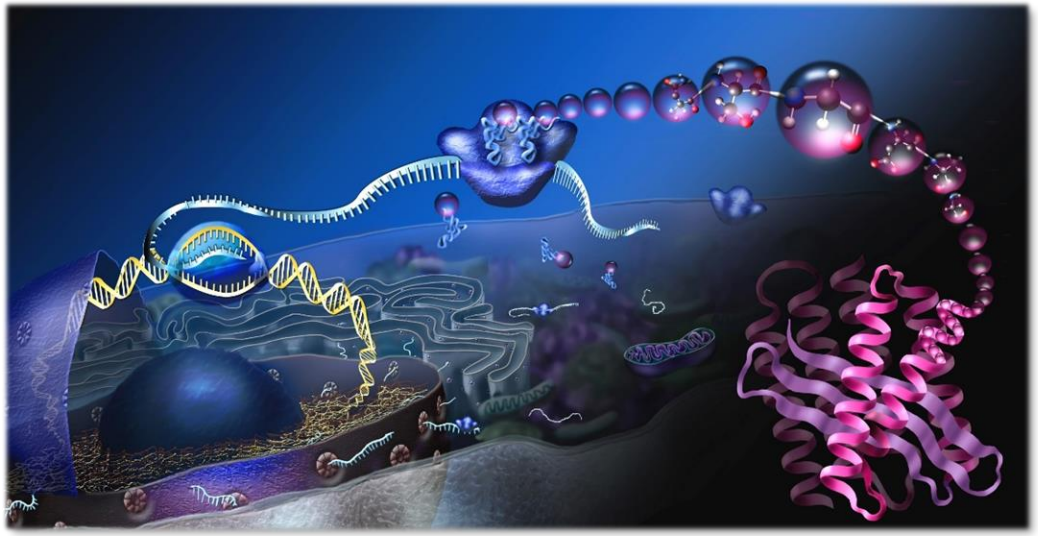
Multi-Omics with Galaxy for Diverse Biological Applications

Tim Griffin and Pratik Jagtap
University of Minnesota



GalaxyP
galaxyp.org

Outline

- Galaxy-P and mass spectrometry-based proteomics multi-omics data analysis
- Multi-omics application 1: Proteogenomics
- Multi-omics application 2: Metaproteomics
- Access and Questions

Acknowledgements

University of Minnesota
Biochemistry, Molecular Biology & Biophysics

GalaxyP

Dr. Pratik Jagtap (Co-leader, Galaxy-P)
Praveen Kumar
Subina Mehta
Caleb Easterly
Ray Sajulga
Andrew Rajczewski
Dr. Shane Hubler
Mark Esler
Dr. Art Eschenlauer
Dr. Candace Guerrero
Matt Chambers
Marie Crane
Emma Leith

Brian Crooker
Wanda Weber
Matt Andrews
Katie Vermillion

Minnesota Supercomputing Institute

James Johnson
Tom McGowan
Dr. Getiria Onsongo
Dr. Michael Milligan

Norwegian University of Life Sciences

Magnus Arntzen
Francesco Delogu
Live H. Hagen
Phil Pope

Benoit Kunath
NMBU/University of Luxembourg

COMMUNITY-BASED SOFTWARE DEVELOPMENT

Bjoern Guening and Bérénice Batut
University of Freiburg, Freiburg, Germany
Harald Barsnes and Marc Vaudel
University of Bergen, Bergen, Norway
Lennart Martens and Bart Mesuere
Ghent University, Ghent, Belgium
Haixu Tang and Sujun Li
Indiana University
Krishanpal Anamika and Priyabrata Panigrahi
Persistent Systems Limited, India
Lloyd Smith and Michael Shortreed
University of Wisconsin-Madison
Tom Doak and Jeremy Fischer
Indiana University
Galaxy Community

Jetstream
Galaxy

Funding

NIH award U24CA199347
NIH NATIONAL CANCER INSTITUTE Informatics Technology for Cancer Research

