

1. I opened a window at NCBI and selected "gene" from the pull down menu and searched for PSRC1.

The screenshot shows the NCBI homepage. At the top, there are navigation links for "Resources" and "How To". The main search bar has "Gene" selected in the dropdown menu, and "PSRC1" is entered in the search field. A "Search" button is visible. Below the search bar, there are sections for "Resources" (with links to NCBI Home, All Resources (A-Z), Chemicals & Bioassays, Data & Software, and DNA & RNA), "Welcome to NCBI" (with a description of the center's mission and a link to "More about the NCBI"), and "Popular Resources" (with links to BLAST, Bookshelf, Gene, Genome, and Nucleotide).

2. I chose the link for the human PSRC1 gene.

The screenshot shows the Entrez Gene search results for "PSRC1". The search bar at the top has "Gene" selected in the dropdown menu, and "PSRC1" is entered in the search field. The results are displayed in a list format. The first result is for the human gene, "PSRC1", which is circled in red. The details for this result include: "Official Symbol PSRC1 and Name: proline/serine-rich coiled-coil 1 [Homo sapiens]", "Other Aliases: RP11-297O4.2, DDA3, FP3214, MGC1780", "Other Designations: OTTHUMP00000013789; OTTHUMP00000013790; OTTHUMP000000205834; differential display and activated by p53; p53-regulated DDA3; proline/serine-rich coiled-coil protein 1", "Chromosome: 1; Location: 1p13.3", "Annotation: Chromosome 1, NC\_000001.10 (109822178..109825771, complement)", "MIM: 613126", "ID: 84722", and a link to "Order cDNA clone". The second result is for the rat gene, "Psrc1", with details: "Official Symbol Psrc1 and Name: proline/serine-rich coiled-coil 1 [Rattus norvegicus]".

The screenshot shows the full report for the human gene "PSRC1". The search bar at the top has "Gene" selected in the dropdown menu, and "PSRC1" is entered in the search field. The report title is "PSRC1 proline/serine-rich coiled-coil 1 [Homo sapiens]" and the Gene ID is 84722, updated on 19-Sep-2010. The "Summary" section is expanded, showing the following information: "Official Symbol PSRC1 provided by HGNC", "Official Full Name proline/serine-rich coiled-coil 1 provided by HGNC", "Primary source HGNC:24472", "Locus tag RP11-297O4.2", "See related Ensembl:ENSG00000134222; HPRD:13127; MIM:613126", "Gene type protein coding", "RefSeq status REVIEWED", "Organism Homo sapiens", and "Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo". The "Table of contents" section on the right lists various sections: Summary, Genomic regions, transcripts, and products, Genomic context, Bibliography, Phenotypes, General gene info, General protein info, Reference sequences, Related sequences, and Additional links. The "Links" section at the bottom includes a link to "Order cDNA clone".

3. On, the human PSRC1 gene page, I scrolled down to the RefSeqs section for the genomic sequence. I selected the FASTA format of the genomic sequence from the Genome Reference Consortium Human Build 37

**RefSeqs of Annotated Genomes: Build 37.1**

The following sections contain reference sequences that belong to a specific genome build: [Explain](#)

**Genome Reference Consortium Human Build 37 (GRCh37), Primary\_Assembly**

**Genomic**

- NC\_000001.10**  
Range: 109822177..109825770, complement  
Download: [GenBank FASTA Sequence Viewer \(Graphics\)](#)
- NT\_032977.9**  
Range: 79794095..79797688, complement  
Download: [GenBank FASTA Sequence Viewer \(Graphics\)](#)

**Alternate assembly (Celera)**

**Genomic**

- AC\_000044.1**  
Range: 108091543..108095136, complement  
Download: [GenBank FASTA Sequence Viewer \(Graphics\)](#)

4. I copied the whole sequence (not all shown here) and then pasted into a word file and saved it. See the word file named psrc1Human.doc.

The screenshot shows the NCBI Nucleotide search interface. The search criteria are set to 'Nucleotide' and the results are displayed in FASTA format. The search results show a 3.59kb region from base 109822177 to 109825770. The sequence is identified as 'Homo sapiens chromosome 1, GRCh37 primary reference assembly'. The FASTA format sequence is displayed, starting with >ref|NC\_000001.10|:c109825770-109822177 Homo sapiens chromosome 1, GRCh37 primary reference assembly. The sequence is shown in a single line, wrapped at 60 characters per line. On the right side, there are options to change the region shown (Whole sequence or Selected region) and to customize the view (Show reverse complement). There are also links to analyze the sequence (Run BLAST, Pick Primers) and to link out to external resources (GeneTests).

This shows testing the local alignment of PSRC1 (human) with ANXA5 (zebrafish) at [http://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi?rm=compare](http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=compare)

I selected FASTA: DNA DNA on the Program pull down menu and then copied and pasted in the whole of the FASTA sequences that I had prepared in my Word files. Then I pressed Compare Sequences...

[Statistical Significance from Shuffles](#)  
[Find Internal Duplications \(lalign/plalign\)](#)

Choose: (A) program and (B, C) sequences to compare:

(A) Program: FASTA: DNA:DNA

Query post-trans modifications  
"\*?#^~+=" included for annotation

(B.1) Enter first (query) sequence: FASTA format Subset range:

```
>ref|NC_000001.10|:c109825770-109822177 Homo sapiens  
chromosome 1, GRCh37 primary reference assembly  
GCTGATTCCC GGGACTAGGTTGCGGGAGAAAGCCTGTTGCGTGGAAGATAAGGCGGCGG  
ACAGGGTGGGCTGGAGGTGTGCTTCGCGTGTGATGCCAGGGGAGTGGAAAGGTGGCTTGG  
ACCCTGAGGATAGTGGCGGGGAGGAAGATGGGGCGTGCAGGCCGCCCCAGTGAGGAGAG  
AGGGCGGCTCCAGGGCAATCTGGGCAGTTGGGGGAGCCCTCCGTATCGTTCCCGGAGGTA
```

[Entrez protein sequence browser](#)

[Entrez DNA sequence browser](#)

(B.2) Or upload sequence from file: Browse...

Protein  DNA (both-strands)  DNA (forward only)  DNA (rev-comp only)

(C.1) Enter the second sequence: FASTA format Subset range:

```
>gnl|ASM:GCF_000000175.2|Zv8_NA7935:c6433-4 Danio rerio  
strain Tuebingen unplaced genomic scaffold, Zv8 NA7935  
CTCCTCATTCTGCTGCTTGTCTGCTTTGTCTGTTGCTCATCCGAACTTAGAAAGTCTT  
GTGCATGAAAATACACATTTAGCCTAAATACCATTACTATAACTTTCTACTCATATTCT  
CAGGTAATATGCTACCTGTCTGAGAACTCAAATTATTTAAATGTTTCAAATTATTATTT  
ACACTTTGTTTATTCAACGTTGTATCTTTGTATATTGTGTAGTATGAATAACTCGACTAT
```

Compare Sequences

Reset Form

(C.2) Or choose file of sequences/accessions: Browse...

Here are the results of the alignment. It seems to have found two regions where there is a short section that looks highly similar -or identical in the second case- which could be some kind of motif.

[Search Databases with FASTA](#) | [Find Duplications](#) | [Search Status](#)

```
# fasta36_t -H -n -c -l -q -w 80 -m 6 -z 11 -Z 10000 -f -12 -g -4 -r +5/-4 TMP.q TMP.q2 6
```

```
FASTA searches a protein or DNA sequence data bank
version 36.06 Aug, 2010
Please cite:
W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448
```

Query: TMP.q

```
l>>>ref|NC_000001.10|:c109825770-109822177 Homo sapiens chromosome 1, GRCh37 primary referenc - 3279 nt
Library: TMP.q2
6430 residues in 1 sequences
```

```
Statistics: (shuffled [500]) MLE statistics; Lambda= 0.1290; K=0.000356
statistics sampled from 1 (1) to 500 sequences
Algorithm: FASTA (3.7 Mar 2010) [optimized]
Parameters: +5/-4 matrix (5:-4) ktup: 6
E-join: 0.1 (0), E-opt: 0.02 (1), open/ext: -12/-4, width: 16
Scan time: 0.960
```

```
The best scores are:
gnl|ASM:GCF_000000175.2|Zv8_NA7935:c6433-4 Danio rerio strain Tueb (6430) [r] 66 23.7 0.78 align
gnl|ASM:GCF_000000175.2|Zv8_NA7935:c6433-4 Danio rerio strain Tueb (6430) [f] 60 22.6 0.96 align
```

>>>ref|NC\_000001.10|:c109825770-109822177, 3279 nt vs TMP.q2 library

```
>>gnl|ASM:GCF_000000175.2|Zv8_NA7935:c6433-4 Danio rerio strain Tuebingen unplace (6430 nt)
rev-comp initn: 66 initl: 66 opt: 66 Z-score: 42.3 bits: 23.7 E(10000): 0.78
banded Smith-Waterman score: 80; 80.0% identity (80.0% similar) in 25 nt overlap (1683-1659:5827-5851)
Entrez Lookup Re-search database General re-search
```

```
1720 1710 1700 1690 1680 1670 1660 1650
ref|N- TTGAGTTTATCATAGACATATACTCAGACACTGTGGTAAAGAGTTTGGTTCTCTTTTTTCATGGTTCATTTAATCCCA
: : : : : : : : : : : : : : : :
gnl|AS CATTAGGTTCCCAACACAATTTCTTAACTTAAATGGCATTTTGGATCTCTTTCTGCTGGTGATACTGTTGTCCTA
5790 5800 5810 5820 5830 5840 5850 5860

1640 1630 1620 1610 1600 1590 1580 1570
ref|N- ACAATTCTAGATGTAGAAATGCTATCCCTGTTTTACCAATAAGGAAACCACTCCCTGCATCATGGCACACACCATGTCGTG
gnl|AS AAAGAAAAAAGGGAAAAACATAAACAAAAATCTTGCATACATTAAAGAAATGGTATTCAGAAAAATTTGGCAGTTTT
5870 5880 5890 5900 5910 5920 5930 5940
```

```
>>gnl|ASM:GCF_000000175.2|Zv8_NA7935:c6433-4 Danio rerio strain Tuebingen unplace (6430 nt)
initn: 60 initl: 60 opt: 60 Z-score: 36.3 bits: 22.6 E(10000): 0.96
banded Smith-Waterman score: 60; 100.0% identity (100.0% similar) in 12 nt overlap (1386-1397:5441-5452)
Entrez Lookup Re-search database General re-search
```

```
1350 1360 1370 1380 1390 1400 1410 1420
ref|NC CTCGGCGGGAGACCTTTGTGCTGAAGGATAGTCTGTCGGAGACCTGCTGCCACCTGTGAACCTTTGACGCGGAGCACC
: : : : : : : : : :
gnl|AS TATATGAAATAAACCTGTTCCCAAGCAGGTTTAAACTTGCAGACCTGCTGCCATGACAACAACACCTGAATGAGTTTCAT
5410 5420 5430 5440 5450 5460 5470 5480

1430 1440 1450 1460 1470 1480 1490 1500
ref|NC CCCTCCCAAGCAGCCTGACGCTCGACTCCGGAGTAATGATAGGAAGGGTTCAGTCAGGGCTCTCCGGGCTACATCTGG
gnl|AS AGAACTACCTTGGATTGTGTCAAATTGTCAATAATAAAAATCCAGTAATCCAGGTATGGAGAATAGGCATGGGCCGGTAT
5490 5500 5510 5520 5530 5540 5550 5560
```

Next step is to see if this region comes up in other sequence comparisons and to track down if this part of the gene sequence has already been annotated with some kind of functionality.  
To be continued...!