
BovineMine Documentation

Release 1.0

Deepak Unni, Aditi Tayal, Colin Diesh, Christine Elsik, Darren Hag

Jan 26, 2018

Contents

1	Tutorial	3
1.1	Overview	3
1.2	Navigation and Searching in BovineMine	4
1.3	Report Page	18
1.4	BLAST	20
1.5	Genomic Region Search	20
1.6	Lists	22
1.7	MyMine	26
1.8	API	27
1.9	Data Sources	28
2	Indices and tables	29

BovineMine is a data mining warehouse for *Bos taurus* genome (UMD3.1.1 assembly), powered by InterMine.

Link to main site: <http://bovinegenome.org/bovinemine>

Link to our YouTube Channel: [Bovine Genome Database on YouTube](#)

Presentations:

- [Generating Customized Integrated Functional Annotation Datasets with BovineMine - ISAG 2017](#)

BovineMine integrates datasets from various data sources such as NCBI RefSeq, ENSEMBL, UniProt, GO annotations, EnsemblCompara, Homologene, Reactome. Additionally, BovineMine also integrates datasets from <http://www.bovinegenome.org>.

Link to the available datasets in BovineMine: <http://bovinegenome.org/bovinemine/dataCategories.do>

The Bovine Genome Database is supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613689, and has been supported by grants 2007-35616-17882 and 2010-65205-20407 from the USDA National Institute of Food and Agriculture. BGD is hosted at the University of Missouri. If you have comments or if you wish to report a problem, please contact the Bovine Genome Database Administrator.

Contents:

BovineMine is powered by InterMine and provides a user-friendly way to access genomic, proteomic, interaction and literature data. This tutorial is aimed at giving users an introduction to different features of BovineMine.

1.1 Overview

This section provides a brief overview of the layout for BovineMine.



The navigation panel highlights different functionalities of BovineMine based on what the user is searching for.

1. **Home** - The homepage for BovineMine
2. **Templates** - Shows the list of templates that users can use based on the nature of their query. Each template is associated with a description which describes what input is expected from the user and what will be the output.
3. **Lists** - The list page offers users a form which they can use to submit or upload a list of genes. Once the list is saved, users can perform enrichment analyses of these genes and export the results.
4. **QueryBuilder** - The QueryBuilder is an interface that provides the user to directly interact with the BovineMine data model. Some users might find the QueryBuilder difficult at first, but this tutorial is aimed at familiarizing users on the functionality of QueryBuilder, how to build a custom template and save their own templates.
5. **BLAST** - The BLAST page is where users can search their input sequence(s) against the several databases provided.
6. **Regions** - The Genomic Region Search tool provides a form where users can enter a series of genomic coordinates, specify flanking regions (if any) and fetch all features that fall within the given interval. The result can be exported, saved as a list for further analyses.

7. **Data Sources** - The Data Sources page provides a summary of all the data that is loaded into BovineMine. It lists a combination of datasets from other databases as well as in-house prepared datasets.
8. **API** - The API provides users an option for programmatically accessing BovineMine by using the InterMine API.
9. **MyMine** - The MyMine serves as a portal for users (those who are logged in) for account management. Users can access their saved templates, most recent queries and saved lists.

1.2 Navigation and Searching in BovineMine

There are three ways for users to query BovineMine.

1.2.1 a. 'Search' Box

The Search box is located on the BovineMine home page.

BovineMine v1.0 An integrated data warehouse and data mining interface for *Bos taurus* genome

Home Templates Lists QueryBuilder BLAST Regions Data Sources Help API MyMine

Search: e.g ATP sy

Search

Search BovineMine. Enter names, identifiers or keywords for genes, proteins, ontology terms, authors, etc. (e.g. *RPL24*, RCAN1, BTN1A1)

NM_001040473.2, 282087, BT13969

SEARCH

Analyse

Enter a list of identifiers.

Gene

e.g. 515523, BT18560, BTG1, ENSBTAG00000006858

advanced

ANALYSE

Welcome Back

BovineMine integrates ma
data for *Bos taurus* (Taxo
can run flexible queries, e
and analyse lists of data.

TUTORIAL

There are two locations (marked) where users can input the name, Gene ID, or keywords related to their gene of interest. The search box is not just limited to Gene but a variety of features such as Transcript, Exons, Interaction, Pathways, GO Annotations and more. Users can make use of wildcard character '*' to get all results that match a particular prefix.

Lets consider gene **rpl24** as an example. Enter rpl24 in the search box and click *Search*. The results page is tabulated and shows a summary about your query, as shown below.

Search our database by keyword

[Back to index](#)

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 4 out of 4 for *rpl24*

0.263s

Type	Details
Gene	282048 BT14245 RPL24 <i>Length:</i> 5503 <i>Chromosome Location:</i> GK000001.2: 46415219-46420721 <i>Organism:</i> B. taurus
Gene	ENSBTAG00000013461 ENSBTAG00000013461 RPL24 <i>Source:</i> Ensembl Data Set <i>Length:</i> 5499 <i>Chromosome Location:</i> GK000001.2: 46415223-46420721 <i>Organism:</i> B. taurus
mRNA	NM_174455.3 <i>Length:</i> 5503 <i>Chromosome Location:</i> GK000001.2: 46415219-46420721

Categories

Hits by Category

- Gene: 2
- mRNA: 2

Hits by Organism

- B. taurus: 4

Users can filter the results based on **Category** and **Organism**. The score column in the result table indicates the similarity of your query to each of the result fetched by BovineMine.

The results page can also be converted to a list. To enable this feature click on **Gene** in **Hits by Category**.

Search our database by keyword

[Back to index](#) - or -

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 2 out of 2 for *rpl24*

Category restricted to Gene ✖

7.981s

<input checked="" type="checkbox"/> Type	Details
<input type="checkbox"/> Gene	282048 BT14245 RPL24 <i>Length:</i> 5503 <i>Chromosome Location:</i> GK000001.2: 46415219-46420721 <i>Organism:</i> B. taurus
<input type="checkbox"/> Gene	ENSBTAG00000013461 ENSBTAG00000013461 RPL24 <i>Source:</i> Ensembl Data Set <i>Length:</i> 5499 <i>Chromosome Location:</i> GK000001.2: 46415223-46420721 <i>Organism:</i> B. taurus

