

# A Multi-omics Visualization Platform (MVP) Plug-in for Galaxy-based Applications

Thomas McGowan<sup>1</sup>; James Johnson<sup>1</sup>; Pratik Jagtap<sup>2</sup>; Getiria Onsongo<sup>1</sup>; Candace Guerrero<sup>2</sup>; Timothy Griffin<sup>2</sup>  
1-University of Minnesota Supercomputing Institute, 2-University of Minnesota Center for Mass Spectrometry and Proteomics

## Overview

- We created visualization platform allowing users to inspect and verify multi-omic datasets
- The MVP platform is tightly integrated with the widely used Galaxy framework
- The MVP plugin is built on HTML5, CSS and JavaScript
- Future work will include expanding the MVP plugin to present metabolomic data in addition to genomic and proteomic data

## Introduction

The Galaxy-P project has extended the popular Galaxy bioinformatics framework into new realms, deploying tools for MS-based proteomics data analysis and integrative "multi-omic" applications. Galaxy-P leverages the many advantages offered by the Galaxy operating environment for informatics and data analysis, including flexibility, transparency and accessibility for bench scientists. The MVP visualization tool extends Galaxy-P's advantages into the visualization of large, complex data sets. This allows researchers to quickly inspect and verify the quality of the results as well offer an overview with visualization and a deeper understanding of underlying spectral data. This can be especially valuable when results include inputs from possibly diverse domains.

## Methods

The core of the MVP is based on standard JavaScript and open-source libraries. It receives data from a documented Galaxy SQLite data provider API. The main visualization is integrated into Galaxy via the Galaxy visualizations registry. Once registered, any dataset of type mz.sqlite will automatically be viewable from the MVP tool. The MVP tool uses

- the DataTables [https://www.datatables.net/] library to manage the presentation, sorting and filtering of data
- the Lorikeet MS/MS viewer [https://github.com/UWPR/Lorikeet, https://github.com/jmchilton/lorikeet.git] to visualize spectra, and
- the IGV.js package [https://github.com/igvteam/igv.js.git] to interactively present features of interest.

This enables a researcher to see, in one HTML window, both genomic and proteomic data as they relate to one another.

## Results

The MVP platform is currently installed within our locally hosted Galaxy-P installation. Also, the tools necessary to create the mz.sqlite data files are in the main Galaxy tool shed. The MVP platform is under active development as we add functionality.

With the incorporation of Integrated Genomics Viewer (IGV) and Lorikeet, the MVP platform is already merging proteomic and genomic results into a single, accessible output. A user can, with relatively few keystrokes, filter and order large datasets down to a manageable subset. Due to the tools use of server-side caching, large data sets are handled as quickly as small datasets.

The MVP tool is now both a producer and consumer of Galaxy data. At invocation, the tool has access to the mz.sqlite dataset via the Galaxy API. Once a user has filtered and sorted the data, the tool can send a subset of the results back to Galaxy. There it is incorporated into the user's history and is available as input for Galaxy tools.

