TargetMine：創薬ターゲットの絞り込みを支援する統合データウェアハウス

水口賢司

http://www.mizuguchilab.org
kenji@nibio.go.jp

医薬基盤研究所

1. 背景

2. 基本概念（data model, query builder and template queries）

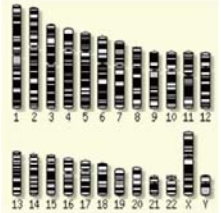
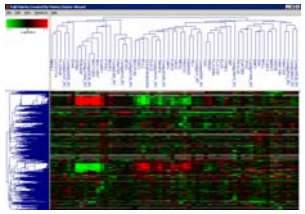
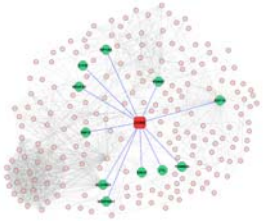
3. 候補遺伝子（タンパク質）絞り込み手法（lists, widgetsその他の機能）

4. C型肝炎ウィルス関連実験データの解析への応用

創薬支援バイオインフォマティクス研究

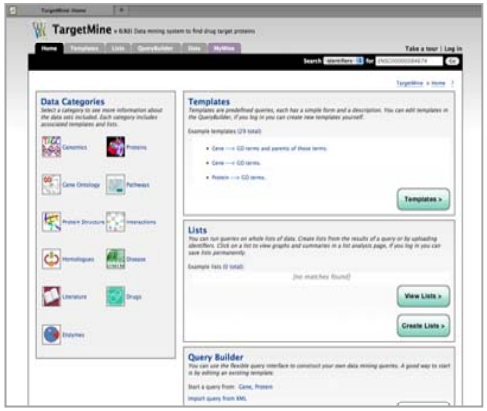
慢性炎症性疾患、呼吸器疾患…

実験データ
疾患関連遺伝子候補のリスト

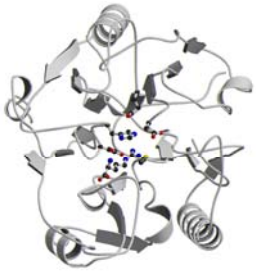




統合データウェアハウスによる候補遺伝子（タンパク質）の絞り込み

TargetMine
Data mining system to find drug target proteins



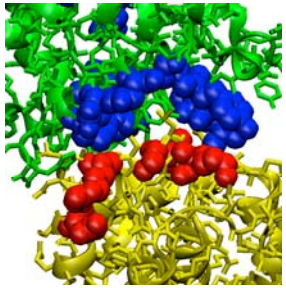
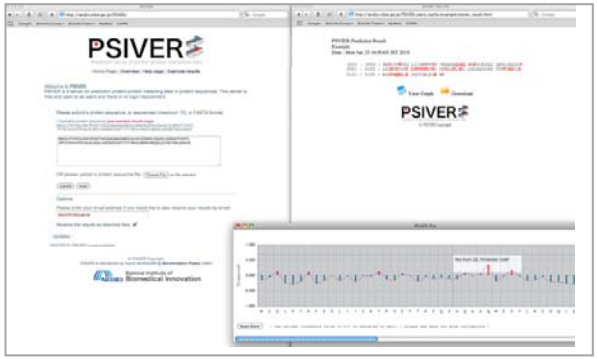
創薬ターゲット候補のバリデーション



立体構造予測

FUGUE

相互作用部位予測

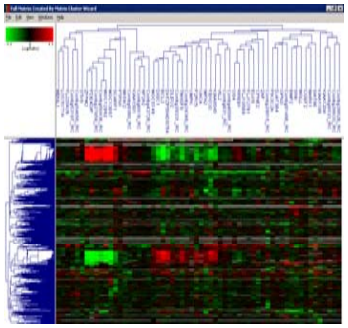



ターゲット開発の現状

新たな創薬ターゲットの発見が急務であるにも関わらず、ゲノミクスなど新規技術を用いたターゲット創出が進んでいない。

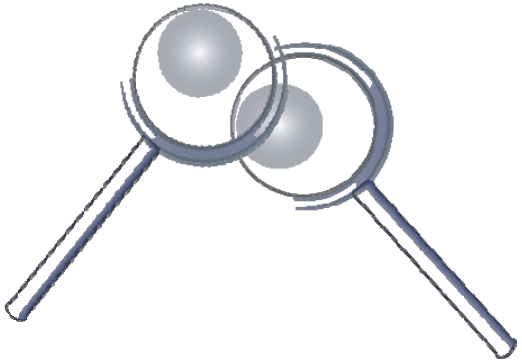
（個々のタンパク質に関する分子機能情報の不足と、疾患との関連付けの難しさから、対象を絞りこんで実験的手法を適応できないのが主な原因）

遺伝子の
発現変動解析



多数のターゲット候補

```
>P23656 CCMO_SYNP6 Hypothetical protein
>P21682 CP79_BP186 Hypothetical 8.8 kDa
>P20343 CYSX_ECOLI Very hypothetical
>P15017 DNU4_RHORU Hypothetical
>P68349 DOF12_ARATH Hypothetical Dof zinc
>P0A3F6 DSDX_ECO57 Hypothetical dsdX-like
>P0A3F7 DSDX_SHIFL Hypothetical dsdX-like
>P06630 EAG_BACSU Hypothetical protein eag
>P30194 EPIY_STAEP Hypothetical 16.7 kDa
…
```



TargetMine

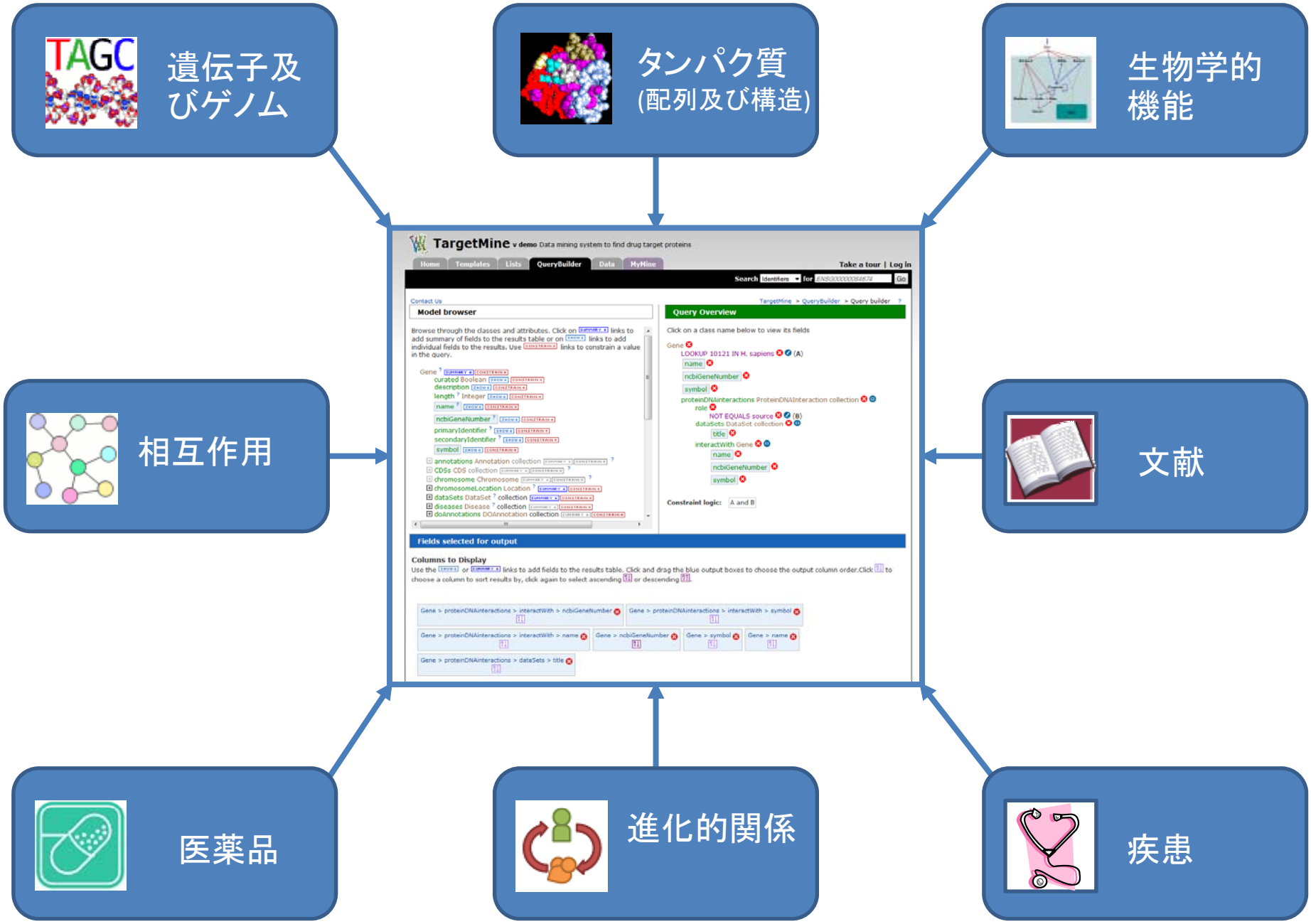
創薬ターゲットの絞り込みを支援する
統合データウェアハウス

- 国際的に広く使用されている主要な16個の公共データベースからのデータを統合
- 既存の公共ツールでは行なうことができない、タンパク質立体構造や医薬品関係データ、転写因子とその作用遺伝子の関係等の情報の統合的検索が可能
- 公共データベースに入っていない、独自のアノテーションや実験結果を組み込む仕組みを提供

