

Computing on Genomes, I

Biol4230

Friday, March 24, 2017

Bill Pearson wrp@virginia.edu 4-2818 Jordan 6-057

- Versions of Genomes, genome coordinates
- Genome data types:
 - sequence
 - alignment
 - features on locations
 - quantitative data
- Looking at a genome (IGV)
- incorporating data
 - The UCSC table viewer
 - importing into IGV

fasta.bioch.virginia.edu/biol4230

1

To learn more:

- UCSC FAQs: <http://genome.ucsc.edu/FAQ/>
- Genome releases:
<http://genome.ucsc.edu/FAQ/FAQreleases.html>
- Genome data file formats:
<http://genome.ucsc.edu/FAQ/FAQformat.html>
- Integrated Genome Viewer:
<https://www.broadinstitute.org/igv/>

fasta.bioch.virginia.edu/biol4230

2

Genome data (UCSC table browser)

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Help](#) [About Us](#)

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: **genome:** **assembly:**

group: **track:** [add custom tracks](#) [track hubs](#)

table: [describe table schema](#)

region: ☐ genome ☒ position [lookup](#) [define regions](#)

identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: Send output to ☐ Galaxy ☐ GREAT ☐ GenomeSpace

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

[get output](#) [summary/statistics](#)

To reset all user cart settings (including custom tracks), [click here](#).

fasta.bioch.virginia.edu/biol4230

5

Genome data (UCSC table browser)

```
chr1 hg38_refGene at_codon 109620269 109620271 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109620269 109620278 0.000000 + 0 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109619813 109620278 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109625303 109625433 0.000000 + 2 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109625303 109625433 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109625662 109625792 0.000000 + 0 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109625662 109625792 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109626160 109626228 0.000000 + 1 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109626160 109626228 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109626319 109626427 0.000000 + 1 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109626319 109626427 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109626726 109626912 0.000000 + 0 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109626726 109626912 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109627175 109627316 0.000000 + 2 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109627175 109627316 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109627429 109627518 0.000000 + 1 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109627429 109627518 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109627774 109627903 0.000000 + 1 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109627774 109627903 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109628083 109628277 0.000000 + 0 gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene exon 109628083 109628277 0.000000 + . gene_id "NM_139156"; transcript_id "NM_139156";
chr1 hg38_refGene CDS 109628364 109628495 0.000000 + 0 gene_id "NM_139156"; transcript_id "NM_139156";
```

GFF/GTF format – Coordinate blocks

fasta.bioch.virginia.edu/biol4230

6

Genome data (UCSC table browser)

```
track type=wiggle_0 name="GM12878" description="H3K4Me1 Mark (Often Found Near  
Regulatory Elements) on GM12878 Cells from ENCODE"  
#bedGraph section chr1:110215375-110255450  
chr1 110230400      110230425      2.72  
chr1 110230425      110230450      3  
chr1 110230450      110230475      3  
chr1 110230475      110230500      3  
chr1 110230500      110230525      3  
chr1 110230525      110230550      3  
chr1 110230550      110230575      3  
chr1 110230575      110230600      1.36  
chr1 110230925      110230950      0.72  
chr1 110230950      110230975      1
```

Wiggle format – continuous data