Entrez Gene Quick Start
An NCBI Mini-Course

NCBI's Entrez Gene provides gene-based information such as chromosome location, sequence, expression, structure, functional, and homology data. Each record represents a single gene from an organism. Entrez Gene includes organisms for which there is a RefSeq genome record.

In this course, we will learn how to obtain information about a human gene such as:

- mRNA, genomic, and protein sequence
- general gene and protein information
- homologs from other eukaryotes
- known SNPs, and whether the SNPs in the coding region alter the function of the protein product
- phenotypes associated with mutations
- protein structure

Entrez Gene is the successor to LocusLink. The course will also cover the advantages of Entrez Gene such as efficient searching options and availability of gene-specific information for all completely sequenced genomes, including bacteria and viruses.

The following handout includes the screen shots of the exercise demonstrated in the mini-course.


Instructor: Wayne Matten (matten@ncbi.nlm.nih.gov)
Problem 1

Retrieve human entries related to "prion protein" in Entrez Gene. Identify the gene for prion protein (PRNP). Name the map location of this gene on the human genome. What is the function of this protein? What are the alternate gene symbols? Name the phenotypes associated with the mutations in this gene.

Is the RefSeq mRNA record reviewed? How many alternatively spliced products have been annotated for the gene?

To obtain information about the homologs from other eukaryotes, click on the Homologene link. Change the Display option to "Alignment Scores". How great is the percent identity between the human and mouse proteins? View the alignment by clicking on the "Blast" link.

Go back to the Entrez Gene report. Identify the variations annotated on this gene by clicking on the geneView in dbSNP link. How many of them are nonsynonymous changes? To determine whether known SNPs in the coding region of a gene are associated with any phenotype, access the OMIM record by clicking on the "Yes" link under the OMIM column in the SNP report. Compare the nonsynonymous changes from the SNP report with the "ALLELIC VARIANTS" in the OMIM record. Are there any SNPs known to cause a change in the function of the prion protein?

Go back to the Entrez gene report. View the list of similar proteins through the "BL" link in the next to the protein NP_000302. To view the site of mutation in the 3D structure, superimpose the protein sequence on the 3D-structure of human prion protein (use BL--3D-structure button--click on the first blue dot--Get 3D Structure Data). Identify and highlight the mutated residue on the 3D structure.
**Gene Information**

**Gene ID:** 5621  
**Primary source:** [HGNC](https://www.genenames.org)  
**Updated:** 07-Aug-2006

**RefSeq status:** Reviewed  
**Total gene size:** 15166 bp

### Genomic regions, transcripts, and products

<table>
<thead>
<tr>
<th>mRNA</th>
<th>Protein</th>
<th>aa</th>
<th>exons</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002311.2</td>
<td>NP_002980.1</td>
<td>254</td>
<td>1</td>
</tr>
<tr>
<td>NM_183079.1</td>
<td>NP_888902.1</td>
<td>254</td>
<td>1</td>
</tr>
</tbody>
</table>

**Exon information:**

- NM_002311.2: length 2468 bp, number of exons 2
- NP_002980.1: length 254 aa, number of exons 1

<table>
<thead>
<tr>
<th>EXON</th>
<th>Coding EXON</th>
<th>INTRON</th>
</tr>
</thead>
<tbody>
<tr>
<td>coords</td>
<td>length</td>
<td>coords</td>
</tr>
<tr>
<td>51 - 90</td>
<td>90 bp</td>
<td>91 - 12788</td>
</tr>
<tr>
<td>12789 - 15166</td>
<td>2378 bp</td>
<td>12799 - 13560</td>
</tr>
</tbody>
</table>

- NM_183079.1: length 2464 bp, number of exons 2
- NP_888902.1: length 254 aa, number of exons 1

### Genomic context

**Chromosome:** 20  
**Location:** 20p13

### Bibliography

1. Sequencing the PRNP gene in Case-2 and HT-29 parental and clonal cell lines, revealed that these cells have a distinct polymorphism at codon 129 and PrPc is expressed in epithelial cells.  
2. A form of polymorphism at codon 129 of prion protein gene confers similar genetic susceptibility to vCJD in British populations as in European and Asian countries.  
3. The presence of Prp variant alleles was found not to be significantly associated to cognitive performance of patients with mental temporal lobe epilepsy with hippocampal sclerosis.  
4. Protein-resistant prion protein (PrPres) in spontaneous lymphoblastic tumors of mice infected with scrapie.  

