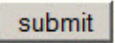# UCSC Genome Browser
Pittsburgh Workshop -- Practical Exercises

We will be using human assembly hg19. These problems will take you through a variety of resources at the UCSC Genome Browser. You will learn how to use the Table Browser, Custom Tracks, Saved Sessions and some navigation processes. 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: submit .

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:** pitt2014
> **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. Visit a session and observe the Publications track. This track was created by aligning sequences appearing in millions of publications.
- Use "Sessions" or "My data, Sessions" pulldown to get to the sessions page.
- Under "Restore Settings" type
    user: example
    session: hg19_GNRH1   [Submit]
- In the "Updated Session" section, click the "Browser" link.
- Observe this gene: 2 isoforms, one with intron in 5'-UTR
- Put your mouse over the items in the Publications track.
- Put your mouse over, then click into Khan2007 or Khan2008 in the middle of the screen (or really, any paper you see there).

1b. Zoom to the exon in the middle of the screen.
- Place your cursor in the Base Position section of the Browser graphic (where the scale bar is).
- Drag it to the right or left to highlight the area of the exon and release when you have the exon (and the Khan papers) centered.

1c. Zoom to codon numbering.
- Drag-and-zoom again to a window about 5 amino acids wide.
- Turn on the codon numbering by clicking the little minibutton to the left of the UCSC Genes track, then checking the "show codon numbering" checkbox.

1d. Jump to the end of the exon.
- Position your cursor over the double-headed arrows to the right and then the left of the two isoforms of GNRH1.
- Click the right arrowheads in the upper isoform and observe the start codon.

1e. Log in and save a session with the full exon.
- Zoom out using 10x button.
- Move the display to the right, using click-and-hold in the graphic and dragging the image, centering the exon.
- In "My data, sessions" log yourself in:
  - user: pitt2014
  - pwd: genome
- In "Save settings"
  - name: hg19_XX_gnrh1Exon   (or something else of your choosing)
  where "XX" is your initials (if someone else has used your initials, use "hg19_XX2_ gnrh1Exon2 or something else)
- Throughout this exercise, feel free to save any sessions you like or to visit the sessions created by others (click the "Browser" link on the right end of the line naming the session).

2a. Go to gene NR5A1 and turn on the "OMIM AV SNPs" track to see the biological function.
- Hide All (below the Browser graphic).
- Type the gene name into the position/search box and select it from the suggestions that appear.
- Turn on the "OMIM AV SNPs" track from the "Phenotype…" group
- Mouseover the SNPs in this track to see the phenotype.
- Click into one or more of them to read details.
- Click into "OMIM Allelic Variant" on details page to see the record at OMIM.
- Repeat this with gene, LHB – how many exons?

2b. Go to mouse homologue of NR5A1.
- Use "View,   In Other Genomes" and go to mouse mm10.
- Is the gene named the same in mouse? Many are not.

3. Find out where in the genome is the amino acid sequence FPVALSCRCGPCRRSTSDCGG.

- Use Blat from the "Tools" pulldown and select Human, hg19.
- Is it unique in the genome?
- Check out both the "Browser" and "details" links.
- From which strand is it transcribed?

4. Set a filter to display only the SNPs from snp141 that change an amino acid in NR5A1.

- Type the gene name into the Position/Search box. Submit.
- Turn on "All SNPs" track.
- Use minibutton on left to access configuration options.
- Scroll far down the page to the section that describes the color conventions used on the page. Note which SNP types are considered "non-synonymous."
- Scroll back up the page and open the Filtering Options.
- Select the classes of SNPs you would classify as non-synonymous.

5. Get a list of all missense SNPs in NR5A1 from the Table Browser into a Custom Track.

- In Table Browser ("Tools" pulldown menu),
      Choose group "Variation and…", track All SNPs(141)
- Type NR5A1 into "Position" box, then "lookup" and choose from list.
- Output format "selected fields from…"   [Get output]
- Choose checkboxes "name" and "func" and the coordinates. [Get output]
- To test that you have retrieved SNPs from the right place, grab an rs# and copy/paste or type into "position and search" box on the Genome Browser. Still in the same gene?
- Repeat this with new output format: "custom track."   [Get output]
- Make up meaningful names at the top. "name" = shortLabel (becomes link above track controls); "description" = longLabel (displayed in graphic)

6a. Use the UCSC Genes details page to access the OMIM record for Hirschsprung Disease.

- Type the disease name into the Position box in the Genome Browser. [go]
- Select the top hit in the list, ECE1 and view in the Browser.
- Click the gene in the graphic to see the details page.
- Find the OMIM link on the details page (in the section, "Sequence and links to…")
- Click through to OMIM and scan the page.

6b. Find Gene Reviews articles for genes related to Hirschsprung Disease.

- Back at the Genome Browser with ECE1 gene in the window, open the track, "GeneReviews" in the "Phenotype and Literature" group.
- Click into the item in the Gene Reviews track to see details.
- Click link, "Hirschsprung-ov" to find a review article written by an expert.

7. Find a list of genes involved in luteinization and pathway information about some of them.

- Type the word "luteinizing" into the Position/Search box on the Genome Browser.
- Find and click on the FSHB gene in the list (or another)
- Click into the gene in the Browser display.
- Click the "Pathways" link on the gene details page.
- Click into both of the links in the KEGG section.
- Click the Biocarta link, too.

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

- Using the Table Browser, find the table, knownCanonical (associated with UCSC Genes track in Genes group).
- Use "position" box and "lookup" button to get coords.
- output format: selected fields from …. [get output]
- Select chrom, chromStart, chromEnd from knownCanonical and geneSymbol and description from kgXref. [get output] [get output]

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

- From the Browser graphical viewer, navigate to "add custom tracks" (or "manage custom tracks" first).
- In a second Firefox (or Chome, or…) window, go to http://bit.ly/pitt2014 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.

