

TargetMine チュートリアルと実習

統合データベース講習会：AJACS 駿河

2013 年 1 月 13 日

Copyright © 2012 The Mizuguchi laboratory (mizuguchilab.org), NIBIO

- このチュートリアルは TargetMine v0.99 に基づいて作成したものです。

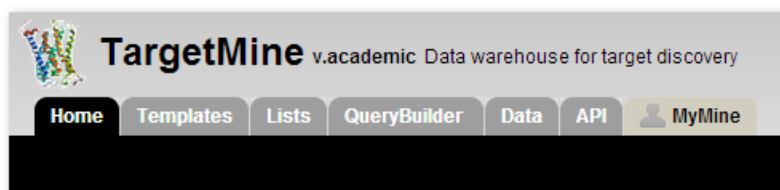
内容

- I. 簡単な検索とレポートページの見方について
- II. リストとリストの解析について
- III. テンプレートクエリ
- IV. クエリビルダー

まずは TargetMine ウェブサイトにアクセスしてください。

<http://targetmine.nibio.go.jp>

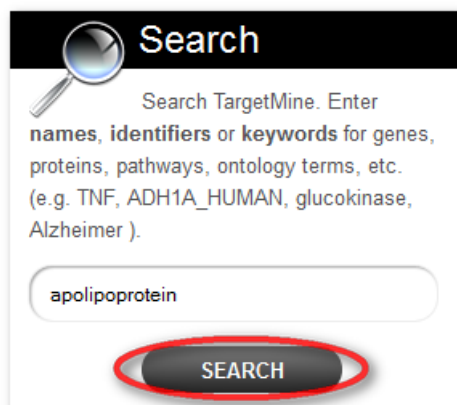
全ての機能は TargetMine のページ上部にあるツールバーから操作できます。



I. 簡単な検索とレポートページの見方について

検索機能はトップページ左の検索ボックス（下図左），または，他のどのページでも右上の検索ボックスからアクセスできます。検索語を入力すれば，（例：遺伝子名，ID，キーワード），TargetMine が持つデータに対して全文検索ができます。

それでは、‘apolipoprotein’ で検索してみてください。



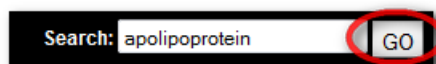
Search

Search TargetMine. Enter names, identifiers or keywords for genes, proteins, pathways, ontology terms, etc. (e.g. TNF, ADH1A_HUMAN, glucokinase, Alzheimer).

