C.elegans Dauer Gene Expression Database (Dauer DB)

- User Manual -

Version 1.0

2009-02-02

Yonsei Proteome Research Center

KOREA

Contents

Abbreviation

GO	Gene Ontology
SOM	Self-Organizing Map
CGC	Caenorhabditis Genetics Center

1. First-Page Menu

You can access the *C.elegans* Dauer Gene Expression Database termed Dauer Metabolic DB using the URL(http://www.dauerdb.org).

- **Total list :** Use the total list browser to explore total dauer gene data in this DB.

    For more information on the total list menu, please refer to "2-1. Browsing Total List".

- **Search :** You can search for detail information on those expressed genes of interest.

    For more information on "Search", please refer to "4. Searching Expression Data".

- **GO browsing :** Use Gene Ontology(GO) classification browser to explore genes with function annotation.

    For more information on GO browser, refer to "2-2. Browsing via Gene Ontology classification".

- **SOM classification :** Browse clustered genes by gene expression pattern.

    For more information on SOM classification, please refer to "2-3. Browsing via SOM clustering".

- **Pathway Map :** Browse by KEGG pathway map

    For more information on Pathway Map, please refer to "2-4. Browsing by Pathway Map".

- **CGC list :** Use this menu to browse CGC gene list.

    For more information on CGC list, please refer to "2-5. Browsing by CGC list".

- **Dauer Comparison Category :** You can compare the expression level of genes between those induced by daumone and those induced by broth culture in this section.

    For more information on Dauer Comparison category, please refer to "2-6. Browsing by Dauer Comparison".

Figure 1-1. First page.

2. Browsing *C.elegans* Dauer Microarray Data

2-1. Browsing Total List.

Use the total list browser to explore total microarray data of *C.elegans* dauer state in this database.

Click "Total list".
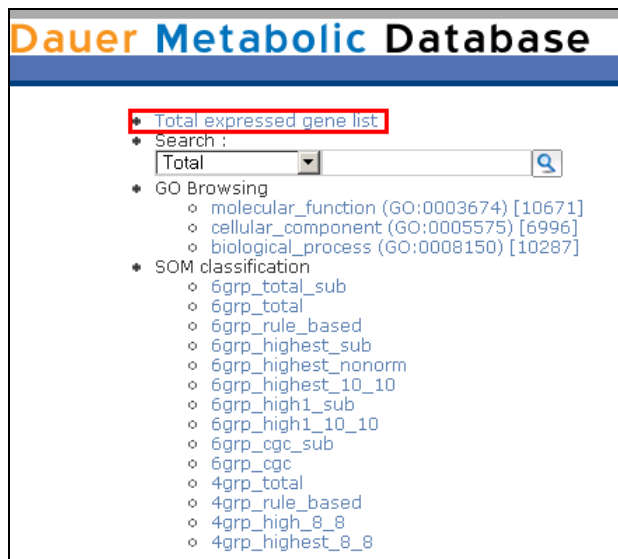
Figure 2-1. Total list menu on first page.

This listing page includes some information as to gene names, expression patterns of those 6-group in S1~S6 (on dauer plate) and those 4-group in L1~L3 and dauer (on broth culture), gene ontology, gene description and SOM cluster information (Figure 2-2).

Figure 2-2. List page of microarray database.

If you want to see more detailed in formation on the expression pattern of individual gene, click Probe ID number on gene information part.

For more information on this part, please refer to "3. Detail Microarray Information".

2-2. Browsing via Gene Ontology Classification

All Genes probed on DNA microarray are categorized Gene Ontology which is a controlled vocabulary used to describe the biology of a gene product in any organism. There are three independent sets of vocabularies, or ontologies, which describe the molecular function of a gene product, the biological process in which the gene product participates, and the cellular component where the gene product can be found. In the first page, you can see this Gene Ontology classified gene list. Gene Ontology has three initial parts – molecular function (GO ID is 3674), cellular component (GO ID is 5575), and biological process (GO ID is 8150).

