The UCSC Genome Browser
Introduction

Osvaldo Graña
Bioinformatics Unit, CNIO
### Organization of genomic data...

#### Genome backbone: base position number
- chromosome band
- sts sites
- gap locations
- known genes
- predicted genes
- microarray/expression data
- evolutionary conservation
- SNPs
- repeated regions
- more...

### Annotation Tracks
- sequence
- Links out to more data
A sample of what we will find:

- **gene details**
- **official sequence**
- **comparisons**
- **SNPs**
Introduction
Basic Searches
Understanding Displays
Get Details or Sequences
Sequence Searches (BLAT)
\textit{in silico} PCR
Proteome Browser
VisiGene Browser
Exercises

UCSC Genome Browser: http://genome.ucsc.edu
The UCSC Homepage: http://genome.ucsc.edu

UCSC Genome Bioinformatics

Genome Browser
- Encode
- Blat
- Table Browser
- Gene Sorter
- In Silico PCR
- Genome Graphs
- Galaxy
- VisiGene
- Proteome Browser
- Utilities
- Downloads
- Release Log
- Custom Tracks
- Mirrors
- Archives
- Training
- Credits
- Publications
- Cite Us
- Licenses
- Jobs
- Staff
- Contact Us

navigate

navigate

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 resources with our tools. The Genome Browser allows you to navigate and scroll over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, knockdown and other information on groups of genes that can be related to the expression of some disease or access to the underlying database. VisiGene lets you browse through those data and then allows you to upload and display genome-wide data sets.

navigate

General information

specific information—new features, current status, etc.

News

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

6 April 2007 - New UCSC Gene Prediction Set Released

We are pleased to announce the release of a new gene prediction set, UCSC Genes, on the latest human Genome Browser (hg18, UCSC Build 36). This annotation, which includes putative non-coding genes as well as protein-coding genes and 99.9% of RefSeq genes, is the next generation of the Known Genes set that has been providing for several years and superseded the existing Known Genes annotation on the hg17 assembly.

The UCSC Genes is a moderately conservative prediction set based on data from RefSeq, GenBank, and Uniprot. Each entry requires the support of one GenBank mRNA sequence plus at least one additional line of evidence, with the exception of RefSeq RNAs, which require no additional evidence. Some of the non-coding transcripts in the set may actually code for protein, but the evidence for the associated protein is weak at best. Compared to RefSeq, this gene set generally has about 10% more protein-coding genes, approximately five times as many putative non-coding genes, and about twice as many splice variants.

The UCSC Genes set is produced using a computational pipeline developed at UCSC by Jim Kent, Chuck Suggest and Mark Dickinson. For detailed information about the process used to construct the genes set, see the track description page. In upcoming months, we plan to release UCSC Genes sets on several organisms in addition to human.

As part of this change, we are now using our own, UCSC Genes accession numbers as the primary key into the underlying knownGene table, rather than the GenBank mRNA accessions used in the previous Known Genes accessions.

We will continue to provide the older Known Genes accession numbers and converting between the two sets:

- knownGeneOfK: new name for table underlying K
- lgxRefChid: new name for table that contains converted converting the knownGeneOfIds to the older UCSC Genes IDs
- lgxTcRgi: data for converting old Known Gene IDs to the newer UCSC Genes IDs

We’d like to acknowledge the many people affiliated with the UCSC Genome Bioinformatics group who worked hard to release this new annotation. developers Jim Kent, Mark Dickinson, and Fan Hua (with technical support from several other engineers in the group), David Haines, our splendid QA team - Archana Thakare, Ann Zwenk, Robert Kohn, Kaela Smith, and Brooke Riedel, our build engineer - Andy Volt, and our system group. We'd also like to thank Chuck Suggest for his input, the people and organizations maintaining the RefSeq, Uniprot, and GenBank databases, and the scientists worldwide who have contributed to them. If you have any questions about this new release, feel free to contact us at genome@genomics.ucsc.edu (general questions) or genome-questions@genomics.ucsc.edu (implementation-specific questions)

