Using the UCSC Genome Browser to Interpret Genetic Disease

Robert M. Kuhn, Angie S. Hinrichs, Fan Hsu, Ann S. Zweig, Galt Barber, David Haussler[1] and W. James Kent
Center for Biomolecular Science and Engineering and [1]Howard Hughes Medical Institute, University of California Santa Cruz (UCSC)

UCSC Genome Browser - http://genome.ucsc.edu

Both clinical geneticists and research scientists are faced with increasing amounts of data to evaluate.

Clinical applications

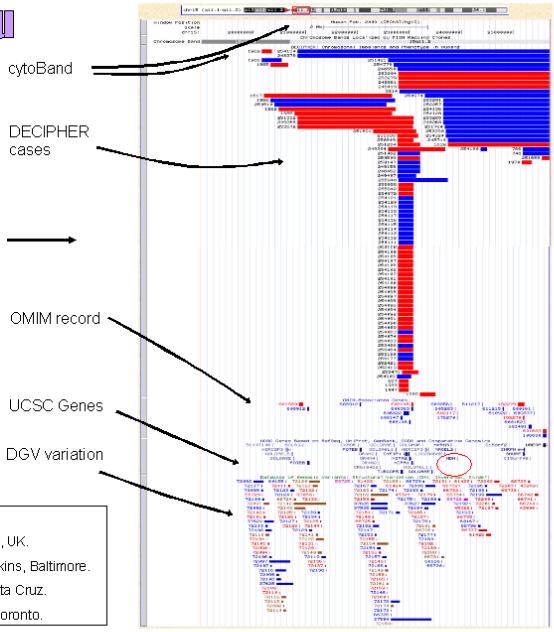
In the clinic, it is necessary to assess whether a chromosomal imbalance is responsible for the clinical symptoms. The clinician needs to evaluate known information about normal variation and clinically significant copy-number changes in an ever-changing informational landscape.

The UCSC Genome Browser allows information from a wide variety of sources to be displayed together in one view at any genomic scale, from whole-chromosome down to single-base-pair resolution. It is possible to display your own cytogenomic data (either from microscopy or CGH arrays) in Custom Tracks, for easy-to-interpret graphical viewing of the data.

Data views may be saved in Sessions and easily shared.

Juxtapose a large number of database resources in one place, AT ANY CHROMOSOMAL RESOLUTION.

E.g., type in "15q11" and select appropriate tracks to view all relevant data aligned by genomic position.



Research applications

Research applications are increasingly using whole-genome or exome sequencing to evaluate patient data. The same advantages apply to visualizing seq data in the context of other Browser tracks as apply to CGH data, with the advantage that the data are at base-pair resolution.

File formats for display of large datasets (bigBed, bigWig and BAM) directly in the browser make visualization easier than ever.

Data redistributed from:

DECIPHER – Nigel Carter, Sanger, UK.

OMIM – Ada Hamosh, Johns Hopkins, Baltimore.

UCSC Genes – Jim Kent, UC, Santa Cruz.

DGV – Steve Scherer, Sick Kids, Toronto.

Data Tracks

cytoBand

Chromosome band information shows landmarks as seen in microscope.

DECIPHER

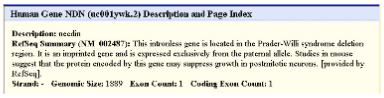
Data from patients around the world. Links on Genome Browser take you back to the DECIPHER record with much more information about phenotype, etc., including photographs.

OMIM

Comprehensive annotation of genetic and molecular data, excellent portal for access to the literature.

UCSC Genes

Gene models. Click-through to details page (red oval above) gives access to mountain of data, including phenotypes, molecular and biological functions, links to orthologs in model organisms and more.



DGV variation

Database of Genomic Variants. Data from numerous large studies. A compendium of benign large deletions and duplications. Helps identify likely candidate genes if trying to identify the source of a phenotype.

SNPs, Arrays (not shown above)

Complete mapping of SNPs from dbSNP (including 1000 genomes).