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

- 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]

11. 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 <u>here</u> 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.

12. Find predicted sites for common restriction enzymes in the gene, NRP1.

- Type gene name in gene box in Genome Browser.
- In Mapping and Sequencing section below the graphic, click on link, "Restr Enzymes"
- Type "EcoRI, BamHI, HindIII," and turn the pulldown to "pack" [Submit].
- Click into your BamHI site in the graphic to see details.
- From the details page, export a list of all BamHI sites in the region using the link "just BamHI"
- Zoom into any BamHI item down to the base level.

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

- 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 you have loaded 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 been seen before and has been submitted to 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 on the new page about how well the two assemblies match, and click the link.
- Turn on "All SNPs (141)" track in hg19.
- Find the rs2507733 SNP and click into it.
- Check out the Allele Frequencies. What do you think of that????

14. 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.

15. 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."
- Now obtain a list of genes:
     Go to http://bit.ly/pitt2014
     Click into the file "genelist" and copy the list
- In the Excel sheet you downloaded, paste the genelist in 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 content, 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).
- 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.

16. Make a Custom Track with these genes in it.
- Copy the list of genes as above or from your spreadsheet.
- 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. chr5
- If you wish to have regular HGNC gene names in your track, use the process from Q8 above to generate a BED4 representation of the data.

17. Make primers that will amplify a fragment from mRNA, but not from DNA, using the large exon and one neighboring exon in the middle of the gene, BRCA1.

- Type the gene name into the gene box.
- Zoom in to include the large exon + a neighbor or two.
- Go to "View…   DNA" in the blue bar.
- Click "Use extended case/color options." To highlight the exons,
      choose "lower case" for default.
      in UCSC Genes row, check "toggle case," and "bold."
      type 255 in your color of choice.
- Open a second window or tab and go to "Tools, In-silico PCR"
- Copy/Paste ~25-30 bp primers from the exons onto PCR page in the separate window.
      Choose "Target: UCSC Genes" and remember to "Flip reverse primer"
- Try the primers on Mouse assembly mm10.

18. 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?
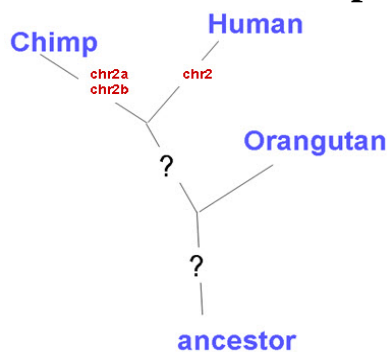
- Hide All, then type "chr2" in Position box. "Go."
- In the Comparative Genomics section below the browser image, click on track control link, "Primate Chain/Net" to configure.
      set: Chains "hide"   Nets "full"   check Chimp only
      Maximum display mode: "full."   Submit.
- Turn on Orangutan Net subtrack. What do you conclude?

19. Obtain a list of SNPs from a single gene (try RET, or some gene you like better) using the Table Browser. Show the human and chimp alleles for each.

- From the Table Browser pulldown menus, select group variation, track All SNPs 141
- Type the gene name into the Position box in the Table Browser, then hit "lookup"
- Output: selected fields from….

20. 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."

21. Are there any Flagged SNPs (141) on chr1 that are not in exons?

- If you find none, how could you check that your method was correct and not just a bug in your process?

22. 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.

23. View the table schema for two tracks you have used: knownGene (UCSC Genes track) and hgdpGeo (HGDP Allele Freq track).

- Go to Table Browser.
- Find the table using the appropriate pulldown menus (group, then track, then table).
- Click the "view table schema" button next to the table pulldown selector.

24. 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.

25. Make a Custom Track that covers the entire footprint of the transcripts from your genelist, rather than exons only.

- Load the list into the Table Browser as before:
    Select group genes, track UCSC Genes, table knownCanonical
    Either paste the genelist or upload it from your machine.
- Be sure that "genome" is selected in "region."
- Output format: "BED – browser extensible data"
- Output file: genelistCoords.txt (or something you like better)
- Get output
- Select chrom, chromStart, chromEnd
- Get output.
- Check "include header." Give it a useful name (= shortLabel for track control) and description (= longLabel for graphic image title).
- Get BED
- Upload the file as a custom track (CT).
- Check it out in the Browser.

26. 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 but UCSC Genes.
- Get the mRNA from an opsin gene on chrX.
- BLAT it back to hg19 and view in the Browser.
- Zoom out 10x.
- Turn on the Segmental Dupes track in the Variation group.

27. Get a list of all HGDP SNPs that are within the transcripts (introns or exons) of the genes in your genelist and make Excel spreadsheet links for them.

- Use the Table Browser to intersect the hgdpGeo table with your new CT.
- Find the hgdpGeo table in the Variation group.
- Create intersection.
- Choose your CT as the secondary table.
- Choose "All HGDP Allele Freq records that have any overlap with genelistCoords." Submit.
- Choose output format: "BED…"
- Output file: genelistGeoSnps.txt or whatever [get output]
- Uncheck "include header" [get BED].   Save file.
- Open file from within Excel. Be sure you get the input in separate columns (Delimited).
- Get the formula from the Excel sheet from before for rs#s and copy it to the appropriate column in the new spreadsheet. Make sure the formula has the **same relationship** to the rs# in the new sheet as it did in the old. That is, the hg19 formula should be two cells to the right of the rs#.

28. 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.

29. 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 track to learn the name of the table for the topmost of the four tracks. You will need to also notice the track group and track name in the controls below the Browser graphic to find the table.
- Go to the Table Browser and set it to look at this 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