

UCSC Genome Bioinformatics



Kate Rosenbloom

Center for Biomolecular Science and Engineering
University of California, Santa Cruz

GMOD User Interface Caucus

January 18, 2007

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Proteome](#) - [FAQ](#) - [Help](#)

[Genome Browser](#)

[ENCODE](#)

[Blat](#)

[Table Browser](#)

[Gene Sorter](#)

[In Silico PCR](#)

[VisiGene](#)

[Proteome Browser](#)

[Utilities](#)

[Downloads](#)

[Release Log](#)

[Custom Tracks](#)

[Mirrors](#)

[Archives](#)

[Training](#)

About the UCSC Genome Bioinformatics Site

This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns.

News

[News Archives](#) ►

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

1 January 2007 - Upcoming Genome Browser Seminars: SF, Seattle, NYC, Cleveland

The UCSC Bioinformatics Group announces four regional seminars and hands-on computer workshops on the UCSC Genome Browser, presented by [OpenHelix](#):

- San Francisco, CA -- Wednesday, 31 January
- Seattle, WA -- Thursday, 1 February
- New York City, NY -- Tuesday, 13 February
- Cleveland, OH -- Wednesday, 14 February

Two sessions will be offered for the New York and Cleveland seminars: 9:00 a.m. to noon and 1 p.m. to 4 p.m. Only the afternoon session will be offered in San Francisco and Seattle.

<http://genome.ucsc.edu>

The UCSC Genome Browser Presents Fully Annotated Genomes

Vertebrates

- human
- chimp
- rhesus macaque
- dog
- cow
- mouse
- rat
- opossum
- chicken
- tetraodon, fugu, zebrafish

Invertebrates

- sea squirt
- sea urchin
- fruitfly (12)
- honeybee
- mosquito
- worm (2)
- yeast

And coming soon...

- cat
- platypus
- medaka, stickleback

Hardware



- **Public Site**

- 8 machines -- redundant
- 64-bit
- 8 Gb RAM
- 1500 Gb storage
- + 15 blat servers

- **Under the hood**

KiloKluster = 1000 CPUs

- Linux Red Hat 9, Apache, Parasol
- 10-Gigabit data transmission
- dual 866 MHz machines x 500
- 1 Gb RAM each

Smaller Clusters

- 100-node cluster: dual Xeon 2.6 GHz
- 400-node cluster

NFS

- 12 machines on RAID arrays
- 4 - 8 Gb RAM
- 20+ Tb storage

Data Contributors

- Human Genome Project
- Genbank/DDJ/EMBL contributors
- ENCODE Consortium
- Novartis GNF foundation
- Affymetrix, Perlegen, SNP Consortium
- SwissProt, Ensembl, EBI and NCBI
- Jackson Labs, RGD, Wormbase, Flybase
- Many contributors of gene prediction and other tracks.

High volume data handling

- All Genbank mRNAs loaded and aligned to the genome nightly; all ESTs weekly (24-48 hours to process).
- At least 6000 - 7000 regular users (separate IP addresses daily).
- 2 - 3 million hits a week
- Consistently #1 or #2 user of bandwidth on the UCSC campus

UCSC Bioinformatics Tools

- Genome Browser
- Table Browser
- Gene Sorter
- VisiGene
- Custom Tracks
- BLAT
- Downloads server, DAS server, mySQL access

Genome Browser

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Help

UCSC Genome Browser on Human Mar. 2006 Assembly

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

position/search chrX:151,073,054-151,383,976 jump clear size 310,923 bp. configure

chrX (q28)

chrX: 151150000 151200000 151250000 151300000 151350000
STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mammalian Gene Collection Full ORF mRNAs
Human mRNAs from GenBank
Spliced ESTs
UniGene Alignments
Vertebrate Multiz Alignment & Conservation (17 Species)
Conservation
mouse
rat
rabbit
dog
armadillo
elephant
opossum
chicken
X_tropicalis
tetraodon
Simple Nucleotide Polymorphisms (dbSNP build 126)
Repeating Elements by RepeatMasker
RepeatMasker

move start < 2.0 > Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options. move end < 2.0 >

default tracks hide all add custom tracks configure 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.

Track configuration & description

Vertebrate Multiz Alignment & Conservation (17 Species)

Display mode:

Pairwise alignments:

mammal

chimp rhesus mouse rat rabbit
 dog cow armadillo elephant tenrec
 opossum

vertebrate

chicken x. tropicalis zebrafish tetraodon fugu

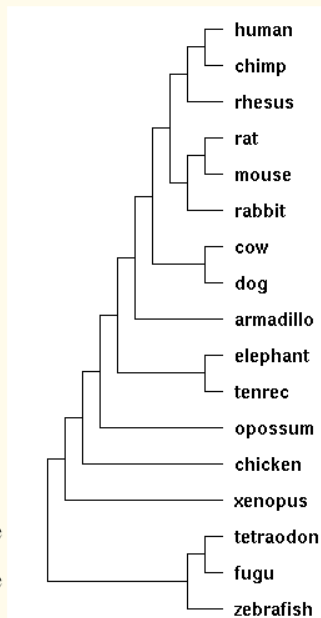
Multiple alignment base-level:

Display bases identical to reference as dots
 Display chains between alignments

Codon Translation:

Default species to establish reading frame:

No codon translation
 Use default species reading frames for translation
 Use reading frames for species if available, otherwise no translation
 Use reading frames for species if available, otherwise use default species



Description

This track shows a measure of evolutionary conservation in 17 vertebrates, including mammalian, amphibian, bird, and fish species, based on a phylogenetic hidden Markov model, phastCons (Siepel *et al.*, 2005). Multiz alignments of the following assemblies were used to generate this track:

- human (Mar. 2006, hg18)
- chimp (Nov 2003, panTro1)
- macaque (Jan 2006, rheMac2)
- mouse (Feb 2006, mm8)

Codon translation uses the following gene tracks as the basis for translation, depending on the species chosen:

Gene Track	Species
Known Genes	human, mouse, rat
RefSeq Genes	chicken
MGC Genes	X. tropicalis
Ensembl Genes	Fugu, chimp
mRNAs	rhesus, rabbit, dog, cow, zebrafish
not translated	armadillo, elephant, tenrec, opossum, Tetraodon

Methods

Best-in-genome pairwise alignments were generated for each species using blastz, followed by chaining and netting. The pairwise alignments were then multiply aligned using multiz, following the ordering of the species tree diagrammed above. The resulting multiple alignments were then assigned conservation scores by phastCons, using a tree model with branch lengths derived from the ENCODE project Multi-Species Sequence Analysis group, September 2005 tree model. This tree was generated from TBA alignments over 23 vertebrate species and is based on 4D sites.

The phastCons program computes conservation scores based on a phylo-HMM, a type of probabilistic model that describes both the process of DNA substitution at each site in a genome and the way this process changes from one site to the next (Eisenstein and Churchill 1996, Yang 1995, Siepel and

Table Browser

[Home](#) [Genomes](#) [Genome Browser](#) [Blat](#) [Tables](#) [Gene Sorter](#) [PCR](#) [FAQ](#) [Help](#)

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. For more complex queries, you may want to use our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: **genome:** **assembly:**

group: **track:**

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).


Gene Sorter

Home	UCSC Human Gene Sorter										Help										
	genome	Human	assembly	May 2004	search	NM_007295	Go!														
	sort by	Expression (GNF Atlas2)	configure	filter (now off)	display	50	output	sequence	text												
#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ T cells	skin	adipocyte	pancreatic islets	heart	lung	kidney	liver	ovary	testis	BLASTP E-Value	Rankprop Score	Genome Position	Description
1	BRCA1	1446																0	1	chr17 38,490,250	breast cancer 1, early onset isoform 1
2	ITGB3BP	97924																n/a	0.00234442	chr1 63,659,669	integrin beta 3 binding protein
3	RFC3	161999																n/a	0.210275	chr13 33,299,938	replication factor C 3 isoform 1
4	Y15758	n/a																n/a	n/a	chr13 51,938,044	LB1 protein.
5	RACGAP1	n/a																n/a	n/a	chr12 48,687,350	Rac GTPase activating protein 1
6	NUP107	n/a																n/a	n/a	chr12 67,394,868	nucleoporin 107kDa
7	MRPL39	99488																n/a	n/a	chr21 25,890,756	mitochondrial ribosomal protein L39 isoform 1
8	POLE2	100419																n/a	n/a	chr14 49,202,356	DNA polymerase epsilon subunit 2
9	RFC4	100881																n/a	0.21743	chr3 187,998,689	replication factor C 4
10	AF290612	102240																n/a	n/a	chr15 39,436,457	Nucleolar and spindle associated protein 1
11	DSCR2	83207																n/a	n/a	chr21 39,473,284	Down syndrome critical region protein 2
12	ZWINT	54581																n/a	n/a	chr10 57,789,122	ZW10 interactor isoform a
13	RBBP8	n/a																n/a	0.402232	chr18 18,814,521	retinoblastoma binding protein 8
14	CR749851	101368																n/a	n/a	chr1 47,459,826	TAL1 (SCL) interrupting locus (Fragment)
15	GMNN	n/a																n/a	n/a	chr6 24,888,700	geminin
16	TRIP13	n/a																n/a	0.211022	chr5 958,581	thyroid hormone receptor interactor 13
17	KIF2C	n/a																n/a	n/a	chr1 44,888,575	kinesin family member 2C
18	KIF14	n/a																n/a	n/a	chr1 197,287,241	kinesin family member 14

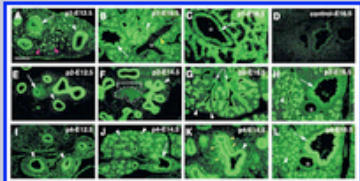
Visigene (a “virtual microscope”)

