Genome Browsers - The UCSC Genome Browser

Background

The UCSC Genome Browser is a well-curated site that provides users with a view of gene or sequence information in genomic context for a specific species, and to control the data viewed in the browser by selecting specific types of data from available menus.

The browser was created by the Genome Bioinformatics Group of the University of California at Santa Cruz. It incorporates data from many reference sequences such as NCBI RefSeq, Ensembl, data from the ENCODE project, and many more sources.

This guide is an introduction to the UCSC Genome browser: how to locate it, manage tracks and annotations and how to get to sequence data from the browser view. It also includes a brief introduction to the table browser, where users can obtain more detailed genomic information in table format that cannot be obtained from the browser view.

The UCSC Genome Browser Entry Page

You can access the UCSC Genome Browser homepage at:

http://genome.ucsc.edu
Access to the genome browser is through the Genomes links in the top menu bar, or through the Genome Browser link at the top of Our tools. This will take you directly to the Human Genome Browser Gateway, which is the default gateway, since the human genome is the most popularly searched at the UCSC Genome Browser.

To change the species, select from the icons on the left, or search using the search box. You can also change the assembly. By default, the current assembly for any given genome is selected. If a new build has just been created for a particular species, not all annotations may be viewable, but they are added over time.

You can search by known genomic position, or search a whole chromosome, or use the search term box to enter a gene symbol, accession number, or any text word to search. The browser gateway has some examples of types of search terms to use on the entry page.

Note: The UCSC Genome Browser saves your session settings as cookies in your web browser's settings. This is a nice feature, because you can move seamlessly between the genome browser and the table browser without having to re-enter your search terms. However, this can be confusing if you want to start a new search for a new region. To do that, click the text link: Reset All User Settings under the Genome Browser menu item on the top menu in the page. This will remove your earlier search terms and let you start a new search.

This search will use the example gene ACE - angiotensin converting enzyme.

Type your search term into the search term box. As you do, you will get suggested results that pop up as you are typing, much like in a Google search. You can select from these suggestions, or ignore them and keep typing your search terms. Click the GO button.

Search Results and the Browser View

Depending on your search terms, you may have to select a specific transcript product from an intermediate page of results grouped by gene source (UCSC, RefSeq, etc). If you have searched for a unique enough term or have used the suggestions from the search term box, you will be taken directly to the browser view.

At the top of the browser view are a number of choices you can use to change your display.
You can

- move up and downstream
- choose from zoom options, including the choice to zoom immediately to the sequence view by selecting the base button.
- add coordinates to the view in the text box to widen your view
- view the location of your view by the red marker on the chromosome graphic

Below the chromosome graphic is the browser map with default tracks displayed. The top section is general gene information, scale, chromosomal coordinates, genes in graphic view.

- Small arrows show the direction the gene is transcribed on the strand
- You will notice that your selected gene's name may be highlighted in black
- Color key for GENCODE Comprehensive Transcript Set (this section used to be called "UCSC Genes")
  - Black - feature has a corresponding protein entry in PDB
  - Dark blue - the transcript has been reviewed or validated by either RefSeq, SwissProt or CCDS
  - Medium blue - other known RefSeq transcripts
  - Light blue - non-RefSeq transcripts
- Don't be concerned if you don't see multiple color options for the transcripts: they are updated and annotated regularly, so some transcripts may change color as they are verified and others may be removed.
- Click the vertical bars on the left of any displayed track to see all the information available about that track's display features and legend
- Click on a transcript line to open the gene’s complete information page

**Detail Views**

When you click any element on the browser's main window, you will open a page with a list of details about that element. Clicking on the black transcript in the GENCODE Transcript track for ACE opens a page that has numerous links and pieces of information about the gene and its protein products. At the top of the screen is a description of the gene and its coding region, followed by a Page Index table.
This index has quick links to portions of the page containing details of the gene. Not all genes will have a large, detailed page. Some well-studied genes, such as p53, have huge pages of links and details. Other less well-known genes will have very few links in the page index, and little detail on the page. Where available, details of microarray expression in various tissues, protein and RNA structure and pathway information is displayed.

Below the Page Index is another table of external links or links to sequence details.

