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.


Course developed by Medha Bhagwat (bhagwat@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.
PRNP prion protein (p27-30) (Creutzfeldt-Jakob disease, Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia) [Homo sapiens]

**Summary**

The protein encoded by this gene is a membrane glycosylphosphatidylinositol-anchored glycoprotein that tends to aggregate into rod-like structures. The encoded protein contains a highly unstable 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 this gene have been associated with Creutzfeldt-Jakob disease, fatal familial insomnia, Gerstmann-Straussler disease, Huntington disease-like 1, and kuru. Alternative splicing results in multiple transcript variants encoding the same protein.
**Genomic context**

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

**Bibliography**

**Related Articles in PubMed**

**GeneReferences into Function**

1. Results suggest that the PRNP genetic variants are not associated with the risk for Alzheimer’s disease in Korean population.
2. The study that provides experimental evidence supporting the hypothesis that the prion gene may be located in normal human brain.
4. Polyposis microsatellite sites within 146 kb of the human prion gene complex, including the genes PRNP, PRNDC, and PRNDC2 are highly conserved.
5. The highest affinity copper (II)-binding mode causes self-association of both peptides, suggesting a role for copper (II) in controlling prion protein self-association in vivo.
6. There is a link between IL-12 and the formation of cytokine from PPE in response to vaccination or cell-mediated immune response.
7. Prion protein may act as an inhibitor of microtubule assembly by inducing formation of stable tubulin oligomers.

---

**HIV-1 protein interactions**

Protein Interaction

1. **Tat**
   - 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>Pubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_000302.1</td>
<td>NP_001455.1</td>
<td>APBB1</td>
<td>HIVD</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>NP_000302.1</td>
<td>NP_05759.3</td>
<td>CLSTN1</td>
<td>HIVD</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>NP_000302.1</td>
<td>NP_00132.2</td>
<td>CLU</td>
<td>HIVD</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>NP_000302.1</td>
<td>NP_01634.2</td>
<td>CNTN1</td>
<td>HIVD</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>PrPC interacts with CSN2A2 (O2 alpha). This interaction was modeled on a demonstrated interaction between bovine PrPC and human CSN2A2 (O2 alpha).</td>
<td>NP_001455.1</td>
<td>CSN2A1</td>
<td>BND</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>PrPC interacts with CSN2A2 (O2 alpha prime). This interaction was modeled on a demonstrated interaction between bovine PrPC and human CSN2A2 (O2 alpha prime).</td>
<td>NP_001455.1</td>
<td>CSN2A2</td>
<td>BND</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>PrPC interacts with CSK2B (O2 beta). This interaction was modeled on a demonstrated interaction between bovine PrPC and human CSK2B (O2 beta).</td>
<td>NP_001455.1</td>
<td>CSK2B</td>
<td>BND</td>
<td>PubMed</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Type</td>
<td>Location</td>
<td>NCBI Reference</td>
<td>Amino Acid Change</td>
<td>Maternal</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>4620402 111587672</td>
<td>ND</td>
<td>ND</td>
<td>g111587672</td>
<td>non-synonymous C</td>
<td>Ala [A]</td>
</tr>
<tr>
<td>4625464 1205832325</td>
<td>ND</td>
<td>Yes</td>
<td>g205832325</td>
<td>syn (T)</td>
<td>Val [V]</td>
</tr>
<tr>
<td>4625216 11800614</td>
<td>ND</td>
<td>Yes</td>
<td>g11800614</td>
<td>non-synonymous A</td>
<td>Lys [K]</td>
</tr>
<tr>
<td>4625338 11800614</td>
<td>ND</td>
<td>Yes</td>
<td>g11800614</td>
<td>non-synonymous A</td>
<td>Lys [K]</td>
</tr>
<tr>
<td>4625535 11800614</td>
<td>ND</td>
<td>Yes</td>
<td>g11800614</td>
<td>non-synonymous A</td>
<td>Lys [K]</td>
</tr>
<tr>
<td>4625545 11800614</td>
<td>ND</td>
<td>Yes</td>
<td>g11800614</td>
<td>non-synonymous A</td>
<td>Lys [K]</td>
</tr>
<tr>
<td>4625625 11800614</td>
<td>ND</td>
<td>Yes</td>
<td>g11800614</td>
<td>syn (G)</td>
<td>Gly [G]</td>
</tr>
</tbody>
</table>