UCSC VisiGene Zoom: [Help](#)

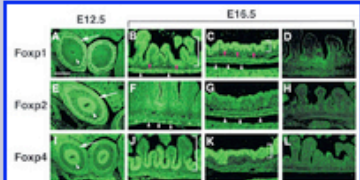
Mouse Foxp2



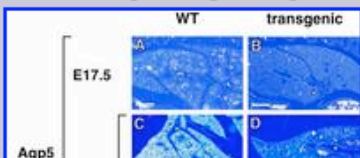
Mouse Foxp2



Mouse Foxp1 Foxp2 Foxp4



Mouse Foxp1 Foxp2 Foxp4



source: [Mahoney Lab](#) source: [MGI](#) Reference: [Mouse Brain Organization Revealed Through Direct Genome-Scale TF Expression Analysis](#).
Year: 2004 Contributors: Gray P.A.,Fu H.,Luo P.,Zhao Q.,Yu J.,Ferrari A.,Tenzen T.,Yuk D.I.,Tsong E.F.,Cai Z.,Alberta J.A.,Cheng L.P.,Liu Y.,Stenman J.M.,Valerius M.T.,Billings N.,Kim H.A.,Greenberg M.E.,McMahon A.P.,Rowitch D.H.,Stiles C.D.,Ma Q.,
Gene: [Foxp2](#) Probe: [RNA from primers](#) GenBank: [NM_053242](#)
Organism: Mus musculus Sex: n/a Strain: C57BL Genotype: wild type
Stage: 10.5 day old embryo (Theiler 17) Body Part: whole
Expression: central nervous system(0.17) Section Type: whole mount



ENCODE Project at UCSC

[Regions](#) - [Data Submission](#) - [Downloads](#) - [Tools](#) - [Terms](#) - [Help](#)

[Regions \(hg17\)](#)

[Regions \(hg16\)](#)

[Data Status](#)

[Downloads](#)

[Submission](#)

[Tools](#)

[Release Log](#)

[Contributors](#)

[Terms of Use](#)

[Genome Browser](#)

About the ENCODE Project

This site contains information related to the [ENCODE project](#) at [NHGRI](#). The UCSC Genome Bioinformatics Group manages the official repository of sequence-related data for the ENCODE consortium and supports the coordination of data submission, storage, retrieval, and visualization. A summary of the status of datasets submitted to UCSC by ENCODE contributors is available on the [ENCODE data status page](#).

UCSC also has a special interest in comparative genomics, and we provide resources for the ENCODE multiple sequence alignment interest group. [Ensembl](#) also provides an [ENCODE resource page](#).

We'd like to thank NHGRI and the contributors of annotations and analyses to this project. The team at UCSC that develops and maintains this ENCODE site is made up of [Daryl Thomas](#), [Kate Rosenbloom](#), [Jim Kent](#), and the [UCSC Genome Bioinformatics staff](#). [Read more](#).

News

[News Archives](#) ►

7 Oct. 2006 - Comparative Genomics Data Release

Twelve tracks of data produced by the ENCODE Multi-Species Sequence Analysis group have been released to the UCSC public server. These tracks contain multiple sequence alignments, conservation, and conserved (constrained) elements produced by four conservation methods (phastCons, binCons, GERP, SCONe) applied to three sequence alignments (TBA, MLAGAN, MAVID), and also an assessment of the agreement among the alignment methods. The alignments were based on genomic sequence in the ENCODE regions of 28 vertebrate species, as defined in the [MSA September 2005 sequence freeze](#).

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

ENCODE Browser

hide [Old \(hg16\)](#)

Region	Description	Chr	Size (~Mb)
ENm001	CFTR	7	1.9
ENm002	Interleukin	5	1.0
ENm003	Apo Cluster	11	0.5
ENm004	Chr22 Pick	22	1.7
ENm005	Chr21 Pick	21	1.7
ENm006	ChrX Pick	X	1.2
ENm007	Chr19 Pick	19	1.0
ENm008	Alpha Globin	16	0.5
ENm009	Beta Globin	11	1.0
ENm010	HOXA Cluster	7	0.5
ENm011	IGF2/H19	11	0.6
ENm012	FOXP2	7	1.0
ENm013	Manual	7	1.1
ENm014	Manual	7	1.2
ENr111	Random	13	0.5
ENr112	Random	2	0.5
ENr113	Random	4	0.5
ENr114	Random	10	0.5
ENr121	Random	2	0.5
ENr122	Random	18	0.5
ENr123	Random	12	0.5
ENr131	Random	2	0.5

Home Genomes Blat Tables Gene Sorter PCR DNA Convert PDF/PS Help

UCSC Genome Browser on Human May 2004 Assembly

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

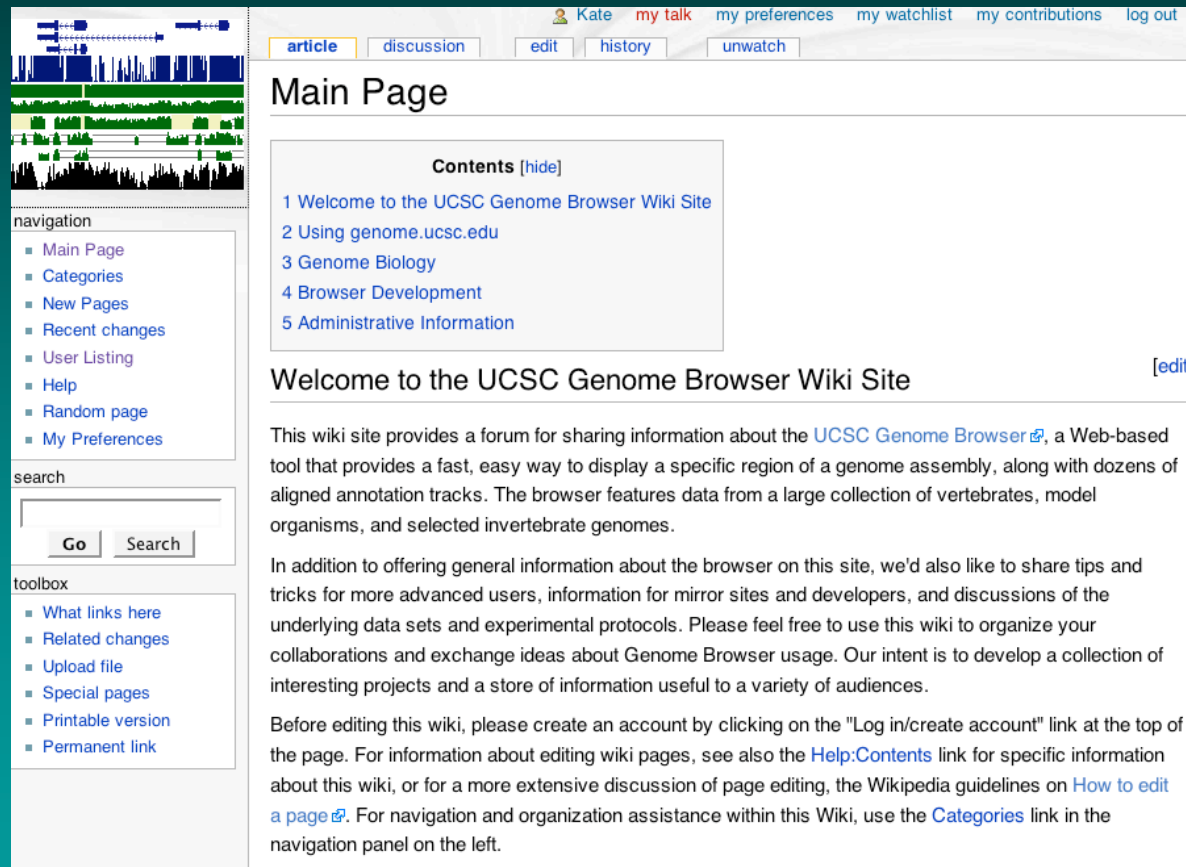
position/search jump clear size 1,877,426 bp.
 configure

chr7 (q31.2-q31.31)

chr7: 115000000 116000000 117000000

STS Markers
 UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
 TES CAV1 MET CAPZA2 ST7 WNT2 CFTR CTTNBP2
 TES AF143171 AF172985 AF234882 BC038954 AK093212 AY009152 AY009153 AY009152
 RefSeq Genes
 Mammalian Gene Collection Full ORF mRNAs
 BC001451 BC005256 BC082246 BC075855 BC126186 BC126188 BC034963
 ExonInphy Human/Mouse/Rat/Dog
 ExonWalk Alt-Splicing Transcripts
 Human mRNAs from GenBank
 Human ESTs That Have Been Spliced
 Spliced ESTs Vertebrate Multiz Alignment & Conservation
 Conservation
 mouse rat dog opossum chicken x_tropicalis tetraodon
 Simple Nucleotide Polymorphisms (dbSNP build 125)
 SNPs
 Repeating Elements by RepeatMasker
 RepeatMasker

New features: Genomewiki



The screenshot shows the main page of the UCSC Genome Browser Wiki. At the top, there is a navigation bar with links for 'Kate', 'my talk', 'my preferences', 'my watchlist', 'my contributions', and 'log out'. Below this is a sub-navigation bar with 'article', 'discussion', 'edit', 'history', and 'unwatch'. The main heading is 'Main Page'. A 'Contents' box lists five items: '1 Welcome to the UCSC Genome Browser Wiki Site', '2 Using genome.ucsc.edu', '3 Genome Biology', '4 Browser Development', and '5 Administrative Information'. Below the contents is the main text area, starting with 'Welcome to the UCSC Genome Browser Wiki Site' followed by an '[edit]' link. The text describes the wiki's purpose and provides information for users and developers. On the left side, there is a 'navigation' panel with links to 'Main Page', 'Categories', 'New Pages', 'Recent changes', 'User Listing', 'Help', 'Random page', and 'My Preferences'. Below that is a 'search' box with 'Go' and 'Search' buttons. At the bottom left is a 'toolbox' with links for 'What links here', 'Related changes', 'Upload file', 'Special pages', 'Printable version', and 'Permanent link'.

article discussion edit history unwatch

Main Page

Contents [hide]

- 1 Welcome to the UCSC Genome Browser Wiki Site
- 2 Using genome.ucsc.edu
- 3 Genome Biology
- 4 Browser Development
- 5 Administrative Information

[edit]

Welcome to the UCSC Genome Browser Wiki Site

This wiki site provides a forum for sharing information about the [UCSC Genome Browser](#), a Web-based tool that provides a fast, easy way to display a specific region of a genome assembly, along with dozens of aligned annotation tracks. The browser features data from a large collection of vertebrates, model organisms, and selected invertebrate genomes.

