

Genome Browsers

Peter FitzGerald. PhD

Head Genome Analysis Unit, CCR, NCI

Talk Outline

- Overview of Genome Browsers
- Different typed of Browsers
- Highlighting three specific Browsers
 - UCSC Genome Browser
 - IGB - Integrated Genome Browser
 - IGV - Integrative Genomics Viewer
- Conclusion

Overview of Genome Browsers

- Tools that provide a graphical view of genomic data
- Provide a view of the spatial relationship between “genes”
- Good for comparing and integrating different data sets
- Drilling down into the raw data (see what cutoffs are doing)

Three different types of browsers based on the way they handle data

- Web-based - Numerous examples but all have these common factors:
 - User interface is through a web browser
 - The data resides remotely (with local hooks possible)
 - Processing and data manipulation is done remotely
- Client-server model - Hybrid (IGB,IGV)
 - Browser runs locally but some the data (genome annotation etc.) resides remotely and some is local and processing and graphics is local.
- Local Client (Vendor Browsers)
 - All data and processing is done locally

Other Web-based Browsers

- NCBI Map Viewer
 - <http://www.ncbi.nlm.nih.gov/mapview/>
- Ensemble
 - <http://www.ensembl.org/>
- Genome Specific

Different browsers have the same underlying genomic sequence (**within the same version**), but annotations are frequently different

Beware of Version Changes

- Different browsers update on different schedules
- Different browsers have different archiving policies
- Annotations are more variable than sequence data
- Nomenclature is different

When comparing data from different sources always make sure you are looking at the same version.

