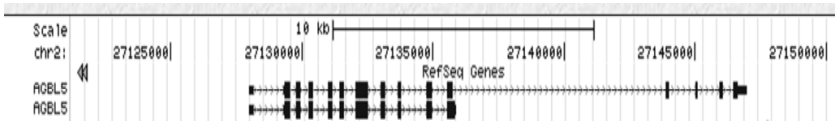
UCSC Genome Browser

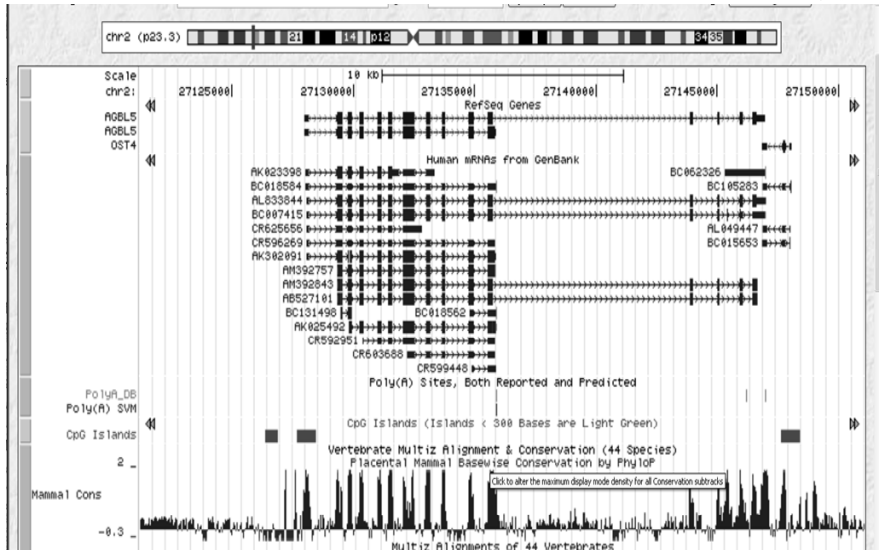
Bingbing Yuan

May 20, 2010

How many isoforms AGBL5 has?

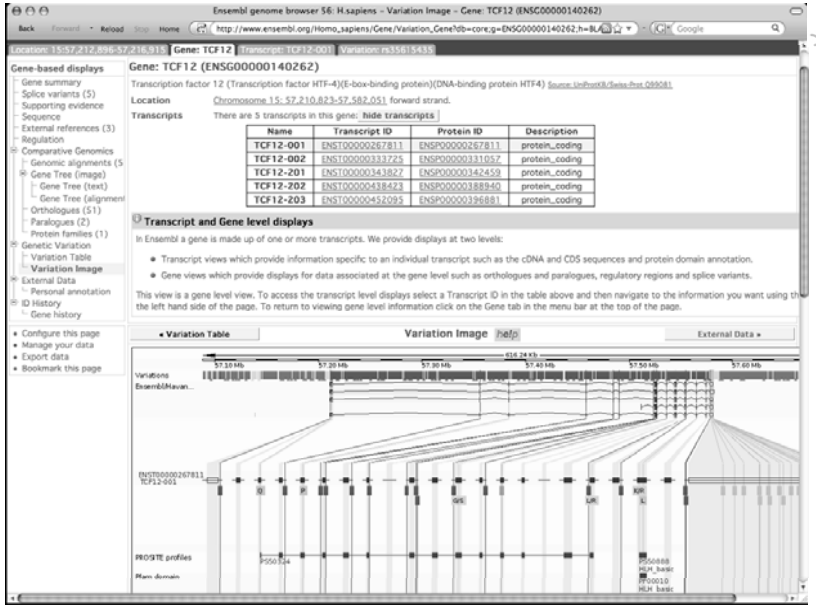


Comprehensive View



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

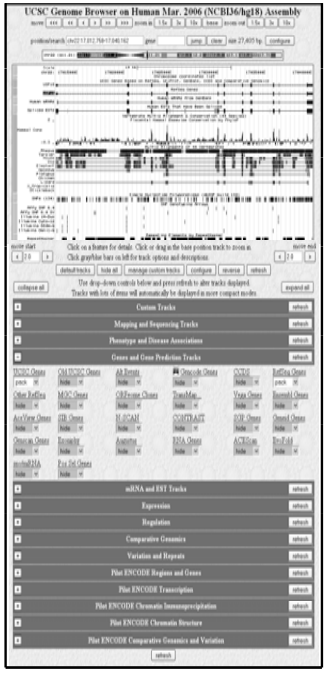
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade: Mammal | genome: Human | assembly: Mar. 2006 (NCBI36/hg18) | position or search term: chr21:36,432,011-36,436,125 | gene | image width: 800 | submit