In addition to offering general information about the browser on this site, we'd also like to share tips and tricks for more advanced users, information for mirror sites and developers, and discussions of the underlying data sets and experimental protocols. Please feel free to use this wiki to organize your collaborations and exchange ideas about Genome Browser usage. Our intent is to develop a collection of interesting projects and a store of information useful to a variety of audiences.

Before editing this wiki, please create an account by clicking on the "Log in/create account" link at the top of the page. For information about editing wiki pages, see also the [Help:Contents](#) link for specific information about this wiki, or for a more extensive discussion of page editing, the Wikipedia guidelines on [How to edit a page](#). For navigation and organization assistance within this Wiki, use the [Categories](#) link in the navigation panel on the left.

navigation

- Main Page
- Categories
- New Pages
- Recent changes
- User Listing
- Help
- Random page
- My Preferences

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

<http://genomewiki.cse.ucsc.edu>

New features: Custom track manager

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Add Custom Tracks

clade genome assembly [hg18]

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [GFF](#), [GTF](#), [WIG](#) or [PSL](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Publicly available custom tracks are listed [here](#). Examples are [here](#).

Paste URLs or data: Or upload:

Optional track documentation: Or upload:

Click [here](#) for a

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Manage Custom Tracks

genome: Human assembly: Mar. 2006 [hg18]

Name	Description	Type	Doc	Items	Pos	delete
regulatory	TeleGene(tm) Regulatory Regions	gff		2	chr22:	<input type="checkbox"/>

New feature: Track reordering

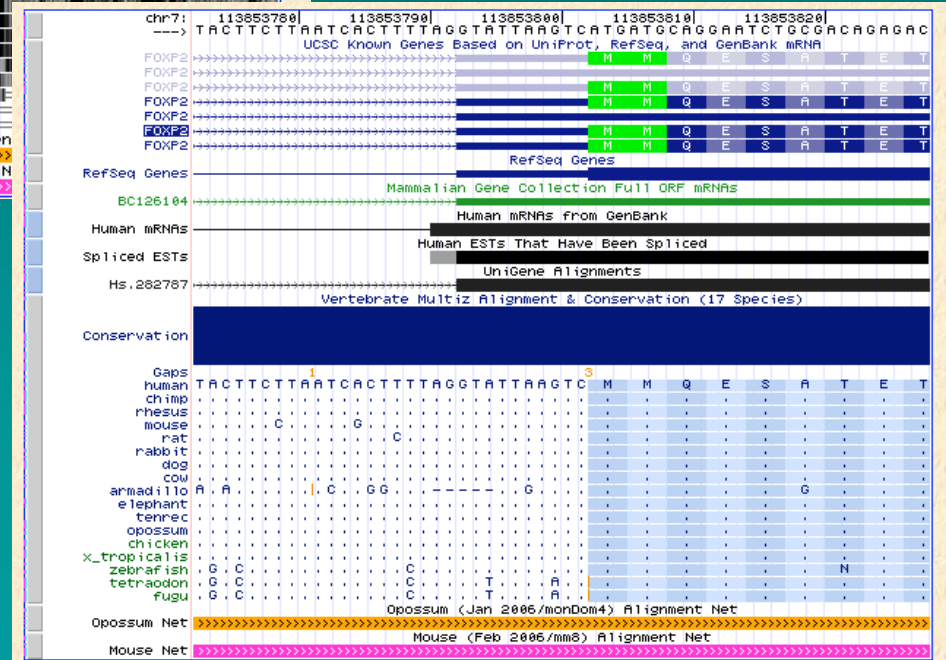
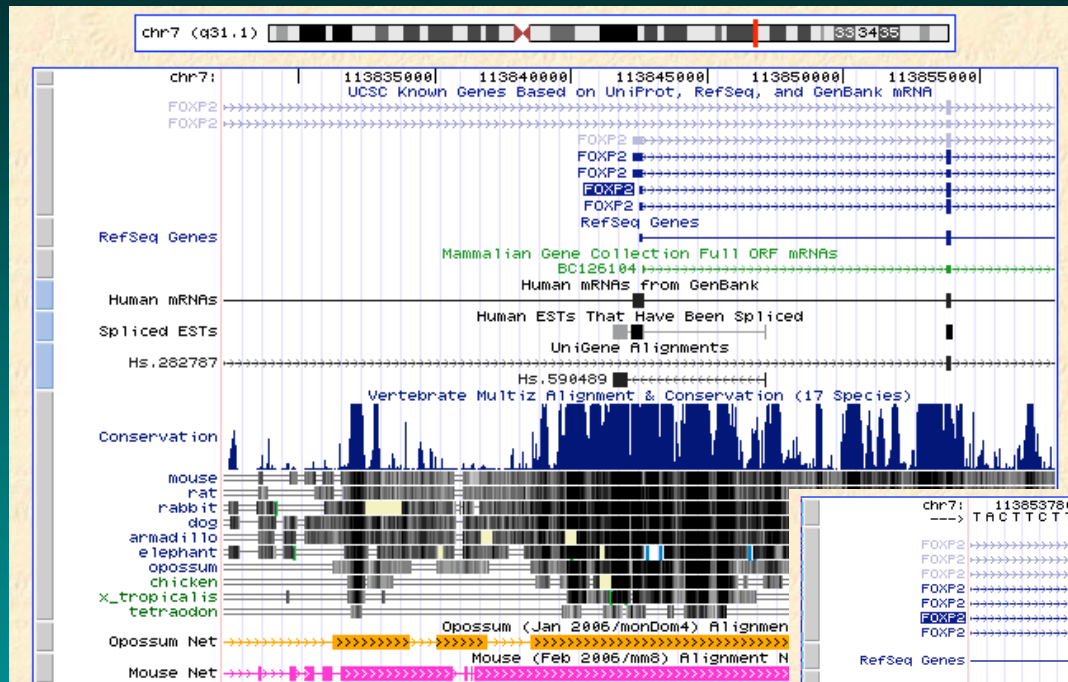
Configure Tracks

Control tracks in all groups here: Control track visibility more selectively below.

Mapping and Sequencing Tracks			<input type="button" value="hide all"/>	<input type="button" value="show all"/>	<input type="button" value="default"/>	<input type="button" value="submit"/>	2	Group
Track Order:								
Base Position	dense	Chromosome position in bases. (Clicks here zoom in 3x)						
Gap	hide	Gap Locations	11	map				
BAC End Pairs	hide	BAC End Pairs	15	map				
GC Percent	hide	Percentage GC in 20,000-Base Windows	23	map				
Short Match	hide	Perfect Matches to Short Sequence (AAAAA)	99	map				
Restr Enzymes	hide	Restriction Enzymes from REBASE	99.9	map				

Genes and Gene Prediction Tracks			<input type="button" value="hide all"/>	<input type="button" value="show all"/>	<input type="button" value="default"/>	<input type="button" value="submit"/>	3	Group
Track Order:								
FlyBase Genes	pack	FlyBase Protein-Coding Genes	34	genes				
RefSeq Genes	dense	RefSeq Genes	35	genes				
FB Noncoding	pack	FlyBase Noncoding Genes	35	genes				
N-SCAN	hide	N-SCAN Gene Predictions	45.1	genes				
Geneid Genes	hide	Geneid Gene Predictions	49	genes				
Genscan Genes	hide	Genscan Gene Predictions	50	genes				
Augustus Genes	hide	Augustus Gene Predictions	51.7	genes				
Human Proteins	pack	Human(hg17) proteins mapped by chained tBLASTn	142	genes				

