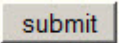# UCSC Genome Browser
ASHG 2014 Workshop

We will be using human assembly hg19.   Some steps may seem a bit cryptic or truncated. That is by design, so you will think about things as you go….

In this document, if you see something in square brackets, e.g, [submit], look for a button with that label: submit.

1. Load Custom Tracks.

- On genome.ucsc.edu, follow "Genome Browser" link to the Browser Gateway.
- Click link: "Click here to reset."
- Navigate to hg19 [submit].   Then: [hide all]
- In the Genes and Gene Prediction bluebar track group, turn on UCSC Genes to "pack" [refresh]
- Below the Browser graphic, find button: [add custom tracks]
      or: [manage custom tracks] then [add custom tracks]
- In a separate window/tab, go to bit.ly/ucscASHG2014 and click into ctExamples.txt.
- Copy either the url for this page or its contents into the upper box on the Custom Track imput page, "Paste URLs or data"   [Submit]
- On the new page, click into top link in the "Pos" column, labeled "chr6"
- Observe the relationship between the items in the Browser graphic and the data you copied from the other page.
- These are a few of the 14 data types available.

2. Visualize a BAM-file Custom Track.

- The Genome Browser display of BAM files shows coverage, mappings of individual sequence reads, base-pair mismatches and quality scores.
- To load a BAM file, it must be indexed (create a .bai file) and placed on a server visible to the Web (and the Browser) via http: protocol.
- We have placed a file on a server for you to load (the filename is in the `bigDataUrl` portion of URL a few steps below).
- Below the Browser graphic, click [add custom tracks]
      or: [manage custom tracks] then [add custom tracks]

- At the top of the page, select:

assembly: hg18

- In the upload (upper) box, type or copy (all on one line):

```
track name=chr21_export type=bam bigDataUrl=http://genome-
test.cse.ucsc.edu/ABRF2010/chr21.bam visibility=pack
```

  [submit]

- New page.   [go to genome browser]
- Back at the Browser, navigate to this interesting location:
       chr21:37,366,714-37,366,825

- Turn on the UCSC Genes and SNPs (130) tracks (This older version works because we are on hg18).
- Zoom out 10x a couple of times. Observe.

- To see this result all at once, you can load a session: Under Restore Settings, type (or click "BAM result" on workshop webpage):
       user: example
       session name: hg18_bamSnps [submit]

- At the top of the page, where it says, "Updated Session," click "Browser."

3a. Table Browser Join: Load a list of genes into the Table Browser and extract info from joined tables.
- In a separate window/tab, go to bit.ly/ucscASHG2014
- Click to genelist and copy contents.
- In genome.ucsc.edu, use the "Tools…" pulldown and choose "Table Browser"
- In Table Browser, choose
       group: Genes and Gene Predictions
       table: knownGene
       track: UCSC Genes
       region: genome
       identifiers: [paste list]
            paste list into box [submit]
       output format: selected fields from primary and related tables [get output]

- On the new page, select from the knownGene table:
       chrom
       txStart
       txEnd
       note: knownGene.name is an internal UCSC identifier and NOT the common
            name for the gene.

- From the kgXref table, select:
    geneSymbol   (THIS is the name of the gene)
    description
- [get output]

3b. Repeat this, getting only one isoform per gene.
- Back at the Table Browser, follow the same process, except choose:
    table: knownCanonical

4. Table Browser Intersection / Subtraction.   We will create a Custom Track of uncommon SNPs by subtracting Common SNPs from All SNPs.
- First, check out how big the All SNPs (dbSNP 138) table is:
- At the bottom of the Genome Browser graphic page, find the Variation bluebar group and "All SNPs (138)" link above its pulldown track controls. Click the link.
- On the new page, click "View table schema." Note the **Row Count:** 65.5 M! … and notice all the other data available for each SNP.
- We'll work with a small subset of SNPs, because it is very slow to do an intersection with such a large table, especially when we are all doing it at once!

LOAD A SESSION
- We have created a pair of tracks that contain only a few thousand SNPs from the All SNPs (138) and Common SNPs (138) tables – covering only 250 kb of chr4, about 0.01% of the genome.
- Starting from the main Browser graphic page, go the pulldown, "My Data" and select "Sessions."
- Under Restore Settings, type:
    user: example
    session name: hg19_ashg2014_SNPs [submit]

- At the top of the page, where it says, "Updated Session," click "Browser."
- Choose a small portion of this 250 kb to zoom and see the difference between these two tracks.