27 March 2007 - Stickelback Assembly Released in Genome Browser: We have released a Genome Browser and Blat server for the Feb 2006 v1.0 draft assembly of Gasterosteus aculeatus produced by the Broad Institute. Read more
Use this Gateway to search by:

- Gene names, symbols
- Chromosome number: chr7, or region: chr11:1038475-1075482
- Keywords: kinase, receptor
- IDs: NP, NM, OMIM, and more…

See lower part of page for help with format
The Genome Browser Gateway
start page choices, April 2007

Make your Gateway choices:
1. Select Clade
2. Select genome = species: search 1 species at a time
3. Assembly: the official backbone DNA sequence
4. Position: location in the genome to examine
5. Image width: how many pixels in display window; 5000 max
6. Configure: make fonts bigger + other choices

Configure Image:
- Image width: 620
- Text size: small
- Display chromosome ideographs in main graphic:
- Show light blue vertical grid
- Display labels to the left of tracks
- Display track description above each track
Sample search: human, March 2006 assembly, tp53

Select from results list

ID search may go right to a viewer page, if unique
Overview of the whole Genome Browser page (mature release)

Genome viewer section

Groups of data

- Mapping and Sequencing Tracks
- Phenotype and Disease Tracks
- Genes and Gene Prediction Tracks
- mRNA and EST Tracks
- Expression and Regulation
- Comparative Genomics
- Variation and Repeats
- ENCODE Tracks
Different species, different tracks, same software

- Species may have different data tracks
- Layout, software, functions the same
Sample Genome Viewer image, TP53 region

UCSC Genome Browser on Human Mar. 2006 Assembly

- base position
- STS markers
- UCSC genes
- RefSeq genes
- MGC clones
- ESTs
- 17 species compared
- single species compared
- SNPs
- repeats
Visual Cues on the Genome Browser

Tick marks; a single location (STS, SNP)

Intron, and direction of transcription <<< or >>>

Track colors may have meaning—for example, UCSC Gene track:
- If there is a corresponding PDB entry, = black
- If there is a corresponding reviewed/validated seq, = dark blue
- If there is a non-RefSeq seq, = lightest blue

For some tracks, the height of a bar is increased likelihood of an evolutionary relationship (conservation track)
Options for Changing Images: Upper Section

- Change your view or location with controls at the top
- Use “base” to get right down to the nucleotides
- Configure: to change font, window size, more…

- Walk left or right
- Zoom in
- Zoom out
- Specify a position
- Click to zoom 3x and re-center
- Fonts, window, more
Annotation Track display options

- Some data is ON or OFF by default
- Menu links to info about the tracks: content, methods
- You change the view with pulldown menus
- After making changes, REFRESH to enforce the change
Annotation Track options, defined

- **Hide**: removes a track from view

- **Dense**: all items collapsed into a single line

- **Squish**: each item = separate line, but 50% height + packed

- **Pack**: each item separate, but efficiently stacked (full height)

- **Full**: each item on separate line
Reset, Hide, Configure or Refresh to change settings

- You control the views
- Use pulldown menus
- Configure options page

enforce any changes (hide, full, squish…)

reset, back to defaults

start from scratch
Annotation Track options, if altered…. important point: the browser remembers!

- Session information (the position you were examining)
- Track choices (squish, pack, full, etc)
- Filter parameters (if you changed the colors of any items, or the subset to be displayed)

...are all saved on your computer. When you come back in a couple of days to use it again, these will still be set. You may—or may not—intend this.

To clear your “cart” or parameters, click default tracks

OR

Click here to reset the browser user interface settings to their defaults.
Click Any Viewer Object for Details

Click the item

New web page opens

Example: click your mouse anywhere on the TP53 line

Many details and links to more data about TP53
Not all genes have this much detail.

Different annotation tracks carry different data.