Data warehouse system

COPYRIGHT IMAGE

A Data warehouse compiles the contents of multiple databases to fit a common data model (Stein 2003, *Nat Rev Genetics* 4, 337-345)


遺伝子及びゲノム
タンパク質(配列及び構造)
生物学的機能
相互作用
文献
医薬品
進化的関係
疾患
TargetMine
Model browser
Query Overview
Fields selected for output
Columns to Display

Sources: Public databases

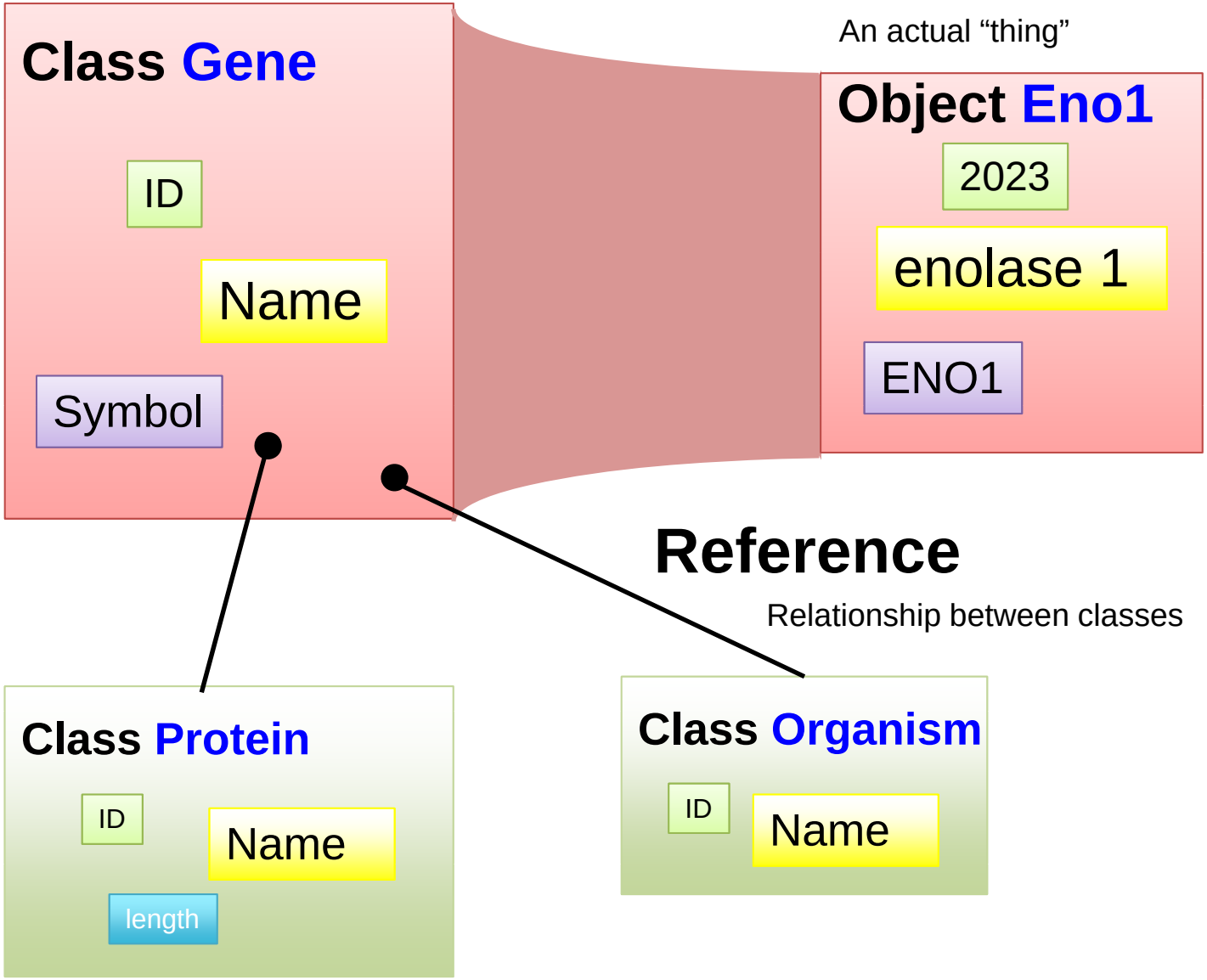
Data type	Database name
Comprehensive gene data	Entrez Gene
Comprehensive protein data	UniProt
Protein domain	InterPro
Pathway	KEGG Pathway
Interaction	BIOGRID, PPIview
GO annotation and ontology	Gene Ontology
Protein structure	PDB, SCOP
Drug	DrugBank
Disease	OMIM, Disease ontology (DO)
Transcription factor	ORegAnno, AMADEUS
Orthologues/Paralogues	KEGG Orthology
Enzyme	The ENZYME database

Organisms in TargetMine

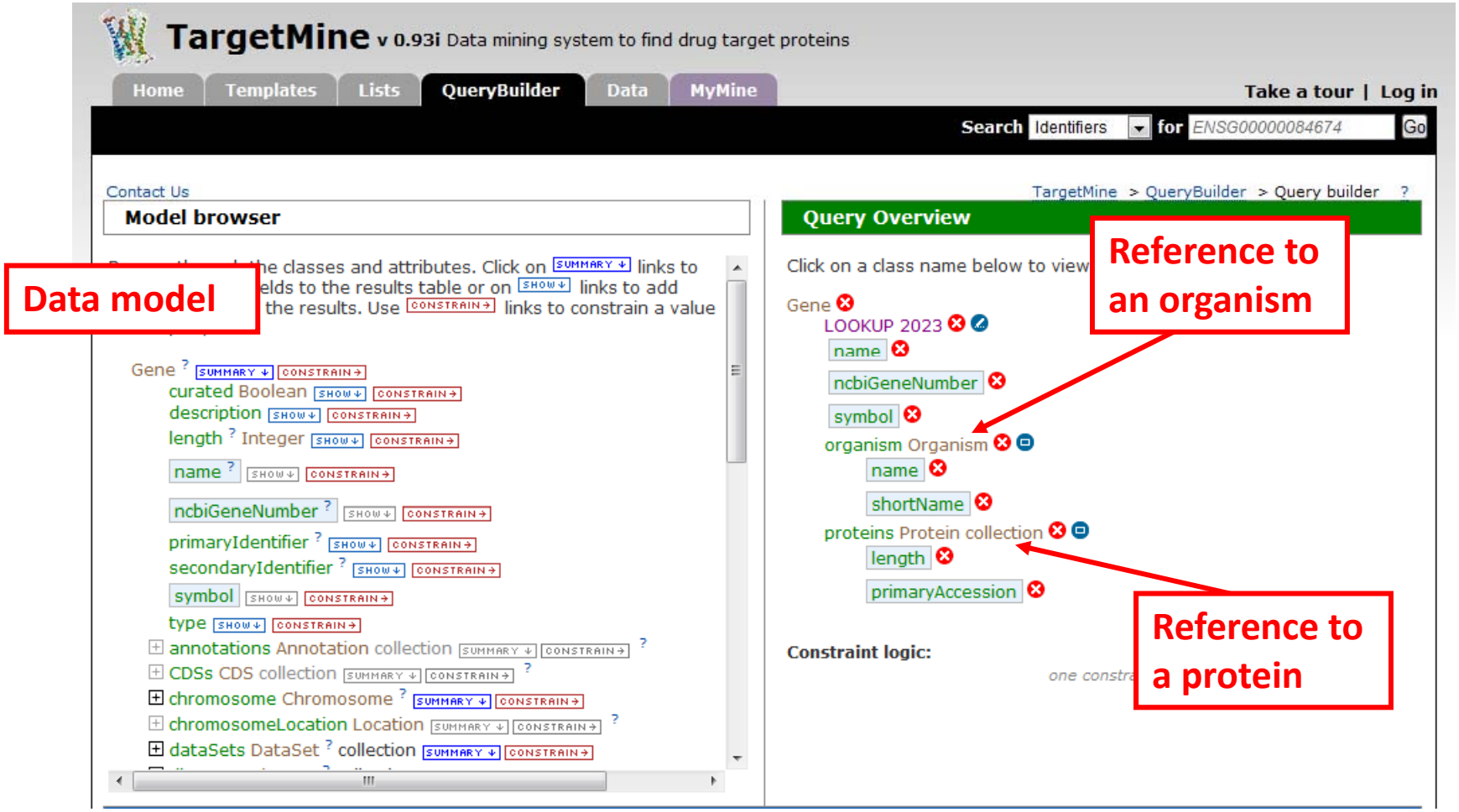
- Protein, Gene, GO, Pathway …
  - *Homo sapiens*
  - *Rattus norvegicus*
  - *Mus musculus*
  - *Drosophila melanogaster*
- PDB related data
  - All organisms are included

Data Model

A holder for “things” with similar properties



Query builder



Output properties list

Fields selected for output

Columns to Display

Use the SHOW or SUMMARY links to add fields to the results table. Click and drag the blue output boxe
choose a column to sort results by, click again to select ascending or descending.

