PhenoMiner: improved interfaces enhance usability of RGD's quantitative phenotype data repository

Kaldunski ML, **Smith JR**, Laulederkind SJF, Hayman GT, Wang S-J, Tutaj M, Vedi M, Demos WM, Gibson AC, Lamers L, Thorat K, Thota J, Tutaj MA, De Pons JL, Dwinell MR, Kwitek AE

Rat Genome Database, Dept. of Physiology, Medical College of Wisconsin, Milwaukee, WI
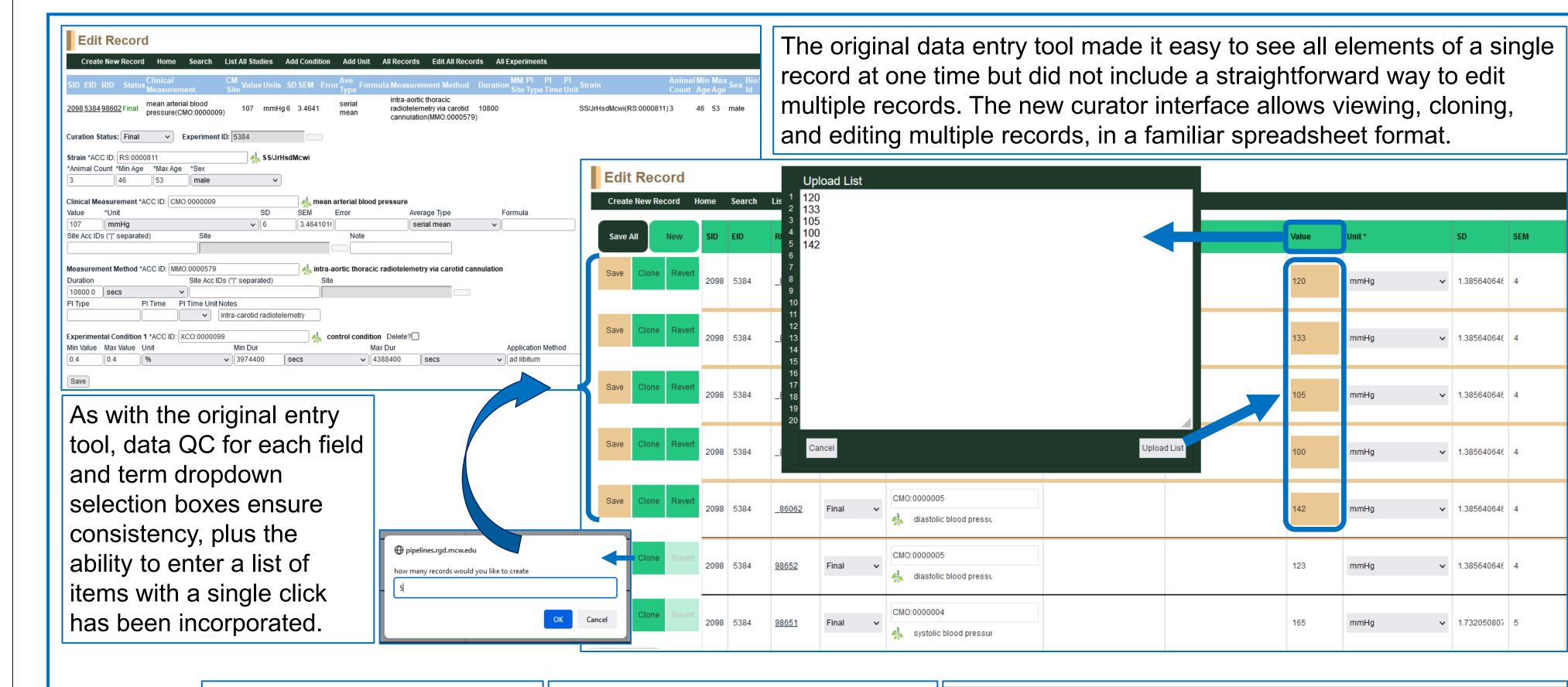
https://rgd.mcw.edu

Abstract:

The Rat Genome Database (RGD, https://rgd.mcw.edu) is the principal resource for data related to rat biomedical research for genome, phenotype, and disease. The data collection is the result of both manual curation by RGD curators, and data importation from other databases through custom pipelines. RGD has developed a growing suite of innovative tools for querying, analyzing, and visualizing this data, making it a valuable resource for researchers worldwide. One recently updated platform is the PhenoMiner data repository with its concomitant data mining tool components. PhenoMiner was developed for rat quantitative phenotype measurement data from both manual curation of scientific literature and direct data submissions by investigators. PhenoMiner enables users to query and visualize quantitative phenotype data across rat strains and multiple studies. The data includes detailed information about what (Clinical Measurement Ontology - CMO), how (Measurement Method Ontology - MMO), and under what conditions (Experimental Conditions Ontology - XCO) phenotypes were measured, and in what animals (Rat Strain Ontology - RS) for each measurement value. A recent project included curation for strains in the Hybrid Rat Diversity Panel, especially those previously underrepresented in the data repository.
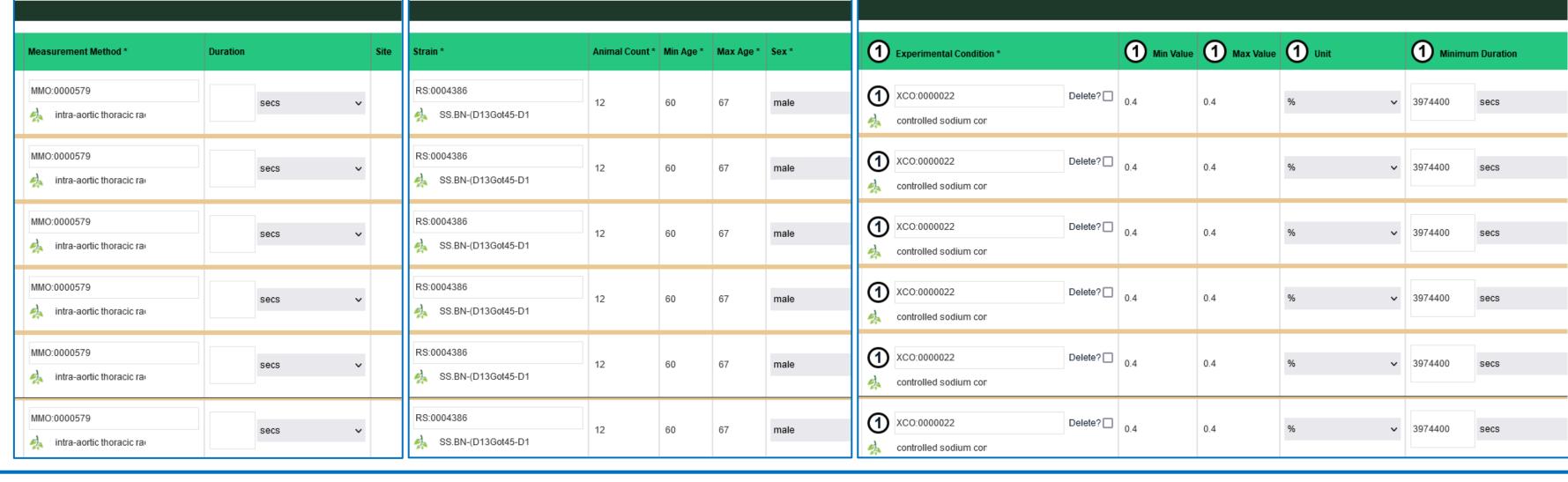
To increase curation efficiency, the data input interfaces have been improved, including enhanced ability to view, clone, and edit multiple records at a time, which decreases the amount of time required for entering study data. Quality control checkpoints in place for individual data entry have been expanded to encompass bulk data loading.

