

## GUIDE TO EXPLORING THE HUMAN GENOME

As scientists around the world sequenced sections of the human genome, they all uploaded their DNA sequences to a centrally located database. These data have always been publicly available through the Web so that the worldwide scientific community can benefit from this information. This process has continued with the genomes of other organisms that are being sequenced (mouse, rat, chimpanzee, corn, etc.)

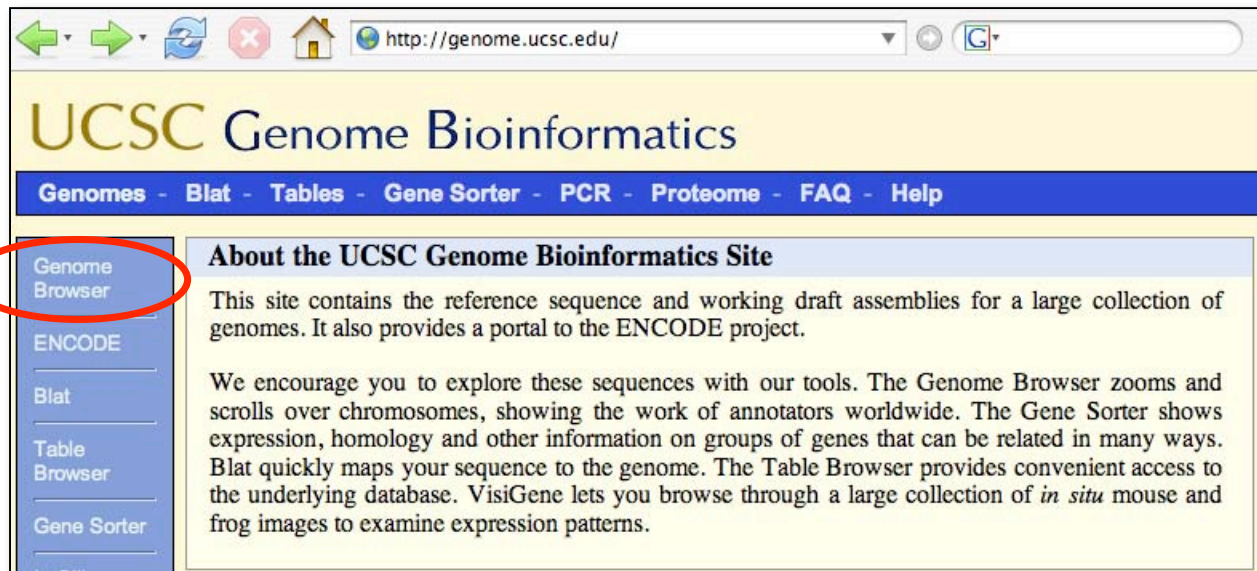
But in its raw form, this information is merely a very large collection of four letters (A, T, C, G) and is difficult to make sense of unless you have a tool to organize it. Many search tools have now been developed to allow you to interpret the A-T-C-G sequences. In this guide, we will focus on just one (relatively) easy-to-use tool — the **University of California at Santa Cruz (UCSC) Genome Browser**:

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

The UCSC Browser can be used in many ways. In this guide, we'll focus on just one of its functions — using the Browser to locate and investigate human genes of interest.

Let's walk through an example search to see how to use the UCSC Genome Browser.

1. Let's suppose that you've been asked to locate and investigate a gene called **RPS11**.
2. First, open a Web browser and type in the UCSC Genome Browser address. Click on the top link in the side navigation bar: **Genome Browser**.



3. This will bring you to a page which allows you to search the human genome (as well as the genomes of other organisms). The top part of that page contains a search form. You can search for a gene by entering your search terms into the "position" form field and then clicking on the Submit button. In our case, we will type in **RPS11**:

**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 RPS11 620 submit

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

add your own custom tracks configure tracks and display clear position

- After a very brief moment of hunting through all 3 billion base pairs of the human genome, the Browser will have found sequences that match the search term you have entered, and a new window will appear with your search results:

**Known Genes**

[RPS11 \(NM\\_001015\) at chr19:54691445-54694756](#) - ribosomal protein S11

**RefSeq Genes**

[RPS11 at chr19:54691445-54694756](#) - (NM\_001015) ribosomal protein S11

- It is best to scan through the links in the first two headings — “Known Genes” and “RefSeq Genes”. The “Known Genes” list is appropriately named as it is a list of genes that are recognized as being related to your search term. The “ReSeq Genes” is the sequence that has been selected by Genome researchers as the “official” version of the gene. Choose the appropriate link in either of these two lists. Either one will open up a new window:

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

position/search chr19:54,691,445-54,694,756 jump clear size 3,312 bp. configure

chr19 (q13.33) 19p13.3 19p13.2 19p12 19q12 19q13.2

|   |          |          |          |          |          |          |
|---|----------|----------|----------|----------|----------|----------|
| chr19:  | 54692000 | 54692500 | 54693000 | 54693500 | 54694000 | 54694500 |
| STS Markers   |          |          |          |          |          |          |
| UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA |          |          |          |          |          |          |
| RefSeq Genes  |          |          |          |          |          |          |
| Human mRNAs   |          |          |          |          |          |          |
| Spliced ESTs  |          |          |          |          |          |          |
| Conservation  |          |          |          |          |          |          |
| RepeatMasker  |          |          |          |          |          |          |

move start < 2.0 > move end < 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.

default tracks hide all custom tracks configure refresh

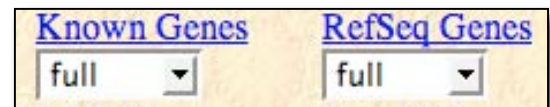
- This new window includes a number of sections. Don’t be surprised if the chart looks a lot more complicated than the one shown above. The Genome Browser is a very powerful genome analysis tool and there is a wealth of information at your fingertips which may be

both exciting and daunting. However the Genome Browser allows you to configure the display to add or remove any information you care to, so you can simplify it in a number of ways. Try this:

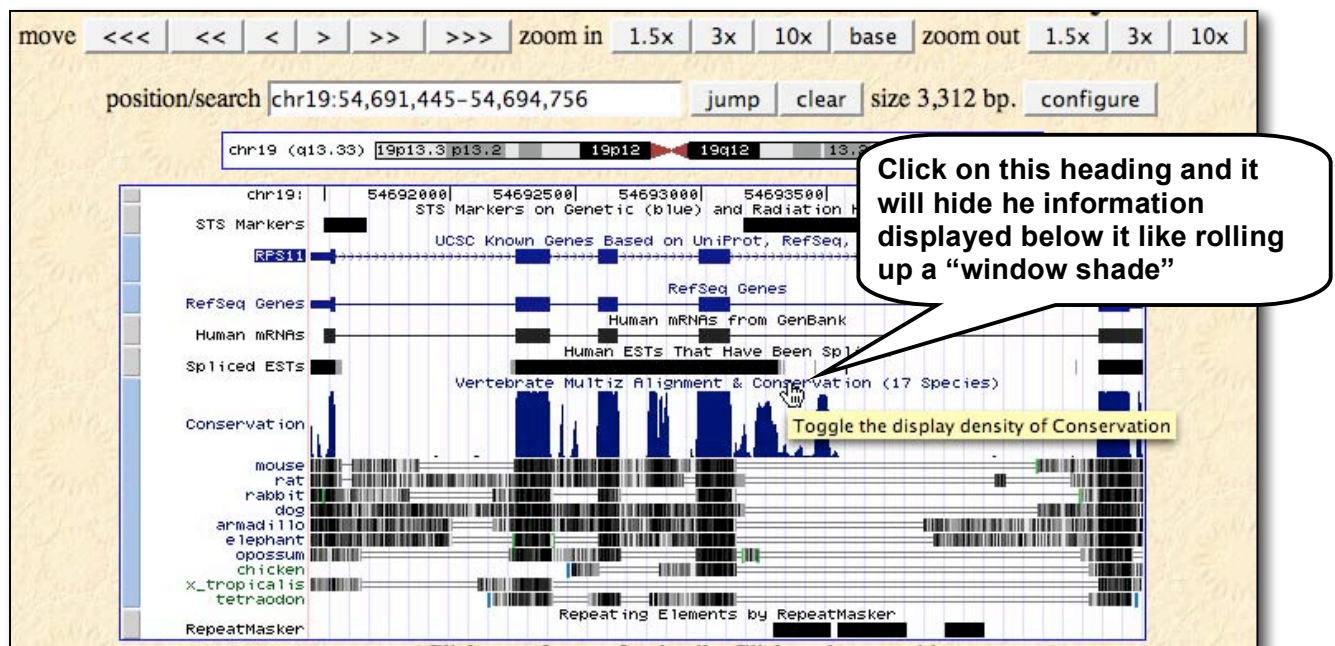
- a. Click on the “Hide All” button. This will close all of the data display options at once.