UCSC “liftover” files can interconvert

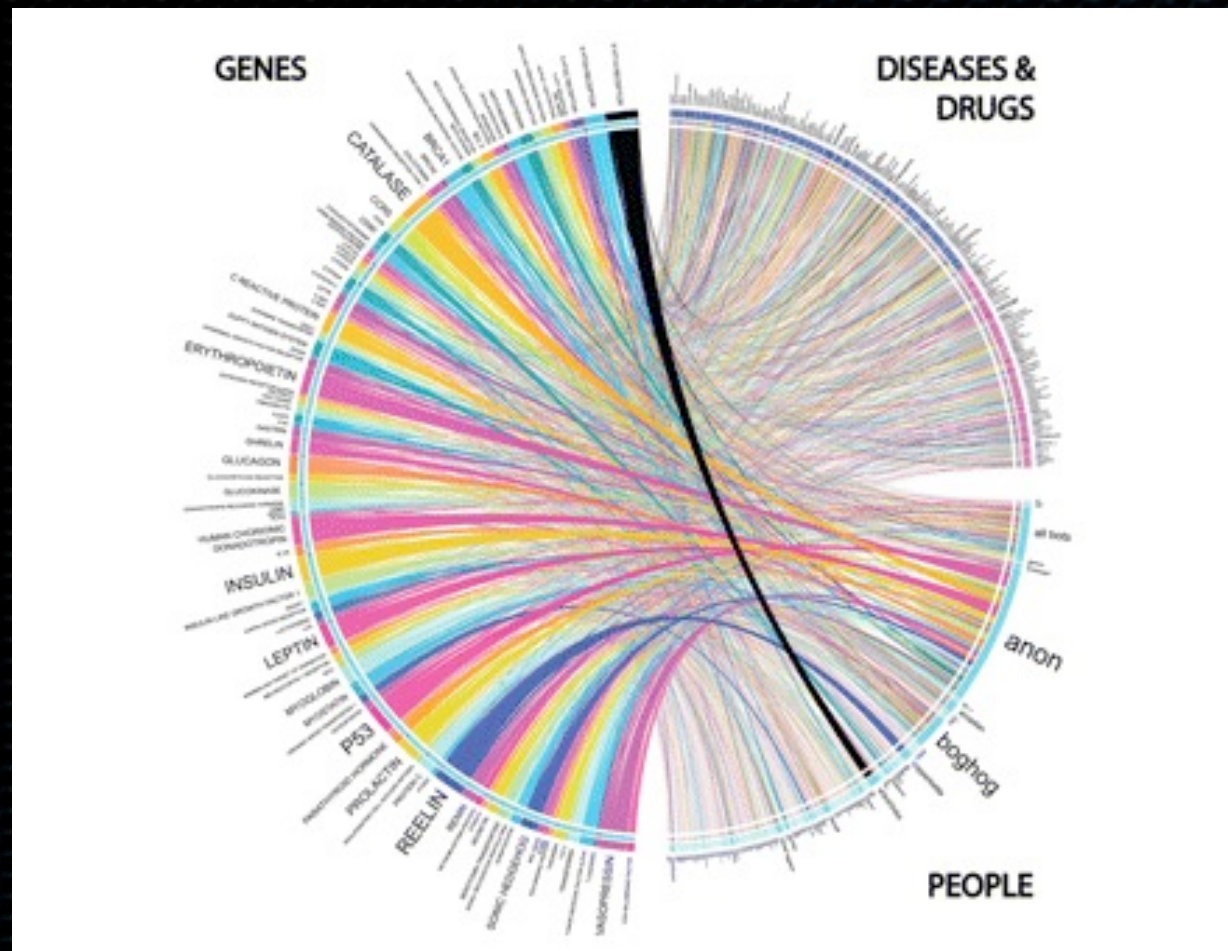
The 2012 Nucleic Acids Research

Database Issue and the online Molecular Biology Database Collection

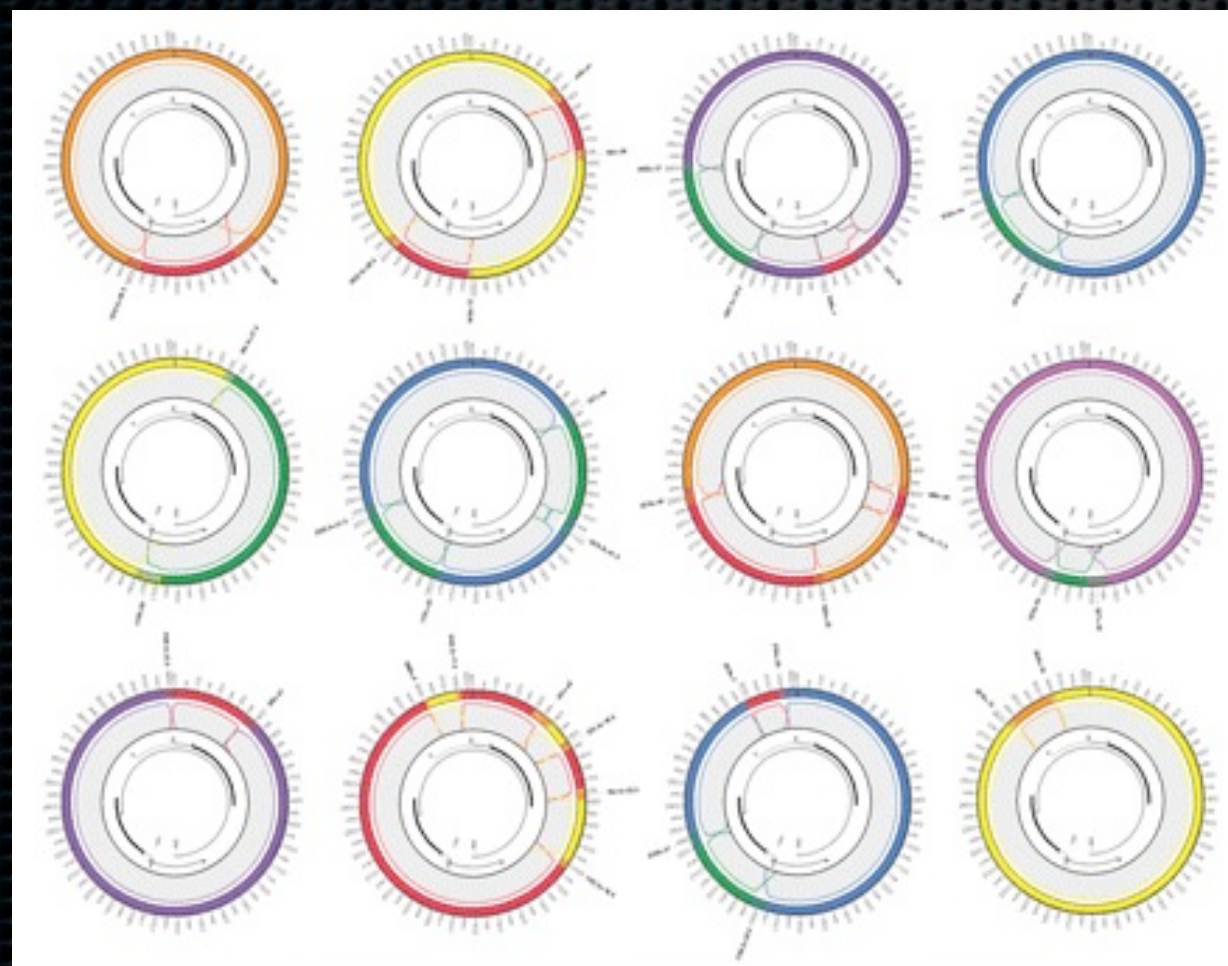
<http://nar.oxfordjournals.org/content/40/D1.toc>

Database Summaries

<http://nar.oxfordjournals.org/content/40/D1/D1/suppl/DC1>



Volume 40 Issue D1 January 2012



Volume 40 Issue W1 July 2012

Web Server issue

<http://nar.oxfordjournals.org/content/40/W1.toc>

Genomes Genome Browser Tools Mirrors Downloads My Data About Us View Help

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr17:41,209,403-41,311,549 102,147 bp. enter position, gene symbol or search terms go

chr17 (q21.31) p13.3 p13.2 p13.1 17p12 17p11.2 17q11.2 17q12 21.2 21.31 21.32 21.33 17q22 q23.2 q24.21 q24.3 17q25.1 17q25.3

Scale chr17: 41,220,000 41,230,000 50 kb 41,240,000 41,250,000 41,260,000 41,270,000 41,280,000 hg19 41,290,000 41,300,000 41,310,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs

Spliced ESTs

Layered H3K27Ac

DNase Clusters

Tbx Factor ChIP

Mammal Cons

Multiz Alignments of 46 Vertebrates

Common SNPs (135)

RepeatMasker

move start Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

UCSC Genome Browser

<http://genome.ucsc.edu/>



Cover of *Nature* human genome issue,
published on 15 February 2001.

Browser was launched in 2000 coinciding with the release of the Human Genome Sequence. Initially built and still managed by Jim Kent, then a graduate student, and David Haussler, professor of Computer Science (now Biomolecular Engineering) at the University of California, Santa Cruz.

UCSC Genome Browser

Other Flavors

- **UCSC Archive**
- **UCSC Microbial**
- **Local mirror of UCSC at NIH (restricted)**
 - NIH local Mirror (Subset) maintained by the Helix Systems (CIT)
<http://genome.cit.nih.gov/>

UCSC Genome Browser

Where to get help

- <http://genome.ucsc.edu/FAQ/> **FAQ**
- <http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html> **Help**
 - **file formats**
- <http://genome.ucsc.edu/contacts.html> **Mailing Lists**
- <http://genomewiki.ucsc.edu/> **Wiki**

UCSC Genome Browser

Genomes 6-63

• **Mammals**

- Human
- Chimp
- Gorilla
- Orangutan
- Gibbon
- Rhesus
- Marmoset
- Mouse
- Rat
- Naked mole-rat
- Guinea pig
- Rabbit
- Pig

