PosMed (Positional Medline) User Manual

RIKEN Bioinformatics And Systems Engineering (BASE) division
Division Director Tetsuro Toyoda

Index

The URL of this website is http://omicspace.riken.jp

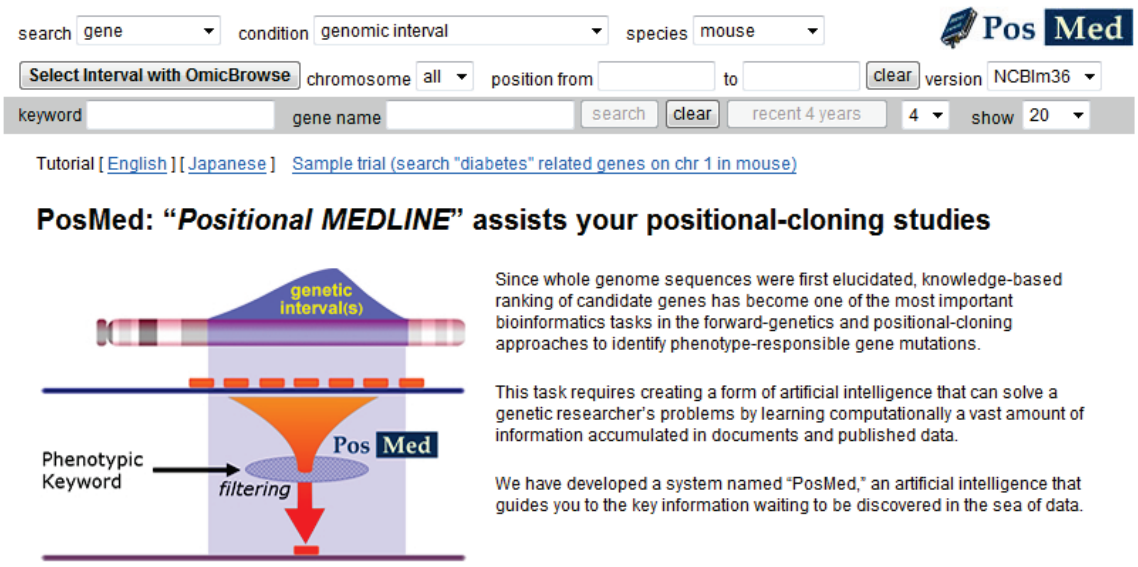
If you have any questions and comments, please feel free to send to
omicspace@gmail.com

January 5th, 2009

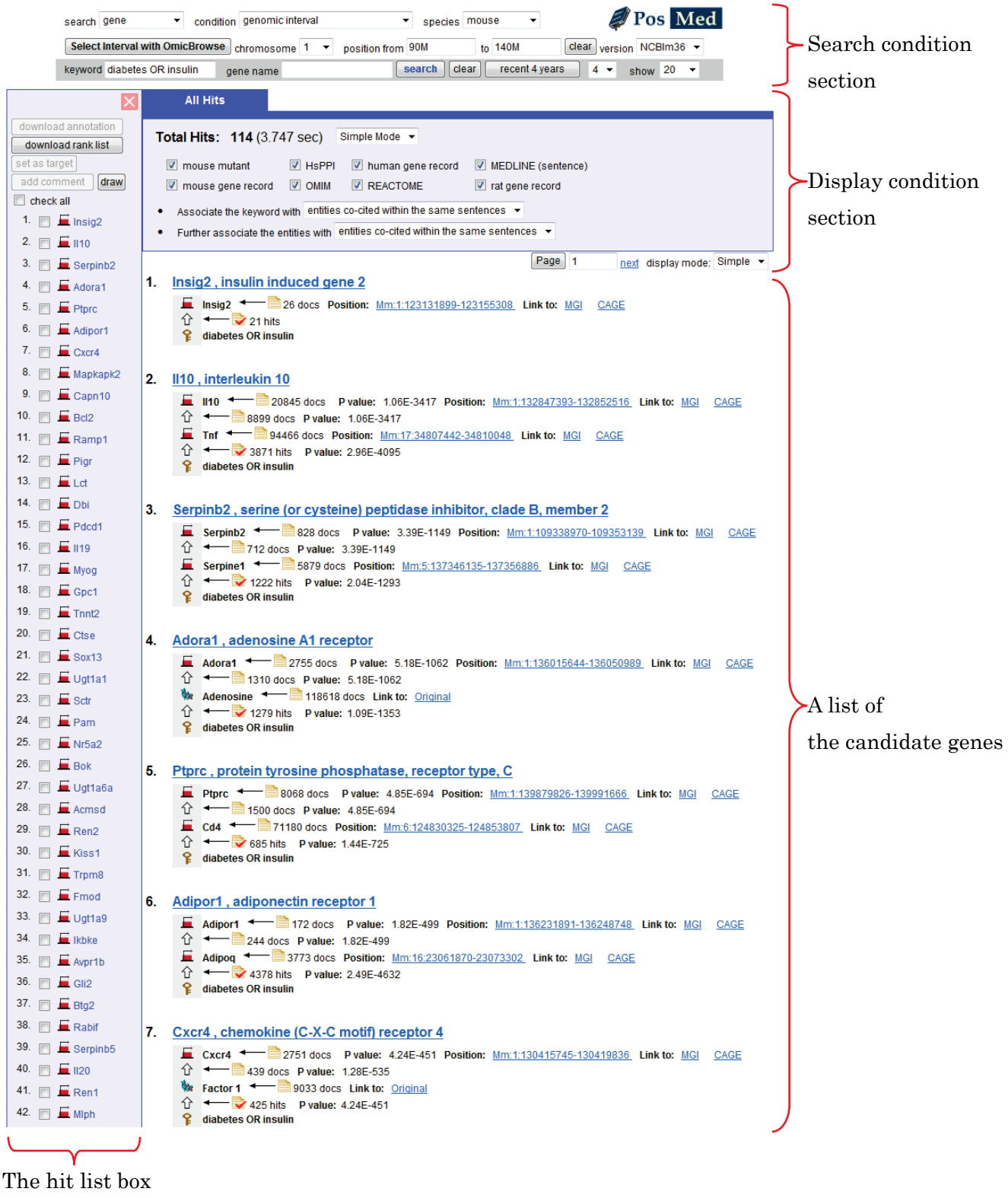
1. Introduction

PosMed is RIKEN BASE's original system which possesses inference-type full text search functions. At first, PosMed performs a full-text search of the document database (containing MEDLINE abstracts etc) by using arbitrary keywords users have entered. And second, PosMed finds significant gene names or symbols that exist within the returned documents with Fisher's exact test and makes a ranked list (in this case, significant means that the P-value is as small as possible.) In the case where a user specifies the chromosomal interval, PosMed makes and displays a ranked list of genes restricted to the given interval. (That is to say, PosMed can search the connection of **Keyword -> Gene A -> Chromosomal interval**.)

Simultaneously, PosMed can also search for appropriate genes within the chromosomal interval that have an estimated gene-gene interaction, automatically. The selected genes will fulfill the conditions of network connections of **Keyword -> Gene B -> Gene C -> Chromosomal interval**. PosMed is a powerful tool which enables us to connect phenotypic functions and genes, because it uses relevant information of not only gene-gene interactions but also other biological objects' interactions (cell-gene, metabolite-gene, mutant mouse-gene, drug-gene, disease-gene, protein-protein interaction and ortholog data.) PosMed enables us to search human and mouse genes, drugs, diseases and document sets (*e.g.* MEDLINE and OMIM.)



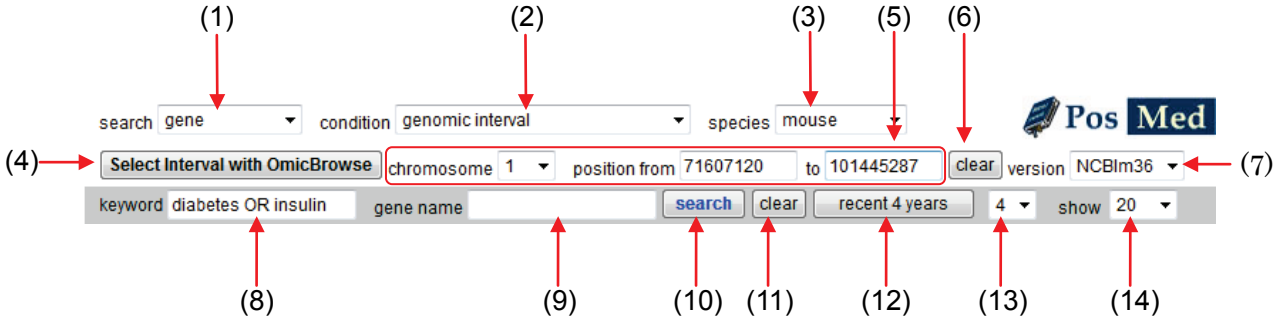
The search result page of PosMed



The maximum number of hit entries is 300 in the hit list box.