apolipoprotein

SEARCH

または

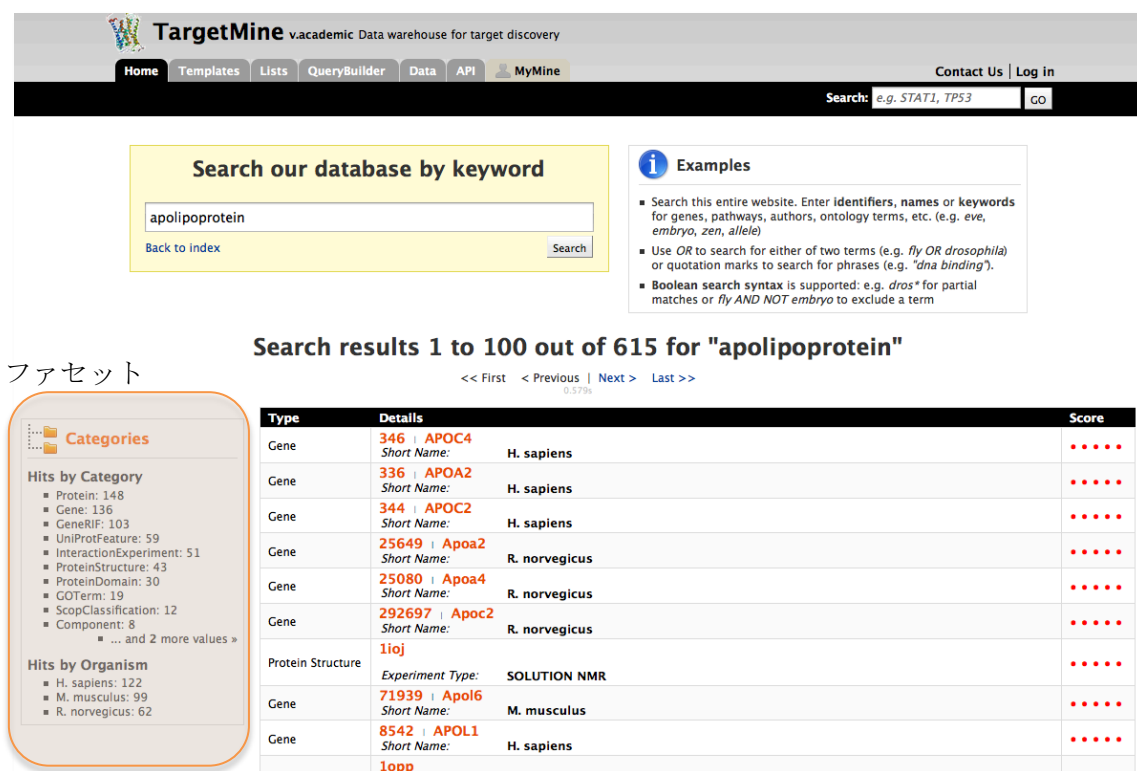


Search: apolipoprotein

GO

ヒント: ワイルドカード (アスタリスク) “*” を使うと部分一致検索ができます。

検索結果は、左側のファセット (Categoriesと書かれた分類) を用いて絞り込みができます。



TargetMine v.academic Data warehouse for target discovery

Home Templates Lists QueryBuilder Data API MyMine

Contact Us | Log in

Search: e.g. STAT1, TP53 GO

Search our database by keyword

apolipoprotein

Back to index Search

Examples

- Search this entire website. Enter identifiers, names or keywords for genes, pathways, authors, ontology terms, etc. (e.g. eve, embryo, zen, allele)
- Use OR to search for either of two terms (e.g. fly OR drosophila) or quotation marks to search for phrases (e.g. "dna binding").
- Boolean search syntax is supported: e.g. dros* for partial matches or fly AND NOT embryo to exclude a term

Search results 1 to 100 out of 615 for "apolipoprotein"

<< First < Previous | Next > Last >>

0.579s

Type	Details	Score
Gene	346 APOC4 Short Name: H. sapiens
Gene	336 APOA2 Short Name: H. sapiens
Gene	344 APOC2 Short Name: H. sapiens
Gene	25649 ApoA2 Short Name: R. norvegicus
Gene	25080 ApoA4 Short Name: R. norvegicus
Gene	292697 Apoc2 Short Name: R. norvegicus
Protein Structure	1ioj Experiment Type: SOLUTION NMR
Gene	71939 Apol6 Short Name: M. musculus
Gene	8542 APOL1 Short Name: H. sapiens
	1opp	

Categories

Hits by Category

- Protein: 148
- Gene: 136
- GeneRIF: 103
- UniProtFeature: 59
- InteractionExperiment: 51
- ProteinStructure: 43
- ProteinDomain: 30
- GOTerm: 19
- ScopClassification: 12
- Component: 8
- ... and 2 more values »

Hits by Organism

- H. sapiens: 122
- M. musculus: 99
- R. norvegicus: 62

- 各検索結果 (オレンジのハイライト) をクリックすると、その“オブジェクト” (特定の遺伝子やタンパク質など) のレポートページが表示されます。

TargetMine にはいろいろなタイプのオブジェクトが格納されていますが、レポートの表示は全て同じ形式です (下図)。

TargetMine v. academic Data warehouse for target discovery

Home Templates Lists QueryBuilder Data API MyMine Contact Us Log in

Search: e.g. STAT1, TP53 GO

Gene : **APOB** *H. sapiens*

NCBI Gene Number 338
Secondary Identifier HGNC:603

DB identifier ENSG00000084674
Name apolipoprotein B (including Ag(x) antigen)

Description apo B-100|apoB-100|apoB-48|apolipoprotein B-100|apolipoprotein B48|mutant Apo B 100
Type protein-coding

Quick Links: Summary Genomics Proteins Gene Ontology Pathways Interactions Disease Enzymes Other

1 Organism

Name	Taxon Id
Homo sapiens	9606

1 Proteins

DB identifier	Primary Accession	Name	Organism . Name	Length
APOB_HUMAN	P04114	Apolipoprotein B-100	Homo sapiens	4563 FASTA...