The number on the right side of Gene Ontology term indicates total gene number which is annotated this Gene Ontology term. In Figure 3-3, total gene numbers annotated with molecular function term are 11,349.

Figure 2-3. GO browsing.

Click gene ontology term to see Gene Ontology terms bellow. Or, click 'number' on the right side to list these genes.

Figure 2-4. GO browsing in molecular function

When you click gene ontology term, you can see gene ontology list with **two numbers**. First number is summation of gene number which annotates exact this term and gene number which annotated bellow GO term. The other number is the gene number which annotates bellow GO term.

"**[dist]**" is the distribution map link of total gene which annotated the GO term.

Distribution map has two parts – stage ratio distribution map and stage distribution map. Stage ratio distribution map shows the expression ratio between one stage and next stage. Stage distribution map shows the distribution of gene expression list. Figure 3-5 shows the distribution of 42 genes with GO annotation (microtubule motor activity).
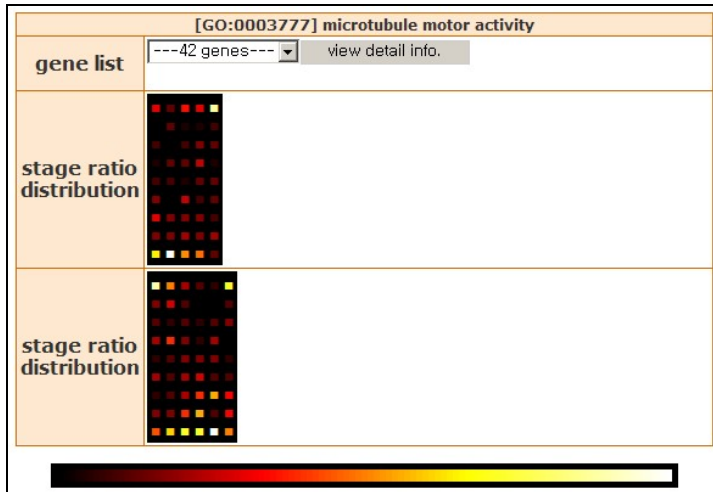
Figure 2-5. The distribution map of genes annotated microtubule motor activity (GO:3777)

2-3. Browsing via SOM clustering

SOM (Self-Organizing Map) is one of the clustering algorithm which is often used in microarray clustering. *C.elegans* dauer gene expression employed the clustering results obtained from the data analysis with GeneCluster 2.0[1]. SOM cluster needs initial seed dimension parameters. In this clustering, we used 10 X 10 or 8 X 8 dimension parameters. In cluster titles, "_sub" cluster is regrouped again by manual annotation after SOM clustering.

Figure 2-6. SOM classification menu in first page**.**
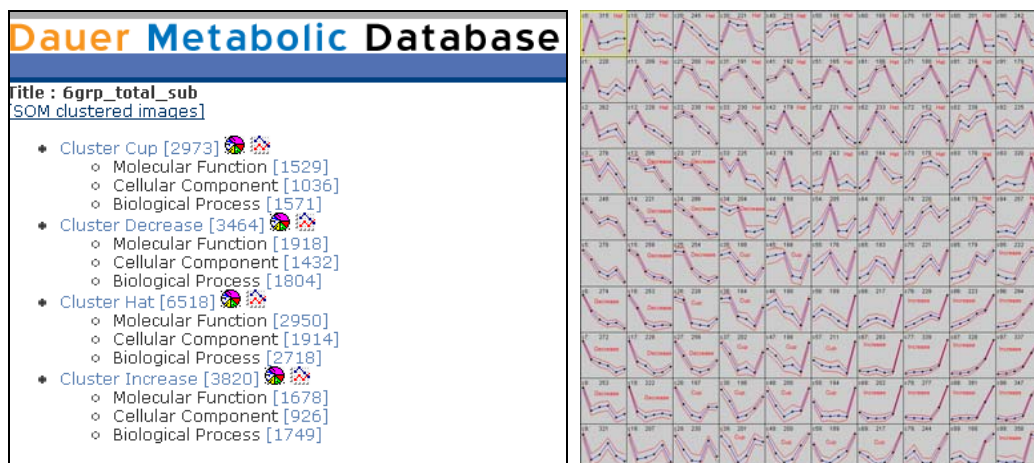
Figure 2-7. The clusters were regrouped by manual annotation and the expression maps of each cluster.