2. Specify the search conditions

The search condition section has some pull-down menus and entry boxes as below.



(1) Search target pull-down menu

The menu items are any, gene, metabolite, drug, cell, mutant, disease, document set, dictionary and database. According to the search target, each search tool shown previously will be displayed.

(2) Search condition pull-down menu

Available conditions are "gene name," "genomic interval," "multiple intervals," "genomic interval (band)," "multiple intervals (band)," "accessions (*e.g.* MGI:107357, HGNC:6079)," "multiple accessions" and "genomic interval and accessions."

(3) Species pull-down menu

The menus are any, mouse and human. (For this manual the "species" we use is "*M.musculus*.")

(4) Select Interval with OmicBrowse button

At the stage of (2), when the menu condition is selected as "genomic interval" or "multiple intervals," this button is displayed with (5). Click this to display OmicBrowse in the PosMed system.

(5) - [1]. Chromosome pull-down menu and base pair entry boxes

- [2]. Genomic band entry box
- [3]. Accessions entry box

At the stage of (2), when the menu condition is selected as "genomic interval" or "multiple intervals," this part is displayed as [1] to specify genomic interval(s). In the case of "genomic interval (band)" or "multiple intervals (band)," this part is displayed as [2] to enter band name (*e.g.* 20p13). In the case of "accession" or "multiple accessions," this part is displayed as [3] to enter accessions. And also in the case of "genomic interval and accessions," both [1] and [2] are displayed at the

same time.

(6) Clear button to reset chromosome pull-down menu and erase the base pair entry boxes

(7) Pull-down menu specifying a genome version. (For this manual the "genome version" we use is "NCBIm36.")

(8) Keyword entry box

PosMed enables us to specify more complex conditional equations using "OR (caps)", "AND (caps)", parentheses, double quotation marks for phrases, wild card (asterisk and question mark), "NOT (caps)" and "WHERE (caps)."

Keyword format in the Genome Browser & PosMed system

Keyword format	**Genome Browser**	**PosMed**
Union operator	OR (caps)	
Set intersection operator	AND (caps)	
Parenthesis expression	Available	
Phrase specification	group with " " *e.g.* "retinitis pigmentosa"	
Treatment of space	as "AND" operator	
Wild card	* (asterisk) *e.g.* diabet* matches *diabetes, diabetic,* etc.	
	? (question mark) *e.g.* ppar? matches *ppara, pparg, ppard,* etc	
Boolean operator	NOT (caps) *e.g.* diabetes NOT "type 1"	
Subquery	―	WHERE (caps)# *e.g.* plant WHERE synapse

#Note : PosMed searches mouse genes by "plant" and "synapse." The search result of "plant" is displayed preferentially.

(9) Gene name entry box to specify as determinate targets.

e.g. Irs1, "insulin receptor substrate 1"

(10) Execute button to search

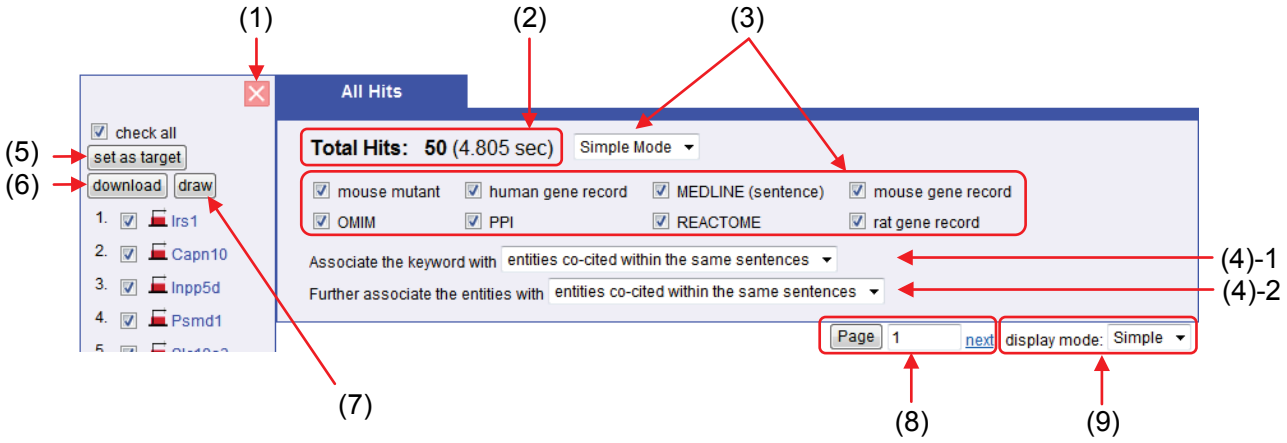
(11) Clear button to erase the keyword entry box and the gene name entry box.

(12) Execute button to search with the specified number of years. The default number is set to 4. This button is linked with the “recent years pull-down menu” as below.

(13) Recent years pull-down menu (Available years are from 2 to 9)

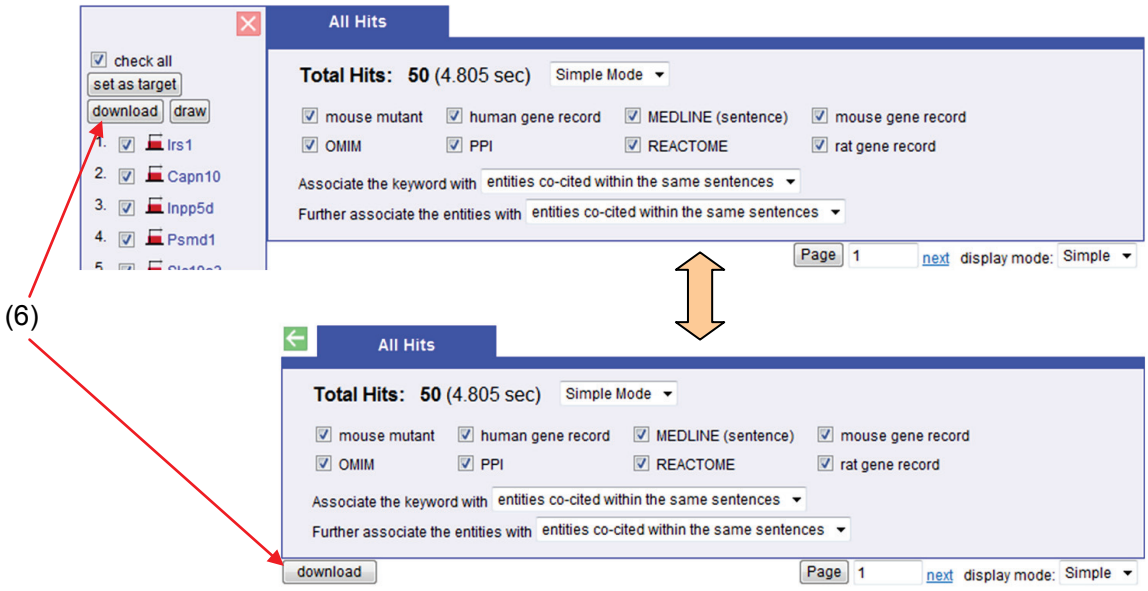
(14) Show pull-down menu (Available numbers are 5, 10, 20, 50, 100, 200 and 500)