New features: Comparative genomics

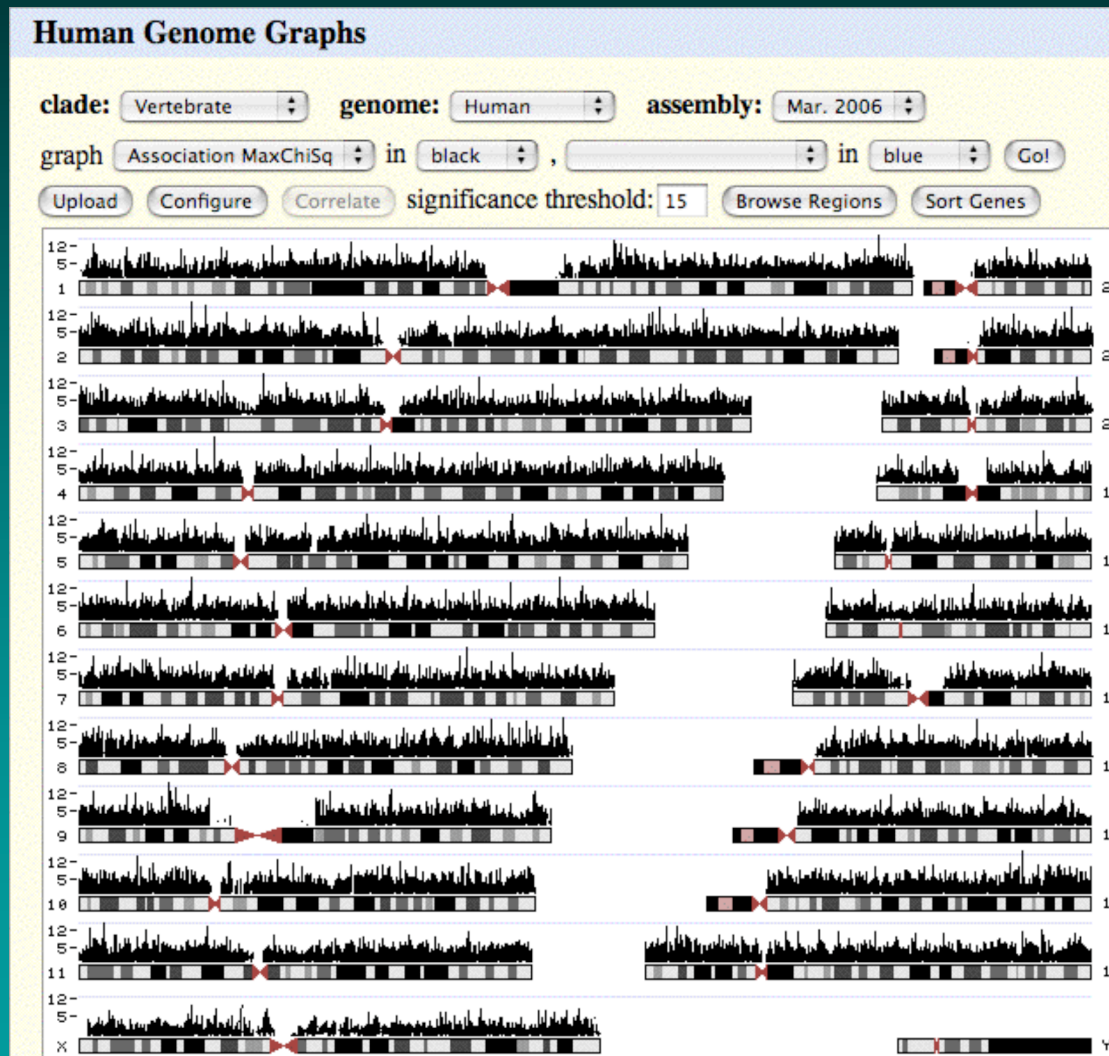


- Gap annotation
- Genomic breaks
- Codon translation at base level

New features (under review): Saving user sessions

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
Sign in to UCSC Genome Bioinformatics
Signing in enables you to save current settings into a named session, and then restore settings from the session later. If you wish, you can share named sessions with other users.
The sign-in page is handled by our wiki system click here to sign in The wiki also serves as a forum for users to share knowledge and ideas.
Session Management
Click here to reset the browser user interface settings to their defaults.
If you sign in , you will also have the option to save named sessions.
Load Settings
Load settings from another user's saved session: user: <input type="text"/> session name: <input type="text"/> <input type="button" value="submit"/>
Load settings from a local file: <input type="text"/> <input type="button" value="Browse..."/> <input type="button" value="submit"/>
Load settings from a URL (http://..., ftp://...): <input type="text"/> <input type="button" value="submit"/>
Save Settings
Save current settings to a local file file: <input type="text"/> file type returned: <input type="text" value="plain text"/> <input type="button" value="submit"/> (leave file blank to get output in browser window)

New features (in development): Whole genome graphing



SNP association study, prepublication data

GMOD Scenario #1: Search for gene by name...

UCSC Genome Bioinformatics

Genomes - Blat - Tables - **Gene Sorter** - PCR - VisiGene -

Home UCSC Human Gene Sorter Help

genome Human assembly Mar. 2006 search **foxp2** Go!

sort by Expression (GNF Atlas2) configure filter (now off) display 50 output sequence text

About the Gene Sorter

This program 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.

Home UCSC Human Gene Sorter Help

genome Human assembly Mar. 2006 search NM_148898 Go!

sort by Expression (GNF Atlas2) configure filter (now off) display 50 output sequence text

#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ Tcells	skin	adipocyte	pancreatic islets	heart	lung	kidney	liver	ovary	testis	BLASTP E-Value	Genome Position	Description
1	FOXP2	40																0	chr7 113,979,951	forkhead box P2 isoform II
2	MYOCD	n/a																n/a	chr17 12,558,831	myocardin
3	CDH26	95426																n/a	chr20 57,994,219	cadherin-like 26 isoform a
4	KCNK16	n/a																n/a	chr6 39,394,435	Pancreatic potassium chann

GMOD Scenario #1: ... and view information page

Human Gene FOXP2 Description and Page Index

Description: forkhead box P2 isoform II

Alternate Gene Symbols: CAGH44, TNRC10

Representative Refseq: [NM_148898](#) **Protein:** [O15409](#) (aka FOXP2_HUMAN or FXP2_HUMAN)

RefSeq Summary: This gene encodes an evolutionarily conserved transcription factor expressed in fetal and adult brain. This transcription factor is a member of the forkhead/winged-helix (FOX) family of transcription factors, and contains a FOX DNA-binding domain and a large polyglutamine tract. Members of the FOX family of transcription factors are regulators of embryogenesis. The product of this gene is thought to be required for proper development of speech and language regions of the brain during embryogenesis. Although a point mutation in this gene has been associated with the KE pedigree segregating developmental verbal dyspraxia, no association between mutations in this gene and another speech disorder, autism, has been found. Four alternative transcripts encoding three different isoforms have been identified.

Position: chr7:113842512-114117390

Strand: +

Genomic Size: 274879

Exon Count: 18 **CDS Exon Count:** 17

Page Index	Quick Links	UniProt Comments	Sequence	Microarray	RNA Struc
Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Methods	

Quick Links to Tools and Databases

Gene Sorter	Genome Browser	Proteome Browser	Table Schema	VisiGene	Allen Brain Atlas
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneLynx
HGNC	HPRD	Jackson Labs	OMIM	PubMed	Stanford SOURC
UniProt	Gepis Tissue				

Comments and Description Text from UniProt (Swiss-Prot/TrEMBL)

ID: [FOXP2_HUMAN](#)

DESCRIPTION: Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat- containing gene 10 protein).

FUNCTION: Transcriptional repressor that may play a role in the specification and differentiation of lung epithelium. May also play a role in developing neural, gastrointestinal and cardiovascular tissues. Involved in neural mechanisms mediating the development of speech and language.

SUBUNIT: Forms homodimers and heterodimers with FOXP1 and FOXP4. Interacts with CTBP1 (By similarity).

SUBCELLULAR LOCATION: Nuclear (Probable).

TISSUE SPECIFICITY: Isoform 1 and isoform 6 are expressed in adult and fetal brain, caudate nucleus and lung.

DEVELOPMENTAL STAGE: Expressed in the brain at 15 and 22 weeks of gestation, with a pattern of strong cortical, basal ganglia, thalamic and cerebellar expression. Highly expressed in the head and tail of nucleus caudatus and putamen. Restricted expression within the globus pallidus, with high levels in the pars interna, which provides the principal source of output from the basal ganglia to the nucleus centrum medianum thalami (CM) and the major motor relay nuclei of the thalamus. In the thalamus, present in the CM and nucleus medialis dorsalis thalami. Lower levels are observed in the nuclei anterior thalami, dorsal and ventral, and the nucleus parafascicularis thalami. Expressed in the ventrobasal complex comprising the nucleus ventralis posterior lateralis/medialis. The ventral tier of the thalamus exhibits strong expression, including nuclei ventralis anterior, lateralis and posterior lateralis pars oralis. Also expressed in the nucleus subthalamicus bilaterally and in the nucleus ruber.

DISEASE: Defects in FOXP2 are the cause of speech-language disorder 1 (SPCH1) [[MIM:602081](#)]; also known as autosomal dominant speech and language disorder with orofacial dyspraxia. Affected individuals have a severe impairment in the selection and sequencing of fine orofacial movements, which are necessary for articulation. They also show deficits in several facets of language processing (such as the ability to break up words into their constituent phonemes) and grammatical skills.

DISEASE: A chromosomal aberration disrupting FOXP2 is a cause of severe speech and language impairment. Translocation t(5;7)(q22;q31.2).

SIMILARITY: Contains 1 C2H2-type zinc finger.

SIMILARITY: Contains 1 fork-head DNA-binding domain.

DATABASE: NAME=Protein Spotlight; NOTE=Issue 51 of October 2004; WWW="http://www.expasy.org/spotlight/back_issues/spltt051.shtml".

GMOD Scenario #1: ... and view information page (2)



mRNA Secondary Structure of 3' and 5' UTRs

Region	Fold Energy	Bases	Energy/Base	Display As
5' UTR	-25.70	150	-0.171	Picture PostScript Text
3' UTR	-33.29	173	-0.192	Picture PostScript Text

The RNAfold program from the [Vienna RNA Package](#) is used to perform the secondary structure predictions and folding calculations. The estimated folding energy is in kcal/mol. The more negative the energy, the more secondary structure the RNA is likely to have.