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

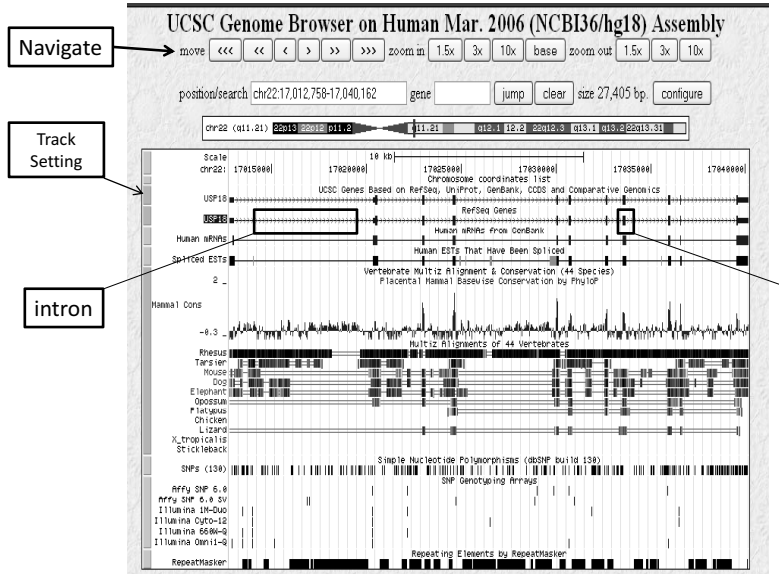
manage custom tracks | configure tracks and display | clear position

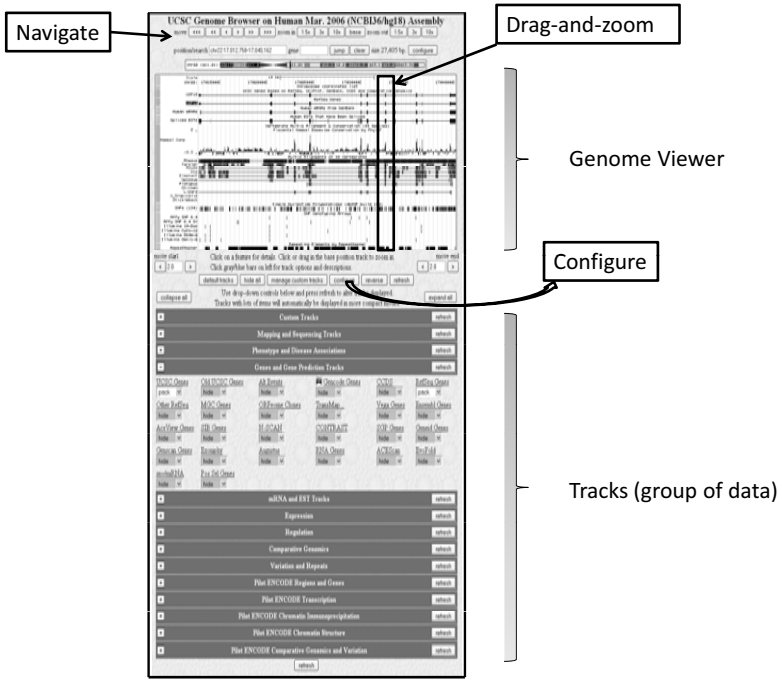
Request:	Genome Browser Response:	
chr7	Displays all of chromosome 7	Chr. region
20p13	Displays region for band p13 on chr 20	
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere	
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000	
RH18061;RH80175	Displays region between STS markers RH18061 and RH80175 or chromosome bands 15q11 to 15q13. This syntax may also be used for other range	
15q11;15q13	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.	identifier
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	keywords
homeobox caudal	Lists mRNAs for caudal homeobox genes	
zinc finger	Lists many zinc finger mRNAs	