Categories

Category: Gene
« show all

Hits by Organism

- B. taurus: 2

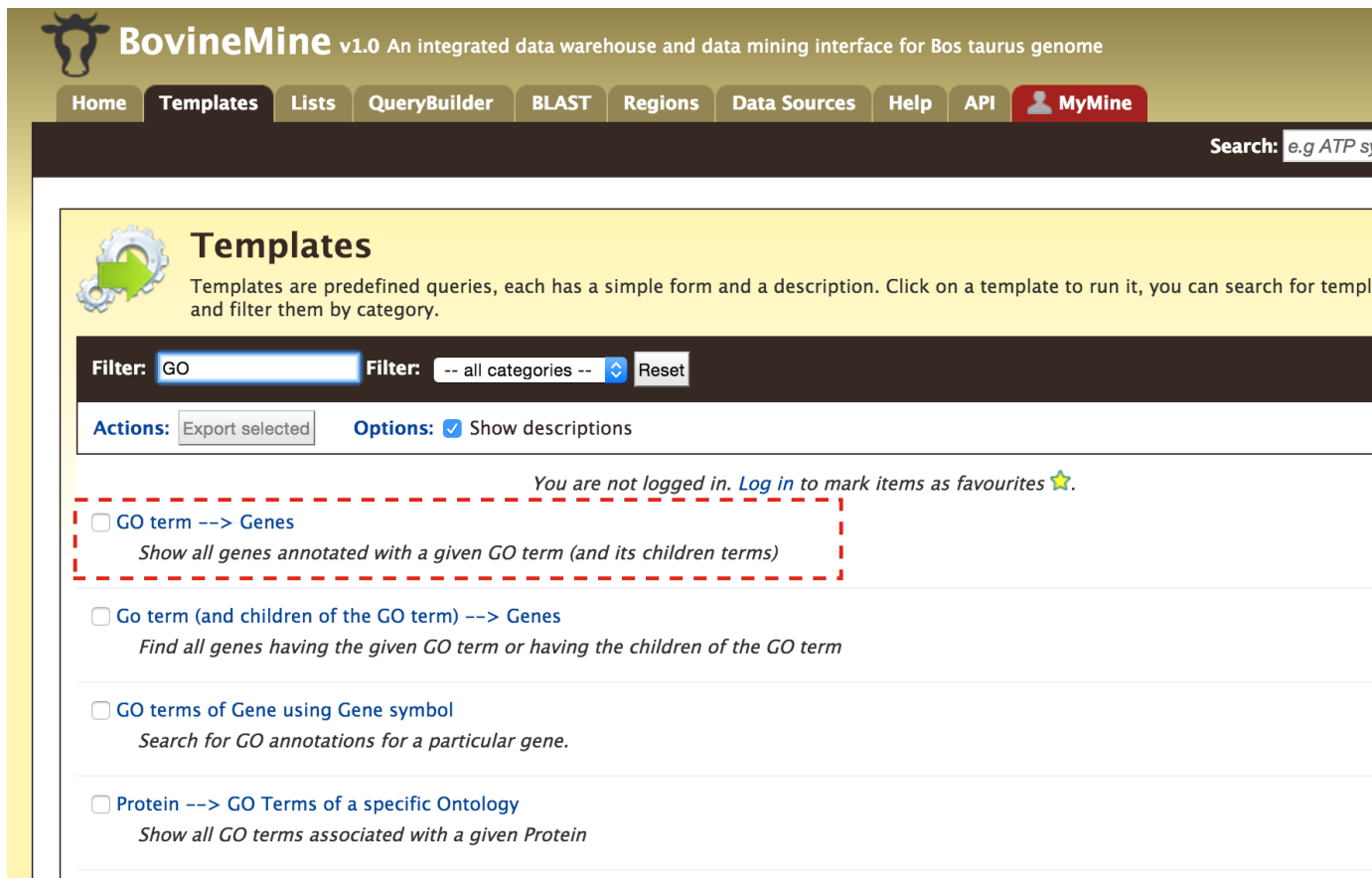
CREATE LIST

This will filter the results for the feature type **Gene** and checkboxes will be available for users to select genes they would like to add to their list. Once the genes are selected, click on **CREATE LIST**.

1.2.2 b. Templates

Apart from the search box, users can make use of predefined templates by clicking on the `Templates` tab.

Users can choose from a list of templates based on the nature of their query.



BovineMine v1.0 An integrated data warehouse and data mining interface for *Bos taurus* genome

Home **Templates** Lists QueryBuilder BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

Templates

Templates are predefined queries, each has a simple form and a description. Click on a template to run it, you can search for templates and filter them by category.


Filter: Filter: -- all categories --

Actions: Options: ☒ Show descriptions

You are not logged in. [Log in](#) to mark items as favourites ☆.

- ☐ **GO term --> Genes**
Show all genes annotated with a given GO term (and its children terms)
- ☐ **Go term (and children of the GO term) --> Genes**
Find all genes having the given GO term or having the children of the GO term
- ☐ **GO terms of Gene using Gene symbol**
Search for GO annotations for a particular gene.
- ☐ **Protein --> GO Terms of a specific Ontology**
Show all GO terms associated with a given Protein


As an example, let's click on `GO terms --> Genes` template. The goal of this template is to query BovineMine to find all genes having a given GO term.



BovineMine v1.0 An integrated data warehouse and data mining interface for Bos taurus genome

Home Templates Lists QueryBuilder BLAST Regions Data Sources Help API MyMine

Search:



GO term → Genes

Show all genes annotated with a given GO term (and its children terms)

GO Term > Name

=

Organism > Name

=

GO Evidence Code > Code

optional
ON | OFF

=

[Show Results](#)

[web service URL](#)
[Perl](#) | [Python](#) | [Ruby](#) | [Java \[help\]](#)
[export XML](#)

Users can also specify a filter for GO evidence code. Lets search for the GO term DNA binding with GO evidence code IEA.

Trail: Query
GO term → **Genes**
 Show all genes annotated with a given GO term (and its children terms)

Manage Columns 4 Filters

1 Create / Add to List 2 Get Code

4 Showing 1 to 25 of 479 rows Rows per page: 25

DB identifier	Secondary Identifier	Symbol	Code Code	Ontology Term Identifier	Ontology Term Name
100124507	BT16412	LIG1	IEA	GO:0003677	DNA binding
100124518	BT21014	ZNF420	IEA	GO:0003677	DNA binding
100125238	NO VALUE	ERCC2	IEA	GO:0003677	DNA binding
100125299	BT18906	EME1	IEA	GO:0003677	DNA binding
100125315	BT28215	TEAD2	IEA	GO:0003677	DNA binding
100125581	BT29166	ZNF75D	IEA	GO:0003677	DNA binding
100139498	BT25769	MCM9	IEA	GO:0003677	DNA binding
100139934	NO VALUE	ZFP90	IEA	GO:0003677	DNA binding
100140288	NO VALUE	NO VALUE	IEA	GO:0003677	DNA binding
100140434	BT20398	ZKSCAN1	IEA	GO:0003677	DNA binding

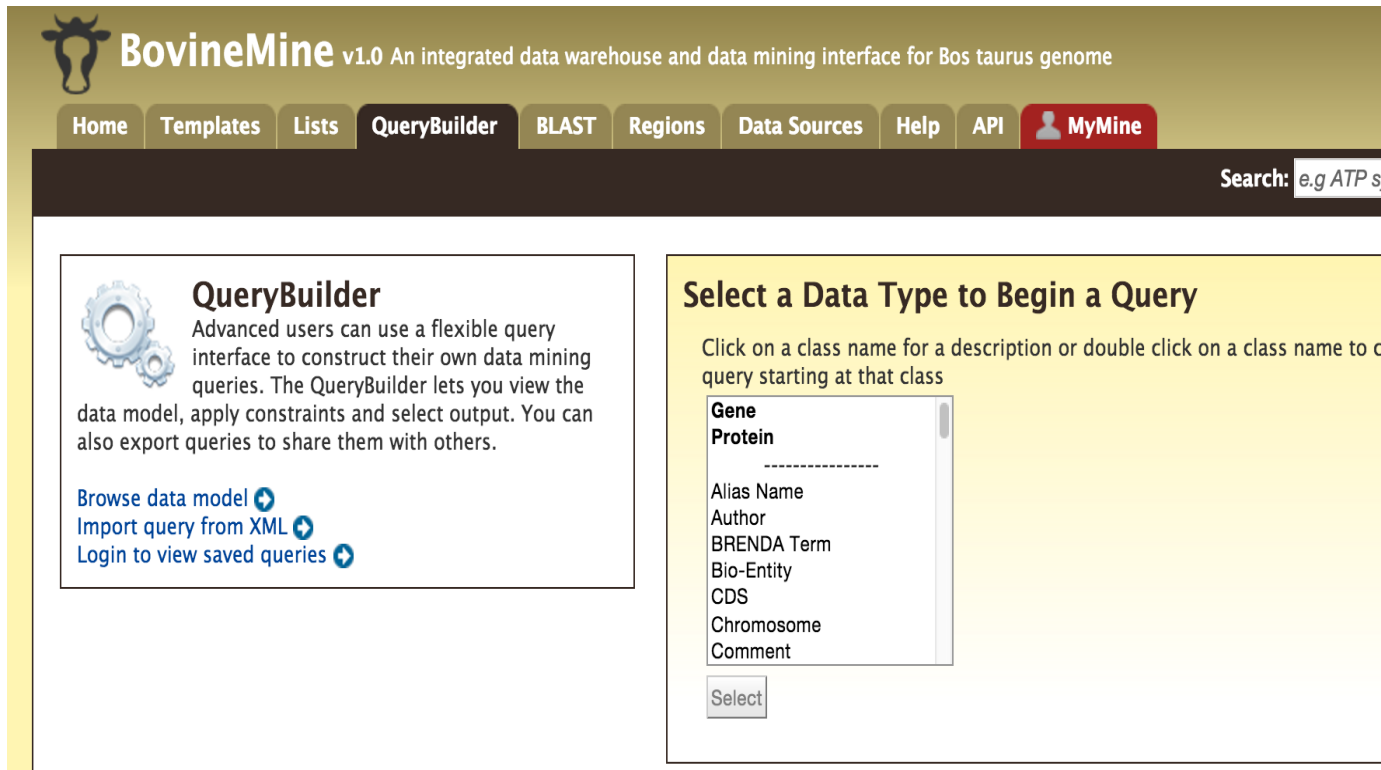
The result page shows all the genes that have been annotated with the Gene Ontology term DNA binding.

