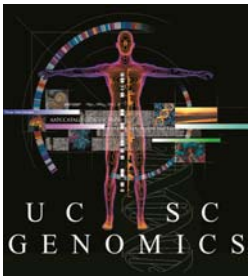
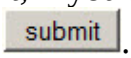


UCSC Genome Browser Advanced Workshop
UC Davis
Davis, CA

We will be using human assembly hg19. These problems will take you through a variety of resources at the UCSC Genome Browser. Some steps may seem a bit cryptic or terse. That is by design, so you will think about things as you go … and so you will 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: 

.

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 "Sessions." Login using:

username: davis2015
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.

1. Load a group of Custom Tracks from a URL.

If you attended the Intermediate workshop, load the session you saved at #9 and skip to the next question here. Otherwise:

- 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/ucDavis2015 and then click into ctExamples.txt.
- Copy the URL of this file into the upper text box (or you could copy the data) [Submit].
- Click the top "chr6" link the "Pos" column.
- 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.

2. Interpret a file of variants in pgSnp format using Variant Annotation Integrator (VAI).

Have a look at the SNPs in the "My SNP Custom Track -- Personal Genome SNP format" track from above. Zoom into a few of the SNPs at high resolution.

- Use the pulldown "Tools…" and select "Variant Annotation Integrator"
- In the "Select Variants" section, select from the pulldown: "My SNP Custom Track – Personal Genome SNP format."
- If you want actual genes names in the output, in "Select Genes" section, choose "Basic Gene Annotation set from GENCODE [latest version]."
- For now accept the default output options, but notice that the VAI is configurable.
- Click the "+" signs in the "Define Filters" section, just to see what's available.
- Select output format: "Variant Effect Predictor (HTML)" [get results]

3. Load a file of variants in VCF format and interpret using VAI.

- Back on the VAI page, in the "Upload Variants" section, click the link: "VCF."
- Scroll down the page and find Example One.
- Find the text with the link that says "click here for a text version" and click that. Or go to this page directly: http://genome.ucsc.edu/goldenPath/help/examples/vcfExampleOne.txt .
- Note the text on the page, "bigDataUrl=http:...." This identifies the location on the Web where the file is located. In practice, with your own VCF data, this would be your http: location.
- On the VAI page, click [add custom tracks] and paste either the contents of the vcfExampleOne.txt page or the URL itself. [Submit]
- Go back to VAI [go to variant annotation integrator] and look at the information about the variants as before.

4. Manipulate your own Custom Tracks.

- Go back to the ctExamples.txt file on http://bit.ly/ucDavis2015
- Copy the wiggle and BED12 "paired ends" custom track sections to a text editor.
- Play around with the data to add a few more items to each.
- Manipulate the colors in your text file. In the "paired ends" track, the color designation is item-by-item in column 9 ("itemRgb"). In the wiggle track, the color is in the track header.
- Load your file into the Browser using "add (or manage) custom tracks"

5. Check out one way a biological sample can differ from the reference assembly.

The reference assembly is just one chromosome of one person. This problem shows how to navigate to a new assembly and also shows that the reference is not necessarily the most common allele.

- Load session (pulldown: My data… , Sessions):
  user: example
  session: hg18_bamSnps
- This is a BAM file from a ChIP-seq experiment (chr21 only) – data courtesy Charles Nicolet, UC Davis (now at USC).
- After loading the session, zoom out 3x, then back in (observe display of mismatches).
- Note the homozygous T in the BAM display near the left side of the graphic.
- Click into the SNP below it – the variant in this sample has a record in dbSNP. dbSNP build 130 is old (alignments at that time were to hg18).
- Copy the rs# (reference SNP): rs2507733
- Go back to the Genome Browser using link in top bluebar.
- Use "View, In Other Genomes (Convert)" link, and go to hg19. [Submit]
- Note the info about how well the two assemblies match, and click the link.
- Turn on "All SNPs (142)" track in hg19.
- Find the rs2507733 SNP and click into it.
- Check out the Allele Frequencies. What do you think of that???? If you don't know what to think of it, let's talk.

6. Look at a region that has changed between genome assemblies.

- [hide all], then load the following Session on human assembly **hg18**:
  user: example
  session: hg18_4bp
- Look at each track very carefully. What does each track show? How do you interpret what you see?
- Now get the DNA from this region ("View… DNA" in top bluebar) and BLAT ("Tools" menu) onto **hg19**. Look at the alignment "details" first, then the go to the Browser via the link produced by BLAT.
- Turn on the Hg18 Diff track in the top track group. Zoom out 10x. How do you interpret this? Read track description of Hg18 Diff.

7. Make a Custom Track with a set of genes in it.

- Now obtain a list of genes:
  Go to http://bit.ly/ucDavis2015
  Click into the file "genelist" and copy the list
- Go to "Tools, Table Browser" and select
  group: Genes and … table: knownGene
  track: UCSC Genes region: genome
- Click "paste list" and paste the list and submit.
- Output format: custom track. "get output"
- Give it a short name and a longer description.
- Click "get custom track in Genome Browser"
- [hide all], then turn your track back on to "pack"
- Type a chromosome name into the position/search box, e.g. chr11

8. Make a spreadsheet of genes with quick links to the Browser.

- In a separate browser window, go to http://genomewiki.ucsc.edu and search for "excel."
- Click through to "Spreadsheet links to Genome Browser views."
- This page gives further details about how you might configure an Excel sheet yourself, but for now, download the spreadsheet template by clicking on the link to the right of the spreadsheet image: "ucscLinks.lxs."
- Copy the list of genes as in the previous question.
- In the Excel sheet you downloaded, paste the genelist into a column to the right of the existing content. Add genes of your choice to the bottom of the column, if you wish.
- In the middle of the spreadsheet, find the two formulas in the section marked "**gene**." Note the relationship between the gene name and the column headed "hg19" (two columns to the right of the gene name).
- Use the Copy command to copy this formula into the **same relationship** with the gene names you pasted into the sheet.
- Click into a few of them.

9. Obtain a list of SNPs from a single gene using the Table Browser. Show the human and chimp alleles for each.

- From the Table Browser pulldown menus, select group variation, track All SNPs 142
- Type some gene name (NF1? FGFR1?) into the Position box in the Table Browser, then hit "lookup"
- Output: selected fields from….

10. Compare UCSC Genes with Ensembl Genes using the intersection function in the Table Browser.

- Set Ensembl Genes as the main track.
- Click into "intersection."
- Choose secondary table as UCSC Genes.
- Set up "All Ensembl Genes that have no overlap with UCSC Genes."
- Export results as a custom track and check it out on the Browser.
- If you feel you have time, repeat this on hg38 (UCSC Genes has been renamed "GENCODE v22" on hg38)

11a. Check to see if Watson and Kriek have any SNPs in common (Table Browser intersection).

- [hide all]
- In the "Variation" track group near the bottom of the page, click into the "Genome Variants" track.
- Under "Select subtracks by project and variant types," hit the "minus" [–] button.
- At "List subtracks," select "all"
- Near the bottom of this long list, check the box and set both Watson and Kriek to "dense." [submit]

- Navigate to gene, FGFR1 (or some other gene of your choice). Zoom out 1.5x. Notice that some variants are common and some unique to one or the other.
- Turn off the FGFR1 splice variants using the configuration options (minibutton on the left side of the screen).
- Use the Table Browser to output a list of variants common between the two datasets, but first, find out the name of the tables involved.
- Click into any of the SNPs to open the track to full visibility, then click again to access the details page.
- "View Table schema" will now allow you to see the structure and the name of the table (pgKriek).
- Go directly from this page to the Table Browser in the Tools pulldown.
- The pgKriek table should be pre-selected in the Table Browser.
- Hit the "Intersection" [create] button.
- Select your secondary table: group variation; track Genome Variants; table, Watson.
- The choice should be preselected as default: "All pgKriek records that have any overlap with Genome Variants" (note that it does not identify the exact table as pgWatson, but we will be able to check out the output to confirm that it is correct.) [submit]
- Back in the Table Browser, choose output format: custom track.
- Name the short and longLabels: Watson+Kriek; SNPs common to Watson and Kriek and set visibility to "dense" [get custom track in genome browser]
- Set the visibility of Watson and Kriek back to dense and inspect your result.
- Your screen should look like this session:
  user: example
  session: hg19_watsonKriek

11b. Get DNA with Watson and Kriek variants marked on it.