Each box for one property.
The order can be changed.

- Gene > ncbiGeneNumber
- Gene > symbol
- Gene > name
- Gene > organism > shortName
- Gene > proteins > primaryAccession
- Gene > proteins > length
- Gene > organism > name

Trail: Query > Results

Create List | Add to List | Add Column | Export | Page size 25 | << First < Previous | Next > Last >>

Gene > ncbiGeneNumber	Gene > symbol	Gene > name	Gene > organism > shortName	Gene > proteins > primaryAccession	Gene > proteins > length	Gene > organism > name
2023	ENO1	enolase 1, (alpha)	H. sapiens	P06733	434	Homo sapiens

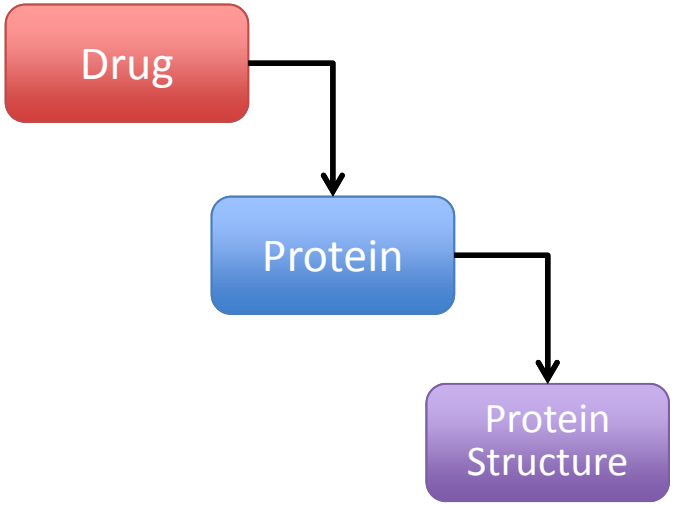
Selected:

<< First < Previous | Next > Last >> | Displaying all 1 rows

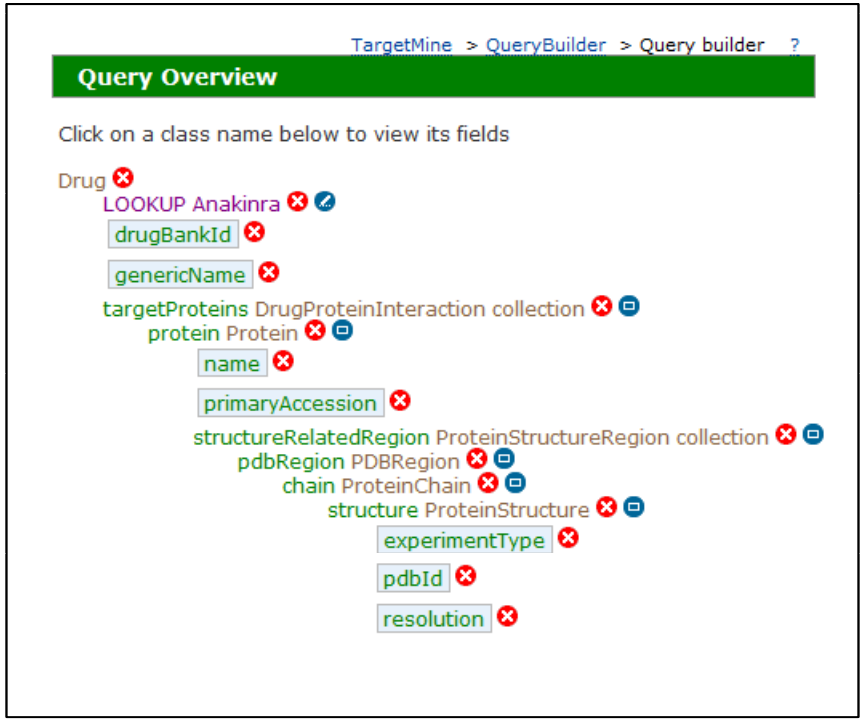
Display of the Results Table

A query example

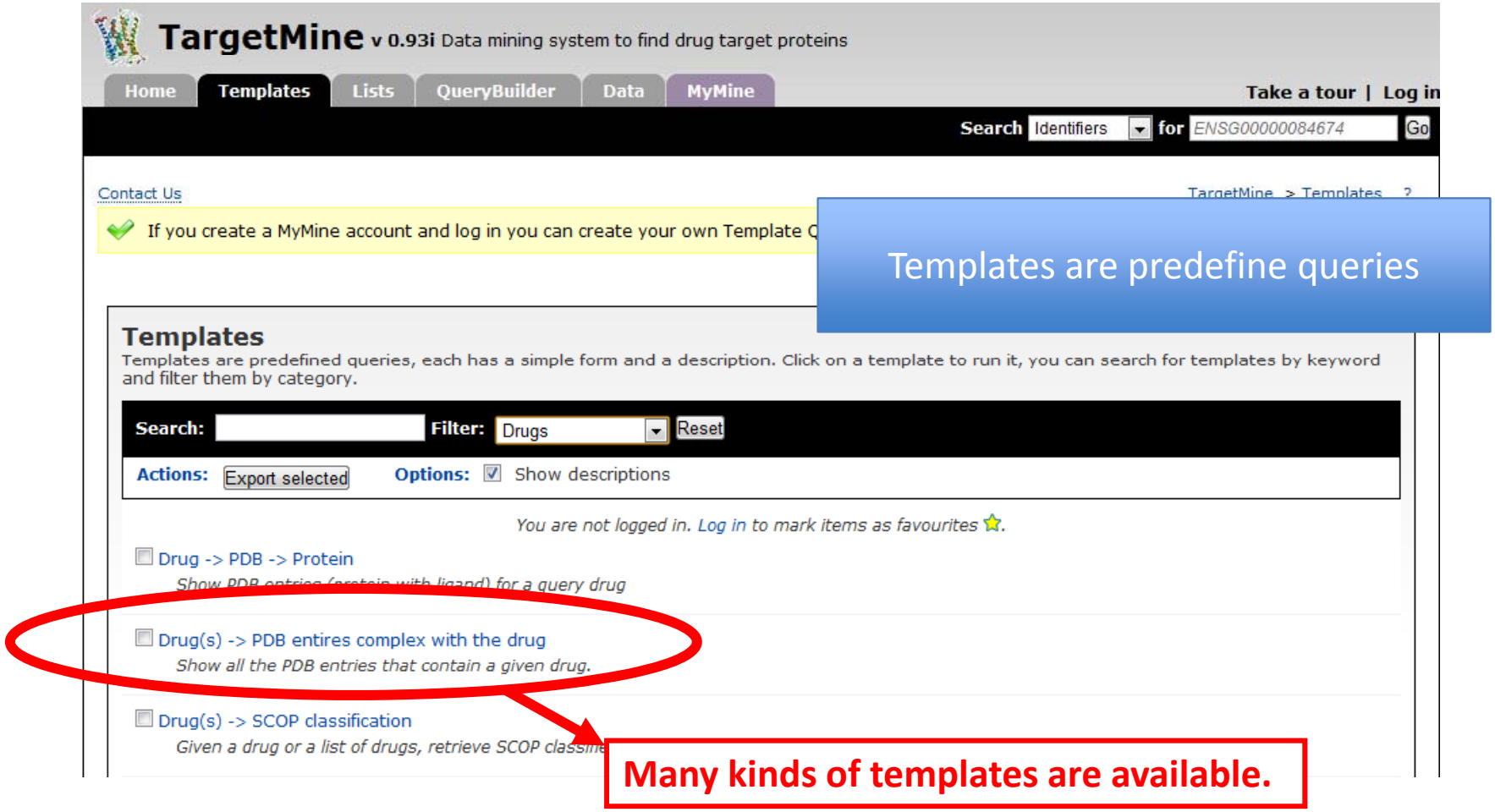
- Show all the protein structures that contain the targets of a given set of drugs (defined in DrugBank).



Complicated queries can be made without writing any SQL query.



Template function



Template function

Template query results

TargetMine v 0.93i Data mining system to find drug target proteins

Home | Templates | Lists | QueryBuilder | Data | MyMine

Take a tour | Log in

Search Identifiers for ENSG00000084674 Go

Contact Us

TargetMine > QueryBuilder > Query results page ?

Trail: Query > Results

Results for template: **Drug(s) -> PDB entires complex with the drug**
Show all the PDB entries that contain a given drug.