Users can,

1. create or add these genes to a list, by clicking on Create/Add to List (Box 1) to perform further analyses.
2. get the code for the query in Perl, Python, Java, Ruby, JavaScript or XML by clicking on Get Code (Box 2).
3. download the search results, by clicking on Download (Box 3), as tab-delimited, comma-separated values, XML or JSON. If the results are genomic features, which is true for the current example, then users can download the results in GFF3 and BED format.
4. customize the layout of the result page by clicking on Manage Columns (Box 4).

1.2.3 c. QueryBuilder

We have provided templates suitable for several common types of queries but if users need more fidelity in their search they can make use of the QueryBuilder. The possibilities of queries using the QueryBuilder are endless. You can format the output the way you want and constrain your queries to perform complex search operations.



The screenshot shows the BovineMine v1.0 web interface. The header includes the BovineMine logo and the text "v1.0 An integrated data warehouse and data mining interface for Bos taurus genome". Below the header is a navigation bar with tabs: Home, Templates, Lists, QueryBuilder (selected), BLAST, Regions, Data Sources, Help, API, and MyMine. A search bar on the right contains the text "Search: e.g ATP s". The main content area is divided into two panels. The left panel, titled "QueryBuilder", features a gear icon and text explaining that advanced users can use a flexible query interface to construct their own data mining queries. It also lists three links: "Browse data model", "Import query from XML", and "Login to view saved queries". The right panel, titled "Select a Data Type to Begin a Query", contains instructions to click on a class name for a description or double click to start a query. It displays a list of data types: Gene, Protein, Alias Name, Author, BRENDA Term, Bio-Entity, CDS, Chromosome, and Comment. A "Select" button is located at the bottom of the list.

BovineMine v1.0 An integrated data warehouse and data mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

QueryBuilder

Advanced users can use a flexible query interface to construct their own data mining queries. The QueryBuilder lets you view the data model, apply constraints and select output. You can also export queries to share them with others.

[Browse data model](#) [Import query from XML](#) [Login to view saved queries](#)

Select a Data Type to Begin a Query

Click on a class name for a description or double click on a class name to start a query starting at that class

Gene
Protein

Alias Name
Author
BRENDA Term
Bio-Entity
CDS
Chromosome
Comment

Select

First let's select *Gene* as a Data Type in the QueryBuilder. Then click on *Select*. This will take you to a Model browser where you can select the attributes, for the feature class 'Gene', which you would want to be shown in your results.

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

Model browser

Browse through the classes and attributes. Click on [SUMMARY](#) links to add summary of fields to the results table or on [SHOW](#) links to add individual fields to the results. Use [CONSTRAIN](#) links to constrain a value in the query.

Gene [SUMMARY](#) [CONSTRAIN](#)

- Length Integer [SHOW](#) [CONSTRAIN](#)
- Gene Identifier [SHOW](#) [CONSTRAIN](#)
- BGD Identifier [SHOW](#) [CONSTRAIN](#)
- Source [SHOW](#) [CONSTRAIN](#)
- Status [SHOW](#) [CONSTRAIN](#)
- Symbol [SHOW](#) [CONSTRAIN](#)
- Alias Alias Name [SUMMARY](#) [CONSTRAIN](#)
- Chromosome Chromosome [SUMMARY](#) [CONSTRAIN](#)
- Chromosome Location Location [SUMMARY](#) [CONSTRAIN](#)
- Data Sets Data Set [SUMMARY](#) [CONSTRAIN](#)
- Duplicate Entities Duplicate Entity [SUMMARY](#) [CONSTRAIN](#)
- GO Annotation GO Annotation [SUMMARY](#) [CONSTRAIN](#)
- Locations Location [SUMMARY](#) [CONSTRAIN](#)
- Ontology Annotations Ontology Annotation [SUMMARY](#) [CONSTRAIN](#)
- Organism Organism [SUMMARY](#) [CONSTRAIN](#)
- Proteins Protein [SUMMARY](#) [CONSTRAIN](#)
- Sequence Ontology Term SO Term [SUMMARY](#) [CONSTRAIN](#)
- Synonyms Synonym [SUMMARY](#) [CONSTRAIN](#)
- Transcripts Transcript [SUMMARY](#) [CONSTRAIN](#)

☐ Show empty fields

Query Overview

no fields constrained

Constraint logic:

no constraints

Lets consider three scenarios for using the QueryBuilder,

i. Querying for Protein Coding genes

Click on Show tab for the attributes Gene Identifier, BGD Identifier, Symbol and Status. This tells the QueryBuilder to show the Gene ID, Bovine Genome Database (BGD) ID and the Symbol for Gene.

Then click on Constrain tab for the attribute Status.

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP sy

Model browser

links to add individual fields to the results. Use **CONSTRAIN** links to constrain a value in the query.

Gene

- Length Integer **SUMMARY** **CONSTRAIN**
- Gene Identifier **SUMMARY** **CONSTRAIN**
- BGD Identifier **SUMMARY** **CONSTRAIN**
- Source **SUMMARY** **CONSTRAIN**
- Status **SUMMARY** **CONSTRAIN**
- Symbol **SUMMARY** **CONSTRAIN**
- Alias Alias Name **SUMMARY** **CONSTRAIN**
- Chromosome Chromosome **SUMMARY** **CONSTRAIN**
- Chromosome Location Location **SUMMARY** **CONSTRAIN**
- Data Sets Data Set **SUMMARY** **CONSTRAIN**
- Duplicate Entities Duplicate Entity **SUMMARY** **CONSTRAIN**
- GO Annotation GO Annotation **SUMMARY** **CONSTRAIN**
- Locations Location **SUMMARY** **CONSTRAIN**
- Ontology Annotations Ontology Annotation **SUMMARY** **CONSTRAIN**
- Organism Organism **SUMMARY** **CONSTRAIN**
- Proteins Protein **SUMMARY** **CONSTRAIN**
- Sequence Ontology Term SO Term **SUMMARY** **CONSTRAIN**
- Synonyms Synonym **SUMMARY** **CONSTRAIN**
- Transcripts Transcript **SUMMARY** **CONSTRAIN**

☐ Show empty fields

Query Overview

Gene

no constraints

Constraint for Gene > Status

Constrain

Choose a filter

Filter query results on this field having a specific value

Gene > Status

= Protein Coding Add to query

Filter query results on this field having any value or not

☒ Has no value ☐ Has a value Add to query

Click on the drop down list for Status and select Protein Coding. Then click on Add to Query.

The Query Overview should resemble the image below,

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

Model browser

links to add individual fields to the results. Use links to constrain a value in the query.

Gene

- Length Integer
- Gene Identifier
- BGD Identifier
- Source
- Status
- Symbol
- Alias Alias Name
- Chromosome Chromosome
- Chromosome Location Location
- Data Sets Data Set
- Duplicate Entities Duplicate Entity
- GO Annotation GO Annotation
- Locations Location
- Ontology Annotations Ontology Annotation
- Organism Organism
- Proteins Protein
- Sequence Ontology Term SO Term
- Synonyms Synonym
- Transcripts Transcript

☐ Show empty fields

Query Overview

Gene

- Gene Identifier
- BGD Identifier
- Status = Protein Coding (A)
- Symbol

Constraint logic: one constraint

Now, click on Show results to see all the genes that have the status Protein Coding.