- Sheep
- Cow
- Horse
- Cat
- Dog
- Panda
- Microbat
- Tenrec
- Elephant
- Opossum
- Tasmanian devil
- Wallaby
- Platypus

• **Vertebrate**

- Chicken
- Turkey
- Zebra finch
- Medium ground finch
- Lizard
- Painted turtle
- X. tropicalis
- Zebrafish
- Tetraodon
- Fugu
- Stickleback
- Medaka
- Lamprey

• **Deuterostome**

- Lancelet
- C.intestinalis
- S.purpuratus
- **Insect**
- D. melanogaster
- D. simulans
- D. sechellia
- D. yakuba
- D. erecta
- D. ananassae
- D. pseudoobscura
- D. persimilis
- D. virilis

- D. mojavensis

- D. grimshawi

- A. gambiae

- A. mellifera

• **Nematode**

- C. elegans

- C. brenneri

- C. briggsae

- C. remanei

- C. japonica

- P. pacificus

• **Other**

- Sea Hare

- S.cerevisiae

UCSC Genome Browser Overview

The screenshot displays the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The main track shows a 2 kb region on chromosome 21 (hg19) with coordinates 33,033,000 to 33,041,000. The track includes various data layers such as RefSeq Genes, Human mRNAs, Spliced ESTs, H3K27ac Mark, DNase Clusters, and Common SNPs. Below the track, there are navigation controls (move start, zoom in/out, move end) and track management options (track search, default tracks, hide all, add custom tracks, track hubs, configure, reverse, resize, refresh). A dropdown menu is open for the 'Assembly' track, showing options: hide, dense, squish, pack, and full. The interface is organized into sections: 'Mapping and Sequencing Tracks' and 'Phenotype and Disease Associations'.

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms go

chr21 (q22.11) 21p13 21p12 21p11.2 21p11 21q21.1 21q21.2 21q21.3 21q22.11 q22.2 21q22.3

Scale 2 kb hg19
chr21: 33,033,000 | 33,034,000 | 33,035,000 | 33,036,000 | 33,037,000 | 33,038,000 | 33,039,000 | 33,040,000 | 33,041,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Layered H3K27ac

DNase Clusters

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

Vertebrate Multiz Alignment (46 Species)

Mammal Cons Multiz Align

Common SNPs (135)

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

RepeatMasker

Repeat Masking Elements by RepeatMasker

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

move start < 2.0 > move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all expand 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.

Mapping and Sequencing Tracks refresh

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate	deCODE Recomb
dense	hide	hide	hide	hide	hide
ENCODE Pilot	Map Contigs	Assembly	GRC Map Contigs	Gap	Publications
hide	hide	hide	hide	hide	hide
BAC End Pairs	Fosmid End Pairs	BU ORChID	GRC Patch Release	Hg18 Diff	GRC Incident
hide	hide	hide	hide	hide	hide
Hi Seq Depth Wiki Track	Mapability	Short Match	Restr Enzymes		
hide	hide	hide	hide		

Phenotype and Disease Associations refresh

GAD View	DECIPHER	OMIM AV SNPs	OMIM Genes	OMIM Pheno Loci	COSMIC
hide	hide	hide	hide	hide	hide

UCSC Genome Browser Overview

The screenshot displays the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The main track shows a 2 kb region on chromosome 21 (hg19) with various annotations including RefSeq Genes, Human mRNAs, Spliced ESTs, H3K27ac Mark, DNase Clusters, and Transcription Factor ChIP-seq. Below the main track, there are controls for zooming and track management. A dropdown menu is open for the 'Assembly' track, showing options: 'hide', 'dense', 'squish', 'pack', and 'full'. The 'dense' option is currently selected.

← Zoom level

UCSC Genome Browser Overview

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms go

chr21 (q22.11) 21p13 21p12 21p11.2 21p11 21q21.1 21q21.2 21q21.3 21q22.11 q22.2 21q22.3

Scale chr21: 33,033,000 33,034,000 33,035,000 33,036,000 33,037,000 33,038,000 33,039,000 33,040,000 33,041,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

Vertebrate Multiz Alignment (46 Species)

Mammal Cons Multiz Align

Common SNPs (135)

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

RepeatMasker

Repeat Masking Elements by RepeatMasker

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

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.

collapse all expand all

Mapping and Sequencing Tracks refresh

Base Position Chromosome Band STS Markers FISH Clones Recomb Rate deCODE Recomb

ENCODE Pilot Map Contigs Assembly GRC Map Contigs Gap Publications

BAC End Pairs Fosmid End Pairs GRC Patch Release Hg18 Diff GRC Incident

Hi Seq Depth Wiki Track BU ORChID Mapability Short Match Restr Enzymes

Phenotype and Disease Associations refresh

GAD View DECIPHER OMIM AV SNPs OMIM Genes OMIM Pheno Loci COSMIC

← Coordinates and search

UCSC Genome Browser Overview

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

chr21:33,031,597-33,041,570 9,974 bp.

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

Ventrate Multiz Alignment (46 Species)

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

RepeatMasker

Mapping and Sequencing Tracks

Phenotype and Disease Associations

← Chromosome position and coordinates

UCSC Genome Browser Overview

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

chr21:33,031,597-33,041,570 9,974 bp.

Scale chr21: 33,033,000 | 33,034,000 | 33,035,000 | 33,036,000 | 33,037,000 | 33,038,000 | 33,039,000 | 33,040,000 | 33,041,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

Vertebrate Multiz Alignment (46 Species)

Human Conserved Multiz Align

Common SNPs (135)

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

RepeatMasker

Repeat Elements by RepeatMasker

move start < 2.0 > move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all expand 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.

