

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".

Screenshot 1 — AmiGO: Term Association Details - Mozilla Firefox

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

Filter by Association
Evidence Code
All / IC / IDA / IEA

View associations
⦿ All ○ Direct associations

[Set filters]
[Remove all filters]

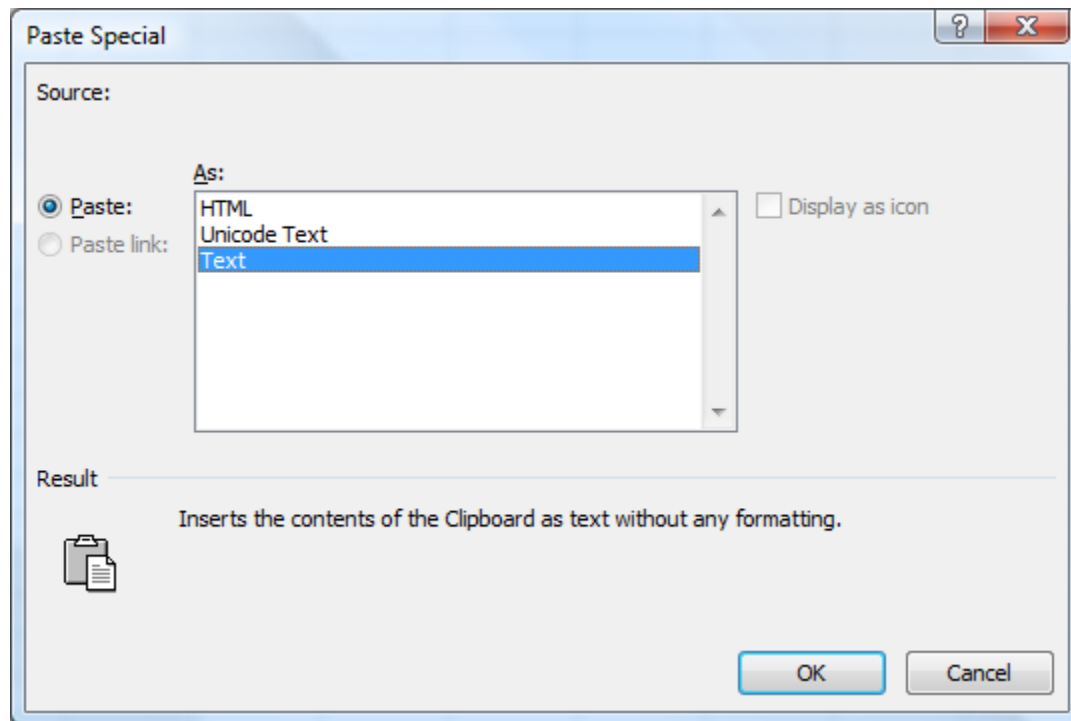**chromatin assembly complex ; GO:0005678** [show def] [view in tree]

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

Done

Screenshot 2 — Mozilla Firefox (gene association format text)

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

UniProtKB	Q13112	CHAF1B	GO:0005678	PMID:7600578	TAS	C	Chromatin assembly factor 1 subunit B	CAF1A|CAF
UniProtKB	P55209	NAP1L1	GO:0005678	PMID:8297347	TAS	C	Nucleosome assembly protein 1-like 1	IPI000238
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	