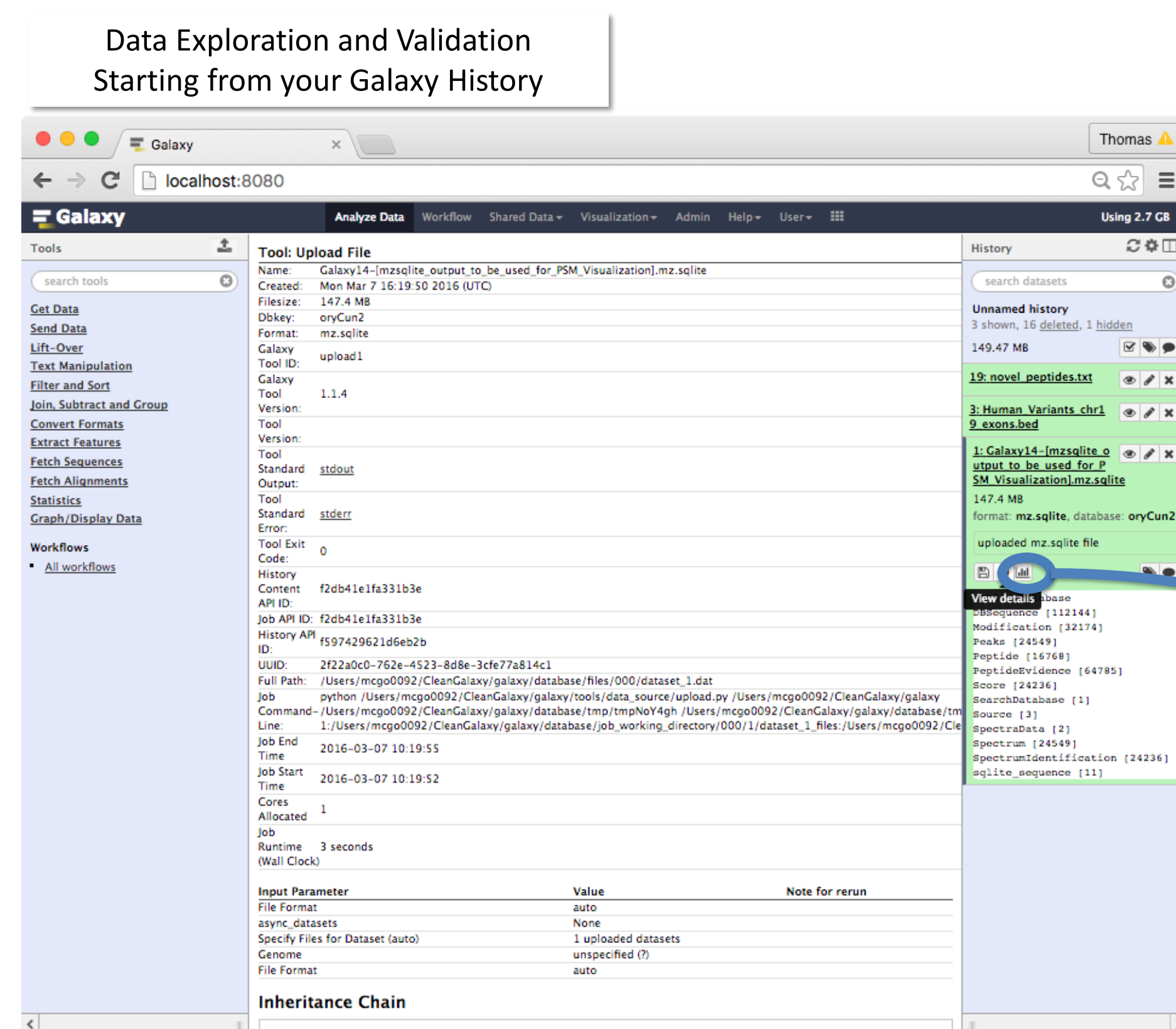
We have also added the ability to save and share the MVP tool's state. Using the Galaxy saved visualization API, a user can save a visualization back to Galaxy in the exact state seen on the browser window. The current filtered and sorted rows, the current MS/MS scans can all be saved to Galaxy. Once saved, the visualization can be shared with other Galaxy users via the standard Galaxy functionality. The shared visualization opens exactly as it was saved, and it is fully functional for further data investigation.

We are actively using the MVP tool for Galaxy workflows applied to a number of biological projects. Most of these integrate genomic/transcriptomic information as well as proteomic data. The tool provides essential functionality for visualizing and interpreting results from these multi-omic studies.

## Conclusions

- Currently, the MVP tool allows a user to quickly
- Visualize large datasets
  - Filter large datasets based on multiple criteria
  - Verify protein identifications