Trail: [Query](#) > Results

Manage Columns

1 Filters

Create / Add to List ▾

Get Co...

Showing 1 to 25 of 19,797 rows

Rows per page: 25

Gene Identifier	BGD Identifier	Symbol	Status
100034674	BT13979	NGFRAP1	Protein Coding
100036590	BT28887	NKG2C	Protein Coding
100036592	NO VALUE	FUT7	Protein Coding
100048947	BT13933	RNASEL	Protein Coding
100101492	BT27065	STK24	Protein Coding
100113475	BT10937	DEFB124	Protein Coding
100124428	BT25811	NCBP1	Protein Coding
100124429	BT14888	ZFP30	Protein Coding
100124430	NO VALUE	ARHGAP4	Protein Coding
100124497	NO VALUE	NO VALUE	Protein Coding

ii. Querying for Protein Coding genes on a particular chromosome

Users can customize the previously run query by adding another constraint for `Chromosome`.

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

Model browser

links to add individual fields to the results. Use **CONSTRAIN** links to constrain a value in the query.

Gene **SUMMARY** **CONSTRAIN**

- Length Integer **SHOW** **CONSTRAIN**
- Gene Identifier **SHOW** **CONSTRAIN**
- BGD Identifier **SHOW** **CONSTRAIN**
- Source **SHOW** **CONSTRAIN**
- Status **SHOW** **CONSTRAIN**
- Symbol **SHOW** **CONSTRAIN**
- Alias Alias Name **SUMMARY** **CONSTRAIN**
- Chromosome** **SUMMARY** **CONSTRAIN**
 - Length Integer **SHOW** **CONSTRAIN**
 - Chromosome Identifier **SHOW** **CONSTRAIN**
 - Chromosome Name **SHOW** **CONSTRAIN**
- Data Sets Data Set **SUMMARY** **CONSTRAIN**
- Located Features Location **SUMMARY** **CONSTRAIN**
- Organism Organism **SUMMARY** **CONSTRAIN**
- Sequence Sequence **SUMMARY** **CONSTRAIN**
- Sequence Ontology Term SO Term **SUMMARY** **CONSTRAIN**
- Chromosome Location Location **SUMMARY** **CONSTRAIN**
- Data Sets Data Set **SUMMARY** **CONSTRAIN**
- Duplicate Entities Duplicate Entity **SUMMARY** **CONSTRAIN**

☐ Show empty fields

Query Overview

Gene

- Gene Identifier **X**
- BGD Identifier **X**
- Status **X**
 - = Protein Coding **X** (A)
- Symbol **X**

Constraint logic:

one constraint

Add a constraint to chromosome

While building the query, click on the Constrain tab for the Chromosome feature class, enter Chr1 in the text box of the pop-up window and click on Add to Query. Then click on Show for Chromosome Identifier and Chromosome Name.

Now, click on Show results and the query will result in all the genes that are of type Protein Coding on Chromosome with ID Chr1.

Note: Since here we are specifying a constrain at a class level, instead of at an individual attribute level, the Query-Builder will look up the input ID chr1 in all attributes.

Trail: [Query](#) > Results

Manage Columns

2 Filters

Create / Add to List ▾

Get C

Showing 1 to 25 of 754 rows

Rows per page: 25

Gene Identifier	BGD Identifier	Symbol	Status	Chromosome . Primary Identifier	Chromosome . Primary Identifier
100125416	BT18065	ACTL6A	Protein Coding	GK000001.2	Chr1
100125779	BT15595	PCYT1A	Protein Coding	GK000001.2	Chr1
100125920	NO VALUE	CCDC54	Protein Coding	GK000001.2	Chr1
100125943	BT19394	PLSCR1	Protein Coding	GK000001.2	Chr1
100138235	BT22670	NO VALUE	Protein Coding	GK000001.2	Chr1
100138320	NO VALUE	NO VALUE	Protein Coding	GK000001.2	Chr1
100138335	BT29782	WDR5B	Protein Coding	GK000001.2	Chr1
100138613	NO VALUE	SRPRB	Protein Coding	GK000001.2	Chr1
100139125	BT20803	MCF2L2	Protein Coding	GK000001.2	Chr1
100139155	NO VALUE	NO VALUE	Protein Coding	GK000001.2	Chr1

iii. Querying for Protein Coding genes on a particular chromosome and theirs exons

Users can customize this query further by configuring the query to show all the exons for each Gene.

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists **QueryBuilder** BLAST Regions Data Sources Help API MyMine

Search: e.g ATP s

Model browser

- Sequence Ontology Term SO Term [SUMMARY] [CONSTRAIN]
- Synonyms Synonym [SUMMARY] [CONSTRAIN]
- Transcripts Transcript [SUMMARY] [CONSTRAIN]
 - Gene Gene
 - Description [SHOW] [CONSTRAIN]
 - Length Integer [SHOW] [CONSTRAIN]
 - Transcript Identifier** [SHOW] [CONSTRAIN]
 - Secondary Identifier [SHOW] [CONSTRAIN]
 - Source [SHOW] [CONSTRAIN]
 - Status [SHOW] [CONSTRAIN]
 - Symbol [SHOW] [CONSTRAIN]
 - Chromosome Chromosome [SUMMARY] [CONSTRAIN]
 - Chromosome Location Location [SUMMARY] [CONSTRAIN]
- Data Sets Data Set [SUMMARY] [CONSTRAIN]
- Exons Exon [SUMMARY] [CONSTRAIN]
 - Transcripts Transcript [SUMMARY] [CONSTRAIN]
 - Length Integer [SHOW] [CONSTRAIN]
 - Exon Identifier** [SHOW] [CONSTRAIN]
 - Source [SHOW] [CONSTRAIN]
 - Chromosome Chromosome [SUMMARY] [CONSTRAIN]
 - Chromosome Location Location [SUMMARY] [CONSTRAIN]
 - Data Sets Data Set [SUMMARY] [CONSTRAIN]
 - Locations Location [SUMMARY] [CONSTRAIN]
 - Organism Organism [SUMMARY] [CONSTRAIN]