---

**0.006 CREUTZFELDT-JAKOB DISEASE [PRNP, GLU200LYS] [g180]**

**FATAL FAMILIAL INSOMNIA INCLUDED**

In 2 patients with Creutzfeldt-Jakob disease (133406) from the same family, Goldgaber et al. (1989) identified a G-to-A transition in the PRNP gene, resulting in a glut200-to-lys (E200K) substitution.

Studying an unusual cluster of cases of CJD in rural Slovakia, Goldgaber et al. (1990) found the E200K mutation in all 11 tested cases of focal CJD, in 12 of 40 healthy first-degree relatives, and in 6 of 25 other relatives. By contrast, no extrafamilial cases or their relatives had the mutation, nor did any unrelated individuals within or outside the cluster regions. One of the healthy individuals with the E200K mutation was the 75-year-old mother of one of the patients. The unusually high incidence of CJD in the Orava and Liptov regions of Slovakia appeared to be of recent origin. Goldgaber et al. (1990) interpreted this as indicating that the mutation is a necessary, but not sufficient, factor in the disease. Another factor such as scrapie-infected sheep was proposed. Mitrov et al. (1990) described the familial occurrence of 3 definite and 2 possible cases of CJD with temporal and spatial separation in the area of focal CJD accumulation in Slovakia. The incubation period appeared to be about 51 years, judging by the interval between the death of the affected mother and the clinical onset in the first affected child. Affected offspring tended to die at the same time, not at the same age. Due to separation of the affected children's possible common exposure to CJD infection was limited to approximately 7 years during their childhood.
**176640**

**PRION PROTEIN: PRNP**

### ALLELIC VARIANTS

(Selected examples)

- 0001 CJD (atypical) [PRNP, extra octapeptide coding repeats]
- 0002 CJD [PRNP, PRO102LEU]
- 0003 REMOVED FROM DATABASE
- 0004 GSS [PRNP, ALA117VAL]
- 0005 PRION DISEASE, SUSCEPTIBILITY TO [PRNP, MET129VAL] [**dbSNP**]
- 0006 CJD [PRNP, GLU200LYS] [**dbSNP**]
- 0007 CJD [PRNP, ASP178ASN AND MET129VAL]
- 0008 REMOVED FROM DATABASE
- 0009 REMOVED FROM DATABASE
- 0010 FAMILIAL NEURODEGENERATIVE DISEASE [PRNP, ASP178ASN AND MET129]
- 0011 GSS [PRNP, PHE198SER]
- 0012 GSS [PRNP, GLN217ARG]
- 0013 REMOVED FROM DATABASE
- 0014 CJD [PRNP, VAL210ILE]
- 0015 GSS [PRNP, PRO102LEU]
- 0016 CJD [PRNP, VAL180ILE]
- 0017 CJD [PRNP, MET129ARG]
- 0018 SPONGIFORM ENCEPHALOPATHY WITH NEUROPSYCHIATRIC FEATURES [PRNP, ASN171SER] [**dbSNP**]
- 0019 CJD [PRNP, GLU200LYS]
- 0020 CJD PROTECTION AGAINST [PRNP, GLU200LYS]
- 0021 CJD PROTECTION AGAINST [PRNP, GLU200LYS]
upstream of a gene which encodes a biochemically and structurally similar protein to the one encoded by this gene. Mutations in this repeat region as well as elsewhere in this gene have been associated with Creutzfeldt-Jakob disease, fatal familial insomnia, Gerstmann-Straussler disease, Huntington disease-like 1, and kuru. Two transcript variants encoding the same protein have been found for this gene.

1. Elevated plasma PrP(C) levels in renal disease were observed, showing that plasma PrP (C) is not a specific marker of neurological disease or Creutzfeldt-Jakob disease.
2. The polymorphism at residue 129 does not change efficiency of conversion to beta-PrP conformation or affect binding of prion fibers, but in a partially desensitized...
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 mutL 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.