<table>
<thead>
<tr>
<th>Genomic Sequence (chr17:61,554,422-61,575,741)</th>
<th>mRNA (may differ from genome)</th>
<th>Protein (1306 aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Sortor</td>
<td>Genome Browser</td>
<td>Protein FASTA</td>
</tr>
<tr>
<td>Ensembl</td>
<td>Entrez Gene</td>
<td>ExonPrimer</td>
</tr>
<tr>
<td>H-INV</td>
<td>HGNC</td>
<td>HPRD</td>
</tr>
<tr>
<td>PubMed</td>
<td>Stanford SOURCE</td>
<td>Treefarm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UniProtKB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wikipedia</td>
</tr>
</tbody>
</table>

The links in green on this table will take you to the corresponding data within the UCSC Genome Browser pages. The links in blue boxes will open a new window or tab in your web browser with information particular to your gene at the source. Especially useful are the links to OMIM, GeneCards, and links to ExonPrimer and other sources where you can determine or obtain reagents for investigation of your gene. We will discuss the sequence view later in this guide.

---

**Getting a Sequence-Fast!**

It's very easy to get the DNA sequence when you are viewing a region in the UCSC Genome browser. There are multiple ways to do it.

- **The fastest way to get the DNA sequence for a specific gene is to right-click** on the graphic gene view in the GENCODE Comprehensive Transcript section and select **Get DNA for [gene name]**

- To get DNA for the whole region you're viewing, go to the top of the page, and use the **View** menu to get DNA for the entire region shown in the browser (not just the gene--unless the gene you're viewing takes up the whole view window)
A new page will open, on this page you can add bases up- or downstream, change masking options and use the button to custom color your DNA sequence.

In this example, I'll color all RefSeq gene exons blue, and underline and color all SNPs red and click submit.
So my sequence is colored to show exons and SNPs:

```
TGTCACCTCGCCCTGATGCTCTCTTGGCCCGAGCAGTCTGTCTGCTGCTG
GCTGTAGCTATAATTAGAAATGTAGAGTGCAACCACAAGAGCAAATTTCCAGCTAGCTG
CAAGACCGCTCGGCTGCCGCCCGCTGAGAAGACCCGACTTACCTGAGGCTGCTCC
AGCCCTGGCCTGACGCTATACCTGACCACGCGCCAGACGCTGGCTCTACCTCG
CTACTCGCCCTGCTGCTACACCGGGCGAGAGATCCTGAGCAACACTCTTACC
AACAGCTAGAGGCCCTCTACGTGAACTTACGCTGTGCCGCGCAGTACCTGAGCC
GATACGGAGACAGATACATCAACTGAGGAGCCATCCCTGGCTCTACCTCGCCCG
```

To get genomic sequence, or mRNA or protein sequences for the region you're viewing in FASTA format, go back to the GENCODE Comprehensive Transcript Set portion of the viewer and click anywhere on a gene transcript to open the detailed Description and Page Index page.

Now go to the section marked Sequence and Links to Tools and Databases

Click on Genomic Sequence to get the DNA sequence
You will be taken to a page called **Get Genomic Sequence Near Gene**. This is very much like the Get DNA page described above, but it has more sequence extension and formatting options, but no color options.

**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 [1000] bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by [1000] bases
- One FASTA record per feature.
- 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]

This output is a FASTA file that you can copy or save for use in other applications.

If you choose mRNA or Protein sequences, you will merely get the FASTA formatted sequence, with no additional options for adding bases or masking.

---

**Viewing and Editing Tracks**

You can control the density of the tracks displayed on your browser map in two ways:

- Right-click on a track to open a pop-up menu that will allow you to change the density of the display, configure the track, or (in some cases) take you to the detail view for that track.
Scroll down the page to the menu areas that allow you to turn on, off or change the density of the tracks.

Select your tracks to add choosing your density of display, and hit the refresh button...

and your tracks will show up in the browser view.

