

## **Module 3: Genes and Sequences (NCBI)**

### **iv - What Information Is Available For a ZGC Clone and its Protein?**

#### **Aims**

- Introduce options to mine data related to a ZGC clone
- Suggest alternative sources of functional information
- Provide examples of orthologs providing functional information
- Show examples of links on nucleotide and protein records

#### **Introduction**

There are a number of alternative ways to determine what is known about a particular Zebrafish Gene Collection (ZGC) cDNA clone. To quickly determine if there are any related records in any of the Entrez databases, you can click on 'Links' in the upper right hand corner of the GenBank page for your cDNA of interest.

Following these links you could answer these questions:

- Is the cDNA clone in a UniGene cluster or Gene record ?
- Are there similar proteins in zebrafish or orthologs whose function is known ? (see also Module 3\_ii)
- Where is this cDNA known to be expressed, based on expression of similar ESTs ? (see also Module4\_i)
- Does this cDNA encode a conserved domain ? (see also Module3\_ii)
- Has this cDNA been annotated on the genome ? (see Module2\_iii)

#### **Exercises**

1. Examine links to related resources found on a Entrez Nucleotide record.
2. Examine links to related resources found on a Entrez Protein record.
3. Examine links to related resources found on a UniGene record to other similarly expressed UniGene records or homologous UniGene records.

## 1. Examine links to related resources found on an Entrez Nucleotide record.

To begin, go to the Entrez Nucleotide (GenBank) record for the cDNA. (for example, BC071422:

<http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=Nucleotide&dopt=GenBank&val=47937949>). From the Nucleotide record, click on the Links pull down menu:

menu:

to view the set of available related links.

The screenshot displays the NCBI Entrez Nucleotide record for BC071422. The top navigation bar includes links for PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, OMIM, and Books. The search bar shows 'Nucleotide' selected. Below the search bar, there are options for 'Display' (GenBank), 'Send' (all to file), and 'Range' (from begin to end). The record details for BC071422 are shown, including the definition: 'Danio rerio zgc:86750, mRNA (cDNA clone MGC:86750 IMAGE:6899054), complete cds.' The 'Links' pull-down menu is open, showing options: Gene, Full text in PMC, Related Sequences, Protein, PubMed, Taxonomy, and LinkOut. An arrow points to the 'Links' menu.

## 2. Examine links to related resources found on an Entrez Protein record.

You could also start at the Entrez Protein page (AAH71422:

<http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=AAH71422.1>)

to view the set of related databases in the Links pull-down menu.

**Navigate to related proteins records such as BLink or the conserved domains database (CDD)**

NCBI Entrez Protein

Search Protein for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display GenPept Send all to file

Range: from begin to end Features:  SNP  CDD  MGC  HPRD  STS

1: [AAH71422](#). Reports Zgc:86750 [Danio ...[gi:47937950]

BLink, Domains, MGC cDNA clone,

**Links**

- Gene
- Full text in PMC
- Related Sequences
- Domain Relatives
- Nucleotide
- OMIM
- PubMed
- Taxonomy
- LinkOut

LOCUS AAH71422 174 aa linear VRT 08-MAR-2005

DEFINITION Zgc:86750 [Danio rerio].

ACCESSION AAH71422

VERSION AAH71422.1 GI:47937950

DBSOURCE accession [BC071422.1](#)

KEYWORDS MGC.

SOURCE Danio rerio (zebrafish)

ORGANISM [Danio rerio](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (residues 1 to 174)

AUTHORS Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T, Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P, Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH, Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW, Villalón DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J, Helton E, Kettman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW, Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U, Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.

CONSRM Mammalian Gene Collection Program Team

3. Examine links to related resources found on a UniGene record to other similarly expressed UniGene records or homologous UniGene records.

### Query the UniGene database:

Click on links found in the drop-down Links menu to go to other UniGene clusters with similar expression patterns (**Co-expressed UniGene**) or homologous UniGene clusters determined by Homologen comparisons (**Homologous UniGene**)

The screenshot shows the NCBI UniGene search interface. At the top, the NCBI logo and UniGene title are visible, along with a user login area for 'schriml'. Below the search bar, the query 'Dr.25683 OR Dr.24802' is entered. The results are displayed in a summary view, showing two items: 'Dr.25683' (ctslb: Cathepsin L, b) and 'Dr.24802' (hspa8: Heat shock protein 8). A 'Links' button is present next to the second result.

This is a 'Links' popup menu for the first result. It contains a list of database links: Gene, HomoloGene, Nucleotide, Protein, PubMed, Taxonomy, Homologous UniGene, and UniSTS.

This is a 'Links' popup menu for the second result. It contains a list of database links: Gene, HomoloGene, Nucleotide, Protein, PubMed, Taxonomy, Co-expressed UniGene, Homologous UniGene, and UniSTS.

