

Sanger Mouse Portal “How to” Guide

Extracting Mouse Genetic Project raw phenotyping data

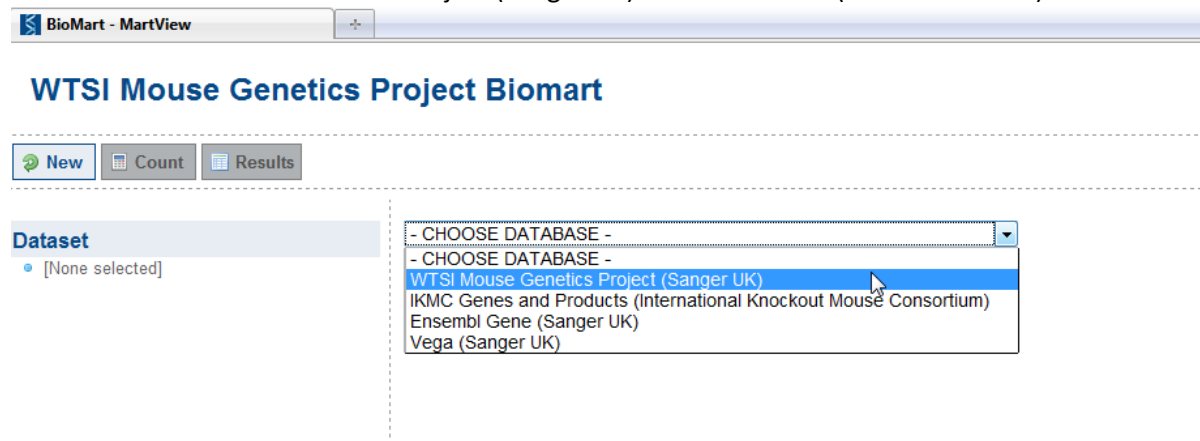
Introduction:

Raw phenotyping data is available through a publicly available BioMart¹ interface. This can be accessed at the following web address <http://www.sanger.ac.uk/htgt/biomart>. Mouse Genetics Project data which has passed all necessary internal quality control checks is automatically published to this BioMart within 24 hours. As of December 2011 this data set comprises approximately 9×10^6 data points and hence it is not recommended that users attempt to extract all data in one bolus. Rather, the query made to the BioMart should be carefully phrased to extract only the data pertinent to answering your specific scientific question. Below is a step-wise example of how to extract the data using the standard BioMart user interface. Further information on using a BioMart can be found at: <http://bowtie-bio.sourceforge.net/recount/biomaRt.pdf>.

Demonstration Example:

The following example, demonstrates the extraction of data associated with the Mysm1 knockout line.

1. Open the BioMart by following the link: <http://www.sanger.ac.uk/htgt/biomart>.
2. Select “WTSI Mouse Genetics Project (Sanger UK) as the database (as shown below).



The screenshot shows the BioMart MartView interface. At the top, there is a browser tab labeled "BioMart - MartView". Below the tab, the main heading is "WTSI Mouse Genetics Project Biomart". Underneath the heading, there are three buttons: "New", "Count", and "Results". The "Dataset" section is visible, showing "[None selected]". A dropdown menu is open, displaying the following options: "- CHOOSE DATABASE -", "- CHOOSE DATABASE -", "WTSI Mouse Genetics Project (Sanger UK)", "IKMC Genes and Products (International Knockout Mouse Consortium)", "Ensembl Gene (Sanger UK)", and "Vega (Sanger UK)".

¹ BioMart is a generic data management system [Smedley *et al* “BioMart – biological queries made easy” (2009) BMC Genomics 10 (22)].

Sanger Mouse Portal “How to” Guide

3. Select “MGP phenotyping” as the dataset (as shown below).

The screenshot shows the BioMart MartView interface. At the top, there is a navigation bar with the BioMart logo and the text "BioMart - MartView". Below this is the title "WTSI Mouse Genetics Project Biomart". There are three buttons: "New", "Count", and "Results". The "Dataset" section shows a dropdown menu with the current selection "WTSI Mouse Genetics Project (Sanger UK)". A secondary dropdown menu is open, showing a list of datasets: "- CHOOSE DATASET -", "- CHOOSE DATASET -", "EUCOMM / KOMP-CSD Gene Targeting Alleles", "EUCOMM Gene Trapping Alleles", "MGP Phenotyping" (highlighted in blue), "MGP Possible MP Terms (internal use only)", "MP Term Export", and "WTSI DNA Clone Resources". A mouse cursor is pointing at the "MGP Phenotyping" option.

4. Click on the Attribute link on the left (as shown below).

The screenshot shows the BioMart MartView interface. At the top, there is a navigation bar with the BioMart logo and the text "BioMart - MartView". Below this is the title "WTSI Mouse Genetics Project Biomart". There are three buttons: "New", "Count", and "Results". The "Dataset" section shows a dropdown menu with the current selection "WTSI Mouse Genetics Project (Sanger UK)". Below the "Dataset" section is the "Filters" section, which shows a dropdown menu with the current selection "[None selected]". The "Attributes" section is highlighted in red, and a tooltip is visible over it that says "Choose data you want to view or save". Below the "Attributes" section is the "Dataset" section, which shows a dropdown menu with the current selection "[None Selected]".