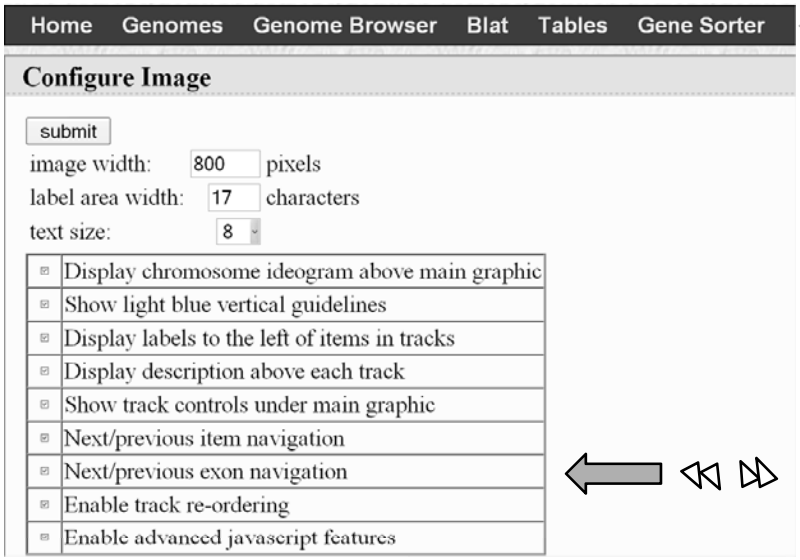


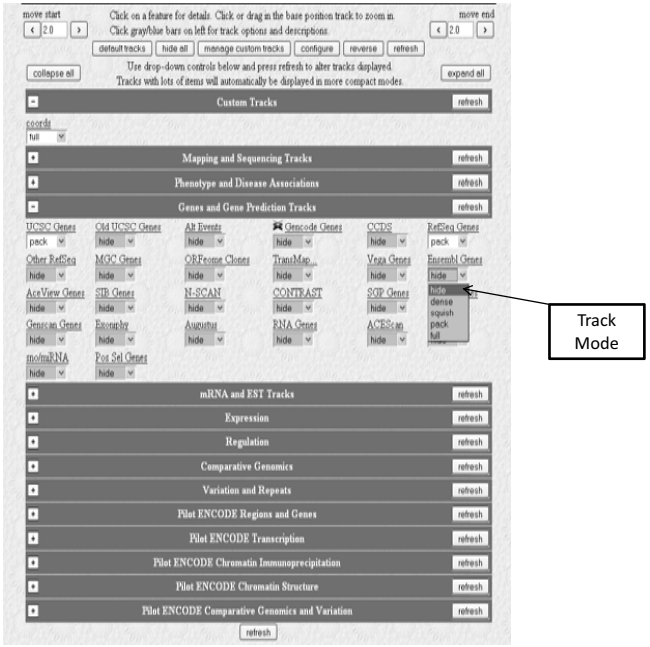
Genome Viewer

Tracks (group of data)





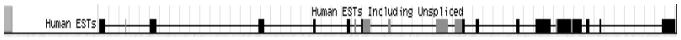




mode of an individual annotation track:

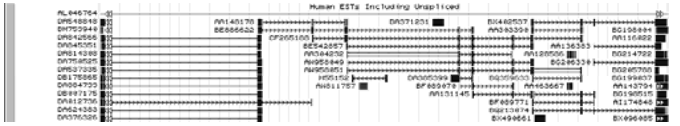
Hide: the track is not displayed at all.

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

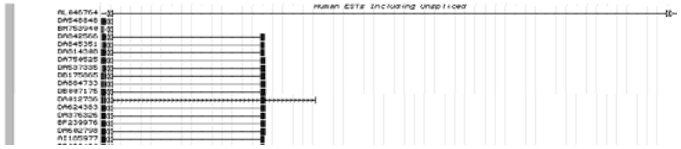


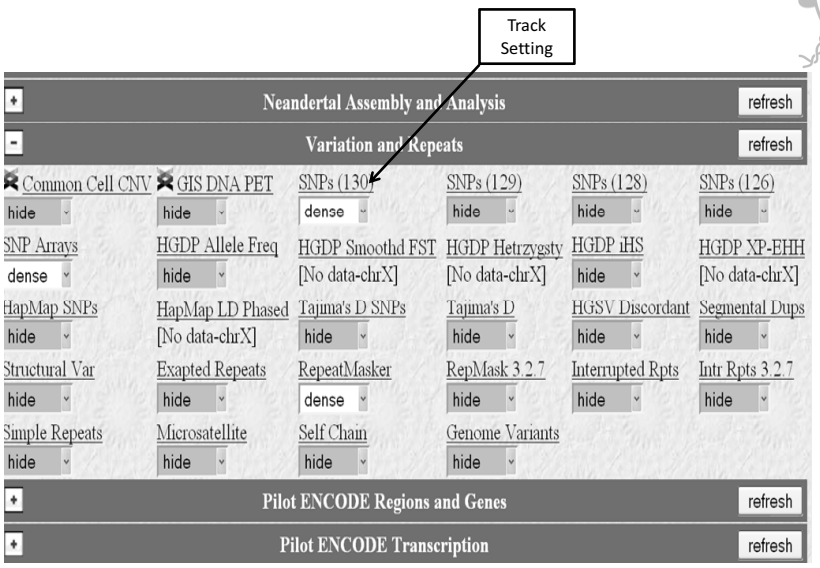
Squish: the track is displayed with each annotation feature shown separately, but at 50% the height of full mode. Features are unlabeled.

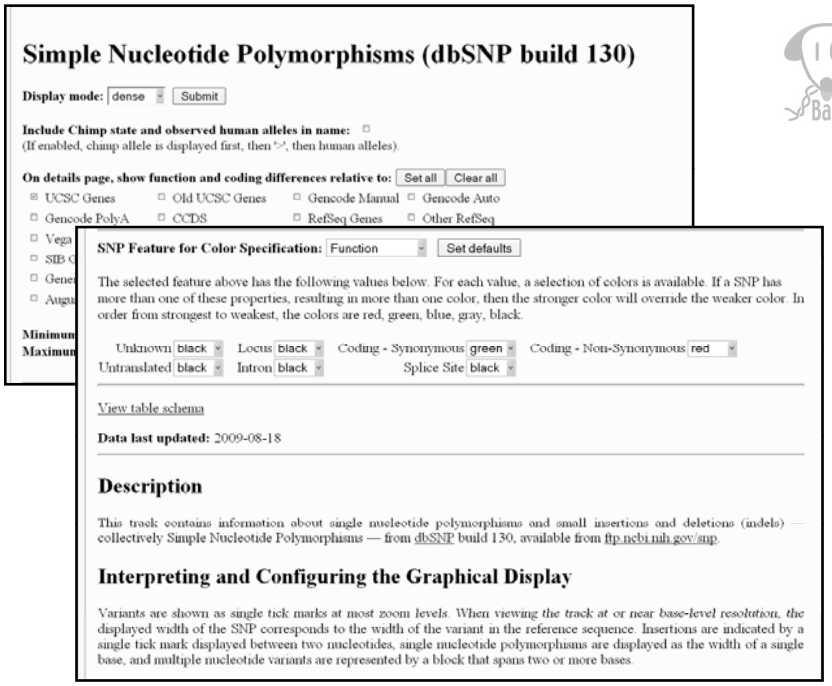
Pack: the track is displayed with each annotation feature shown separately and labeled

