GALAXY-P: RECENT DEVELOPMENTS AND EMERGING APPLICATIONS

Pratik D. Jagtap1, James E. Johnson1, Thomas McGowan1, Innocent Onsongo1, Benjamin Lynch1, Candace R. Guerrero1, Kevin Murray1, Lloyd M. Smith2, Michael R. Shortreed2, Anthony J. Cesnik2, Lennart Martens3, Adrian D. Hegeman1 and Timothy J. Griffin1

1University of Minnesota, Minneapolis, MN; 2University of Wisconsin-Madison, Madison, WI; 3Ghent University/VIB, Ghent, Belgium

Introduction. The Galaxy for Proteomics (Galaxy-P) project was launched several years ago, with the objective of extending the genomics-centric Galaxy bioinformatics framework (Genome Biol. 11: R86) to employ proteomics informatics tools. Since its inception, the Galaxy-P project has moved its focus from proteomics tools to integrative analysis across different `omic domains (i.e. multi-omics). The Galaxy software framework offers numerous advantages as a platform for multi-omics data analysis and informatics (Nat Biotechnol 33: 137). These include the flexibility to implement and integrate disparate software programs that cross `omic domains, scalability for large dataset volumes and compute-intensive operations, and easy sharing of tools and complete workflows, even those comprised of complicated, multiple step processes.

Results. Here we outline the current state of the Galaxy-P project, summarizing recent developments and emerging future plans in multi-omic data analysis and informatics. Areas of active development described include:

- Results visualization and interpretation
- MS-based metabolomics data analysis and informatics
- Integrative genomic-proteomic data analysis and informatics
- Collaborations and outreach activities

Conclusions. Galaxy-P has enabled numerous research studies in multi-omics. Emerging and future developments will focus on enhancing its capabilities, especially in the realm of visualization and interpretation of results, and dissemination of the high-value workflows and tools to the community.

Current status. MVP has been developed as a Galaxy-compatible plug-in for visualization and interpretation of multi-omics results generated in Galaxy data analysis workflows. MVP utilizes standard JavaScript and open-source libraries, receiving data from a Galaxy SQLite data provider API. The plugin integrates with the Galaxy visualizations registry, such that any registered data of type mz.sqlite will be viewable from the MVP tool. Functions within MVP currently include:

- Sorting and organization of data by peptide sequence, protein accession or other annotation
- Visualization of MS/MS data via the Lorikeet viewer
- Integration with the Integrated Genome Viewer (IGV) utilizing the IGV JavaScript package for mapping and viewing peptide sequences against genomes and transcriptomes.

Future extensions will include:

- Enhanced functionalities for filtering peptide sequence matches and post-translational modifications
- Added functionalities for viewing and interpreting MS-based metabolomics data
- Connectivity to web-based informatics resources to enable results interpretation (e.g. CBioPortal for cancer informatics, NDEx for pathway analysis)

Emerging and future work. MVP has been developed with an eye toward extensibility, enabling not only visualization of data and results, but also connectivity to informatics resources to aid in interpretation. Future extensions will include:

- Flexibility and accessibility for improved proteogenomic analysis
- Implementation of additional tools to address challenges in metaproteomics
- Deployment of additional tools to address challenges in metaproteomics
- Integration with the Galaxy visualizations registry, such that any registered data of type mz.sqlite will be viewable from the MVP tool
- Functions within MVP currently include:
  - Sorting and organization of data by peptide sequence, protein accession or other annotation
  - Visualization of MS/MS data via the Lorikeet viewer
  - Integration with the Integrated Genome Viewer (IGV) utilizing the IGV JavaScript package for mapping and viewing peptide sequences against genomes and transcriptomes.

Future extensions will include:

- Enhanced functionalities for filtering peptide sequence matches and post-translational modifications
- Added functionalities for viewing and interpreting MS-based metabolomics data
- Connectivity to web-based informatics resources to enable results interpretation (e.g. CBioPortal for cancer informatics, NDEx for pathway analysis)

Emerging and future work. MVP has been developed with an eye toward extensibility, enabling not only visualization of data and results, but also connectivity to informatics resources to aid in interpretation. Future extensions will include:

- Enhanced functionalities for filtering peptide sequence matches and post-translational modifications
- Added functionalities for viewing and interpreting MS-based metabolomics data
- Connectivity to web-based informatics resources to enable results interpretation (e.g. CBioPortal for cancer informatics, NDEx for pathway analysis)

Collaborations, outreach and training

Emerging and future work. A main focus going forward is making the Galaxy-based multi-omics tools accessible by more researchers. Two main avenues for these dissemination include:

- We are actively working with Globus Genomics to implement our Galaxy-P instance in cloud infrastructure backed by Amazon Web Services. The Globus instance will be used for training purposes and as a scalable option for collaborative, large-scale studies
- We are leveraging the Docker technology, to create Galaxy-P "Flavours", which are customized instances that can be downloaded and easily installed on local infrastructure or implemented in cloud-based infrastructure

Acknowledgements

We thank Mark Vaudel, Harald Barnsnes, Björn Grüning and Ira Cooke for assistance in optimizing and deploying software tools. We thank the Galaxy development community and core Galaxy team for continued innovation and improvement of the framework. We also acknowledge NSF grants 1147079 and 1458524, and NIH grant 1U24CA199347 for funding support.