Mapping and Sequencing Tracks refresh

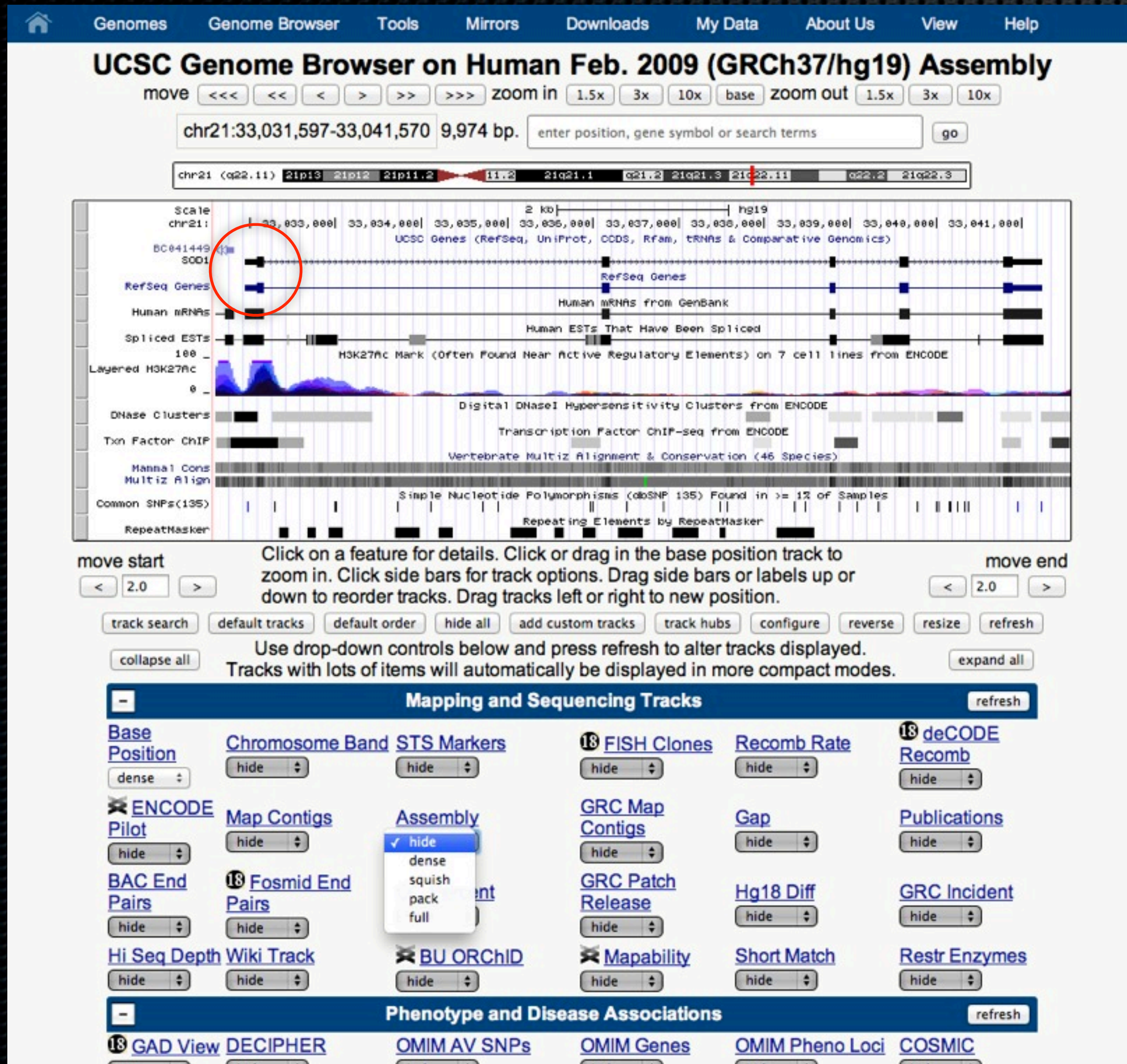
Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate	deCODE Recomb
dense	hide	hide	hide	hide	hide
ENCODE Pilot	Map Contigs	Assembly	GRC Map Contigs	Gap	Publications
hide	hide	hide	hide	hide	hide
BAC End Pairs	Fosmid End Pairs	BU ORChID	GRC Patch Release	Hg18 Diff	GRC Incident
hide	hide	hide	hide	hide	hide
Hi Seq Depth Wiki Track	Mapability	Short Match	Restr Enzymes		
hide	hide	hide	hide		