TABLE BROWSER INTERSECTION
- We will be subtracting Common SNPs from All SNPs, using our small set of a few thousand, and saving the result in a new Custom Track.
- Go to the Table Browser (via "Tools…" pulldown)
- Choose:
    group: Custom Tracks
    track: 4q21.1 all SNPs   (primary table)
    table: (accept default)
    region: genome
    intersection: [create]

- On new page:
  - **group:** Custom Tracks
  - **track:** 4q21.1 common   (secondary table)
  - **table:** (accept default)

  select the second button:
  All 4q21.1 all SNPs records that have no overlap with 4q21.1 common
  [submit]

- Back at Table Browser main page:
  - **output format:** custom track   [get output]

- On the new page, give the track a name ("uncommon" ?) and long label.
  - **visibility:** dense   [get custom track in genome browser]

- If you are lost, you can see the result here (or click "SNP result" on the workshop webpage):
  - **user:** example
  - **session name:** hg19_ashg2014_snpResult
    - zoom out a few times

5. Make a filter in the Table Browser to obtain only the data that exceed some value.

- We will make a track that contains only the strongest vertebrate conservation signal at the FGFR1 locus. We will confine ourselves to a small region to avoid slow processing times.
- Start by loading a session that looks at the 3' end of the FGFR1 gene, with the Conservation track turned on (or click "Filter input" on the workshop webpage):
  - **user:** example
  - **session name:** hg19_FGFR1cons

- Note that the maximum value displayed for the PhyloP Conservation track is 4.88 (left margin), and that some of the peaks have a pink "hat" on top, indicating that the value extends off the scale. A reasonable cutoff to get the tallest peaks is 4.0.
- Click into one of the peaks to go to the details page and then note the table name at the top: phyloP100wayAll .
- Go to the Table Browser ("Tools …") and set it to that table:
  - **group:** Comparative Genomics
  - **track:** Conservation
  - **table:** phyloP100way (the default)
  - **position:** (button)   (accept the default coordinates which match the region on
          Browser graphic).
          PLEASE DO NOT LEAVE "genome" SELECTED !
  - **filter:** [create]

- On the new page:
    data value   is: > 4.0   [submit]

- Back at Table Browser page:
    output format: custom track [get output]

- Make up a name for short and long labels ("topOfPeaks" ?). [get custom track in genome browser]

ADJUST SCALES OF CUSTOM TRACK AND PHYLOP TRACK TO MATCH
- To make the two tracks look like they belong together and have the same vertical scale, we will configure the scales and windowing to match each other. (Windowing: Because the display must combine the data for more than one nucleotide per pixel, a choice must be made about how to display the data. The default "windowing" on the phyloP track is "mean," which smoothes out the values -- especially any negative values, as you will see).
- Right-click the phyloP track.   Choose "Configure 100 Vert. Cons"
    Vertical viewing range: min: -2   max -4 (to match cutoff we used in filter)
    Windowing function: maximum

- Right-click and configure your new Custom Track.
    Track height: 50   (to match phyloP default)
    Vertical viewing range: min: 4   max: 10   (to match phyloP range of 6)
    Data view scaling: use vertical viewing range setting
    Windowing function: maximum

- To see what this display should look like at this point, load (or click "Filter result" on the workshop webpage):
    user: example
    session name: hg19_FGFR1consResults

6. Use the Variant Annotation Integrator (VAI) to classify SNP variants.
- The VAI tool will take files of variants in VCF or pgSnp format and return information about the biochemical effects of the variants.
- Start by loading a session that reloads the Custom Tracks in #1 above (or click "Custom Track result" on the workshop webpage):
    user: example
    session name: hg19_ctExamplesResult

- One of the Custom Tracks, My SNPs -- Personal Genome SNP format, is in pgSnp format, and is suitable for the VAI (the other supported format is VCF).

- If you open the VAI in the next step in a separate tab or window, you will be able to compare the results from the VAI to the Browser graphic.
- In the "Tools…" pulldown, select "Variant Annotation Integrator"
- In the **Select Variants** section:
  variants: My SNPs – Personal Genome SNP format   (should be the default)

- In the **Select Genes** section, choose: Basic Gene Annotation Set from GENCODE Version 19.
- Notice the large number of annotation options, including Regulatory Annotations. Select any option (or accept default), then near the bottom, under **Configure Output**,
  output format: Variant Effect Predictor (HTML)   [get results]

- Note the types of information provided: Gene; Consequence (misssense, frameshift, splice_donor_variant, {up,down}stream_gene_variant (and distance)); Co-located Variation (+dbSnp rs#, if available); SIFT, PolyPhen scores, etc.

- Coming in VAI 2.0:   sorting by variant type.