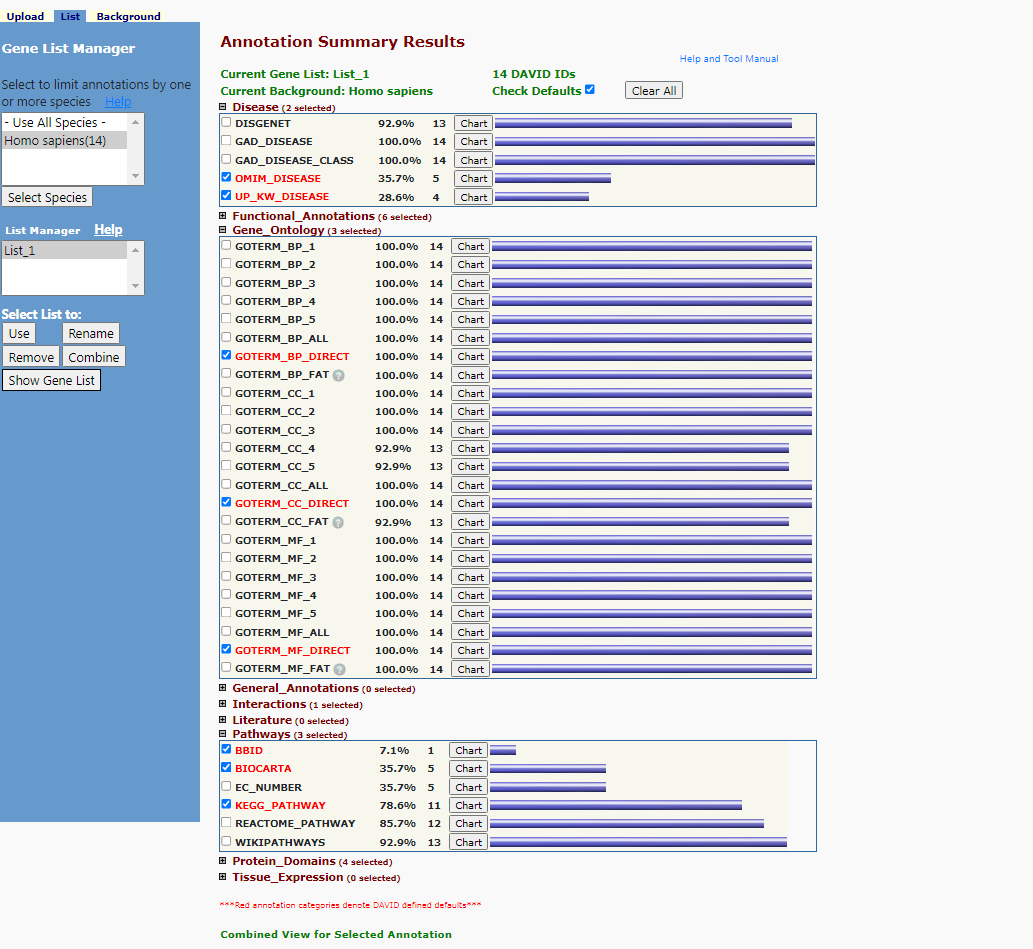
**Worksheet in class GENE SETS and FUNCTIONAL ANALYSIS (in class or online portion of the presentation grade)**