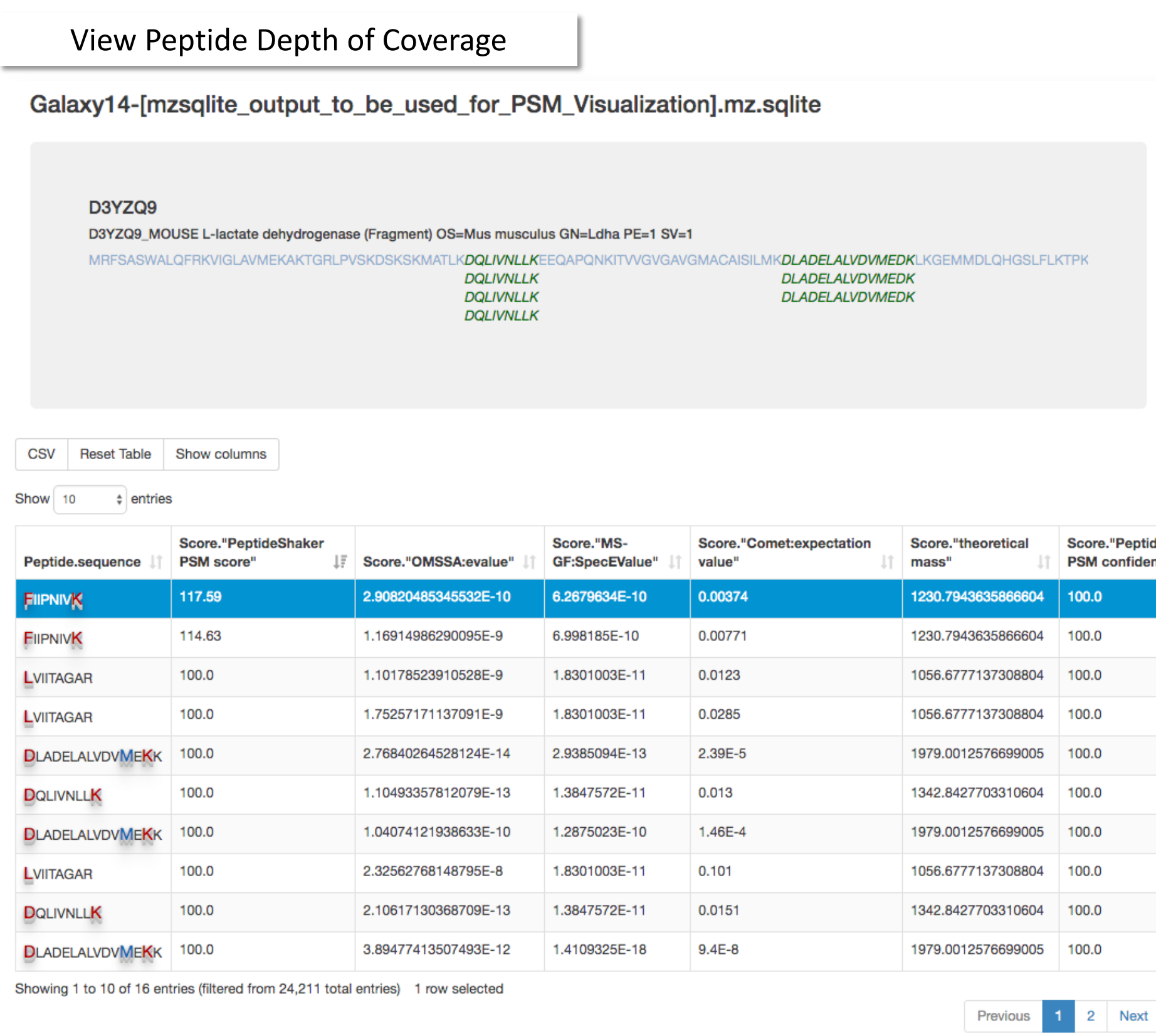
- Future work will include
- Addition of metabolomics tools
  - More robust interaction with Galaxy via saving and sharing visualization state



MVP Viewer is invoked from the existing Galaxy visualization framework

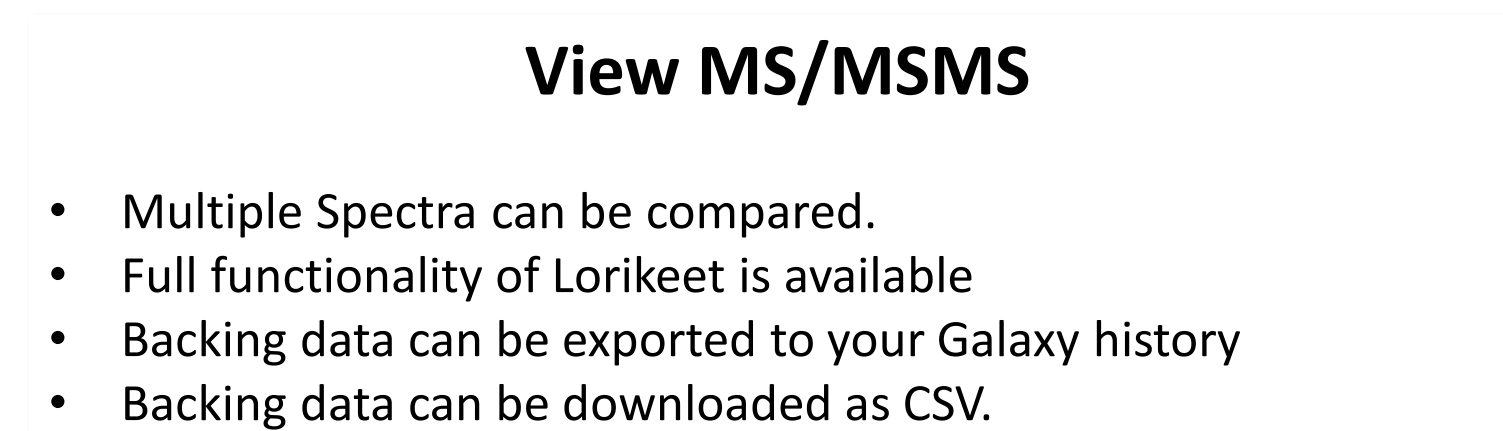
The MVP Viewer has API access back to the originating Galaxy history.