3. Specify the display conditions



(1) Open all hits button

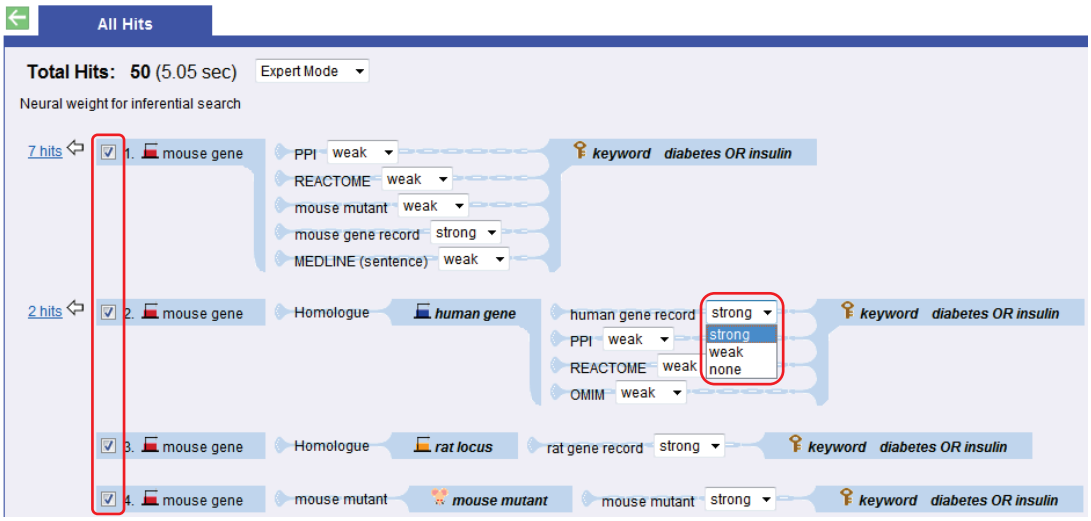
Click this button to show or close a list of genes. If you check some of the hits, you can set them as the next search targets by clicking “set as target” and download them as an excel file by clicking “download.”



(2) Number of hits and time taken for the search

(3) Search path display mode pull-down menu and path list

- Simple Mode : Users can check document sets for inference search process.
- Expert Mode : The display condition part expands with all search path with neural weight pull-down menu. Users can check path and select neural weight for inferential search process.



(4) Association pull-down menu

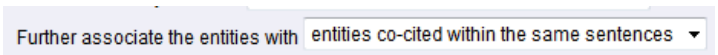
As above, PosMed can search the connection of **Keyword -> Gene A -> Gene B -> Chromosomal interval**. User can select the association status between the keyword and a single entity (gene A etc.) or two entities. The association status has can be set to either “co-cited within the same sentences” or “documents”.

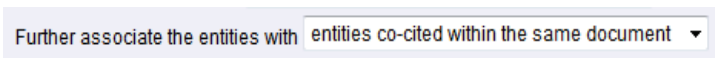
(4)-1 represents the association status between the keyword and Gene A.

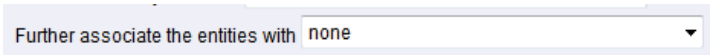
The default is set to the association of “entities co-cited within the same sentences.”

(4)-2 represents the association status between Gene A and Gene B.

The default is set to the association of “none.”



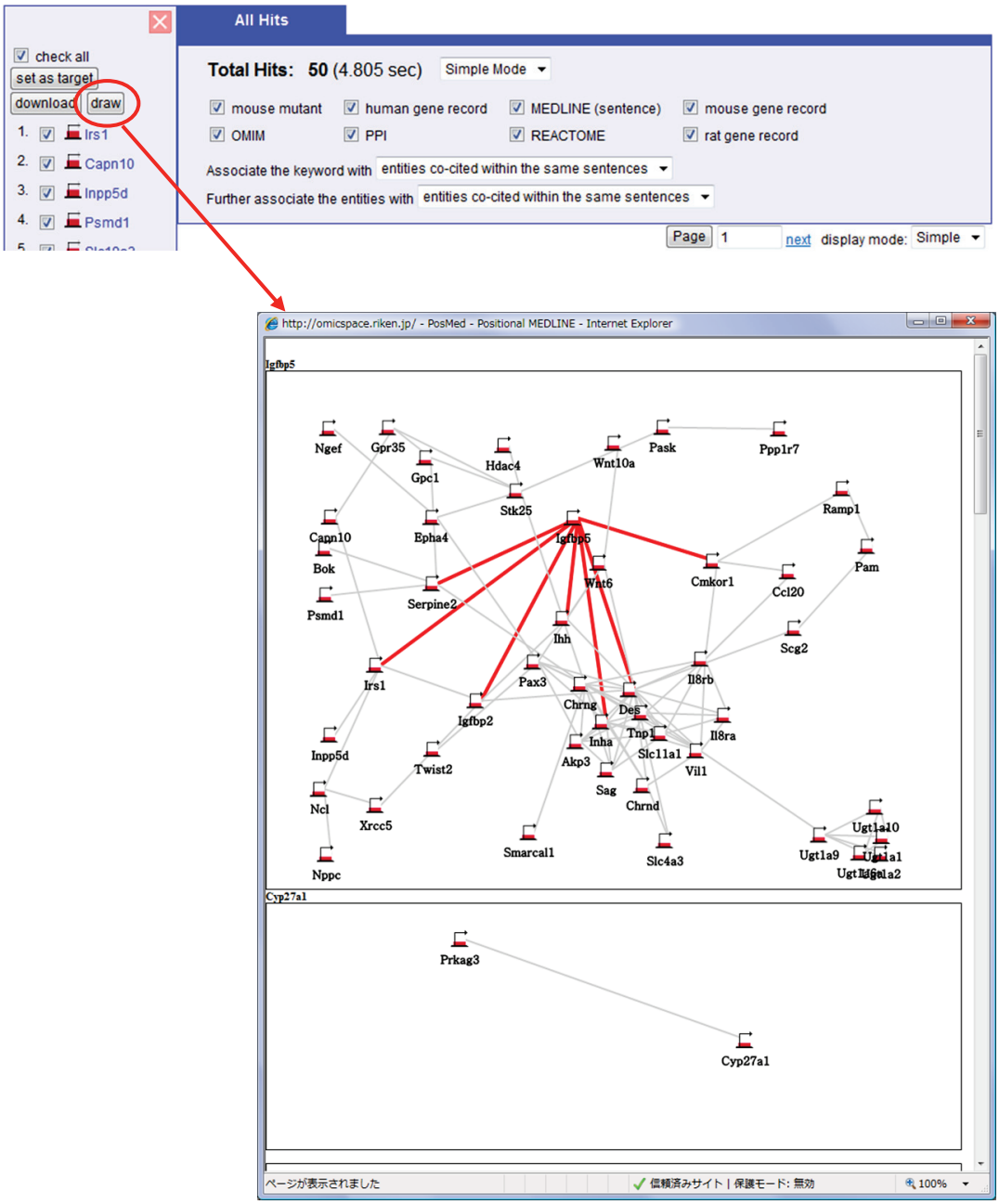




(5) Set as target button to set them as the next search targets.

(6) Download button to download all hits as an Excel file.

(7) Draw button to draw the co-citation networks of hit entries in another window.



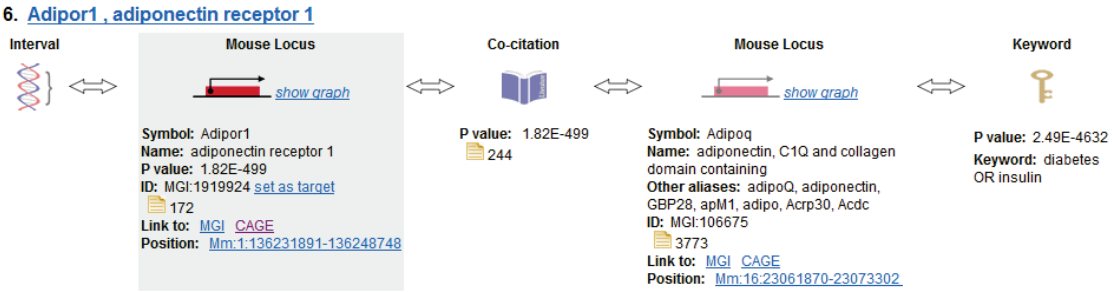
(8) Page entry box and link to next (previous) page

(9) Search result display mode (mentioned in next section)

- "Simple" display mode shows a compact list of the search results with minimum information. Each result indicates the search path from keyword (bottom) to gene (top).

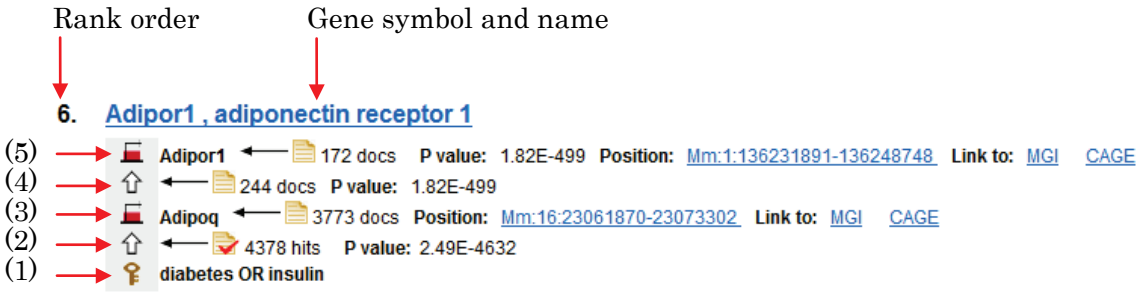
- "Tree" display mode shows a compact list of the search results in a hierarchical structure with minimum information similar to the "Simple mode."

- "Graph" display mode shows a list of the search result with more information. Each result indicates the search path from keyword (right) to gene (left).



4. Understand the relationship between the candidate genes and the keyword

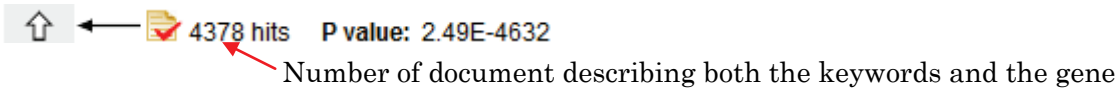
4-1. Relationship displayed in the “Simple mode”



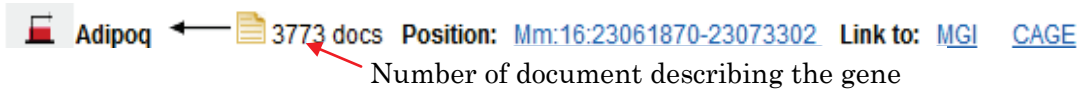
The order from (1) to (5) indicates the search path from keywords (bottom) to genes (top).

(1) A keyword icon indicates the keyword(s) entered in the keyword entry box previously.

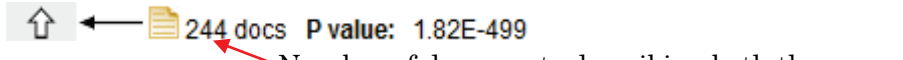
(2) An arrow icon indicates the relationship between the keyword(s) and a gene with a p value.



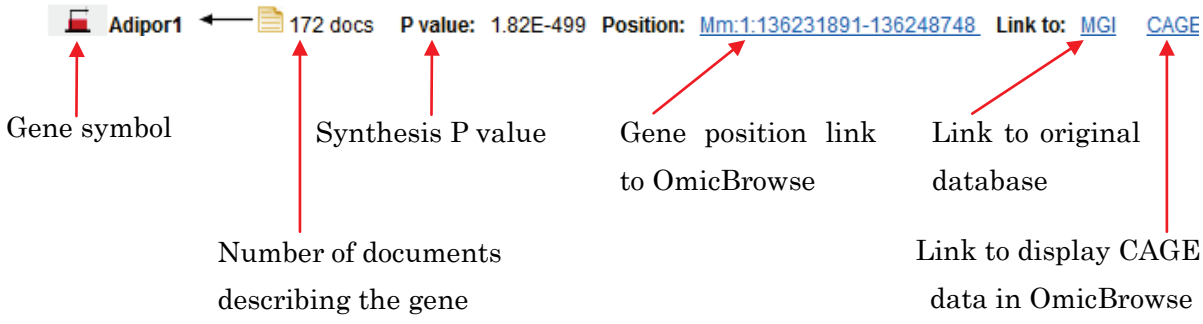
(3) A Mouse gene icon indicates an intermediate path has relationships with both the keyword(s) and gene.



(4) An arrow icon indicates relationships between two mouse genes.



(5) A Mouse gene icon indicates a gene within the interval.

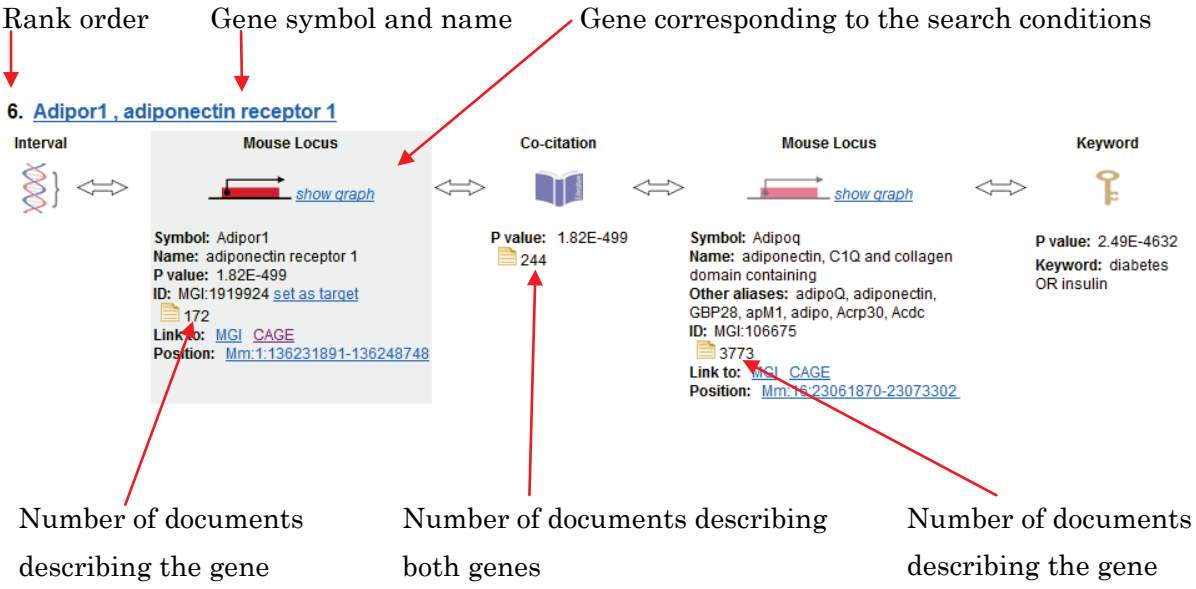


4-2. Relationship displayed in the “Tree mode”

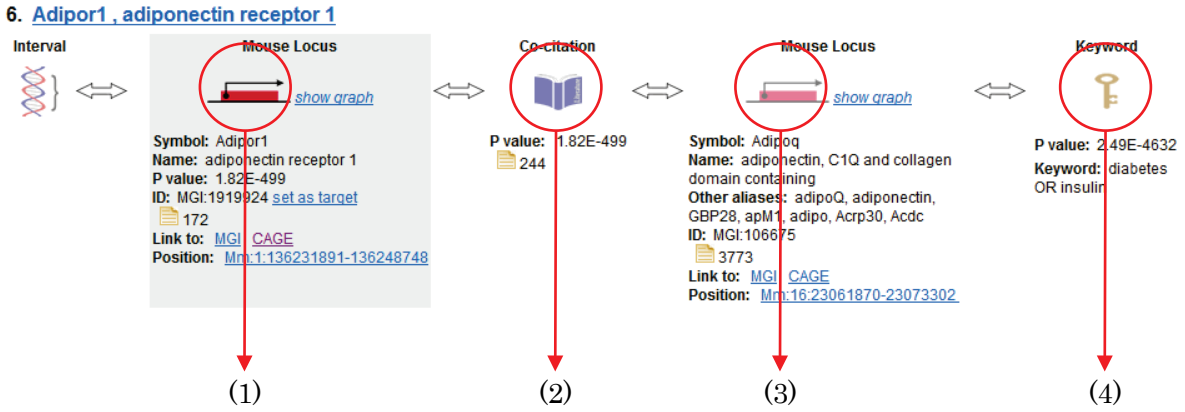
In the “Tree mode”, the hierarchy structure indicates the search path from keywords (bottom) to genes (top) in much the same way as the “Simple mode.”

4-3. Relationship displayed in the “Graph mode”

In the “Graph mode”, each result indicates the search path from keywords (right) to genes (left).

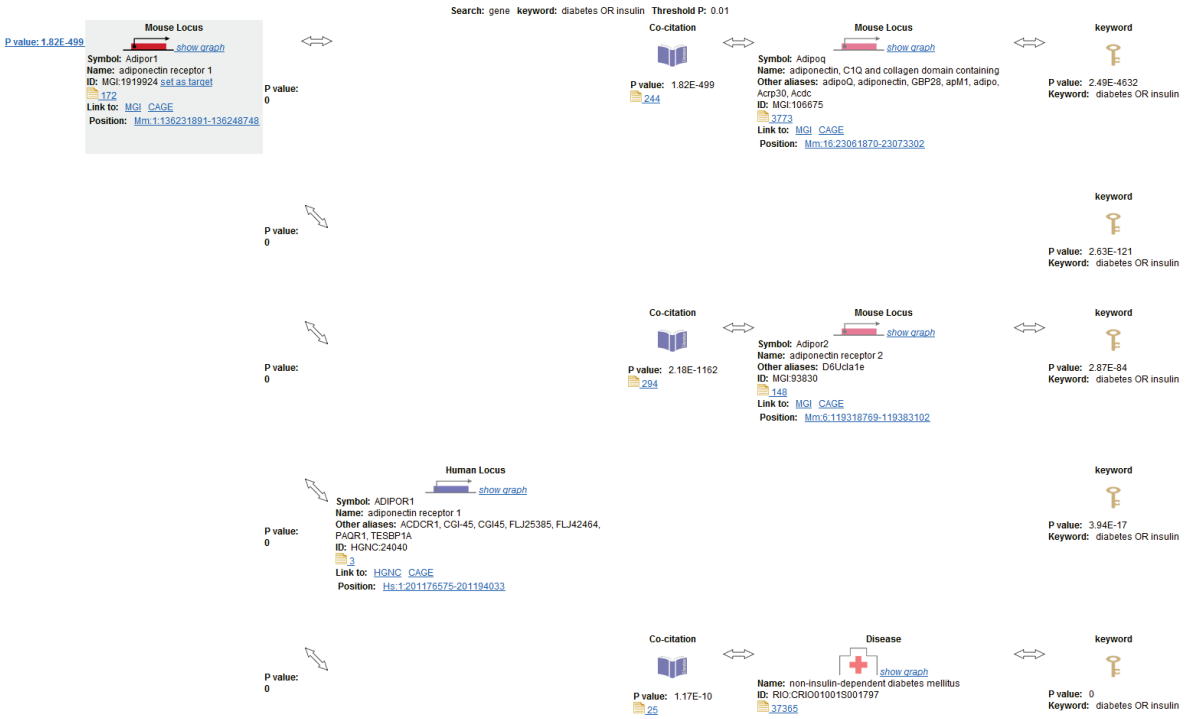


5. Browsing the details of hit results



5-1. Relationships between the gene and the keywords

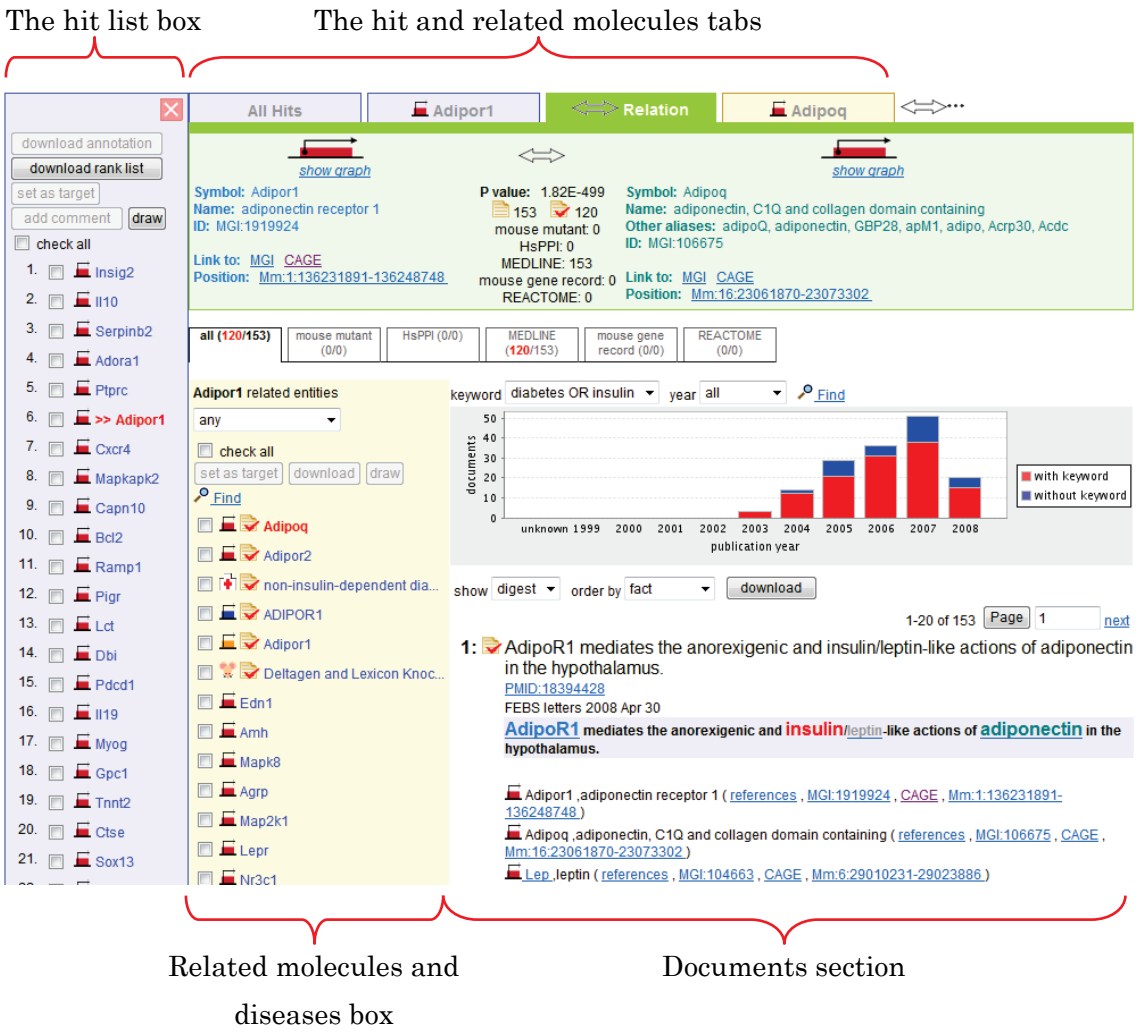
When you click (1), the gene, various inference relationships from the gene to the corresponding biological objects are displayed.



5-2. Relationship between the gene and another gene

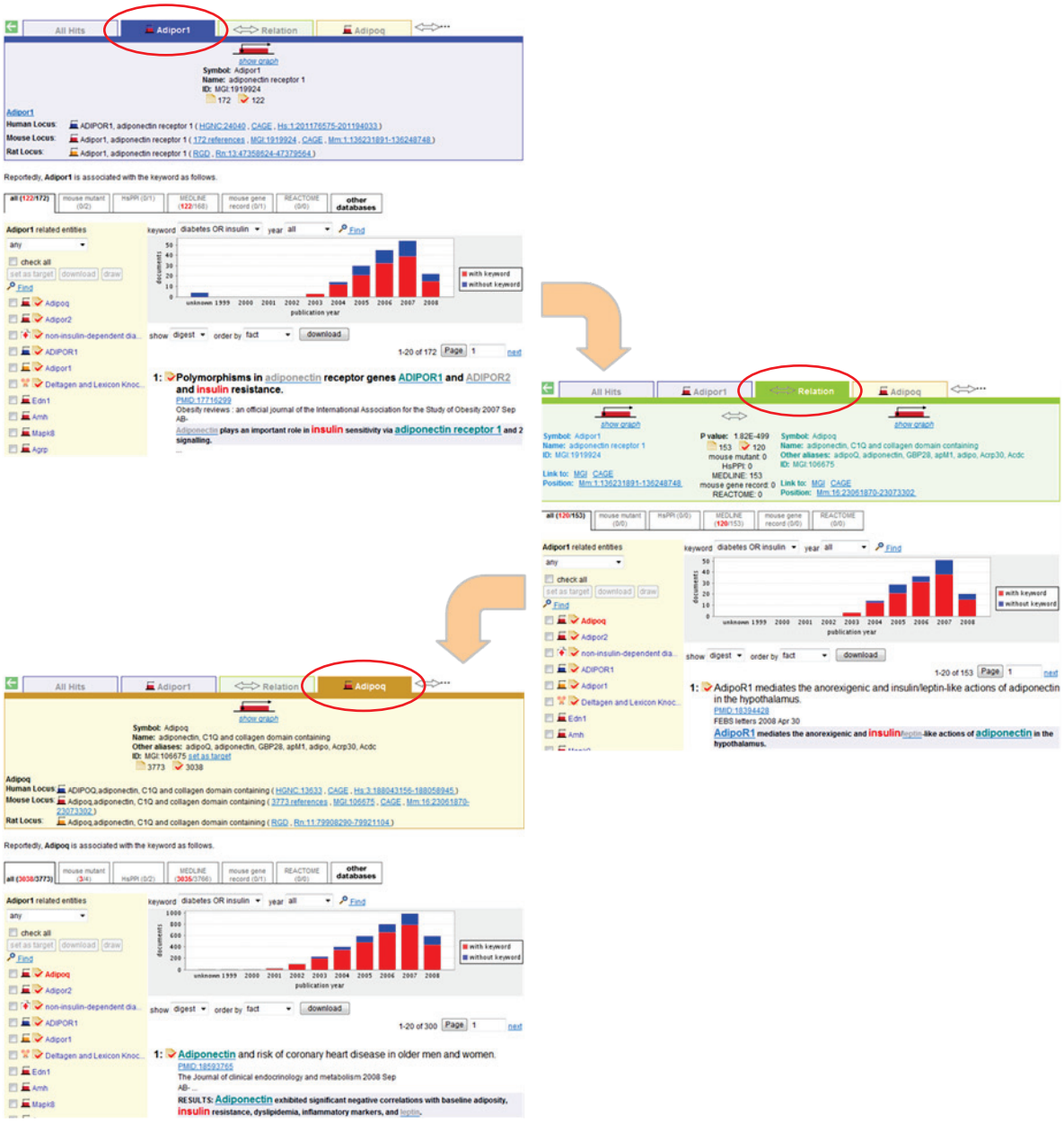
When you click (2), the relationship, document digests that include both the genes are displayed.

In the document digest display, the gene names and the keywords are highlighted with light blue (Adipor1 etc.), light green (Adipoq etc.) and red (diabetes or insulin) respectively. Additional biological objects are displayed in gray. In the default display sequence, the documents including the keyword(s) are ranked higher. Click a document number in order to transfer to PubMed at NCBI.



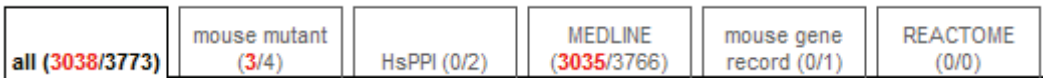
5-2-1. Hit and related molecules tabs

These tabs display the descriptions of hit molecules (Adipor1 etc.), related molecules (Adipoq etc) and their relationships. Click a tab to display each description and change information of the “related molecules and disease box” and the “documents section” respectively.

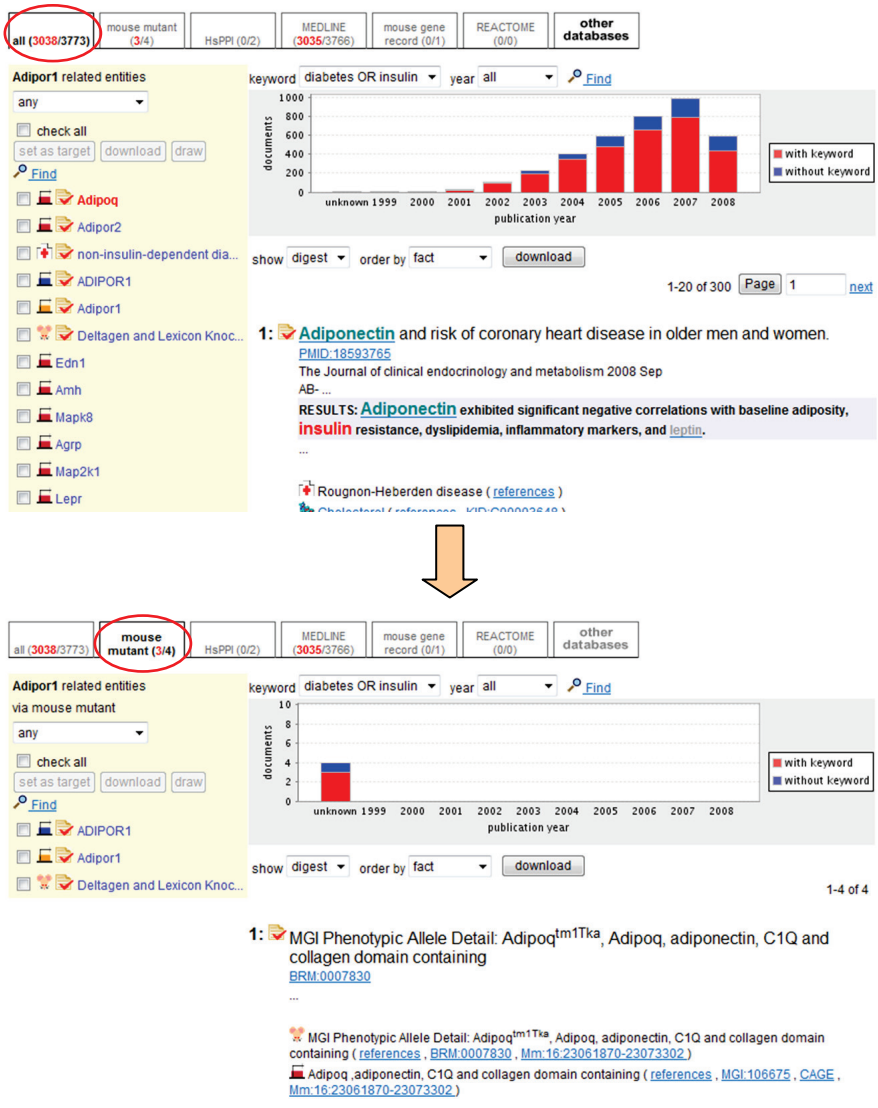


5-2-2. Hit document tabs

These tabs indicate the number of hit documents in each document set. In the case of mouse, the document sets for the inference search process are mutant information, MEDLINE, gene description of MGI, PPI description of the HPRD and REACTOME description of BioPAX. The number of hit documents which contain the keywords are highlighted in red. The default setting is all.

