**Assignment**

1. **Entrez Searching: Controlled Vocabularies**
	1. **Term translation in Entrez PubMed**
	2. On the [NCBI Home Page](http://www.ncbi.nlm.nih.gov/), enter cancer in the search box and click Go. (Leave the database menu set at All Databases.)
	3. Click on the PubMed.
	4. Scroll down, Click the Search details.
	5. How did Entrez translate this query? What new terms do you see? What field limits are used? Look in particular for the field [MeSH Terms].
	6. *Medical Subject Headings (MeSH) is a controlled vocabulary used to index all PubMed abstracts.*
	7. Edit your query in the text box so that only the term "neoplasms"[MeSH Terms] remains. Click the Search button under the text box to run the modified search. The resulting set only contain abstracts relevant to neoplastic disease. (Please note total hits number that decreased)



**1.2 Term translation in Entrez Protein**

* 1. Select Protein from the Database pulldown menu. Clear the search box and type cancer. Click Go.
	2. See the search details.
	3. How did Entrez translate this query?
	4. *Entrez Taxonomy, searched with the field [organism], is a controlled vocabulary used to index all molecular biology data at NCBI. In the sequence databases (eg. nucleotide, protein, genome, popset, snp), Entrez translates this query to retrieve records from the crustacean genus Cancer rather than records corresponding to neoplastic disease.*
	5. To retrieve only sequences from the genus *Cancer*, edit your query in the text box so that only the term "cancer"[organism] remains. Click the Search button under the text box to run the modified search.
	6. To retrieve sequences more related to neoplastic disease, change the query to cancer[title] and click Go. This query retrieves records that contain the word "cancer" in their definition lines (titles).

**1.3 Term translation in Entrez Taxonomy (*study by yourself*)**

* 1. Select Taxonomy from the Database pulldown menu. Clear the search box and type cancer. Click Go.
	2. You should see the single record for the genus *Cancer*. Click the Details tab above the search results.
	3. How did Entrez translate this query? What field limits are used?
	4. *In Taxonomy, Entrez simply searches for unfielded query terms using [All Names].*
	5. Clear your search box and search Taxonomy for rock crab. Click the name of the resulting hit. Do you see why the query "rock crab" found this record?
	6. *Entrez Taxonomy automatically assoicates scientific names with common names annotated on the taxonomy record.*
1. **Nucleotide Data in Entrez: Using Filters, Fields and Links**



1. Retrieve all zebrafish mRNAs
	1. On the [NCBI Home Page](http://www.ncbi.nlm.nih.gov/), set the Database pulldown to Nucleotide and enter zebrafish [organism] in the search box. Click Go.
	2. Click on the “Additional filters” as show on your left hand side.
	3. Select the Molecule types as mRNA and the “Source database” as RefSeq (for NCBI Reference Sequences). Notice that all the resulting records are in the Nucleotide database.
	4. *This set corresponds to the current transcriptome for zebrafish (a non-redundant set of all mRNAs).*
2. Limit the set to estrogen receptors
	1. Click the “Advanced” (at the above of page).
	2. Now we need to select the subset database. Our records are in Nucleotide. You should see your zebrafish[orgn] search, limited to "mRNA,RefSeq", in your history list.
	3. Select Title from the All Fields menu, and type estrogen receptor in the text box to the right. Click the “Show index list”.
	4. Click the first term in the list, estrogen receptor, and then click the AND button above the list. Notice that the term "estrogen receptor"[title] is now in your search box at the top of the page.
	5. Locate the history number of your zebrafish set limited to mRNA and RefSeq (the number will be preceded by a # sign). In the search box at the top of the page, type AND followed by the history number of the zebrafish set, preceded by a #. Your final query should look something like "estrogen receptor"[title] AND #5. Click Preview to see how many records this query retrieves. (you can use “Add” to builder too.)
	6. Click on the number of records retrieved (to the right of the query in your history).
3. View the FASTA sequences for the estrogen receptor mRNAs
	1. Which of these records are actually estrogen receptors (not predicted)? These should correspond to the three records that are NM RefSeqs.
	2. Click the checkboxes to the left of each of the three estrogen receptors. Then choose FASTA from the Display menu at the end of page. FASTA sequences for the three records should appear.
4. Find the Gene records for the estrogen receptor mRNAs
	1. Go back in your browser to the Entrez results page. Make sure the three estrogen receptor mRNAs are still checked.
	2. Study “Display setting” and “Sent to”.
5. Retrieve all nucleotide records associated with a single estrogen receptor gene
	1. Locate the gene record for the type 1 receptor with symbol esr1.
	2. Click Nucleotide in the Links menu to the right of the esr1 gene.
	3. In the “Display setting” menu at the above of page, click Revision History. How many times has the sequence of this record been updated?
6. **Protein Data in Entrez: Sequences, Structures and Domains**
7. Retreive a structure with bound gleevec
	1. On the [NCBI Home Page](http://www.ncbi.nlm.nih.gov/), type gleevec in the search box and click Go.
	2. Click on the hits to PubChem Compound.
	3. Click on the image of the resulting chemical (CID 5291), and then on the Protein Structures link to the right of the image on the summary page.