☐ Show empty fields

Query Overview

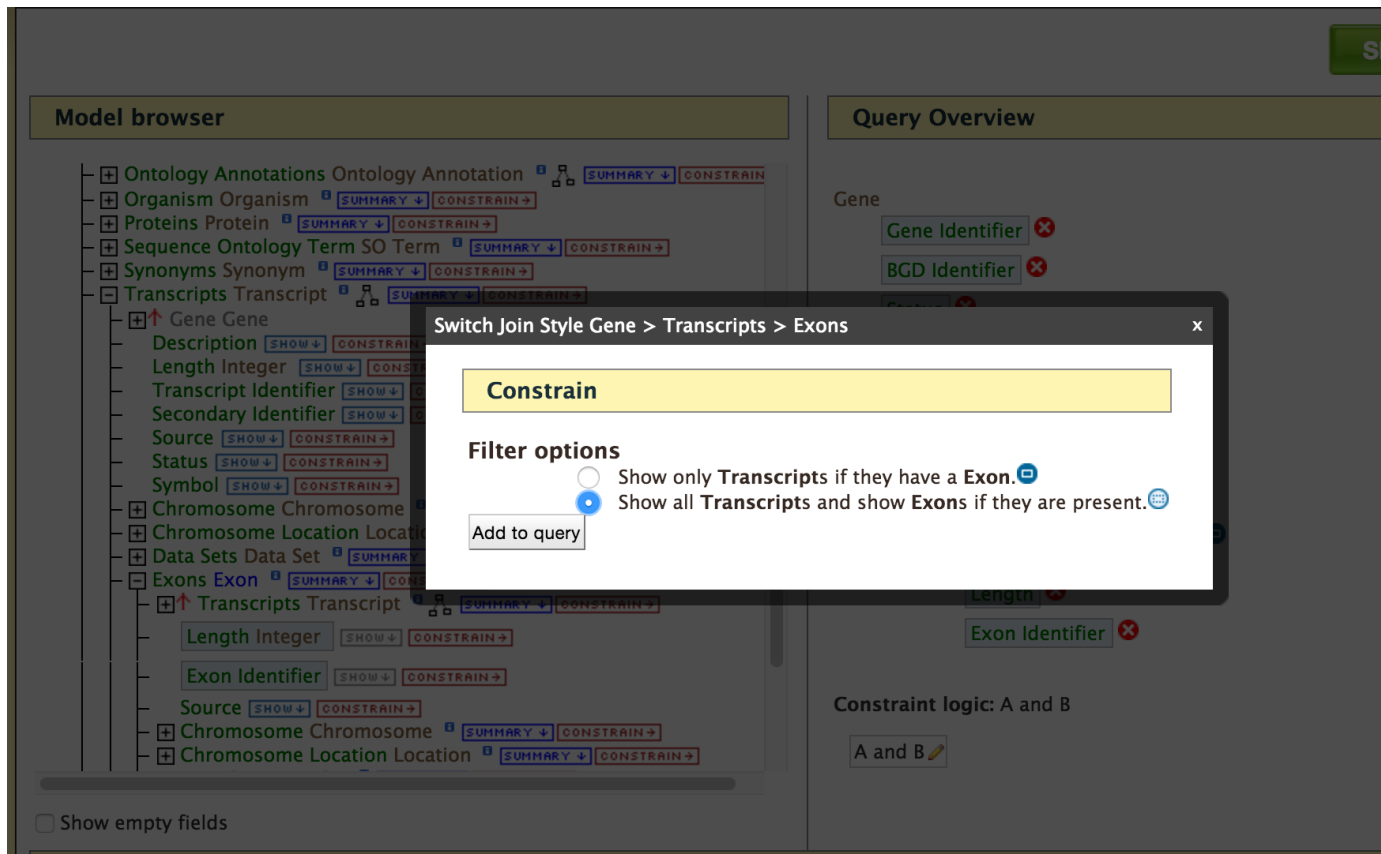
Gene

- Gene Identifier [X]
- BGD Identifier [X]
- Status [X]
 - = Protein Coding [X] (A)
- Symbol [X]
- Chromosome Chromosome [X]
 - LOOKUP chr1 [X] (B)
- Chromosome Identifier [X]
- Chromosome Name [X]
- Transcripts Transcript collection [X]
 - Transcript Identifier [X]
- Exons Exon collection [X]
 - Length [X]
 - Exon Identifier [X]

Constraint logic: A and B

A and B

1. Expand the Transcript subclass for Gene
2. Expand the Exon subclass for Transcript
3. Select Exon Identifier and Length.
4. To make the query more relaxed, click on the blue square icon (labelled with No. 4) near the Exon collection in the Query Overview. You will see a pop as shown below,



Select Show all Genes and show Exons if they are present and click on Add to query.
Now run the query and you should see the following results page,

Trail: [Query](#) > Results

Manage Columns

2 Filters

Create / Add to List

Get C

Showing 1 to 25 of 1,923 rows

Rows per page: 25

Gene Identifier	BGD Identifier	Symbol	Status	Chromosome . Primary Identifier	Chromosome » Chromosome Name	Transcripts » Exons
100125416	BT18065	ACTL6A	Protein Coding	GK000001.2	Chr1	14 Exons
100125416	BT18065	ACTL6A	Protein Coding	GK000001.2	Chr1	14 Exons
100125779	BT15595	PCYT1A	Protein Coding	GK000001.2	Chr1	10 Exons
						<div>Exon IdentifierLength</div> <div>exon11055270</div> <div>exon11056127</div> <div>exon11057100</div> <div>exon11058117</div> <div>exon11059152</div> <div>exon1106079</div> <div>exon11061143</div> <div>exon11062189</div> <div>exon11063187</div> <div>exon11064258</div>
100125779	BT15595	PCYT1A	Protein Coding	GK000001.2	Chr1	9 Exons

There are 1923 genes on Chromosome 1 and the gene with NCBI Gene ID 100125416 has two transcripts, each with 14 exons. Users can click on the 14 exons to expand the table with additional rows describing the Exon ID and Length for each exon.

1.3 Report Page

Every query result has a report page and the layout of the report page depends on the data available for a given query. Continuing with the example of **rp124**, the report page for this gene is shown below.

Gene : RPL24 *B. taurus*

Gene Identifier [282048](#) BGD Identifier [BT14245](#)
Source [NCBI RefSeq](#) Status [Protein Coding](#)

[SHARE](#)

Quick Links: **Summary** **Genes** **Proteins** **Function** **Homology** **Interactions** **Other**

Genome feature

Region:	gene	Length:	5503
Location:	GK000001.2:46415219-46420721 reverse strand		

1 Organism
Trail: [Gene](#)

[Manage Columns](#) [1 Filters](#) [Create / Add to List](#) [Get Code](#) [Download](#)

Showing 1 to 1 of 1 rows Rows per page: 25

Organism » Name	Organism » Taxon Id
Bos taurus	9913

1 Proteins
Trail: [Gene](#)

[Manage Columns](#) [1 Filters](#) [Create / Add to List](#) [Get Code](#) [Download](#)

Showing 1 to 1 of 1 rows Rows per page: 25

Proteins » DB identifier	Proteins » Primary Accession	Proteins » Organism . Name	Proteins » Length
RL24_BOVIN	Q86211	Bos taurus	157

The report page provides a complete description for gene rpl24. The header of the report page shows the Gene Identifier (NCBI Gene ID) and status indicating the type of gene, in this case a protein coding gene.

On the right hand side of the report page there are external links that links out to other Mines and databases.

The contents of the report page is divided into categories based on the type of information provided,

Summary

Provides a summary about a gene such as length, chromosome location and strand information. Users can also get the complete FASTA sequence of the gene by clicking on the FASTA tab.

Genes

This section provides information about the gene model. It displays all the transcripts and exons for a gene. The FASTA sequence of each transcript or exon can be downloaded by clicking on the FASTA tab. JBrowse is used to visualize the gene model.

Proteins

This section provides information about the protein product of gene rpl24. The comments section gives a brief description about the protein along with the UniProt accession.

Function

This section provides Gene Ontology annotations for a gene and annotations are divided into three categories,

- Biological process
- Molecular function
- Cellular Component

The GO terms are displayed along with the evidence code indicating how the annotations were derived.

Homology

This section provides information on all the homologues for *rpl24*. The first part shows a summarized view of the homologues in different organisms. The table below gives a detailed information about the homologue, the type of homologue and from which dataset the information was obtained.

Interaction

This section provides information about interactions. For *rpl24* there are no interaction information available but for genes that do have interaction information, a network is displayed showing all interactors for the current gene. The network displayer is a Cytoscape plugin.

Others

This section provides additional information such as,

- Child features – lists all the features that are sub features of the current gene
- Flanking regions – lists all the features flanking the current gene
- Overlapping features – lists all the features that overlap with current gene
- Pathways – lists all the pathways in which the current gene is a part of
- Publications – Publications related to the current gene

1.4 BLAST

Users can perform BLAST against the *Bos taurus* genomic, CDS or protein sequences using the BLAST page.

1.5 Genomic Region Search


The Genomic Region Search is a tool to fetch features that are within a given set of genomic coordinates or to fetch features that are within a given number of bases flanking a given set of genomic coordinates.