- Highlight the region including 3 exons on the 3'-end of the gene
- Your screen should look like this session:
  user: example
  session: hg19_watsonKriek2
- Zoom into the highlighted region.
- In the pulldown menus, go to "View… DNA"
- Choose [extended case/color options]
- Set default to lower case
- Toggle case for All three datasets
- Set Watson+Kriek and Genome Variants to Bold
- Set Watson+Kriek to Red 255
- Set UCSC Genes to Green 255
- Set Genome Variants to Blue 255
- [submit]. The Variants may be hard to see, as they are single base-pairs. If so, go back to the Browser graphic and zoom in to a region with a small number of annotations.

12a. How many genes in the human genome are larger than one megabase?

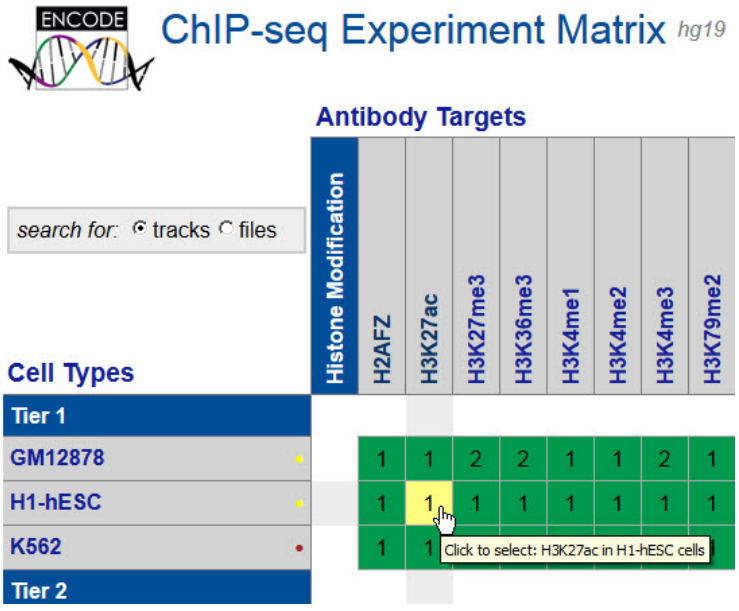
- Using the Table Browser, select the knownCanonical table from the UCSC Genes track.
- Be sure “genome” is selected rather than “position.”
- Set a filter with a MySQL “WHERE” (The “WHERE” is implied) using the “Free-form query” input box: Type “chromEnd - chromStart > 1000000”.

12b. How many spliced protein-coding genes are smaller than 1 kb?

- Using the Table Browser, select the knownGene table from the UCSC Genes track (knownCanonical will not let you filter out protein-coding genes)
- Be sure “genome” is selected rather than “position.”
- Set a filter: Type “exonCount > 1 AND cdsStart != cdsEnd AND txEnd – txStart < 1000”.
- What does each part of the filter do for your query (HINT: Each part of the quesry corresponds to one part of the title sentence – in order: spliced, protein-coding, small than….)?

13. Find regions with a significant amount of acetylation on Histone 3 Lysine 27 (H3K27ac) in embryonic stem cells (use any region near genes that interests you).

- Use the ENCODE Human Experiment Matrix (accessible first via Home page) to identify the proper dataset.
- Open the Histone ChIP-seq matrix.
- Find in the matrix the intersection of H1-hESC and the desired histone:



- Choose both tracks that appear: Signal and Peaks and compare them to each other.

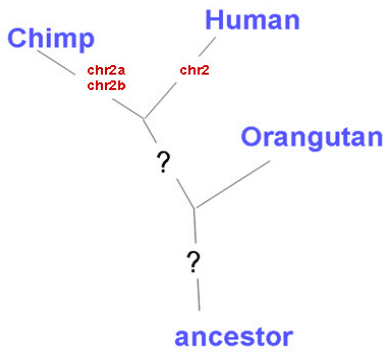
14. Make a Custom Track out of RNA-seq data from the ENCODE project that have expression levels above some threshold.

- Use the ENCODE data matrix to find RNA-seq data from Cold Spring Harbor Laboratory (CSHL).
- Turn on both replicates for Long poly-A+ RNA signal for both strands for K562 whole cells.
- [Hide all]. Go to gene TP53. Then zoom out 1.5x.
- Note the pink tops on the signal that extends above the level of 100 in the graphic.
- Use your right mouse button on each track to configure the track (one by one) to view them at 50 pixels high.
- You can see how this should look in this session:
  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. The Table Browser should be all set up with the proper group, track and table when you navigate there from this details page but look for the tablename in the light blue bar at the top of the details page.
- 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.
- Adjust the viewing height of the track in the graphic to roughly the same scale as the parent track and drag it so it sits atop the parent track.
- Result (new custom track):
  user: example
  session: hg19_p53RnaPeaks

15. Sex-linked colorblindness is a well-known and common condition. Use the mRNA from one of the opsin genes on chrX along with BLAT to explain why it is so common.

- [hide all]. Search for opsin, click into a gene, then click into a transcript.
- On the gene details page, click into the mRNA
- BLAT it back to hg19 and view in the Browser (“Tools…”).
- Zoom out 10x.
- Set the Segmental Dupes track in the Variation group to “pack.”

16. 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], then 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 Orangutan Net subtrack. 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).
- 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 genome.
- Look at the chromosome ideogram above the main graphic. What are the implications for evolution of the human and chimp lineages?

17. Use the Human Genome Diversity Project data to look at world-wide distribution of some SNP variants.

- [hide all]
- Turn on HGDP Allele Freq track (in Variation group) to full.
- Type gene name, wt1 into position/search box.
- Click into the first 5 (or more) SNPs in the wt1 gene…. Interesting? So is the 7th SNP: rs1569776.
- Click one of the links marked "Population key:" below the map.
- Also check out SNP rs6619104 (just type it into the position/search box). What do you make of the fact that there are two links in most of the SNP tracks you see?
- And rs2229989. And rs3775291.

18. If you were about to clone a piece of DNA from the wt1 gene and need to use an EcoRI site, how might you use the SNP tracks to be sure your sample would cut the way the reference assembly predicts? Which SNP track would you use? Why would you care about SNPs?

- Remember earlier when you exported a list of restriction sites from the details page? Try that, then make a Custom Track out of it.
- Then use the Table Browser to intersect the SNP track you chose with your track.

19. Autism track hub shows no differences in histone H3K4me3 modification between autistic and control brains.

Trimethylation of H3K4 is associated with active promotors. Some genes do show a difference between neuronal and non-neuronal (glial) cell controls. The authors report little difference in autistic brains when compared with control brains.

- [hide all]
- Go to "My data…" "Track Hubs" and [connect] to UMass ZHub (autistic brains)
- Back at hg19, type "autism" in the search box above the Browser graphic.
- Visit the first gene on the list, AUTS2 and note that there is little difference among the three data tracks.
- Click into the track description using the minibutton to the left of the track and read about the display. You should see that the blue dataset is non-neuronal (NeuN-, glia), the green is neuronal (NeuN+) and the orange is autistic brain. [submit] to return to the graphic page.
- Look at several of the other genes by searching with "autism" again. Try "autistic" to see if you get a different set. SEZ6L2 gives a nice difference between neuronal and glial cells. FOXP1 shows some tissue-specific alternative splicing.