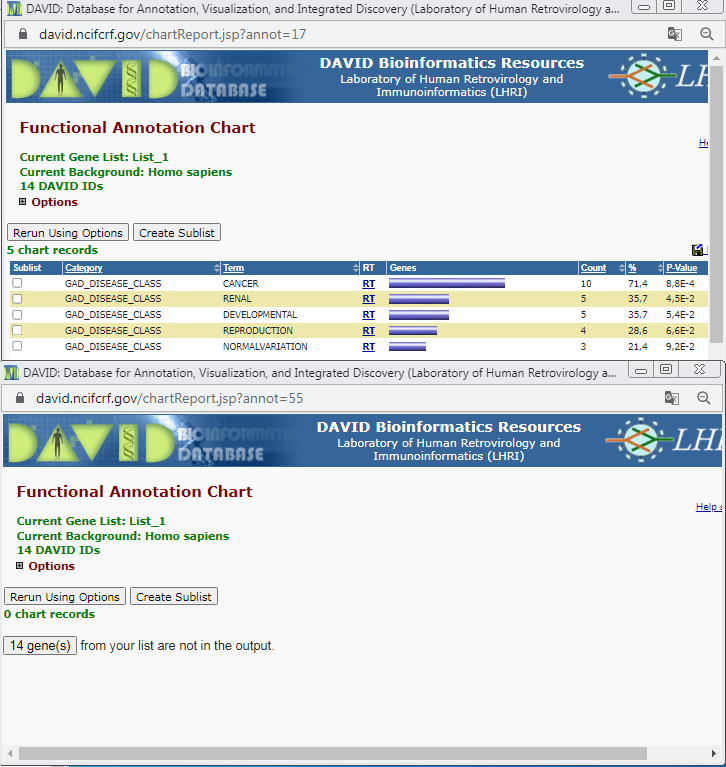
***You have to be in class or online working on the following questions starting from 16:30-17:20 and upload your work to get part of your presentation grade. If you are not present in class you have to go to the tutorial pages of each tool or google short tutorial videos online and study them between 15:30-16:20. If you are in class I will go over these tools before you work on the following questions***.

1. Go to David database [DAVID: Functional Annotation Tools (ncifcrf.gov)](https://david.ncifcrf.gov/tools.jsp) and use the list of Affymetrix IDs provided in the start analysis to examine if the gene list is associated significantly with any a) DISEASE; b) KEGG PATHWAY and c) GO:BP? Write your answer here with screenshots.
2. As a I see from the charts of GAD\_DIESEASE\_CLASS, it is related with cancer and other diseases.

For Annotation Summary:

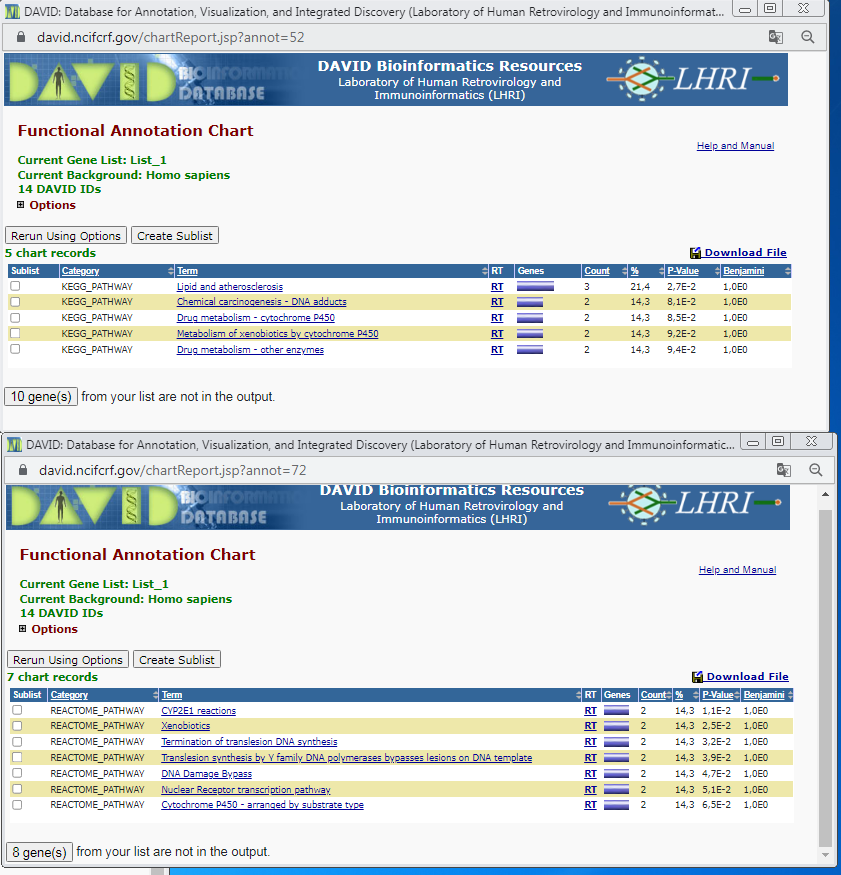


For the diseases example:



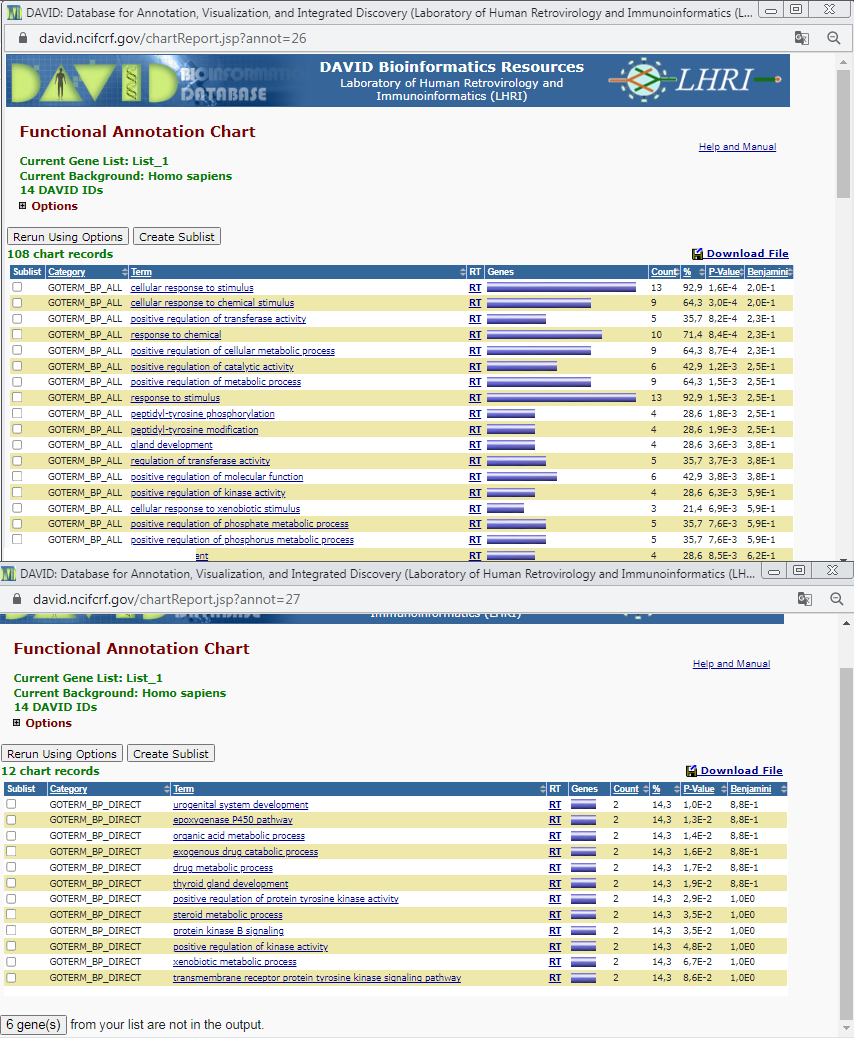
1. KEGG\_PATHWAY chart shows that it has association with the atherosclerosis.

For KEGG PATHWAY example:

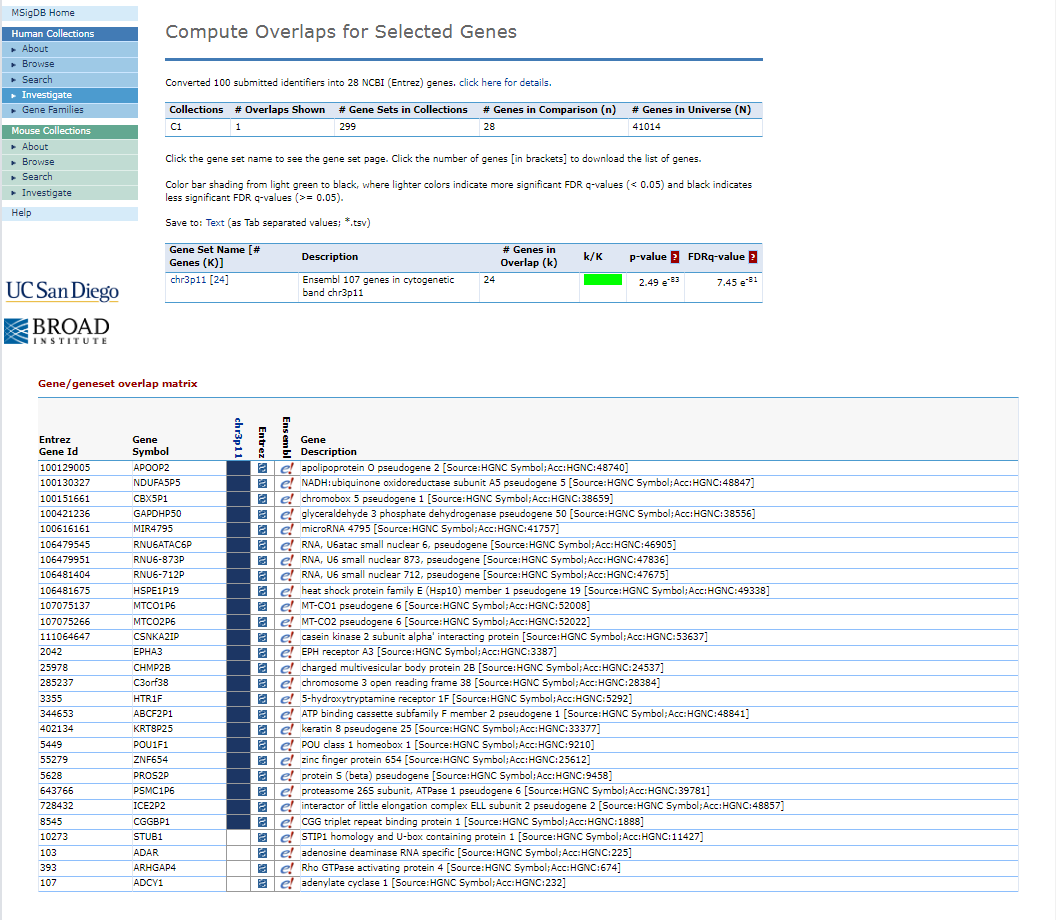


1. As seen from the charts it is related with the urogenital development etc. I did not see any disease in this part.

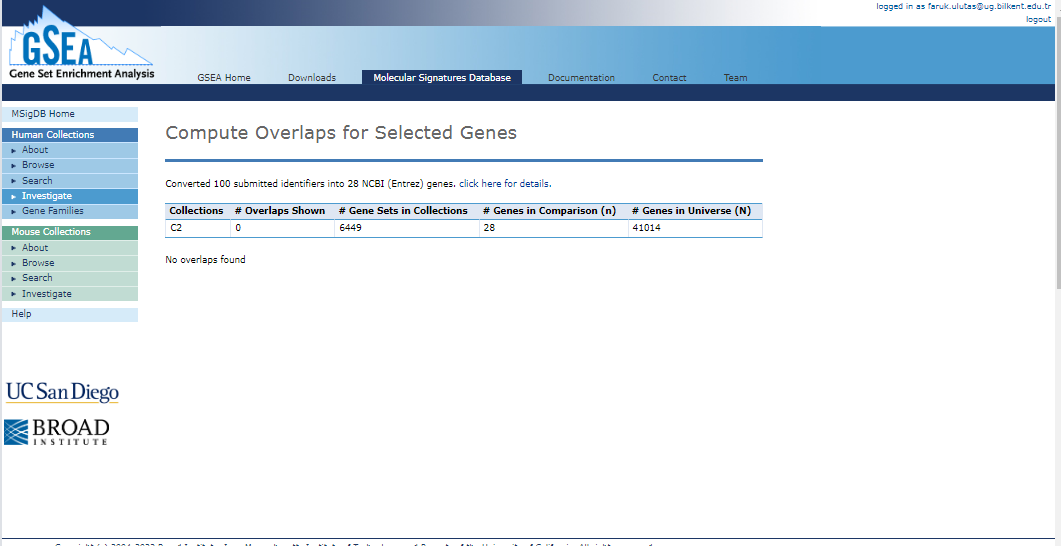
For GO:BP example:



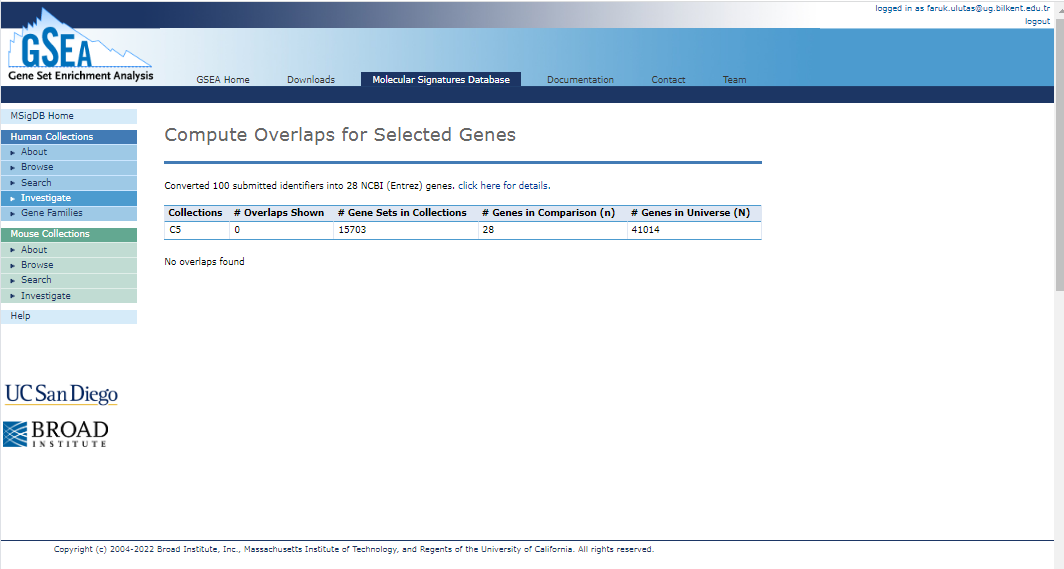
1. Go to Msigdb database [GSEA | MSigDB | Investigate Human Gene Sets (gsea-msigdb.org)](http://www.gsea-msigdb.org/gsea/msigdb/human/annotate.jsp)and select any chromosomal region when you click on C1 lists, open up the gene list link and copy the gene names (tsv file to open) and paste in the investigate genes box of Msigdb; and search if this set of genes overlaps with any a) C1:positional gene sets (sanity check); b) C2: curated gene sets; and c) C5:ontology gene sets? Use screenshots and explain in a few short sentences
2. It overlaps with 24 other genes which are listed in the screenshot’s overlap matrix.