The coordinates have to be of the format,

```
chromosome_name:start..end
```

```
chromosome_name:start-end
```

```
chromosome_name start end
```

BovineMine v1.0 An integrated data warehouse and mining interface for Bos taurus genome

Home Templates Lists QueryBuilder BLAST **Regions** Data Sources Help API MyMine

Search:

Overlap features search from a new list of Genomic Regions

Search for features that overlap a list of genome coordinates you enter or upload, e.g. GK000001.2:50000..100000

[Genome coordinates help](#)

- Select Organism: *genome build: not available*
- Select Feature Types:

<input checked="" type="checkbox"/> Exon	<input checked="" type="checkbox"/> rRNA	<input checked="" type="checkbox"/> Transcript
<input checked="" type="checkbox"/> Gene	<input checked="" type="checkbox"/> snRNA	<input checked="" type="checkbox"/> UTR
<input checked="" type="checkbox"/> mRNA	<input checked="" type="checkbox"/> snoRNA	
<input checked="" type="checkbox"/> miRNA	<input checked="" type="checkbox"/> tRNA	
- Type/Paste in genomic regions in ☒ base coordinate ☐ interbase coordinate
 (click to see an example)▼
 (example for GWAS hits in feed efficiency)▼

GK000001.2:7901376..7901377
 GK000003.2:80105316..80105317
 GK000003.2:88904960..88904961
 GK000004.2:7139260..7139261
 GK000004.2:75484332..75484333
 GK000005.2:47594268..47594269
 GK000005.2:47866991..47866992
 GK000005.2:48623407..48623408
 GK000005.2:48876680..48876681
 GK000005.2:49341986..49341987

 or Upload genomic regions from a .txt file...
 No file chosen
- Extend your regions at both sides:

1k

10k

100k

1M

10M

Click on [click to see an example](#) for a representative set of genomic coordinates. Users can extend the regions on either side of a given coordinate using the slider or using the textbox.

Users can also select the type of coordinate system they would like to use: **base coordinate** or **interbase coordinate**.

Lets try and example. Click on the [click to see an example](#) and extend the region search by 500 bp and click on [Search](#). The result page will give a list of features that are present in each of the genomic interval provided in the input.

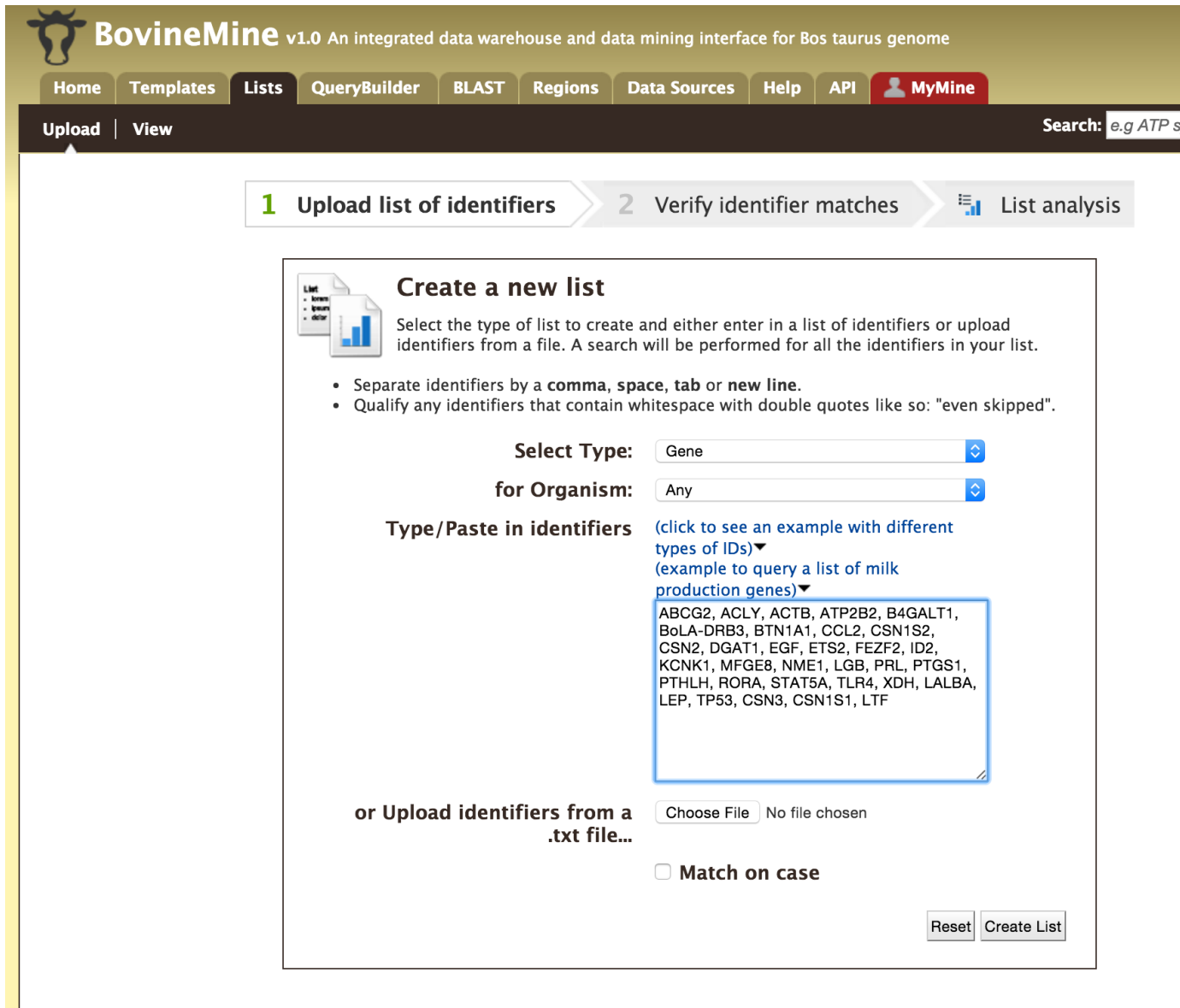
Export for all regions: TAB CSV GFF3 FASTA BED FASTA... or Create List by feature type: Exon Go Page size 10

GENOME REGION	FEATURE	FEATURE TYPE	LOCATION
GK000001.2:7901376..7901377	No overlap features found		
GK000003.2:80105316..80105317 FASTA...	LEPR 497205	Gene	GK000003.2:80070983..80070983
TAB CSV GFF3 FASTA BED	LEPR XM_010803434.1	mRNA	GK000003.2:80070983..80070983
Create List by Gene Go	OB-R ENSBTAG00000005910	Gene	GK000003.2:80071689..80071689
	LEPR NM_001012285.2	mRNA	GK000003.2:80071689..80071689
	OB-R ENSBTAT00000007764	mRNA	GK000003.2:80071689..80071689
	LEPR XM_010803431.1	mRNA	GK000003.2:80073801..80073801
	LEPR XM_010803433.1	mRNA	GK000003.2:80073801..80073801
	LEPR XM_010803430.1	mRNA	GK000003.2:80073801..80073801
	LEPR XM_010803432.1	mRNA	GK000003.2:80073801..80073801
	OB-R ENSBTAT00000043464	mRNA	GK000003.2:80075988..80075988
GK000003.2:88904960..88904961 FASTA...	DAB1 538818	Gene	GK000003.2:88373429..88373429
TAB CSV GFF3 FASTA BED	DAB1 XM_010803545.1	mRNA	GK000003.2:88373429..88373429
Create List by Gene Go	DAB1 XM_010803548.1	mRNA	GK000003.2:88758287..88758287
GK000004.2:7139260..7139261 FASTA...	ABCA13 100300398	Gene	GK000004.2:7095106..7195106
TAB CSV GFF3 FASTA BED	ABCA13 XM_002686635.3	mRNA	GK000004.2:7095106..7195106
Create List by Gene Go	ABCA13 ENSBTAG00000003531	Gene	GK000004.2:7096630..7486630
	ABCA13 ENSBTAT000000061018	mRNA	GK000004.2:7096630..7486630

The results can be exported as tab-delimited and comma-separated values. If the results have genomic features then they can be exported in GFF3 or BED format. Users can also export FASTA sequences of the features. If users are interested in creating a list of particular features from the result page then they can filter based on feature type, shown in red box, and click on ‘Go’.

1.6 Lists

Users can create a list of features. The input can either be gene IDs, transcript IDs, gene symbols, etc.



BovineMine v1.0 An integrated data warehouse and data mining interface for Bos taurus genome

Home Templates **Lists** QueryBuilder BLAST Regions Data Sources Help API MyMine

Upload | View Search: e.g ATP s

1 Upload list of identifiers **2 Verify identifier matches** List analysis

Create a new list

Select the type of list to create and either enter in a list of identifiers or upload identifiers from a file. A search will be performed for all the identifiers in your list.