Some notes about track density:

- Sometimes you only have the option of **show** or **hide**. When this is the case, first show the tracks, then right-click on tracks in the browser itself to extend your density view options.
- Depending on which genome build you choose, not all options seen on this guide will be available. As more annotations are added, more options will show up.
- Here are the differences between dense, squish, pack and full:
If you have added or edited tracks past the point where the browser is easy to view, and you wish to return to the default tracks and start again, or even hide all of the tracks and only show a few of interest to you, there are buttons for this located between the browser map and the display menus.
Search Genomes with BLAT

You can use sequences that you retrieve from the browser or nucleotide sequences that you have obtained from other sources to run a BLAT on the UCSC Genome Browser site. BLAT is not quite the same as a genomic BLAST. It works by keeping an index of the target genome in memory, rather than the entire genome sequence itself.

To try a BLAT from the ACE gene view we’ve been working with:

- Click on the transcript graphic in the GENCODE Comprehensive Transcript Set part of the browser view to open the Description and Page Index page.

- Now click on either the Genomic Sequence or the mRNA sequence links in the Sequence menu.

Note: If you click on Genomic Sequence, you can add bases up- or downstream, but be sure to keep your sequence total length under 25,000 bases. BLAT cannot align anything larger than that. This is why it is sometimes easier to align the mRNA sequence instead of the genomic sequence.

- Once you have a view of the FASTA formatted sequence(s) in your web browser window, use a "select all" command to highlight the FASTA sequence, then copy it.
- Now click on the Tools menu in the UCSC Genome Browser menu bar and choose Blat (first option in the Tools menu)
- Paste your sequence in the text box, and select the species genome, assembly, query type (DNA, RNA, protein, or let BLAT guess it for you); select your sort output and output type. The hyperlink output type is most useful for viewing results.
- Click submit
  - Note: if you click "I'm feeling lucky", like Google, you will only get the best result, viewed in the Genome Browser graphic view. Though cute, this is not always very helpful.
  - You can also use the File Upload field to upload your own sequence file, in plain text format.
If you have used sequence from a known gene or known genomic region, you will likely have one very high-scoring match, followed by others of lesser scores.

Note that you have matches on both forward and reverse strands of the genome in the region defined.

Click on the browser link for any match to see a graphic view of the BLAT alignment in the Genome Browser.

Click on the details link to see a series of "blocks" of matching sequence, followed by side-by-side alignments of the query sequence (your input) and the target hit from the genome.

---

**Table Browser**

The UCSC Genome Browser is built on a series of SQL tables. However, the details in all of these tables cannot be fully viewed in the graphical browser view. For that reason, UCSC gives users a way to search or browse the tables to extract data that cannot be gathered from the browser view.

To get to tables directly related to your area of interest, from the Genome Browser window, go to the top menu bar, select **Tools -> Table Browser**. To access the table browser without starting from a genomic region, you can use the Table Browser links from the UCSC Genome Browser home page left menu.

The Table Browser set-up page has many options:

- Set your **species** of choice and **assembly**
- Choose your targets from the group menu: default is **Genes & Gene Predictions**, but you can access table data on **variation**, **regulation**, **disease** annotation and much more
- Use the **region** you have been viewing (indicated by the position radio button and text box) or chose the entire
 genome (not recommended, due to massive amounts of data from whole genome tables)
    - Use the lookup button to look up an area by a gene name
    - Use the define regions button to paste or upload a file of genomic coordinates
- Paste or upload a list of identifiers (works only for specific tracks, like RefSeq accession numbers for RefSeq gene tracks)
- Filter by inputting specific gene symbols to search under, or by entering specific sequence for short repeats, or by many other filter options. The filter create button will open a new window where you can set your filter preferences
- Create an intersection between two tables within the genome browser’s database
- Output files as simple text, hyperlinks, BED files, FASTA alignments, for Galaxy analysis suite and more
- Use the summary/statistics button to get a preview of how much data will be returned in your tables before you actually request the output

The Table Browser can be particularly useful for finding transcription start and end sites, which are not always the same as the CDS regions, plus exact coordinates for each exon start and end site for a particular gene. These features are not available in the browser view.

More Information and Links

This guide is only an introduction to the features and functionality of the UCSC Genome Browser. You can also create custom tracks and upload them to the browser, view other users’ custom tracks and more.

For some other useful guides and materials on the UCSC Genome Browser, check out these links:

- The UCSC Genome Browser Training pages
- Contact the Biosciences & Bioinformatics Librarian at Galter Health Sciences Library for individual assistance

Printed: Friday, March 8, 2019 1:32 AM