- b. You now want to open two data display sections on the chart: “Known Genes” and “Base Position”. Scroll down a little below the chart and make these selections in the drop down menus in the section entitled, “Genes and Gene Prediction Tracks”. Then click on the “Refresh” button and your window will look simpler.



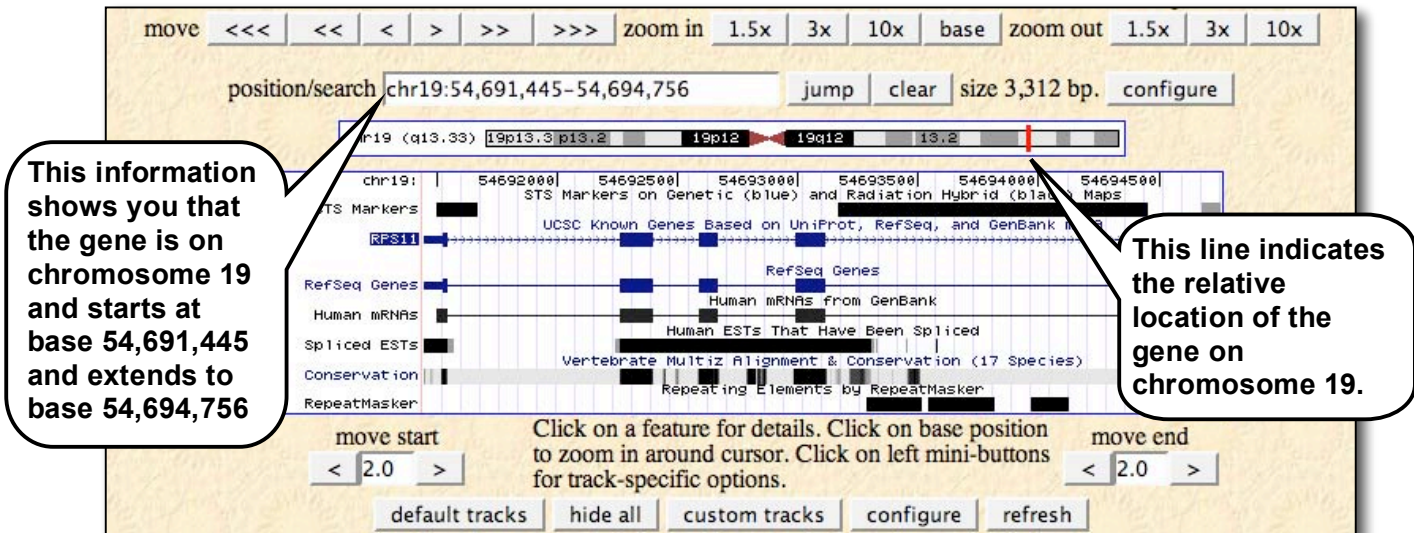
- c. If you want to make the information on the chart even more manageable, you can “window shade” sections in the chart to shrink down how much information is displayed. Click on any heading within the chart and it will hide that information displayed below the heading:



- d. Once you have done this your chart should look like our original “simplified” version above (in Step #5.)

