Introduction

You have a list of genes differentially expressed in skeletal muscles of young DMD patients and controls. The following two exercises are meant to help you interpret long lists of differentially expressed genes in terms of biological function.
The first exercise uses DAVID, one of the most comprehensive tools to test for functional categories that are overrepresented in a list of differentially expressed genes. It uses annotation from curated databases like GO, KEGG, InterPro and OMIM.
The second exercise uses Anni, a tool that mines the biomedical literature to derive information about genes.

Please split up in two groups; the first group starts off with DAVID, the second group with Anni. This to avoid server time-outs.

Entrez gene IDs work well for almost all annotation programs. The list you have has been converted from a list of differentially expressed genes using David: http://david.abcc.ncifcrf.gov/conversion.jsp

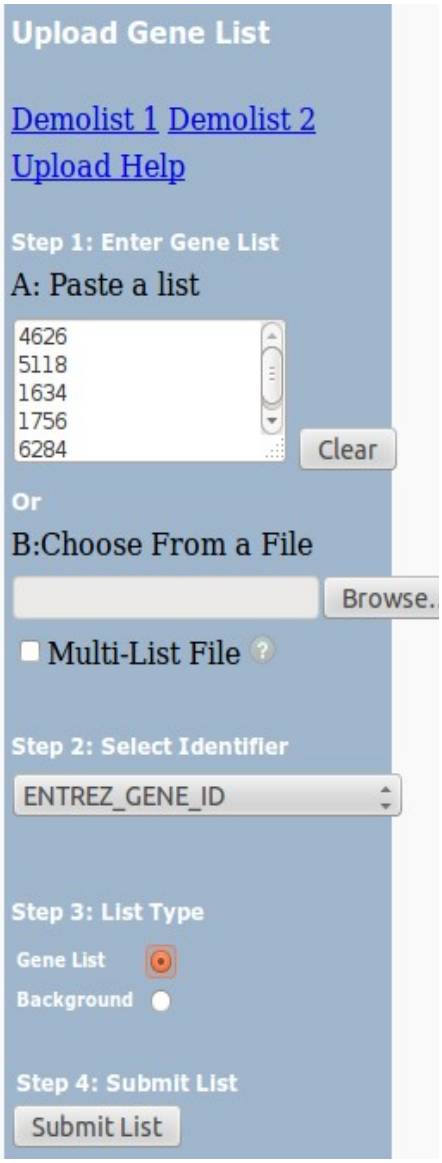
Exercise 1: Enrichment analysis of biological pathways with DAVID

DAVID compares a list of selected genes versus all genes (in an organism or on an array) and asks whether your selected list is enriched for genes belonging to a specific pathway.

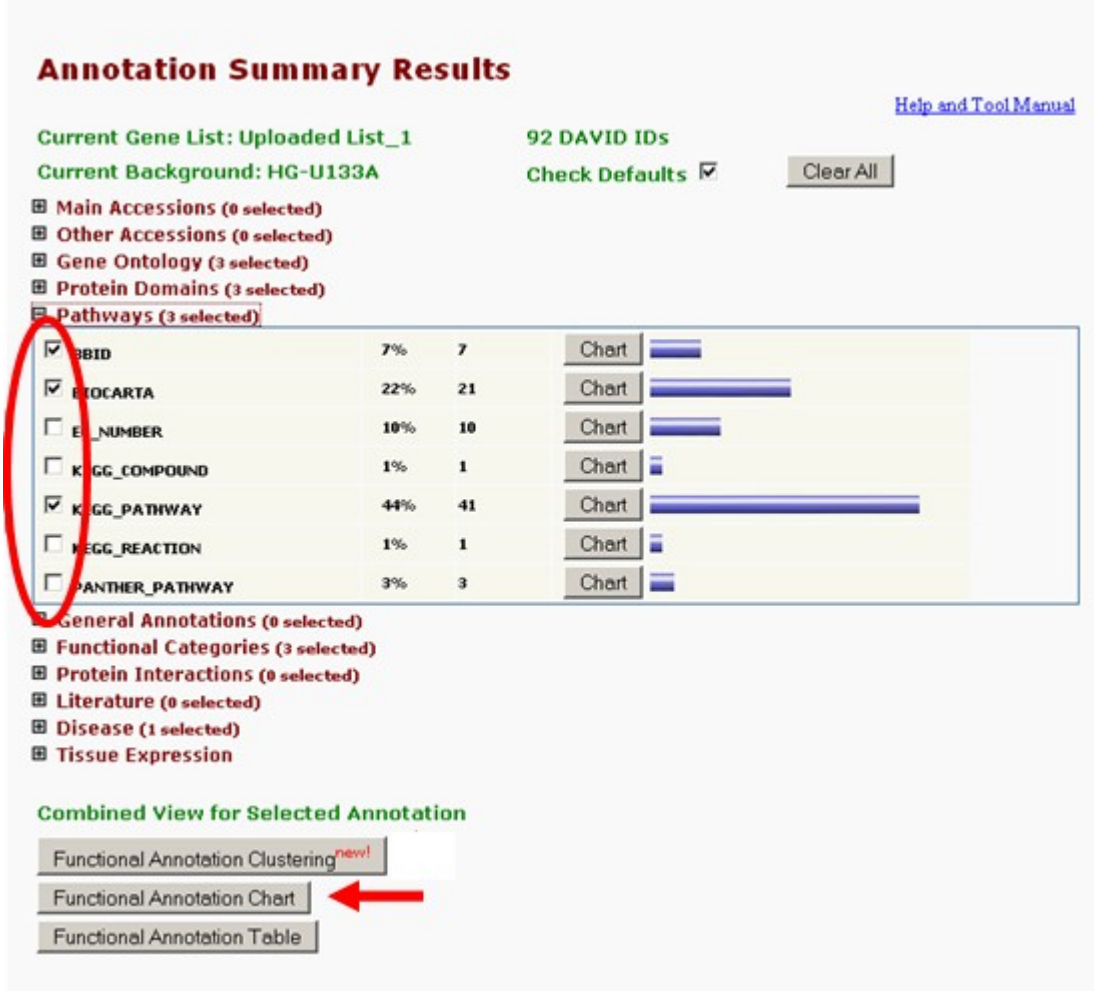
1. Go to DAVID's Functional Annotation tool at http://david.abcc.ncifcrf.gov/summary.jsp

2. Start with uploading your Entrez gene identifiers , using either copy-paste or the "Choose From a File" option to upload a tab-delimited text file containing identifiers.

3. Make a background list of all the genes represented on your platform. Under tab "Background" background lists from all Affymetrix and Illumina platforms are already available. The background appropriate here is "Human Genome U133A Array".

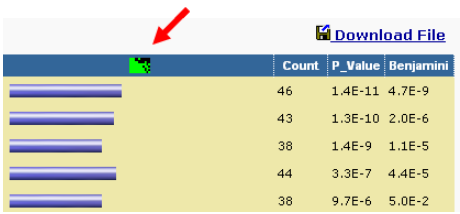


4. Select the databases that you want to include for functional annotation and select “Functional Annotation Chart”. You will obtain a table giving the data source where the annotation comes from, the term itself, the number of genes annotated with that term (you will be able to see the genes by clicking on the bar), and the enrichment p-value (hypergeometric test, uncorrected and corrected for multiple (pathway) testing by the Benjamini-Hochberg method.



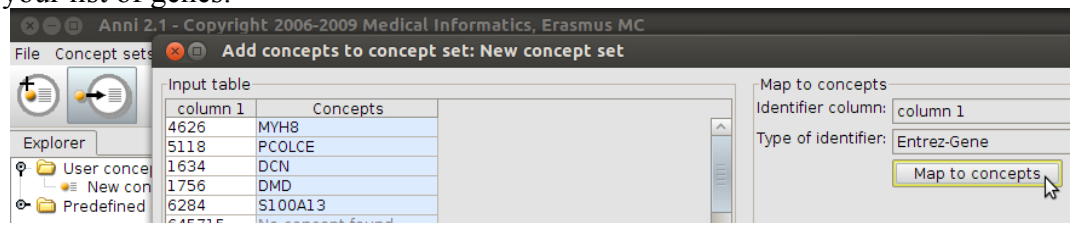
In the default setting, the most informative sources are automatically selected. To compare your results with the results from Exercise 2, perform the functional annotation chart again but now only select the Gene Ontology GOTERM_BP_ALL.

5. There is a lot of redundancy (overlap) in the annotations. By choosing the “Functional Annotation Clustering” function in the window above, the annotations represented by highly similar sets of genes are grouped. By hitting the clustering icon in the result table, you will be able to see which genes are annotated with which terms.

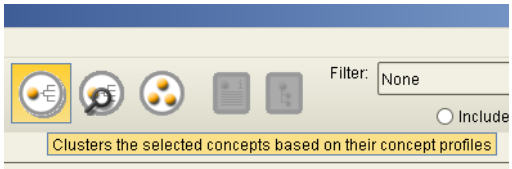


Exercise 2: Annotation of biological functions using Anni

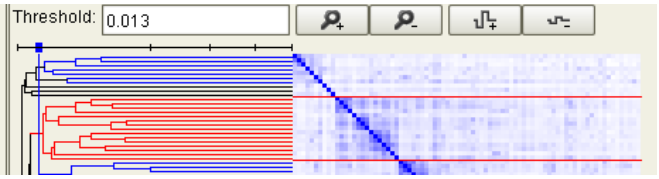
1. Go to http://www.biosemantics.org/index.php?page=anni-2-0.
2. Optional: use the Anni tutorial to get acquainted with the program.
3. Load your list of genes.