Done

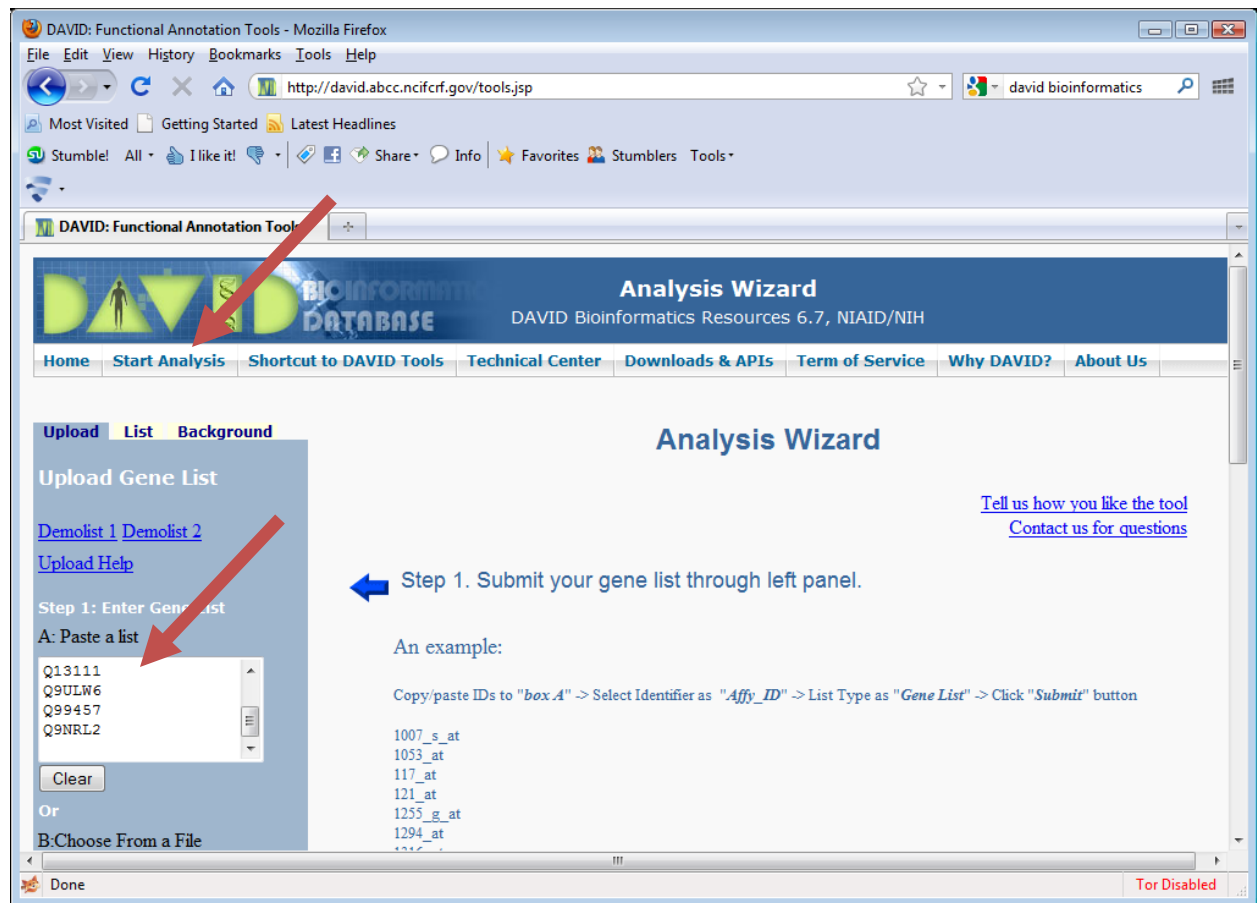Copy and paste this to an Excel spread sheet (or similar). Right-click, paste special, as text.

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:
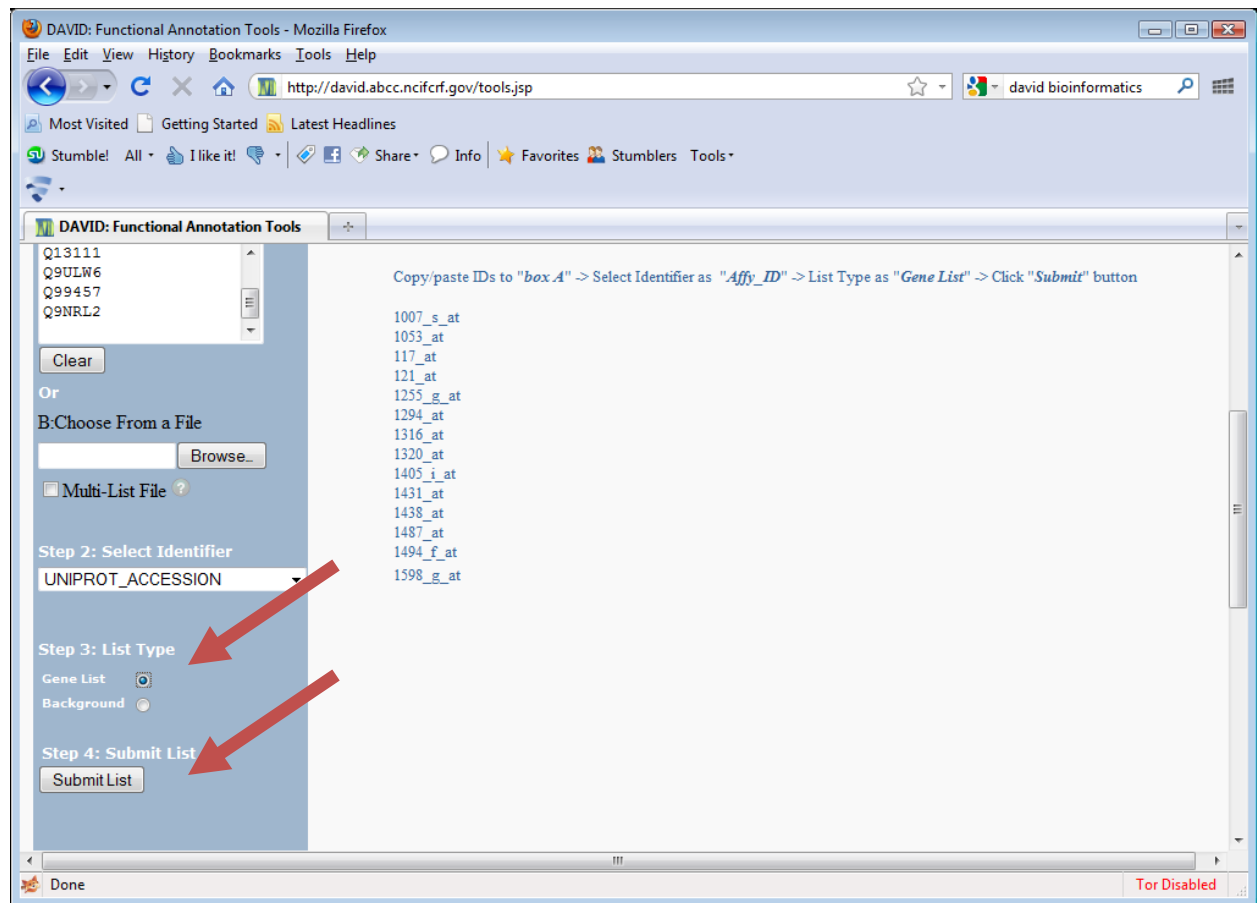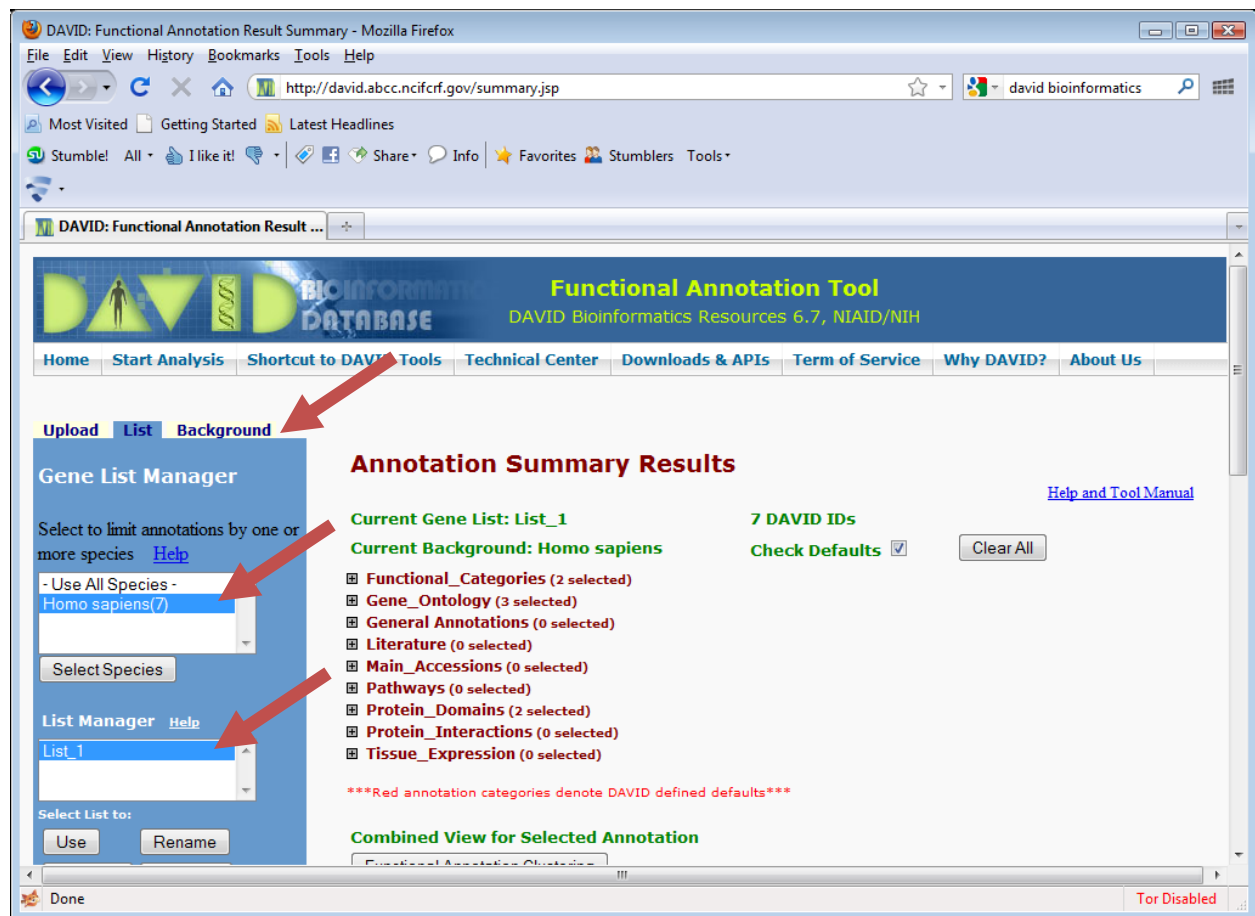