Phenotype and Disease Associations refresh

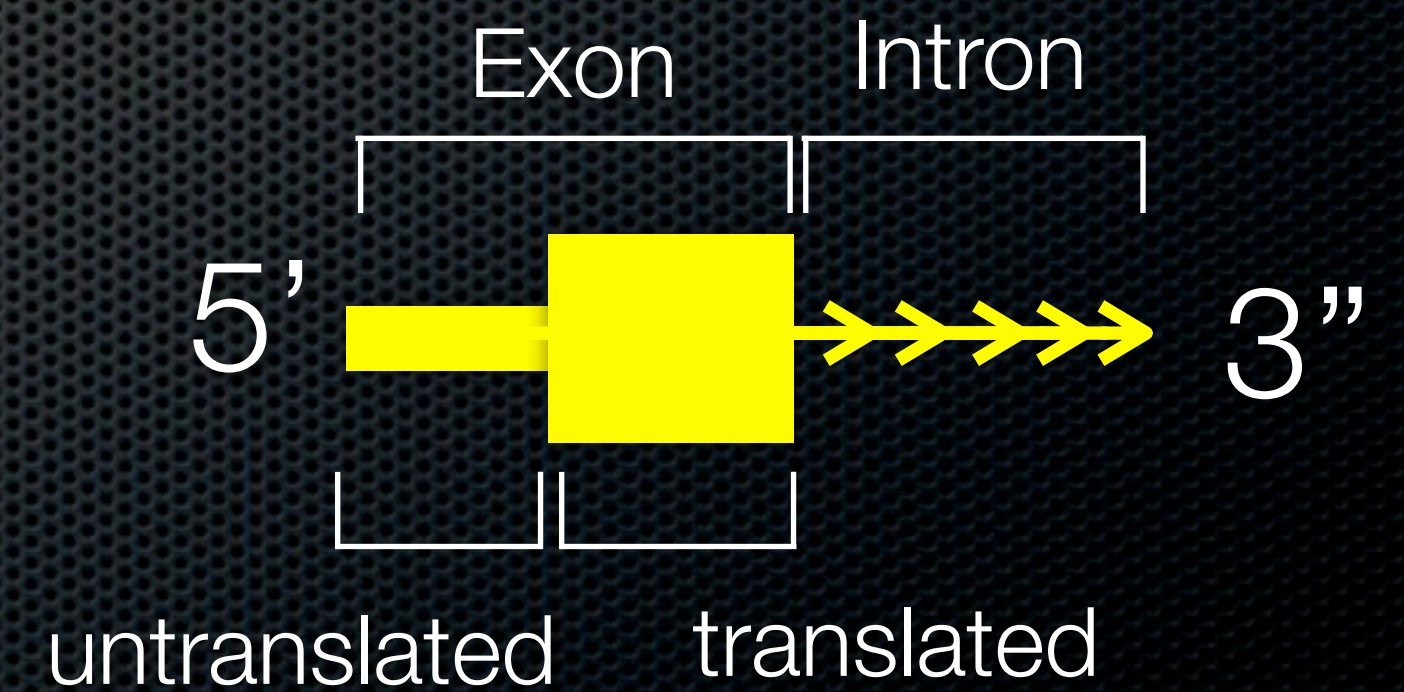
GAD View	DECIPHER	OMIM AV SNPs	OMIM Genes	OMIM Pheno Loci	COSMIC
hide	hide	hide	hide	hide	hide

← Main Display (Image)

UCSC Genome Browser Overview



← Main Display (Image)



UCSC Genome Browser Overview

The screenshot displays the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The main track shows a 2 kb region on chromosome 21 (hg19) with coordinates 33,033,000 to 33,041,000. The track includes various genomic features such as RefSeq Genes, Human mRNAs, Spliced ESTs, H3K27ac Mark, DNase Clusters, Transcription Factor ChIP, and Common SNPs. Below the track, there are navigation controls (move start, zoom in, zoom out, move end) and track selection options (track search, default tracks, default order, hide all, add custom tracks, track hubs, configure, reverse, resize, refresh). A dropdown menu is open for the 'Assembly' track, showing options: hide, dense, squish, pack, and full. The interface also includes sections for 'Mapping and Sequencing Tracks' and 'Phenotype and Disease Associations'.

← Select controls

UCSC Genome Browser Overview

The screenshot displays the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The main track shows a 2 kb region on chromosome 21 (hg19) with coordinates 33,033,000 to 33,041,000. The track includes various data layers: RefSeq Genes (BC041449, SOD1), Human mRNAs from GenBank, Spliced ESTs, H3K27ac Mark (layered), Digital DNase I Hypersensitivity Clusters from ENCODE, Transcription Factor ChIP-seq from ENCODE, Vertebrate Multiz Alignment & Conservation (46 Species), Common SNPs (135), and RepeatMasker. The interface also features navigation controls (move, zoom in/out), track search, and track configuration options (collapse/expand all, track search, default tracks, etc.).


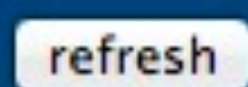

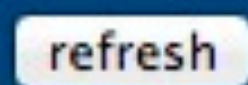



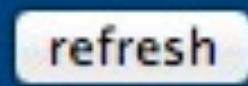







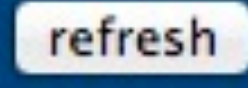

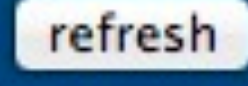
← Track controls

UCSC Genome Browser

Tracks

- Vary from Genome to Genome and possibly version
- Many current versions may lack tracks that you expect

Current Human Genome Tracks

	Mapping and Sequencing Tracks	
	Phenotype and Disease Associations	
	Genes and Gene Prediction Tracks	
	mRNA and EST Tracks	
	Expression	
	Regulation	
	Comparative Genomics	
	Neandertal Assembly and Analysis	
	Variation and Repeats	



UCSC Genome Browser Tracks

Mapping and Sequencing Tracks

refresh

Base Position dense	Chromosome Band hide	STS Markers hide	18 FISH Clones hide	Recomb Rate hide	18 deCODE Recomb hide
X ENCODE Pilot hide	Map Contigs hide	Assembly hide	GRC Map Contigs hide	Gap hide	Publications hide
BAC End Pairs hide	18 Fosmid End Pairs hide	GC Percent hide	GRC Patch Release hide	Hg18 Diff hide	GRC Incident hide
Hi Seq Depth hide	Wiki Track hide	X BU ORChID hide	X Mapability hide	Short Match hide	Restr Enzymes hide

Phenotype and Disease Associations

refresh

18 GAD View hide	DECIPHER hide	OMIM AV SNPs hide	OMIM Genes hide	OMIM Pheno Loci hide	COSMIC hide
GWAS Catalog hide	ISCA hide	18 RGD Human QTL hide	18 RGD Rat QTL hide	18 MGI Mouse QTL hide	GeneReviews hide

UCSC Genome Browser Tracks

Genes and Gene Prediction Tracks refresh

UCSC Genes pack	GENCODE... hide	Old UCSC Genes hide	Alt Events hide	CCDS hide	RefSeq Genes pack
Other RefSeq hide	MGC Genes hide	ORFeome Clones hide	TransMap... hide	Vega Genes hide	Ensembl Genes hide
AceView Genes hide	SIB Genes hide	N-SCAN hide	SGP Genes hide	Geneid Genes hide	Genscan Genes hide
Exoniphy hide	Yale Pseudo60 hide	tRNA Genes hide	H-Inv 7.0 hide	17 EvoFold hide	sno/miRNA hide
IKMC Genes Mapped hide	lincRNAs... hide				

mRNA and EST Tracks refresh

Human mRNAs dense	Spliced ESTs dense	Human ESTs hide	Other mRNAs hide	Other ESTs hide	18 H-Inv hide
UniGene hide	Gene Bounds hide	SIB Alt-Splicing hide	18 Poly(A) hide	PolyA-Seq hide	18 CGAP SAGE hide
Human RNA Editing hide					