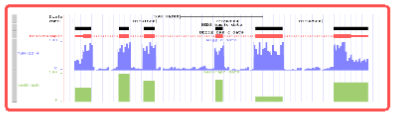
Probe mappings from many common microarray platforms.

Custom Tracks

Load your own data alongside hosted data tracks

Custom tracks are user-supplied data that can be privately uploaded to the Genome Browser. Any data that maps with genomic coordinates can be displayed down to base-pair resolution.

Multiple formats allow simple block display, gene-structures or plots of values mapped to coordinates.



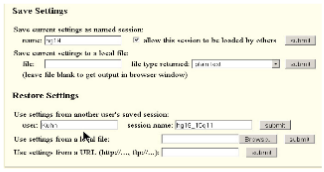
Sharing Sessions

Share your insight with colleagues

One you have configured the Browser to show exactly the regions and data tracks that make your point (including Custom Tracks, if any), save it as a session to share with colleagues. Load the session using the interface shown here or send a URL and a single click opens the session.

Create a separate session for each patient

Each patient record can be accompanied by a separate Session. While we do everything we can to preserve the data indefinitely, it would be safest to back this up in a file saved on your local system.



High-throughput Sequence Data

How do you visualize your Next-Gen Sequencing data?

Single-read data processed from Next-Generation Sequencing (NGS) machines generate millions of short reads which can be aligned to the reference genome using a variety of tools.

Raw Reads → aligner / SAMtools → SAM / BAM → index → .bai → Custom Track

Using the SAMtools suite, the alignments are converted to BAM format (Binary SAM files), often with millions of alignments to the reference assembly. BAM alignment files from NGS sequencing runs are far too large to upload as Custom Tracks without a connection timeout. Even files that do not timeout can take a very long time to upload.

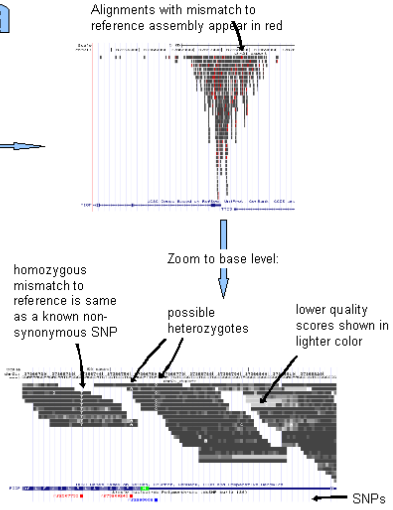
Solution to the problem of displaying HUGE data file:

1. Convert aligner output to SAM format.
2. Index the file using SAMtools to create a .bai file.
3. Make both files available to the web (http:, https: or ftp:).
4. Create the Custom Track by uploading the location of the data.

   track name=trackName type=bam bigDataUrl=http://path/file.bam
5. The Browser fetches only the portion of the file it needs to make the display.
6. Interact with the track yourself at: http://genome.ucsc.edu/cgi-bin/hgTracks?

   hgS_doOtherUser=submit&hgS_otherUserName=Example&
   hgS_otherUserSessionName=hg18_hiThruput

Sample track data courtesy Charles Nicolet, UC Davis



More Information

Click "**Help**" link to get tool-specific help pages.

Search for answers to questions:
http://genome.ucsc.edu/contacts.html

Or **email** your question to the actively monitored public list:
genome@soe.ucsc.edu

OpenHelix provides free training material:
http://www.openhelix.com/downloads/ucsc/ucsc_home.shtml
and also offers training seminars (some free or discounted).

Acknowledgements

•This work was funded by National Human Genome Research Institute (2 P41 HG002371-06 to UCSC Center for Genomic Science, 3 P41 HG002371-06S1 ENCODE supplement to UCSC Center for Genomic Science); National Cancer Institute (Contract No. N01-CO-12400 for Mammalian Gene Collection).

• Center for Biomolecular Science and Engineering, University of California Santa Cruz (UCSC). Santa Cruz, CA, 95064, USA.

Reference

The UCSC Genome Browser Database: update 2011. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ. Nucleic Acids Res. 2011 Jan;39(Database issue):D876-82.

Genomic Disorders, 2011. Hinxton, UK. March 23-25, 2011