In figure 2-7, we regrouped 4 categories by expression pattern – cup, hat, suppression and induction shape. The number in "[ ]" is the total gene number in the cluster. The each cluster is classified by Gene Ontology.

For more information GO clustering, refer to "2-2 Browsing via gene ontology classification"

2-4. Browsing by Pathway Map
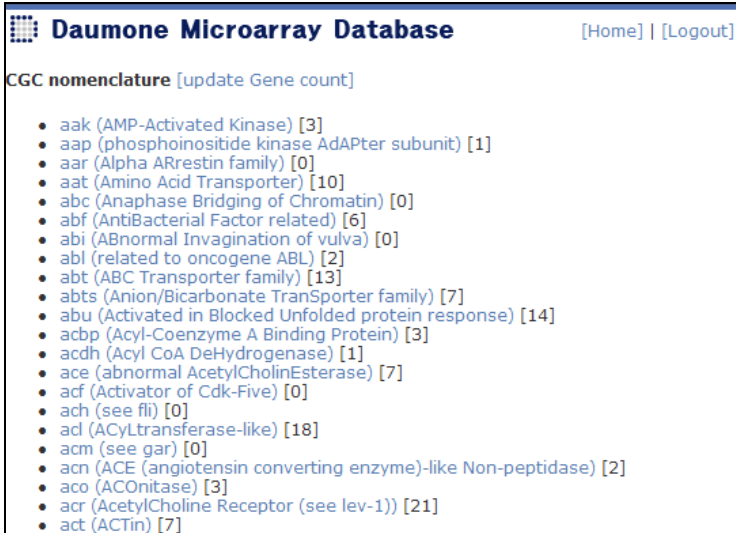
Dauer Gene DB has various KEGG pathways that are linked to microarray data. In figure 2-8, the green box is the enzyme which is located in *C.elegans*. Click this green box, you can see those liked microarray data.
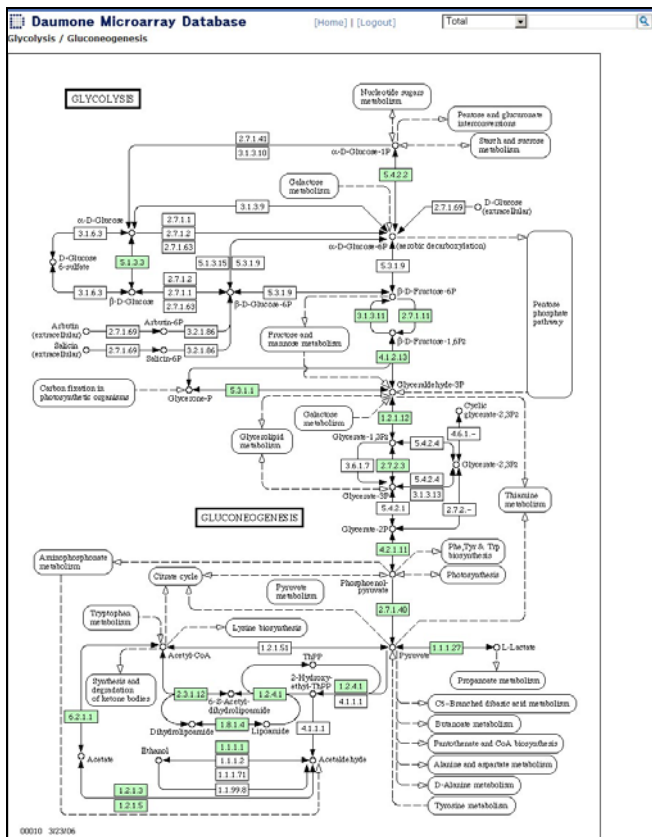
Figure 2-8. Glycolysis pathway from KEGG map. The green box is the enzyme located in *C. elegans*.