Informative description
Other resource links
Links to sequences
Microarray data
mRNA secondary structure
Protein domains/structure
Homologs in other species
Gene Ontology™ descriptions
mRNA descriptions
Pathways
Get DNA, with Extended Case/Color Options

- Use the DNA link at the top
- Plain or Extended options
- Change colors, fonts, etc.
Get Sequence from Details Pages

Click a track, go to **Sequence** section of details page

![Sequence section on detail page](image)

Click the item
UCSC Genome Browser Agenda

- Introduction and Credits
- Basic Searches
- Understanding Displays
- Get Details or Sequences
- Sequence Searches (BLAT)
- in silico PCR
- Proteome Browser
- VisiGene Browser
- Exercises

UCSC Genome Browser:  http://genome.ucsc.edu
Accessing the BLAT tool

BLAT = BLAST-like Alignment Tool
- Rapid searches by INDEXING the entire genome
- Works best with high similarity matches
- See documentation and publication for details
  - Kent, WJ. Genome Res. 2002. 12:656
BLAT tool overview: www.openhelix.com/sampleseqs.html

- Make choices
- Paste one or more sequences
- Or upload

DNA limit 25000 bases
Protein limit 10000 aa
25 total sequences

Submit
Results with demo sequences, settings default; sort = Query, Score

- Score is a count of matches—higher number, better match
- Click **browser** to go to Genome Browser image location (next slide)
- Click **details** to see the alignment to genomic sequence (2\(^{nd}\) slide)
From **browser** click in BLAT results

A new line with **your Sequence from BLAT Search** appears!

Watch out for reading frame! Click - - - -> to flip frame

Base position = full **and** zoomed in enough to see amino acids
**BLAT results, alignment details**

Your query

Genomic match, color cues

Side-by-side alignment

*Aligned Blocks with gaps <= 8 bases are merged for this display when only one sequence is shown. In such cases, sequences are of the same size.*
UCSC Genome Browser Agenda

- Introduction and Credits
- Basic Searches
- Understanding Displays
- Get Details or Sequences
- Sequence Searches (BLAT)
- in silico PCR
- Proteome Browser
- VisiGene Browser
- Exercises

UCSC Genome Browser:  http://genome.ucsc.edu
In-Silico PCR: find genomic sequence using primers

- Select genome
- Enter primers
- Minimum 15 bases
- Flip reverse primer?
- Submit

(note: the tool does not handle ambiguous bases at this time—don’t use Ns)
In Silico PCR: results

- Genomic location shown, links to Genome Viewer
- Product size shown
- Your primers displayed, flipped if necessary
- Predicted genomic sequence shown
- Primer melting temperatures provided

Primers:

```python
>chr4:75082832-75083261
GCCACAGTGCTCCGGA AGATTACTGATCATTT
```

Primer Melting Temperatures:

- Forward: 60.5°C gccacagtgcctccgga
- Reverse: 34.6°C aatgatcaagtaactt

The temperature calculations are done assuming 50 mM salt and 50 nM annealing oligo concentration. The code to calculate the melting temp comes from Primer3.
- Access from homepage or UCSC Gene pages
- Exon diagram, amino acids...
- Many protein properties (pI, mw, composition, 3D...)

Proteome Browser

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

UCSC Proteome Browser Gateway

The UCSC Proteome Browser was created by the Genome Bioinformatics Group of the Software Copyright © The Regents of the University of California. All rights reserved.

Enter a gene symbol or a Swiss-Prot/TrEMBL protein ID: brca1

more data
VisiGene: Biological image data

Expression in cells and tissues; mRNA or protein

Search by symbol, author, body parts, IDs, stages...
UCSC Genome Browser Agenda

- Introduction and Credits
- Basic Searches
- Understanding Displays
- Get Details or Sequences
- Sequence Searches (BLAT)
- *in silico* PCR
- Proteome Browser
- VisiGene Browser
- Exercises

UCSC Genome Browser:  http://genome.ucsc.edu
Notice:

- The materials and slides offered are for non-commercial use only. Reproduction, distribution and/or use for commercial purposes is strictly prohibited.

- Copyright 2006, OpenHelix, LLC