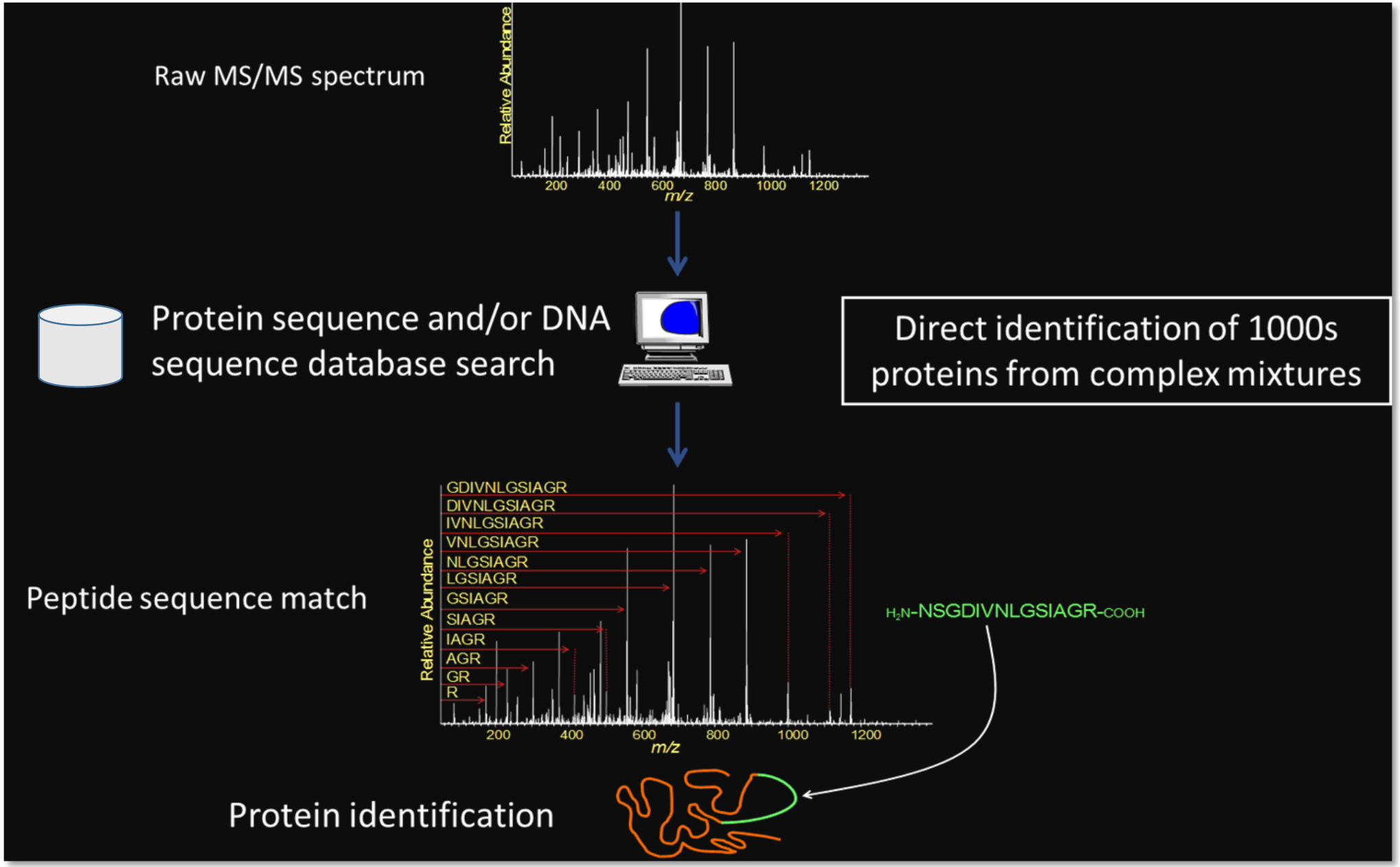
NSF award 1458524
NSF

Norwegian American Centennial Chair

twitter.com/usegalaxyp

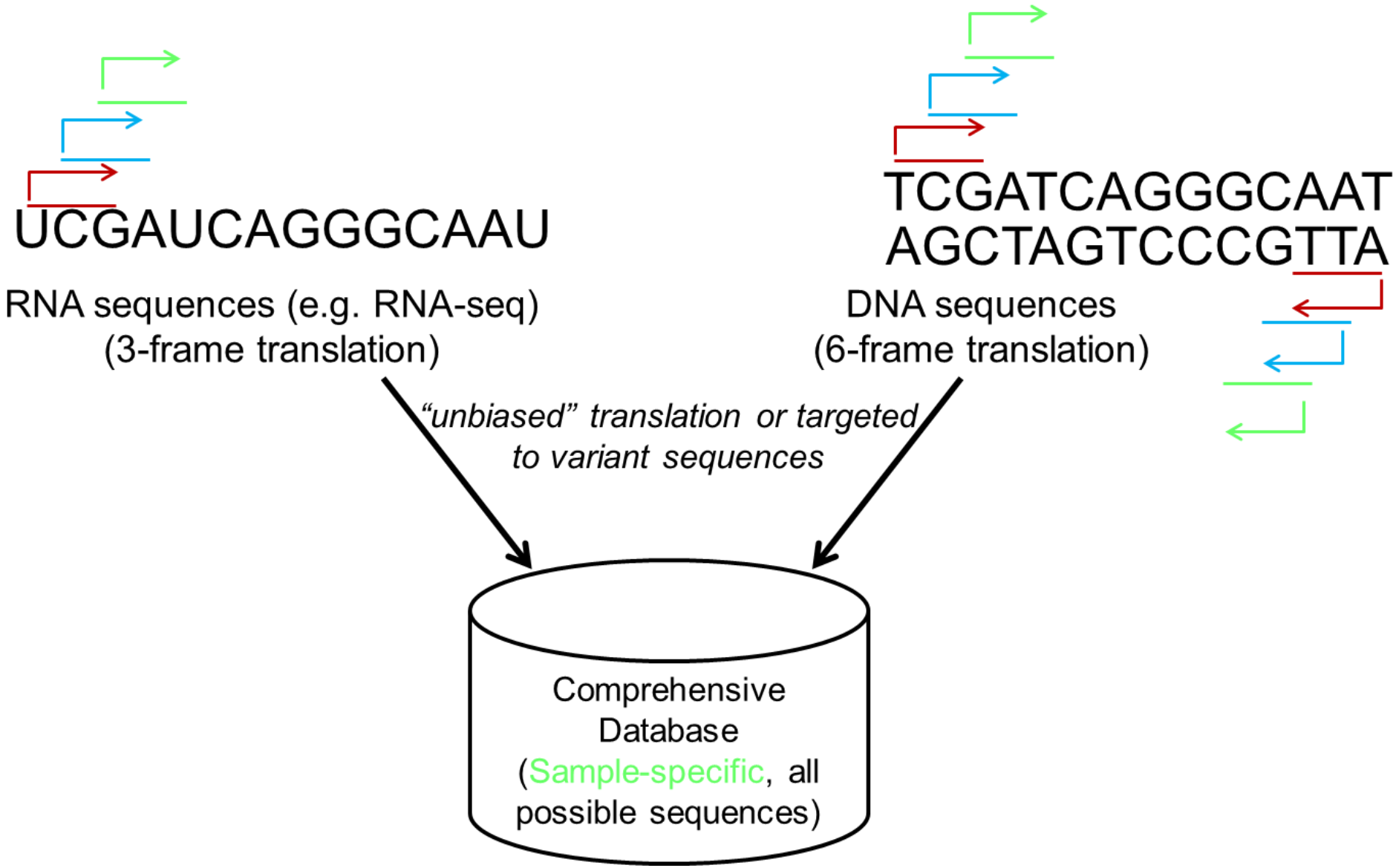
GalaxyP
galaxyp.org

Mass spectrometry-based proteomics as a center-piece of multi-omics

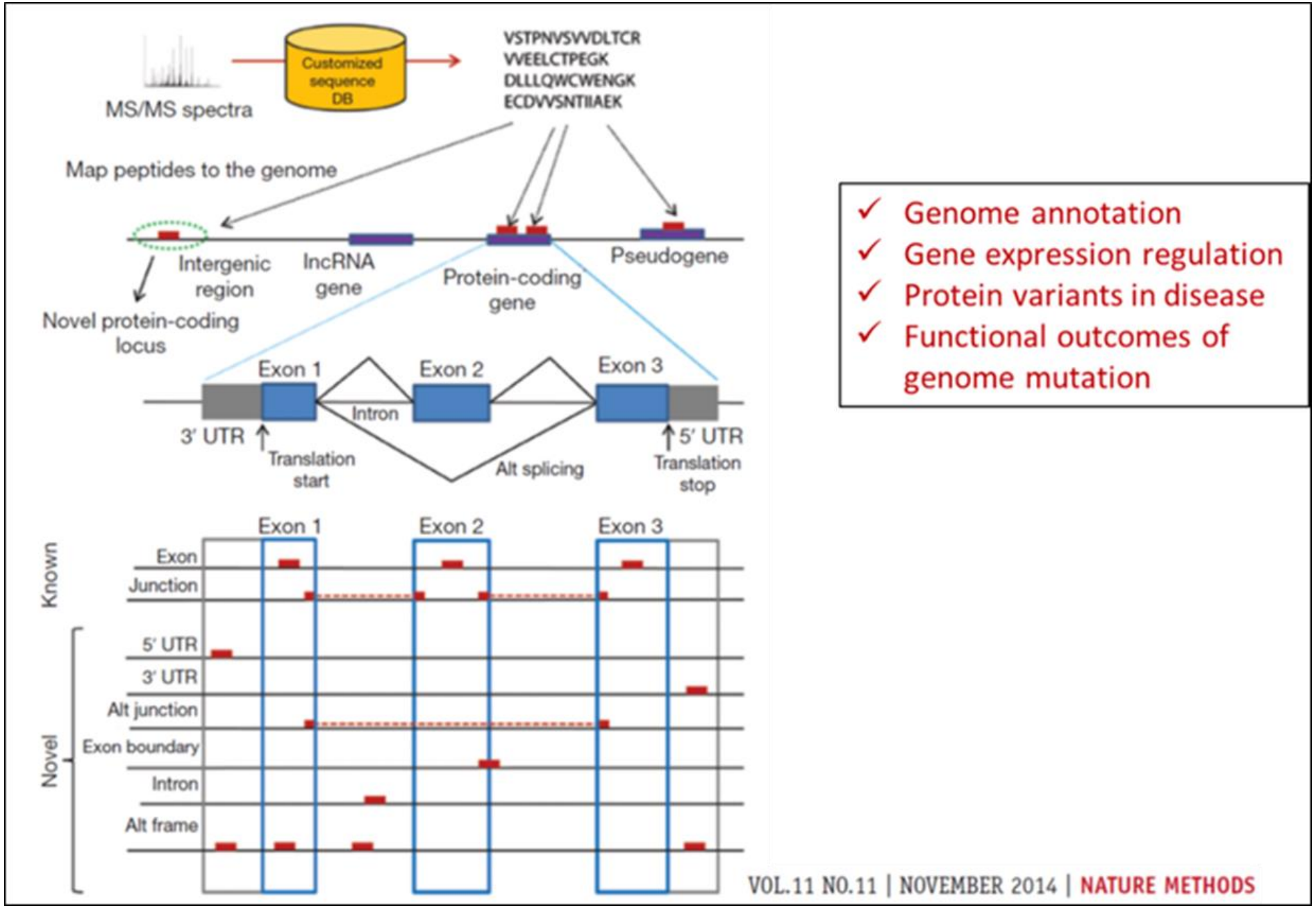


GalaxyP