[PubMed links]
95. Polymorphism of the codon 129 of the prion protein (PrP) gene and neuropathology of cerebral aging

96. a specific prion protein fragment has a role in dodecaside stability by retarding the rate of fibril formation

97. Shiels' syndrome of PrP (195-213)

### HIV-1 protein interactions

HIV-1 Tat binds to a stem-loop structure in the mRNA of prion protein (PrP) that is similar to HIV-1 TAR RNA and infection of astrocytes with HIV-1 results in an increased level of PrP mRNA, suggesting Tat upregulates PrP expression

<table>
<thead>
<tr>
<th>Description</th>
<th>Interactant</th>
<th>Other Gene</th>
<th>Complex</th>
<th>Source</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTR_061158.1</td>
<td>APEB1</td>
<td></td>
<td></td>
<td>HPRD</td>
<td>PubMed</td>
</tr>
<tr>
<td>NTR_065720.1</td>
<td>CLSTN1</td>
<td></td>
<td></td>
<td>HPRD</td>
<td>PubMed</td>
</tr>
<tr>
<td>NTR_065720.1</td>
<td>CLU</td>
<td></td>
<td></td>
<td>HPRD</td>
<td>PubMed</td>
</tr>
<tr>
<td>NTR_061158.1</td>
<td>CTN1</td>
<td></td>
<td></td>
<td>HPRD</td>
<td>PubMed</td>
</tr>
</tbody>
</table>

PrP is interacts with CSNK2A1 (CK2 alpha). This interaction was modeled on a demonstrated interaction.

### General gene information

**Markers**
- WI-18738(e-PCR) (Links: Unigene:10817)
- Alternate names: IUA: 25, EHIS:2301, STS-D00015
- SGC44304(e-PCR) (Links: Unigene:2335)
- Alternate names: EST498246; BH579429
- D20S1014(e-PCR) (Links: Unigene:21619)
- Alternate names: G60-677-676, GDB-677676; RH14068, RH6750, SHGC-12813;
- UTR-10221, WL-7784, gdb:102720, sting:10911
- RH1030(e-PCR) (Links: Unigene:36672)
- Alternate names: GDB:177793; strG20232
- RH10249(e-PCR) (Links: Unigene:43453)
- Alternate name: T27631
- RH1030(e-PCR) (Links: Unigene:58320)
- Alternate names: GDB:513003; strG20379
- GDB-185261(e-PCR) (Links: Unigene:155287)
- GDB-185261(e-PCR) (Links: Unigene:158256)
- PRNP-2(e-PCR) (Links: Unigene:224000)
- PMC150957P1(e-PCR) (Links: Unigene:270809)

**Gene Ontology**

Provided by GOA

<table>
<thead>
<tr>
<th>Function</th>
<th>Evidence</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>copper ion binding</td>
<td>TAS</td>
<td>PubMed</td>
</tr>
<tr>
<td>guanidinium binding</td>
<td>DA</td>
<td>PubMed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process</th>
<th>Evidence</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>copper ion homeostasis</td>
<td>NAS</td>
<td>PubMed</td>
</tr>
<tr>
<td>metabolism</td>
<td>TAS</td>
<td>PubMed</td>
</tr>
<tr>
<td>response to oxidative stress</td>
<td>ISS</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Evidence</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golgi apparatus</td>
<td>ISS</td>
<td></td>
</tr>
<tr>
<td>endoplasmic reticulum</td>
<td>ISS</td>
<td></td>
</tr>
<tr>
<td>extracellular to membrane</td>
<td>TAS</td>
<td>PubMed</td>
</tr>
<tr>
<td>lipid raft</td>
<td>ISS</td>
<td></td>
</tr>
</tbody>
</table>
Homology:
    Mouse, Rat