UCSC Genome Browser Tracks

The image shows a screenshot of the UCSC Genome Browser interface, specifically the 'Genes and Gene Prediction Tracks' and 'mRNA and EST Tracks' sections. The 'Genes and Gene Prediction Tracks' section is divided into two rows of tracks. The first row includes UCSC Genes (set to 'pack'), GENCODE... (set to 'hide'), Old UCSC Genes (set to 'hide'), Alt Events (set to 'hide'), CCDS (set to 'hide'), and RefSeq Genes (set to 'pack'). The second row includes Other RefSeq (set to 'hide'), MGC Genes (set to 'hide'), ORFeome Clones (set to 'hide'), TransMap... (set to 'hide'), Vega Genes (set to 'hide'), and Ensembl Genes (set to 'hide'). The third row includes AceView Genes (set to 'hide'), SIB Genes (set to 'hide'), N-SCAN (set to 'hide'), SGP Genes (set to 'hide'), Geneid Genes (set to 'hide'), and Genscan Genes (set to 'hide'). The fourth row includes Exoniphy (set to 'hide'), Yale Pseudo60 (set to 'hide'), tRNA Genes (set to 'hide'), H-Inv 7.0 (set to 'hide'), 17 EvoFold (set to 'hide'), and sno/miRNA (set to 'hide', circled in red). The fifth row includes IKMC Genes Mapped (set to 'hide') and lincRNAs... (set to 'hide'). The 'mRNA and EST Tracks' section is also divided into two rows. The first row includes Human mRNAs (set to 'dense'), Spliced ESTs (set to 'dense'), Human ESTs (set to 'hide'), Other mRNAs (set to 'hide'), Other ESTs (set to 'hide'), and 18 H-Inv (set to 'hide'). The second row includes UniGene (set to 'hide'), Gene Bounds (set to 'hide'), SIB Alt-Splicing (set to 'hide'), 18 Poly(A) (set to 'hide'), PolyA-Seq (set to 'hide'), and 18 CGAP SAGE (set to 'hide'). The 'Human RNA Editing' track is set to 'hide'. A 'refresh' button is present at the end of each track section.

hide
dense
squish
pack
full

UCSC Genome Browser Tracks

The image displays two sections of the UCSC Genome Browser interface, each with a blue header bar and a 'refresh' button on the right. The top section is titled 'Expression' and contains 18 tracks arranged in a 3x6 grid. The bottom section is titled 'Regulation' and contains 18 tracks arranged in a 4x5 grid. Each track name is a blue hyperlink, and each has a control button below it, either 'hide' or 'show', with a small up/down arrow icon. Some tracks have a circled number next to the name, indicating a count or version.

Expression

- Affy Exon Array (hide)
- Affy GNF1H (hide)
- Affy RNA Loc (hide)
- Affy U133 (hide)
- Affy U133Plus2 (hide)
- Affy U95 (hide)
- Allen Brain (hide)
- Burge RNA-seq (hide)
- CSHL Small RNA-seq (hide)
- ENC Exon Array... (hide)
- ENC ProtGeno... (hide)
- ENC RNA-seq... (hide)
- GIS RNA PET (hide)
- GNF Atlas 2 (hide)
- Illumina WG-6 (hide)
- RIKEN CAGE Loc (hide)
- Sestan Brain (hide)

Regulation

- ENCODE Regulation... (show)
- CD34 Dnase1 (hide)
- CpG Islands (hide)
- ENC Chromatin... (hide)
- ENC DNA Methyl... (hide)
- ENC DNase/FAIRE... (hide)
- ENC Histone... (hide)
- ENC RNA Binding... (hide)
- ENC TF Binding... (hide)
- FSU Repli-chip (hide)
- ORegAnno (hide)
- Stanf Nucleosome (hide)
- SUNY SwitchGear (hide)
- SwitchGear TSS (hide)
- TFBS Conserved (hide)
- TS miRNA sites (hide)
- UMMS Brain Hist (hide)
- UW Repli-seq (hide)
- Vista Enhancers (hide)
- NKI Nuc Lamina... (hide)
- UCSF Brain Methyl (hide)

UCSC Genome Browser Tracks

[-]
Comparative Genomics
refresh

Conservation <input type="text" value="full"/>	18 Cons Indels MmCf <input type="text" value="hide"/>	GERP <input type="text" value="hide"/>	18 Evo Cpg <input type="text" value="hide"/>	Primate Chain/Net <input type="text" value="hide"/>	Placental Chain/Net <input type="text" value="hide"/>
Vertebrate Chain/Net <input type="text" value="hide"/>					

[-]
Neandertal Assembly and Analysis
refresh

18 H-C Coding Diffs <input type="text" value="hide"/>	18 Sel Swp Scan (S) <input type="text" value="hide"/>	18 5% Lowest S <input type="text" value="hide"/>	18 S SNPs <input type="text" value="hide"/>	18 Cand. Gene Flow [No data-chr21]	Neandertal Seq <input type="text" value="hide"/>
18 Neandertal Mito [No data-chr21]					