- User can send data back to their history
- User can access other datasets from their Galaxy history
- User will be able to save the entire MVP, including state, back to Galaxy
- User will be able to share a saved MVP with other Galaxy users

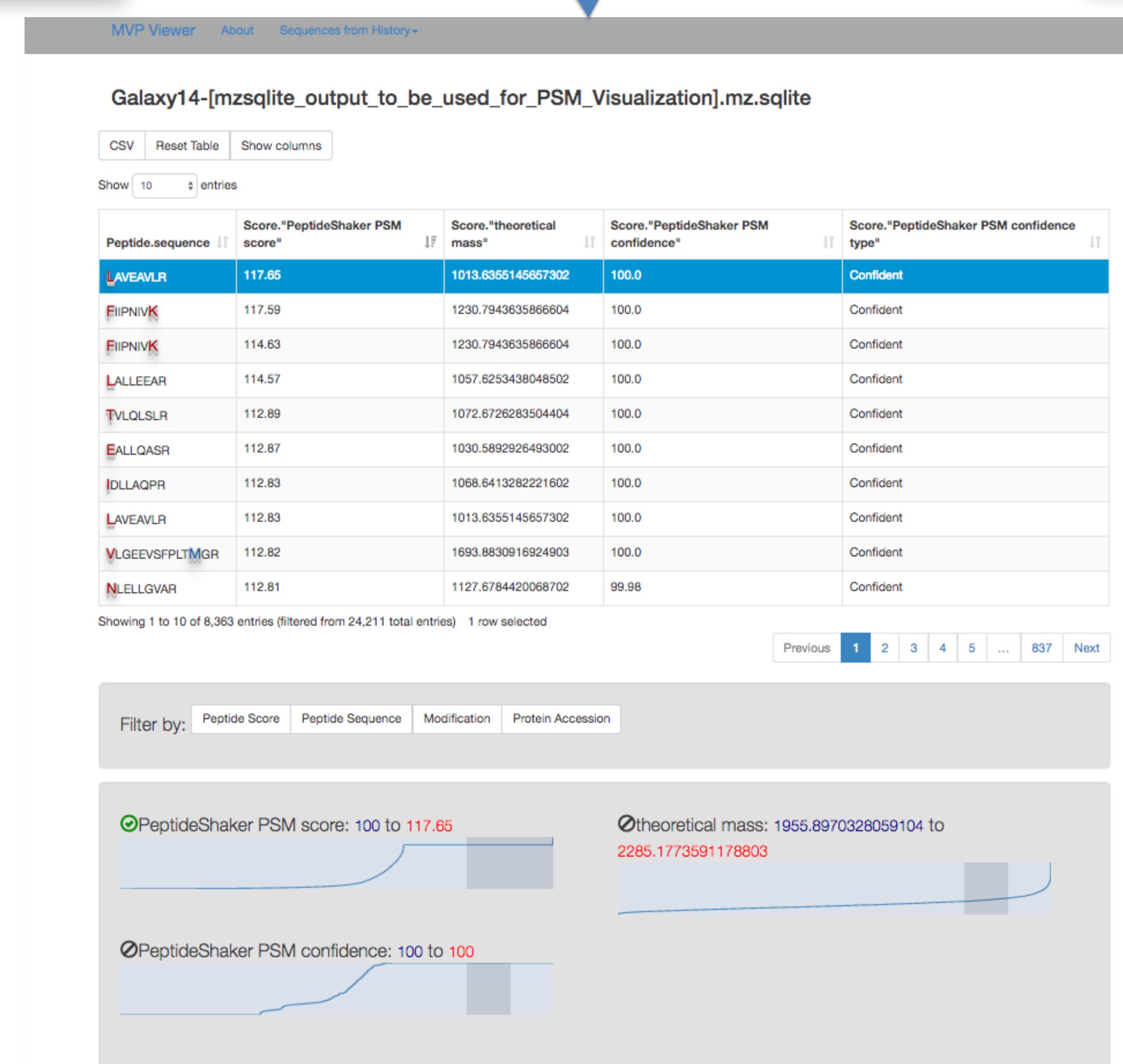


Filter on Galaxy Data Sets

