TargetMine Exercise and Tutorial

AIST KANSAI

January 20th 2012
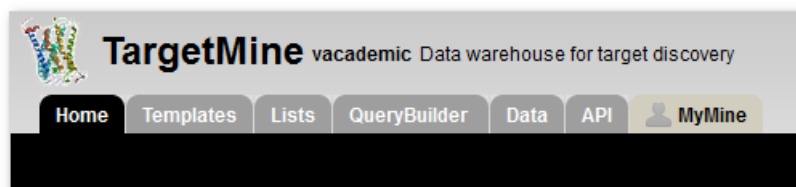
CONTENTS

Go to the TargetMine web site: targetmine.nibio.go.jp

All features are accessible from the toolbar on each TargetMine webpage:
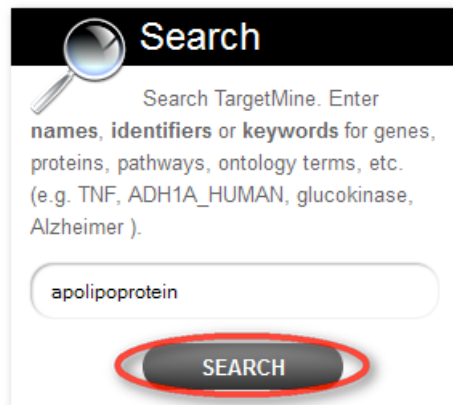

I. Quick Search and report pages

A search function may be found at the upper left corner of the top page or appears at the left side of menu bar on all pages. A user may enter any text (such as a gene name, identifier or a keyword) in the quick search box to search for the related content in TargetMine.

For example,

Run a quick search for a word 'apolipoprotein':
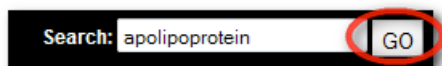
or

NOTE: Use wildcard character '*' for partial matching.

The search results may then further be filtered using different features. Clicking on the search result will direct the user to the `report page` for that object. The report page for a particular object in the TargetMine database provides all the stored information for that object. All report pages have the same basic structure:


- The upper half displays the basic information about the object. The lower parts provide more detailed information corresponding to the data

categories in TargetMine (which can be viewed on the home page). This information includes the classes associated with the object and the query results of a set of predefined queries known as templates (see below) related to specific data categories which are executed for the object on the fly.

II. Lists and List Analysis pages

The 'list' function in TargetMine enables queries on user-supplied list(s) of objects (e.g., a list of genes/proteins). The user may either save lists from results pages or create them by uploading lists of identifiers. Lists can be used to constrain queries, either a template query or in the query builder (see below), to perform logical operations (union, intersect, subtract) and exported. The users may also search and copy the lists.

The Lists page is accessible via the 'Lists' tab on the TargetMine menu bar. The 'Upload' tab on the Lists page permits the user to create a list of objects either by 'Type/Paste in identifiers' or upload a file containing a list of objects.
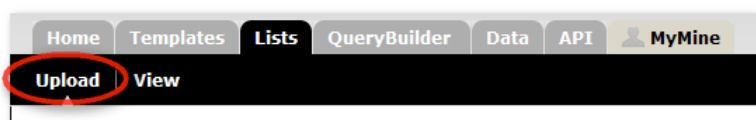
A list can be created of only one 'type' of objects (e.g., genes, proteins), which can be set from the drop-down list. However, it is possible for the list to contain a mixture of identifier types (e.g., gene symbols and gene IDs). Optionally, the user may also set the organism for the list of identifiers.

This tutorial will deal with uploading an example list of genes. The list will then be used to view the 'List Analysis page', run a template query, save a further list from the results and employ the lists for a customised query using the query builder.

• Click on 'click to see an example' link, to input an example gene list.

- Click on the "Create List" button.

- **List Confirmation**

Before a list is created it is checked against the TargetMine database and any discrepancies are reported in the '`Verify identifier matches`'. The user may either assign a name to the list or use the default name generated by the system. For the purpose of this tutorial, the users are advised to name the list as 'demo', and click on the save button.

- **List Analysis Page**

Details of the contents of the lists together with additional information about a list can be viewed through the '**List Analysis**' page. The analysis page is automatically displayed when a list is created. It can also be accessed at other times by clicking on the list name on the '`list page`'.

List Analysis page consists of three sections:
- A **summary** of the contents of the list includes a table with the first 10 objects of the list, a list description, List Info, Convert tools and Link outs (if available). Above the table, there are

Display/Export/Use options and a find box. Searching a list is case sensitive and wild cards cannot be used.

- A series of '**widgets**' providing more information about the list, for example, overrepresented KEGG pathways, Gene Ontology (GO) terms, Drug and Disease Ontology associations, Protein-protein interactions (PPIs) and Associated Drugs.

- A set of **template queries** that have been pre-run on the set of objects in the list.

III. Template queries

To enable the users to quickly and easily carry out diverse queries across data sources, TargetMine includes a library of **predefined Template queries**. These consist of a query form in which certain variables can be changed by imposing constraints (filters) on certain data types so as to restrict the queries to certain subsets of data.

Template queries can be accessed in different ways

- Each `data category` page has a set of templates related to that data type.
- Templates appear on `report pages` and `list analysis pages`.
- Templates can be searched using the **Quick Search** available on every TargetMine page, by changing the drop-down list from 'Identifiers' to 'Templates'.
- The `templates page`, accessible from the TargetMine home page or via tabs in the menu bar, allows the users to search or browse for templates.
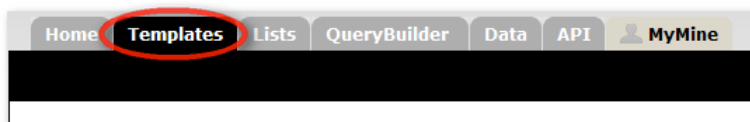
For example, search the template queries for 'transcription':
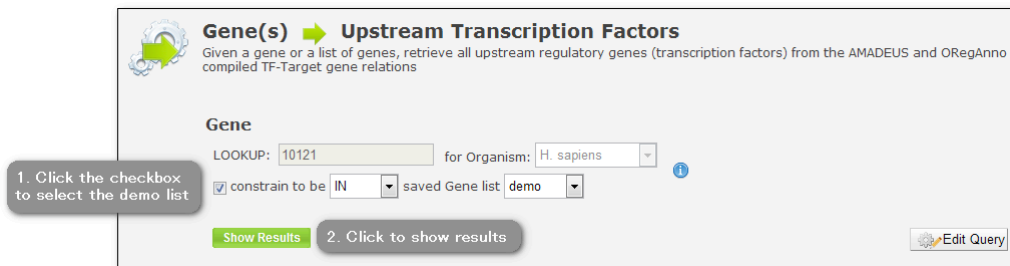

and select the template 'Gene(s) --> Upstream Transcription Factors':
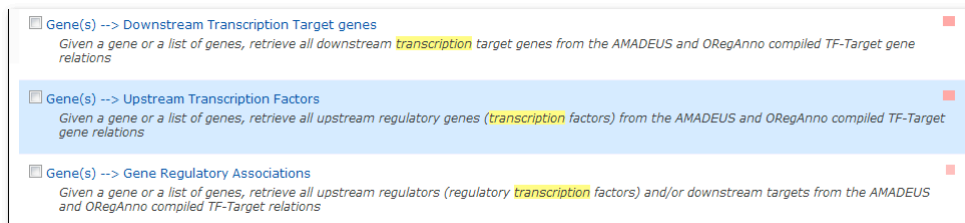
Now that we have created a list the template query gives us the option to use it instead of entering an identifier.

Click the checkbox to select the demo list "demo" saved earlier, and then click on "Show Results" to run the query.

The results table shows the transcription factors that target the genes from our original list.

The user can also create new lists from results tables. We can save a list of the transcription factors from this query to view in the list analysis page or to use in more templates.

Click on the checkbox by 'Source > ncbiGeneNumber' to select all genes
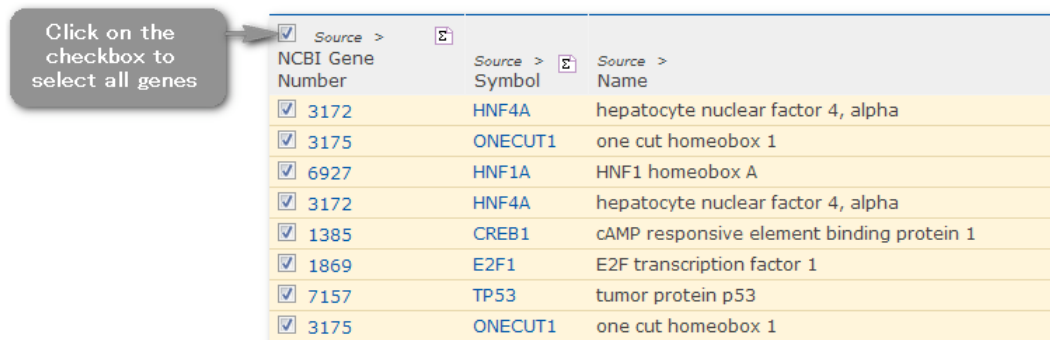
Click on the tab 'CREATE LIST' and choose a name for the saved list. For example, name it as 'upstream_TFs'. Then click 'Save selected' to save.
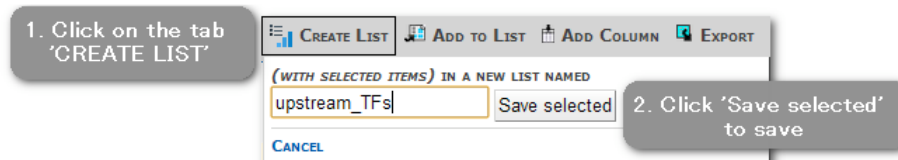
Using lists is an excellent way to retrieve information about a whole list of data. For example the users can also retrieve:

- Interactions for these genes
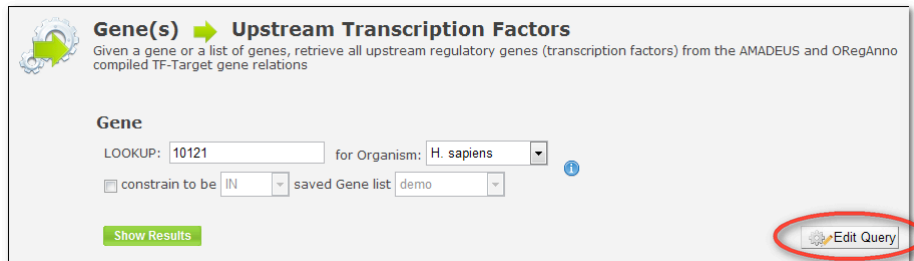- Disease Ontology (DO) annotation for these genes

Please try it yourself.

IV. Query Builder

Aimed mostly at advanced users, the 'Query Builder' allows the users to
- construct new queries across multiple data sources in TargetMine
- modify the existing template queries to change the output or add a constraint as may be understood from the following example.

Find the template ' Gene(s) --> Upstream Transcription Factors' as described in previous section. Click 'Edit Query' to bring up the Query Builder:


The 'Model browser' on the left permits the user to navigate the TargetMine data model. In this example, the model browser displays the initial Gene class and all its attributes (e.g., Name, Symbol) and related classes (e.g., Organism, Proteins). TargetMine uses an object based data model; biological types are termed 'classes' and individual entries are termed 'objects'.

The user can click ⊞ to open related classes and see their attributes. The SHOW↓ tab may be used to add fields to the output. The CONSTRAIN→ tab allows the user to add constraints (filters).
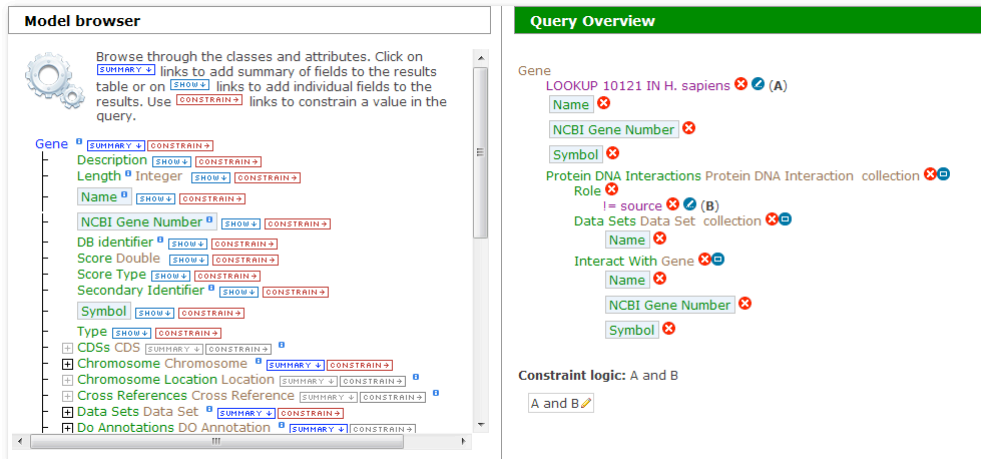
Underneath the model browser, the '**Fields selected for output**' displays the fields selected by the user to be displayed in the results. Each field is represented by a box which eventually corresponds to a column in the results table. The order of the fields in the column may be changed by drag and drop.
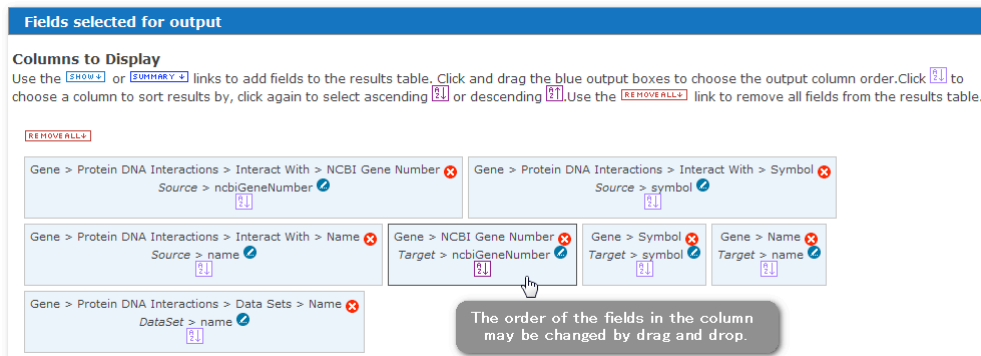
The '**Query Overview**' section shows the constraints applied to this query. Notice that one of them says 'Gene' then '**LOOKUP 10121 IN H. sapiens**'. This indicates that the query is restricted to a human gene with the identifier 10121.

What if the user wants to search for **all** genes in TargetMine for **any** upstream transcription factor that may regulate their gene expression? The user may simply remove this constraint and run the query again.

Click on the ❌ button next to '**LOOKUP 10121 IN H. sapiens**' to remove the constraint:
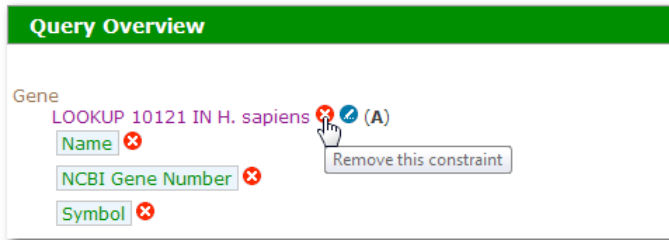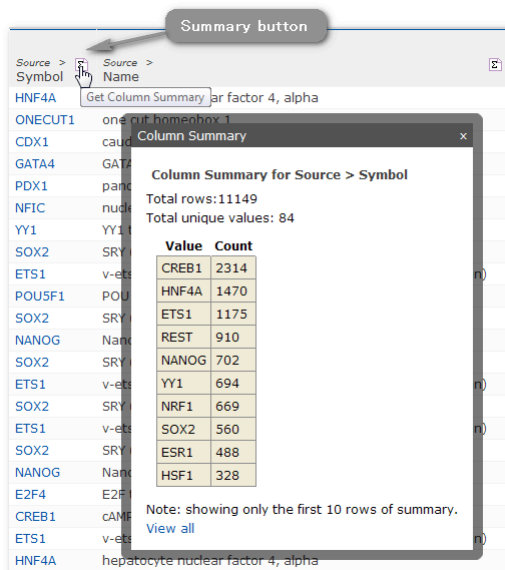
Now click 'Show results' to run the query:
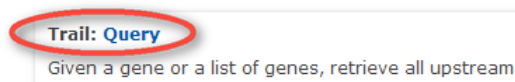

The results table now shows all the upstream transcription factors in TargetMine and their target genes.

The user may click the summary button Σ on the 'Source > symbol' and 'Target > symbol' columns to find more information about the column.
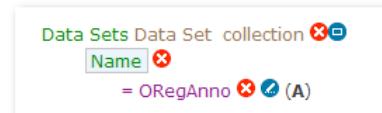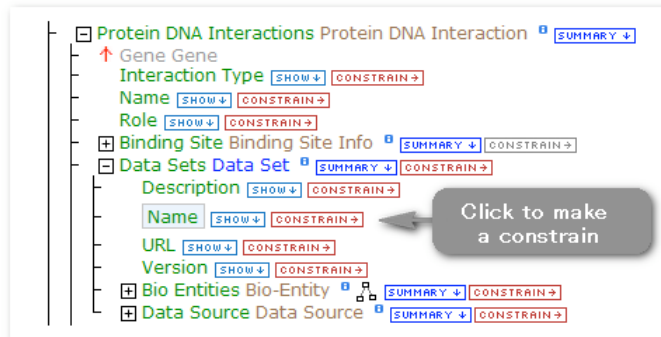

Let's try something further.
Click on 'Query' in the Trail to revert to the query builder.

Now, try to use the 'Model browser' and find the attribute at the following path 'Gene -> Protein DNA Interactions -> Data Sets -> Name'. Click the

`CONSTRAIN→` tab beside to bring out a pop-up window. Select 'ORegAnno' from the dropdown list then click 'Add to query' to make the constraint.

Model browser   Query overview

Click 'Show results' to run the query again. The results will only show the transcription factors retrieved from the ORegAnno database.

The users are highly encouraged to play around with the QueryBuilder themselves.