In response to user feedback, the public-facing data mining tool was reworked to improve the user interface (UI) for data interactivity. Improvements in the search functionality allow straightforward navigation of complex datasets, and the new query results page simplifies filtering to facilitate tailoring of specific query results. Users are no longer required to return to the front page or to start a new query to remove some conditions or other components. New functionality makes it possible to view results for related terms with the same unit of measurement in the same graph. Data can be downloaded either as filtered results or all results matching the user's original query, providing users with data at multiple levels. Upcoming planned improvements to the user interface will include the ability to visualize as well as download imported high throughput phenotyping data from individual inbred and outbred (e.g., heterogeneous stock) rats. RGD is advancing the utility of PhenoMiner by streamlining curation with increased ability to input data accurately and efficiently, and via improvements in the public UI enabling better data access, filtering, and visualization, with continued ability to download data for further analyses.

PhenoMiner is a tool for ontology-based standardization, storage and mining of quantitative phenotype data for the laboratory rat. The ontologies used to organize the data in the database include:
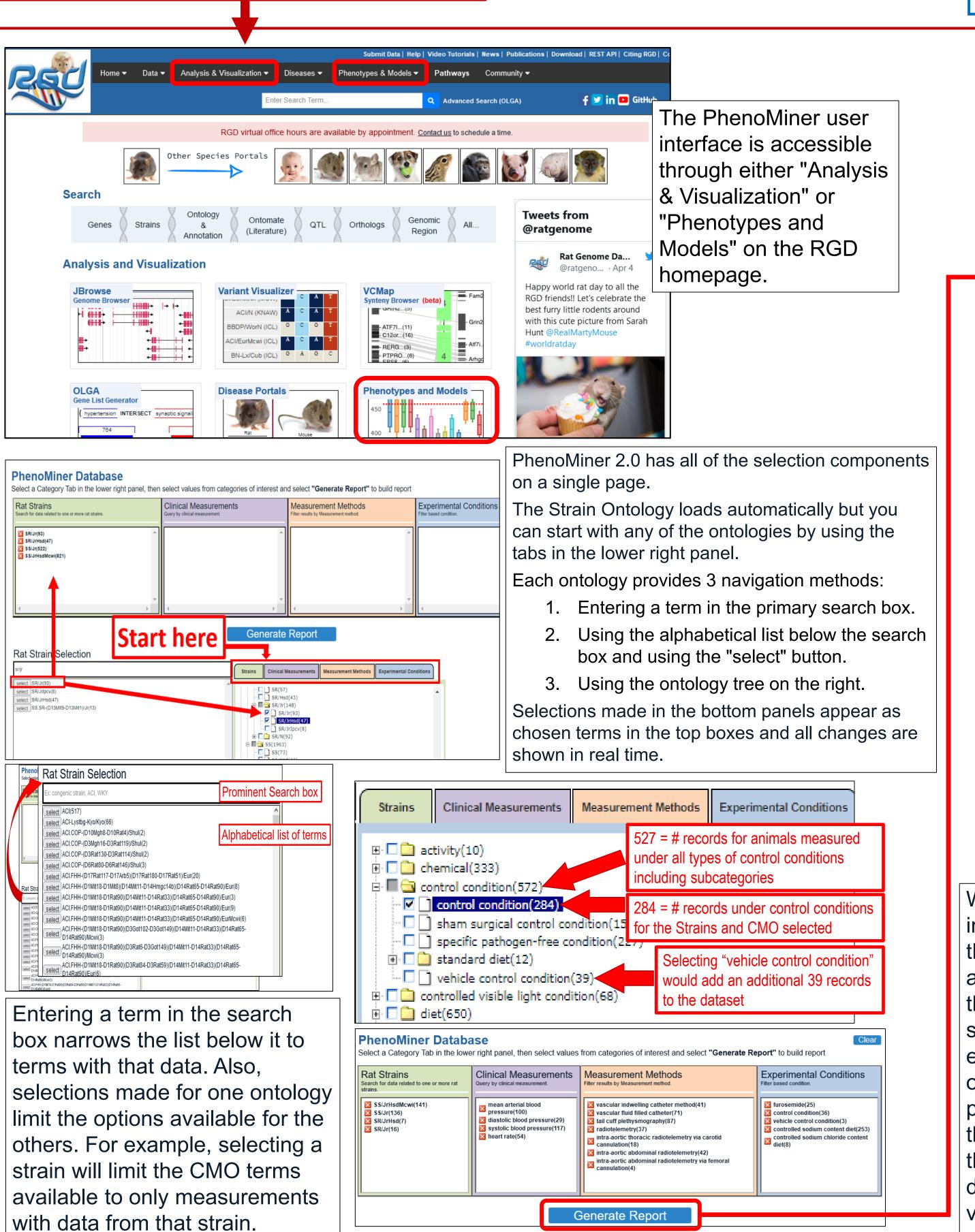
- **Rat Strain Ontology** (RSO) for animals measured (e.g., SHR, WKY).
- **Vertebrate Trait Ontology** (VTO) for organizing related phenotypes within a study (e.g., arterial blood pressure trait).
- **Clinical Measurement Ontology** (CMO) for what specific phenotype was measured (e.g., systolic blood pressure).
- **Measurement Method Ontology** (MMO) for how the phenotype was measured (e.g., indwelling catheter, tail-cuff).
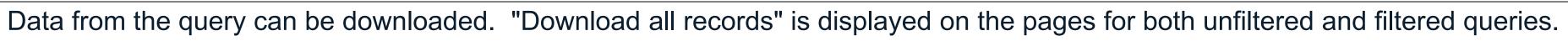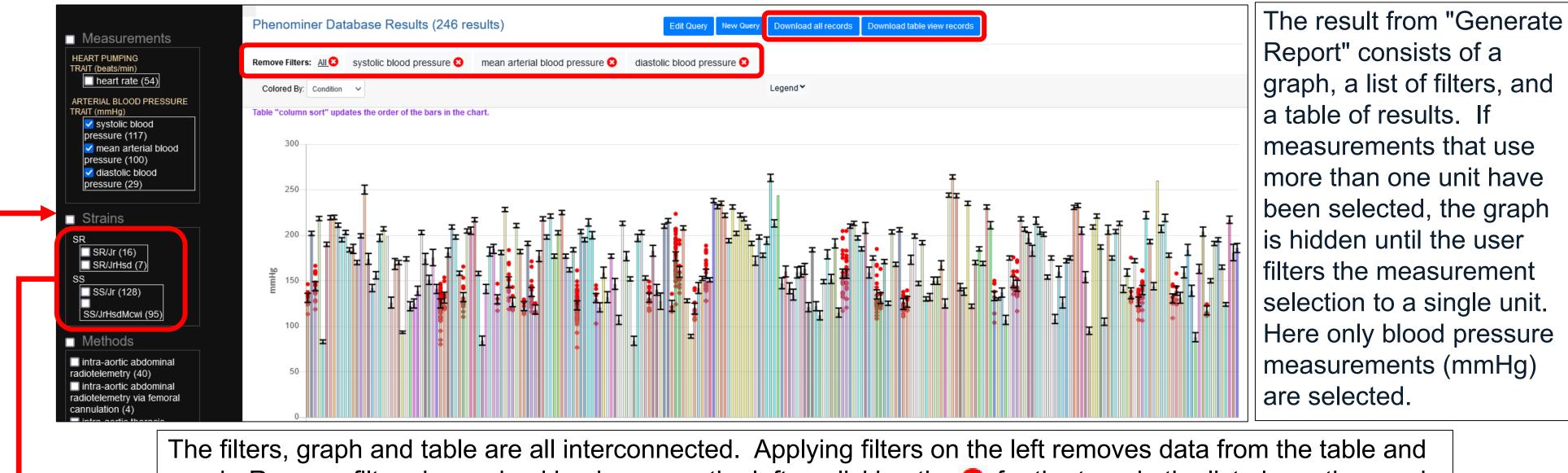- **Experimental Condition Ontology** (XCO) for recording the experimental conditions (e.g., control, salt diet).