[-]
Variation and Repeats
refresh

Common SNPs(135) <input type="text" value="dense"/>	Flagged SNPs(135) <input type="text" value="hide"/>	Mult. SNPs(135) <input type="text" value="hide"/>	All SNPs(135) <input type="text" value="hide"/>	Common SNPs(132) <input type="text" value="hide"/>	Flagged SNPs(132) <input type="text" value="hide"/>
Mult. SNPs(132) <input type="text" value="hide"/>	All SNPs(132) <input type="text" value="hide"/>	SNPs (131) <input type="text" value="hide"/>	GIS DNA PET <input type="text" value="hide"/>	HAIB Genotype <input type="text" value="hide"/>	SNP/CNV Arrays <input type="text" value="hide"/>
HGDP Allele Freq <input type="text" value="hide"/>	18 HapMap SNPs <input type="text" value="hide"/>	DGV Struct Var <input type="text" value="hide"/>	Segmental Dups <input type="text" value="hide"/>	RepeatMasker <input type="text" value="dense"/>	Interrupted Rpts <input type="text" value="hide"/>
Simple Repeats <input type="text" value="hide"/>	Microsatellite <input type="text" value="hide"/>	Self Chain <input type="text" value="hide"/>	18 Genome Variants <input type="text" value="hide"/>	NumtS Sequence <input type="text" value="hide"/>	

UCSC Genome Browser

“Genes”

- Different annotation tracks for the same “gene” have different amounts of information
 - UCSC track typically shows the most extensive data
 - Sequence, crossreferences, structure, expression, gene ontology, pathways, **orthologs**
 - Different amounts of data are available for different genes

Extract Sequence

- Download specific Sequence Data
 - Corresponding to the display
 - Corresponding to a feature
- Download Genomic DNA, RNA, Protein, Promoter sequence
- Download feature via the Table Browser



UCSC Genome Browser

Extract Sequence

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Get DNA in Window (hg19/Human)

Get DNA for

Position

Note: This page retrieves genomic DNA for a single region. If you would prefer to get DNA for many items in a particular track, or get DNA with formatting options based on gene structure (introns, exons, UTRs, etc.), try using the [Table Browser](#) with the "sequence" output format.

Sequence Retrieval Region Options:

Add extra bases upstream (5') and extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

- All upper case.
- All lower case.
- Mask repeats: to lower case to N
- Reverse complement (get '-' strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".

Options for
get DNA
from View
menu or
mouse-over
feature

UCSC Genome Browser

Extract Sequence

- Sequence and Links to Tools and Databases					
Genomic Sequence (chr21:32,490,736-32,649,224)			mRNA (may differ from genome)		Protein (1452 aa)
Gene Sorter	Genome Browser	Protein FASTA	VisiGene	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneNetwork
Gepis Tissue	H-INV	HGNC	HPRD	Human Cortex Gene Expression	Jackson Lab
MOPED	OMIM	PubMed	Reactome	Stanford SOURCE	Treefam
UniProtKB	Wikipedia				

Options for getting sequence data from Gene view

UCSC Genome Browser

Extract Sequence

- Sequence and Links to Tools and Databases					
Genomic Sequence (chr21:32,490,736-32,649,224)			mRNA (may differ from genome)		Protein (1452 aa)
Gene Sorter	Genome Browser	Protein FASTA	VisiGene	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneNetwork
Gepis Tissue	H-INV	HGNC	HPRD	Human Cortex Gene Expression	Jackson Lab
MOPED	OMIM	PubMed	Reactome	Stanford SOURCE	Treefam
UniProtKB	Wikipedia				

Options for getting sequence data from Gene view

UCSC Genome Browser

Extract Sequence

Genomic Sequence Near Gene

Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the [Table Browser](#) using the output format sequence.

Sequence Retrieval Region Options:

- Promoter/Upstream by bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by bases
- One FASTA record per gene.
- One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra downstream (3')
 - Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats: to lower case to N

Options for getting Genomic sequence

UCSC Genome Browser

Extract Sequence

Options for getting
sequence data from
Protein Fasta item

Protein Alignments for knownGene uc011adk.1

MAF table:

Formatting options:

- Separate into exons
- Show nucleotides
- Output lines with just dashes
- Format output as table Truncate headers at characters (enter zero for no headers)

Species selection: Defaults

Primate

- chimp gorilla orangutan rhesus baboon
- marmoset tarsier mouse lemur bushbaby

Placental Mammal

- tree shrew mouse rat kangaroo rat guinea pig
- squirrel rabbit pika alpaca dolphin
- cow horse cat dog microbat
- megabat hedgehog shrew elephant rock hyrax
- tenrec armadillo sloth

Vertebrate

- wallaby opossum platypus chicken zebra finch
- lizard x. tropicalis tetraodon fugu stickleback
- medaka zebrafish lamprey

For information about output data format see the [User's Guide](#)

```
>uc011adk.1_hg19 1453 chr21:32496787-32639288-
MGNAESQHVEHEFYGEKHASLGRKHTSRSLRSLSHKTRRRTRHASSGKVIHRNSEVSTRSSSTPSIPQSLAENGLPFSQDGTLEDGSPWVDRVDMGLR
>uc011adk.1_panTro2 1453 chr21:30868972-31012391-
MGNAESQHVEHEFYGEKHASLGRKHTSRSLRSLSHKTRRRTRHASSGKVIHRNSEVSTRSSSTPSIPQSLAENGLPFSQDGTLEDGSPWVDRVDMGLR
>uc011adk.1_gorGor1 1453 Supercontig_0003507:449-38073+;Supercontig_0134003:870-956-;Supercontig_00
-----HASLGRKHTSRSLRSLSHKTRRRTRHASSGKVIHRNSEVSTRSSSTPSIPQSLAENGLPFSQDGTLEDGSPWVDRVDMGLR
>uc011adk.1_rheMac2 1453 chr3:15243594-15386973+
```

UCSC Genome Browser

Tools - BLAT

(BLAST-Like Alignment Tool)

Search by sequence **similarity**

- BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 25 (sometimes 20) bases or more.
- It may miss more divergent or shorter sequence alignments.
- BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more.