37 Synonyms

Value
ENSG00000084674
HGNC:603
APOB
FLDB
LDLQC4
NM_000384.2
NM_000384
ENST00000233242

Lists

This Gene isn't in any lists.
Upload a list.

Link to other Mines

YeastMine
No results found.

FlyMine
No results found.

modMine
D. melanogaster
Rfabg CG15828

metabolicMine
H. sapiens
APOB
M. musculus
Apob

RatMine
R. norvegicus
Apob

External Links

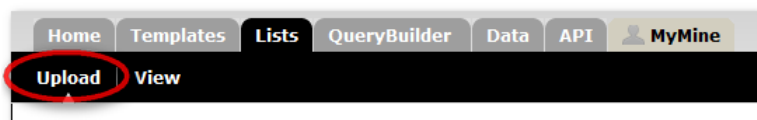
ensembl
ENSG00000084674
Entrez Gene: 338

- ページ上部にはオブジェクトの基本的な情報が表示され、下部にはオブジェクトのカテゴリ（遺伝子やタンパク質、化合物などの分類分け、TargetMine 内では“クラス”と呼ばれる）に応じて、より詳細な情報が表示されます。例えば、関連するクラスの情報や、クラス毎に予め用意された検索例（“テンプレート”と呼ばれる）など（これらの詳細は後述）。

II. リストとリストの解析について

リスト機能を使うと自分で指定したリスト（例：遺伝子／タンパク質のリスト）を作成することができます。リストは検索結果の一覧からも作成できますし、ID を使ってアップロードすることもできます。リストは検索の絞り込みや、クエリビルダー（詳細は以下に示します）のテンプレートとして、また、論理演算を使った検索 (Union, Intersect, Subtract) にも使用することができます。リストの検索やコピーも可能です。

リストページは上部メニューバーの“Lists”タブから開くことができます。“アップロード”タブはリストを作る専用のページで、ID の入力、またはファイルのアップロードからリストを作ることができます。



1つのリストには、ドロップダウンメニューから1つのタイプ（例：遺伝子，タンパク質）を指定してください（タイプの混在したリストは作れません）。オプションで生物種を指定することもできます。また、複数のタイプの ID を混在させることは構いません（例：遺伝子記号と遺伝子 ID）。

このチュートリアルでは、例として遺伝子のリストを作成し、解析します。‘List Analysis page’ で得られた解析結果は、クエリビルダーを用いて新たなクエリを作るために保存します。

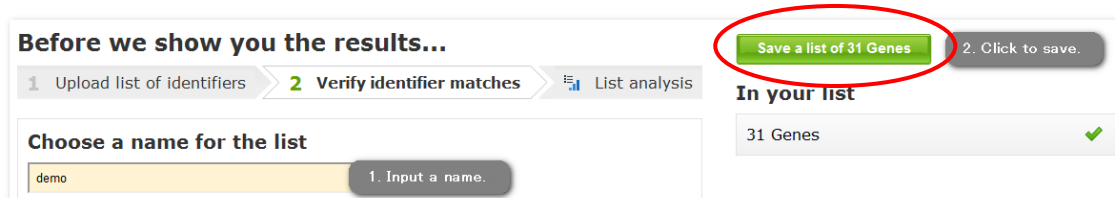
‘click to see an example’ をクリックしてください。クリックするとサンプルの ID 群が自動的に入力されます。

次は “Create List” ボタンをクリックしてください。

● リストの確認

“2 Verify identifier matches” では、入力された ID が TargetMine 中に存在するか、または曖昧さが残らないかなどのチェックがなされます。特にメッセージが表示されなければ、そのままリストに名前をつけます。任意の名前でもデフォルトの名前のままでも構いません。

今回はチュートリアルなので 'demo' と入力し、'Save a list of 31 Genes' ボタンをクリックしてください。



Before we show you the results...

1 Upload list of identifiers 2 Verify identifier matches List analysis

Choose a name for the list

demo 1. Input a name.

