LEARN TO USE BIOINFORMATICS RESOURCES

Exercises for UCSC Genome Browser
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; filtering

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 Exercises, version 23b. Correspond to the data available in October 2015.

Step-by-Step instructions for 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	Access the Table Browser by clicking the **Tables or Table Browser link** from the homepage. **Click the "reset all user cart settings" link near the bottom of the panel** to clear prior activity.	
3	Select **Mammal** as the clade, **Human** as the genome. Choose the **Dec. 2013 (GRCh38/hg38)** assembly.	
4	Choose the data to access: • in "group", select **Variation** • in "track", select **Common SNPs** (the most recent SNP assembly) • in "table", select **snps142Common** (or the most recent SNP assembly)	
5	In the "region" area, **set the radio button to "position". In the position text box, type Clock.**	
6	**Click the "lookup" button** to locate the Clock coordinates. On the results page, **click the top-most CLOCK** gene result hyperlink. *The browser will return you to the table browser interface, and load the appropriate coordinates in the position box. Ensure the radio button still indicate "position".*	
7	For now, leave the filter, intersect, and other items as default, for none.	
8	In the **"output format" area**, **choose "selected fields from primary and related tables". Click the "get output" button.**	
9	Here we will select the fields we want. **Choose checkboxes for: chrom, chromStart, chromEnd, name, strand, observed, and func.** When this is done, **click "get output".**	
10	Scroll through the list of SNPs. Note that not all are "single" nucleotides. Some are found in untranslated regions, some in coding, some in introns.	
11	You can return to the table browser interface and "filter" for just missense SNPs. **Back button to the Table Browser query page.**	
12	**Click the filter: "create" button**. In the "func" or function area, **select the "missense" checkbox**. **Click "submit" for the filter.**	
13	**Re-run the query by clicking "get output" here, and again on the next page.** Note all the results are now missense SNPs.	
Note: UCSC calls the track Simple Nucleotide Polymorphisms--not just "*single*"--because this data may include insertions, deletions, and alterations that are larger than one single nucleotide position.		

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.** *(reset cart if you have filters in place)*	
2	Enter the Table Browser, by clicking either of the **Table Browser links** from the homepage.	
3	Choose "**human**" and the "**Dec. 2013**" assembly.	
4	Choose a table: Choose "**Genes and Gene Predictions Tracks**" in "**group"** pull-down menu, choose "**GENCODE**" (current assembly) in the "**track" menu** and **"knownGene** in the **"table"** menu**.**	
5	Type in **"chr22:40000000-50000000" in the position box.** [seven zeros]	
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 GENCODE 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	Access the Table Browser by clicking either of the homepage **Table Browser links or the Tools menu**. *(reset cart if you have filters in place)*	
2	**Choose** "**human**" and the "**Dec. 2013**" assembly.	
3	**Choose "Genes and Gene Prediction Tracks" group and the track "GENCODE" (current assembly)** The table you will need first is **"knownGene".** *Note: 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	**Click "clear"** to remove any intersection remaining from exercise 2. 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 using the cross-reference table **hg38.kgXref fields**.	
9	In hg38.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 hg38.knownGene" in the top section of the page. Your results will display UCSC IDs, chromosome, protein ID, GO IDs, and gene symbols, as selected.	
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**.	