Protein Domain and Structure Information

InterPro Domains: [Graphical view of domain structure](#)

[IPR001766](#) - Fork head transcription factor

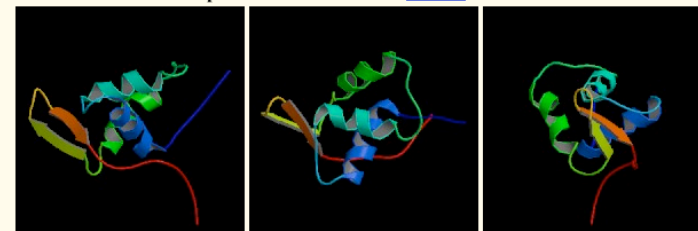
[IPR011991](#) - Winged helix repressor DNA-binding

[IPR007087](#) - Zinc finger, C2H2-type

Pfam Domains:

[PF00250](#) - Fork head domain

ModBase Predicted Comparative 3D Structure on [O15409](#)



Front

Top

Side

GMOD Scenario #1: ... and view information page (3)

Homologous Genes in Other Species (BLASTP Best Hit)

Mouse	Rat	Zebrafish	D. melanogaster	C. elegans	S. cerevisiae
Genome Browser	Genome Browser	Genome Browser	Genome Browser	Genome Browser	Genome Browser
Gene Details	Gene Details		Gene Details	Gene Details	Gene Details
Gene Sorter	Gene Sorter		Gene Sorter	Gene Sorter	Gene Sorter
Jackson Lab	RGD	Ensembl	FlyBase	WormBase	SGD
Protein Sequence	Protein Sequence		Protein Sequence	Protein Sequence	Protein Sequence
Alignment	Alignment		Alignment	Alignment	Alignment

Gene Ontology (GO) Annotations with Structured Vocabulary

Molecular Function:

[GO:0003676](#) nucleic acid binding
[GO:0003677](#) DNA binding
[GO:0003700](#) transcription factor activity
[GO:0008270](#) zinc ion binding
[GO:0043565](#) sequence-specific DNA binding
[GO:0046872](#) metal ion binding

Biological Process:

[GO:0006350](#) transcription
[GO:0006355](#) regulation of transcription, DNA-dependent

Cellular Component:

[GO:0005634](#) nucleus

Descriptions from all associated GenBank mRNAs

[AK131266](#) - Homo sapiens cDNA FLJ16201 fis, clone CTONG2008721, highly similar to Homo sapiens CAGH44 mRNA.
[AF454830](#) - Homo sapiens forkhead transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[CR749236](#) - Homo sapiens mRNA; cDNA DKFZp686H1726 (from clone DKFZp686H1726).
[AF467252](#) - Homo sapiens clone FCA1 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[BC018016](#) - Homo sapiens forkhead box P2, mRNA (cDNA clone IMAGE:4285527), complete cds.
[AY144615](#) - Homo sapiens brain forkhead/winged helix transcription factor FOXP2 isoform mRNA, complete cds; alternatively spliced.
[AF493430](#) - Homo sapiens FOXP2 short isoform (FOXP2) mRNA, complete cds.
[AF337817](#) - Homo sapiens putative forkhead/winged-helix transcription factor (FOXP2) mRNA, complete cds.
[AF467257](#) - Homo sapiens clone HF2B2 forkhead/winged helix transcription factor (FOXP2) mRNA, complete cds; alternatively spliced.
[AF467253](#) - Homo sapiens clone P22 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[U80741](#) - Homo sapiens CAGH44 mRNA, partial cds.
[AF467258](#) - Homo sapiens clone AMYG 2a.2 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[AF467259](#) - Homo sapiens clone AMYG 4a.4 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[BC126104](#) - Homo sapiens forkhead box P2, mRNA (cDNA clone MGC:161382 IMAGE:8991820), complete cds.
[AF467255](#) - Homo sapiens clone 3RACE700 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[AF467256](#) - Homo sapiens clone BA4 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[AF467254](#) - Homo sapiens clone STR7 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
[AF086040](#) - Homo sapiens full length insert cDNA clone YX52E07.
[DQ778626](#) - Homo sapiens forkhead box P2 variant 3 mRNA, complete cds, alternatively spliced.

GMOD Scenario #2 (sort of): Search by keyword

Home Genomes Blat Tables Gene Sorter PCR FAQ Help

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	genome	assembly	position or search term	image width
Vertebrate	Human	Mar. 2006	zinc ion binding	620

[Click here to reset](#) the browser user interface settings to their defaults.

[manage custom tracks](#) [configure tracks and display](#) [clear position](#)

About the Human Mar. 2006 (hg18) assembly [\(sequences\)](#)

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Genome Sequencing Consortium.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an STS marker, or a cytological band, a chromosomal coordinate range, or keywords from description of an mRNA. The following list shows examples of valid position queries for the See the [User's Guide](#) for more information.

Known Genes

PAPPA2 (AF342989) at chr1:174698930-175078591 - pappalysin 2
PAPPA2 (NM 020318) at chr1:174698930-175078591 - pappalysin 2 isoform 1
S100A1 (NM 006271) at chr1:151867497-151871137 - S100 calcium binding protein A1
S100B (NM 006272) at chr2:46842959-46849424 - S100 calcium-binding protein, beta
PAPPA (NM 002581) at chr9:117955892-118204420 - pregnancy-associated plasma protein A
TNFSF10 (NM 003810) at chr3:173706159-173723963 - tumor necrosis factor (ligand) superfamily, CALR (NM 004343) at chr19:12910423-12916303 - calreticulin precursor
PRIM1 (NM 000946) at chr12:55411637-55432413 - DNA primase small subunit, 49kDa
MMP16 (NM 005941) at chr8:89118578-89408825 - matrix metalloproteinase 16 isoform 1
ADAM21 (NM 003813) at chr14:69993970-69996374 - ADAM metalloproteinase domain 21 preproprotein
CALR3 (NM 145046) at chr19:16450888-16468003 - calreticulin 3
ADAM20 (AF029899) at chr14:70058832-70061255 - ADAM metalloproteinase domain 20
ADAM20 (NM 003814) at chr14:70058832-70071485 - ADAM metalloproteinase domain 20 preproprotein
BMX (NM 203281) at chrX:15428873-15484572 - BMX non-receptor tyrosine kinase
MMP1 (NM 002421) at chr11:102165861-102174104 - matrix metalloproteinase 1 preproprotein
APP (NM 000484) at chr21:26174733-26465003 - amyloid beta A4 protein precursor, isoform A
ADH6 (NM 000672) at chr4:100344903-100359424 - class V alcohol dehydrogenase 6
RPS27 (NM 001030) at chr1:152229853-152231247 - ribosomal protein S27
ADAM10 (NM 001110) at chr15:56675802-56829469 - ADAM metalloproteinase domain 10
SOD3 (NM 003102) at chr4:24405153-24411561 - superoxide dismutase 3, extracellular
RAG1 (NM 000448) at chr11:36546139-36557871 - recombination activating gene 1
TP73L (NM 003722) at chr3:190831910-19109756 - tumor protein p73-like
TP73L (AB016073) at chr3:190831971-191095181 - tumor protein p73-like
ADAM12 (NM 003474) at chr10:127693415-128067055 - ADAM metalloproteinase domain 12 isoform 1
PDE5A (NM 001083) at chr4:120634999-120769429 - phosphodiesterase 5A isoform 1
MBTPS2 (NM 015884) at chrX:21767675-21810794 - membrane-bound transcription factor protease, TP53 (NM 000546) at chr17:7512465-7531642 - tumor protein p53
S100A3 (NM 002960) at chr1:151786433-151788358 - S100 calcium binding protein A3
STAMBIP (NM 213622) at chr2:73909655-73943518 - STAM binding protein
RIMS1 (NM 014989) at chr6:72653448-73167137 - regulating synaptic membrane exocytosis 1
TP73 (NM 005427) at chr1:3558989-3639716 - tumor protein p73
EWSR1 (NM 005243) at chr22:27994283-28026505 - Ewing sarcoma breakpoint region 1 isoform EWS

Non-Human RefSeq Genes

AT1G22440 at chr6:66603562-66604647 - (NM 102093) oxidoreductase/ zinc ion binding
AT4G22110 at chr6:66603859-66604647 - (NM 118332) oxidoreductase/ zinc ion binding
AT4G22110 at chr6:66603859-66604647 - (NM 202860) oxidoreductase/ zinc ion binding
AT5G24760 at chr1:79759536-79760470 - (NM 203102) oxidoreductase/ zinc ion binding
AT5G24760 at chr1:79759536-79760470 - (NM 122385) oxidoreductase/ zinc ion binding
AT5G242250 at chr6:66603952-66604491 - (NM 123590) oxidoreductase/ zinc ion binding
AT2G32600 at chr19:21944339-2196550 - (NM 128819) nucleic acid binding / zinc ion binding
APC11 at chr17:77445680-77451230 - (NM 11461) APC11; protein binding / zinc ion binding
AT1G32780 at chr6:66603790-66604488 - (NM 103012) oxidoreductase/ zinc ion binding
AT1G32780 at chr4:100216469-100568747 - (NM 103012) oxidoreductase/ zinc ion binding

GMOD Scenario #3: Customized report on aspects of gene

Home Gene Sorter Filter

On this page you can restrict which genes appear in the main table based on the values in any column. Click the *submit* button to return to the main Gene Sorter page with the current filter settings applied.

Quickly obtain a list of gene names that pass the filter:

Filter Controls for Displayed Columns:

Name - Gene Name/Select Gene

Name search (including * and ? wildcards):

Include if any words in search term match.