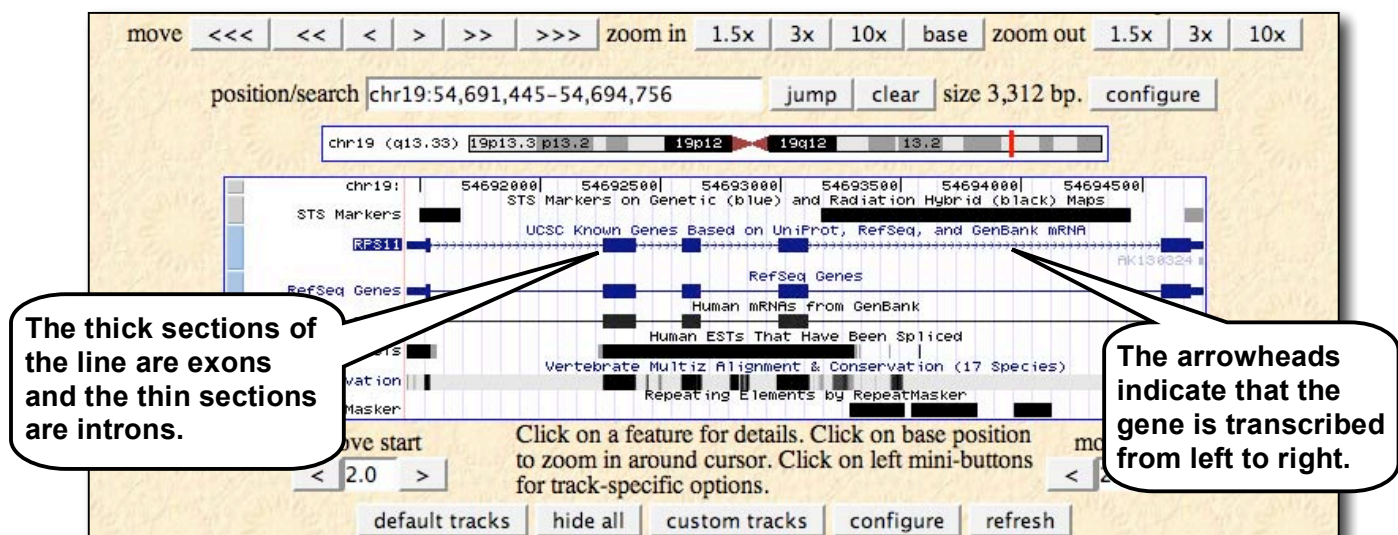
7. As you can see, the chart displays a map of the gene — **RPS11** — showing its exact position (on chromosome 19). It also displays a lot of information in a characteristic fashion and once you learn how to interpret the Genome Browser interface, you can decipher a lot about any genes you look up here.

- a. Above the chart is a horizontal drawing of a chromosome. This indicates to you — with a red line — the location of the gene on the chromosome. Also the information now written in the “position/search” form field is the exact location of the gene on the chromosome (written in number of bases).



So we now know our RPS11 gene is located towards the end of chromosome 19 and it starts at base 54,691,445 and extends to 54,694,756. So the gene is 3,312 bases

- b. At the very top of the window are “Move” and “Zoom” tools. These allow you to explore the chromosome further: The “Move” arrows allow you to slide to the left or the right along the chromosome and the “Zoom” buttons allow you to zoom in or out to look at the DNA “neighborhood” around the RPS11 gene.
- c. And there is more information about the gene coded in this chart. The arrowheads along the line representing the gene tell you that the DNA is transcribed from left to right when this gene is read. The thick parts of the line are exons and the skinny parts (with the arrowheads) are introns. So, RPS11 has 5 exons and 4 introns.



8. For detailed information, click on the label RPS11 on the left side of the chart. A new window will open with a brief, but informative description of the gene and its protein product.

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

position/search chr19:54,691,445-54,694,756 jump clear size 3,312 bp. configure

chr19 (q13.33) 19p13.3 19p13.2 19p12 19q12 19q13

chr19: 54692000 | 54692500 | 54693000 | 54693500 | 54694000 | 54694500

STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps

UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Vertebrate Multiz Alignment & Conservation (17 Species)

Repeating Elements by RepeatMasker

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

alt tracks hide all custom tracks configure refresh

move end < 2.0 >

**Click on the name of the gene and it will bring you to more detailed information about its known functions.**

9. As you can see below, the description on this page holds the basic facts about the protein, which turns out to be a protein in the small ribosome subunit.

**Human Gene RPS11 Description and Page Index**

**Description:** ribosomal protein S11  
**Alternate Gene Symbols:** BC007283, BC007603, BC007945, BC010028, BC016378, BC018829, BC070224, BC100025, X06617  
**Representative Refseq:** [NM\\_001015](#) **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.  
**Position:** chr19:54691446-54694756  
**Genomic Size:** 3311

|                                   |                               |                                  |                                   |                            |                               |
|-----------------------------------|-------------------------------|----------------------------------|-----------------------------------|----------------------------|-------------------------------|
| <a href="#">Page Index</a>        | <a href="#">Quick Links</a>   | <a href="#">UniProt Comments</a> | <a href="#">Sequence</a>          | <a href="#">Microarray</a> | <a href="#">RNA Structure</a> |
| <a href="#">Protein Structure</a> | <a href="#">Other Species</a> | <a href="#">GO Annotations</a>   | <a href="#">mRNA Descriptions</a> | <a href="#">Pathways</a>   | <a href="#">Methods</a>       |

10. To find even more about RPS11, click on the other links in this section — these “**Page Index**” links are “quick links” to lower sections on this same Web page.
  - a. The “**Sequence**” link brings you to the DNA sequence of the gene, the mRNA sequence of the transcript, and the amino acid sequence of the protein product. It is interesting to click on each of these links to see how the information changes from gene to mRNA transcript to protein product.

**Sequence**

[Genomic \(chr19:54,691,446-54,694,756\)](#)
[mRNA \(may differ from genome\)](#)
[Protein \(158 aa\)](#)

b. Click on the “**Genomic**” Sequence link and this will open up a new page that allows you to customize how you want this gene sequence presented to you. Try different configurations:

- The 5’UTR and 3’UTR Exons are the Untranslated Regions at the beginning and end of a gene — they are transcribed to mRNA but are not translated into protein because they lie before the “Start” codon and after the “Stop” codon.
- The CDS Exons are the actual coding sequences of DNA.
- Most people find it easier to interpret a gene sequence if introns and exons are distinguished by using lowercase vs. capital letters.

**Genomic Sequence Near Gene**

**Get Genomic Sequence Near Gene**

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

**Sequence Retrieval Region Options:**

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

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

**Sequence Formatting Options:**

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

Try different settings to see how it changes the display of the gene coding sequence.

These settings especially help to distinguish different parts of the sequence. Try different options.

- After you click “Submit”, a window will open with your gene sequence. You can go back and forth to the previous Web page and reconfigure the display in different ways to see how the coding sequence changes if you only view exons, or remove the UTRs, or if you include the promoter region, etc.

```
>hg18_knownGene_NM_001015_range=chr19:54691446-54691446
ACTTCCTCCTCCAGCCCTCGGTAATCGATAAGGAACCCGGACGCTGCTG
CCCCTTTCTTTTTTCAGCGCGCCGGGAAGATGGCGGACATTCAGgtgcg
gactcgggggtggatgccagggtgcggggtccgcttggccttcaggggc
ggtccgggagggcagagcccggcgccaaagtccgggggttaattccggg
cgtccgtgattatcttagagattaactggagtgaggggcgcttgcctt
tgcttctagatgagtgcttccctaatccctgggggctccctggcctaatgg
aggctcaagcgttttaggacttaccttcttaaaaccaagttaagaagg
ttttttggagttctccggtcttcagggttaattctgtggaacaggactc
ctaagggttaatccaggagcctgcccgtctggtatttaattctgaatcct
tatgtactacataaggagtgaatcctgggtggcgatcttgggtccttggg
gaatgatcctcgctgaaagggtgctccagggtgtgagagggttgatcctggt
ttagaggatgaattttgcaatccctgagaggccttctggagtataaaat
tgaataaactcgggcttagatctgagatgaattacaggtaaggctcagg
gtggtcagtctttagagggtgattctggggcaccaggccttggacggcg
cgctttgcaagtaaatgtagagcatggggctctgggaggttctgggaagg
ctctagggtgggtggctgcccgttcaacttagcagctcatcagtttt
ctcctcataatctgtagACTGAGCGTGCCTACCAAAAGCAGCCGACCATC
TTTCAAAACAAAGAGGGTCTCTGCTGGGAGAACTGGCAAGGAGAAGCT
CCCCGGTACTACAAGAACATCGGCTCTGGGCTTCAAGACACCAAGGAGG
tgcggggaacctcagaagaagaagggaacctggcgttccctgcaactgt
gcccacgacgagttgccctgctcgcattcaagtggcttctggggctgctg
ggaattgtagttgcttccctgaggccacgcccctggctcttttaaggaa
cggccgcccgaaggctcaactcctttatcttctctatccttccagGCTATTG
AGGGCACCTACATTGACAAGAAATGCCCTTCACTGGTAATGTGTCCATT
CGAGGGCGGATCCTCTCTGtaagtggggagttactggtgctctggggcc
tgaataactgaaagaagggtcttggggccagactcctgggtcctgggtg
agaggggtggttaggaatgcaaacctgtagggtccaggtacattggcagat
gatgtttgtttcagatggtcttcagatgcccacgtgggcaactgctgag
aaagccacttggtaaaactgatgcccgaatggggctttttgggatccct
gctcagctgcttctgagtcaccagcatgcccctgggttacctatggccctct
ttcccattgggacctgacctatgatcggcccccgctcctagCGTGGTGAC
CAAGATGAAGATGCAGAGGACCATTGTCATCCGCCGAGACTATCTGCACT
ACATCCCGCAAGTACAACCCCTTCGAGAAGCGCCACAAGACATGTCTGTGA
CACCTGTCCCTGCTTCAGgtgagcgcagtgcccatacaggttgcctcag
gcccagctctctcagccttcagattccagatcggaccaatttaaggccaa
ctgaggggagggaaagactgaggtggcattctctgggaagcccagaatctc
aggatttgtatatacatatgttcttcagaatcaaaagcgttctacaggggtga
atatctctacctgaaatgcttggaaaccagattttggaatgttctcgcata
catagttagatgttttggggatgtgacataggtctaaacaagaagctcat
ttgtttcataatacaccttatccacatagcctgaaagtgtcttaccagcagt
ggttttataaattctgtgtgaaacaaggtttgcgctacattgaaccattgg
aaagcaaaagggtgacagctggtattgaaagaagtttcagatcttagccagg
caccgtggtcgaagccagtaatcccagcactgtaggaggctgagggcaggga
ggatcccttgagcccaggagtttgagaccagcctgggcaacatggcaaaa
```

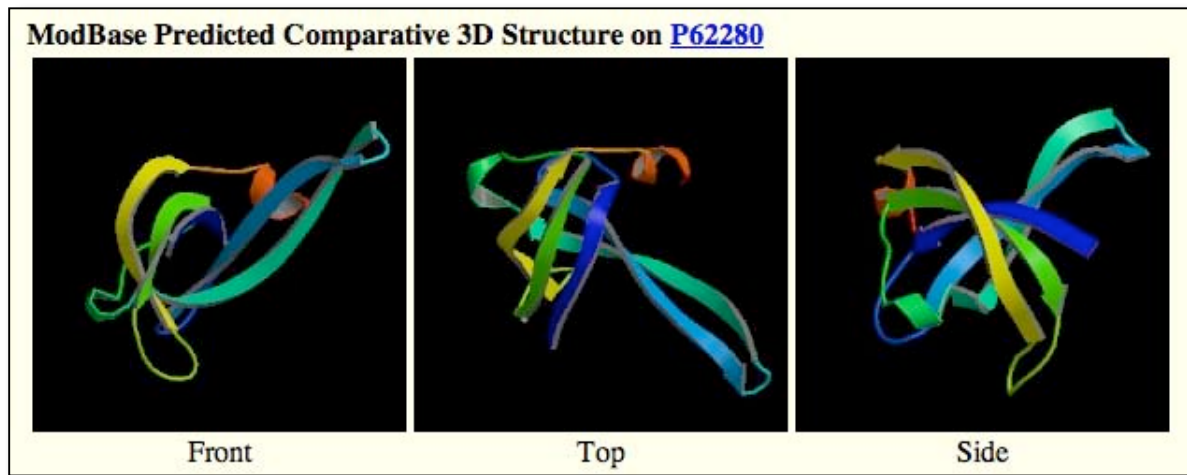
- c. Go back to the “Sequence” links and click on the “mRNA” Sequence link to view the sequence of the mRNA transcript that is produced after splicing out introns. Oddly, this sequence is expressed in DNA nucleotides, so you will see thymine in the sequence instead of uracil. Don’t be confused; it’s a matter of scientific convention: The mRNA sequences were completed by sequencing cDNA copies made from the extracted mRNA.

```
>NM_001015 (RPS11)
acttctctccagccctgcgtaatcgataaggaacccggagcgtgctgccccttctt
ttttcagggcgccggggaagatggcggacatcagactgagcgtgctaccaaaagcagc
cgaccatcttcaaaaacaagaagaggggtcctgctgggagaaactggcaaggagaagctcc
cgcggtactacaagaacatcggtctgggcttcaagacacccaaggaggctattgagggca
cctacattgacaagaaatgcccctcactggtaattgtgctcattcagggcggtacctct
ctggcgtggtgaccaagatgaagatgcagagggaccattgtcatccgccgagactatctgc
actacatccgcaagtacaaccgcttcgagaagcggccacaagaacatgtctgtacacctgt
ccccctgctcagggacgtccagatcggtgacatcgtcacagtggcgagtgccggcctc
tgagcaagacagtgcgcttcaacgtgctcaaggtcacaagcctcggcgccaccaagaagc
agttccagaagttctgaggctggacatcggcccgtccccaccaatgaaataaagtattt
tctcattccaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
```