2-5. Browsing by CGC list

Daumone microarray database has 1,594 CGC (Caenorhabditis Genetics Center) gene list. If you are familiar with CGC nomenclature, you can browse the "CGC list" and find the gene of interest. [2]

Figure 2-9. CGC nomenclature gene list. Number in parenthesis indicates the total number of gene expressed.

2-6. Browsing by Dauer Comparison

To compare those gene expressed in the dauer larvae that were induced by daumone (6-groups) and the dauer larvae that were induced by long-term broth culture (4-groups).
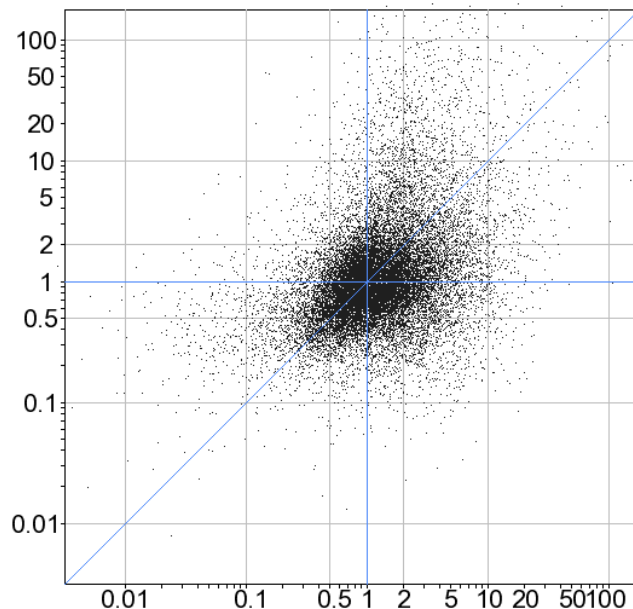
Figure 2-10. The 2-D distribution graph shows dauer comparison. X-axis indicates the expression ratio of $Log_{10}(Da[S6]/S1)$. Y-axis is the expression ratio of Log_{10} (Da[L]/S1).

Da[S6] : dauer larvae induced by daumone at S6, Da[L] : dauer larvae induced by broth culture.

S1 is the sample taken at 12h after L1 synchronized worms which were placed on the daumone plate.

Figure 2-11. GO classification of dauer comparison.

In dauer comparison, we focus on 4 categories.

- **1 stage < Dauer, 1 stage < 6 stage :** Gene expression of Da[L] that is induced (more than 2 fold) and those dauer larvae (Da[S6]) grown on the daumone plate that is induced.

- **1 stage > Dauer, 1 stage < 6 stage :** Gene expression of Da[L] that is suppressed (more than 2 fold) and Da[S6] that is induced.

- **1 stage < Dauer, 1 stage > 6 stage :** Gene expression of Da[L] that is induced (more than 2 fold) and Da[S6] that is suppressed.

- **1 stage > Dauer, 1 stage > 6 stage :** Gene expression of Da[L] that is suppressed (more than 2 fold) and Da[S6] that is suppressed.