The tell DAVID that you have entered UniProt Accessions. Like this
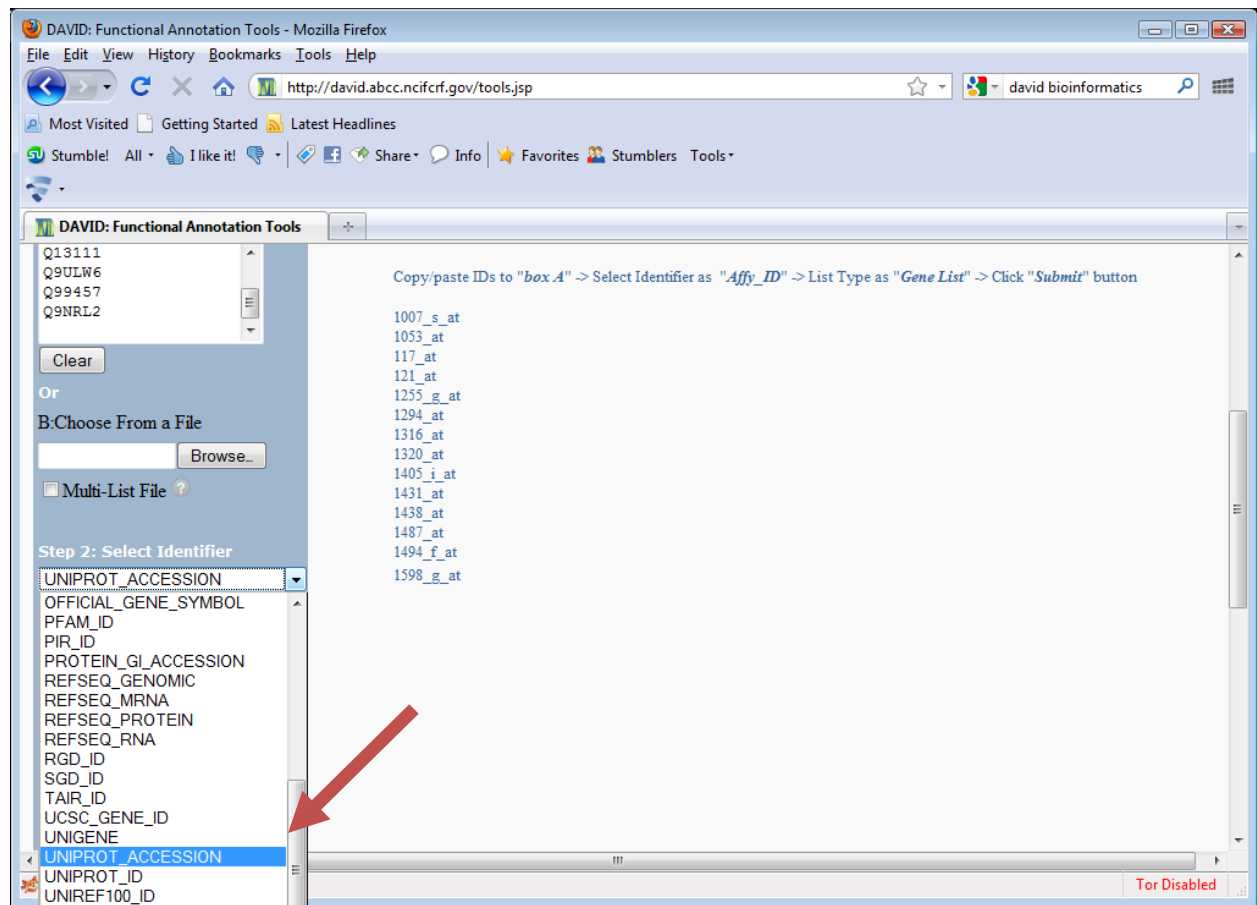
Then select "Gene List" under List type and click on Submit list:

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:
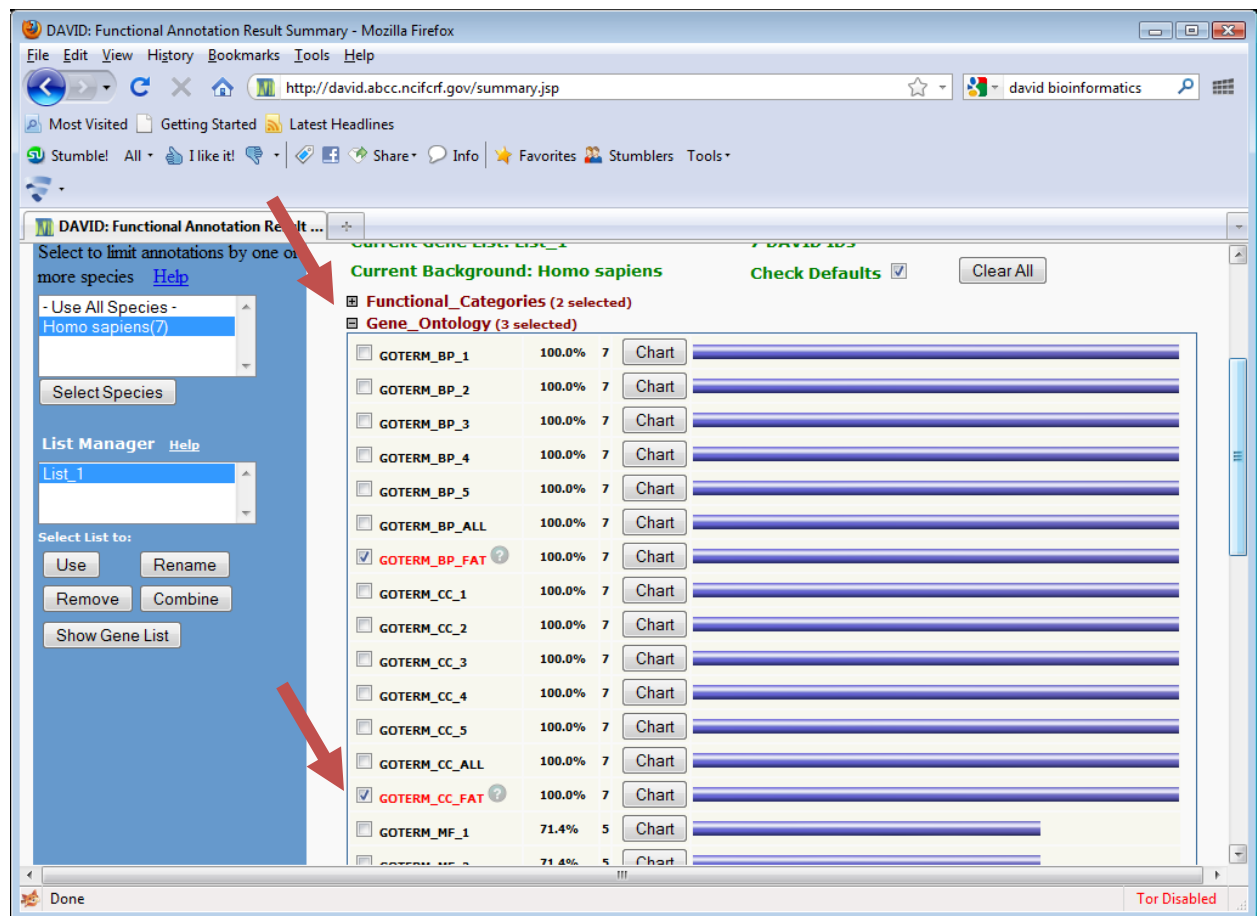
If you then click on the "Chart" button, you see a new window open like this:
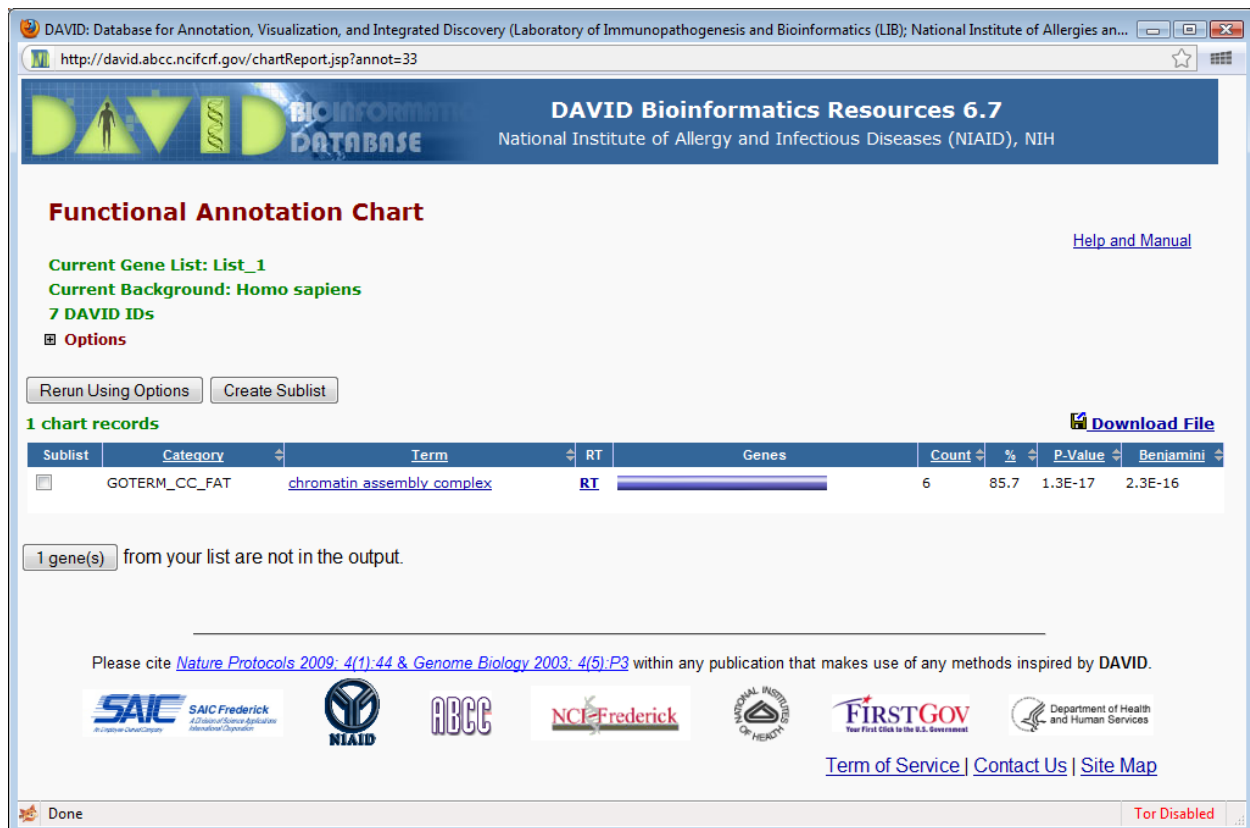
6 out of the 7 genes in your list (85.7%) are annotated with the GO term for chromatin assembly complex. The probability of randomly choosing 7 genes from the human genome where 6 of them all have this associated GO term is 1.3E.17. When you correct for multiple hypothesis testing (Benjamini), the probability 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