Save a list of 31 Genes 2. Click to save.

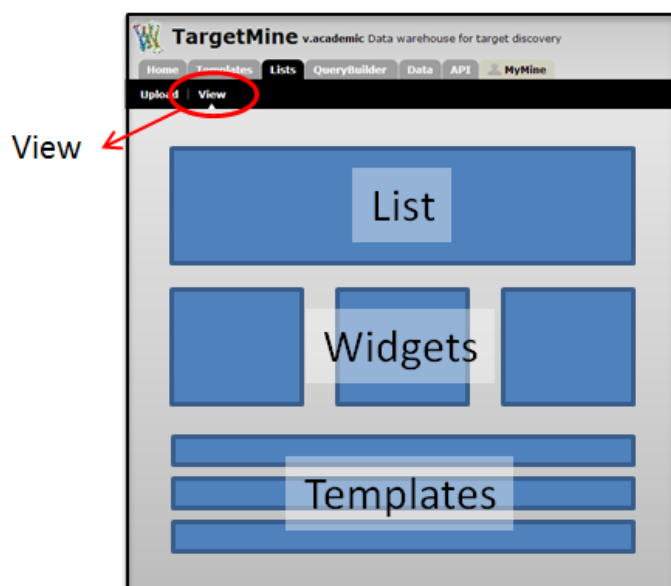
In your list

31 Genes ✓

● List Analysis Page

'List Analysis' ページでは、リスト中の各メンバーの詳細など多くの情報を得ることができます。リストが作成されると自動的に 'List Analysis' ページが表示されます。一度作ったリストについては、'Lists' タブの 'View' をクリックすると、いつでもこのページに戻ることができます（下図）。

List Analysis は3つのセクション (List, Widgets, Templates) に分かれています。



- 画面上部のボックス (List) には概要として、リスト中のメンバーの最初の 10 個が表示されます。そこには、リストの内容や情報、変換ツールやリンク先 (存在する場合のみ) が表示されます。これらの表の上部には Export/Edit オプションと検索ボックスがあります。

- 画面中部 (Widgets) にある一連のボックスからはリストについてのより多くの情報を得ることができます。例えば、このリストと特に強く関係付けられる KEGG パスウェイ、遺伝子オントロジー (GO) ターム、創薬や疾患に関するオントロジー、タンパク質間相互作用 (PPIs) や関連する薬の情報を得ることができます。

The screenshot displays two side-by-side widgets from the TargetMine interface.

Disease Ontology Enrichment: This widget shows a table of enriched disease terms. It includes a 'View' and 'Download' link. The table has columns for 'DO Term', 'p-Value', and a count. The data is as follows:

DO Term	p-Value	Count
pancreatitis [DOID:4989]	3.3624E-21	14
pancreas disease [DOID:26]	2.3565E-19	15
endocrine system disease [DOID:28]	5.6695E-19	19
alcoholic pancreatitis [DOID:4988]	6.0003E-18	9
gastrointestinal system disease [DOID:77]	1.1705E-12	18
acute pancreatitis [DOID:2913]	1.1857E-9	7
hypersensitivity reaction type II disease [DOID:417]	1.5571E-9	15
diabetes mellitus [DOID:9351]	1.7371E-9	15
periodontitis [DOID:824]	1.9324E-9	8
vascular disease [DOID:178]	2.33E-9	17

Associated Drugs: This widget shows a list of drugs associated with the genes in the list. It includes a 'View' and 'Download' link. The table has columns for 'Drug.drugBankId' and 'Targeted Genes'.

Drug.drugBankId	Targeted Genes
DB00143	5
DB00157	2
DB01213	2
DB01296	2
DB01834	2
DB01907	2
DB03431	2
DB04077	2
DB04184	2
DB04412	2
DB00005	1
DB00031	1
DB00051	1

- 画面下部(Templates)には、このリストについて、あらかじめ実行された各種検索（テンプレートクエリ）の結果が表示されます。

The screenshot shows the 'Template results for 'demo'' section. It lists four template queries with their respective result counts and a 'Show first 10 rows' link for each.

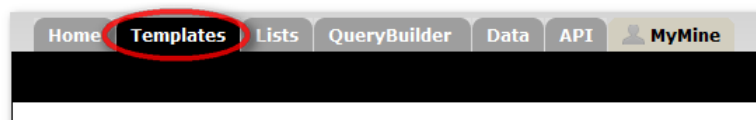
Template Query	Results
Gene --> Enzyme	18 results
Gene --> GO terms.	603 results
Gene --> Pathway	619 results
Gene --> Proteins.	39 results

III. テンプレートクエリ

多様な検索（解析）を素早く簡単に実行させるために、TargetMine では事前に定義されたテンプレートクエリのライブラリを持っています。一般に、TargetMine の検索は、あるクラスについて条件を指定して（例えば、“遺伝子名が TGFB1 である”）、関連するクラスのデータ（例えば、知られている全ての相互作用タンパク質）を表示する、という形式をとります。テンプレートライブラリは、そのような検索を多数用意したもので、検索例中の一部を変更することもできます。（TargetMine では、条件を指定することを「制限をかける」constrain という言葉で表現します。）

テンプレートクエリのアクセス方法は複数あります:

- テンプレートページには TargetMine のトップページ, もしくは上部のメニューバーのタブからアクセスしてテンプレートの検索や閲覧をすることが可能です。
- テンプレートはレポートページと List Analysis Page にも表示されます。



例としてテンプレートクエリに ‘transcription’ と入力し, 検索します。

Filter: transcription Filter: -- all categories -- Reset

☐ Gene(s) --> Downstream Transcription Target genes
Given a gene or a list of genes, retrieve all downstream transcription target genes from the AMADEUS and ORegAnno compiled TF-Target gene relations

☒ Gene(s) --> Upstream Transcription Factors
Given a gene or a list of genes, retrieve all upstream regulatory genes (transcription factors) from the AMADEUS and ORegAnno compiled TF-Target gene relations

☐ Gene(s) --> Gene Regulatory Associations
Given a gene or a list of genes, retrieve all upstream regulators (regulatory transcription factors) and/or downstream targets from the AMADEUS and ORegAnno compiled TF-Target relations

この検索は, 与えられた遺伝子または遺伝子リストについて, 上流の転写因子 (AMADEUS または OregAnno データベースに収納されている情報) をすべて取得するというものです。

オブジェクト (特定の遺伝子やタンパク質など) またはリストを指定する (「特定のオブジェクトまたはリストに制限する」) という型のテンプレートクエリでは, オブジェクトの ID を入力するボックスと共に, リストを制限するためのオプションが提供されます。

そして, テンプレートは ‘Gene(s) --> Upstream Transcription Factors’ をクリックしてください。

Gene(s) → Upstream Transcription Factors
Given a gene or a list of genes, retrieve all upstream regulatory genes (transcription factors) from the AMADEUS and ORegAnno compiled TF-Target gene relations

Gene
LOOKUP: 10121 for Organism: H. sapiens

☒ constrain to be IN saved Gene list demo

Show Results 2. Click to show results

Edit Query

チェックボックスをクリックし, 先ほど保存したサンプルリスト “demo” を選択してください。次に, “Show Results” ボタンを押し, 検索を実行します。

その結果、先ほど作ったリストの遺伝子をターゲットとする転写因子が表示されます。

結果の一覧から新しいリストを作ることもできます。この検索結果から **List Analysis** ページや他のテンプレートに使用するための転写因子のリストを保存できます。