3. Detail Microarray Information

Detailed view of one gene expression data contains gene abstract information, normalized or intact expression data, genechip information, functional annotation and DNA sequence.
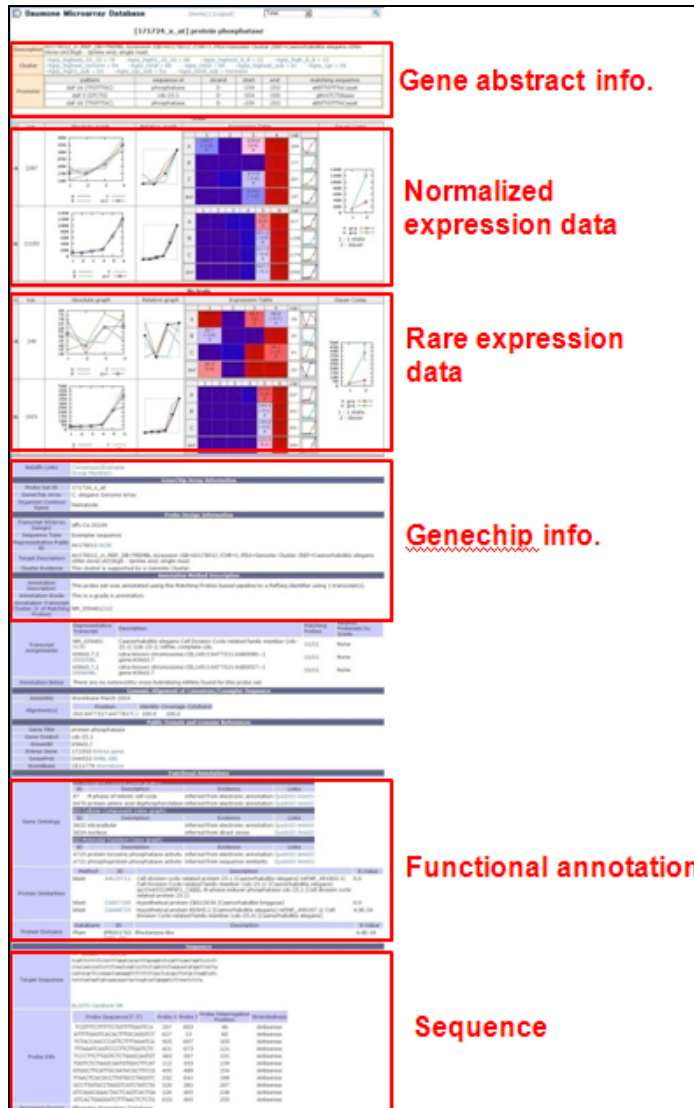

Figure 3-1. Detail expression information of a single gene.

Detailed view of expression data includes both the normalized data and rare expression data to estimate the effect of normalization. Normalized data is on "Scale" part and rare data is on "No Scale".

3-1. General Information

	[171724_x_at] protein phosphatase					
Description	AV178012_rc /REP_DB=TREMBL Accession /GB=AV178012 /CHR=1 /FEA=Genomic Cluster /DEF=Caenorhabditis elegans cDNA clone:yk536g8 : 3prime end, single read.					
Cluster	·6grp_highest_10_10 > 78 ·6grp_high1_10_10 > 88 ·4grp_highest_8_8 > 21 ·4grp_high_8_8 > 15 ·6grp_highest_nonorm > 94 ·6grp_total > 88 ·4grp_total > 08 ·6grp_highest_sub > In ·6grp_cgc > 59 ·6grp_high1_sub > Da ·6grp_cgc_sub > Da ·6grp_total_sub > Increase					

	pattern	sequence id	strand	start	end	matching sequence
Promoter	daf-16 (TTGTTTAC)	phosphatase	D	-259	-252	atttTTGTTTACaaat
	daf-3 (GTCTG)	cdc-25.1	D	-554	-550	gttcGTCTGtaaa
	daf-16 (TTGTTTAC)	phosphatase	D	-259	-252	atttTTGTTTACaaat

Figure 3-2. General information of gene expression data

- **Description :** Gene description with accession number,
- **Cluster :** SOM cluster which this gene belongs to

