UCSC Genome Browser

Bingbing Yuan
May 20, 2010
How many isoforms AGBL5 has?
Comprehensive View
Genome Browser

- Ensembl Browser: http://www.ensembl.org/index.html
  - BLAST/BLAT: low similarity
  - Genetic Variation
- UCSC Browser: http://genome.ucsc.edu/
  - Easy to use
  - lots of tracks
- GBrowse: http://gmod.org/wiki/Gbrowse
  - private data
  - long setup & maintenance time
- IGV: http://www.broadinstitute.org/igv/
  - easy navigation
  - good for large data analysis
Mining Data from Genome Browsers 2010 by Tyra Wolfsberg from NHGRI
# Human (Homo sapiens) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz. Software Copyright (c) The Regents of the University of California. All rights reserved.

<table>
<thead>
<tr>
<th>clade</th>
<th>genome</th>
<th>assembly</th>
<th>position or search term</th>
<th>gene</th>
<th>image width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammal</td>
<td>Human</td>
<td>Mar. 2006 (NCBI36/mm9)</td>
<td>chr21:36,432,011 35,438,125</td>
<td></td>
<td>800</td>
</tr>
</tbody>
</table>

Click here to reset the browser user interface settings to their defaults.

- manage custom tracks
- configure tracks and display
- clear position

## Request:

| chr7 | Displays all of chromosome 7 |
| 20p13 | Displays region for band p13 on chr 20 |
| chr3:11000000 | Displays first million bases of chr 3, counting from p-arm telomere |
| chr3:10000000+2000 | Displays a region of chr3 that spans 2000 bases, starting with position 1000000 |
| RH18061:RH80175 15q11,15q13 | Displays region between STS markers RH18061 and RH80175 or chromosome bands 15q11 to 15q13. This syntax may also be used for other range queries, such as between uniquely-determined ESTs, mRNAs, refSeqs, etc. |
| D16S3046 | Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well. |
| AA205474 | Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17 |
| AC008101 | Displays region of clone with GenBank accession AC008101 |
| AF083811 | Displays region of mRNA with GenBank accession number AF083811 |
| PRNP | Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP |
| NM_017414 | Displays the region of genome with RefSeq identifier NM_017414 |
| NP_059110 | Displays the region of genome with protein accession number NP_059110 |
| pseudogene mRNA | Lists transcribed pseudogenes, but not cDNAs |
| homeobox | Lists mRNAs for homeobox genes |
| zinc finger | Lists many zinc finger mRNAs |
Genome Viewer

Tracks (group of data)
Configure Image

<table>
<thead>
<tr>
<th>Button</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>submit</td>
<td></td>
</tr>
<tr>
<td>image width</td>
<td>800 pixels</td>
</tr>
<tr>
<td>label area</td>
<td>width: 17 characters</td>
</tr>
<tr>
<td>text size</td>
<td>8</td>
</tr>
</tbody>
</table>

- Display chromosome ideogram above main graphic
- Show light blue vertical guidelines
- Display labels to the left of items in tracks
- Display description above each track
- Show track controls under main graphic
- Next/previous item navigation
- Next/previous exon navigation
- Enable track re-ordering
- Enable advanced javascript features
mode of an individual annotation track:

Hide: the track is not displayed at all.

Dense: the track is displayed with all features collapsed into a single line.

Squish: the track is displayed with each annotation feature shown separately, but at 50% the height of full mode. Features are unlabeled.

Pack: the track is displayed with each annotation feature shown separately and labeled.

Full: the track is displayed with each annotation feature on a separate line.
## Neandertal Assembly and Analysis

### Variation and Repeats

<table>
<thead>
<tr>
<th>Track Setting</th>
<th>Common Cell CNV</th>
<th>GIS DNA PET</th>
<th>SNPs (130)</th>
<th>SNPs (129)</th>
<th>SNPs (128)</th>
<th>SNPs (126)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hide</td>
<td>hide</td>
<td>dense</td>
<td>hide</td>
<td>hide</td>
<td>hide</td>
</tr>
<tr>
<td>SNP Arrays</td>
<td>dense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HapMap SNPs</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural Var</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple Repeats</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HapMap LD Phased</td>
<td></td>
<td>HGDP Allele Freq</td>
<td>HGDP Smoothd FST [No data-chrX]</td>
<td>HGDP Hetzygeny [No data-chrX]</td>
<td>HGDP iHS hide</td>
<td>HGDP XP EHH [No data-chrX]</td>
</tr>
<tr>
<td>Exapted Repeats</td>
<td>hide</td>
<td>RepeatMasker</td>
<td>dense</td>
<td>hide</td>
<td>hide</td>
<td>hide</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>hide</td>
<td>Self Chain</td>
<td>Genome Variants</td>
<td>hide</td>
<td>hide</td>
<td>hide</td>
</tr>
<tr>
<td>Tajima's D SNPs</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tajima's D</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGSV Discordant</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrupted Rpts</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmental Dups</td>
<td>hide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Pilot ENCODE Regions and Genes

## Pilot ENCODE Transcription
Simple Nucleotide Polymorphisms (dbSNP build 130)

Display mode: [dense] [Submit]

Include Chimp state and observed human alleles in name: [ ]
(If enabled, chimp allele is displayed first, then '>', then human alleles).

On details page, show function and coding differences relative to: [Set all] [Clear all]
- UCSC Genes
- Old UCSC Genes
- Gencode Manual
- Gencode Auto
- Gencode PolyA
- CCDS
- RefSeq Genes
- Other RefSeq
- Vega
- SIB Genes
- Gene
- Augustus

SNP Feature for Color Specification: [Function] Set defaults

The selected feature above has the following values below. For each value, a selection of colors is available. If a SNP has more than one of these properties, resulting in more than one color, then the stronger color will override the weaker color. In order from strongest to weakest, the colors are red, green, blue, gray, black.

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Locus</th>
<th>Coding - Synonymous</th>
<th>Coding - Non-Synonymous</th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td></td>
<td>green</td>
<td>red</td>
</tr>
</tbody>
</table>

View table schema

Data last updated: 2009-08-18

Description

This track contains information about single nucleotide polymorphisms and small insertions and deletions (indels) — collectively Simple Nucleotide Polymorphisms — from dbSNP build 130, available from ftp.neb.nlm.nih.gov/snp.

Interpreting and Configuring the Graphical Display

Variants are shown as single tick marks at most zoom levels. When viewing the track at or near base-level resolution, the displayed width of the SNP corresponds to the width of the variant in the reference sequence. Insertions are indicated by a single tick mark displayed between two nucleotides, single nucleotide polymorphisms are displayed as the width of a single base, and multiple nucleotide variants are represented by a block that spans two or more bases.
Simple Nucleotide Polymorphisms (dbSNP build 130)

**dbSNP build 130 rs4987121**

**dbSNP:** rs4987121  
**Position:** chr21:36440549-36440549  
**Band:** 21q22.12

- **Chimp allele:** A  
- **Chimp strand:** +  
- **Chimp position:** ch
- **Orangutan allele:** A  
- **Orangutan strand:** +  
- **Orangutan position:** ch
- **Macaque allele:** A  
- **Macaque strand:** -  
- **Macaque position:** ch

**Class:** single  
**Validation:** by-cluster, by-frequency, by-hapmap  
**Function:** missense  
**Molecule Type:** genomic  
**Average Heterozygosity:** 0.015 +/- 0.084  
**Weight:** 1

**Coding annotations by dbSNP:**  
NM_001236: missense M (ATG) --> L (TTG)

**UCSC's predicted function relative to selected gene tracks:**  
UCSC Genes AB004851 (uc002yvf.1) intron  
UCSC Genes CBR3 (uc002yve.1) missense M (ATG) --> L (TTG)  
UCSC Genes BC047014 (uc002yvd.1) intron  
UCSC Genes CR594732 (uc002yve.1) intron

**HapMap SNP**

**Mappings to PDB protein structures**

- 2hrb  
- X-Ray  
- LS-SNP  
- Chimera

**Chimera help**
Direct to add track site
Manage Custom Tracks

**genome:** Human  **assembly:** Mar. 2006 (NCBI36/hg18) [hg18]

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
<th>Doc</th>
<th>Items</th>
<th>Pos</th>
<th>delete</th>
<th>add custom tracks</th>
<th>go to genome browser</th>
<th>go to table browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Track</td>
<td>User Supplied Track</td>
<td>bed</td>
<td>352191</td>
<td>chr16</td>
<td>□</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Add Custom Tracks

- clade [Mammal]
- genome [Human]
- assembly [Mar. 2006 (NCBI36/hg18)]

Display your own data as custom annotation tracks in the browser. Data must be formatted in **BED**, **bigBed**, **BEDGRAPH**, **GFF**, **GTF**, **WIG**, **bigWig**, **MAF**, **BAM** or **PSL** formats. To configure the display, set **track** and **browser** line attributes as described in the **User’s Guide**. URLs for data in the bigBed and bigWig formats must be embedded in a track line in the box below. Publicly available custom tracks are listed [here](#). Examples are [here](#).

Paste URLs or data: 
Or upload: 

Optional track documentation: 
Or upload: 

Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.
Extract DNA sequence
Genome browser

>chr22:34304505+34304954 450bp TAACAGATTGATGATGATGAAATGGG CCCATGAGTGCTCCTAAAGCAGCTGC
TtACAGATTGATGATGATGAAATGGGgggtggccagggtgggggggtga
gaccttcagaggaagccagggctgtggttcataacaaacgttgcgtcacaat
tatgacacqctgaagttttcagggctgtggtgagcccaagtggagaaggtaag
tacacagaacatcctagagaacccctacatcttagaagatgaatcataaattaatatag

gaccttgctgctgcgtagaagggatgttgattatctctatatgttagaataattcttgta
tccagaatgggttaaccacacacatgtgcgaagatgtgtgaacgtaatctcaca
agcaagctctctcctcagcagagaaacaccagccgctacaggaagcaag
aaatggtgctcaatatttaaggtgaatccacagacccagatgtcagagctcc
aagcactttgctctcagctccacgCAAGCTGCTTTTAGAGCACCACTGATGaG

Primer Melting Temperatures

Forward: 66.7 °C taacagattgatgatgatgaaatggg
Reverse: 73.8 °C ccacatgagtgccctctaaacagcagcgtg

The temperature calculations are done assuming 50 mM salt and 50 nM annealing oligo concentration. The code to calculate the melting temp comes from Primer3.
Encyclopedia of DNA Elements (ENCODE) Consortium
Founded by National Human Genome Research Institute (NHGRI)
ENCODEx Histone Modifications by Broad Institute ChIP-seq

Select views (help):
- Peaks
- Signal
- Regions

Signal Configuration
- Type of graph: bar
- Track height: 32 pixels (range: 16 to 100)
- Vertical viewing range: min: 1, max: 20 (range: 0 to 15647)
- Data view scaling: use vertical viewing range setting
- Always include zero: OFF
- Transform function: Transform data points by: NONE
- Windowing function: mean
- Smoothing window: OFF
- Draw y indicator lines: at y = 0 OFF

Select subtracks by cell line and antibody:

<table>
<thead>
<tr>
<th>Antibody</th>
<th>GM12878</th>
<th>H1-hESC</th>
<th>HepG2</th>
<th>HMEC</th>
<th>HSMM</th>
<th>HUVEC</th>
<th>K562</th>
<th>NHEK</th>
<th>NHIK</th>
<th>Cell Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCF</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - CTCF</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K4me1</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K4me2</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K4me3</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K9ac</td>
</tr>
<tr>
<td>H3K9me1</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K9me1</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K27ac</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K27me3</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
<td>+ - H3K36me3</td>
</tr>
</tbody>
</table>
### List subtracks:

- **only selected/visible**
- **all**

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Antibody</th>
<th>Views</th>
<th>Restricted Until</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM12878</td>
<td>CTCF</td>
<td>Peaks</td>
<td>ENCODE Histone Mods, Broad ChIP-seq Peaks (CTCF, GM12878)</td>
</tr>
<tr>
<td>GM12878</td>
<td>CTCF</td>
<td>Signal</td>
<td>ENCODE Histone Mods, Broad ChIP-seq Signal (CTCF, GM12878)</td>
</tr>
<tr>
<td>GM12878</td>
<td>H3K4me1</td>
<td>Peaks</td>
<td>ENCODE Histone Mods, Broad ChIP-seq Peaks (H3K4me1, GM12878)</td>
</tr>
<tr>
<td>GM12878</td>
<td>H3K4me1</td>
<td>Signal</td>
<td>ENCODE Histone Mods, Broad ChIP-seq Signal (H3K4me1, GM12878)</td>
</tr>
<tr>
<td>GM12878</td>
<td>H3K4me2</td>
<td>Peaks</td>
<td>ENCODE Histone Mods, Broad ChIP-seq Peaks (H3K4me2, GM12878)</td>
</tr>
</tbody>
</table>

- NHLF H4K20me1 Signal ENCODE Histone Mods, Broad ChIP-seq Signal (H4K20me1, NHLF) | 2010-06-28 |
- NHLF Input Control Signal ENCODE Histone Mods, Broad ChIP-seq Signal (NHLF control) | 2010-06-29 |

24 of 177 selected

Submit

Downloads

**Data version:** through the ENCODE Jan 2010 Freeze

### RESTRICTED

<table>
<thead>
<tr>
<th>until</th>
<th>File</th>
<th>Size</th>
<th>Submitted</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-10-05</td>
<td>wgEncodeBroadChIPSeqPeaksGm12878Ctcf broadPeak.gz</td>
<td>366K</td>
<td>2009-01-05</td>
<td>cell=GM12878, dataType=ChIPSeq, antibody=CTCF, lab=Broad, type=Peak</td>
</tr>
<tr>
<td>2009-10-05</td>
<td>wgEncodeBroadChIPSeqRawDataRep1Gm12878Ctcf.fastq.gz</td>
<td>774M</td>
<td>2009-01-05</td>
<td>cell=GM12878, dataType=ChIPSeq, antibody=CTCF, lab=Broad, type=Replicate</td>
</tr>
<tr>
<td>2009-10-05</td>
<td>wgEncodeBroadChIPSeqRawDataRep2Gm12878Ctcf.fastq.gz</td>
<td>517M</td>
<td>2009-01-05</td>
<td>cell=GM12878, dataType=ChIPSeq, antibody=CTCF, lab=Broad, type=Replicate</td>
</tr>
<tr>
<td>2009-10-05</td>
<td>R Bibi H3K4me1 ChIP-Seq ChIP-Seq H3K4me1</td>
<td>232M</td>
<td>2009-01-05</td>
<td>cell=GM12878, dataType=ChIPSeq, antibody=H3K4me1, lab=ChIP-Seq, type=Replicate</td>
</tr>
</tbody>
</table>
References

OpenHelix: http://www.openhelix.com/ucsc/

MIT library: Bioinformatics Tutorial Series
http://libguides.mit.edu/bits

Tyra Wolfsberg, Ph.D. Current Topics in Genome Analysis 2010
http://www.youtube.com/watch?v=7BN0T7AQqmY


UCSC Genome Browser Wiki site:
http://genomewiki.ucsc.edu/index.php/Main_Page