1. There is no overlaps for C2.

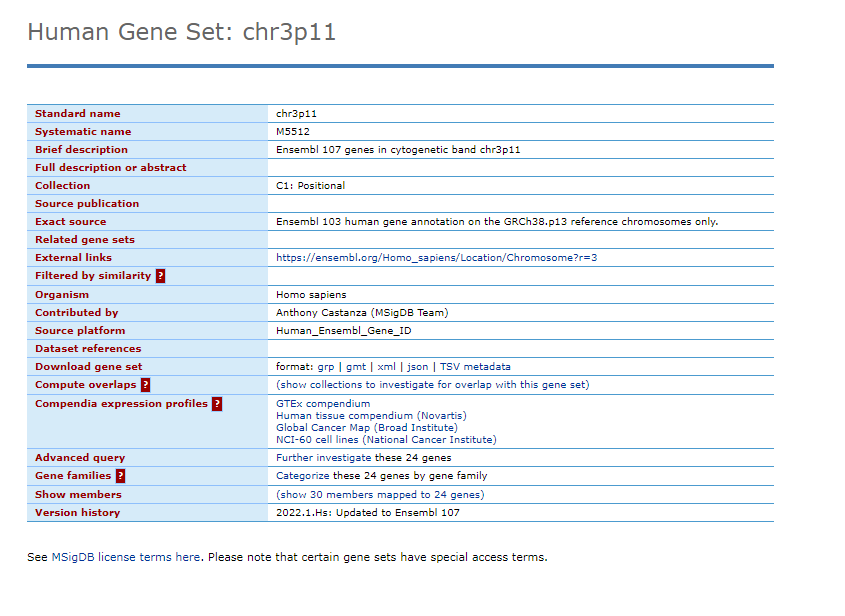


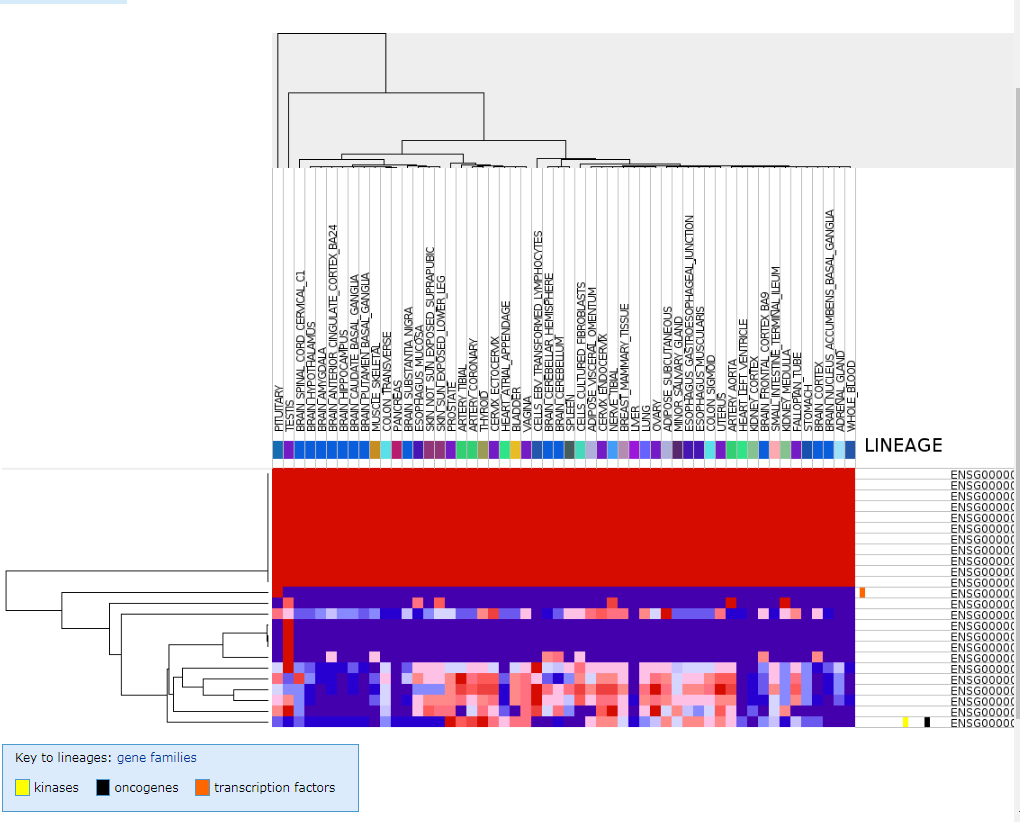
1. There is no overlaps for C5.



1. In Msigdb, for your chromosomal gene set you selected, examine their expression profiles in GTEX database? Is there any cluster of genes that are expressed similarly and in a tissue specific manner? Use screenshots and explain using a few short sentences.

Yes, In GTEx database, there are several brain related parts are listed.





1. Go to string database [STRING: functional protein association networks (string-db.org)](https://string-db.org/) and click search and on the next page, select from the left menu, “multiple proteins” and paste your choromosomal gene list you got from Msigdb (use ENSEMBL IDs, separated by comma) in the box and select Homo sapiens as the species and submit. This will bring you the mapped proteins, click continue. What you will get is a gene/protein network (nodes, vertices) and edges representing the relationship between the nodes. These interactions could be protein-protein (PPI) or functional (shared GO or KEGG or keywords) or cooccurrence in a pubmed article. More the edges between any two nodes there are more connections between them. Click on analysis button below. This will bring you tables of the enriched terms if there are any. Take a screenshot of the network as well as the enrichment table, and paste them here and explain shortly your results based on the following questions. A) Is the chromosomal region (and its genes) you have selected enriched with any terms? B) Does the network has significantly more number of edges than expected (random network)?
2. Yes. Since some of they have sort of edges with others this is sort of enrichment.
3. Yes. As far as I see from others screen some of the networks has only few or zero networks with others.