Map Viewer

Phenotypes
    Creutzfeldt-Jakob disease MIM: 123400
    Gerstmann-Sträussler disease MIM: 137440
    Huntington disease-like 1 MIM: 603212
    Insomnia, fatal familial MIM: 680072
    Prion disease with protracted course MIM: 606688

Pathways
    KEGG pathway: Neuro-degenerative Disorders 01510
    KEGG pathway: Prion disease 05060

General protein information

Names: prion protein

CD230 antigen, prion protein PrP, major prion protein, prion-related protein, prion protein (p27-30)
(Creutzfeld-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia)

NCBI Reference Sequences (RefSeq)

mRNA Sequence NM_000311

Transcriptional Variant
    Transcript Variant: This variant (1) represents the longer transcript. Variants 1 and 2 encode the same protein.

Source Sequence BC002532, BG077054, M13899

Product NP_000302 prion protein preproprotein

Consensus CDS (CCDS) CCDS13080_1 (preliminary)

Conserved Domains (1) summary

PRNPSF_SPL11824 (PRNP, prion protein, p27-30)

Summary

Official Symbol: PRNP and Name: prion protein (p27-30) (Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia) [Homo sapiens]

GeneID: 5621 Primary source: HUMAN 9449 updated 07-Aug-2006

Organism: Homo sapiens

Lineage: Bacteria; Archaea; Eukaryota; Chordata; Craniata; Vertebrata; Tetrapoda; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo


Summary: The protein encoded by this gene is a membrane glycoprophosphatidylinositol-anchored glycoprotein that tends to aggregate into rod-like structures. The encoded protein contains a highly unstable
BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.14 [Apr-08-2006]


Length = 253 (1..253)


Length = 254 (1..254)

NOTE: Expect and evalue are calculated based on the size of the nr database.
<table>
<thead>
<tr>
<th>#</th>
<th>Gene ID</th>
<th>Location</th>
<th>Description</th>
<th>Reference</th>
<th>p.Ran</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4620403</td>
<td>g.11638777</td>
<td>N.D.</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>4620464</td>
<td>g.8993335</td>
<td>N.D.</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>4620521</td>
<td>g.1800014</td>
<td>N.D.</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>4620538</td>
<td>g.6052772</td>
<td>N.D.</td>
<td>Yes</td>
<td>synonymous</td>
<td>T</td>
</tr>
<tr>
<td>5</td>
<td>4620549</td>
<td>g.11852079</td>
<td>N.D.</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>4620523</td>
<td>g.11538777</td>
<td>N.D.</td>
<td>Yes</td>
<td>nonsynonymous</td>
<td>G</td>
</tr>
</tbody>
</table>

**OMIM**

*176640 PRION PROTEIN; PRNP*

**Alternative titles; symbols**

PRP

**PRION-RELATED PROTEIN; PRIP**

Gene map locus: 20pter-p12
**176640**

**PRION PROTEIN: PRNP**

**ALLELIC VARIANTS**

*(selected examples)*

- 0001 CREUTZFELDT-JAK OR DISEASE (PRNP, EXTRA OCTAPEPTIDE CODING REPEATS)
- 0002 GROETSMANN-STRAUSSLER DISEASE (PRNP, PR012LBU)
- 0003 REMOVED FROM DATABASE
- 0004 GROETSMANN-STRAUSSLER DISEASE (PRNP, ALA117VAL)
- 0005 PRION DISEASE SUSCEPTIBILITY TO (PRNP, MET129VAL) dbSNP
- 0006 CREUTZFELDT-JAK OR DISEASE (PRNP, GLU200LYS) dbSNP
- 0007 CREUTZFELDT-JAK OR DISEASE (PRNP, ASP178ASN AND MET129VAL)
- 0008 REMOVED FROM DATABASE
- 0009 REMOVED FROM DATABASE
- 0010 FATAL FAMILIAL DEMYELINATION (PRNP, ASP178ASN AND MET129)
- 0011 GROETSMANN-STRAUSSLER DISEASE (PRNP, PHE198SER)
- 0012 GROETSMANN-STRAUSSLER DISEASE (PRNP, GLN171ARG)
- 0013 REMOVED FROM DATABASE
- 0014 CREUTZFELDT-JAK OR DISEASE (PRNP, VAL210ILE)
- 0015 GROETSMANN-STRAUSSLER DISEASE (PRNP, PR010LBU)
- 0016 CREUTZFELDT-JAK OR DISEASE (PRNP, VAL180ILE)
- 0017 CREUTZFELDT-JAK OR DISEASE (PRNP, MET122ARG)
- 0018 SPONGIFORM ENCEPHALOPATHY WITH NEUROPSYCHIATRIC FEATURES (PRNP, ASN171SER) dbSNP
- 0019 CREUTZFELDT-JAK OR DISEASE, PROTECTION AGAINST (PRNP, GLU219LYS)
- 0020 CREUTZFELDT-JAK OR DISEASE, PRION-LIKE PRP
Lineage: Bacteria; Metazoa; Chordata; Cenozoic; Primates; Catarrhini; Homo

