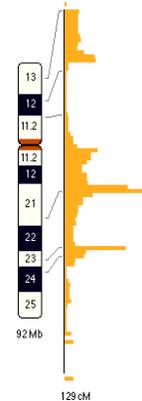




Guide to Exploring the Human Genome

**Kenneth R. Miller
Brown University**

(last update: June 7, 2006)



The DNA sequences of the human genome are now accessible by more than a hundred databases and searching tools on the World Wide Web. In this brief guide, however, I'll focus on just one (relatively) easy-to-use tool, the UC Santa Cruz (UCSC) Genome Browser:

Primary Resource:

The UCSC Genome Browser:

<http://genome.ucsc.edu/cgi-bin/hgGateway>

The UCSC Browser can be used in a number of ways. In this guide, I'll focus on just one — using the Browser to locate and investigate genes of interest. As an example, let's suppose that you've been asked to locate and investigate a gene called **RS11**. First, open a web browser and type in the UCSC Genome Browser address:

<http://genome.ucsc.edu/cgi-bin/hgGateway>

A window will open, the top part of which looks something like this:

Since we're looking for **RS11**, we'll type "RS11" into the "position" window (as shown above) and hit the **Submit** button. After a few seconds the Browser will have found sequences that match the string you've entered, and a new window will appear:

Known Genes

[RS11_HUMAN at chr19:54691499-54694754](#) - (RS11_HUMAN) 40S ribosomal protein S11.

It's usually best to take the appropriate link under the first heading that appears. That may be either the "Known Genes" heading, or "RefSeq Genes," as shown below for the "Clock" gene:

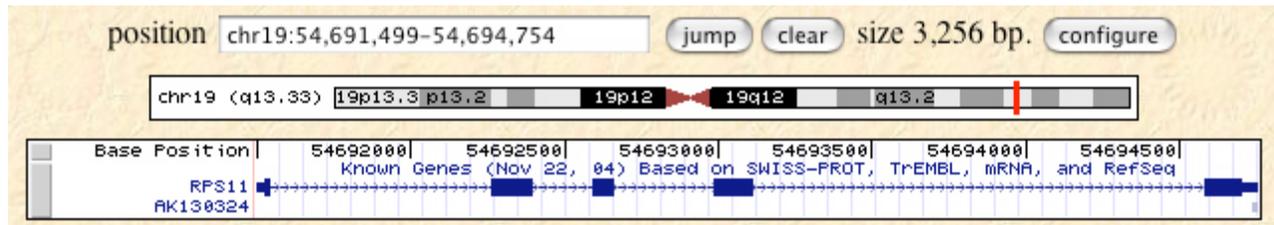
Known Genes

[CLOCK at chr4:56214201-56291736](#) - (AB002332) clock homolog (mouse)

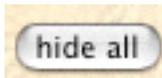
RefSeq Genes

[CLOCK at chr4:56214201-56328539](#) - (NM_004898) clock

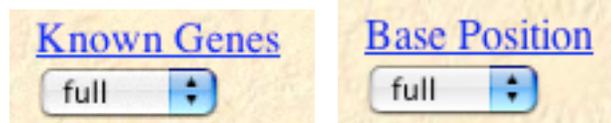
Click on either the "Known" or "RefSeq" link, and a window will open showing the location of the gene you've chosen. Here's the window for RS11:



Don't be surprised if your window looks a lot more complex than the one shown above. The Browser allows you to display so much information that it can be truly confusing. So, try the following. First, hit the "Hide All" button:



This will close all of the data display options. And now, open just two of them that you're going to use: The "Known Genes" button and the "Base Position" option:



Hit the "Refresh" button, and your window should now look just like the one above, displaying a manageable amount of information. As you can see, there's a map of our gene (actually called RPS11) showing its exact position (on chromosome 19), and several little blocks showing that it consists of 5 exons and is transcribed from left to right. The window also allows you to move to the left or the right along the chromosome, and to zoom in or out to look at the "neighborhood" around the RPS11 gene.

For detailed information, click on RPS11 and a new window will open with a brief, but informative description of the gene and (in this case) its protein product:

Human Gene RPS11 Description and Page Index

Description: 40S ribosomal protein S11.
Alternate Gene Symbols: NM_001015
Representative mRNA: [BC007283](#) **Protein:** [P62280](#) (aka RS11_HUMAN)
RefSeq Summary: Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S17P family of ribosomal proteins. It is located in the cytoplasm. The gene product of the E. coli ortholog (ribosomal protein S17) is thought to be involved in the recognition of termination codons. This gene is co-transcribed with a small nucleolar RNA gene, which is located in its third intron. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome.

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

As you can see, the description on this page gives us the basic facts about the protein, which turns out to be one of the proteins in the small ribosome subunit. To find even more about RPS11, click on any of the other links on the page.

The "Sequence" links give the DNA sequence of the gene, a comparison of the human version of the gene with its corresponding gene in the mouse ("Comparative"), the mRNA sequence, and the Amino Acid sequence of the protein.

To get detailed information about the gene, click on one of the "Tools and Databases" links:

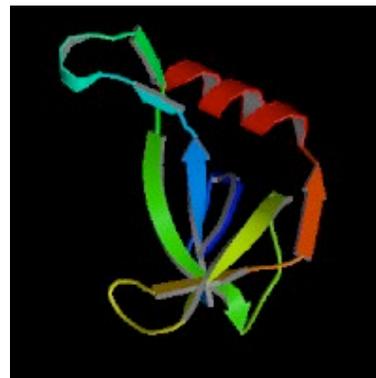
Quick Links to Tools and Databases					
Genome Browser	Gene Sorter	Proteome Browser	UniProt	Entrez Gene	PubMed
OMIM	GeneLynx	GeneCards	CGAP	Stanford SOURCE	ExonPrimer
Ensembl	Jackson Labs	H-INV			

The "OMIM" (On-Line Mendelian Inheritance in Man) or the Stanford "SOURCE" links are especially useful, and will link you to a wealth of information about the gene. Here's what the OMIM window for our RS11 search:

The screenshot shows the OMIM website interface. At the top, there's a search bar with "OMIM" selected and "180471" entered. Below the search bar, there are navigation tabs for "Nucleotide", "Protein", "Genome", "Structure", "PMC", and "Taxonomy". The search results show "180471" and "RIBOSOMAL PROTEIN S11; RPS11". A "TABLE OF CONTENTS" section lists links for "TEXT", "REFERENCES", "CONTRIBUTORS", "CREATION DATE", and "EDIT HISTORY". There is also a "Gene map locus 19q" link and a "TEXT" section with a reference to "180466 for general information about ribosomal proteins."

Either of these will tell you the scientific background of the gene, list any genetic disorders associated with the gene, and refer you directly to scientific publications on the gene or its protein product.

The "Structure" link will even show you the three-dimensional structure of the protein product (if known):



Happy Gene Hunting!