Sanger Mouse Portal “How to” Guide

5. Select Published Graph Data (as shown below).

The screenshot shows the BioMart MartView interface for the WTSI Mouse Genetics Project. The browser address bar shows 'BioMart - MartView'. The page title is 'WTSI Mouse Genetics Project Biomart'. The bio::mart logo is in the top right. Below the title are buttons for 'New', 'Count', and 'Results', and a row of buttons for 'URL', 'XML', 'Perl', and 'Help'. The main content area is titled 'Please select columns to be included in the output and hit 'Results' when ready'. It contains several sections:

- Dataset:** MGP Phenotyping
- Filters:** [None selected]
- Attributes:** A list of 25 attributes including Marker Symbol, Allele Name, Colony Prefix, Strain, Adult LacZ Expression, Auditory Brainstem Response, Body Composition (DEXA), Brain Histopathology, Citrobacter Challenge, Dysmorphology, Embryo LacZ Expression, Eye Histopathology, Eye Morphology, Fertility, General Observations, Glucose Tolerance (ip), Grip Strength, Haematology (CBC), Heart Histology, Heart Weight, Hot Plate, Indirect Calorimetry, MicroCT & Quantitative Faxitron, Micronuclei, Modified SHIRPA, Non-Invasive Blood Pressure, and Open Field.
- Feature Selection:** A list of features and phenotyping tests with checkboxes. The 'Published Graph Data' option is selected and circled in red. Other options include Heatmap Data, Heatmap Graphs (Deprecated), Parameter Level Heatmap Data, Published Images, Fertility, Homozygote Viability, Adult Expression, and Collaborator Data.
- Allele/Colony Details:** A section with checkboxes for Marker Symbol, Allele Name, Colony Prefix, Strain, Finalised Population Count, Significant Call Count, and Mouse Line Summary.
- Phenotyping Tests:** A section with checkboxes for various tests such as Adult LacZ Expression, Auditory Brainstem Response, Body Composition (DEXA), Brain Histopathology, Citrobacter Challenge, Dysmorphology, Embryo LacZ Expression, Eye Histopathology, Eye Morphology, Fertility, General Observations, Glucose Tolerance (ip), Grip Strength, Haematology (CBC), Micronuclei, Modified SHIRPA, Non-Invasive Blood Pressure, Open Field, Peripheral Blood Lymphocytes, Plasma Chemistry, Plasma Immunoglobulins, Prepulse Inhibition, Recessive Lethal Study, Rotarod, Salmonella Challenge, Skin Histopathology, Stress Induced Hyperthermia, and Tail Epidermis Wholemount.

Sanger Mouse Portal “How to” Guide

6. Now click on Filters to access a dialog box to restrict the query (as shown below).

The screenshot shows the BioMart MartView interface for the WTSI Mouse Genetics Project Biomart. The browser title is "BioMart - MartView". The page header includes the "WTSI Mouse Genetics Project Biomart" title and the "biomart" logo. Below the header are buttons for "New", "Count", "Results", "URL", "XML", "Perl", and "Help".

The main content area is divided into two columns. The left column contains a "Dataset" section with "MGP Phenotyping" selected, a "Filters" button (circled in red), and an "Attributes" list with the following items: Allele Name, Pipeline, Membership, Mouse Name, Gender, Genotype, Genetic Background, Observation, X Value, Y Value, and Colony Prefix. Below this is another "Dataset" section with "[None Selected]".

The right column contains a selection panel titled "Please select columns to be included in the output and hit 'Results' when ready". It lists several data types with radio buttons: Heatmap Data, Heatmap Graphs (Depricated), Parameter Level Heatmap Data, Published Graph Data (selected), Published Images, Fertility, Homozygote Viability, Adult Expression, and Collaborator Data. Below this is a "Features" section with a "Published Graph Data" sub-section containing checkboxes for: Allele Name, Pipeline, Membership, Mouse Name, Gender, Genotype, Genotype String, Genetic Background, Observation, X Value, Y Value, Graph Type, and Colony Prefix. At the bottom is a "MIG IDs (Internal Use Only)" section with checkboxes for Mouse ID and Population ID.