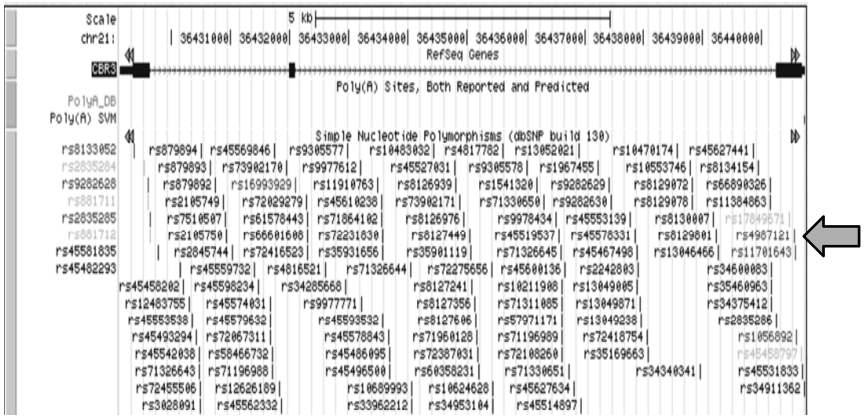


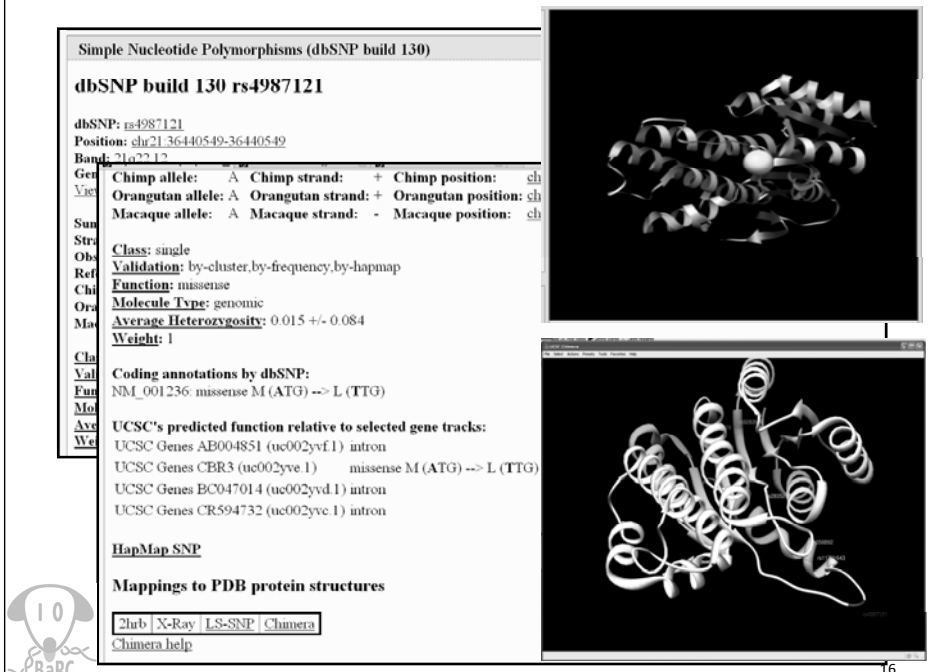
Full: the track is displayed with each annotation feature on a separate line.

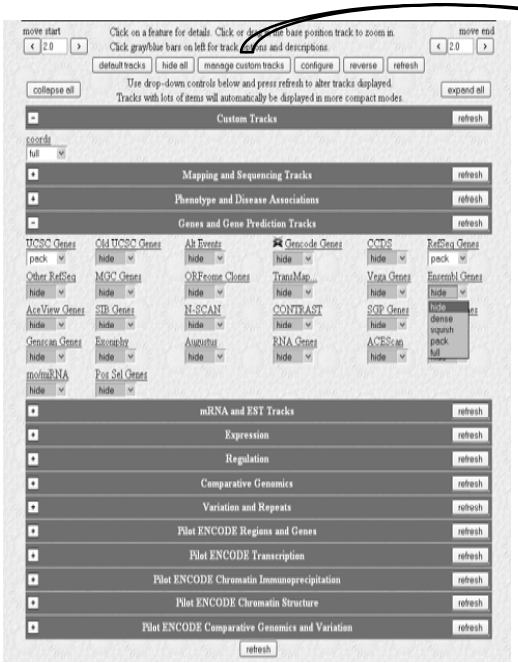


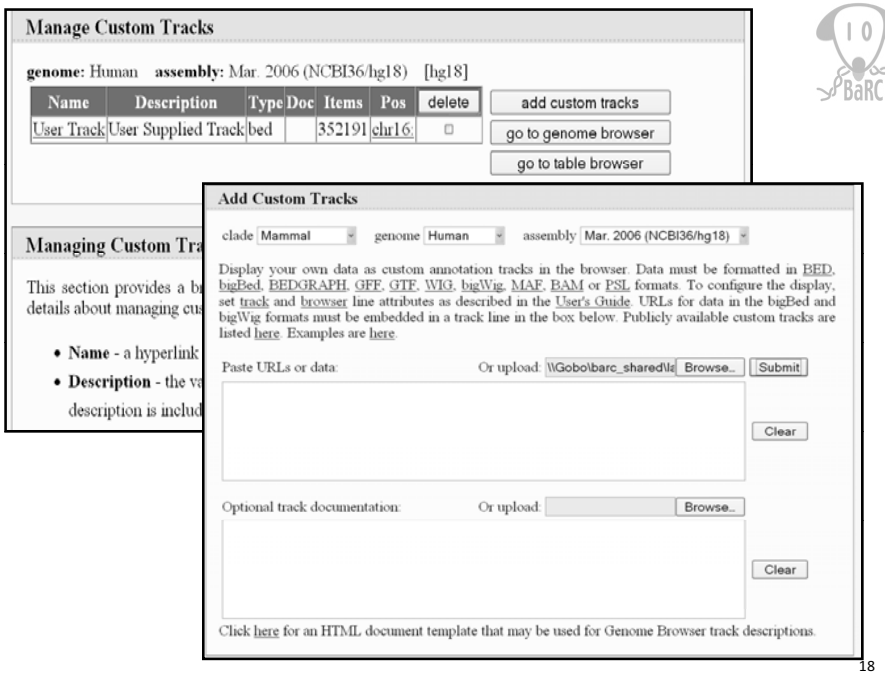

Track Setting
Neandertal Assembly and Analysis
refresh
Variation and Repeats
refresh
Common Cell CNV
GIS DNA PET
SNPs (130)
SNPs (129)
SNPs (128)
SNPs (126)
hide
hide
dense
hide
hide
hide
SNP Arrays
HGDP Allele Freq
HGDP Smoothd FST
HGDP Hetrzygsty
HGDP iHS
HGDP XP-EHH
dense
hide
[No data-chrX]
[No data-chrX]
hide
[No data-chrX]
HapMap SNPs
HapMap LD Phased
Tajima's D SNPs
Tajima's D
HGSV Discordant
Segmental Dups
hide
[No data-chrX]
hide
hide
hide
hide
Structural Var
Exapted Repeats
RepeatMasker
RepMask 3.2.7
Interrupted Rpts
Intr Rpts 3.2.7
hide
hide
dense
hide
hide
hide
Simple Repeats
Microsatellite
Self Chain
Genome Variants
hide
hide
hide
hide
Pilot ENCODE Regions and Genes
refresh
Pilot ENCODE Transcription
refresh

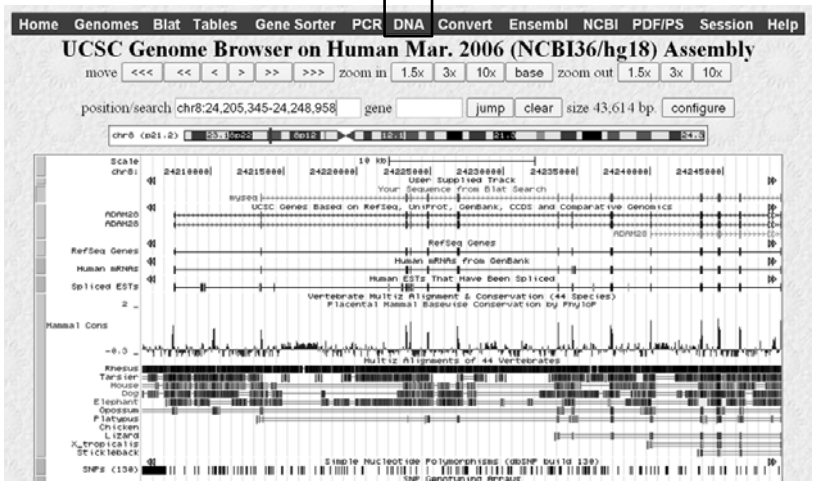

Simple Nucleotide Polymorphisms (dbSNP build 130)
Display mode: dense Submit
Include Chimp state and observed human alleles in name:
(If enabled, chimp allele is displayed first, then '>', then human alleles).
On details page, show function and coding differences relative to: Set all Clear all
UCSC Genes
Old UCSC Genes
Gencode Manual
Gencode Auto
Gencode PolyA
CCDS
RefSeq Genes
Other RefSeq
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.
Unknown black
Locus black
Coding - Synonymous green
Coding - Non-Synonymous red
Untranslated black
Intron black
Splice Site black
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.ncbi.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.