Limit to items (no wildcards) in list:

- Exon count
- GO terms
- Description

Home Configure Gene Sorter Help

Columns: Settings:

Expression ratio colors: Show all splicing variants:

Name	On	Position	Description	Configuration
#	<input type="checkbox"/>	▼	Item Number in Displayed List/Select Gene	n/a
Name	<input checked="" type="checkbox"/>	▲▼	Gene Name/Select Gene	n/a
UniProt	<input type="checkbox"/>	▲▼	UniProt (SwissProt/TrEMBL) Protein Display ID	n/a
UniProt Acc	<input type="checkbox"/>	▲▼	UniProt (SwissProt/TrEMBL) Protein Accession	n/a
PDB	<input type="checkbox"/>	▲▼	Protein Data Bank	n/a
Gene Ontology	<input checked="" type="checkbox"/>	▲▼	Gene Ontology (GO) Terms Associated with Gene	n/a
M. Vidal P2P	<input type="checkbox"/>	▲▼	Human Protein-Protein Interaction Network from Marc Vidal	n/a
E. Wanker P2P	<input type="checkbox"/>	▲▼	Human Protein-Protein Interaction Network from Erich Wanker	n/a
HPRD P2P	<input type="checkbox"/>	▲▼	Human Protein-Protein Interaction Network from the Human Reference Protein Database	n/a
Description	<input checked="" type="checkbox"/>	▲	Short Description Line/Link to Details Page	n/a

Home UCSC Human Gene Sorter Help

genome assembly search

sort by display output

#	Name	Exon Count	Gene Ontology
1	FOXP2	18	'nucleic acid binding' 'DNA binding' 'transcription factor activity' 'nucleus' 'transcription' 'regulation of transcription, DNA-dependent'

```
#num  name  exonCount  go  description
1     FOXP2    18      'nucleic acid binding','DNA binding','transcription factor activity','nucleus','transcription','regulation of transcription, DNA-dependent'
```

GMOD Scenario #3 Alternate: Customized report on aspects of gene

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. For more complex queries, you may want to use our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: genome: assembly:

group: track:

table:

region: genome position

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Select Position

Known Genes

[FOXP2 \(NM_148898\) at chr7:113842512-114117390](#) - forkhead box P2 isoform II
[FOXP2 \(NM_014491\) at chr7:113842512-114117390](#) - forkhead box P2 isoform I
[FOXP2 \(CR749236\) at chr7:113513851-114097274](#) - Hypothetical protein DKFZp686H1726.
[FOXP2 \(CR749236\) at chr7:113513851-114097274](#) - Hypothetical protein DKFZp686H1726.

[FOXP](#)
[FOXP](#)
[FOXP](#)
[CTBP](#)
[CTBP](#)

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Table Browser

Re: Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. For more complex queries, you may want to use our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: genome: assembly:

group: track:

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

- Exon count
- GO terms
- Swiss-Prot disease description

GMOD Scenario #3: Customized report on gene, cont.

Linked Tables

<input checked="" type="checkbox"/>	go	goaPart	
<input type="checkbox"/>	hg18	all_mrna	Summary info about a patSpace alignment
<input type="checkbox"/>	hg18	bioCycPathway	BioCyc Pathway to Known Gene cross reference
<input type="checkbox"/>	hg18	blastKGPep04	A predicted peptide - linked to a predicted gene.
<input type="checkbox"/>	hg18	blastKGRef04	Mapping table for tblastn protein tracks
<input type="checkbox"/>	hg18	ceBlastTab	Tab-delimited blast output file
<input type="checkbox"/>	hg18	cgapAlias	C <input type="checkbox"/> hg18 knownIsoforms Links together various transcripts of a gene into a cluster
<input type="checkbox"/>	hg18	dmBlastTab	T: <input type="checkbox"/> hg18 knownToAllenBrain Map known gene to some other id
<input type="checkbox"/>	hg18	drBlastTab	T: <input checked="" type="checkbox"/> hg18 knownToEnsembl Map known gene to some other id
<input type="checkbox"/>	hg18	dupSpMrna	D <input type="checkbox"/> hg18 knownToHprd Map known gene to some other id
<input type="checkbox"/>	hg18	foldUtr3	In <input type="checkbox"/> hg18 knownToLocusLink Map known gene to some other id
<input type="checkbox"/>	hg18	foldUtr5	In <input type="checkbox"/> hg18 knownToPfam Map known gene to some other id
<input type="checkbox"/>	hg18	gbCdnalInfo	<input type="checkbox"/> hg18 knownToRefSeq Map known gene to some other id
<input type="checkbox"/>	hg18	gbSeq	<input type="checkbox"/> hg18 knownToU133 Map known gene to some other id
<input type="checkbox"/>	hg18	gbStatus	<input type="checkbox"/> hg18 knownToU133Plus2 Map known gene to some other id
<input type="checkbox"/>	hg18	gnfAtlas2Distance	D <input type="checkbox"/> hg18 knownToU95 Map known gene to some other id
<input type="checkbox"/>	hg18	gnfU95Distance	D <input type="checkbox"/> hg18 knownToVisiGene Map known gene to some other id
<input type="checkbox"/>	hg18	imageClone	G <input type="checkbox"/> hg18 mmBlastTab Tab-delimited blast output file
<input type="checkbox"/>	hg18	keggPathway	K <input type="checkbox"/> hg18 mrnaOrientInfo Extra information on ESTs - calculated by polyInfo program
<input type="checkbox"/>	hg18	kgAlias	L: <input type="checkbox"/> hg18 mmaRefseq Cross reference table between refseq and mRNA IDs based on LocusLink
			<input type="checkbox"/> hg18 rmBlastTab Tab-delimited blast output file
			<input type="checkbox"/> hg18 scBlastTab Tab-delimited blast output file
			<input type="checkbox"/> hg18 seq
			<input type="checkbox"/> hg18 spMrna The best representative mRNA for a protein
			<input checked="" type="checkbox"/> proteome spDisease A cross-reference table between Swiss-Prot IDs and disease description.
			<input type="checkbox"/> proteome spOldNew
			<input type="checkbox"/> proteome uniProtAlias
			<input type="checkbox"/> uniProt displayId Relate ID and primary accession. A good table to use just get handle on all records.

Allow Selection From Checked Tables

Select Fields from hg18.knownGene

<input type="checkbox"/>	name	Name of gene
<input type="checkbox"/>	chrom	Reference sequence chromosome or scaffold
<input type="checkbox"/>	strand	+ or - for strand
<input type="checkbox"/>	txStart	Transcription start position
<input type="checkbox"/>	txEnd	Transcription end position
<input type="checkbox"/>	cdsStart	Coding region start
<input type="checkbox"/>	cdsEnd	Coding region end
<input type="checkbox"/>	exonCount	Number of exons
<input type="checkbox"/>	exonStarts	Exon start positions
<input type="checkbox"/>	exonEnds	Exon end positions
<input type="checkbox"/>	proteinID	SWISS-PROT ID
<input type="checkbox"/>	alignID	Unique identifier for each (known gene, alignment position) pair

go.goaPart fields

<input type="checkbox"/>	dbObjectId
<input checked="" type="checkbox"/>	dbObjectSymbol
<input type="checkbox"/>	notId
<input type="checkbox"/>	goId
<input type="checkbox"/>	aspect

hg18.knownToEnsembl fields

<input checked="" type="checkbox"/>	name	Same as name field in known gene
<input type="checkbox"/>	value	Other id

proteome.spDisease fields

<input type="checkbox"/>	accession	SWISS-PROT accession number
<input type="checkbox"/>	displayID	SWISS-PROT display ID
<input checked="" type="checkbox"/>	diseaseDesc	disease description

GMOD Scenario #3: Report on aspects of gene, cont.(2)

hg18.kgXref fields

<input type="checkbox"/>	kgID	Known Gene ID
<input type="checkbox"/>	mRNA	mRNA ID
<input type="checkbox"/>	spID	SWISS-PROT protein Accession number
<input type="checkbox"/>	spDisplayID	SWISS-PROT display ID
<input checked="" type="checkbox"/>	geneSymbol	Gene Symbol
<input type="checkbox"/>	refseq	RefSeq ID
<input type="checkbox"/>	protAcc	NCBI protein Accession number
<input type="checkbox"/>	description	Description

proteome.spDisease fields

<input type="checkbox"/>	accession	SWISS-PROT accession number
<input type="checkbox"/>	displayID	SWISS-PROT display ID
<input checked="" type="checkbox"/>	diseaseDesc	disease description

Select Fields from hg18.knownGene

<input checked="" type="checkbox"/>	name	Name of gene
<input type="checkbox"/>	chrom	Reference sequence chromosome or scaffold
<input type="checkbox"/>	strand	+ or - for strand
<input type="checkbox"/>	txStart	Transcription start position
<input type="checkbox"/>	txEnd	Transcription end position
<input type="checkbox"/>	cdsStart	Coding region start
<input type="checkbox"/>	cdsEnd	Coding region end
<input checked="" type="checkbox"/>	exonCount	Number of exons
<input type="checkbox"/>	exonStarts	Exon start positions
<input type="checkbox"/>	exonEnds	Exon end positions
<input type="checkbox"/>	proteinID	SWISS-PROT ID
<input type="checkbox"/>	alignID	Unique identifier for each (known gene, alignment position) pair

go.term fields

<input type="checkbox"/>	id
<input checked="" type="checkbox"/>	name
<input type="checkbox"/>	term_type
<input type="checkbox"/>	acc
<input type="checkbox"/>	is_obsolete
<input type="checkbox"/>	is_root

- Exon count
- GO terms
- Swiss-Prot disease description

```
#hg18.knownGene.name hg18.knownGene.exonCount go.term.name hg18.kgXref.geneSymbol proteome.spDisease.diseaseDesc
CR749236 23 n/a FOXP2 n/a
CR749236 23 nucleic acid binding,nucleus,zinc ion binding, FOXP2 n/a
BC018016 3 n/a FOXP2 n/a
AY144615 18 nucleic acid binding,DNA binding,transcription factor activity,nucleus,transcription,regulation of transcription, DNA-dependent,zinc ion binding,s
AY144615 18 n/a FOXP2 n/a
NM_148898 18 nucleic acid binding,DNA binding,transcription factor activity,nucleus,transcription,regulation of transcription, DNA-dependent,zinc ion binding,s
NM_014491 17 nucleic acid binding,DNA binding,transcription factor activity,nucleus,transcription,regulation of transcription, DNA-dependent,zinc ion binding,s
```

GMOD Scenarios 4 & 5:

Bulk queries and external data integration;

Compare user gene set to UCSC Known Genes

- How many user genes are not in Known Genes ?
- How well conserved across different species are the genes unique to the user gene set ?