- d. Now click on the “Protein” Sequence link to view the sequence of amino acids that construct the protein product from this gene. This sequence is expressed in the single letter code sometimes used to symbolize amino acids. A dictionary for these codes is available on our “Human Genome Scavenger Hunt” Web site.

```
>NM_001015 (RPS11)
MADIQTERAYQKQPTIFQNKRRVLLGETGKEKLPYYKNIQLGFKTPKEAIEGTYIDKKC
PFTGNVSIIRGRLLSGVVTMKMQRTIVIRRDYLYHIRKYNRFEKRHKHNMVHLSPCFRDV
QIGDIVTVGECRPLSKTVRFNVLVKTKAAGTKKQFQKF
```

11. Go back to the “Page Index” links and click on the “**Protein Structure**” link; this shows you the three-dimensional structure of the protein product (if known).



12. You can also get further detailed information about the gene by clicking on one of the “**Tools and Databases**” links. For example, 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. 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.

| Quick Links to Tools and Databases |                                 |                                  |                              |                         |                             |
|------------------------------------|---------------------------------|----------------------------------|------------------------------|-------------------------|-----------------------------|
| <a href="#">Genome Browser</a>     | <a href="#">Gene Sorter</a>     | <a href="#">Proteome Browser</a> | <a href="#">Table Schema</a> | <a href="#">UniProt</a> | <a href="#">Entrez Gene</a> |
| <a href="#">PubMed</a>             | <a href="#">OMIM</a>            | <a href="#">GeneLynx</a>         | <a href="#">GeneCards</a>    | <a href="#">HGNC</a>    | <a href="#">CGAP</a>        |
| <a href="#">Reactome</a>           | <a href="#">Stanford SOURCE</a> | <a href="#">Jackson Labs</a>     | <a href="#">H-INV</a>        |                         |                             |

Here’s the **OMIM** page returned from our RPS11 search:

NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC

Search OMIM for 180471 Go

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

All: 2  
Items 1 - 2 of 2

1: [\\*180471](#)  
**RIBOSOMAL PROTEIN S11; RPS11**

**TABLE OF CONTENTS**

- [TEXT](#)
- [REFERENCES](#)
- [CONTRIBUTORS](#)
- [CREATION DATE](#)
- [EDIT HISTORY](#)

Entrez  
OMIM  
Search OMIM  
Search Gene Map  
Search Morbid Map

Help  
OMIM Help  
How to Link

FAQ  
Numbering System  
Symbols  
How to Print  
Citing OMIM  
Download

OMIM Facts  
Statistics

**Happy Gene Hunting!**