“Source” から全ての遺伝子を選択してください。

Click on the checkbox to select all genes

Results for template
Total rows: 26

CREATE LIST ADD TO LIST ADD COLUMN CSV FASTA GFF3 PAGE SIZE 25 << FIRST < PREVIOUS | NEXT > LAST >>

Source	Source	Source	Target	Target	Target		
		Name	NCBI Gene Number	Symbol	Name		
<input checked="" type="checkbox"/>	3172	HNF4A	hepatocyte nuclear factor 4, alpha	<input type="checkbox"/>	125	ADH1B	alcohol dehydrogenase (class I)
<input checked="" type="checkbox"/>	3175	ONECUT1	one cut homeobox 1	<input type="checkbox"/>	125	ADH1B	alcohol dehydrogenase (class I)
<input checked="" type="checkbox"/>	6927	HNF1A	HNF1 homeobox A	<input type="checkbox"/>	125	ADH1B	alcohol dehydrogenase (class I)
<input checked="" type="checkbox"/>	3172	HNF4A	hepatocyte nuclear factor 4, alpha	<input type="checkbox"/>	217	ALDH2	aldehyde dehydrogenase (class 2)
<input checked="" type="checkbox"/>	1385	CREB1	cAMP responsive element binding protein 1	<input type="checkbox"/>	3107	HLA-C	major histocompatibility class I, C
<input checked="" type="checkbox"/>	1869	E2F1	E2F transcription factor 1	<input type="checkbox"/>	3265	HRAS	v-Ha-ras H-Ras
<input checked="" type="checkbox"/>	7157	TP53	tumor protein p53	<input type="checkbox"/>	3265	HRAS	v-Ha-ras H-Ras
<input checked="" type="checkbox"/>	3175	ONECUT1	one cut homeobox 1	<input type="checkbox"/>	3304	HSPA1B	heat shock protein 70
<input checked="" type="checkbox"/>	3297	HSF1	heat shock transcription factor 1	<input type="checkbox"/>	3304	HSPA1B	heat shock protein 70

“CREATE LIST” タブをクリックし、名前をつけてリストを保存してください。(例: ‘upstream_TFs’)

1. Click on the tab 'CREATE LIST'

(WITH SELECTED ITEMS) IN A NEW LIST NAMED
upstream_TFs Save selected

2. Click 'Save selected' to save

CANCEL

一群のオブジェクト全体の情報を得るために、リストの使用は大変効果的です。この方法を使うと以下の様なリストも作成できます。

- 選択した遺伝子（産物）の相互作用タンパク質
- これらの遺伝子の疾患オントロジー (DO) アノテーション


他にも色々試してみてください。

IV. クエリビルダー

ここからは上級者向け応用編です。‘クエリビルダー’を使用すると、以下のようなことができます。

- TargetMine 上の複数のデータ・ソースを任意に組み合わせた新しいクエリの構築

- 既存のテンプレートクエリを修正することで出力を変え、**Constraints** を加える。
以下のような例から実際に操作をしてみてください。

 **Gene(s) → Upstream Transcription Factors**

Given a gene or a list of genes, retrieve all upstream regulatory genes (transcription factors) from the AMADEUS and ORegAnno compiled TF-Target gene relations

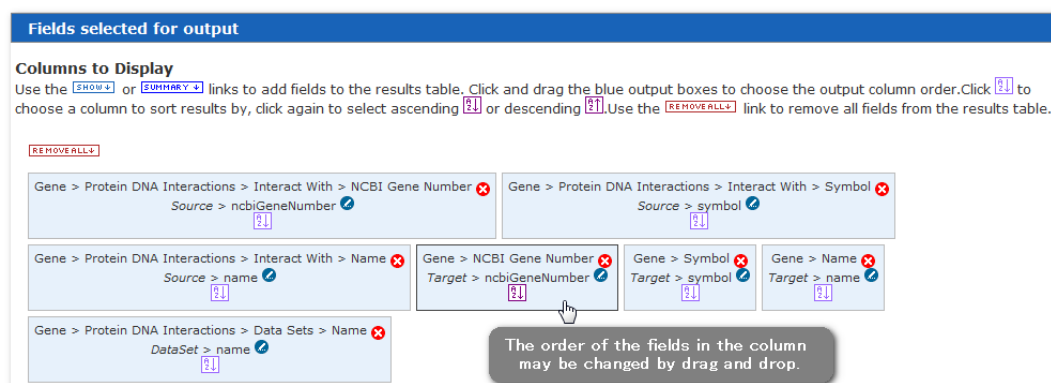
Gene

LOOKUP: for Organism:

☐ constrain to be saved Gene list

関連するクラスと属性を見るためには  をクリックしてください。  タブは出力にフィールドを追加するために使用します。  タブでは絞り込み（フィルタ）を追加することができます。