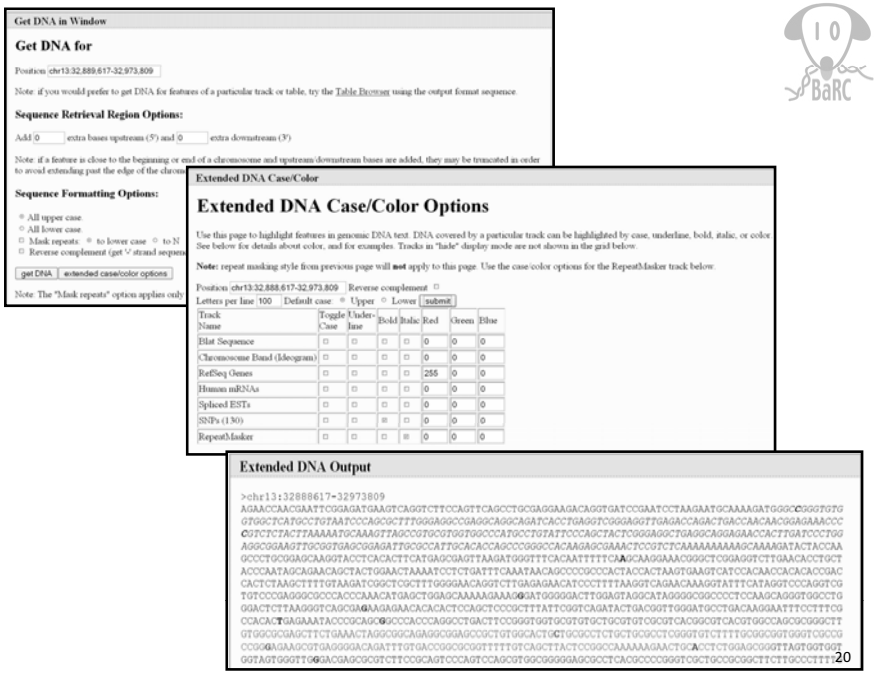

Scale
5 kb
chr21:
36431000 36432000 36433000 36434000 36435000 36436000 36437000 36438000 36439000 36440000
RefSeq Genes
CBR3
Poly(A) Sites, Both Reported and Predicted
PolyA_DB
Poly(A) SVM
Simple Nucleotide Polymorphisms (dbSNP build 130)
rs8133052
rs2835284
rs9282628
rs881711
rs2835285
rs881712
rs45581835
rs45482293
rs879894 rs45569846 rs9305577 rs10483032 rs4817782 rs13052021 rs10470174 rs45627441
rs879893 rs73902170 rs9977612 rs45527031 rs9305578 rs1967455 rs10553746 rs8134154
rs879892 rs16993929 rs11910763 rs8126939 rs1541320 rs9282629 rs8129072 rs66890326
rs2105749 rs72029279 rs45610238 rs73902171 rs71330650 rs9282630 rs8129078 rs11304863
rs7510507 rs61578443 rs71864102 rs8126976 rs9978434 rs45553139 rs8130007 rs17849671
rs2105750 rs66601608 rs72231830 rs8127449 rs45519537 rs45578331 rs8129801 rs4987121
rs2845744 rs72416523 rs35931656 rs35901119 rs71326645 rs45467498 rs13046466 rs11701643
rs45559732 rs4816521 rs71326644 rs72275656 rs45600136 rs2242803 rs34600083
rs45458202 rs45598234 rs34285668 rs8127241 rs10211908 rs13049005 rs35460963
rs12483755 rs45574031 rs9977771 rs8127356 rs71311085 rs13049871 rs34375412
rs45553638 rs45579632 rs45593532 rs8127606 rs57971171 rs13049238 rs2835286
rs45493294 rs72067311 rs45578843 rs71960128 rs71196989 rs72418754 rs1056892
rs45542038 rs58466732 rs45486095 rs72387031 rs72100260 rs35169663 rs45458792
rs71326643 rs71196988 rs45496500 rs60358231 rs71330651 rs34340341 rs45531833
rs72455586 rs12626189 rs10689993 rs10624628 rs45627634 rs34911362
rs3028091 rs45562332 rs33962212 rs34953104 rs45514897

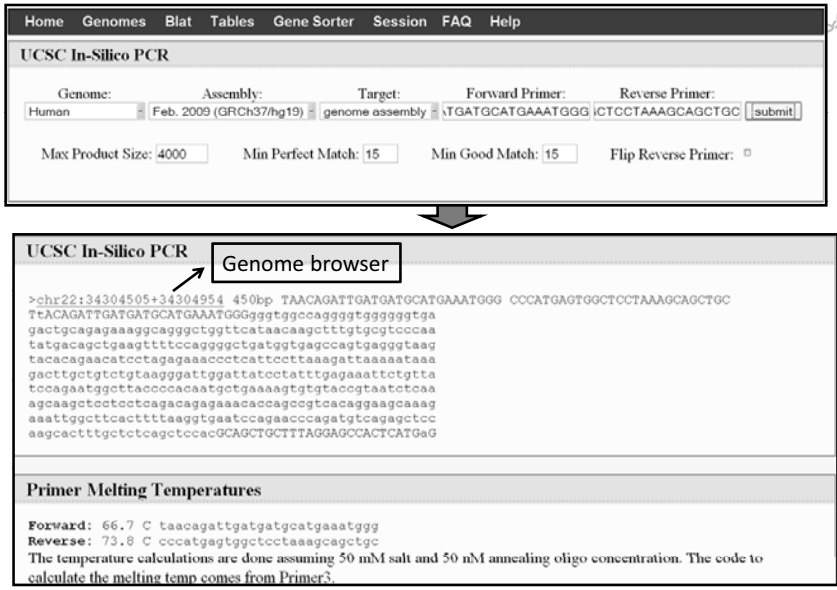

Simple Nucleotide Polymorphisms (dbSNP build 130)
dbSNP build 130 rs4987121
dbSNP: rs4987121
Position: chr21:36440549-36440549
Chimp allele: A Chimp strand: + Chimp position:
Orangutan allele: A Orangutan strand: + Orangutan position:
Macaque allele: A Macaque strand: - Macaque position:
Class: single
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 (uc002yvc.1) intron
HapMap SNP
Mappings to PDB protein structures
2hub X-Ray LS-SNP Chimera
Chimera help

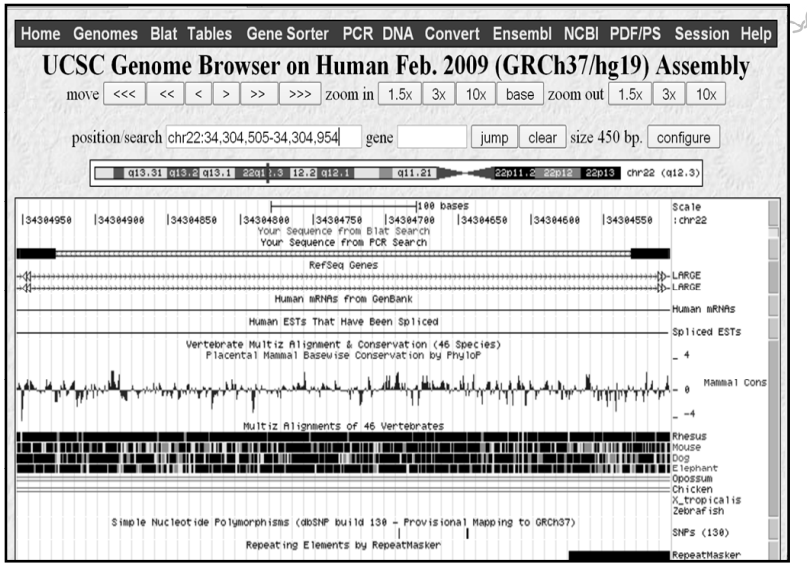

Direct to add track site
move start
move end
Click on a feature for details. Click or drag in the base position track to zoom in.
Click gray/blue bars on left for track options and descriptions.
default tracks
hide all
manage custom tracks
configure
reverse
refresh
collapse all
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
expand all
Custom Tracks
Mapping and Sequencing Tracks
Phenotype and Disease Associations
Genes and Gene Prediction Tracks
mRNA and EST Tracks
Expression
Regulation
Comparative Genomics
Variation and Repeats
Pilot ENCODE Regions and Genes
Pilot ENCODE Transcription
Pilot ENCODE Chromatin Immunoprecipitation
Pilot ENCODE Chromatin Structure
Pilot ENCODE Comparative Genomics and Variation


Manage Custom Tracks
genome: Human assembly: Mar. 2006 (NCBI36/hg18) [hg18]
Name
Description
Type
Doc
Items
Pos
delete
User Track
User Supplied Track
bed
352191
chr16:
add custom tracks
go to genome browser
go to table browser
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:
Browse..
Submit
Clear
Optional track documentation:
Click here for an HTML document template that may be used for Genome Browser track descriptions.


Extract DNA sequence
Home
Genomes
Blat
Tables
Gene Sorter
PCR
DNA
Convert
Ensembl
NCBI
PDF/PS
Session
Help
UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly
move
zoom in
zoom out
position/search chr8:24,205,345-24,248,958
gene
jump
clear
size 43,614 bp.
configure

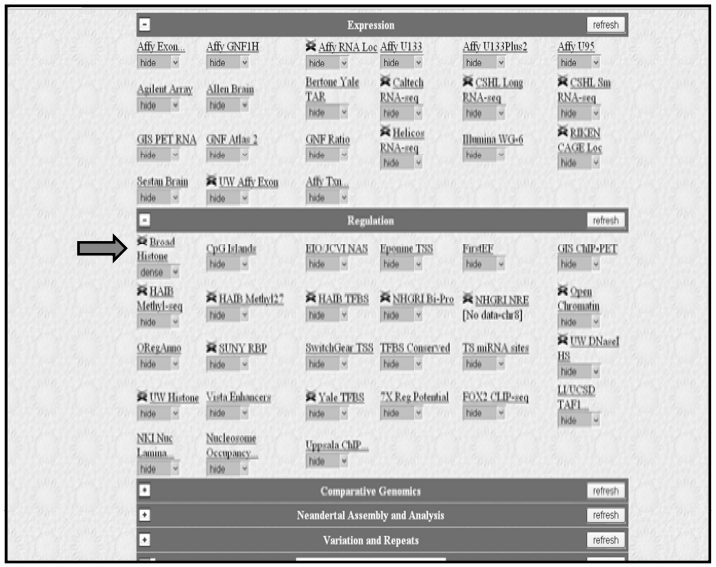

Get DNA in Window
Get DNA for
Sequence Retrieval Region Options:
Sequence Formatting Options:
Extended DNA Case/Color
Extended DNA Case/Color Options
Extended DNA Output