- Access tabular data sets from history
- Select single or multiple sequences for data search

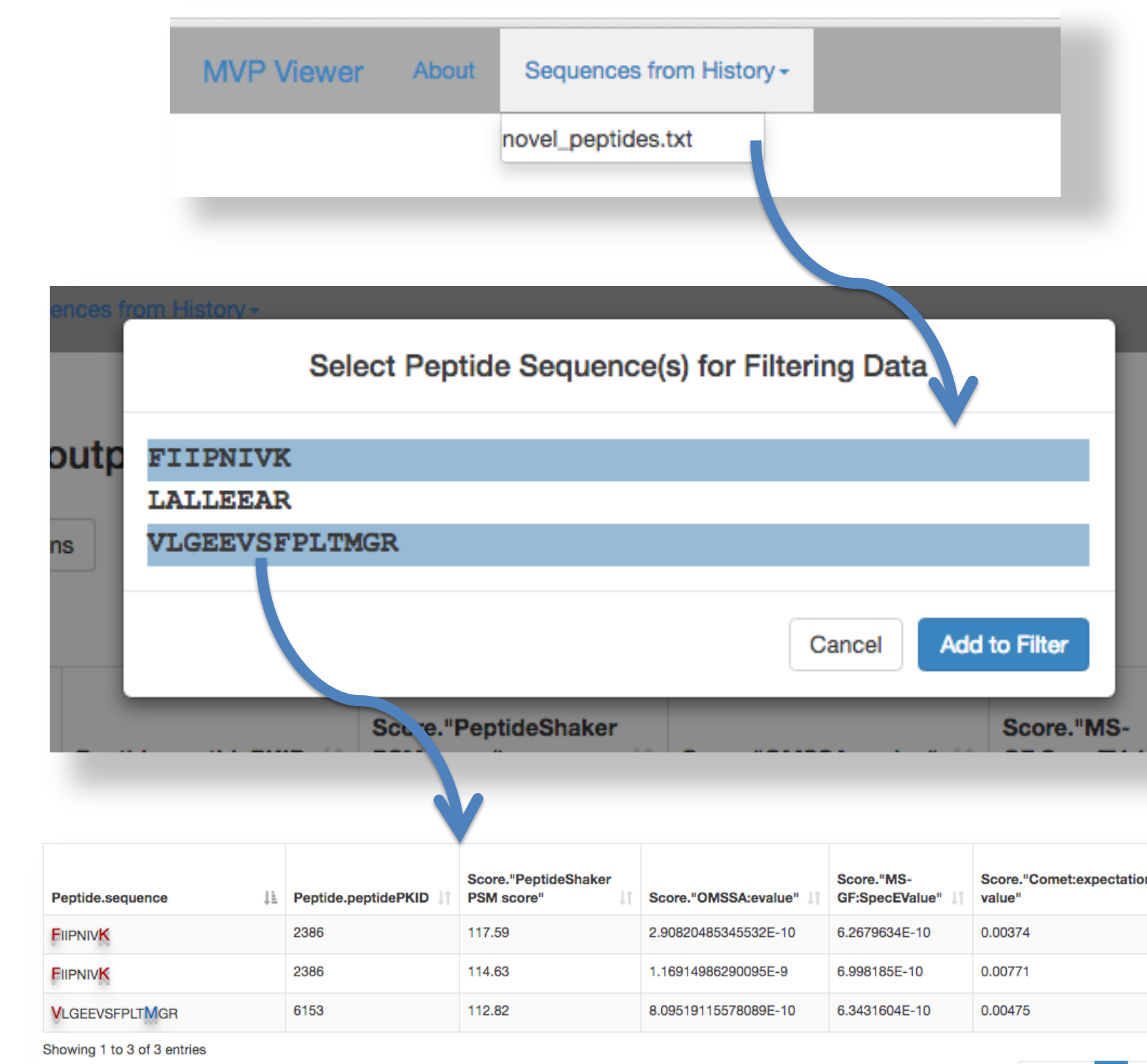


- Multiple Spectra can be compared.
- Full functionality of Lorikeet is available
- Backing data can be exported to your Galaxy history
- Backing data can be downloaded as CSV.

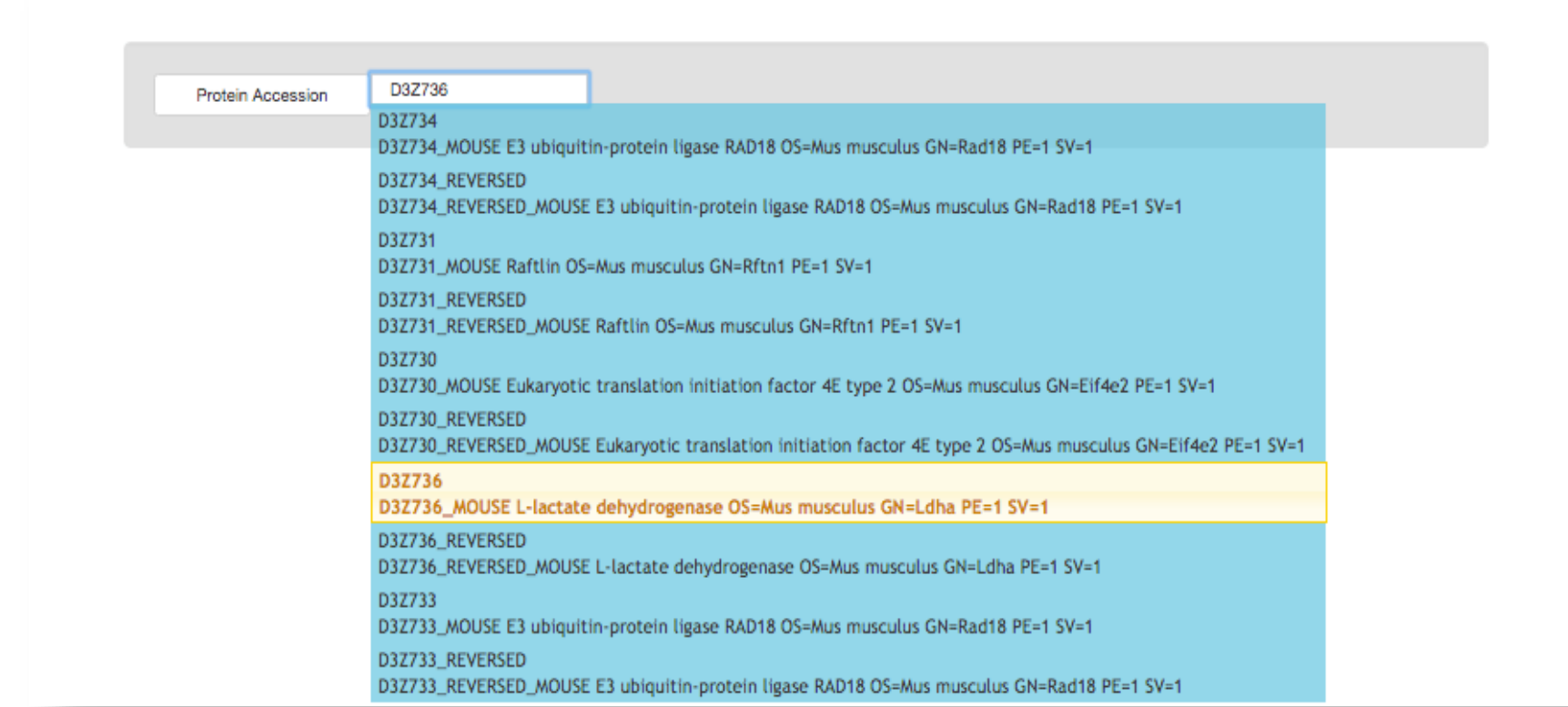


Filter on Modifications

- Score fields are dynamic
- Score distributions are graphed
- User can choose which scores are viewed/filtered
- Score filters are connected via AND



Filter on Accessions



- Filter on Modifications
- All modifications from the data set are presented
- Filtering is based on point and click