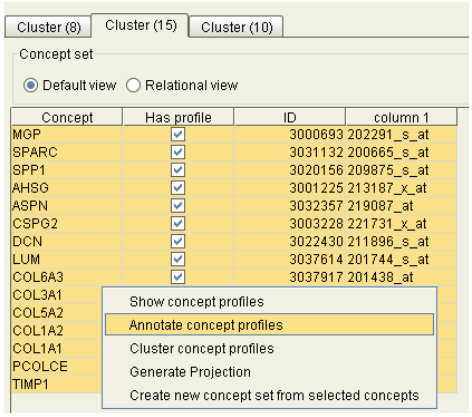


4. How many of the candidate genes have concept profiles? Why do not all genes have concept profiles?
5. Select all genes and cluster genes based on their literature similarity.

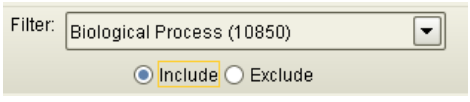


6. Select a few subclusters by clicking on the tree, select the genes in the panel below and right-click for "Annotation" to find the terms associated with the selected genes.

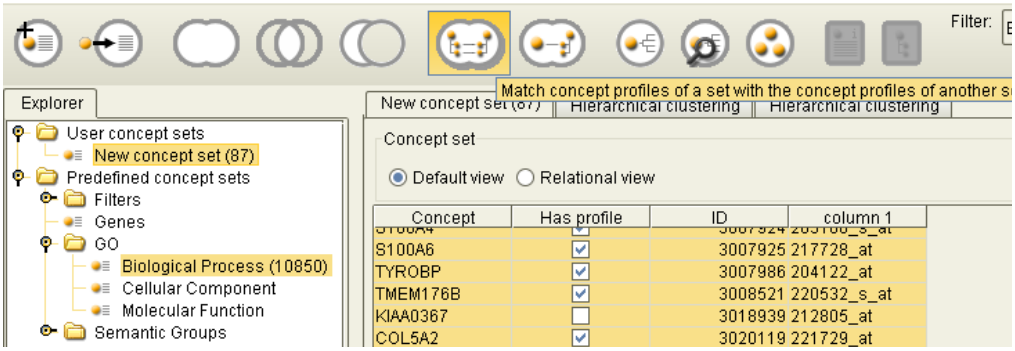


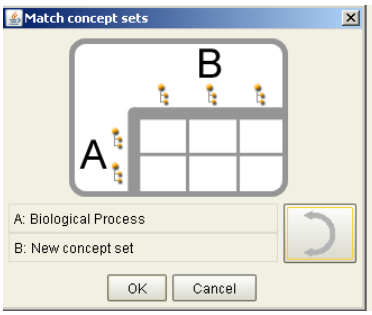


7. Do the same thing but now *only* annotate with the concept profiles from the semantic group "Biological Process". Annotation gives you now a ranked list of biological processes.



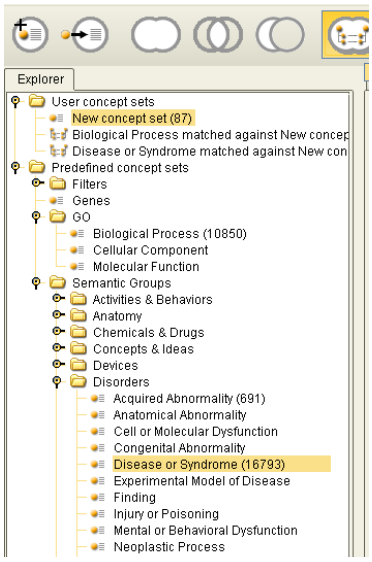
8. Undo the filter option. Select the “New concept set” and the “Biological Process” concept set in the left panels by clicking while holding the SHIFT key. Match the two concept sets. This will take a while!