Gene aliases: C3T, OSS, PFR, ASCB, PRF, PFR; CD260, MGC26679, PFR27-30; PFR33-35C

Summary: The protein encoded by this gene is a membrane glycoprophosphatidylinositol-anchored glycoprotein that tends to aggregate into rod-like structures. The encoded protein contains a highly variable region of five tandem octapeptide repeats. This gene is found on chromosome 20, approximately 20 kb upstream of a gene which encodes a biochemically and structurally similar protein to the one encoded by this gene. Mutations in the repeat region as well as elsewhere in the gene have been associated with Creutzfeldt-Jakob disease, familial amyloid polyneuropathy, German-Steinert disease, Huntington disease-like 1, and locus. Two transcript variants encoding the same protein have been found for this gene.

Genomic regions, transcripts, and products

RefSeq below

MT-611987.8

Genomic context

Chromosome: 20, Location: 20pter-p12

NCBI Blast search results:

Query: g(HM80113) prion protein prionprotein [Homo sapiens]

Blast hits with 40 unique species:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA001142</td>
<td>prion protein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001142</td>
<td>prion protein prionprotein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001149</td>
<td>prion protein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001149</td>
<td>prion protein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001149</td>
<td>prion protein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001149</td>
<td>prion protein [Homo sapiens]</td>
</tr>
<tr>
<td>AA001149</td>
<td>prion protein</td>
</tr>
</tbody>
</table>
Problem 2

Retrieve human entries related to "colon cancer" in Entrez Gene. Identify the gene MLH1. Name the map location of this gene on the human genome. What is the function of this protein? What are the alternate gene symbols? Name the phenotypes associated with the mutations in this gene.

Is the RefSeq mRNA record reviewed? How many alternatively spliced products have been annotated for the gene?

To obtain information about the homologs from other eukaryotes, click on the Homologene link. Change the Display option to "Alignment Scores". How great is the percent identity between the human and mouse proteins? View the alignment by clicking on the "Blast" link.

Go back to the Entrez Gene report. Identify the variations annotated on this gene by clicking on the geneView in dbSNP link. How many of them are nonsynonymous changes? To determine whether known SNPs in the coding region of a gene are associated with any phenotype, access the OMIM record by clicking on the "Yes" link under the OMIM column in the SNP report. Compare the nonsynonymous changes from the SNP report with the "ALLELIC VARIANTS" in the OMIM record. Are there any SNPs known to cause a change in the function of the MLH1 protein?

Go back to the Entrez gene report. View the list of similar proteins through the "BL" link in the next to the protein NP_000240. To view the sites of mutations in the 3D structure, superimpose the protein sequence on the 3D-structure of E.coli multL protein 1BKNB (use BL--3D-structure button--click on the second blue dot--Get 3D Structure Data). Identify and highlight the amino acid corresponding to the human MLH1 isoleucine 32 on the 3D structure. What is the amino acid at this position in the E.coli protein? Based on this information, do you think the I32V mutation in the human protein will alter its function? Confirm your findings through the OMIM record for MLH1.