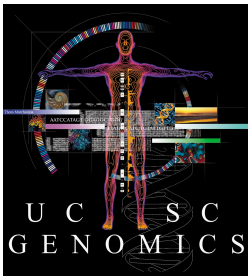


UCSC Genome Browser Workshop Practical
University of Pennsylvania School of Veterinary Medicine
Children's Hospital of Philadelphia

presented by Robert Kuhn Consulting

We will primarily be using human assembly hg19. The Browser operates the same way for all organisms and assemblies, but hg19 is the best annotated. The Browser functionality you will see here is applicable to mouse or other organisms. These problems will take you through a variety of resources at the UCSC Genome Browser. You will learn how to use the Browser navigation, the Table Browser and Saved Sessions. Experiment! YOU WILL NOT BREAK ANYTHING.

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

We will use the main Browser site, **http://genome.ucsc.edu** . It is also possible to use the European mirror **http://genome-euro.ucsc.edu ,** which can be faster at certain times of the day.

At any point in this workshop, you can login to the Sessions tool and save your session. Go to the pulldown menu, "My Data…" and choose "My Sessions." Login using:

username: philly2023
password: genome

This will be a shared login! Please use your initials and/or some other unique text string and do not overwrite the sessions of anyone else.

1a. Find the gene, Fibroblast Growth Factor Receptor 1 (FGFR1).

- Go to the home page, select Genome Browser, then Reset All User Settings, then hg19 human assembly. [Go]
- [hide all]
- Set UCSC Genes to "pack" [refresh]
- Type FGFR1 into the position box above the Browser graphic, grab it from the dropdown [go]

1b. Turn off splice variants.

- To the left of the gene track, locate the little gray bar and click it.
- Turn off the splice variants by unchecking the box [submit]

1c. Zoom to a few exons.

- Put your cursor into the graphic near the coordinates at the top, above the data tracks. Drag it to the right or left to select a few exons, then let go. Choose the [zoom in] option. OR: type the following coordinates into the Position Box: chr8:38,278,721-38,285,067 and hit [go].

1d. Turn on SNPs tracks.

- In the Variation bluebar group at the bottom of the page, turn the dbSNP 155 track to "pack." [refresh] – but note that the track is in "dense" mode, not "pack."
- This is a quirk of the nested composite tracks.
- Click the gray bar to the left of the SNP track and notice that there are several subtracks, only one of which is on, and that one (Common SNPs) is set to dense.
- Set the Common SNPs to "hide" and also turn on ClinVar SNPs to "pack."
- In a separate window, see these results in the following session (your exact window may vary):
  user: example
  session: hg19_fgfr1SNPs

1e. Export DNA showing exons, introns and SNPs.

Note: If you loaded the session listed in the previous instruction, the extended cases/color options described here should be already selected.

- In the top menu, find "View > DNA."
- [extended case / color options]
- Set "default case" to "lower."
- Check "toggle case" for the Genes track.
- For color, choose one for each track:
- dbSNPs 155: Red 255
- UCSC Genes: 255 Blue
- [submit]
- Observe the annotation of the SNPs and the exons on the DNA sequence.

1f. Zoom to codon level.

- Return to the main graphic at the pulldown menu, Genome Browser.
- Find Exon 8 by moving your mouse over the exon in the middle of the screen.
- Now drag-and-zoom as you did before to select just this exon. OR: go to these coordinates: chr8:38,281,993-38,282,314.
- At this resolution you see the amino acid single-letter code for the exon.
- Zoom again using the 10x "Zoom in" button.
- Observe the codon numbers.
- Go find a codon table at Google (search: image codon table).
- Note that your codons do not match.

- Click the little arrow at the left side of the DNA sequence to get the revere complement, which should match the codon.

1g. Turn other isoforms back on.

- Using the gray configuration button, turn the isoforms back on as before. [Submit]
- Notice that the numbering for an amino acid is different for the various forms, depending on how many and which exons upstream form the annotation.

1h. Click into the details page of the SNP (rs727505369).

- Click into the red SNP (rs727505369). If it is not on your screen type or copy it into the search box.
- On the details page, note the change in amino acid caused by the variant: E (glutamate) to G (glycine).
- Return to the Browser using the link in the top bluebar.

1i. Find out some phenotypic information about the variant via HGMD and UniProt Variants tracks.

- In the Phenotype and Literature bluebar group beneath the Browser graphic, UniProt Variants and turn it on to "pack."
- Hover your mouse over, the UniProt record to see the phenotype of the variant: Hypogonadotropic hypogonadism 2 with or without anosmia; also found in family member with isolated anosmia.

1j. Read a full authoritative article on this gene.

- In the Phenotype and Literature bluebar group beneath the Browser graphic, turn on the Gene Reviews track.
- Click into the green bar in the GeneReviews track – this track points to articles solicited by NCBI.
- From the details page, you can click into several different articles relevant to this gene, such as:

  craniosynostosis FGFR-Related Craniosynostosis Syndromes

- The paper will open in a new window.

1k. Look at Conservation data for this gene.

- Go back to the Browser using the Genome Browser link in the top bluebar.
- In the Comparative Genomics bluebar section below the graphic, turn the Conservation track to "pack." [refresh] . This track shows 100 species in alignment to human (though only 8 are displayed – you can turn the rest on using the configuration button you've seen above).
- Zoom out 100x to see the region with both exons and introns.
- Drag the Conservation track up to just below the gene track: Put your cursor to the left of the track and hold down the mouse button to drag.
- Notice how the conservation signal is strong at the exons and weak in introns – exactly what you would expect if a gene is selected during evolution based on the amino-acid

content (functional units). Introns are free to evolve away – the farther away from human, the less homology.
- Click into the gray bar to turn on all 100 species.

1m. Make a Session from this view to share with others.

- Using "My data… My Sessions" pulldown, name your session.
  session: hg19_<yourInitials>_FGFR1
- Remember that we are sharing this sessions login. Please do not overwrite some else's session! Use some other name if there is a collision with your initials.
- You can now easily get back to this view by loading the session. You can also share it with others. Use your right mouse button to get the URL to share.

1n. Jump to the mouse homologue of this region.

- In the top bluebar, pull down "View…, In Other Genomes."
- Choose "Mouse" and accept mm10 assembly [Submit]
- Click into the link of mouse regions. Note that 63% of the bases match, mostly in the exons.
- You are now looking at the FGFR1 gene in mouse. Note the similarity of structure.

1o. Make a Session from this mouse view to share with others.

- Using "My data… My Sessions" pulldown, name your session.
  session: mm10_<yourInitials>_FGFR1
- Remember that we are sharing this sessions login. Do not overwrite some else's session! Use some other name if there is a collision with your initials.
- You can now easily get back to this view by loading the session. You can also share it with others.

2. Find all the genes in the chromosome cytological band, 4p16.3, their coordinates and a brief description.

- From the main page of the Genome Browser, go to "Tools…, Table Browser."
- Select Human genome assembly hg19 and find the table, knownCanonical (associated with UCSC Genes track in Genes group). (note: if you use the table, knownGene, you get info about multiple isoforms for each gene).
- Use "position" box and "lookup" button to get coordinates (type in "4p16.3").
- output format: selected fields from …. [get output]
- On the new page, select chrom, chromStart, chromEnd from knownCanonical and geneSymbol and description from kgXref. [get output]

3. Load a group of Custom Tracks from a file or URL.

- [hide all] Turn on UCSC Genes track to "pack" [refresh].
- From the Browser graphical viewer, navigate to "add custom tracks" (or "manage custom tracks" first).
- In a second Firefox (or Chrome, or…) window, go to http://bit.ly/ucscPhilly2023 and then click into ctExamples.txt.
- Copy the data from this file into the upper text box of the custom track page – OR use the URL to the file. [Submit].

- Click the "chr6" link at the top of the "Pos" column.
  Your screen should look like this session:
  user: example
  session: hg19_ctExamples

To get to the sessions menu faster, you can use a keyboard shortcut. To find the shortcut, type "?" anywhere on the page (except in the search box). Notice in the menu of shortcuts that "My Sessions" is "s s."

- Compare the data on the .txt page and the display in the Genome Browser to see how the text determines the properties of the tracks.
- When you have completed this, save your session under "My Data > My Sessions," giving it some unique name.

4. Load a BAM file.

During a sequencing experiment, the first step in a processing pipeline is to align the raw reads from the machine to the genome reference assembly. The result is a BAM (Binary Alignment Mapping) file.

A BAM file can be loaded into the Browser to visualize the coverage of the sequencing, but only from a location already on the web somewhere. The BAM and its associated .bai file must be together in the same directory on your server. To read more about loading BAM files: go to the Add Custom Tracks page and click the "BAM" link at the top of the page.

