593

UCSC Genome Browser and Variant Annotation Integrator

Robert M. Kuhn, Angie S. Hinrichs, Ann Zweig, Matt Speir, Brian Raney, Brian T. Lee, Steve Heitner, Maximillian Haeussler, Luvina Guruvadoo, Pauline Fujita, Hiram Clawson, Jonathan Caspar, Donna Karolchik, Galt Barber and W. James Kent

UC Santa Cruz Genomics Institute, University of California Santa Cruz (UCSC)

Baskin Engineering UC SANTA CRUZ

UCSC GENOMICS

Genome Browser - Displaying variants

http://genome.ucsc.edu

Display variants in the Genome Browser

NORMAL VARIATION

Normal variation is contained in these data tracks:
- Common SNPs
- 1000 Genomes
- Personal Genome Variants
- Exome Variant Server
- Database of Genomic Variants (DGV)
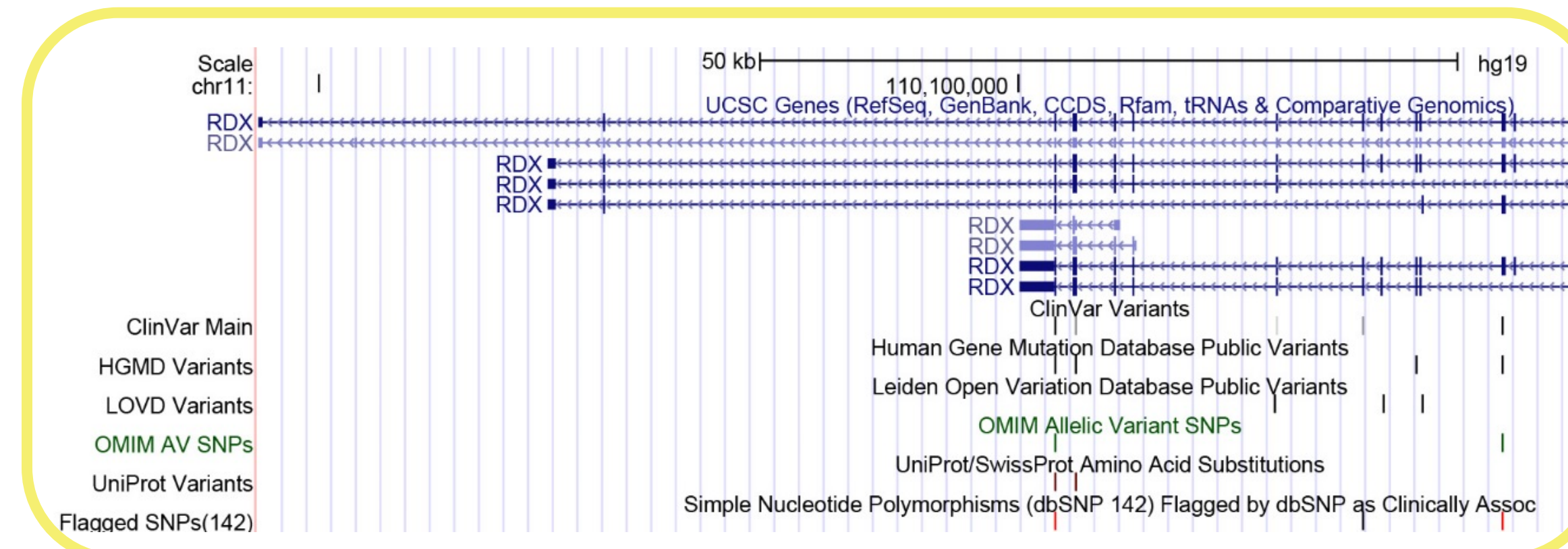
Coming: ExAC 66K exomes data

DISEASE-ASSOCIATED

Variation associated with disease is viewed by aggregating multiple resources:
- ClinVar
- Human Mutation Database (HGMD – public version only)
- Leiden Open Variation Database (LOVD)
- Online Mendelian Inheritance in Man (OMIM AV SNPs)
- UniProt (amino-acid-level data only)
- dbSNP (Clinical-flag SNPs)

- Load these views as live sessions:
- user Example
  - session hg19_variants3normal
  - session hg19_variants3

Variant Annotation Integrator

Determines molecular consequence of variants

Configurable using different annotation sets (genesets, regulation, etc)

1. Upload file in pgSNP format   (or post VCF file on the Web)

```
track type=pgSnp db=hg19 name="pgSnp"
description="My SNP Custom Track -- Personal Genome SNP format"
 visibility=pack
# pgSNP format:
# chrom chromStart chromEnd name alleleCount alleleFreq alleleScores
chr6   31691111   31691112   G/A     2   30,60   90,70
chr6   31691167   31691168   T/C     2   20,70   90,70
chr6   31691181   31691182   T/G     2   20,70   90,70
chr6   31691277   31691278   T/G/A   3   9,60,7  80,80,30
```

2. Find VAI in the Tools menu.

3. Select annotation filters

4. Obtain a molecular prediction for each transcript/isoform: gene, nucleotide and amino-acid substitutions, co-located SNPs, SIFT, PolyPhen…

Track Data Hubs

Why Track Data Hubs?

It is useful to be able to see your own data on the Genome Browser. For small datasets, direct upload to the Browser is straightforward.

Large datasets, however, cause a timeout during upload and cannot be displayed. This includes whole-exome variant sets.

Track Hubs allow the user to host the data on an http:-accessible site, where the Genome Browser can pull only data needed for display.

To use the Variant Annotation Integrator (VAI) on a large file, the data must be in VCF format (files in pgSnp format may be uploaded directly to the Browser).

Details for working with VCF:
http://genome.ucsc.edu/goldenPath/help/vcf.html

Other data types supported by track hubs: bigBed, bigwig, BAM, bigGenePred

Genome Browser-in-a-Box

Bring the Browser to your desktop – free for academic use

Download an image of the Browser and build a Virtual Machine on your desktop, right next to your data. Our data will be fetched from the Web as needed.

YOUR data never leaves the house.

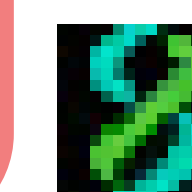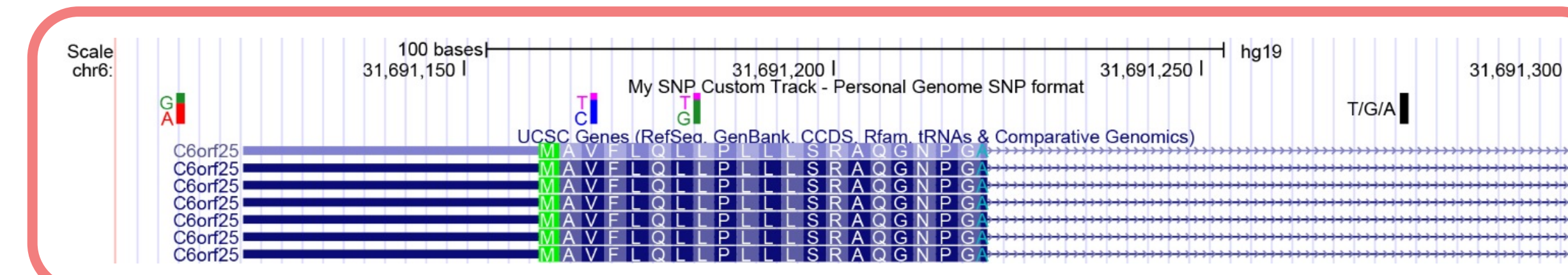GBiB runs on Windows, Mac or Linux systems via a Virtual Machine (VM).

It reverses the paradigm of data handling – most of the standard Browser data are loaded as needed through the Web, while your own data are treated locally rather than uploaded to the Browser.

1. Download and install a virtual machine to your system: virtualbox.org
2. Download the GBiB from genone-store.ucsc.edu
3. Install the GBiB – requires about 20 Gb of disk space for read/write, but occupies only 7 Gb as downloaded.

You can make local copies of Browser data that you use regularly to increase speed.

Learn more about the GBiB:

http://genome.ucsc.edu/goldenPath/help/gbib.html

Future Directions

New Data Integrator

E.g., All SNPs in a chosen group of genes. An expansion of Table Browser, the Data Integrator will allow you to keep both the gene names and the SNP names that go with them.

Exon-only display

Slices the genome into a display showing only exons.

… or showing alt-sequences from hg38

… or chromosome rearrangements

More Information

Click "**Help**" link (upper right) to get tool-specific help pages.

Click the "**Contact Us**" link (lower left of home page) to **search** for answers in the mailing-list archives, or email your question to genome@soe.ucsc.edu, our actively monitored public list.

OpenHelix provides free training material:
http://www.openhelix.com/downloads/ucsc/ucsc_home.shtml.

New! YouTube channel: UCSC Genome Browser

@GenomeBrowser   GenomeBrowser

workshops:
http://genome.ucsc.edu/training

Download this poster: http://bit.ly/acmg2015

Reference

The UCSC Genome Browser Database: 2015 update.

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ. *Nucleic Acids Res.* 2015 Jan 28;43(Database issue):D670-81. doi: 10.1093/nar/gku1177. Epub 2014 Nov 26. PMID: 25428374

ACMG 2015.   Salt Lake City.   March 25-27, 2015