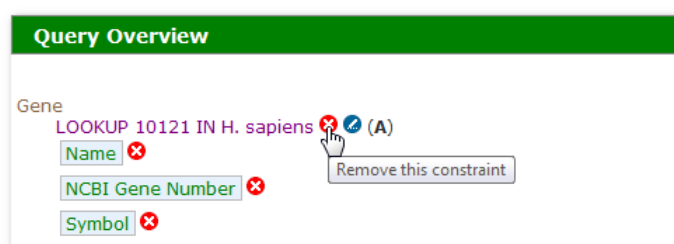
Model Browser の下には、ユーザーが選択したフィールドが **‘Fields selected for output’** として表示されます。それぞれのフィールドは、検索結果ページの各列に対応します。フィールドの順序はドラッグ・アンド・ドロップで変更することができます。



右側の **‘Query Overview’** では、クエリに適用された絞り込み条件を示しています。その内の1つは、**‘Gene’** で **‘LOOKUP 10121 IN H. sapiens’** と書かれているものです。これは、ID 10121 を持つヒトの遺伝子のみにクエリが絞り込まれた（制限された）ということを示しています。

もし、この遺伝子をターゲットにするものだけでなく、全ての転写因子について調べたくなった時には、絞り込み条件を外して、もう一度クエリを走らせれば解除できます。

絞り込み条件を解除するためには、**‘LOOKUP 10121 IN H. sapiens’** の隣にある **✖** ボタンをクリックしてください。

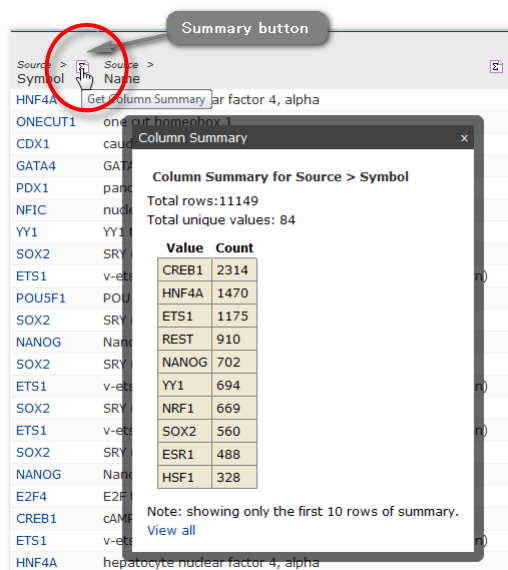


そして、**‘Show results’** をクリックしてください。



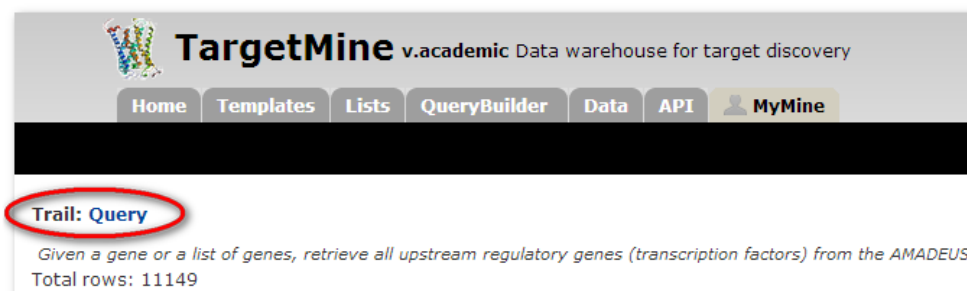
そうすると、（ TargetMine に格納されている）全ての転写因子とそのターゲット遺伝子を得ることができます。