BLAT is not BLAST

Be cautious of matches, and no-match means little

UCSC Genome Browser

Tools - Table Browser

- Retrieve the data associated with a track in text format
- Calculate intersections between tracks
- Retrieve DNA sequence covered by a track

Tools - GeneSorter

- Displays a sorted table of genes that are related to one another. The relationship can be one of several types, including protein-level homology, similarity of gene expression profiles, or genomic proximity.
- One of the most powerful features of the Gene Sorter is its filtering capabilities. Use the filter to fine-tune the list of displayed genes to a subset based on a selection of detailed and flexible criteria. For example, the filter may be used to select all human genes over-expressed in the cerebellum that have GO-annotated G-protein coupled receptor activity.
- The Gene Sorter offers two options for displaying and downloading sequence associated with the genes in the table.
 - **sequence button** will fetch associated protein, mRNA, promoter, or genomic sequence.
 - **text button** will dump the table into a simple tab-delimited format suitable for import into a spreadsheet or relational database.

Tools - Others

- Genome Graphs - is a tool for displaying genome-wide data sets
- InSilico PCR - searches a sequence database with a pair of PCR primers
- Liftover - converts genome coordinates and genome annotation files between assemblies.
- VisiGene - This tool converts genome coordinates and genome

UCSC Genome Browser

MyData - Sessions

- Way of saving your “View” of the genome for later use or sharing
- Requires an account
- Saved for 4 months after last access
- Can be shared with others
- Can include Custom Tracks - Your Data

MyData - Track Hubs

- Way of importing data from outside sources
- Generally consists of **collections** of tracks
- Publically created Hubs
- My Hubs - “your” data hosted on Web server or FTP site

MyData - Custom Tracks

Way of importing data from an outside source as a TRACK.

- Upload annotation data from your own computer
 - Only visible from same location (lasts 48 hours)
- Upload data to a Web or FTP site
 - Can be shared with anyone (forever - or data life time)
- Create session with custom track and share
 - Can be shared with anyone (lasts 4 months or data life time)
 - Contribute track to UCSC

MyData - Custom Tracks

Source of information about many different file formats

- **General Formats**
- Axt
- **BAM**
- **BED**
- BED detail
- bedGraph
- bigBed
- bigWig
- Chain
- GenePred table
- **GFF**
- GTF
- MAF
- Microarray
- Net
- Personal Genome SNP
- PSL
- **VCF**
- **WIG**
- **ENCODE-specific formats:**
- ENCODE broadPeak
- ENCODE gappedPeak
- ENCODE narrowPeak
- ENCODE pairedTagAlign
- ENCODE peptideMapping
- ENCODE RNA elements
- ENCODE tagAlign
- **Download only formats:**
- 2bit
- **fasta** format
- **fastQ** format
- nib format
-

MyData - Custom Tracks

BED format

1. **chrom** - name of the chromosome
2. **chromStart** - Start of feature (0-based)
3. **chromEnd** - End of the feature (not included in display)
9 more optional columns
4. **name** - a label for the feature
5. **score** - a score (0-1000)
6. **strand** - which strand the feature on (+/-)

chr1	15000	20000	gene1	50	+
chr2	106000	108000	gene2	400	-

UCSC Genome Browser

MyData - Custom Tracks

BEDGraph format

1. **chrom** - name of the chromosome
2. **chromStart** - Start of feature (0-based)
3. **chromEnd** - End of the feature (not included in display)
4. **score** - a score (integer or real positive / negative number)

chr1	15000	20000	1.00
chr2	106000	108000	0.75

Format Conversion Utilities

- Galaxy (<http://galaxy.psu.edu/> - <http://galaxy.cit.nih.gov/>)
 - Galaxy is an open, web-based platform for data intensive biomedical research. Whether on the free public server or your own instance, you can perform, reproduce, and share complete analyses.
- Samtools (<http://samtools.sourceforge.net>)
 - SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format. Also, note TABIX for indexing generic tab delimited files.
- Picard (<http://picard.sourceforge.net/>)
 - Picard comprises Java-based command-line utilities that manipulate SAM files, and a Java API (SAM-JDK) for creating new programs that read and write SAM files. Both SAM text format and SAM binary (BAM) format are supported.
- UCSC Utilities (<http://hgdownload.cse.ucsc.edu/admin/exe/>)

UCSC Genome Browser

MyData - Custom Tracks

Indexed formats

Indexed binary file formats are much more efficient.

Only the portions of the files needed for the region currently displayed are transferred and loaded into the Browser. Thus for large data sets they are considerably faster than regular files. (e.g. bigBED, bigWIG, *BAMindexed*)

UCSC Genome Browser

Downloads

- Genome Data - download any and all data behind the browsers in large chunks or specific files
- Source Code - download the code used to drive the browser (mirrors)
- Utilities- download the code for ancillary programs
- FTP - download data via our FTP site at **<ftp://hgdownload.cse.ucsc.edu/>**
- MYSQL Access - Direct MySQL access to the UCSC Browser database **<genome-mysql.cse.ucsc.edu>**.

Integrated Genome Browser

IGB

- Integrated Genome Browser (IGB) is an easy-to-use, highly customizable genome browser you can use to visualize and explore genomic data and annotations, including RNA-Seq, ChIP-Seq, tiling array data, and more
- IGB (GenoViz) <http://genoviz.sourceforge.net/>
- Support Forum <http://sourceforge.net/projects/genoviz/forums/forum/439787>
- IGB User's Guide (PDF) http://genoviz.sourceforge.net/IGB_User_Guide.pdf

The Integrative Genomics Viewer

IGV

- The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.
- IGV <http://www.broadinstitute.org/igv/>
- FAQ <http://www.broadinstitute.org/software/igv/FAQ>
- User's Guide <http://www.broadinstitute.org/software/igv/UserGuide>
- Support Forum <https://groups.google.com/forum/?fromgroups#!forum/igv-help>