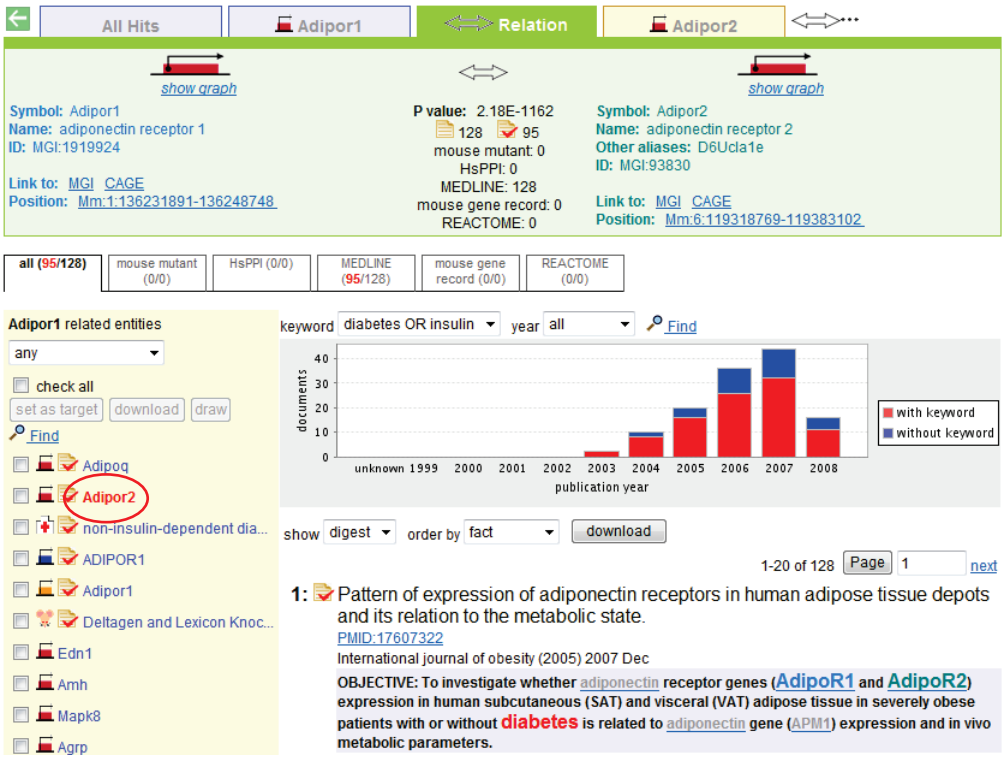


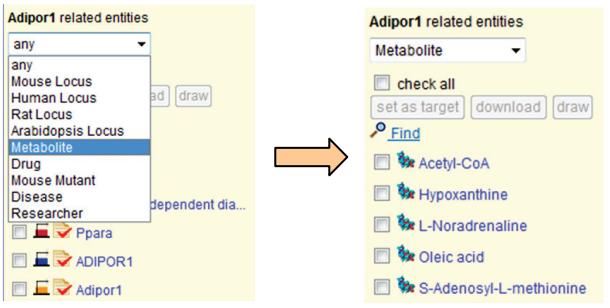
Click a tab to display the corresponding document digests as described above.
(*e.g.* The transition from "all" to "mouse mutant.")



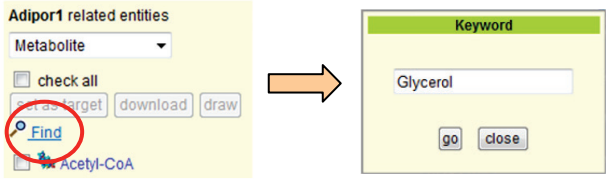
5-2-3. Related molecules and diseases box

Other related molecules or diseases are listed as images in the left yellow box. If the documents contain the keywords, the related objects will be displayed with hit images . You can click a name (*e.g.* Adipor2) to display the relationships, document digests that include the keywords (diabetes or insulin), the gene name (Adipor1 etc.) and the biological object (*e.g.* Adipor2 etc.) in the documents section. This function is the same as the "Hit and related molecules tabs" described previously.





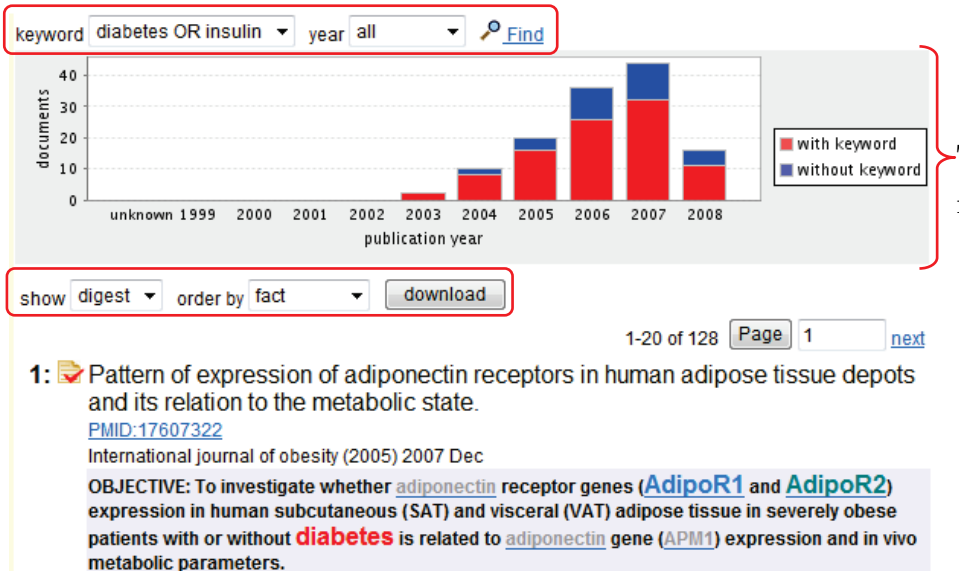
Top fifty related molecules and diseases are displayed as biological objects initially. You can select a particular condition from the pull down menu. (*e.g.* Metabolite)



Click the "Find" image to display the keyword search box. Click the go button to find a biological object of interest in this box. (*e.g.* Glycerol)

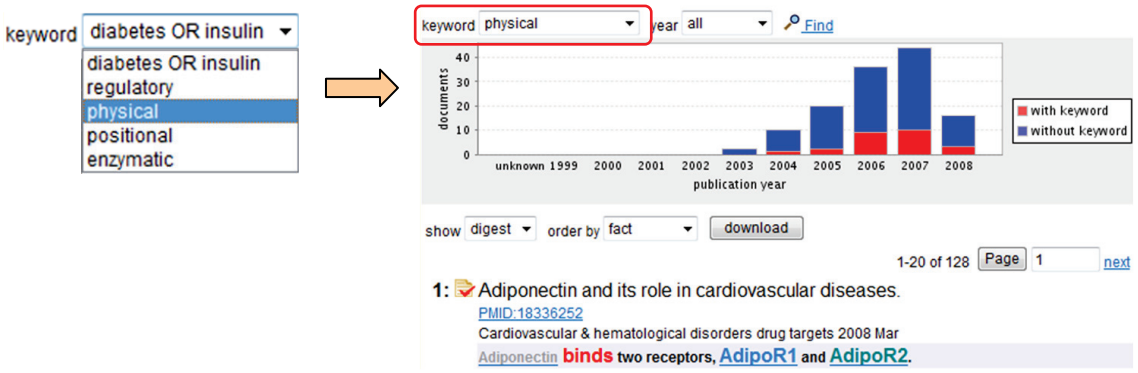
5-2-4. Documents section

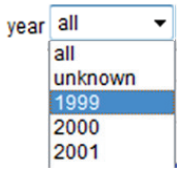
The documents section has three convenient functions to find desired documents. The functions are the “show” pull-down menu, “order by” pull-down menu, “year” pull-down menu and “Find.”



The charts indicate the number of documents.

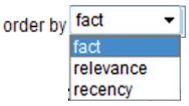
The “keyword” pull-down menu enables us to change the display sequence of the hit documents by keyword(s) or biological terms. The default is the previously entered keyword(s) in this case “diabetes OR insulin.” Select “physical” to superordinate documents including physical terms (*e.g.* binds) and highlight these terms.





The published year of documents is narrowed down by the “year” pull-down menu. You can select years from 1999 to the current year or any with unknown published years. The default setting is “all.”

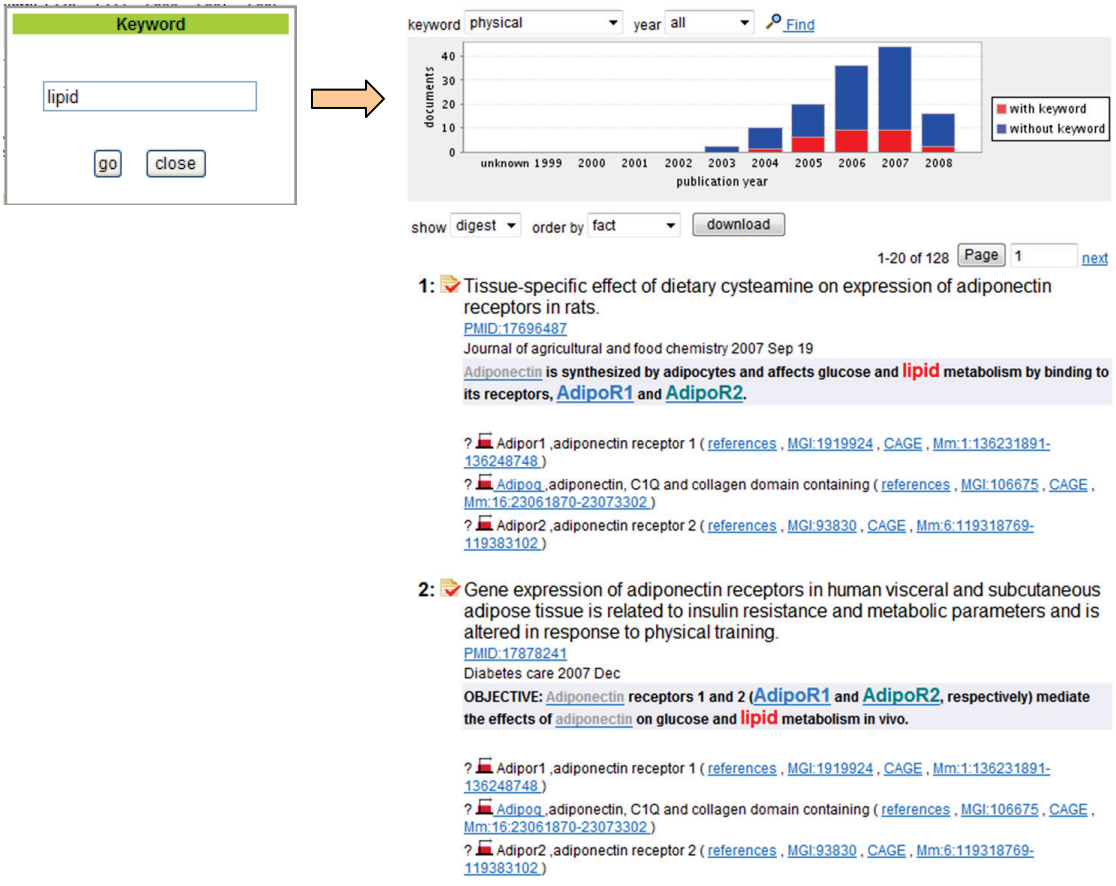
The “show” pull-down menu has two modes: the default is the digest mode to display the minimum sentences including keyword(s) and biological objects, and the other one is the all mode to display whole text.