Create List | Add to List | Add Column | Export | Page size 25 | << First < Previous | Next > Last >>

Drug > drugBankId	Drug > genericName	ProteinStructure > pdbId	ProteinStructure > name	ProteinStructure > experimentTyp
DB00819	Acetazolamide	1azm	DRUG-PROTEIN INTERACTIONS: STRUCTURE OF SULFONAMIDE DRUG COMPLEXED...	X-RAY DIFFRACT
DB00819	Acetazolamide	1dmy	COMPLEX BETWEEN MURINE MITOCHONDRIAL CARBONIC ANYHDRASE V AND THE...	X-RAY DIFFRACT
DB00819	Acetazolamide	1jd0	CRYSTAL STRUCTURE OF THE EXTRACELLULAR DOMAIN OF HUMAN CARBONIC...	X-RAY DIFFRACT
DB00819	Acetazolamide	1rj6	Crystal Structure of the Extracellular Domain of Murine Carbonic...	X-RAY DIFFRACT
DB00819	Acetazolamide	1yda	STRUCTURAL BASIS OF INHIBITOR AFFINITY TO VARIANTS OF HUMAN...	X-RAY DIFFRACT
DB00819	Acetazolamide	1ydb	STRUCTURAL BASIS OF INHIBITOR AFFINITY TO VARIANTS OF HUMAN...	X-RAY DIFFRACT
DB00819	Acetazolamide	1ydd	STRUCTURAL BASIS OF INHIBITOR AFFINITY TO VARIANTS OF HUMAN...	X-RAY DIFFRACT
DB00819	Acetazolamide	1zsb	CARBONIC ANHYDRASE II MUTANT E117Q, TRANSITION STATE ANALOGUE...	X-RAY DIFFRACT
DB00819	Acetazolamide	2h4n	H94N CARBONIC ANHYDRASE II COMPLEXED WITH ACETAZOLAMIDE	X-RAY DIFFRACT
DB00819	Acetazolamide	2uy4	ScCTS1_acetazolamide crystal structure	X-RAY DIFFRACT
DB00819	Acetazolamide	3czv	Crystal structure of the human carbonic anhydrase XIII in complex...	X-RAY DIFFRACT
DB00819	Acetazolamide	3dc3	Use of Carbonic Anhydrase II, IX Active-Site Mimic, for the Purpose...	X-RAY DIFFRACT

Selected:

<< First < Previous | Next

Click for detailed information (Summary page)

Examples for a summary page

Contact Us

Trail: Query > Results > Drug

Summary for selected Drug

drugBankId	DB00819
genericName	Acetazolamide
casRegistryNumber	59-66-5
description	One of the carbonic anhydrase inhibitors that is sometimes effective against absence seizures. It is sometimes useful also as an... [View all]
fdaLabelIssuedDate	Tue Mar 01 09:00:00 JST 2005
hetId	AZM
keggDrugId	D00218

⊞ brandNames 3
⊞ drugTypes 2

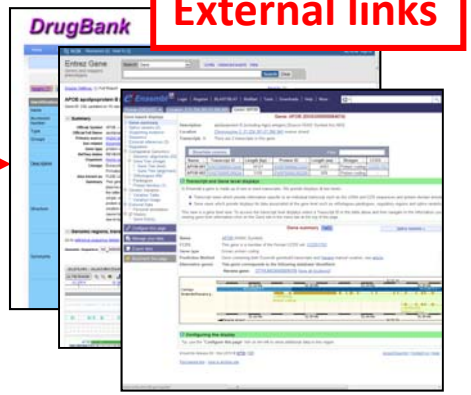
Further information for this Drug

DrugBank Drug Card: DB00819

Lists

Add this object to existing list: --------

External links



An example of a Drug object

Trail: Query > Results > Gene

Summary for selected Gene

ncbiGeneNumber ?	2012
primaryIdentifier ?	ENSG00000134531
secondaryIdentifier ?	HGNC:3333
symbol	EMP1
name ?	epithelial membrane protein 1
chromosomeLocation	[unknown]
organism.shortName	H. sapiens
description	-

⊞ organism 1 Organism [details...]
⊞ proteins 1 Protein [details...]
⊞ publications 13 Publication
⊞ synonyms 8 Synonym

Further information for this Gene

ensembl: ENSG00000134531

Entrez Gene: 2012

Lists

Lists in which this Gene can be found: genes

Add this object to existing list: -------- Add

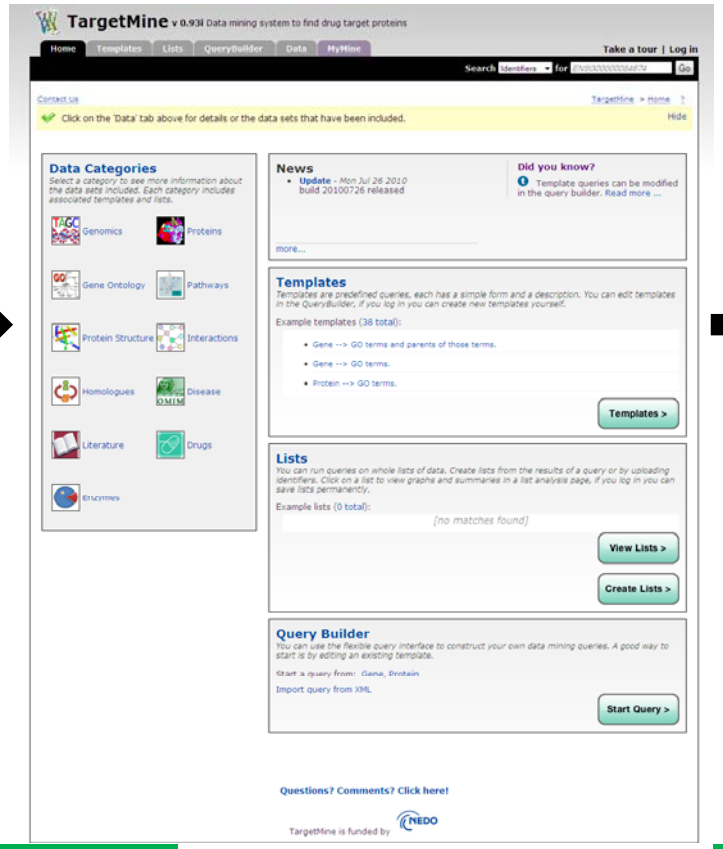
An example of a Gene object

Target selection using TargetMine

INPUT

Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
....
....
....
....
Gene 200

Biological process of interest



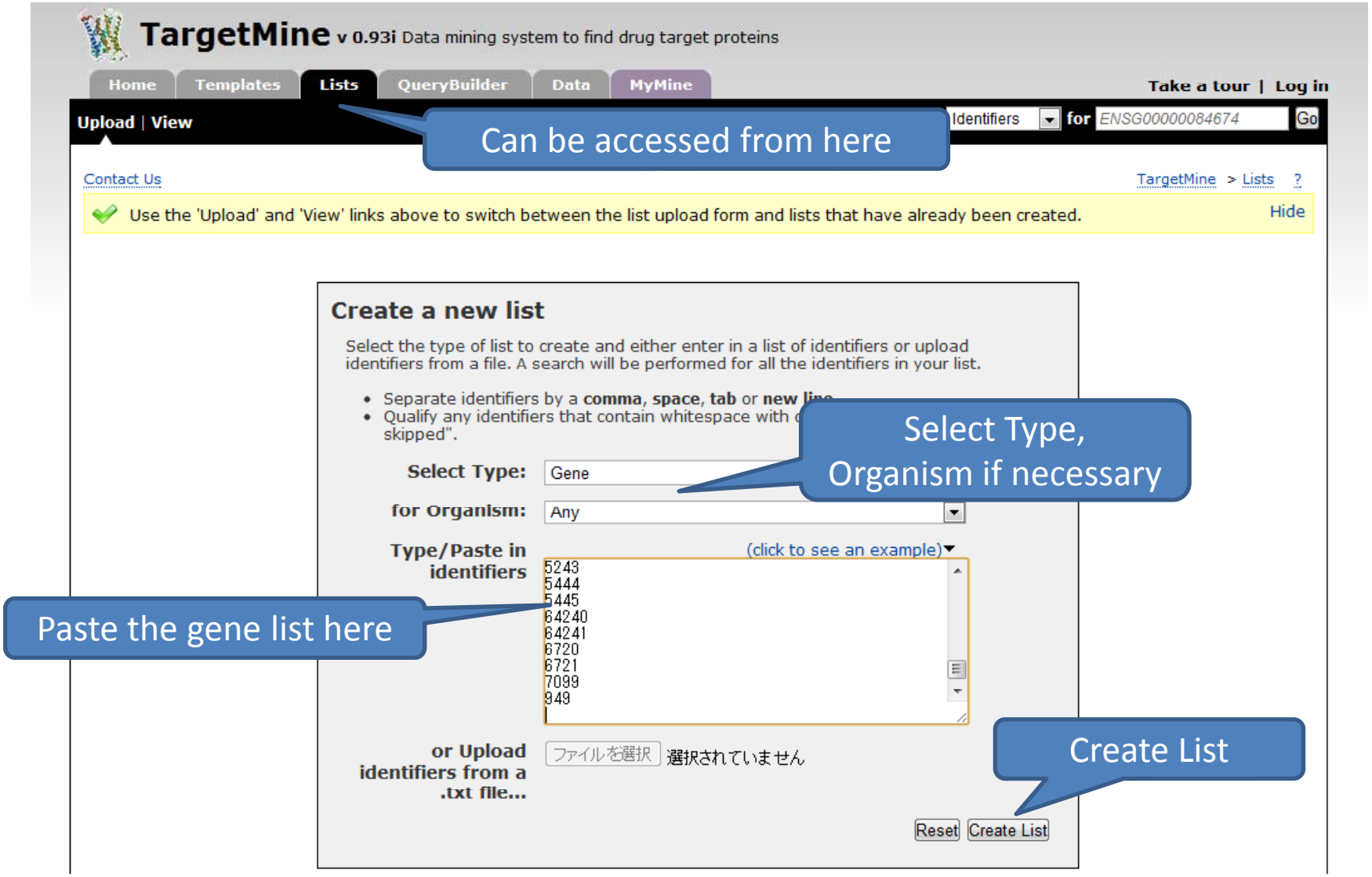
OUTPUT

Gene 4
Gene 17
Gene 45
Gene 87

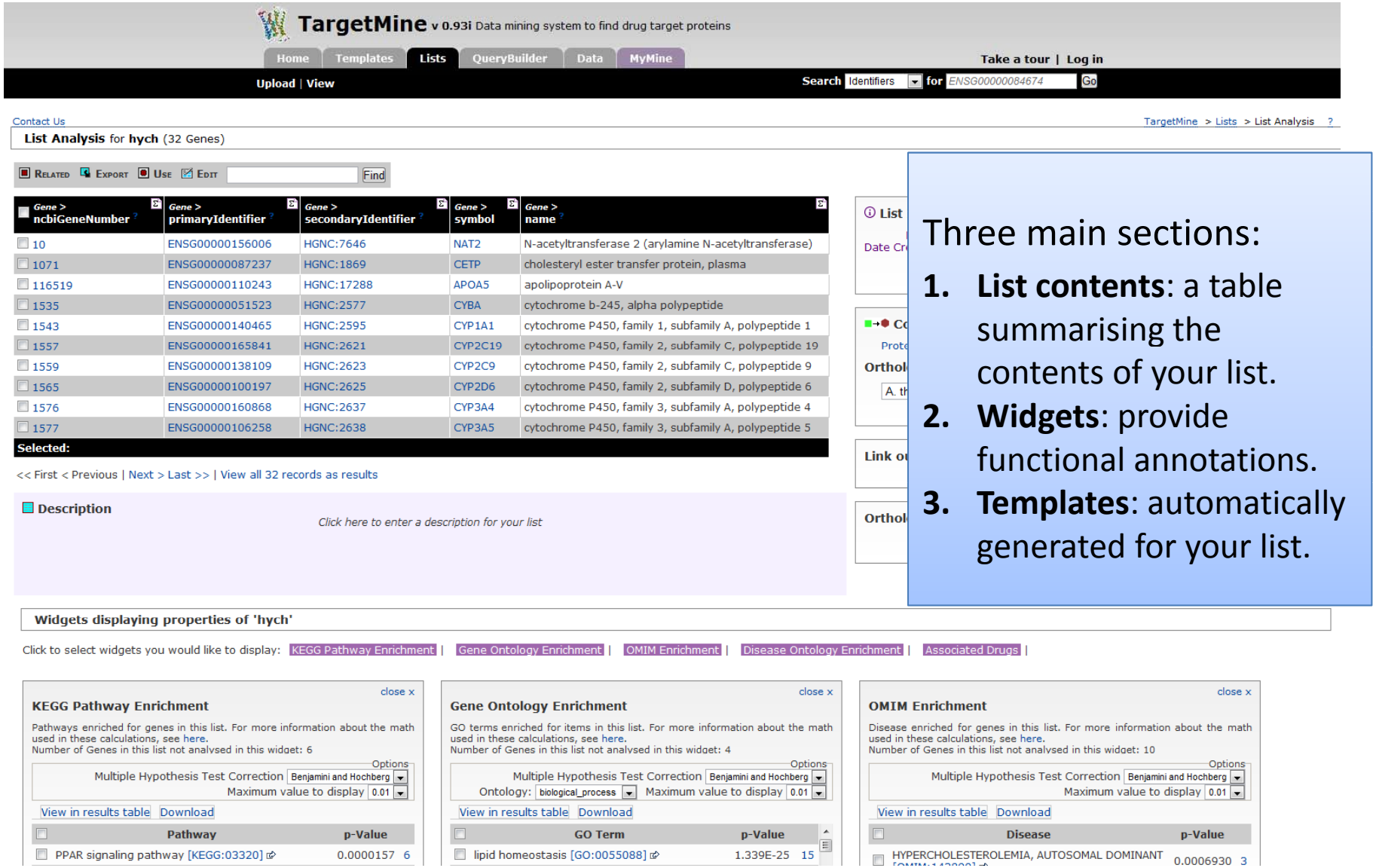
Many Candidates

Selected Targets

List function



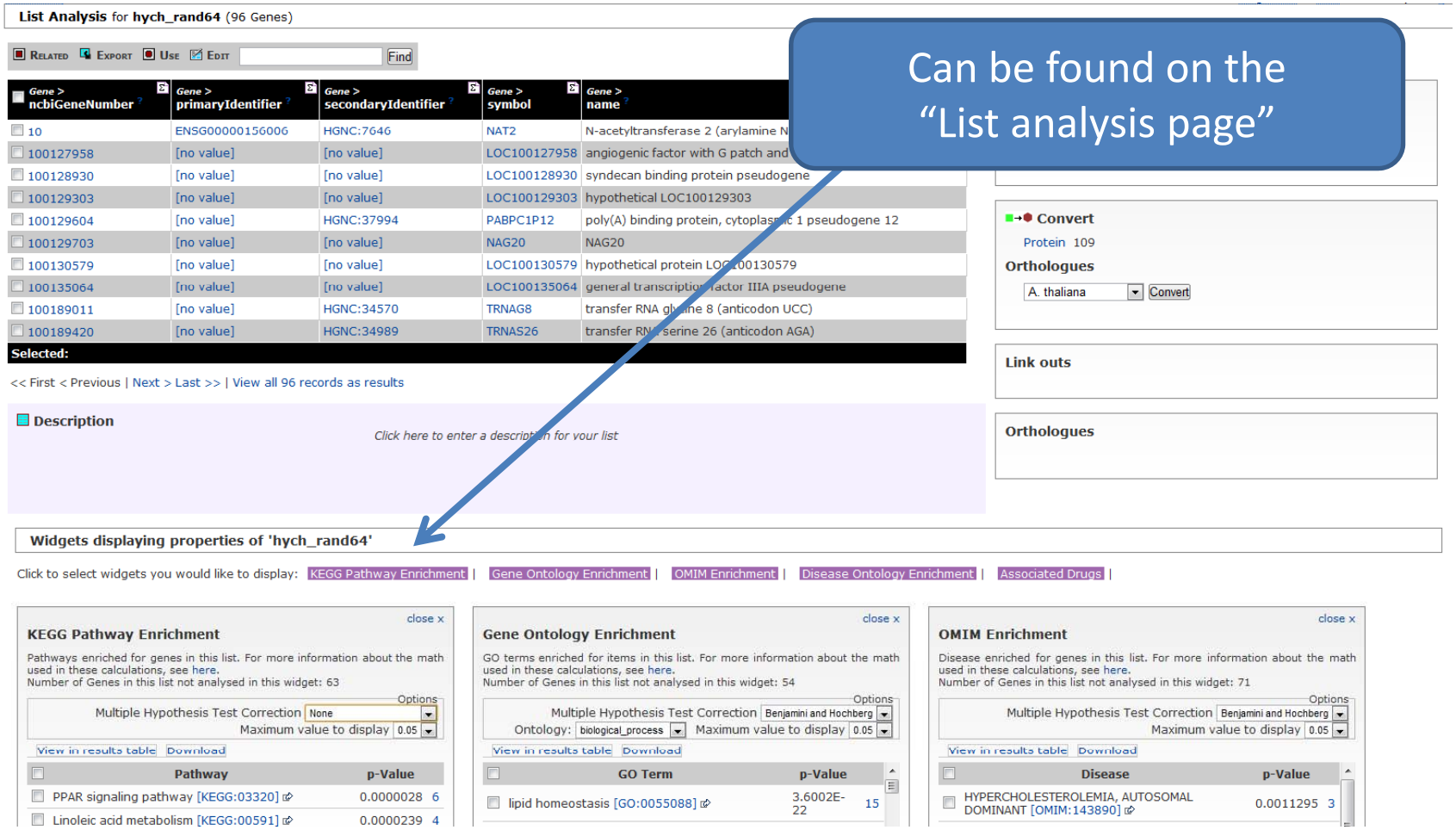
List analysis page



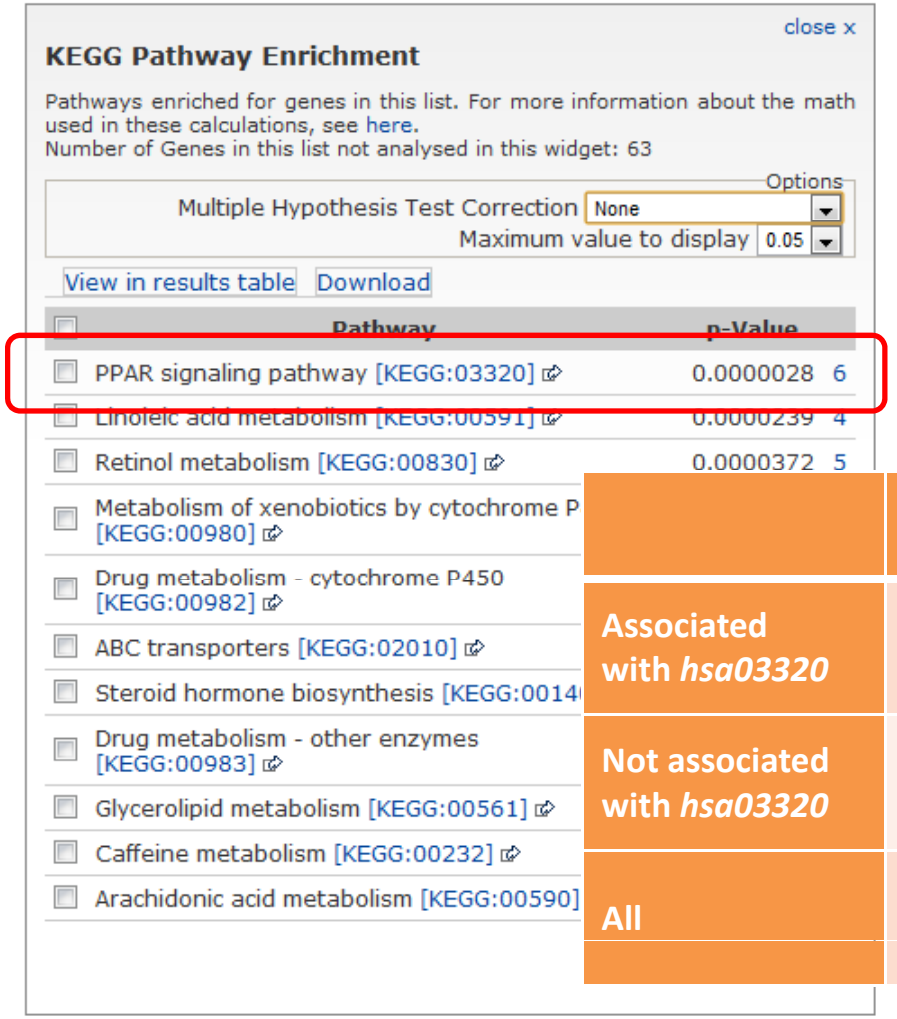
Three main sections:

1. **List contents**: a table summarising the contents of your list.
2. **Widgets**: provide functional annotations.
3. **Templates**: automatically generated for your list.

Enrichment widgets



Enrichment analysis



$$p = \frac{C_6^{69} \times C_{27}^{4979}}{C_{33}^{5048}}$$

$$= 0.00000278955\ldots$$

	In the list	Not in the list	All
Associated with *hsa03320*	6	63	69
Not associated with *hsa03320*	27	4952	4979
All	33	5015	5048

Enrichment widgets

close x

KEGG Pathway Enrichment

Pathways enriched for genes in this list. For more information about the math used in these calculations, see here.
Number of Genes in this list not analysed in this widget: 63

Options

Multiple Hypothesis Test Correction: Benjamini and Hochberg
Maximum value to display: 0.05

View in results table | Download

	Pathway	p-Value	
☑	PPAR signaling pathway [KEGG:03320]	0.0001088	6
☑	Linoleic acid metabolism [KEGG:00591]	0.0004652	4
☑	Retinol metabolism [KEGG:00830]	0.0004841	5
☑	Drug metabolism - cytochrome P450 [KEGG:00982]	0.0005064	5
☑	Metabolism of xenobiotics by cytochrome P450 [KEGG:00980]	0.0005546	5
☐	ABC transporters [KEGG:02010]	0.0008172	4
☐	Steroid hormone biosynthesis [KEGG:00140]	0.0017757	4
☐	Drug metabolism - other enzymes [KEGG:00983]	0.0169648	3

Check the checkboxes of the top 5 enriched items (pathways) in the list

close x

KEGG Pathway Enrichment

Pathways enriched for genes in
used in these calculations, see
Number of Genes in this list not

Click on "view in results table"

Maximum value to display: 0.05

View in results table | Download

Display checked items in results table
Display all items in results table
Cancel

Show results in a new results table

TargetMine v 0.93i Data mining system to find drug target proteins

Home | Templates | Lists | QueryBuilder | Data | MyMine

Take a tour | Log in

Search Identifiers for ENSG00000084674 Go

Contact Us