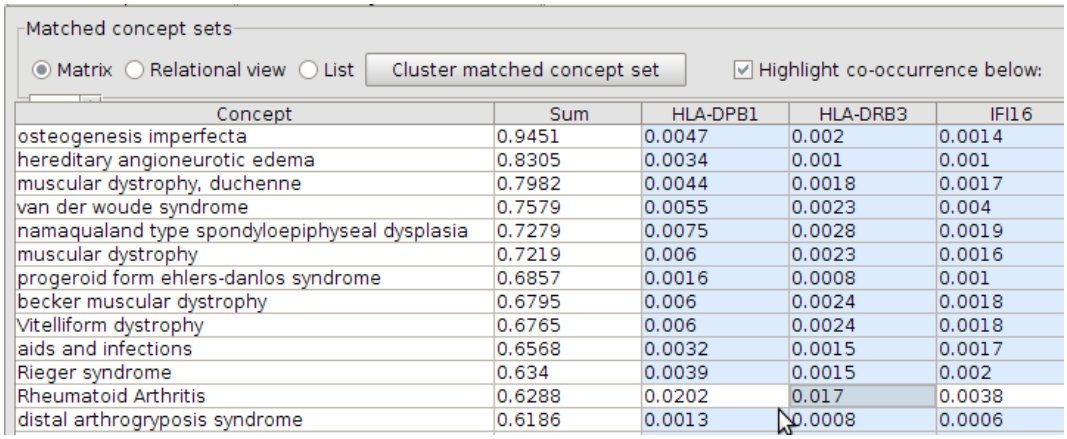


You will obtain a ranked list of biological processes associated with your gene list. This list can be compared to the list obtained with DAVID. Note that the list you created now is not based on curated annotation from GO but just on literature associations!

9. Perform the same matching but now with concept profiles from “Disease or Syndrome”.

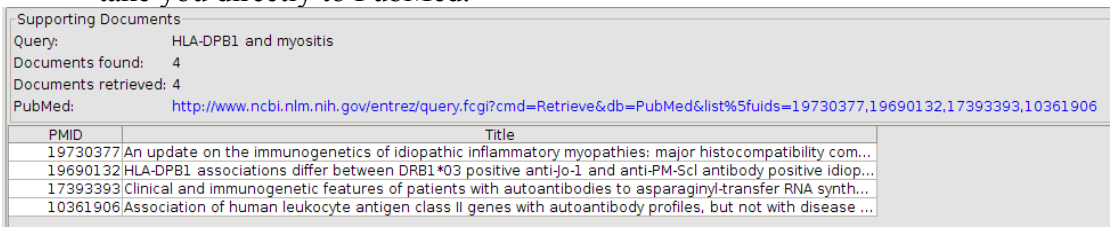


10. Check “Highlight co-occurrence below: 1”. All the white boxes have at least one abstract where the association between the gene and disease is mentioned. You can directly access the underlying literature by right-clicking on the box and select “Get co-occurring documents”.



Concept	Sum	HLA-DPB1	HLA-DRB3	IFI16
osteogenesis imperfecta	0.9451	0.0047	0.002	0.0014
hereditary angioneurotic edema	0.8305	0.0034	0.001	0.001
muscular dystrophy, duchenne	0.7982	0.0044	0.0018	0.0017
van der woude syndrome	0.7579	0.0055	0.0023	0.004
namaqualand type spondyloepiphyseal dysplasia	0.7279	0.0075	0.0028	0.0019
muscular dystrophy	0.7219	0.006	0.0023	0.0016
progeroid form ehlers-danlos syndrome	0.6857	0.0016	0.0008	0.001
becker muscular dystrophy	0.6795	0.006	0.0024	0.0018
Vitelliform dystrophy	0.6765	0.006	0.0024	0.0018
aids and infections	0.6568	0.0032	0.0015	0.0017
Rieger syndrome	0.634	0.0039	0.0015	0.002
Rheumatoid Arthritis	0.6288	0.0202	0.017	0.0038
distal arthrogryposis syndrome	0.6186	0.0013	0.0008	0.0006

11. All the blue boxes describe associations between genes and disease that have not been mentioned explicitly in Medline abstracts (no co-occurrence). To check the nature of these associations, right click on the box and select “Explain score”. A new window appears with a ranked list of concepts shared by the gene and disease. Again the underlying literature can be explored by right-clicking and selecting “Find supporting documents”. The hyperlink will take you directly to PubMed.



PMID	Title
19730377	An update on the immunogenetics of idiopathic inflammatory myopathies: major histocompatibility com...
19690132	HLA-DPB1 associations differ between DRB1*03 positive anti-Jo-1 and anti-PM-Scl antibody positive idiop...
17393393	Clinical and immunogenetic features of patients with autoantibodies to asparaginyl-transfer RNA synth...
10361906	Association of human leukocyte antigen class II genes with autoantibody profiles, but not with disease ...

12. You may want to finish by comparing your findings with DAVID and Anni with the those reported in Pescatori et al, FASEB J 2007, 21:1210-26. PMID: 17264171