The “order by” pull-down menu enables us to change the display sequence of the hit documents by biological objects, keyword(s) or their recentness. The default is the “fact” mode to display sentences including previously selected biological objects, in this case AdipoR1 (Adipor1) and the biological object AdipoR2 (Adipor2). Select “relevance” to superordinate documents including keywords (diabetes or insulin) and highlight these terms. The “recency” mode superordinates the latest sentences.

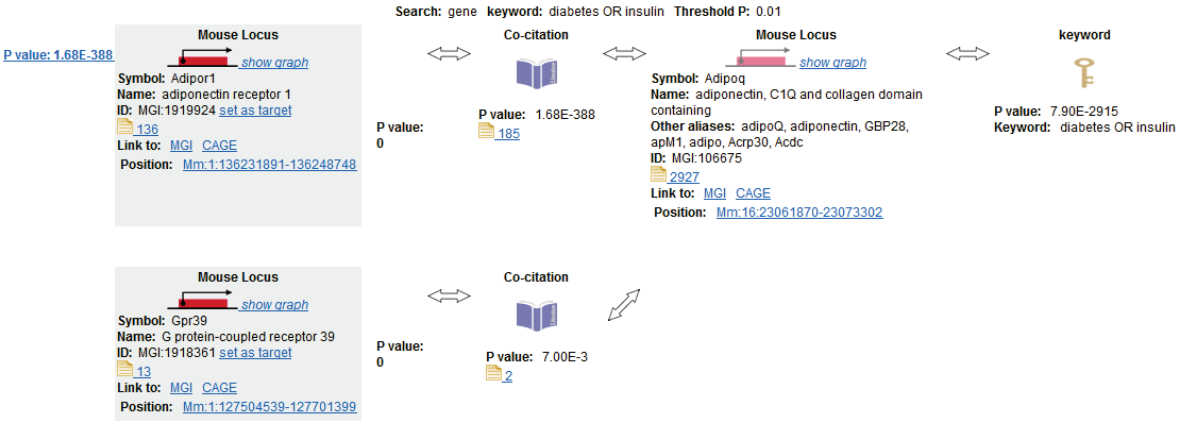
You can download the list of all documents by clicking the “download” button. The list has information about the PMID (PubMed identifier), title, name of the journal, published year and date, abstract and links to NCBI.

Click the “Find” image to display the keyword search box. Click the go button to superordinate documents including the highlighted keyword of interest. (*e.g.* lipid)



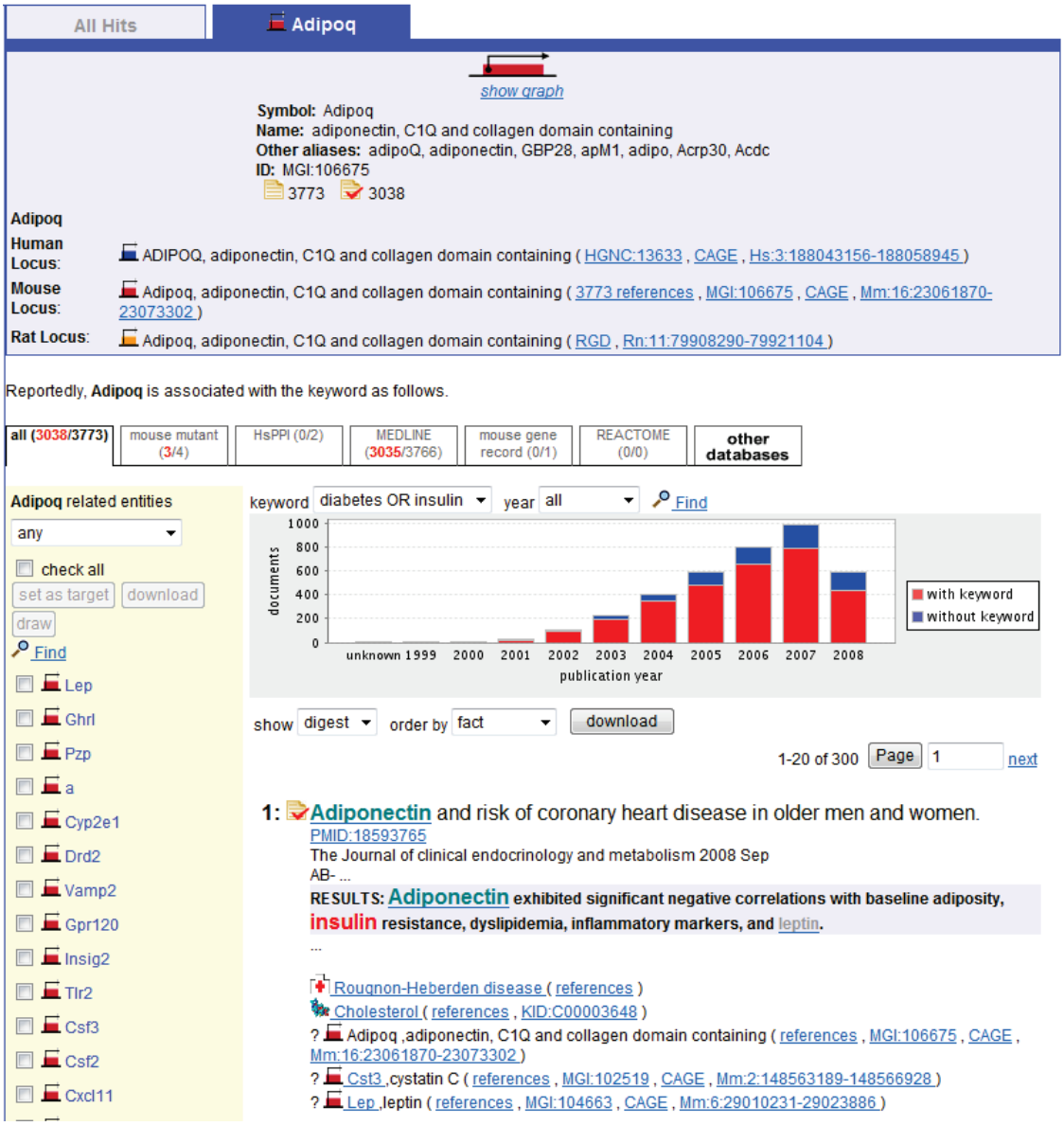
5-3. Relationship to gene cluster associated with the gene

When you click (3), gene (Adipoq), a list of relationships to genes associated with the gene is displayed.



5-4. Relationship between the gene and the keywords

When you click (4), the keyword icon, document digests describing both the keywords and the gene (Adipoq, adiponectin etc.) are displayed. In the document digest display, the gene name and the keywords are highlighted with light green (adiponectin) and red (diabetes or insulin) respectively. In the default display sequence, documents including the keyword(s) are ranked higher. Click a document number in order to transfer to PubMed at NCBI.



Additional biological objects are highlighted in gray. (*e.g.* leptin) Other related molecules are listed in the left yellow box. Click the images to display the relationships and to find further related biological objects.