📌 サマリーボタンをクリックするとその列に関する概要の情報を得ることができます。



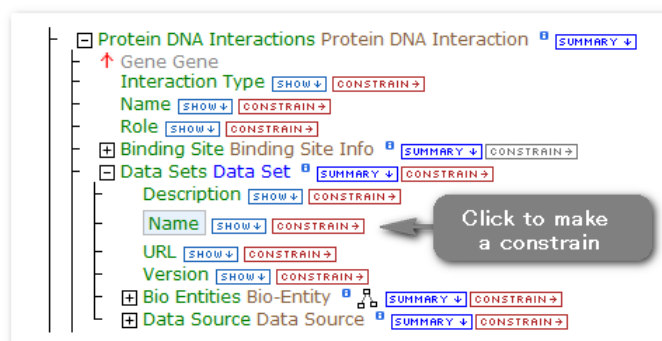
もう少し他の操作もしてみましょう。

‘Query’ をクリックして、クエリビルダーに戻ります。

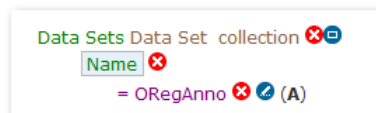


それでは、“Gene → Protein DNA Interactions → Data Sets → Name” の順でパスを実行してみてください。次に Name の横の **CONSTRAIN** タブをクリックするとポップアップが表示されます。‘ORegAnno’ をドロップダウンリストから選択し ‘Add to query’ ボタンをクリックします。

そうするとデータセットに ‘ORegAnno’ データベースの絞り込みをかけることができます。



Model browser



Query overview

‘Show results’ をクリックし、もう一度実行してみましょう。

Given a gene or a list of genes, retrieve all upstream regulatory genes (transcription factors) from the ANADeUS and ORegAnno compiled TF-Target gene relations
Total rows: 1175

Source	Source	Source	Target	Target	Target	Target	Target
Name	Name	Name	NCBI Gene Number	Symbol	Symbol	Symbol	Symbol
1044	CDX1	caudal type homeobox 1	100	ADA	adenosine deaminase	ORegAnno	
2628	GATA4	GATA binding protein 4	100	ADA	adenosine deaminase	ORegAnno	
3851	PDX1	pancreatic and duodenal homeobox 1	100	ADA	adenosine deaminase	ORegAnno	
4782	NFIC	nuclear factor I/C (CCAAT-binding transcription factor)	100	ADA	adenosine deaminase	ORegAnno	
7528	YY1	YY1 transcription factor	100	ADA	adenosine deaminase	ORegAnno	
5978	RE1	RE1-silencing transcription factor	100130418	CECR7	cat eye syndrome chromosome region, candidate 7 (non-protein coding)	ORegAnno	
5978	RE1	RE1-silencing transcription factor	100130873	CARM1P1	coactivator-associated arginine methyltransferase 1 pseudogene 1	ORegAnno	
5978	RE1	RE1-silencing transcription factor	1002	CDH4	cadherin 4, type 1, R-cadherin (retinal)	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10021	HCN4	hyperpolarization activated cyclic nucleotide-gated potassium channel 4	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10044	SH2D3C	SH2 domain containing 3C	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10135	NAMPT	nicotinamide phosphoribosyltransferase	ORegAnno	
5978	RE1	RE1-silencing transcription factor	1014	CDH16	cadherin 16, KSP-cadherin	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10143	CLEC3A	C-type lectin domain family 3, member A	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10218	ANGPTL7	angiopoietin-like 7	ORegAnno	
4782	NFIC	nuclear factor I/C (CCAAT-binding transcription factor)	1026	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	ORegAnno	
6667	SP1	Sp1 transcription factor	1026	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	ORegAnno	
6670	SP3	Sp3 transcription factor	1026	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	ORegAnno	
6772	STAT1	signal transducer and activator of transcription 1, 91kDa	103	ADAR	adenosine deaminase, RNA-specific	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10307	APBB3	amyloid beta (A4) precursor protein-binding, family B, member 3	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10368	CACNG3	calcium channel, voltage-dependent, gamma subunit 3	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10369	CACNG2	calcium channel, voltage-dependent, gamma subunit 2	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10381	TUBB3	tubulin, beta 3 class III	ORegAnno	
5978	RE1	RE1-silencing transcription factor	10394	PRG3	proteoglycan 3	ORegAnno	
7157	TP53	tumor protein p53	104	ADARB1	adenosine deaminase, RNA-specific, B1	ORegAnno	
6772	STAT1	signal transducer and activator of transcription 1, 91kDa	10410	IFITM3	interferon induced transmembrane protein 3	ORegAnno	

<< First < Previous | Next > Last >> | Displaying rows 1 to 25 | Total rows: 1175

検索結果には、ORegAnno データベースから取得した転写因子の情報のみが表示されます。

このようにクエリビルダーを使うと細かく条件を指定して調べることが出来るので、ぜひ利用してみてください。