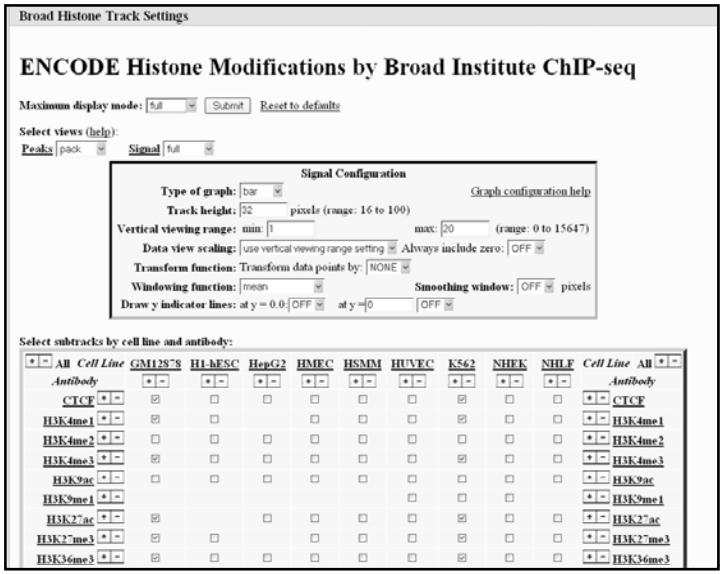




Encyclopedia of DNA Elements (ENCODE) Consortium

Founded by National Human Genome Research Institute (NHGRI)





List subtracks: ○ only selected/visible ◉ all

	Cell Line[1]	Antibody[2]	Views[3]			Restricted Until
☑	GM12878	CTCF	Peaks	ENCODE Histone Mods, Broad ChIP-seq Peaks (CTCF, GM12878) ...	schema	2009-10-05
☑	GM12878	CTCF	Signal	ENCODE Histone Mods, Broad ChIP-seq Signal (CTCF, GM12878) ...	schema	2009-10-05
☑	GM12878	H3K4me1	Peaks	ENCODE Histone Mods, Broad ChIP-seq Peaks (H3K4me1, GM12878) ...	schema	2009-10-05
☑	GM12878	H3K4me1	Signal	ENCODE Histone Mods, Broad ChIP-seq Signal (H3K4me1, GM12878) ...	schema	2009-10-05
☐	GM12878	H3K4me2	Peaks	ENCODE Histone Mods, Broad ChIP-seq Peaks (H3K4me2, GM12878) ...	schema	2009-10-05
☐	NHLF	H4K20me1	Signal	ENCODE Histone Mods, Broad ChIP-seq Signal (H4K20me1, NHLF) ...	schema	2010-06-28
☐	NHLF	Input Control	Signal	ENCODE Histone Mods, Broad ChIP-seq Signal (NHLF control) ...	schema	2010-06-29

24 of 177 selected

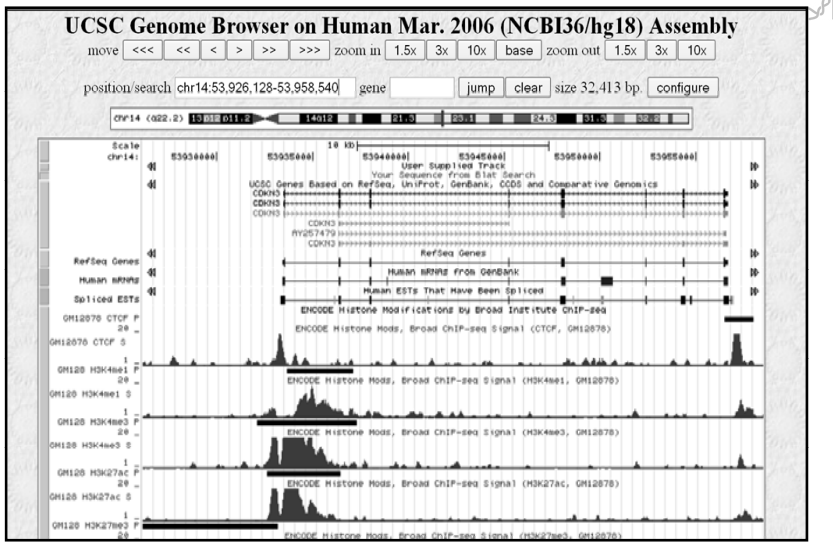
Submit

Downloads

Data version: through the ENCODE Jan 2010 Freeze

RESTRICTED

until	File	Size	Submitted	Details
2009-10-05	wgEncodeBroadChipSeqPeaksGm12878Ctcf.broadPeak.gz	363K	2009-01-05	cell=GM12878; dataType=ChipSeq; antibody=CTCF; lab=Broad; typ
2009-10-05	wgEncodeBroadChipSeqRawDataRep1Gm12878Ctcf.fastq.gz	774M	2009-01-05	cell=GM12878; dataType=ChipSeq; antibody=CTCF; lab=Broad; typ
2009-10-05	wgEncodeBroadChipSeqRawDataRep2Gm12878Ctcf.fastq.gz	517M	2009-01-05	cell=GM12878; dataType=ChipSeq; antibody=CTCF; lab=Broad; typ



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

ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res*. 2010 Jan;38(Database issue):D620-5. Epub 2009 Nov 17.

The UCSC Genome Browser database: update 2010. *Nucleic Acids Res*. 2010 Jan;38(Database issue):D613-9. Epub 2009 Nov 11.

The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res*. 2008 Jan;36:D773-9.

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