Sanger Mouse Portal “How to” Guide

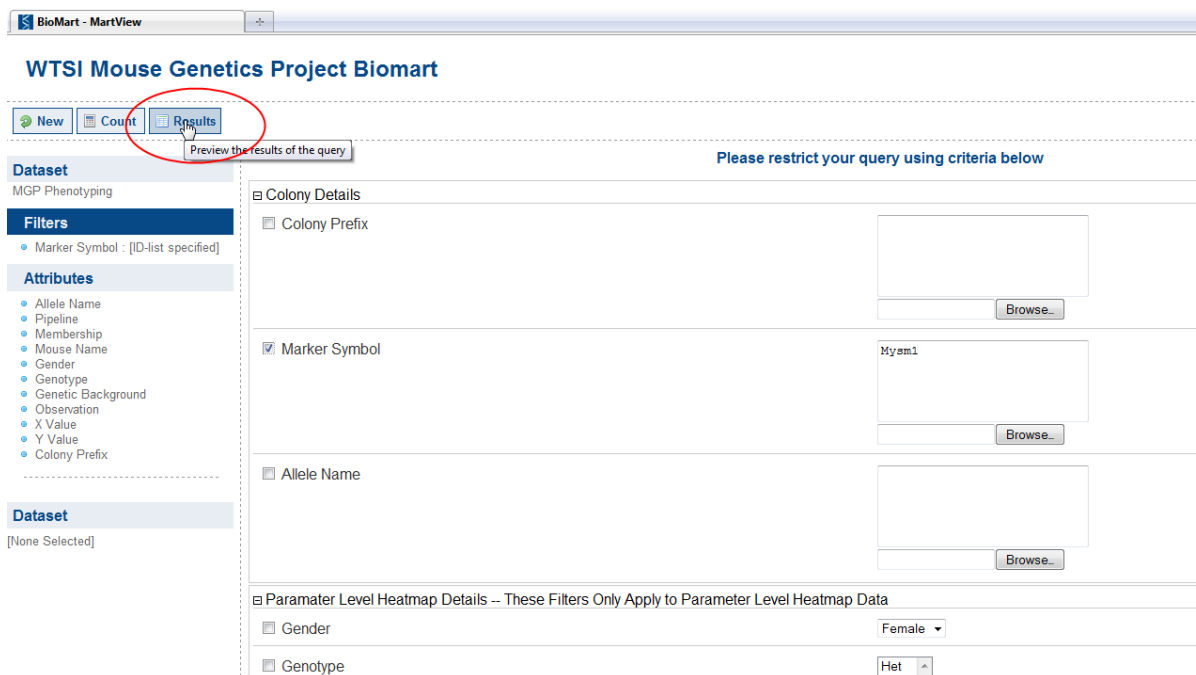
7. We can filter the data by a gene name e.g. Mym1 (a.k.a. Marker symbol) (as shown below).

The screenshot displays the WTSI Mouse Genetics Project Biomart interface. On the left, the 'Dataset' is identified as 'MGP Phenotyping'. Under the 'Filters' section, 'Marker Symbol' is selected and circled in red. The 'Attributes' list includes Allele Name, Pipeline, Membership, Mouse Name, Gender, Genotype, Genetic Background, Observation, X Value, Y Value, and Colony Prefix. The main query area, titled 'Please restrict your query using criteria below', contains several filter sections: 'Colony Details' with a 'Colony Prefix' field; 'Allele Name' with an empty field; 'Parameter Level Heatmap Details -- These Filters Only Apply to Parameter Level Heatmap Data' with dropdowns for 'Gender' (set to Female), 'Genotype' (set to Het), and 'Pipeline' (set to MGP Citrobacter Challenge); and 'Test' with a list of test types including Adult LacZ Expression, Auditory Brainstem Response, Body Composition (DEXA), Brain Histopathology, Citrobacter Challenge, and Dysmorphology. The 'Marker Symbol' field is set to 'Mym1' and is also circled in red. Navigation buttons for 'New', 'Count', 'Results', 'URL', 'XML', 'Perl', and 'Help' are visible at the top.

Sanger Mouse Portal “How to” Guide

8. Click on the Results button (as shown below) to filter as specified.

 Note if the program is running for a long time and no results appear, it is likely you have asked for too big a slice of the data in one query. Please close and reconsidered your filters. A poorly formulated query can cause so much data to be sent to your computer that the web browser (and potentially your computer) will become overwhelmed and stop responding. It should also be noted that the volume of data available through this interface is roughly two orders of magnitude greater than supported by Excel. As such careful pre-planning of queries is advised.



BioMart - MartView

WTSI Mouse Genetics Project Biomart

New Count **Results** Preview the results of the query

Dataset
MGP Phenotyping

Filters
• Marker Symbol : [ID-list specified]

Attributes
• Allele Name
• Pipeline
• Membership
• Mouse Name
• Gender
• Genotype
• Genetic Background
• Observation
• X Value
• Y Value
• Colony Prefix

Dataset
[None Selected]

Please restrict your query using criteria below

Colony Details

Colony Prefix

Marker Symbol
Mysm1

Allele Name

Parameter Level Heatmap Details -- These Filters Only Apply to Parameter Level Heatmap Data

Gender Female

Genotype Het

Sanger Mouse Portal “How to” Guide

11. Results can be exported using the export facility (as shown below).

BioMart - MartView

WTSI Mouse Genetics Project Biomart

New Count Results

Dataset
MGP Phenotyping

Export all results to File [CSV] Unique results only **Go**

Email notification to

Filters
Marker Symbol : [ID-list specified]

Attributes
Allele Name
Pipeline
Membership
Mouse Name
Gender
Genotype
Genetic Background
Observation

View
200 rows as HTML Unique results only

| Allele Name | Pipeline | Membership | Mouse Name | Gender | Genotype | Genetic Background | Observation | X Value |
|-------------|----------|------------|------------|--------|----------|--------------------|-------------|---------|
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |