Enhancing the Multi-omics Visualization Platform (MVP) Plug-in for Galaxy-based Applications

Thomas McGowan1; James Johnson1; Pratik Jagtap1; Praveen Kumar1; Getinia Onsongo2; Candace Guerrero3; Timothy Griffin1,4

1University of Minnesota Supercomputing Institute, 2Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 3BICB (Bioinformatics and Computational Biology), University of Minnesota, 4University of Minnesota Center for Mass Spectrometry and Proteomics

Overview

- We created visualization platform allowing users to inspect and verify multi-omics datasets.
- The MVP platform is tightly integrated with the widely used Galaxy framework.
- We have enabled the MVP application to orient a user between an MSMS scan, a peptide, a protein and a genomic exon location on one screen.
- In addition, we now visualize differences between a reference genome sequence and a variant search sequence. This includes substitutions, deletions, additions and multiple exon origins.

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 as offer an overview with visualization and a deeper understanding of underlying spectral data.

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. This year, we have been focusing on performance gains via code optimization and adding some functionality on the server.

Preliminary Data

The MVP platform is currently installed within our locally hosted Galaxy-P installation. The tools necessary to create the mz.sqlite data files are in the main Galaxy tool shed. This year, we have been developing a second SQLite schema to house genome to protein to peptide mapping. The MVP tool uses this schema to navigate between genome and the proteome. With the incorporation of Integrated Genomics Viewer (IGV) and Lorikeet, the MVP platform is already merging proteomics and genome results into a single, accessible output. A user can, with relatively few keystrokes, filter and order large datasets down to a manageable subset. Large data sets are handled as quickly as small datasets. The MVP tool allows for quick entry into experimental results from either a peptide sequence view, a PSM scoring view or a protein-centric view. The peptide view allows for rapid filtering based on sequences of interest. The PSM score view allows for fast filtering of scores yielding the lowest decoy hit rate. And the protein-centric view allows for an immediate examination of proteins of interest. This may be of particular interest to users who have curated variant proteomes.

Acknowledgement

We acknowledge NSF grant 145924, and NIH grant U24CA199347 for funding support to the Galaxy-P team at the University of Minnesota.