**Computer Lab/Assignment 1 Due Wednesday, Sep 1 at 1:30 pm**

**1. PubMed – 5 pts for completion**

Perform a search for a biological subject of your choice across all of the Entrez databases (from the main Entrez search page). If your search results do not produce hits in both PubMed and Nucleotide (under Genomes), search using a different term that produces hits in both of these databases.

a) what is your search term? **Candida**

b) how many articles does your search return in PubMed? **75,630**

c) how many articles does your search return in PubMed if only the MeSH terms are used? (look in the Search details box on the right column; what is the meaning of "MeSH terms"?) **48.041**

d) Of the results in part c), how many articles have free full-text, and how many are reviews? **Free full text: 14,821. Reviews: 2469**

**2. Nucleotide – 5 pts for completion**

a) Pick a eukaryotic organism that you are unfamiliar with. What is the taxonomic name of the species, and what is its taxonomic ID number? **Cocos nucifera, Taxonomic ID - 13894**

b) How many nucleotide sequence entries for this organism are present in the main nucleotide database? **213264**

c) How many sequence entries for this species were released in the nucleotide database in the past 12 months? **2**

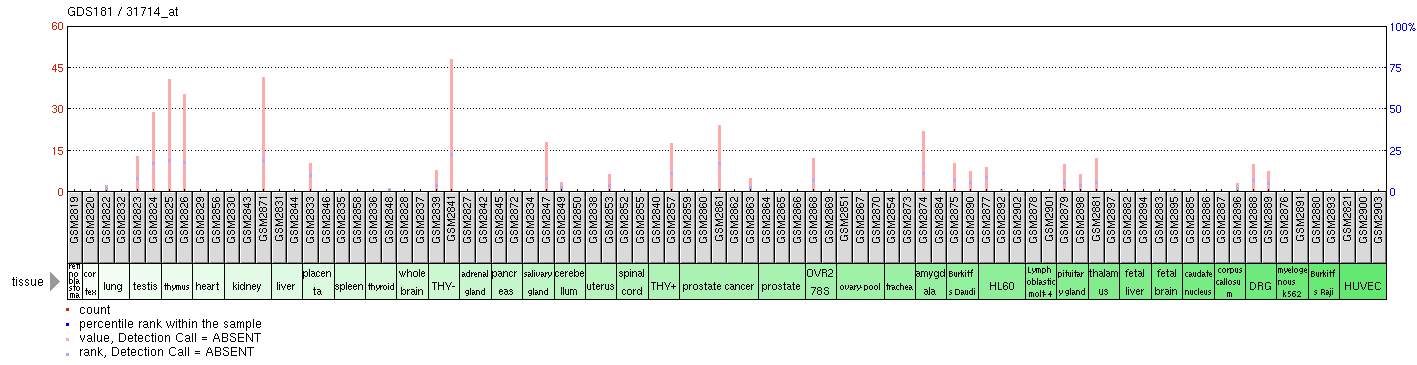
**3. OMIM, NCBI Gene and Gene Expression – 10 pts for completion**

Perform a search for a human gene of your choice in the OMIM database. **Gata6**

a) What are the cytogenetic location and the genomic coordinates? (if the gene lacks this information, pick another gene). **Cytogenetic locations: 18q11.2 Genomic coordinates (GRCh38):**[**18:22,169,588-22,202,527**](https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&position=chr18:22169588-22202527&dgv=pack&knownGene=pack&omimGene=pack)

b) What disease phenotype(s) is or are associated with variants of this gene, if any? **Atrial Septal defects, Atrioventricular septal defect 5, pancreatic agnesis and congenital heart defects and Tetralogy of fallots**

c) From "Gene Info" on the right column, follow the "NCBI Gene" link. Explore links from this page. Show a screenshot of a page showing the expression profile of your gene in different tissues/organs, in GEO profiles for GDS181 (large-scale analysis of the human transcriptome) microarray dataset. It should look something like this:



**For Gata6: A screenshot of a computer

Description automatically generated with medium confidence**

**4. Variation – 10 pts for correct answers**

Access the NCBI variation resources at <http://www.ncbi.nlm.nih.gov/variation> and review the resources listed.

a) If you wanted to know which clinical conditions are associated with variants in a particular gene, what tools or resources could you use?  **ClinVar , MedGen , Medical Genetics Summaries(MGS)**

b) What tool or database should you use to find studies and data on genotype-phenotype associations? **PheGenI , dbGaP**

**5. UCSC Genome Browser – 10 pts for completion**

Search for your human gene in the UCSC Genome Browser.

Look at the many different display options for annotation tracks. Adjust the display to show:

Base position (dense)

Gencode genes (pack)

ENCODE Regulation (show)

Conservation (Full)

Common SNPs (dense)

Repeat Masker (dense)

Adjust zoom to show your complete gene plus about 10,000 nucleotides of upstream sequence (look carefully at the Gencode gene annotations; arrows indicate the direction of transcription).

1. What non-exonic regions (if any) are conserved?

**The adjusted zoom showing complete gene + 10,000 nucleotide of upstream with highlighted conserved regions is shown below:**

Timeline

Description automatically generated

b) Which histone modifications are associated with the transcription start site? **Layered H3K4Me1, Layered H3K4Me3 and Layered H3K27Ac**

Paste a screenshot of the histone modifications at the transcription start site of your gene.

Timeline

Description automatically generated