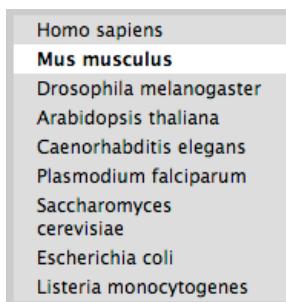


GOme Instructions for Users

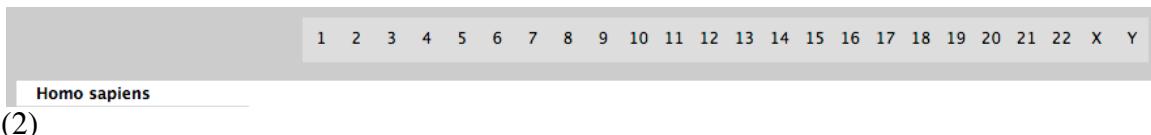
The screenshot shows a user interface for selecting an organism and a chromosome. On the left, a list of organisms is shown, with **Homo sapiens** selected. To the right of the list is a dropdown menu for choosing a chromosome, currently set to "window 5". Further right are three radio button options: "most commonly appearing GO terms" (selected), "most clustered GO terms (p-value 10^{-6})", and "custom: adjust p-value". Below these options is a text input field for "GO number search" containing the value "5882", with "Submit" and "Erase" buttons nearby. At the top of the interface is a row of numbered buttons from 1 to 22, followed by "X" and "Y".

The Gome database is a browsable repository for gene ontology clusters as depicted in the Clustering of Gene Ontology Terms in Genomes manuscript by Tiirikka et al. (submitted). One can choose the organism from a drop-down menu (1) e.g *Homo sapiens* and specify the desired chromosome by clicking at it on grey box on top (2).



(1)

The numbers and letters express the chromosome set of the chosen organism from the left white box.



(2)

It is also possible to use a fixed limit for p-values or instruct the database to only look for most common terms (sorted by frequency) or just have the listing of the most clustered (the statistically most significant) terms. (3)

most commonly appearing GO terms
 most clustered GO terms (p -value 10^{-6})
 custom: adjust p -value

(3)

The results are sortable by clicking the column name e.g. P-value; this sorts the data either in ascending or descending manor (ASC (ascending) / DESC (descending)). (4)

Results 1 – 20 of 48 (Page 1/3) for <i>Caenorhabditis elegans</i> : window 5 – Most commonly appearing GO terms					
Chromosome	Start	End	P-value	Ontology	Name
ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC
X	6323253	6377643	0.000000111537	40032	"The process by which the anatomical structures of the post-embryonic soma are generated and organized. Morphogenesis pertains to the creation of form." [GOC:ems, ISBN:0140512888]
2	11527148	11542412	0.0000014279	9790	"The process whose specific outcome is the progression of an embryo from zygote formation until the end of its embryonic life stage. The end of the embryonic stage is organism-specific and may be somewhat arbitrary. For example, it would be at birth"
X	6279632	6332567	0.0000152119	40032	"The process by which the anatomical structures of the post-embryonic soma are generated and organized. Morphogenesis pertains to the creation of form." [GOC:ems, ISBN:0140512888]

(4)

The columns in the results are for

- Chromosome
- Window start position
- Window end position
- P-value
- GO code having a hyperlink to AmiGO service)
- The definition of the GO term.

Results are organized to pages. The user can select how many rows are shown on the display. (5)

2	45040	71279	0.000852805	40007	"The increase in size or mass of an entire organism, a part of an organism or a cell." [GOC:bf, GOC:ma]
2	9955098	9981660	0.000878962	40035	"The process whose specific outcome is the progression of the hermaphrodite genitalia over time, from formation to the mature structures." [GOC:ems, ISBN:0140512888]
First Page	Prev	1 2 3		Next	Last Page

(5)

When one wants to establish a fixed p-value threshold it can be done using the “custom: adjust p-value” radio button. When this button is clicked a query box with title “Adjust p-value” pops up to the right hand side. After setting the threshold value the query will be executed by clicking the submit button. (6)

The screenshot shows a search interface for biological data. At the top, there is a navigation bar with links labeled 1 through 16. Below this is a list of organisms: Homo sapiens, Mus musculus, Drosophila melanogaster, Arabidopsis thaliana, Caenorhabditis elegans, Plasmodium falciparum, Saccharomyces cerevisiae, Escherichia coli, and Listeria monocytogenes. A dropdown menu next to the organism list shows options for window sizes: window 5, window 10, window 15, window 20, window 25, window 30, window 35, and window 40. The "window 5" option is selected. To the right of the dropdown are three radio buttons: "most commonly appearing GO terms", "most clustered GO terms (p-value 10^{-6})", and "custom: adjust p-value". The "custom: adjust p-value" option is selected. Below these buttons is a text input field for "GO number search:" and a text input field for "Adjust p-value:" containing the value "0.0003". There are also "Submit" and "Erase" buttons. The results section below the search controls displays the following table:

Results 1 – 20 of 138 (Page 1/7) for Saccharomyces cerevisiae : window 5 – Custom p-value: p-value ≤ 0.0003					
Chromosome	Start	End	P-value	Ontology	Name
ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC	ASC / DESC
10	472675	486111	0.000041132	5515	"Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules)." [GOC:g]
10	472675	490727	0.000041132	5515	"Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules)." [GOC:g]
10	472675	486111	0.00000572822	3723	"Interacting selectively with an RNA molecule or a portion thereof." [GOC:mah]

(6)

If one is interested only in certain GO term it can be searched using the “GO number search:” box. (7) In this case the GO:0000000 syntax of 7 digits and “GO:” prefix is not necessary. User can only input the non-zero part of the GO code like 5882 for GO:0005882.

Chromosome	Start	End	P-value	Ontology	Name
ASC / DESC			ASC / DESC	ASC / DESC	ASC / DESC
21	44699835	44899004	0.00000105586	5882	intermediate filament
21	44835577	45118169	0.00000105586	5882	intermediate filament
21	44844925	45130571	0.00000105586	5882	intermediate filament
21	44856424	45165232	0.00000105586	5882	intermediate filament
21	44871468	45179316	0.00000105586	5882	intermediate filament

(7)

The flat files contain information on the organisms mentioned on the manuscript (8).

The flatfiles from the database are downloadable from these links [human](#), [fruit fly](#), [roundworm](#), [E. coli](#), [mouse](#), [yeast](#), [rockcress](#)

(8)

The syntax of the flat files is:

CHROMOSOME : {CONTIG : GENE NAME: GENE START POSITION : GENE END POSITION} : K : k : N : n : E-VALUE : P-VALUE : GO CODE : GO NAME

The parameters are defined in the article and {} part is repeated for the other genes in the window .

An example from human window 5.

```
10 NT_008705.15 219749 38278801 38300744 NT_008705.15 7581
38345803 38388310 NT_008705.15 389950 38389706 38397559
NT_008705.15 340947 38408990 38423144 NT_008705.15 7587
38423280 38452273 K 98 k 3 N 983 n 5 0.498474059003052
0.00828939 5634 nucleus
```