GMOD Scenarios 4 & 5: Loading external data

Add Custom Tracks

clade genome assembly [hg18]

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [GFF](#), [GTF](#), [WIG](#) or [PSL](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Publicly available custom tracks are listed [here](#). Examples are [here](#).

Paste URLs or data: Or upload:

Optional track documentation: Or upload:

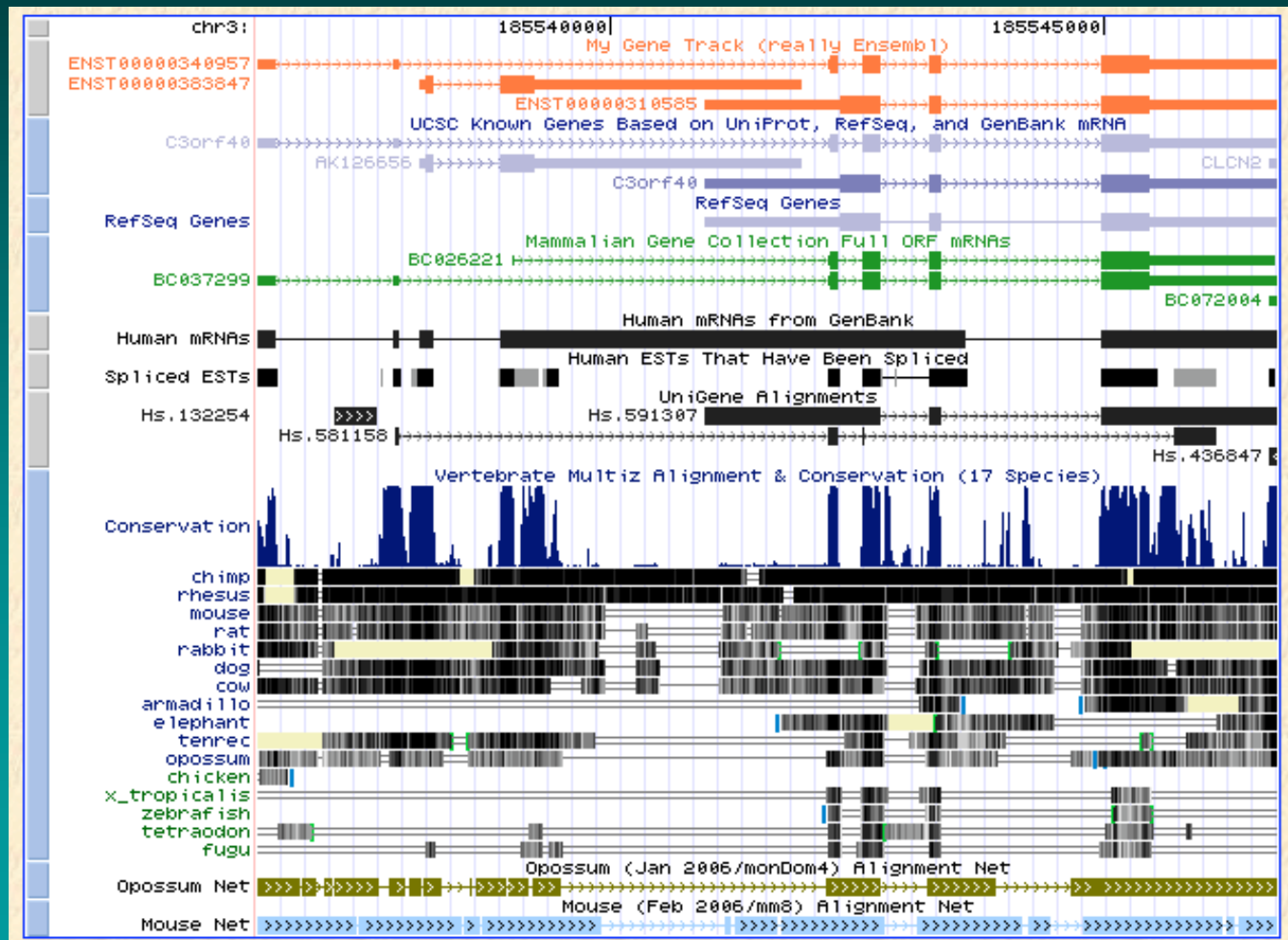
Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.

Manage Custom Tracks

genome: Human assembly: Mar. 2006 [hg18]

Name	Description	Type	Doc	Items	Pos	delete	update
My Genes	My Gene Track (really Ensembl)	gff	Y	59069	chr3:	<input type="checkbox"/>	<input type="checkbox"/>

GMOD Scenarios 4 & 5: Loading external data, cont.



GMOD Scenarios 4 & 5: Intersection on whole dataset

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. For more complex queries, you may want to use our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: genome: assembly:

group: track:

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: plain text gzip

Intersect with My Genes

Select a group, track and table to intersect with:

group: track:

table:

These combinations will maintain the gene/alignment structure (if any) of My Genes:

- All My Genes records that have any overlap with Known Genes
- All My Genes records that have no overlap with Known Genes
- All My Genes records that have at least % overlap with Known Genes
- All My Genes records that have at most % overlap with Known Genes

GMOD Scenarios 4 & 5: Intersection on whole dataset, cont.

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See [Using the Table Browser](#) for a description of the controls in this form. For more complex queries, you may want to use our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: genome: as
group: track:
table:
region: genome position
identifiers (names/accessions):
filter:
intersection with knownGene:
correlation:
output format:
output file: (leave blank)
file type returned: plain text gzip compressed

Note: Intersection doesn't work with all fields or selected fields

Output ct_MyGenes as Custom Track

Custom track header:

name=

description=

visibility=

url=

Create one BED record per:

- Whole Gene
- Upstream by bases
- Exons plus bases at each end
- Introns plus bases at each end
- 5' UTR Exons
- Coding Exons
- 3' UTR Exons
- Downstream by bases

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

Kent's UI Guidelines

- Keep it reliable
- Keep it fast
- Label everything in plain English
- Put the most commonly used controls on the top of the page
- Keep it as simple as possible (but no simpler)
- Try to make options work together in an orthogonal way
- Remember your users are *intelligent* professionals. Don't dumb things down; complexity comes with the territory
- Don't change the site unnecessarily once people have gotten used to it.

User interface challenges: User-configurable ordering

Home Configure Gene Sorter

submit Columns: Settings:

Expression ratio colors: Show all splicing variants:

Name	On	Position	Description	Configuration
#	<input checked="" type="checkbox"/>	▼	Item Number in Displayed List/Select Gene	n/a
Name	<input checked="" type="checkbox"/>	▲▼	Gene Name/Select Gene	n/a
UniProt	<input type="checkbox"/>	▲▼	UniProt (SwissProt/TrEMBL) Protein Display ID	n/a
UniProt Acc	<input type="checkbox"/>	▲▼	UniProt (SwissProt/TrEMBL) Protein Accession	n/a
RefSeq	<input type="checkbox"/>	▲▼	NCBI RefSeq Gene Accession	n/a
Entrez Gene	<input type="checkbox"/>	▲▼	NCBI Entrez Gene/LocusLink ID	n/a
GenBank	<input type="checkbox"/>	▲▼	GenBank mRNA Accession	n/a
Ensembl	<input type="checkbox"/>	▲▼	Ensembl Transcript ID	n/a
GNF Atlas 2 ID	<input type="checkbox"/>	▲▼	ID of Associated GNF Atlas 2 Expression Data	n/a
VisiGene	<input checked="" type="checkbox"/>	▲▼	UCSC VisiGene In Situ Image Browser	n/a
Allen Brain	<input type="checkbox"/>	▲▼	Allen Brain Atlas In Situ Images of Adult Mouse Brains	n/a
U133 ID	<input type="checkbox"/>	▲▼	ID of Associated Affymetrix U133 Expression Data	n/a
U133Plus2 ID	<input type="checkbox"/>	▲▼	ID of Associated Affymetrix U133 Plus 2.0 Expression Data	n/a
U95 ID	<input type="checkbox"/>	▲▼	ID of Associated Affymetrix U95 Expression Data	n/a

Find:

Configure Tracks

Control tracks in all groups here: Control track visibility more selectively below.