galaxyp.org

Proteogenomics: enabling more comprehensive identification of proteomes



Potential results generated by proteogenomics

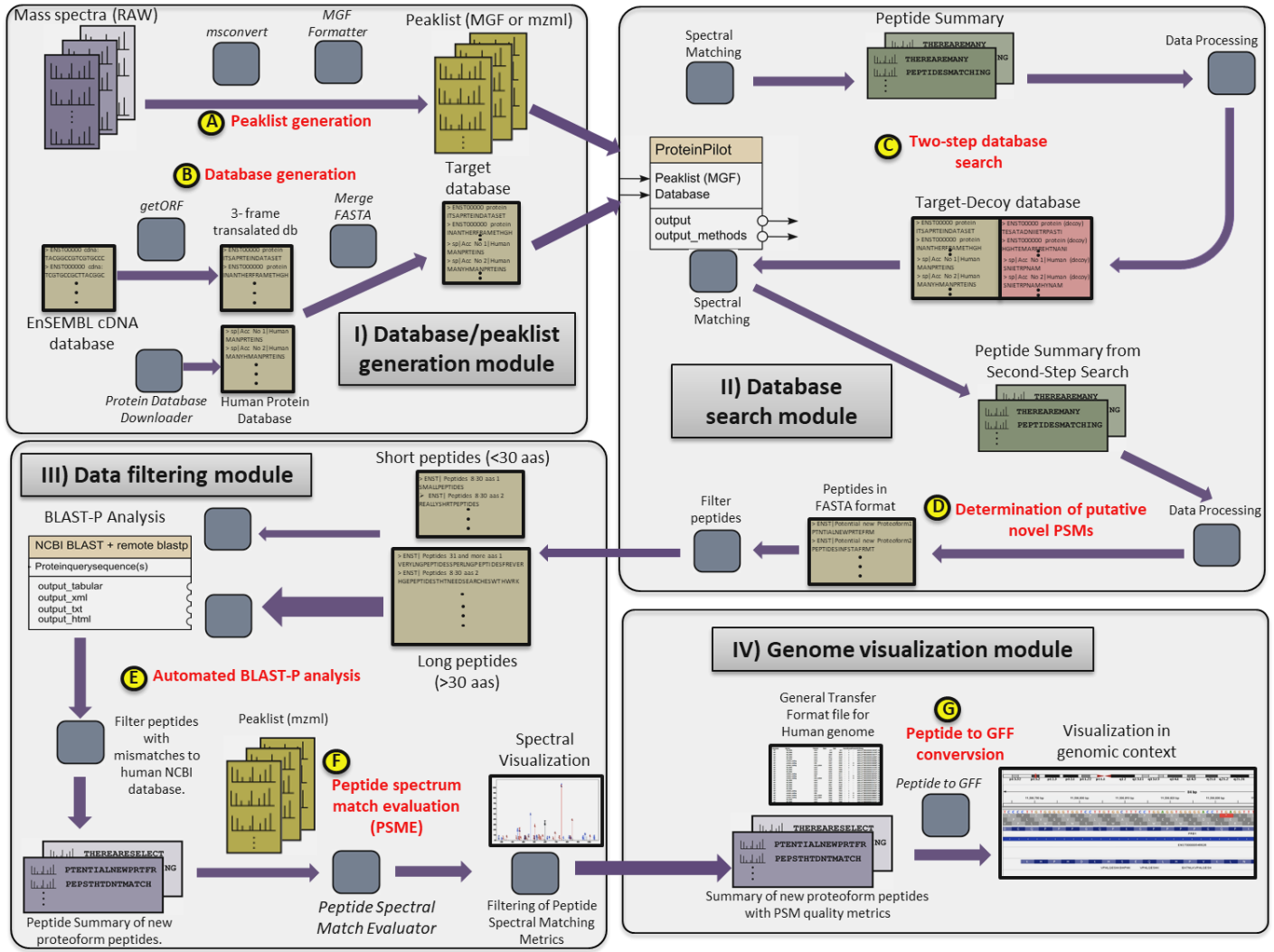


- ✓ Genome annotation
- ✓ Gene expression regulation
- ✓ Protein variants in disease
- ✓ Functional outcomes of genome mutation