TargetMine > QueryBuilder > Query results page

Saved your query as "query_9"

You can see a summary of any results column by clicking the sigma icon in the header. Hide

Trail: List > Results

Selected Genes from the list 'hych_rand64' for the widget 'KEGG Pathway Enrichment'

Create List | Add to List | Add Column | Export | Page size 25 | << First < Previous | Next > Last >>

Gene > ncbiGeneNumber	Gene > symbol	Gene > name	Gene > pathways > identifier	Gene > pathways > name
1557	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	00591	Linoleic acid metabolism
1559	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	00591	Linoleic acid metabolism
1576	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	00591	Linoleic acid metabolism
1577	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	00591	Linoleic acid metabolism
1543	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	00830	Retinol metabolism
1557	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	00830	Retinol metabolism
1559	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	00830	Retinol metabolism
1576	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	00830	Retinol metabolism

Trail: List > Results

Selected Genes from ... athway Enrichment'

Check the checkbox to select all genes

Create ... Add to List Add Column Export Page size 25 << First < Previous | Next > Last >>

Gene > ncbiGeneNumber	Gene > symbol	Gene > name	Gene > pathways > identifier	Gene > pathways > name
1557	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	00591	Linoleic acid metabolism
1559	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	00591	Linoleic acid metabolism
1576	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	00591	Linoleic acid metabolism
1577	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	00591	Linoleic acid metabolism

Click on "Create List"

Trail: List > Results

Selected Genes from ... list 'hych_rand64' for the widget 'KEGG Pathway Enrichment'

Create List Add to List Add Column Export Page size 25 << First < Previous | N

(with selected items) in a new list named

kegg_top_5 Save selected

Cancel

Save as a new list

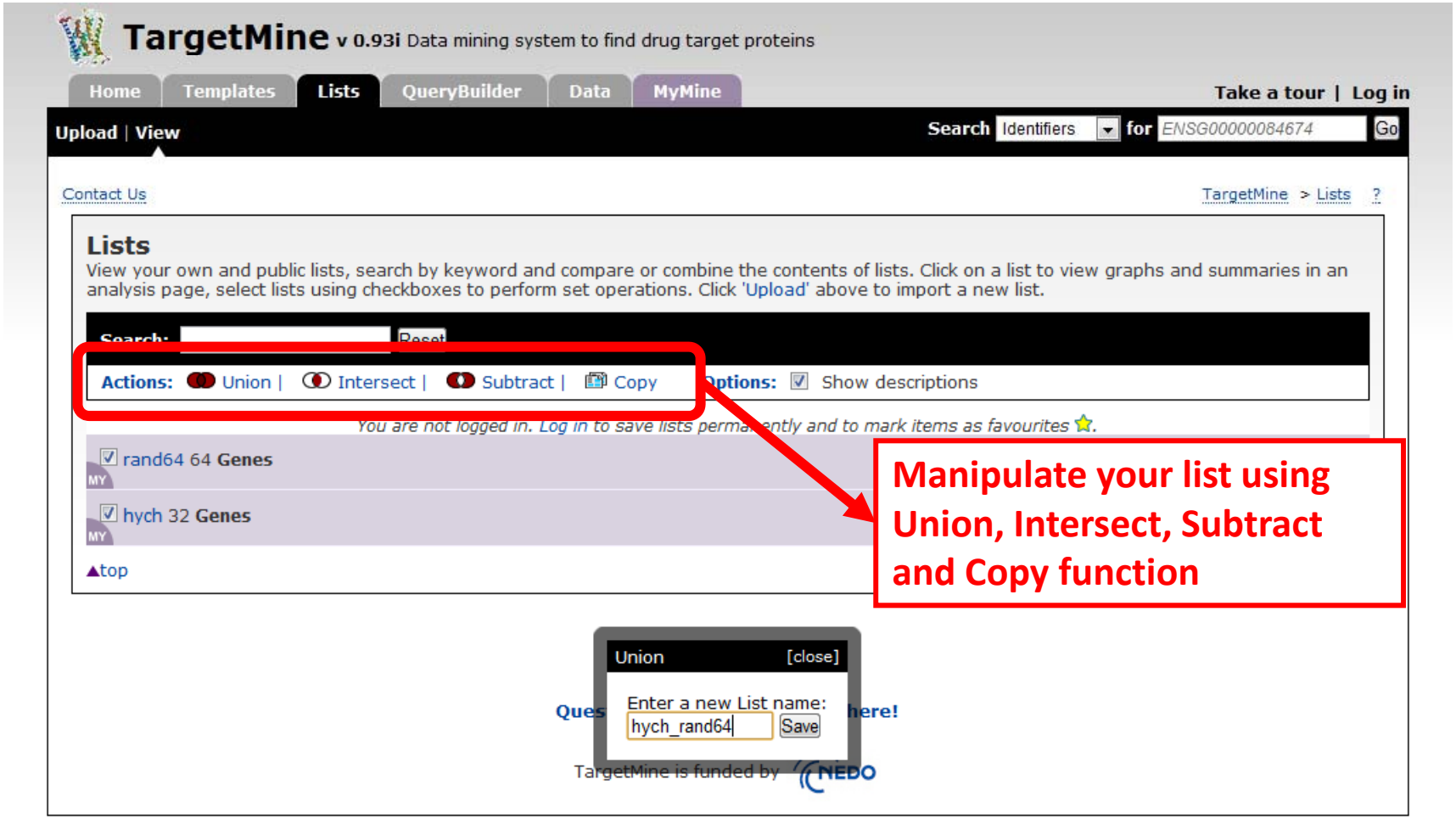
Do the same operation for **GO** and **OMIM**, and get 3 lists

omim_top_5 23 **Genes**

go_top_5 16 **Genes**

kegg_top_5 12 **Genes**

Organise your list



Manipulate your list using Union, Intersect, Subtract and Copy function

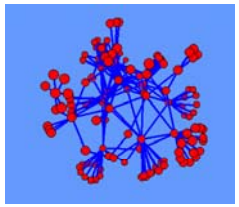
C型肝炎新規ターゲット候補の同定

【背景】C型肝炎ウィルス(HCV)コアタンパク質とヒトPA28γとの相互作用が、肝癌の発症に重要な役割を果たしていることが見いだされたが、詳細なメカニズムは不明

【本解析以前の状況と問題点】

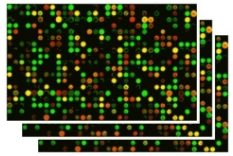
酵母2ハイブリッド法

HCVの病原性及び複製機構を解明する目的で、HCVコアタンパク質及び、NS4Bタンパク質と相互作用する宿主のタンパク質を網羅的に同定した。



タンパク質マイクロアレイ

HCVによる癌化のモデルマウス間で、肝臓でのタンパク質発現量に差のあるタンパク質を同定した。

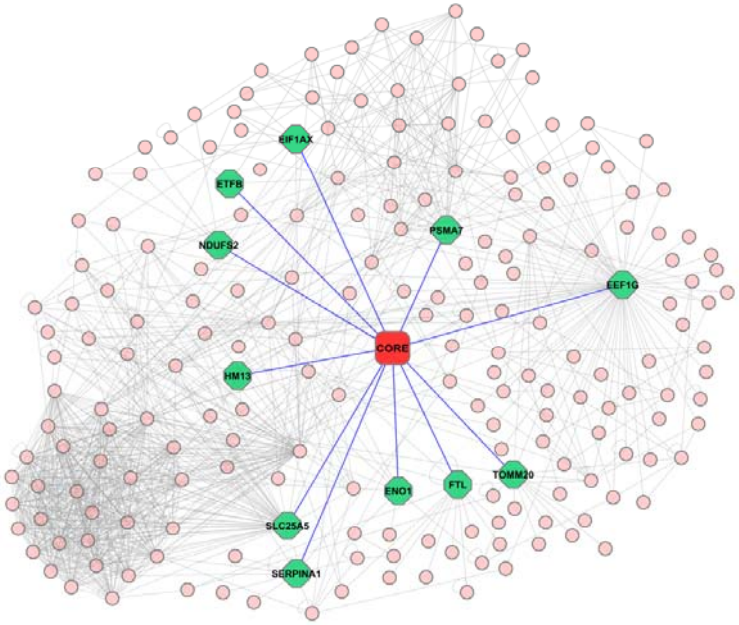


→ 多数の候補遺伝子（タンパク質）は同定されたが、意味付けは難しく、次の実験へ進めなかった

【目的】これらの遺伝子リストを解析して絞り込みを行ない、実験的に検証可能な仮説を提唱する

共同研究者：森石　恒司、松浦　善治（大阪大学微生物病研究所）

C型肝炎新規ターゲット候補の同定



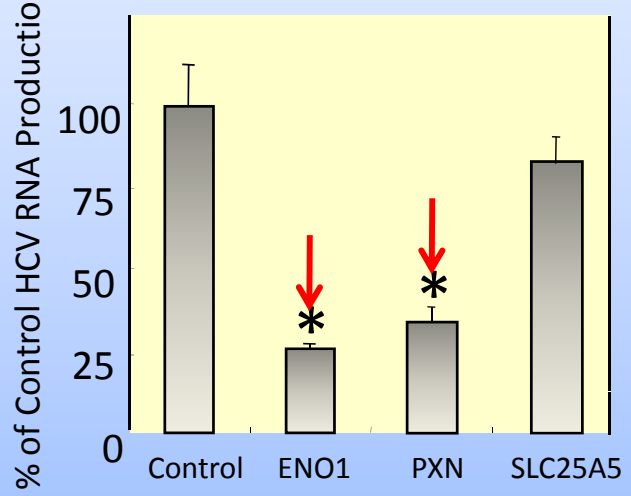
ウィルスタンパク質と宿主の細胞ネットワークを結びつけている459 の遺伝子

TargetMine

Data mining system to find drug target proteins

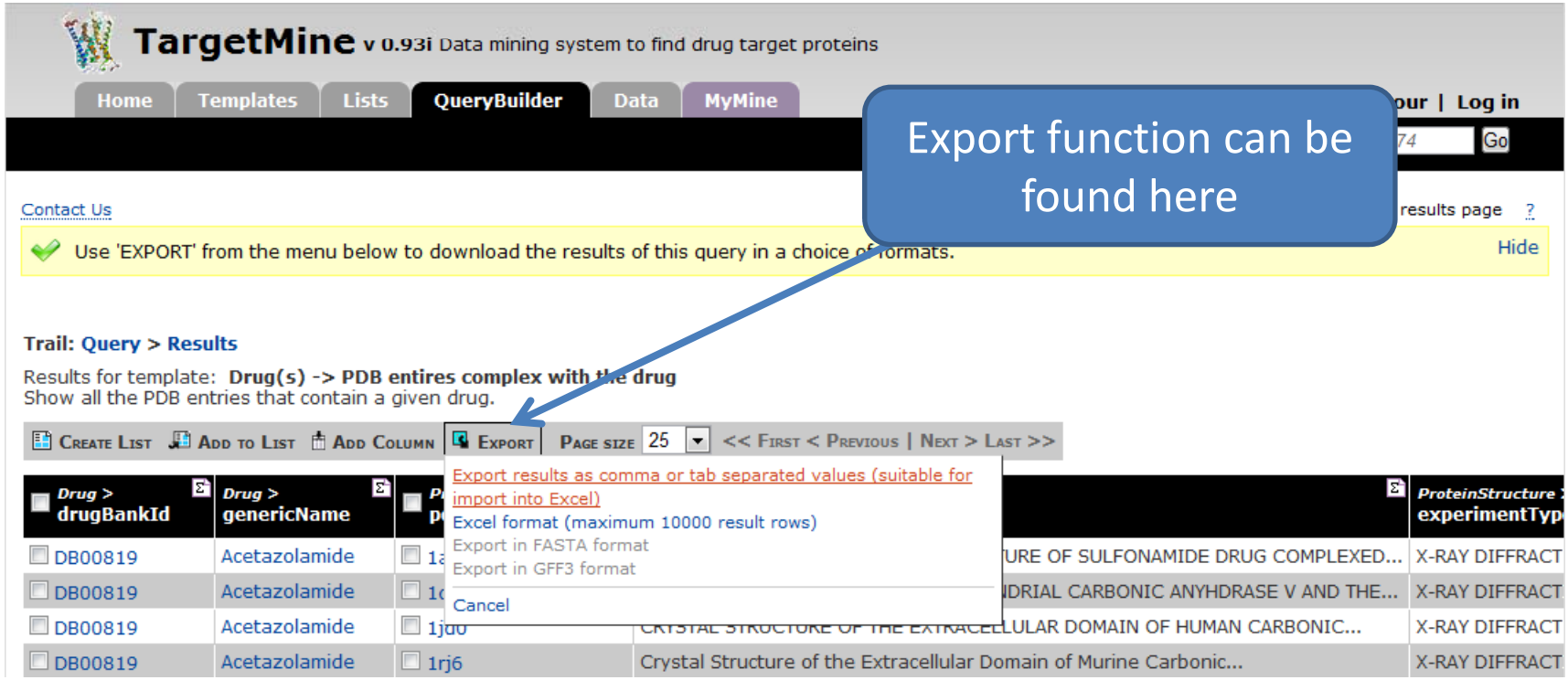
4遺伝子を絞り込んで、実験検証

siRNAによる遺伝子ノックダウンで、HCVの複製が減少

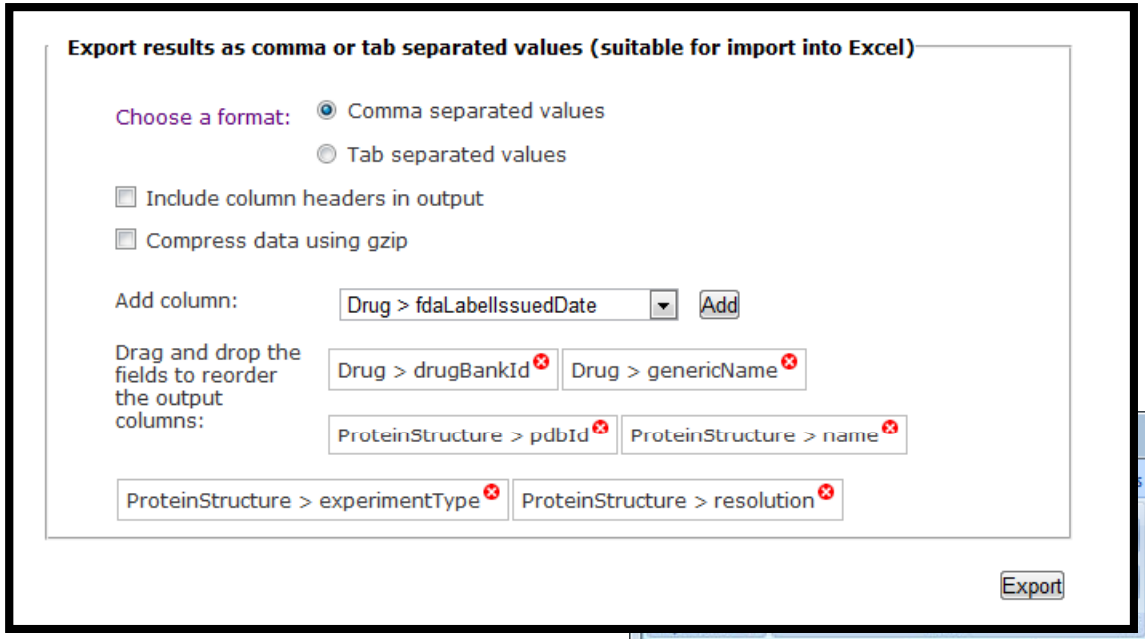


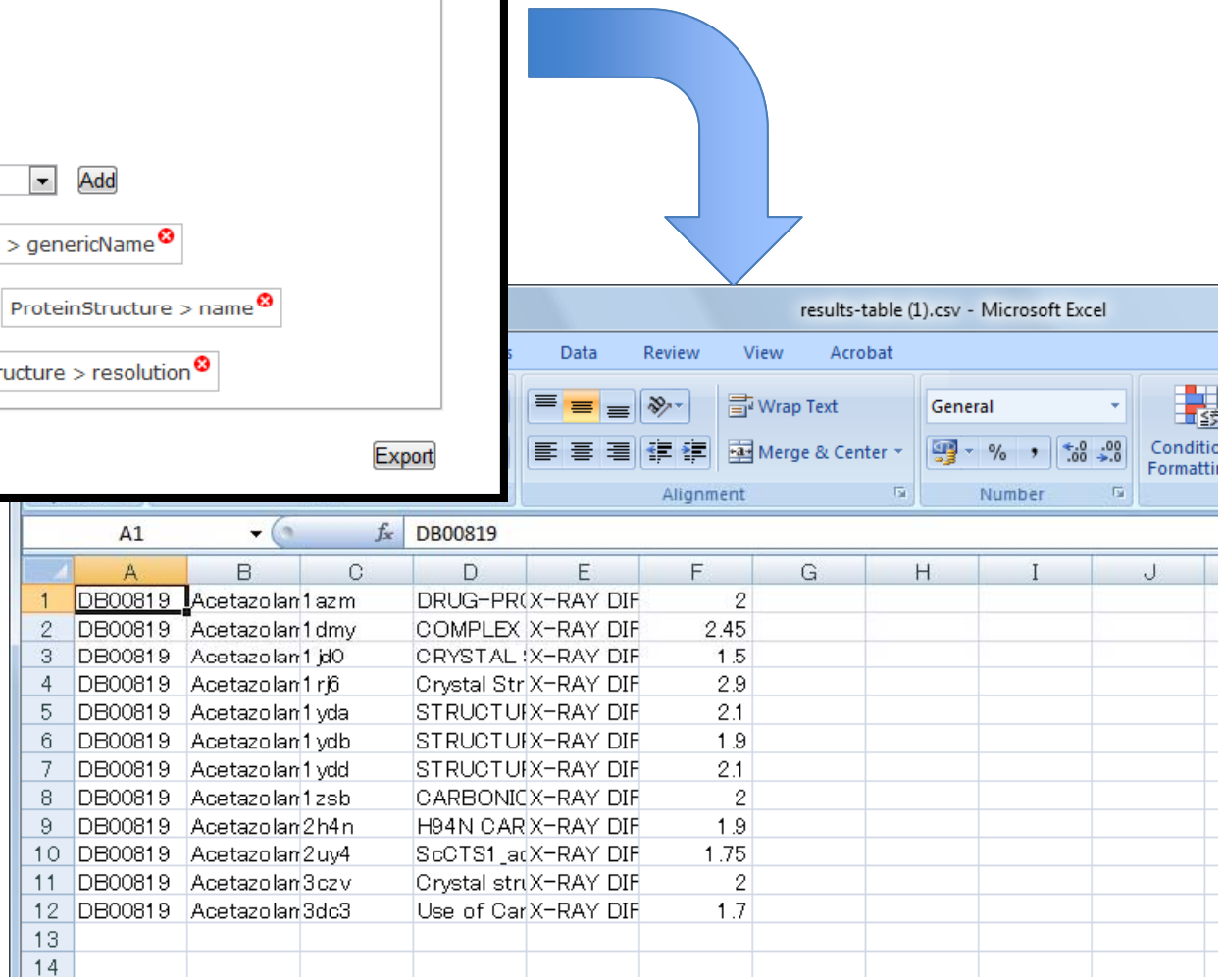
Tripathi et al., Mol. Biosyst. (2010)

Export function



Export function

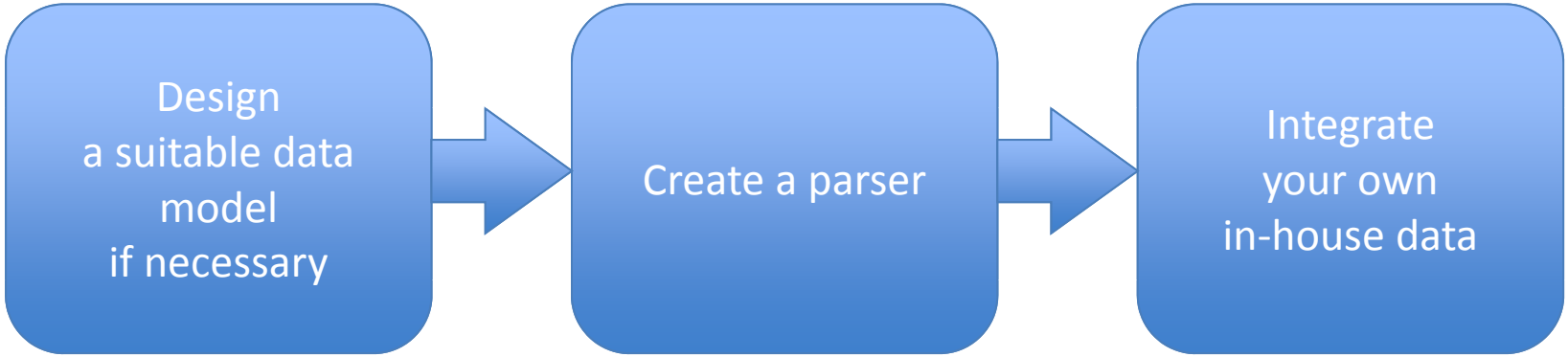




	A	B	C	D	E	F
1	DB00819	Acetazolam	1azm	DRUG-PR(	X-RAY DIF	2
2	DB00819	Acetazolam	1dmy	COMPLEX	X-RAY DIF	2.45
3	DB00819	Acetazolam	1jd0	CRYSTAL :	X-RAY DIF	1.5
4	DB00819	Acetazolam	1rj6	Crystal Str	X-RAY DIF	2.9
5	DB00819	Acetazolam	1yda	STRUCTUI	X-RAY DIF	2.1
6	DB00819	Acetazolam	1ydb	STRUCTUI	X-RAY DIF	1.9
7	DB00819	Acetazolam	1ydd	STRUCTUI	X-RAY DIF	2.1
8	DB00819	Acetazolam	1zsb	CARBONIC	X-RAY DIF	2
9	DB00819	Acetazolam	2h4n	H94N CAR	X-RAY DIF	1.9
10	DB00819	Acetazolam	2uy4	ScCTS1_ac	X-RAY DIF	1.75
11	DB00819	Acetazolam	3czv	Crystal stru	X-RAY DIF	2
12	DB00819	Acetazolam	3dc3	Use of Car	X-RAY DIF	1.7
13						
14						

Data customisation

- Integration of in-house experimental data or customised datasets is possible.



まとめ

- 各種生物学的概念がclass間のreferenceという形で結びつけられ、それらをmodel browserでたどる事で、高度な検索が可能。
- 既存の公共ツールでは行なうことができない、タンパク質立体構造や医薬品関係データ、転写因子とその作用遺伝子の関係等の情報の統合的検索。
- 多くの検索例をtemplatesとして提供。
- Enrichment widgetsを用いた、対話的な遺伝子絞り込み。
- 公共データベースに入っていない、独自のアノテーションや実験結果を組み込む仕組み。