PhenoMiner: RGD's quantitative phenotype data repository has data update and data mining tool improvements
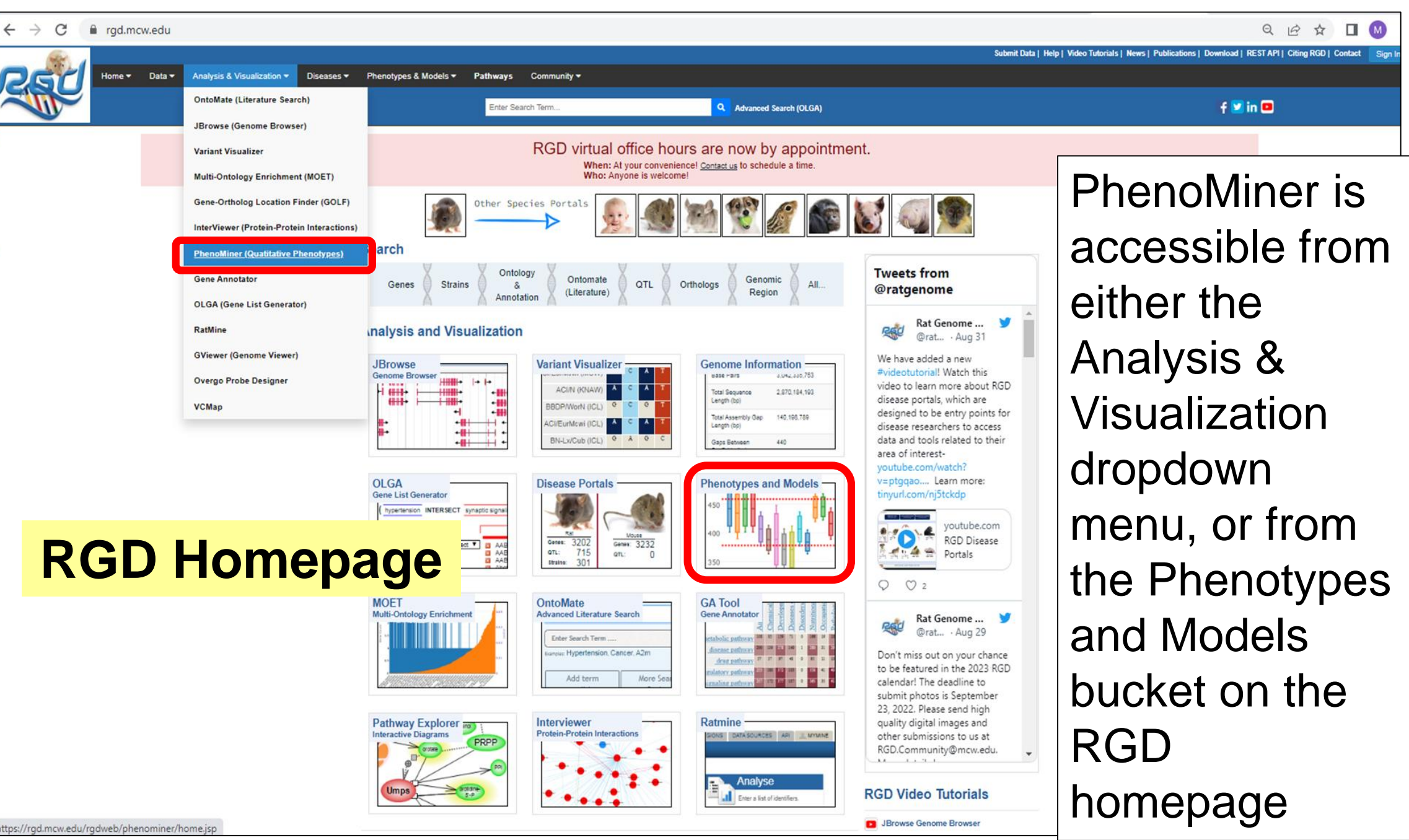
[1]***Kaldunski, M.L.,** [1]Smith, J.R., [1]Thota, J., [1]Tutaj, M.A., [1]Laulederkind, S.J.F., [1]Hayman, G.T., [1]Wang, S-J., [1]Tutaj, M., [1]Vedi, M., [1]Demos, W.M., [1]Nalabolu, H.S., [1]Lamers, L., [1]Gibson, A.C., [1]Thorat, K., [1]Brodie K.C., [1]Zacher S.M., [1]De Pons, J.L., [1,2]Dwinell, M.R., [1,2]Kwitek, A.E.

[1]Rat Genome Database, Dept. of Biomedical Engineering, Medical College of Wisconsin, Milwaukee, WI, [2]Department of Physiology, Medical College of Wisconsin, Milwaukee, WI
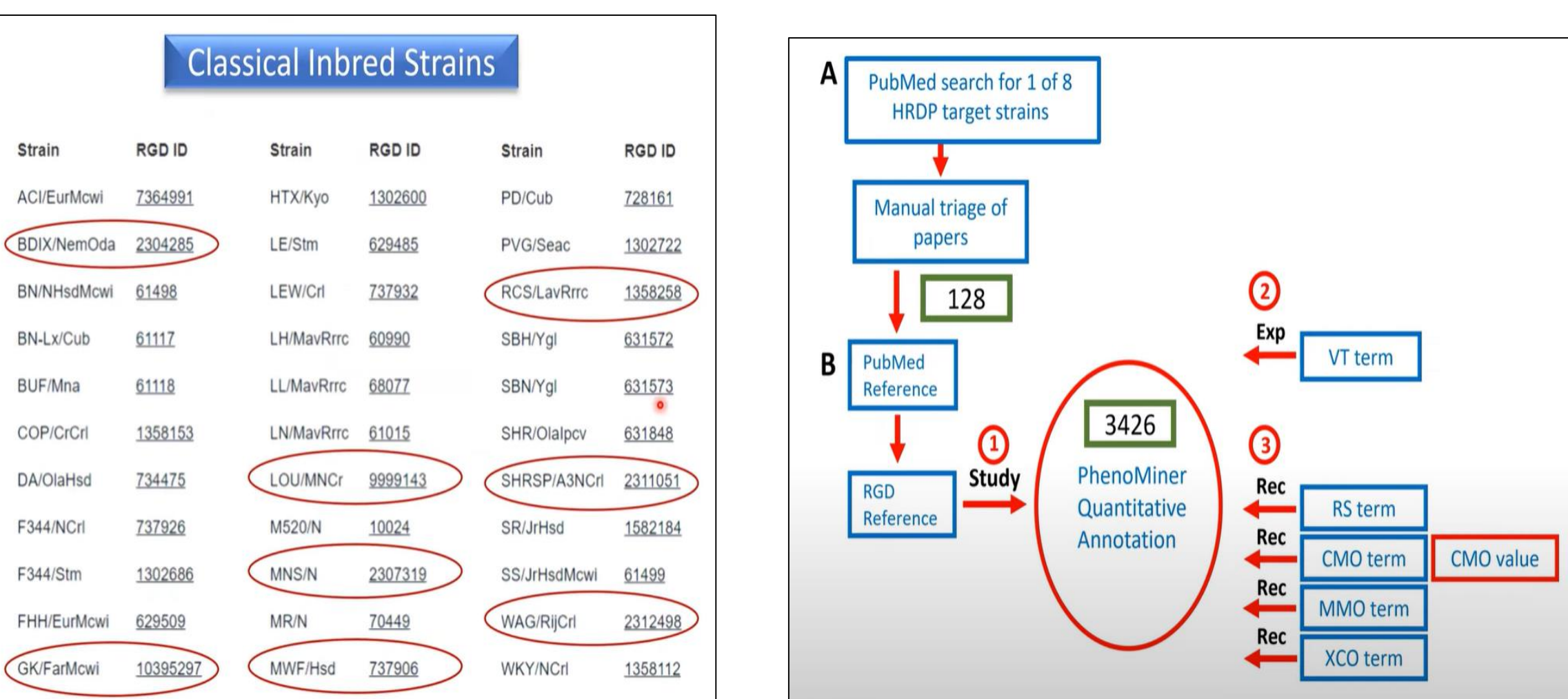
https://rgd.mcw.edu

ABSTRACT

RGD (https://rgd.mcw.edu) is the primary online resource for laboratory rat genetic, genomic, disease and phenotypic data. The PhenoMiner database and mining tool components were developed for rat quantitative phenotype measurements, from both manual curation of scientific literature as well as uploaded data provided by investigators. Data includes detailed information about what (CMO, clinical measurement ontology), how (MMO, measurement method ontology), and under what conditions (XCO, experimental conditions ontology) phenotypes were measured in what animals (species, strain) for each measurement value. Recently, a concerted effort of manual data curation was undertaken to increase data from scientific literature for HRDP founder strains, especially those underrepresented in the data repository. In addition, based on user feedback, the data mining tool was reworked to improve the user interface for data interactivity. In particular, improvements in the search functionality help make navigating the complex datasets more straightforward. The query results page was redesigned to simplify filtering of the data returned, to facilitate tailoring specific query results. New functionality makes it possible to view results for related terms with the same unit of measure in the same graph. The tool provides the ability to download either filtered results or all results matching the user's original query, providing users with documentation at multiple levels. Upcoming planned improvements will include the availability of imported high throughput phenotyping data from individual inbred and outbred (i.e., heterogeneous stock) rats that will utilize the new graphing functions. The improvements in the PhenoMiner user interface will enable better data access, filtering, and visualization, and continued ability to download data for further analyses.

PhenoMiner is an ontology-structured database of <u>quantitative</u> phenotype data for the laboratory rat from scientific literature (manual curation) and from high-throughput phenotyping projects (imported, provided by investigators).

RGD Homepage

PhenoMiner is accessible from either the Analysis & Visualization dropdown menu, or from the Phenotypes and Models bucket on the RGD homepage
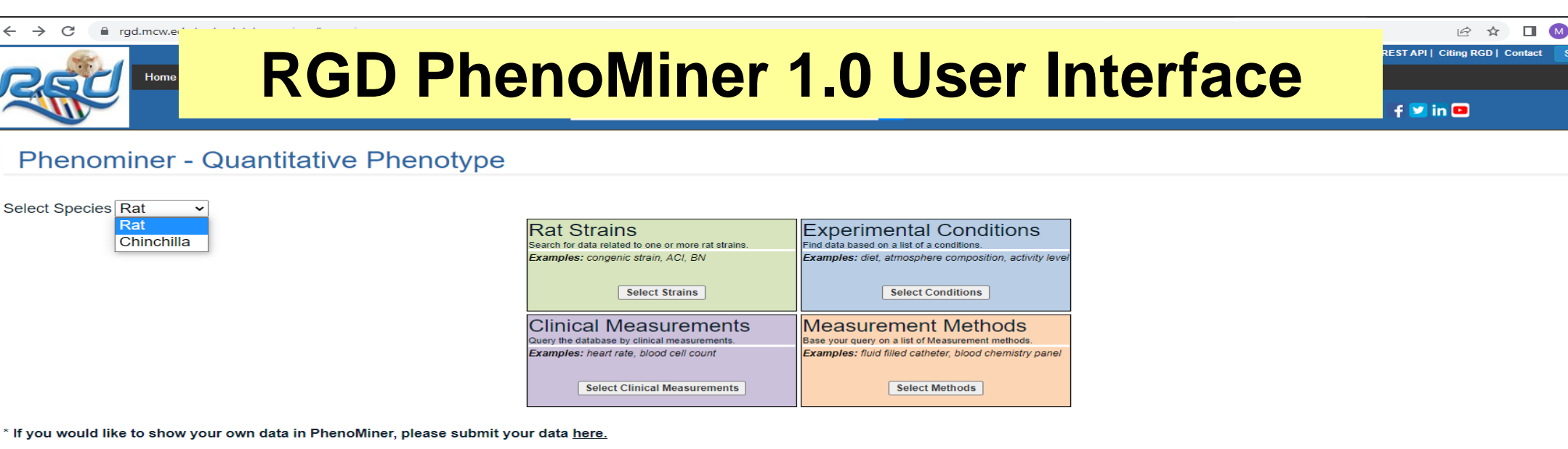
Manual data curation was undertaken to increase data from the scientific literature for HRDP founder strains, especially those eight circled below which were underrepresented or absent in the data repository.

Classical Inbred Strains

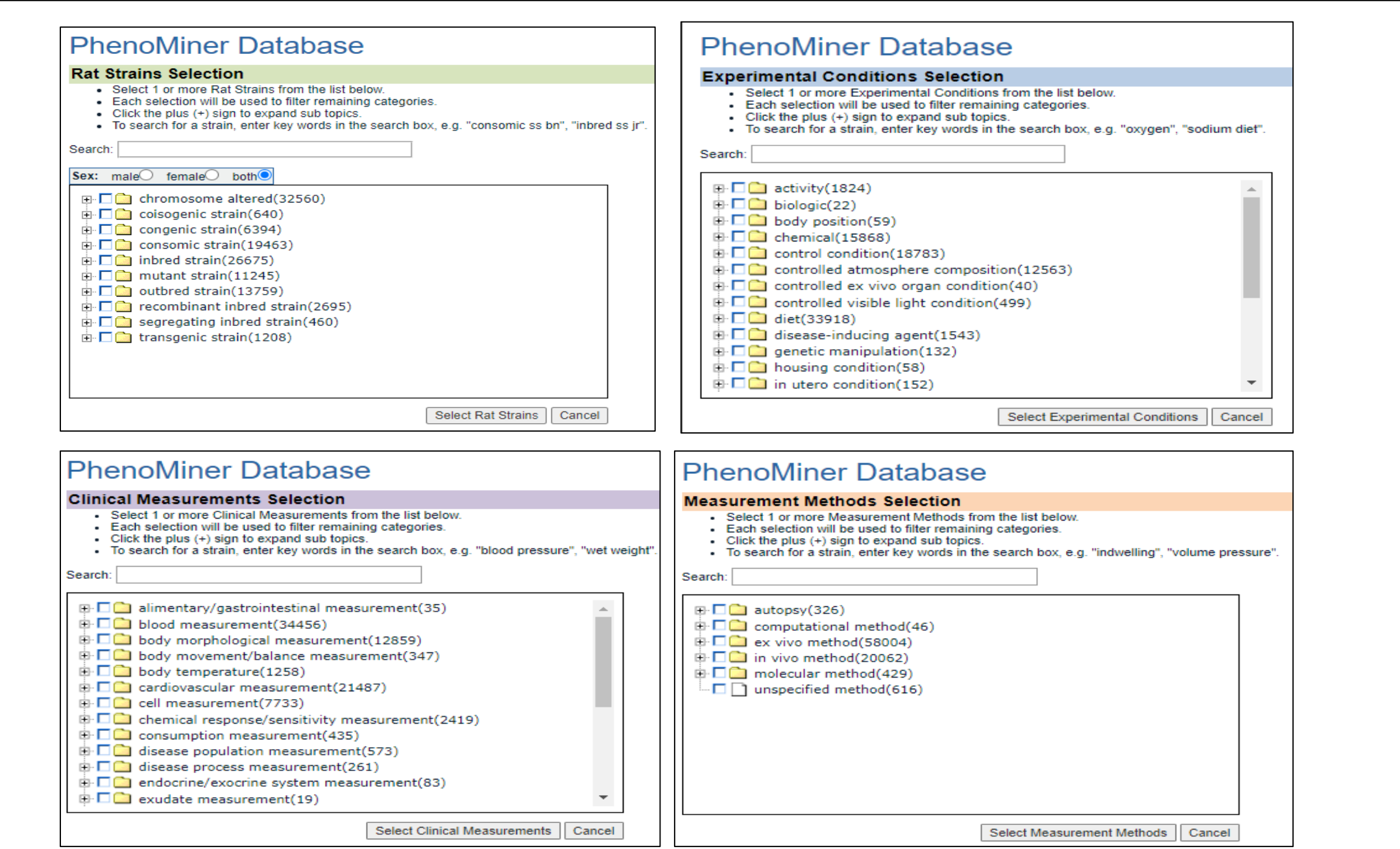| Strain	RGD ID	Strain	RGD ID	Strain	RGD ID
ACI/EurMcwi	7364991	HTX/Kyo	1358204	PD/Cub	
BDIX/NemOda	2304265	LE/Stm	629485	PVG/Seac	1302722
BN/NHsdMcwi	61498	LEW/Crl	737932	RCS/LavRrrc	1358268
BN-Lx/Cub	61117	LH/MavRrrc	68077	SBN/Ygl	631573
BUF/Mna	61118	LL/MavRrrc	68077	SHR/NCrl	631665
COP/Cr/Crl	1358343	LN/MavRrrc	61113	SHR/Olalpcv	631848
DA/OlaHsd	734475	LOU/MWcj	9999143	SHRSP/A3NcjC	2311051
F344/Mcwi	737299	MS20/N	10024	SR/JrHsd	1582184
F344/Stm	1302808	MNS/N	2307319	SS/JrHsdMcwi	61498
FHH/EurMcwi	629509	MR/N	70449	WAG/RijCrl	2312499
GK/FarMcwi	10395297	MWF/Hsd	737306	WKY/NCrl	1358112

Data mining tools for flexible querying are built into the database.

RGD PhenoMiner 1.0 User Interface

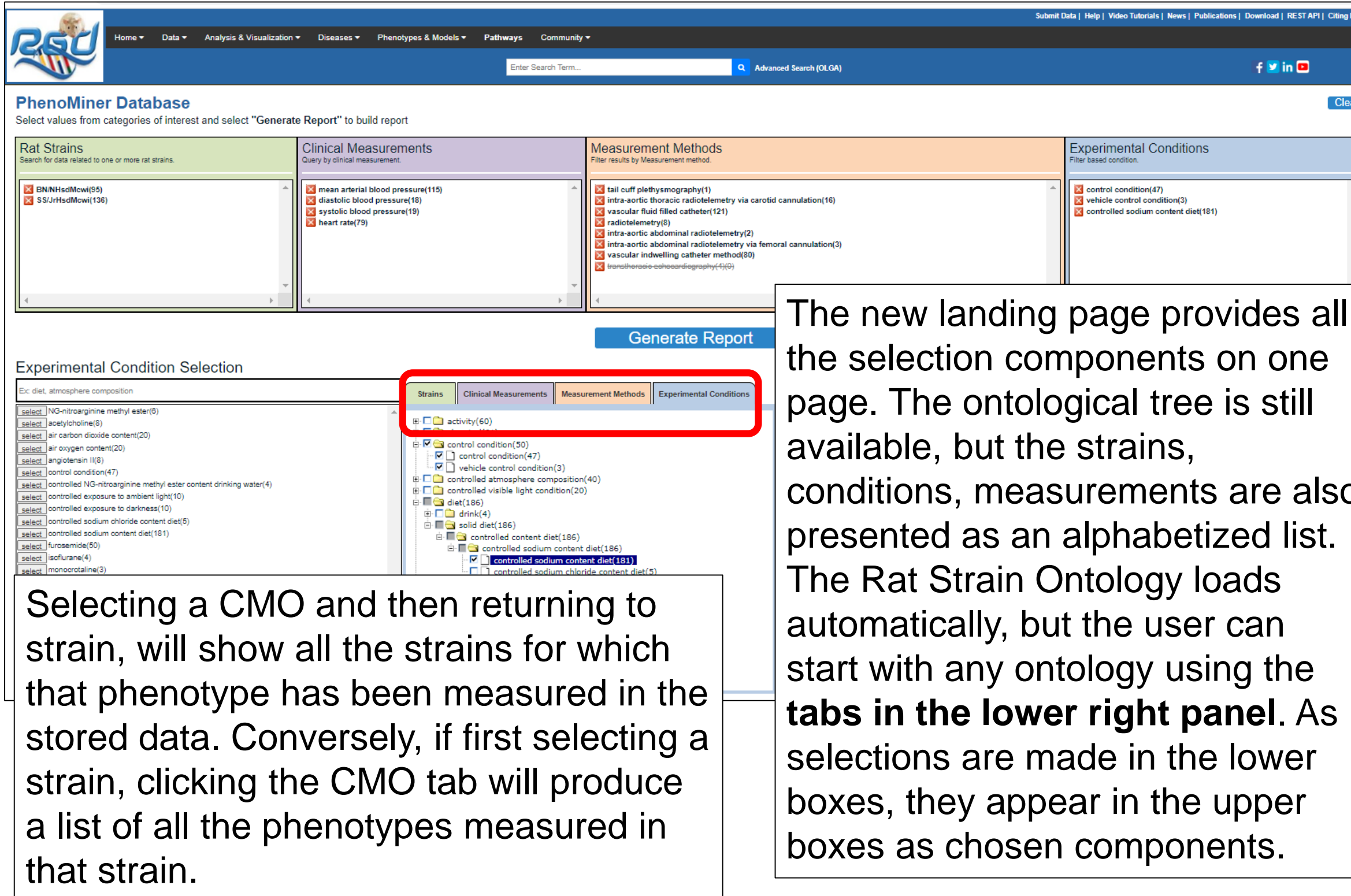The ontologies used to organize the data in the database include:
- Rat Strain Ontology (RSO) for animals measured (SHR, WKY, etc.).
- Vertebrate Trait Ontology (VTO) for organizing related phenotypes within a study (arterial blood pressure trait, etc.).
- Clinical Measurement Ontology (CMO) for what specific phenotype was measured (systolic blood pressure, etc.).
- Measurement Method Ontology (MMO) for how the phenotype was measured (indwelling catheter, tail-cuff, etc.).
- Experimental Condition Ontology (XCO) for recording the experimental conditions (control, salt diet, etc.).
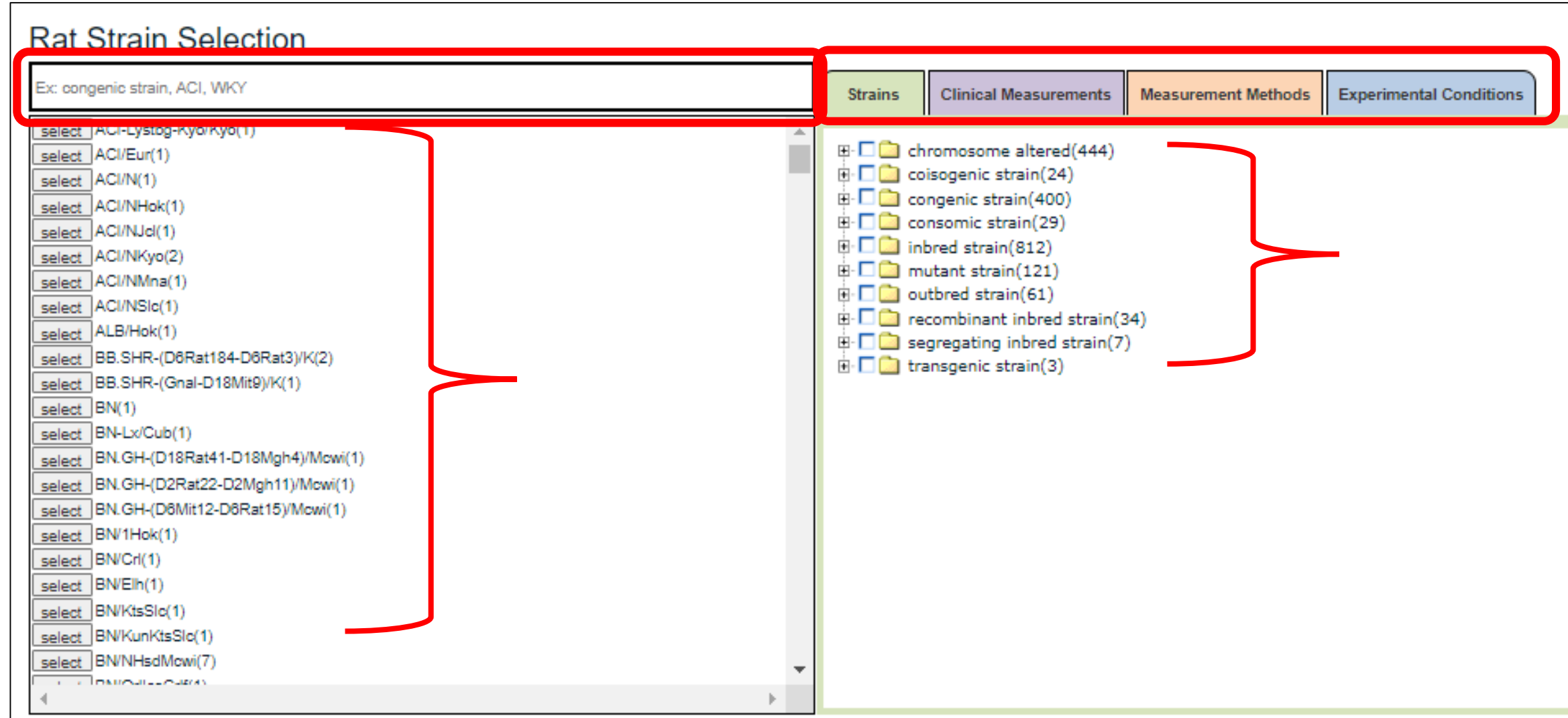

RGD PhenoMiner 2.0 User Interface

The original PhenoMiner landing page was flexible but required the user to navigate the ontology tree for each data category, iteratively. Based on user feedback, the data mining tool has been reworked to improve the user interface for data interactivity.
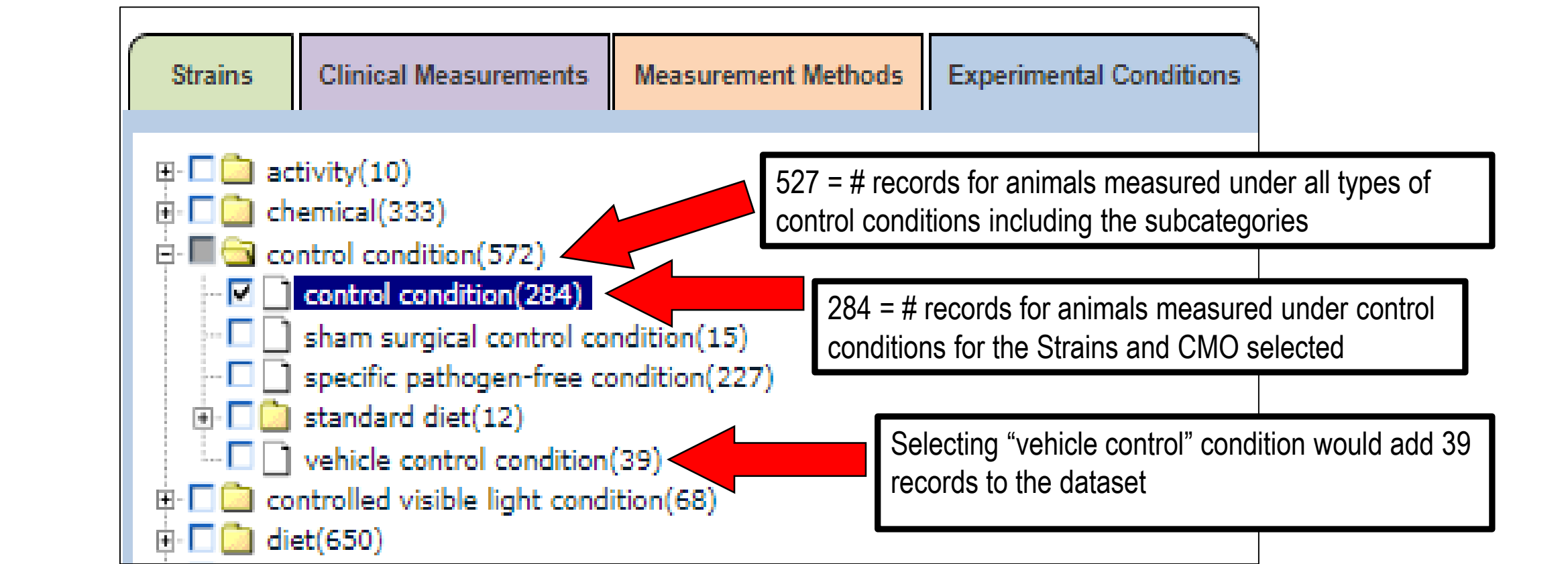

The new landing page provides all the selection components on one page. The ontological tree is still available, but the strains, conditions, measurements are also presented as an alphabetized list. The Rat Strain Ontology loads automatically, but the user can start with any ontology using the **tabs in the lower right panel**. As selections are made in the lower boxes, they appear in the upper boxes as chosen components.

Selecting a CMO and then returning to strain, will show all the strains for which that phenotype has been measured in the stored data. Conversely, if first selecting a strain, clicking the CMO tab will produce a list of all the phenotypes measured in that strain.

Each ontology, selected using the tabs in the right-hand box, provides 3 navigation methods: entering a term in the primary search box, or using the alphabetical list below the search box and using the "select" button, or using the ontological tree in the right side box.
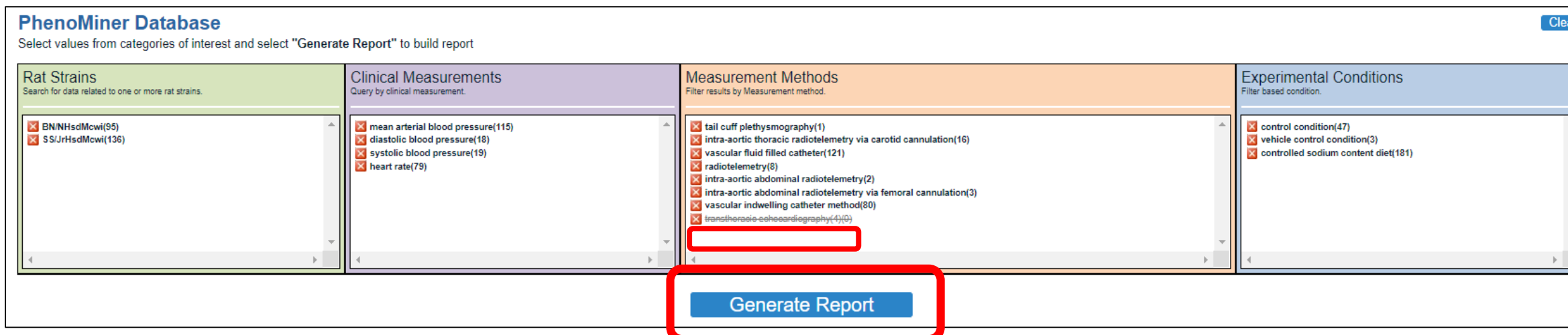

527 = # records for animals measured under all types of control conditions including the subcategories

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

Selecting "vehicle control condition" would add 39 records to the dataset
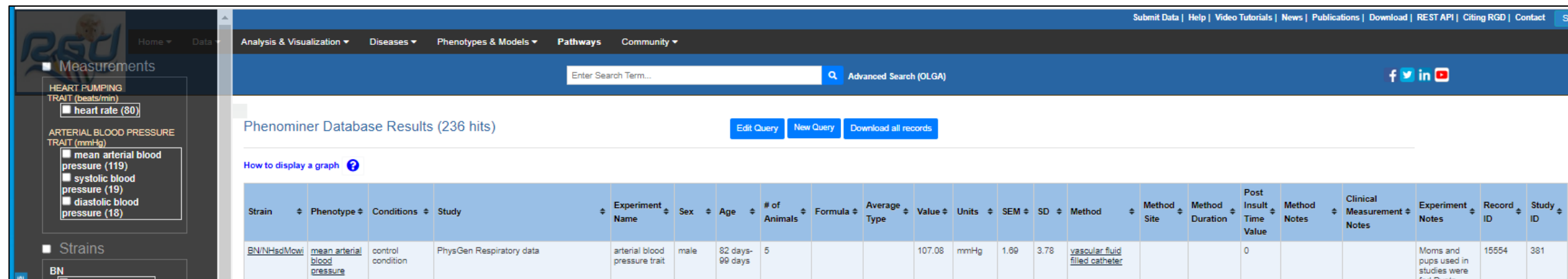
Further tailoring of the data query can be accomplished utilizing the red X boxes that if clicked will remove that term from the report results. Note that a term for which there is no data after selecting terms in a different category, will show in the bucket with a strike out line.
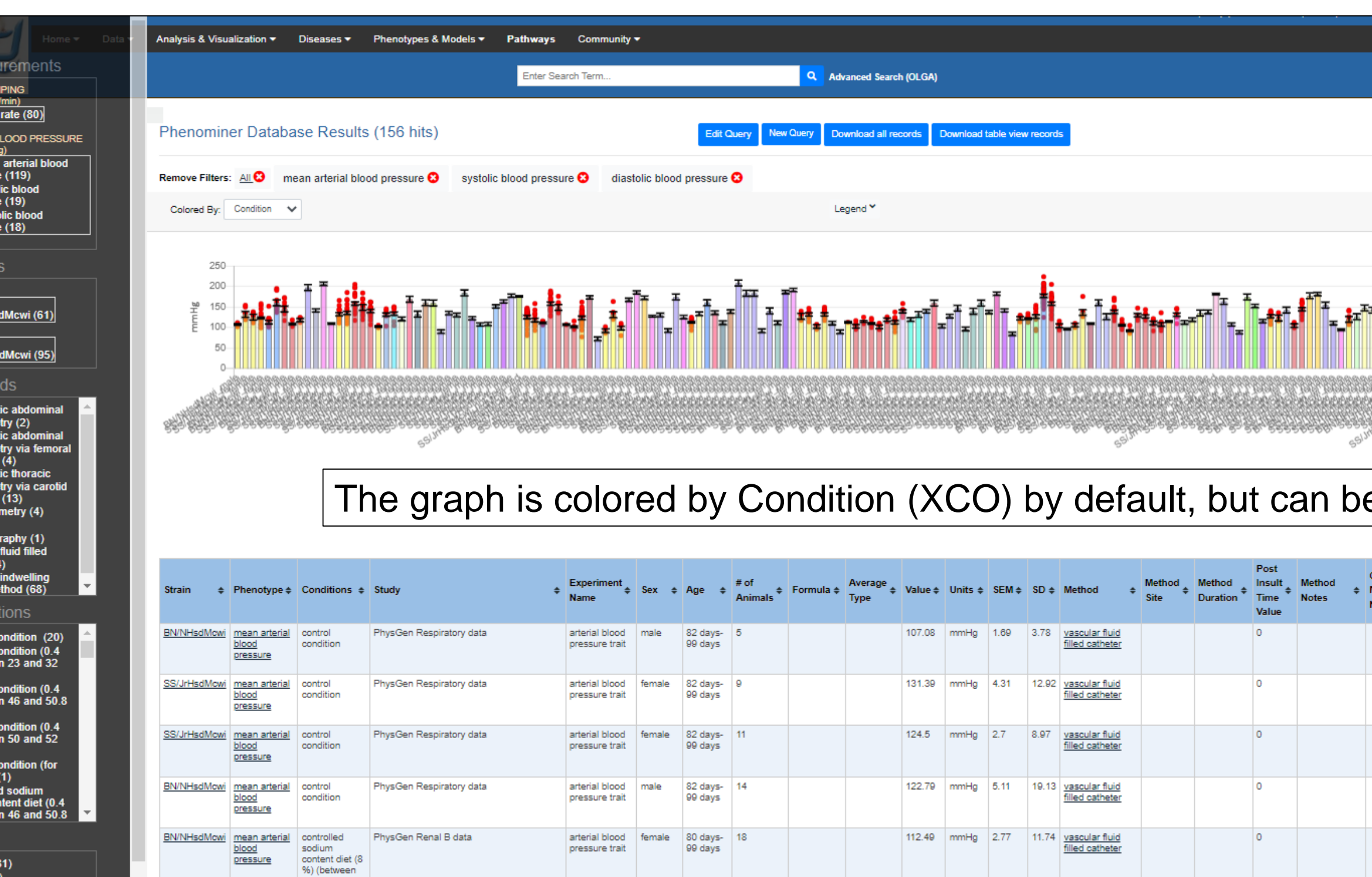Once satisfied with the query parameters, click the Generate Report button.

The report page cannot generate the graph if CMO terms with more than one unit of measure are selected.
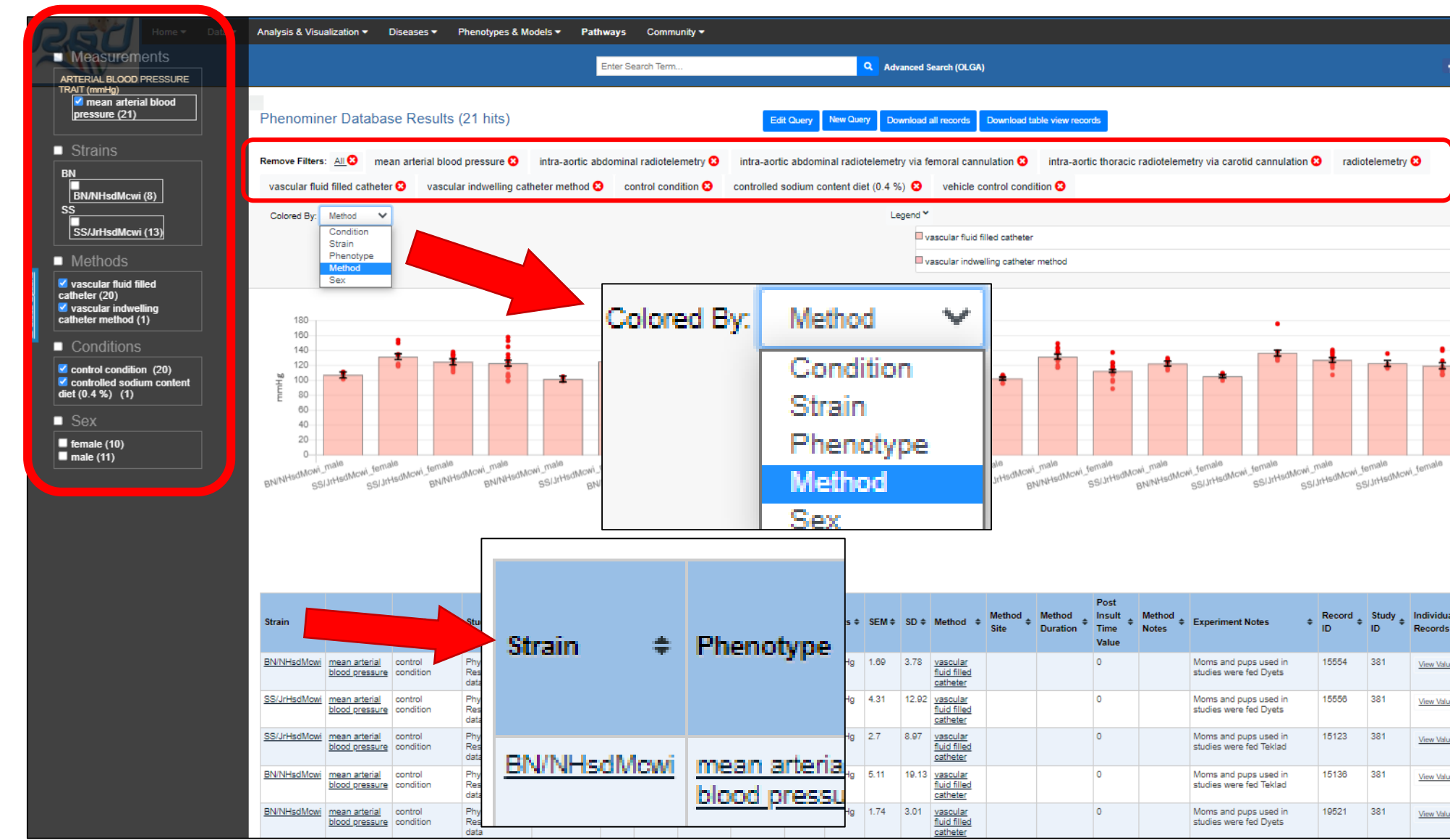
In this case, using the left side menu to remove the heart rate term and select only blood pressure terms allows the graph to be generated, even though it is still multiple terms.
The report page consists of the graph of the CMO term(s), a table of the displayed data along with all other fields stored for those datapoints, and a menu of terms on the left to facilitate further filtering of results.
Note that individual rat phenotype values are graphed with red dots, in addition to the mean and SEM for the bar.

The graph is colored by Condition (XCO) by default, but can be changed.

Without leaving the report page or generating a new query, the user can further tailor the results using the menu on the left. Choosing an ontology and unselecting items will alter the data in the table and graph. Also, the red X's on the filters list at the top can be used to remove terms from the display and table. In this way the user can tailor the data captured.
When selections are made in two or more different categories in the left panel, the terms listed are reduced to only those that apply. To get all the options from the original query back into the lists, use the upper filters and click the red X next to **Remove Filters: All**.
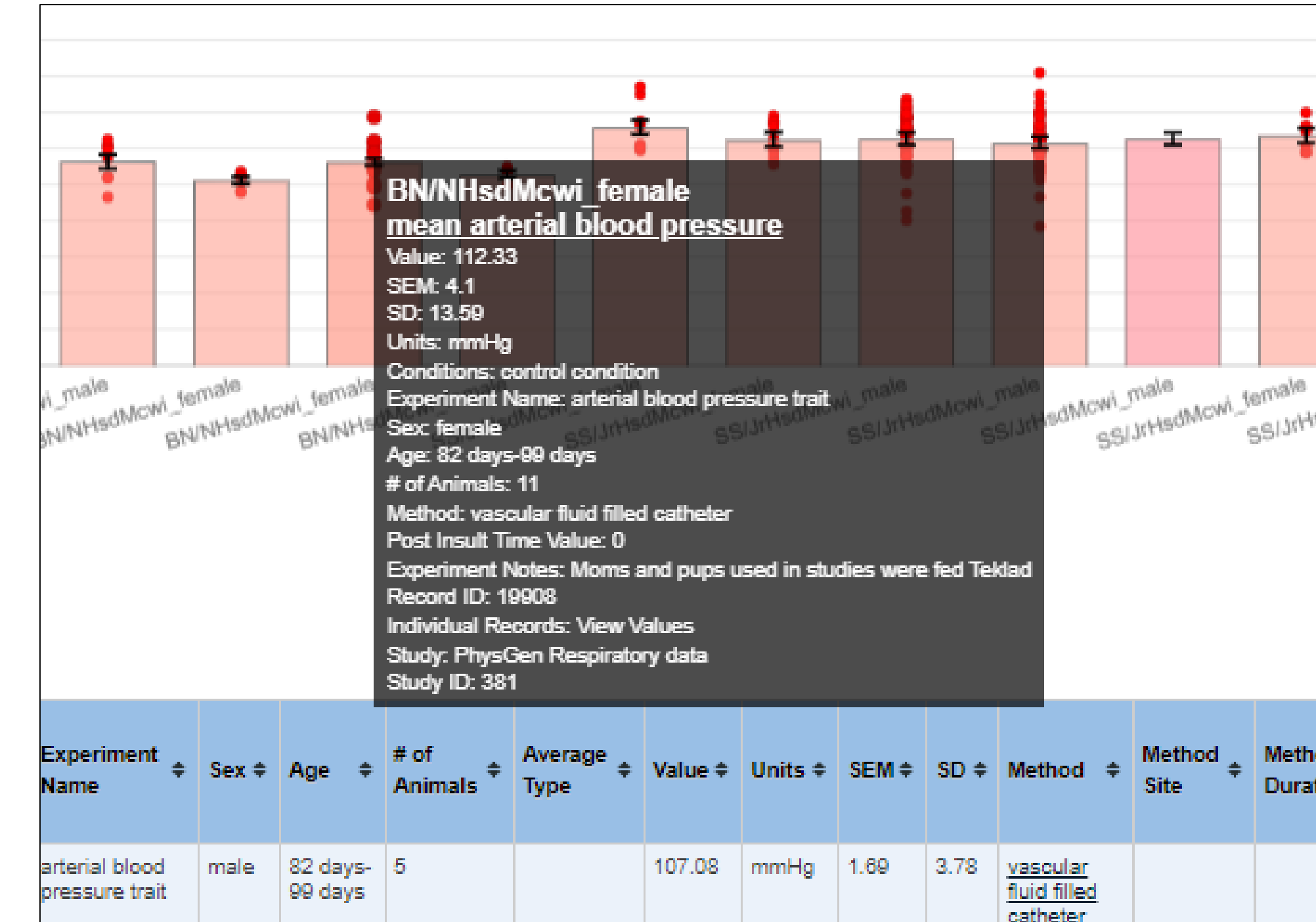

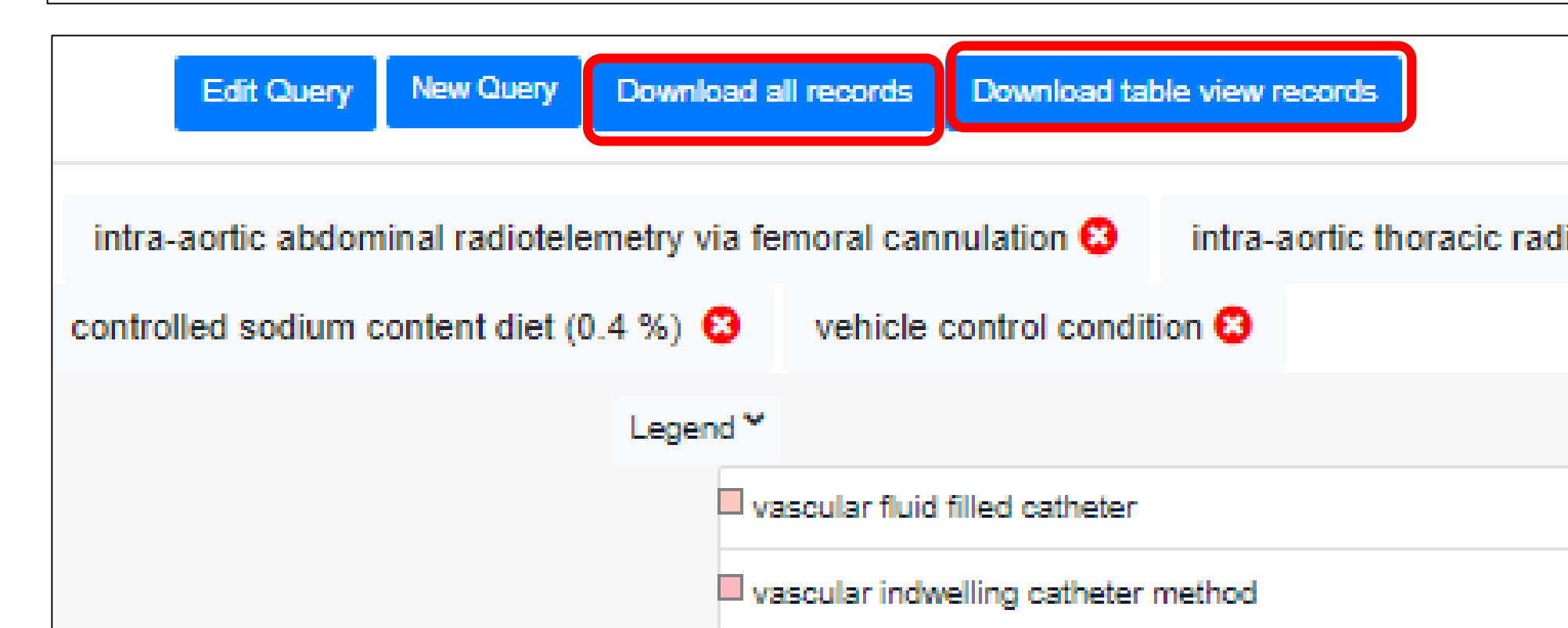Changing the "colored by" will update the graph and legend.

Using the table to alphabetically sort the data by Strain, will also reorder the graph. Sorting the table by Value is also useful.

Hovering the mouse over a dot or bar will display a pop-up with all the related fields stored for that data value.
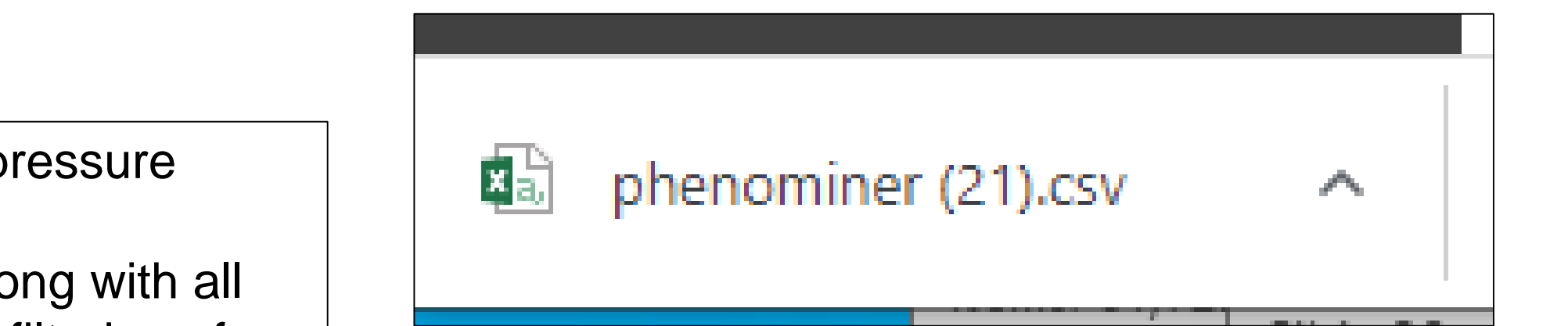

The query report page includes a <u>Download all records</u> button that will create a CSV format file of the report page table. This table includes all the data and parameters stored for the terms selected in the original query. That is, **all** the terms that were in the selection boxes at the time the user clicked the Generate Report button.
The <u>Download table view records</u> button will generate the CSV file for the filtered dataset.
The spreadsheet includes the citation information for each datapoint in the table, to allow the user to retrieve the published papers from which the data was compiled.

Upcoming planned improvements will include the ability to download the high throughput phenotyping data from individual rats. Also inbred and outbred (i.e., heterogeneous stock) rat data is being imported and will utilize the new graphing functions.

All of the improvements in the PhenoMiner user interface will enable better data access, filtering, and visualization, and continued ability to download data for further analyses.