  For more information, refer to "3-3. Browsing via SOM clustering"
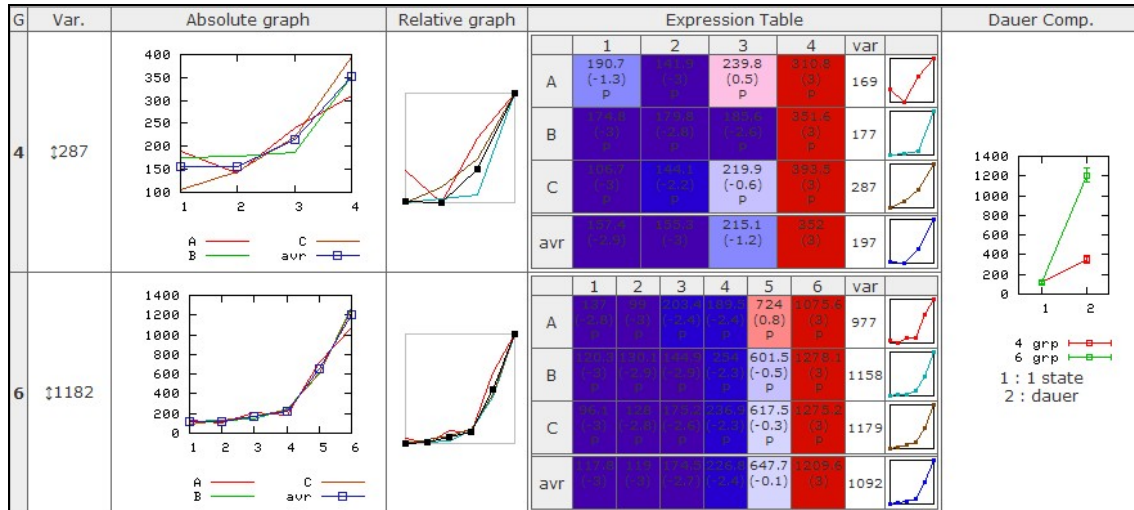
3-2. Expression Data

Figure 3-3. Expression part on detail view.

- **Group (G)** : Sample data group. (4 or 6 group)
- **Var. (Variance)** : Var. = Max (expression data) – Min (expression data)
- **Absolute graph** : graph by expression value.
- **Relative graph** : graph by the relative expression data. Relative expression data is generated by normalization of expression data.
- **Expression table** : Table for triple set (A, B, and C) and time series. (4 or 6 group)
- **Dauer comparison graph** : Graph for comparing Da[L] from broth culture and Da[S6] from daumone plate.

4. Searching Expression Data

4.1 Searching through Detailed Information

Insert searching term in searching textbox to search through detailed information. You can confine searching field to select option bar.
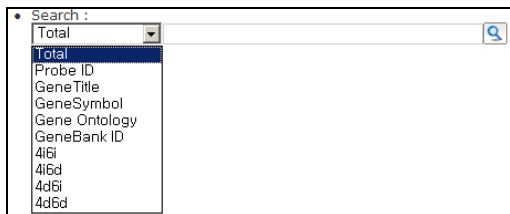
Figure 4-1. Searching field and searching bar.

Searching field option

- **Total :** total expression data list.
- **Probe ID :** Probe ID is the unique identification number of microarray chip.

  ex) 171724_x_at , 186744_at, 176986_s_at and 177008_s_at
- **Gene Title :** CGC gene name

  ex) daf-28, daf-3 and vit-4
- Gene Symbol :
- **Gene Ontology :** Gene ontology
- **GeneBank ID :** Identification number from GeneBank
- **4i6i :** 4-group induced and 6-group induced gene list in dauer comparison
- **4i6d :** 4-group increase and 6-group suppressed gene list in dauer comparison
- **4d6i :** 4-group suppressed and 6-group induced gene list in dauer comparison
- **4d6d :** 4-group suppressed and 6-group suppressed gene list in dauer comparison

Reference

[1] M. Reich , K. Ohm , M. Angelo , P. Tamayo and J. P. Mesirov (2004) Bioinformatics 20(11), GeneCluster 2.0: an advanced toolset for bioarray analysis

[2] CGC nomenclature : http://biosci.umn.edu/CGC/CGChomepage.htm