Exercises for UCSC Advanced Topics:
Table Browser and Custom Tracks

1) Obtain a list of SNPs in a single gene (Clock) using the Table Browser.

Skills: basic table search menus and options; choosing format; downloading sequence

2) Find CpG islands in known genes on the last part of chromosome 22 of the human genome. Obtain this sequence as one FASTA record per region.

Skills: basic table search menus and options; intersecting tables, choosing format, downloading sequence

3) From a list of UCSC genes, add gene symbols and GO IDs for additional information about the gene set. Bonus step: add GO terms.

Skills: basic table search menus and options; using tables, choosing related tables and selected fields

---

UCSC Advanced Topics Exercises, version 15b.
Correspond to the data available in October 2008.

Step-by-Step instructions for the UCSC Genome Browser Advanced Topics exercises

1) Obtain a list of SNPs in a single gene (Clock) using the Table Browser.

Step	Action	✓
1	Go to the UCSC Genome Browser homepage, **genome.ucsc.edu.**	
2	Enter the Table Browser, by clicking either of the **Table Browser links** from the homepage.	
3	Choose "**human**" and the "**May 2004**" assembly.	
4	Choose a table: Choose group "**Variation and Repeats**" in the "**group**" pull-down menu, "**SNPs'** in the "**track**" menu and "**snp125**" in the "**table**" menu.	
5	Type in **"CLOCK" in the position box.**	
6	Click the "**lookup**" button.	
7	You will see a list of records with "Clock" in the record. **Click the first link "CLOCK (NM_004898)".** The position of this gene will appear in position box. (Alternatively, you could paste in a known accession number and choose that option, but we wanted to show how the 'look up' works here.). This should give you the position: **chr4:56139588-56253925.**	
8	Leave buttons for filter and intersection as default (none).	
9	Under "**Output Format**" choose "**selected fields from primary and related tables**" in pull-down menu. Click **"get output."**	
10	In the resulting menu, **check the boxes for these fields: chrom, chromStart, chromEnd, name, strand, observed and func.** Click **"get output."**	
11	You can now copy/paste or download the resulting file for more study. Clock is an interesting gene that encodes a protein associated with circadian rhythm sleep disorders.	

2) Find CpG islands in known genes on the last part of chromosome 22 of the human genome. Obtain this sequence as one FASTA record per region.

Step	Action	✓
1	Go to the UCSC Genome Browser homepage, **genome.ucsc.edu.**	
2	Enter the Table Browser, by clicking either of the **Table Browser links** from the homepage.	
3	Choose "**human**" and the "**May 2004**" assembly.	
4	Choose a table: Choose "**Genes and Gene Predictions Tracks**" in "**group**" pull-down menu, choose "**Known genes**" in the "**track" menu** and **"knownGene** in the **"table"** menu**.	
5	Type in **"chr22:40000000-49396972" in the position box.**	
6	Click the **intersection** "**create**" button.	
7	On the resulting page, choose "**Regulation**" in the **group** menu and **"CpG Islands**" in the **track** menu. Leave other options as default (**"all Known Genes records that have any overlap with CpG Islands"**) and **click submit.**	
8	On the resulting page (back to table browser interface), choose "**sequence**" on the **output format** menu. Click "**Get Output."**	
9	On resulting screen, choose "**genomic'** and click **submit.**	
10	Make sure only "**5' UTR Exons, CDS Exons, 3' UTR Exons**" options are chosen **(unclick introns).** Then click on the **One FASTA record per region** option and leave rest of the sequence retrieval options as default. Click **"Get Sequence."**	
11	You can now copy/paste or download the resulting file (a list of CpG islands in known genes) for more study. The resulting file will be large. In cases like this, it is best to type in a file name in the "**Output File**" box. This will save a FASTA formatted text file to your computer.	

3) From a list of UCSC genes, add gene symbols and GO IDs for additional information about the gene set. Bonus step: add GO terms.

Step	Action	✔
1	Go to the UCSC Genome Browser homepage, **genome.ucsc.edu.** Enter the Table Browser by clicking either of the homepage **Table Browser links**.	
2	**Choose** "**human**" and the "**Mar 2006**" assembly.	
3	**Choose "Genes and Gene Prediction Tracks" group and the track "UCSC Genes."** The table you will need first is **"knownGene".** *Note: For this exercise and similar searches, it is useful to know which tables contain the data you need. If you choose the table and click the "describe table schema" button to examine the data fields within. This will also list all the tables linked to this table and joining fields.*	
4	**Choose the "position" radio button** and **type chr7** as the location. **Click "lookup"** to add the nucleotide range quickly. This just limits our set of data for this example. *You can choose genome-wide, if that's what you need, later.*	
5	Leave all other choices as default and **choose "selected fields from primary and related tables" in the "output format" menu**.	
6	**Click "get output."**	
7	On the next page you choose the items available for the output. At the top is our table choice. **Select "name" and "chrom" and "proteinID"** for our purposes. *You may find other times you want more data.*	
8	We now need to add data from linked tables. In the lower area, select one table from which we need data, by **clicking the checkbox: "kgXref"** (a cross-reference table for various identifiers). **Now click "Allow Selection from Checked Tables"** at the bottom of the page to view choices for table.	
9	In hg18.kgXref fields box, **choose "kgID," "geneSymbol" & "refseq" fields**.	
10	By making the kgXref table available, new associated tables are now also available from the "Linked Tables" area below. One of the new choices is **go**. **Click the checkbox next to go for the "goaPart" table. Click "Allow Selection from Checked Tables"** to view the choices for that table.	
11	In the new table box "go.goaPart Fields," **select the field "gold."**	
12	**Click "get output"** in the box "Select Fields from hg18.knownGene" in the top section of the page. Your results will display UCSC IDs, chromosome, protein ID, GO IDs, and gene symbols, which we built from the series of table choices.	
13	Extra credit: return to the last checkbox page and add GO terms by **checking the other "go" box ("term")** in the "Linked Tables" area. **Then check "Allow Selection from Checked Tables," and add "name" from go.term fields. Get output**.	