The original data entry tool made it easy to see all elements of a single record at one time but did not include a straightforward way to edit multiple records. The new curator interface allows viewing, cloning, and editing multiple records, in a familiar spreadsheet format.

As with the original entry tool, data QC for each field and term dropdown selection boxes ensure consistency, plus the ability to enter a list of items with a single click has been incorporated.

Improved Curation Tool

Improved User Interface

The PhenoMiner user interface is accessible through either "Analysis & Visualization" or "Phenotypes and Models" on the RGD homepage.

PhenoMiner 2.0 has all of the selection components on a single page.
The Strain Ontology loads automatically but you can start with any of the ontologies by using the tabs in the lower right panel.

Each ontology provides 3 navigation methods:
1. Entering a term in the primary search box.
2. Using the alphabetical list below the search box and using the "select" button.
3. Using the ontology tree on the right.

Selections made in the bottom panels appear as chosen terms in the top boxes and all changes are shown in real time.

Entering a term in the search box narrows the list below it to terms with that data. Also, selections made for one ontology limit the options available for the others. For example, selecting a strain will limit the CMO terms available to only measurements with data from that strain.

527 = # records for animals measured under all types of control conditions including subcategories

284 = # records under control conditions for the Strains and CMO selected

Selecting "vehicle control condition" would add an additional 39 records to the dataset

Data from the query can be downloaded. "Download all records" is displayed on the pages for both unfiltered and filtered queries.
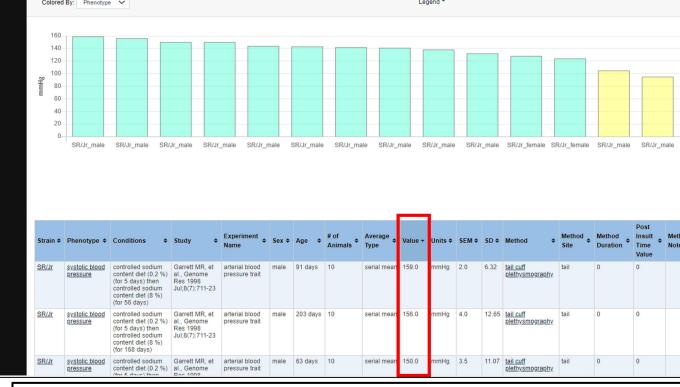
The result from "Generate Report" consists of a graph, a list of filters, and a table of results. If measurements that use more than one unit have been selected, the graph is hidden until the user filters the measurement selection to a single unit. Here only blood pressure measurements (mmHg) are selected.

The filters, graph and table are all interconnected. Applying filters on the left removes data from the table and graph. Remove filters by unchecking boxes on the left or clicking the ❌ for the term in the list above the graph.

Select the parameter to color the bars. A dropdown legend is provided to disambiguate the colors.

Sort the table using the up/down arrows in the column headers. Bars in the graph are reordered accordingly. Here the "Values" are sorted in descending order

The combination of sorting and coloring can highlight important or unexpected results.

Where data for individual rats exists, the graph shows the average as the bar with the individual values shown as red dots for each phenotype. Hover over the bar to see a popup with details about the average value for that datapoint or over a dot to see that individual value.

Mouse over "View Values" to see the list of animal IDs and individual values for the biological replicates included in the corresponding summary/mean value.