GalaxyP
galaxyp.org

Bioinformatic requirements of proteogenomics

- Software – sophisticated, multi-step workflows

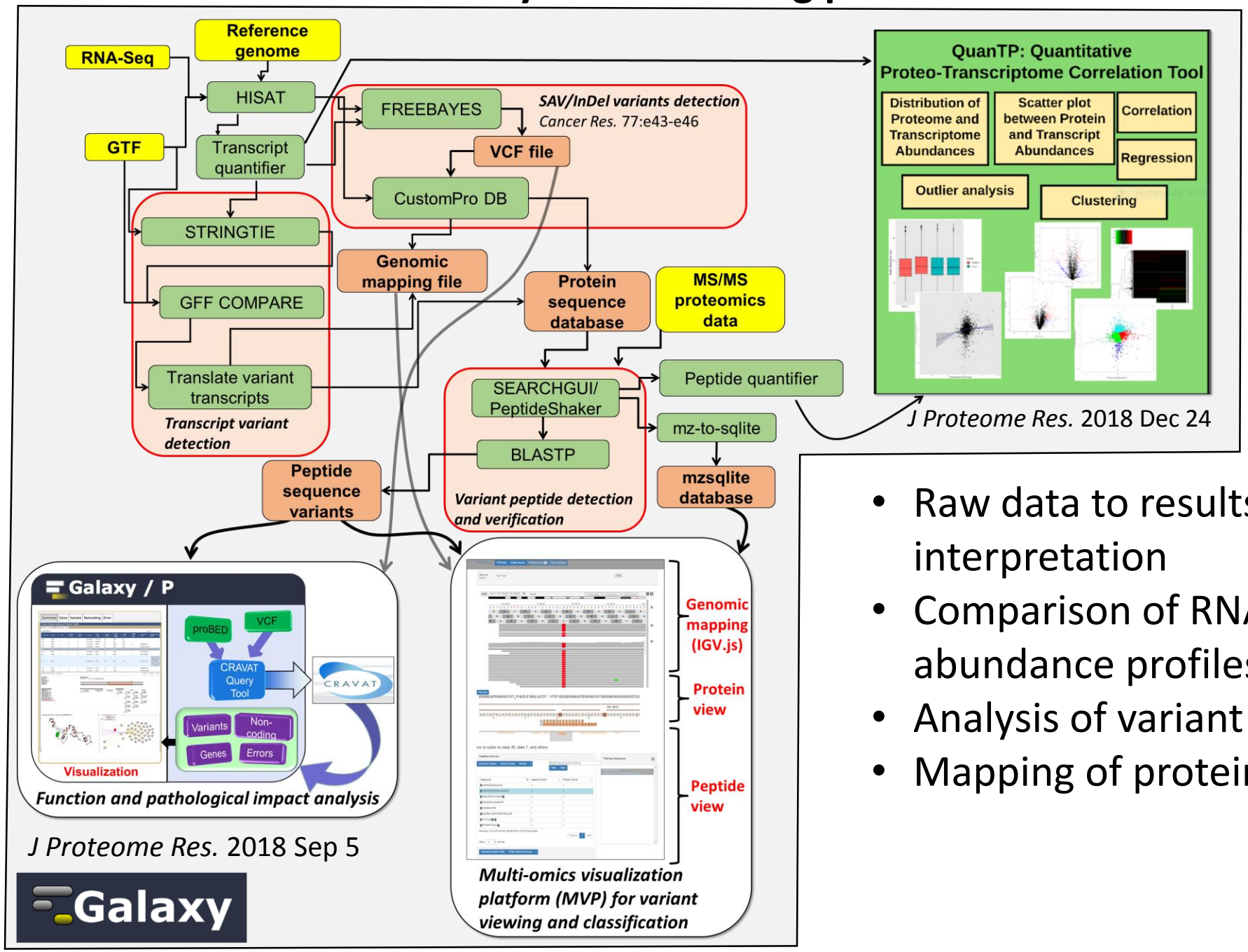


- Customized dB generation
- Matching sequences to MS/MS data
- Filtering and QC!
- Interpretation! Beyond a list....

J. Proteome Res., 2014, 13, pp 5898–5908

GalaxyP
galaxyp.org

Galaxy as an enabling platform

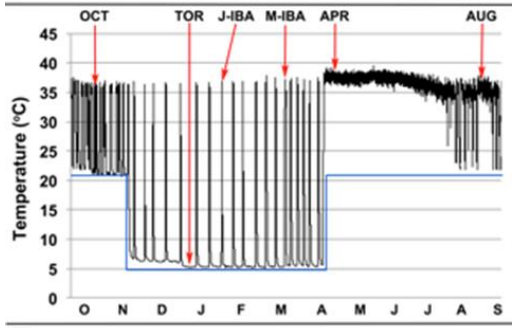


- Raw data to results viewing and interpretation
- Comparison of RNA and protein abundance profiles
- Analysis of variant impact
- Mapping of proteins to genomes

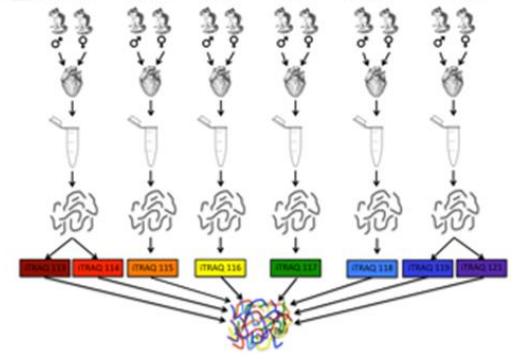
GalaxyP
galaxyp.org

Example Applications -- Proteogenomics of Ground Squirrel Hibernation

A Ground Squirrel Circannual Cycle



B



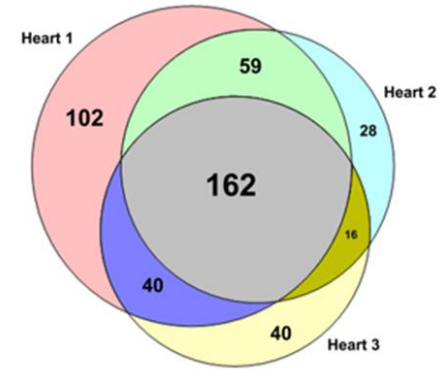
C Identification Summary Table

	Heart 1	Heart 2	Heart 3	Total
Total Spectra	119,123	80,972	89,824	289,919
Spectra (1% global FDR)	54,291	34,621	36,786	127,051
Proteins (1% global FDR)	1,747	1,443	1,379	2,007
Novel Proteoforms	363	265	258	445

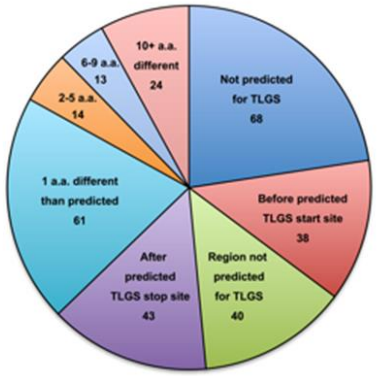
D Differentially Expressed Proteins

OCT	TOR	J-IBA	M-IBA	APR	
98	119	153	123	108	Total
12	14	23	14	20	Unique
30	52	57	60	56	Upregulated
36	47	58	38	15	Downregulated

A Novel Peptide Sequences by iTRAQ Run

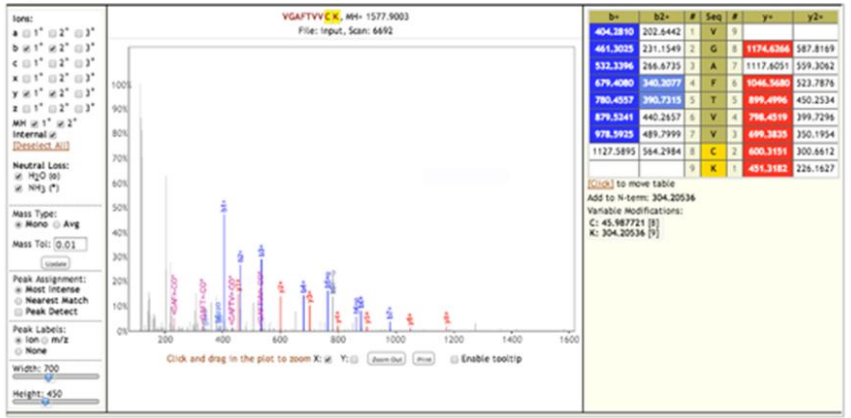


B Novel Peptide Sequences by Category

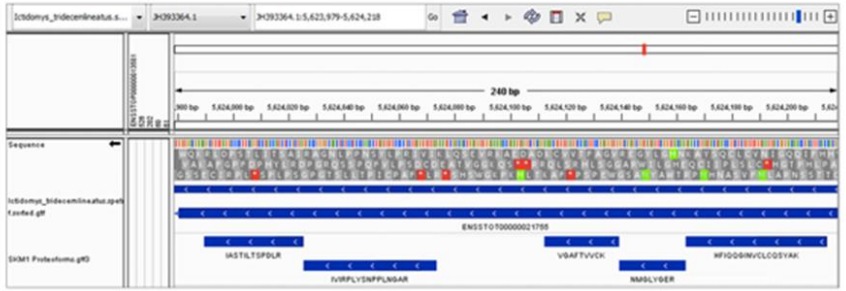


Mapping of proteins to genome for annotation

C Spectral Evaluation



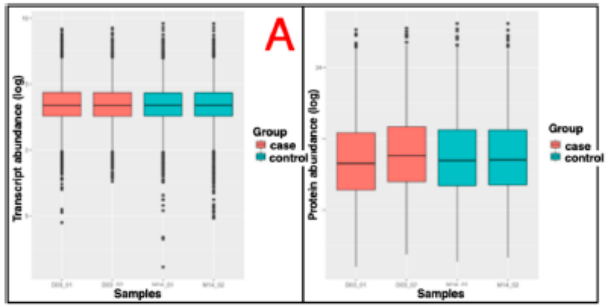
D Visualization on the Genome

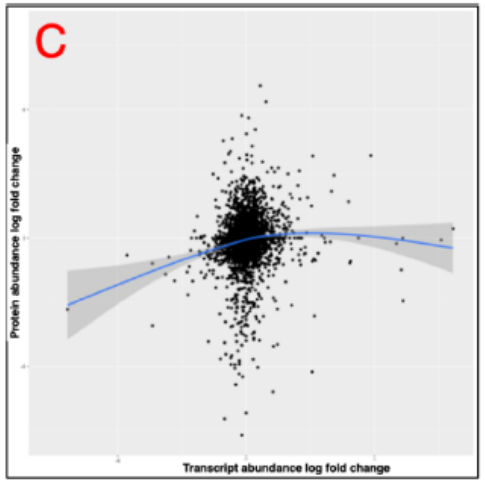


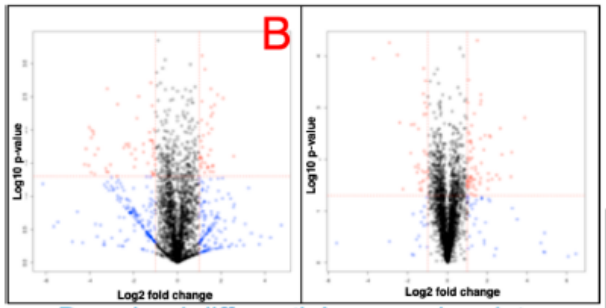
J Proteome Res. 2015 14:4792-804. (Matt Andrews, Katie Vermillion, UMN-Duluth)

GalaxyP
galaxyp.org

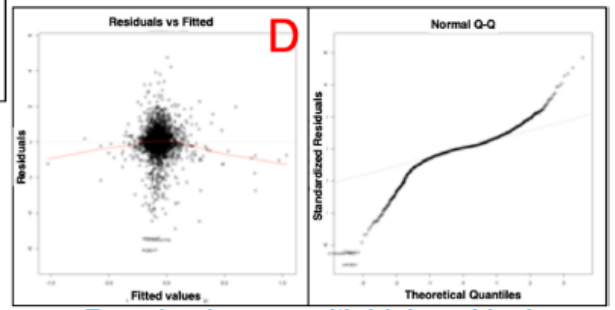
Example application: comparative proteo-transcriptomics in bovine



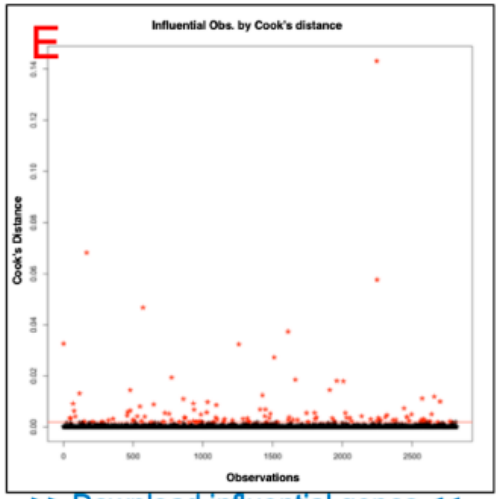




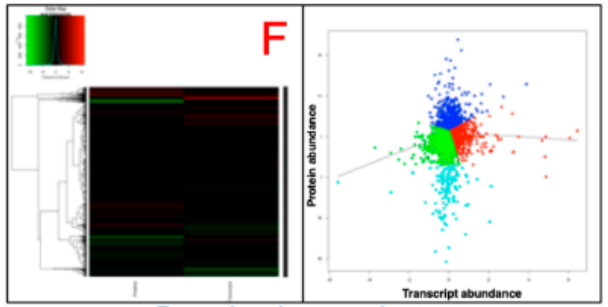
[>> Download differential expression data <<]



[>> Download genes with high residuals <<]



[>> Download influential genes <<]



[>> Download gene clusters <<]

- *Bos taurus* liver tissues compared precalving and postcalving, analyzed by quantitative RNA-Seq transcriptomics and label-free MS-based proteomics
- Correlation between RNA and protein abundance response via QuanTP in Galaxy-P

J Proteome Res. 2018 Dec 24
(Brian Crooker, Wanda Weber, UMN)

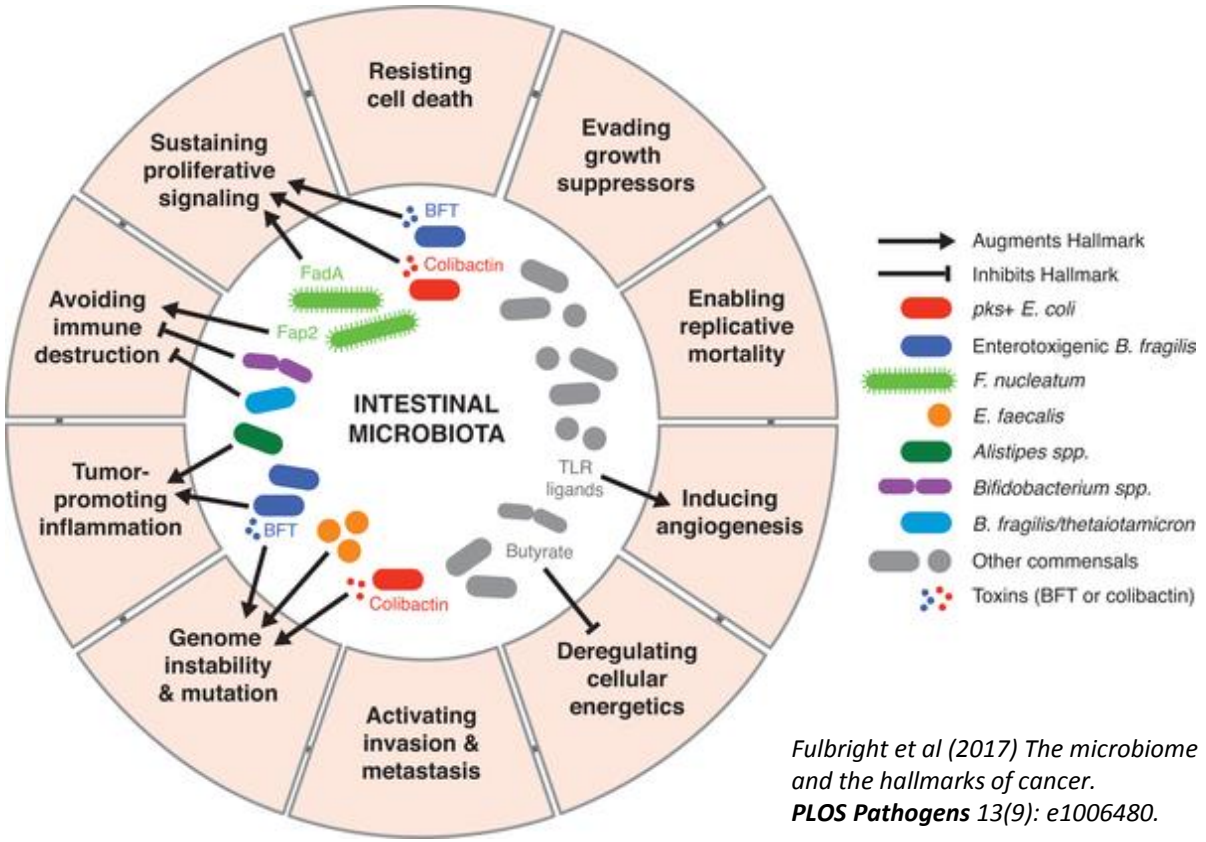
NIH NATIONAL CANCER INSTITUTE
Informatics Technology for Cancer Research

GalaxyP
galaxyp.org

METAPROTEOMICS

Microbiome: Microbial genetic potential and response

Multiple studies have shown **correlation of microbial composition with physiological conditions.**



Fulbright et al (2017) The microbiome and the hallmarks of cancer. ***PLOS Pathogens*** *13(9): e1006480.*

Microbial-derived signals modulate numerous hallmarks of cancer through diverse mechanisms.

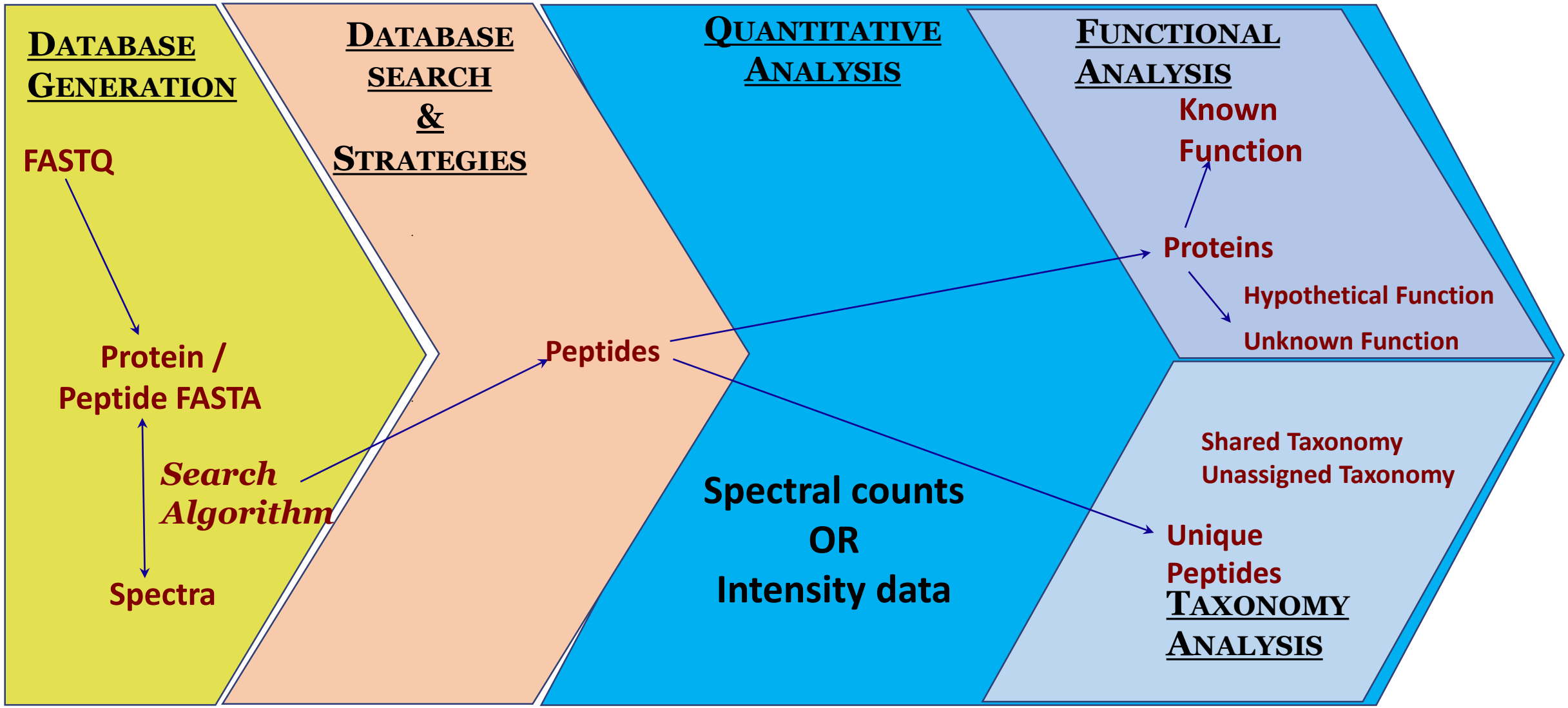
Metagenomics: DNA Sequencing identifies species present within complex community (16S rRNA and Whole Genome Sequencing).

Metatranscriptomics: RNA Sequencing identifies species present and possible functions within complex communi (RNASeq).

Metaproteomics:
The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time. Potential to unravel the **mechanistic details of microbial interactions with host / environment** by analyzing the **functional dynamics of the microbiome.**

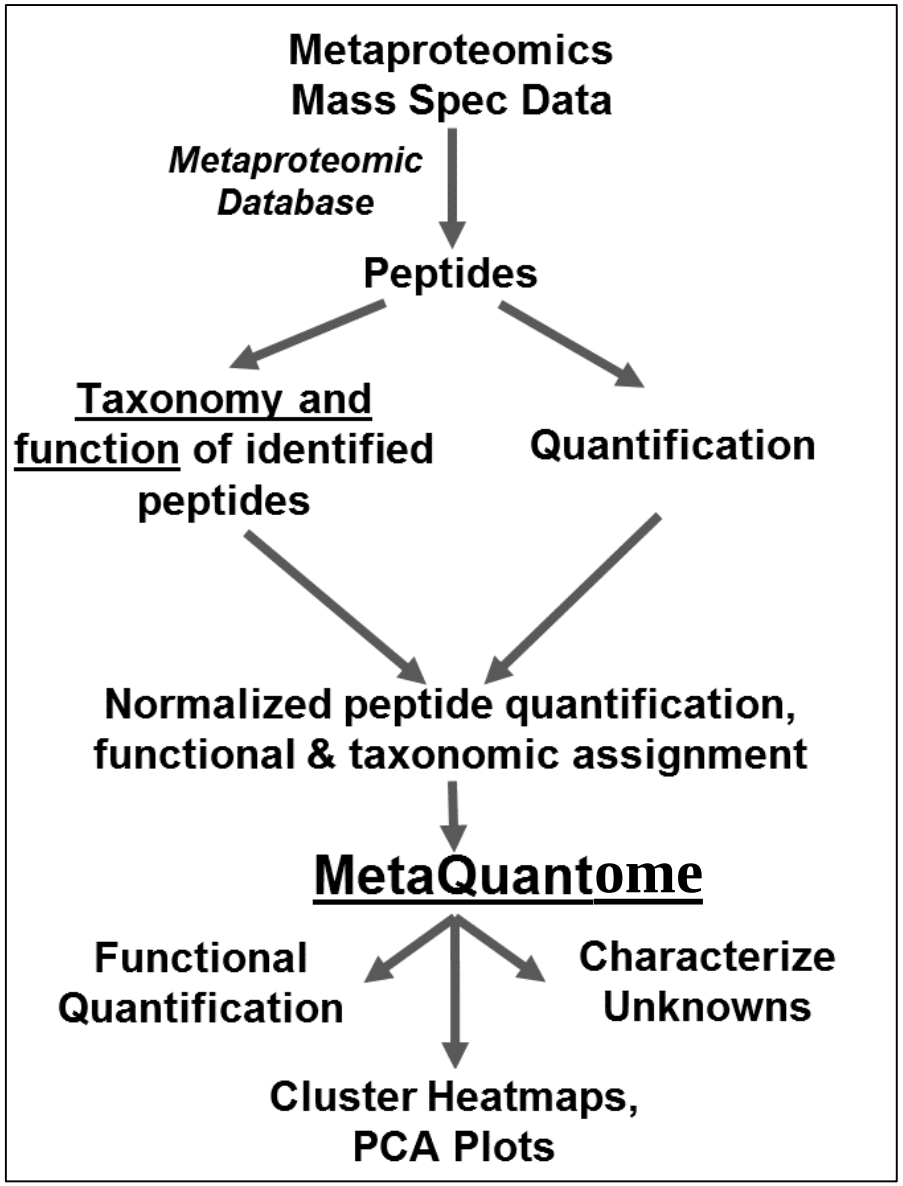
GalaxyP
galaxyp.org

Metaproteomics Workflow



MetaQuantome

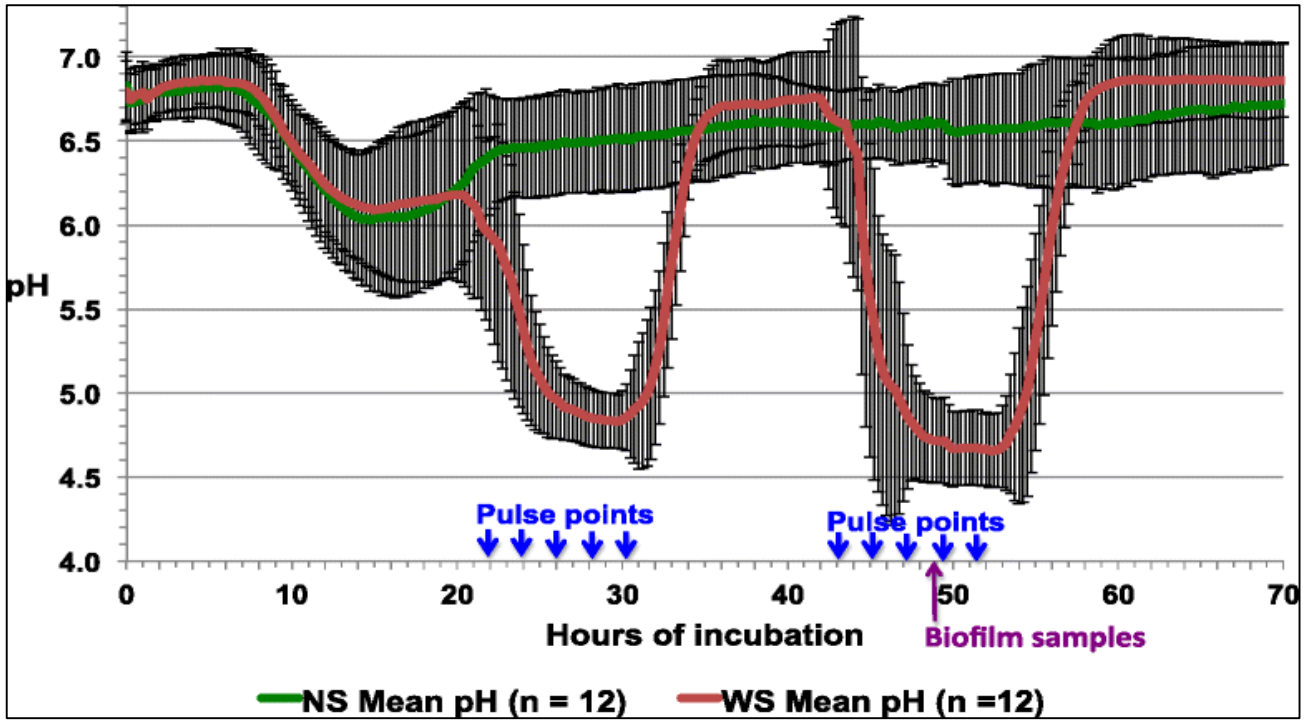
- **metaQuantome allows for robust quantitative functional & taxonomic analysis from metaproteomics datasets.**
  - **Quantitative**: *Supports analysis of multiple samples, including comparison across multiple experimental conditions*
  - **Support for function-taxonomy interaction analysis**: *Leverages taxonomic and functional information of the same dataset*
  - **Flexible** & **Accessible**: *Free and open source – available on Github, Python Package Index, Bioconda, and Galaxy*