Mapping and Sequencing Tracks		<input type="button" value="hide all"/>	<input type="button" value="show all"/>	<input type="button" value="default"/>	<input type="button" value="submit"/>	2	Group	
Track Order:								
Base Position	dense ▼	Chromosome position in bases. (Clicks here zoom in 3x)						
Gap	hide ▼	Gap Locations	11	map ▼				
BAC End Pairs	hide ▼	BAC End Pairs	15	map ▼				
GC Percent	hide ▼	Percentage GC in 20,000-Base Windows	23	map ▼				
Short Match	hide ▼	Perfect Matches to Short Sequence (AAAAA)	99	map ▼				
Restr Enzymes	hide ▼	Restriction Enzymes from REBASE	99.9	map ▼				

Genes and Gene Prediction Tracks		<input type="button" value="hide all"/>	<input type="button" value="show all"/>	<input type="button" value="default"/>	<input type="button" value="submit"/>	3	Group
Track Order:							
FlyBase Genes	pack ▼	FlyBase Protein-Coding Genes	34	genes ▼			
RefSeq Genes	dense ▼	RefSeq Genes	35	genes ▼			
FB Noncoding	pack ▼	FlyBase Noncoding Genes	35	genes ▼			
N-SCAN	hide ▼	N-SCAN Gene Predictions	45.1	genes ▼			
Geneid Genes	hide ▼	Geneid Gene Predictions	49	genes ▼			
Genscan Genes	hide ▼	Genscan Gene Predictions	50	genes ▼			
Augustus Genes	hide ▼	Augustus Gene Predictions	51.7	genes ▼			
Human Proteins	pack ▼	Human(hg17) proteins mapped by chained tBLASTn	142	genes ▼			

User interface challenges: Track grouping to avoid overload

The image displays a complex genomic browser interface with numerous tracks organized into several categories. Each track includes a 'hide' dropdown menu, illustrating the challenge of track grouping to avoid overload.

Mapping and Sequencing Tracks				
Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense	hide	dense	hide	hide
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
hide	hide	hide	hide	hide
Fosmid End Pairs	GC Percent	WSSD Duplication	Short Match	Restr Enzymes
hide	hide	hide	hide	hide

Phenotype and Disease Associations		Comparative Genomics			
RGD OTL	Locus Variants	Conservation	Fugu Blat	Fugu Chain	Fugu Net
hide	hide	pack	hide	hide	hide
Genes and Gene Prediction Tracks		Tetraodon Chain	Tetraodon Net	Zebrafish Chain	Zebrafish Net
Known Genes	CCDS	RefSeq Genes	Other Ref	X_tropicalis Chain	X_tropicalis Net
pack	hide	dense	hide	hide	hide
Vega Genes	Vega Pseudogenes	Ensembl Genes	AceView	Chicken Chain	Chicken Net
hide	hide	hide	hide	hide	hide
N-SCAN	SGP Genes	Geneid Genes	GeneScan	Cow Chain	Cow Net
hide	hide	hide	hide	hide	hide
Augustus Genes	Retroposed Genes	Superfamily	Yale Pseu	Dog Net	Rat Chain
hide	hide	hide	hide	hide	hide
sno/miRNA	ExonWalk			Rhesus Chain	Rhesus Net
hide	dense			hide	hide

Variation and Repeats				
SNPs	SNP Arrays	HapMap LD	Tajima's D	Tetraodon Ecores
dense	hide	hide	hide	hide
SNP Recomb Rates	SNP Recomb Hots	Segmental Dups	Structural V	X_tropicalis Chain
hide	hide	hide	hide	hide
Simple Repeats	Microsatellite	RIPs	Self Chain	Opossum Chain
hide	hide	hide	hide	hide

ENCODE Chromatin Immunoprecipitation				
Affy pVal	Affy Sites	Affy Strict pVal	Affy Strict Sig	Affy Strict Sites
hide	hide	hide	hide	hide
GIS ChIP-PET	LI ChIP Various	LI gIF ChIP	LI Ng gIF ChIP	LI Ng TAFI
hide	hide	hide	hide	IMR90
LI Ng Validation	Sanger ChIP	Sanger ChIP Hits	Stanf ChIP	Stanf ChIP Score
hide	hide	hide	hide	hide
UCD Ng ChIP	UT-Austin ChIP	Uppsala ChIP	Uppsala ChIP Butyrate	Yale STAT1 pVal
hide	hide	hide	hide	hide
Yale STAT1 Sig	Yale STAT1 Sites	Yale ChIP pVal	Yale ChIP Signal	Yale ChIP Sites
hide	hide	hide	hide	hide

ENCODE Chromosome, Chromatin and DNA Structure				
BU ORChID	Duke/NHGRI DNase	Stanf Meth	Stanf Meth Score	UNC FAIRE
hide	hide	hide	hide	hide
UT-Austin STAGE	UVa DNA Rep	UVa DNA Rep Seg	UVa DNA Rep Ori	UVa DNA Rep TR50
hide	hide	hide	hide	hide
UW/Reg OCP DNaseI Sens	UW DNase GM			
hide	hide			

ENCODE Transcript Levels				
Affy RNA Signal	Affy Transfrags	BU First Exon	Riken CAGI	
hide	hide	hide	hide	

ENCODE Comparative Genomics				
Consens Elements	TBA Alignment	TBA Cons	TBA Elements	MLAGAN Alignment
hide	hide	hide	hide	hide
MLAGAN Cons	MLAGAN Elements	MAVID Alignment	MAVID Cons	MAVID Elements
hide	hide	hide	hide	hide
Align Agree	Align Gaps	DLESS		
hide	hide	hide		

ENCODE Variation

User interface challenges: Composite tracks to group similar data

Affymetrix ChIP/Chip (retinoic acid-treated HL-60 cells) Sites

Display mode:

Select subtracks:

All

Timepoint

0hrs

2hrs

8hrs

32hrs

Factor

Brg1

CEBPe

CTCF

H3K27me3

H4Kac4

P300

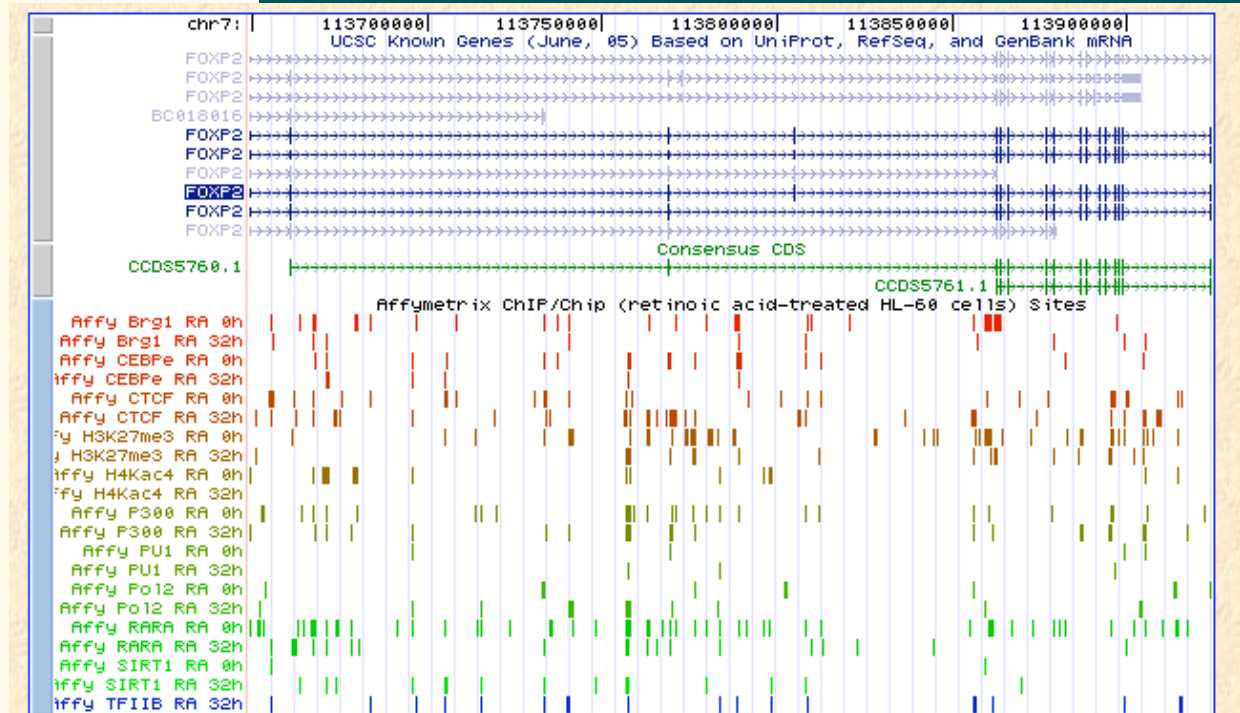
PUI

RARA

Pol2

SIRT1

TFIIB



Show checkboxes for: Only selected subtracks All subtracks

Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 0hrs) Sites

Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 2hrs) Sites

Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 8hrs) Sites

User Support and Training

- FAQs: <http://genome.cse.ucsc.edu/FAQ/>
- questions? genome@soe.ucsc.edu
archived answers:
<http://genome.ucsc.edu/contacts.html>
- OpenHelix: <http://www.openhelix.com/>
 - Classes, seminars
 - Free online tutorial
 - Quick reference cards

Thanks!

- UCSC Genome Browser Team:
 - David Haussler – PI
 - Jim Kent – Browser Concept, BLAT, Team Leader
 - Donna Karolchik – Engineering Mgr, Docs & Training
 - Mark Diekhans, Fan Hsu, Angie Hinrichs, Kate Rosenbloom, Hiram Clawson, Rachel Harte, Heather Trumbower, Galt Barber, Andy Pohl - Engineering
 - Robert Kuhn (mgr), Ann Zweig, Kayla Smith, Brooke Rhead, Archana Thakkapallayil – QA/Support
 - Jorge Garcia, Chester Manuel, Victoria Lin, Erich Weller, Paul Tatarsky – KiloKluster, Sys-admin
- Funding:
 - National Human Genome Research Institute
 - Howard Hughes Medical Institute
 - National Cancer Institute