- Separate identifiers by a **comma, space, tab or new line**.
- Qualify any identifiers that contain whitespace with double quotes like so: "even skipped".

Select Type: Gene

for Organism: Any

Type/Paste in identifiers [\(click to see an example with different types of IDs\)](#)
[\(example to query a list of milk production genes\)](#)

ABCG2, ACLY, ACTB, ATP2B2, B4GALT1, BoLA-DRB3, BTN1A1, CCL2, CSN1S2, CSN2, DGAT1, EGF, ETS2, FEZF2, ID2, KCNK1, MFGE8, NME1, LGB, PRL, PTGS1, PTHLH, RORA, STAT5A, TLR4, XDH, LALBA, LEP, TP53, CSN3, CSN1S1, LTF

or Upload identifiers from a .txt file... Choose File No file chosen

☐ Match on case

Reset Create List

The list tool tries to lookup the query throughout the database and will attempt to convert the identifiers to the type selected in the list 'Select Type' option. Lets try the examples provided. Click on 'Click to see example' link (highlighted in red box) and click on 'Create List'.

Before we show you the results ...

Choose a name for the list

(e.g. Smith 2013)

Add additional matches


You entered: 32 identifiers

We found: 3 Genes

Why are the numbers different? See below.

Duplicates found - which one(s) do you want?

Page 1 of 12  **1** 2 3 4 5 ... 12 

Identifier you provided	Matches 							
	primary identifier	symbol	source	name	length	class	secondary identifier	organism short
ACLY	511135	ACLY	NCBI RefSeq		44373	Gene	BT11987	B. taurus
	ENSBTAG00000016740	ACLY	Ensembl 79		44266	Gene	ENSBTAG00000016740	B. taurus
DGAT1	282609	DGAT1	NCBI RefSeq		9414	Gene		B. taurus
	ENSBTAG00000026356	DGAT1	Ensembl 79		9212	Gene	ENSBTAG00000026356	B. taurus
ABCG2	536203	ABCG2	NCBI RefSeq		127704	Gene	BT11761	B. taurus


Summary

The list tool does a lookup of the identifiers and shows you the results. If there are any duplicates, users can decide to add the relevant entries individually.

The summary section provides information regarding those identifiers that had a direct hit without any duplicates.

Click on 'Add all' and then click on 'Save a list of 62 genes'.

This will take users to a List Analysis page,


List Analysis for My New List (62 Genes)

Manage Columns
1 Filters
Create / Add to List
Get Code
Download

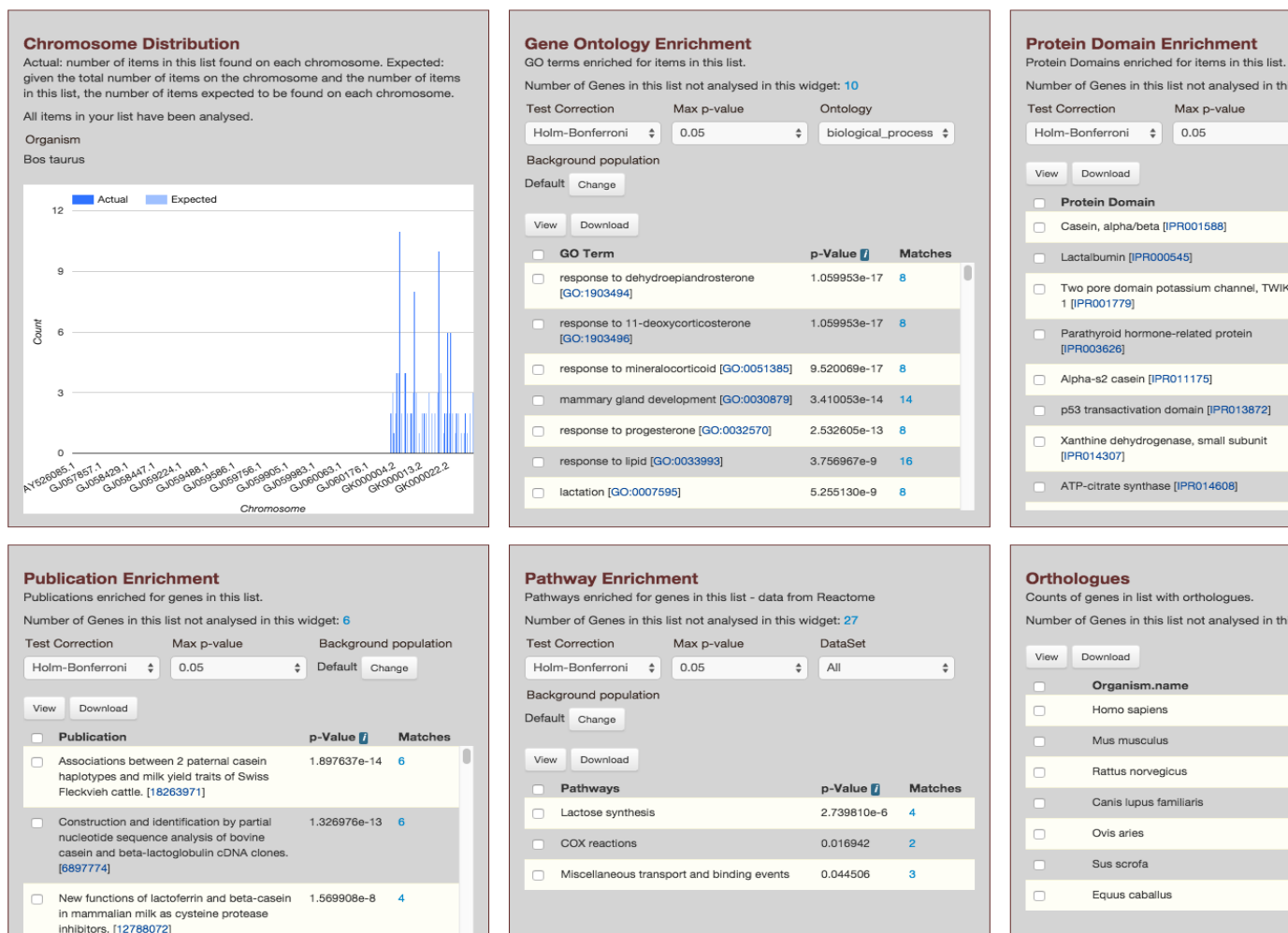
Showing 1 to 10 of 62 rows
Rows per page: 10

Gene ID	BGD Identifier	Symbol	Name	Source	Length	Chromosome . Primary Identifier	Chromosome Location . Start	Chromosome Location . End	Organism
280836	BT11292	LEP	NO VALUE	NCBI RefSeq	16833	GK000004.2	93249792	93266624	B. taurus
280846	BT26175	LTF	NO VALUE	NCBI RefSeq	41310	GK000022.2	53514781	53556090	B. taurus
280901	BT12480	PRL	NO VALUE	NCBI RefSeq	8616	GK000023.2	35105135	35113750	B. taurus
280960	NO VALUE	XDH	NO VALUE	NCBI RefSeq	60080	GK000011.2	14176297	14236376	B. taurus
280979	NO VALUE	ACTB	NO VALUE	NCBI RefSeq	3451	GK000025.2	39343599	39347049	B. taurus
281043	BT10752	CCL2	NO VALUE	NCBI RefSeq	1867	GK000019.2	16232956	16234822	B. taurus
281099	BT11954	CSN2	NO VALUE	NCBI RefSeq	8693	GK000006.2	87179499	87188191	B. taurus
281148	NO VALUE	ETS2	NO VALUE	NCBI RefSeq	19191	GK000001.2	152872775	152891965	B. taurus
281536	BT12122	TLR4	NO VALUE	NCBI RefSeq	11018	GK000008.2	108828899	108839916	B. taurus
281542	BT10936	TP53	NO VALUE	NCBI RefSeq	12395	GK000019.2	27985492	27997886	B. taurus

Orthologues
C. lupus familiaris (125) E. ...
H. sapiens (100) M. muscul...
R. norvegicus (108) S. scrof...

External Links
No external links.

This page provides users with widgets to perform analyses on gene lists that they have created.



Currently available widgets are,

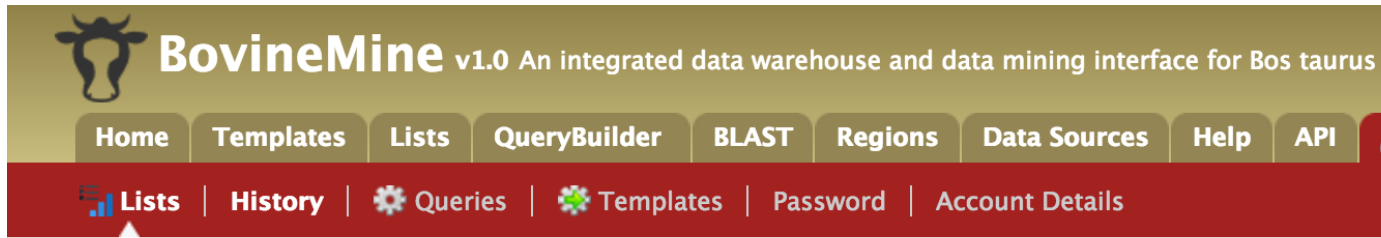
1. Chromosome Distribution
2. Gene Ontology Enrichment
3. Protein Domain Enrichment
4. Publication Enrichment
5. Pathway Enrichment
6. Orthologues

1.7 MyMine

Users can create an account on BovineMine which enables them to store their lists, custom templates and keep track of their sessions. To create a login, click on the 'Log in' on the top-right corner of any page on BovineMine.

New users can create a new account with the username being their Email ID.

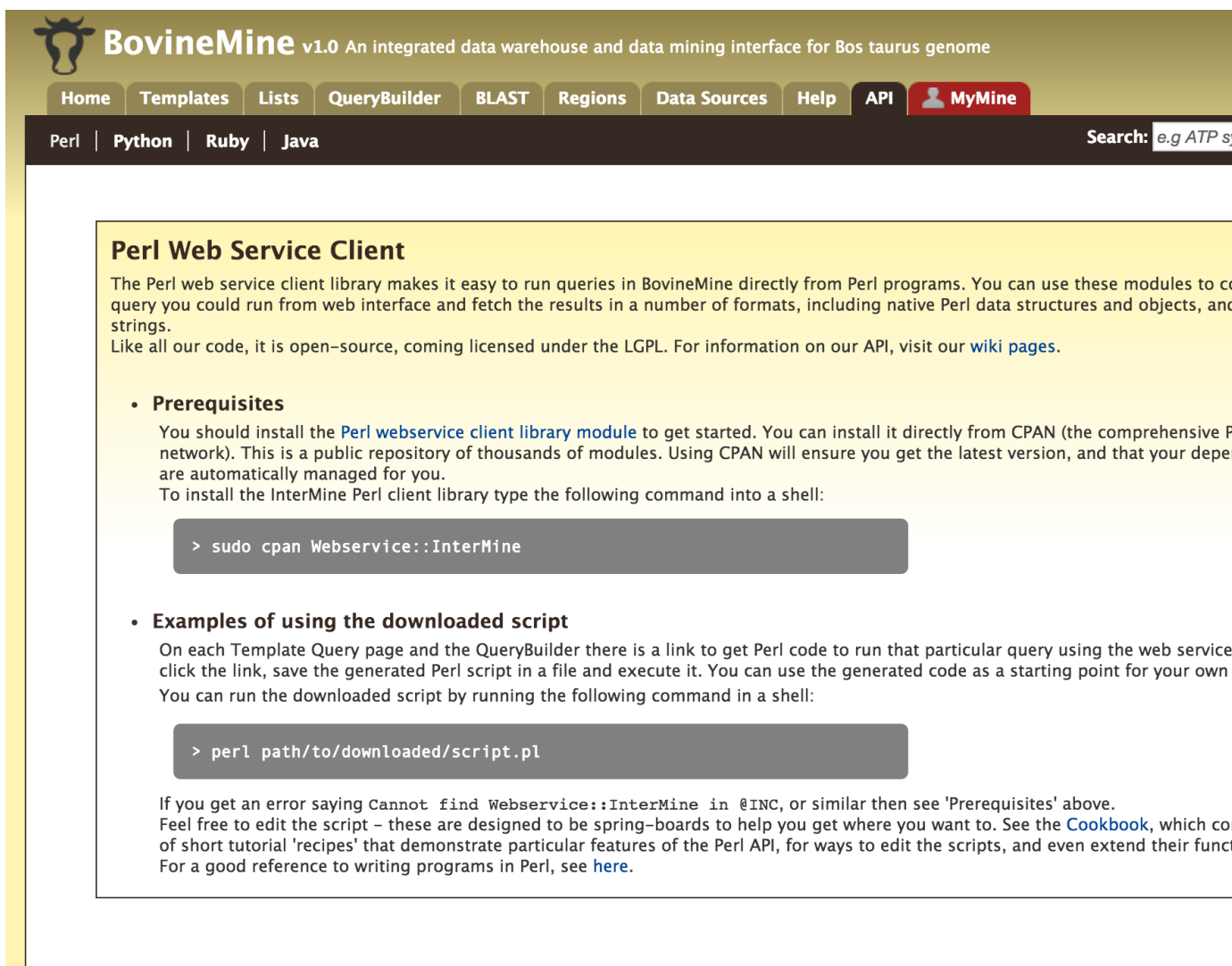
BovineMine provides MyMine for users to manage their lists, queries, templates, and account details.



1. Lists – lists saved by the user.
2. History of queries by user – shows a list of most recent queries performed by the user.
3. Templates – Templates created by the user or existing templates that are marked as ‘favorite’ by the user.
4. Password change – change the password for the user’s account.
5. Account details – for updating user preferences.

1.8 API

For users who would like to programmatically access BovineMine they can make use of the InterMine API.



Perl Web Service Client

The Perl web service client library makes it easy to run queries in BovineMine directly from Perl programs. You can use these modules to convert a query you could run from web interface and fetch the results in a number of formats, including native Perl data structures and objects, and strings.

Like all our code, it is open-source, coming licensed under the LGPL. For information on our API, visit our [wiki pages](#).

- Prerequisites**

You should install the [Perl webservice client library module](#) to get started. You can install it directly from CPAN (the comprehensive Perl network). This is a public repository of thousands of modules. Using CPAN will ensure you get the latest version, and that your dependencies are automatically managed for you.

To install the InterMine Perl client library type the following command into a shell:

```
> sudo cpan Webservice::InterMine
```
- Examples of using the downloaded script**

On each Template Query page and the QueryBuilder there is a link to get Perl code to run that particular query using the web service. Click the link, save the generated Perl script in a file and execute it. You can use the generated code as a starting point for your own scripts. You can run the downloaded script by running the following command in a shell:

```
> perl path/to/downloaded/script.pl
```

If you get an error saying `Cannot find Webservice::InterMine in @INC`, or similar then see 'Prerequisites' above. Feel free to edit the script – these are designed to be spring-boards to help you get where you want to. See the [Cookbook](#), which contains a series of short tutorial 'recipes' that demonstrate particular features of the Perl API, for ways to edit the scripts, and even extend their functionality. For a good reference to writing programs in Perl, see [here](#).

Perl, Python, Ruby and Java are the 4 languages in which the InterMine API is available. For more information on the details on the API visit the InterMine Wiki.

1.9 Data Sources

Provides a description of the datasets that are integrated into BovineMine along with the location from which these datasets were downloaded, their version or release, citations wherever applicable and any additional comments.

For more generic examples on how to use InterMine, click [here](#). These are tutorials created by FlyMine that showcase the different features of InterMine.

CHAPTER 2

Indices and tables

- `genindex`
- `modindex`
- `search`