

Hands on practices

Week 4: UCSC Genome Browser

UCSC Genome Browser(<http://genome.ucsc.edu>) contains the reference sequences and working draft assemblies for a large collection of genomes. It also contains functional annotations for these genomes and links to other biomedical resources. In this week, we will start with an unknown sequence, and learn how to use UCSC Genome Browser to check the function of this sequence and more.

1. BLAT search in UCSC Genome Browser.

BLAT

1.1 What is BLAT?

A BLAST-like alignment tool. It can be 50-500 times faster than BLAST. BLAT on DNA was designed to search sequences of 95% or greater similarity with length 25 bases or more. It can be accessed in the menu of UCSC genome browser ("Tools --> Blat").

1.2 Run BLAT

The query sequence is listed below. Please use the BLAT tool in the UCSC Genome Browser to determine the location of this DNA sequence in the human genome.

>seq1

```
gctgccgggacgggtccaagatggacggccgctcaggttctgcttttacctgcggcccag
agccccattcattgccccggtgctgagcggcgcccgagtcggcccaggcctccgggga
ctgccgtgccgggaggagaccgcatggcgaccctggaaaagctgatgaaggccttcga
gtccctcaagtccctccagcagcagcagcagcagcagcagcagcagcagcagcagcagc
gcagcagcagcagcagcagcaacagccgccaccgcccggccgcccggccgcccctcctca
gcttctcagccgcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc
gcccggcccccggccaccggcccggctgtggctgaggagccgctgcaccgaccaaaagaa
agaactttcagctaccaagaaagaccgtgtgaatcattgtctgacaatatgtgaaaacat
agtggcacagtctgtcagaaattctccagaatttcagaaacttctgggcatcgctatgga
actttttctgctgtgtagtgatgacgcagagtcagatgtcaggatggtggctgacgaatg
cctcaacaaagtatatcaagctttgatggattctaatttccaaggttacagctcgagct
ctataaggaataaaaaagaatggtgcccctcggagtttgctgctgccctgtggaggtt
tgctgagctggctcacctggttcggcctcagaaatcaggecttacctggtgaaccttct
gccgtgacctgactcgaacaagcaagagaccgaagaatcagtcaggagaccttggtgc
agctgttccaaaattatggcttcttttgcaattttgcaaatgacaatgaaattaaggt
tttgtaaaaggccttcatagcgaacctgaagtcaagctccccaccattcggcgacagc
ggctggatcagcagtgagcatctgccagcactcaagaaggacacaatatttctatagttg
gctactaaatgtgctccttaggcttactcgcttctgtcgaggatgaacactccactctgct
gattcctggcgtgctgctcacctgaggtatttggtgcccttgctgcagcagcaggtcaa
```

```
ggacacaagcctgaaaggcagcttcggagtgacaaggaaagaatggaagtctctccttc
tgcagagcagcttgtccaggtttatgaactgacgttacatcatacacagcaccaagacca
```

How many regions were hit by this query sequence? Choose the best alignment, and try to answer the following questions: Is there a gene overlapping with this sequence in the human genome? If yes, what is the size of the gene? How many exons does this gene contain?

2. The tracks in UCSC Genome Browser

Each dataset is called a track in UCSC Genome Browser. Any annotation (genome-wide or not, functional annotation or not) can be represented as a track. The current file format supported in UCSC Genome Browser include BED, bigBed, bedGraph, GFF, GTF, WIG, bigWig, MAF, BAM, BED, Personal Genome SNP, VCF, broadPeak, narrowPeak, and PSL.

Check the human genome in UCSC Genome Browser. How many tracks are there for the human genome? There are 10 classes of different tracks. Check the drop-down boxes to see how it will impact the visualization of the result in the genome browser.

3. The tables underlining the tracks.

Each track is in fact a table file in the background. Check the corresponding table for a track, for example, the Refseq annotation track. Please find it in the Table Browser.

You can access the Table Browser by click "Tools" in the top level menu and select the "Table Browser" sub-menu.

4. Add your own track: custom tracks.

In the Table Browser, input "chr4:2976408-3417992" in the box for the "region" section and select the output format as "custom track", and then click "get output" button. You will be asked to input some information for your own custom track, such as the name and description of this track. You can get the custom track in the table browser or in the genome browser. Give it a try.

5. Track comparison.

You can compare any two tracks (including custom tracks) in the table browser.

In the Table Browser page, select a track that will be compared and click "Intersection: create" button (above the "output format" input box). An intersection creation page will show to let you select the other track that you want to compare with. You can select "Repeats" in the "Group" box, and pick "RepeatMasker" track. Then, you can click "submit" button to create the intersection comparison. Finally, you need to get output as for any Table, i.e., select proper output format and click the "get output" button to get the result of the comparison. You can always click the "summary/statistics" button before you click "get output" button to get some feeling about the size of the output.

6. Links to outside databases

Go back the Genome Browser and check the region hit by the query sequence in step 1. Check the Refseq track annotation of this region. Double click the gene named "HTT". You will see a page showing "Refseq Gene HTT" followed by some links to outside databases, such as OMIM, PubMed. You can click those links to see the information for this gene.

7. Function of the query sequence

Please answer this question by finding the function of the gene overlapping with the query sequence we give in step 1.