- Go to the workshop webpage and click into the link, bamTracks.txt
- Note three blocks of text representing three tracks. Each has a "browser" line and a "track" line. Note the variable "type=." Each block points to data which can be anywhere on the web.
- Note that only the bedGraph file is human readable. *.bam and *bw (bigwig) are binary indexed files and not human readable.
- Copy/Paste either the contents of the page or the URL as you did above into the Custom Tracks input text box in the Browser.
- Click into the "chr20" link in the "Pos" column to view the tracks in the Browser.
- Zoom into an exon to see details. The BAM track will show mismatches between the sample an the reference as red tick marks. The other two tracks are coverage tracks.

4b. Observe the reads in a BAM file as a wiggle density graph.

Above you viewed the individual reads and observed mismatches between your sample and the reference assembly. Now let's view the data as a continuous density graph (mimics the bedGraph track).

- Click into the configuration page using either the gray bar to the left of the track in the graphic or by clicking the name of the track in the bluebar group below the Browser graphic, above the track's pulldown menu.
- Look for "Display data as a density graph" and check the box. New options appear.
- Be sure the track height is set to its maximum.
- At the top of the page, set the visibility control to full and [Submit].
- There are a number display options to allow customization of the image. Experiment with more of them.

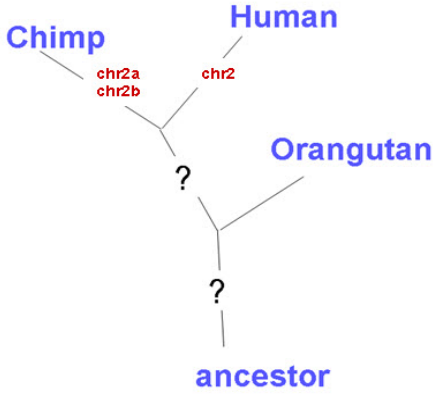
- Save a session with your favorite settings. You may want to check them out later when you are working with your own data.

5. Make a Custom Track out of RNA-seq data from the ENCODE project with expression levels above some threshold.

This problem shows you how to use a filter in the Table Browser and make a new dataset based on your values. It is useful for removing the noise or cherry-picking the best peaks from a dataset. The method works on your own RNA-seq data as well.

- Load the following session, which shows RNA-seq data from Cold Spring Harbor Lab as part of the ENCODE Project:
  user: example
  session: hg19_p53Rna
- What is the cutoff signal value for this track (in the left-side margin of the graphic)? That is, how high must be the signal before it runs off the top of the track and gets a pink hat?
- Click into the topmost of the four tracks and learn the name of the table for that track. Look for the tablename in the light blue bar at the top of the details page (wgEncode…).
- Go to the Table Browser using the Tool… pulldown and confirm it is looking at the proper table.
- Set a filter to only pick up signal greater than that shown in the graphic (essentially picking up signal represented by the pink caps on the peaks).
- Output custom track with a useful name and set visibility to full.
- Drag the track down so it sits atop the source track.
- Using the configuration button on the left side of the track, adjust the viewing height to the same scale as the parent track (2 data points per pixel).
- Result (new custom track):
  user: example
  session: hg19_p53RnaPeaks
- When you have completed this, save your session under "My Data > My Sessions," giving it some unique name that includes your initials.

6. Human chr2 shares homology with two different chimp chromosomes. Was there a fusion event in the human lineage? Or was there a split of the ancestral chr2 into two new chroms in the chimp lineage?



- [hide all] Type “chr2” in Position box. “Go.”
- In the Comparative Genomics section below the browser image, click on track control link, “Primate Chain/Net” to configure.
  set: Chains “hide” Nets “full” check Chimp only
  Maximum display mode: “full.” Submit.
- Your screen should look like this session:
  user: example
  session: hg19_chr2chimp
- Turn on the Gorilla and Orangutan Net subtracks. What do you conclude? Compare to:
  user: example
  session: hg19_chr2primates
- Zoom into the region on the long arm of chr2 where you see many small alignments clustered together (these are typically repeat regions or regions that have rearranged in one lineage or the other).
- With this region in your window on the graphical viewer, go the menus at the top of the page, “View… In Other Genomes (Convert)” and navigate to the chimp panTro5 genome. Select the best alignment at the top of the page of links: “chr2B:…..”
- Look at the chromosome ideogram above the main graphic: The region that has rearranged in human is the centromere region of chimp. What are the implications for evolution of the human and chimp lineages?
- When you have completed this, save your session under “My Data > My Sessions,” giving it some unique name that includes your initials.