GalaxyP
galaxyp.org

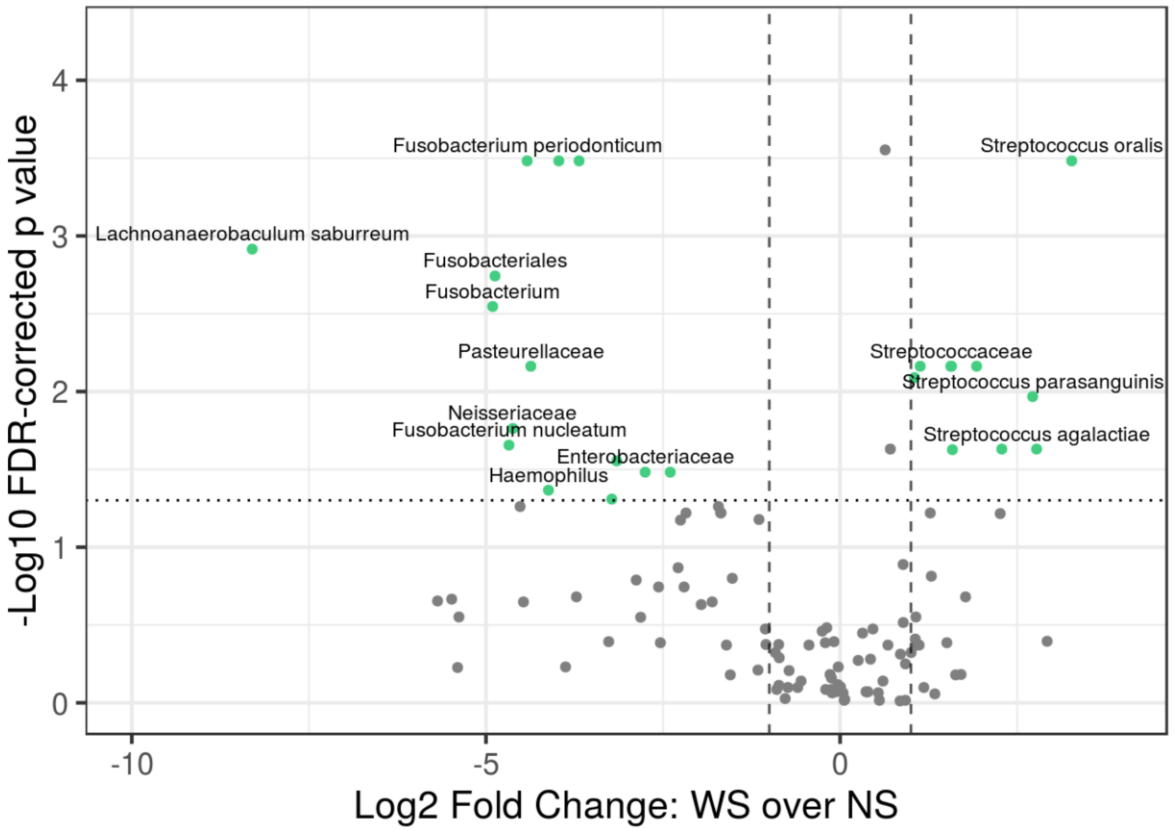
Case Study : Sucrose-induced oral dysbiosis

- Mass spectral data was acquired from plaque samples from **twelve subjects at high risk for dental caries grown in biofilm reactor** in the presence (With Sucrose, or **WS**) and absence of sucrose (No Sucrose, or **NS**) (12 in each group, 24 total samples)
- Mass spectra were searched against the Human Oral Microbiome database (HOMD) to identify microbial peptides.
- Quantitation, functional annotation, and taxonomic assignment was performed in Galaxy; **metaQuantome** was used to analyze the results.

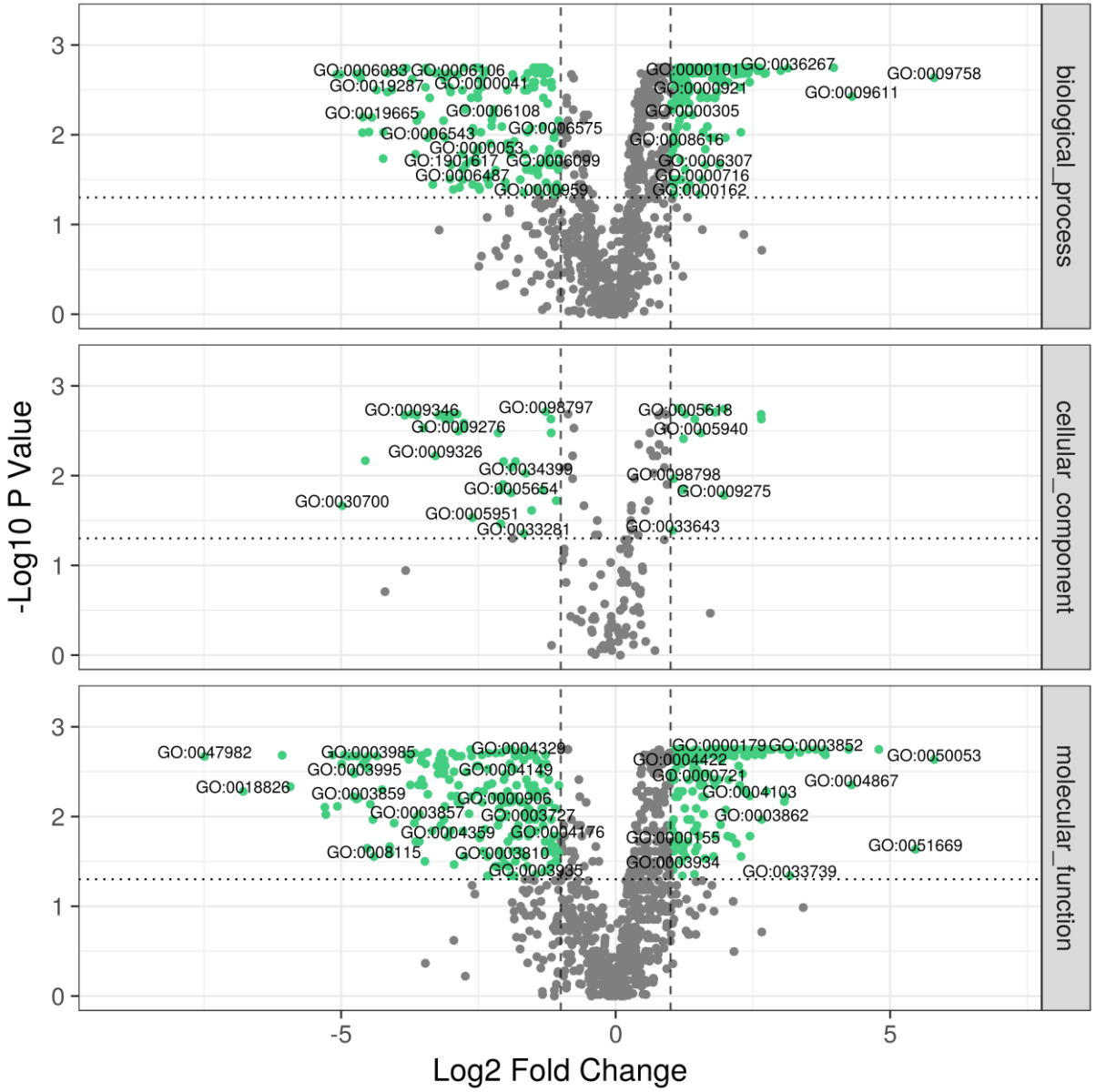


Rudney et al., BMC Microbiome [DOI: 10.1186/s40168-015-0136-z](https://doi.org/10.1186/s40168-015-0136-z)

Oral dysbiosis results: volcano plots

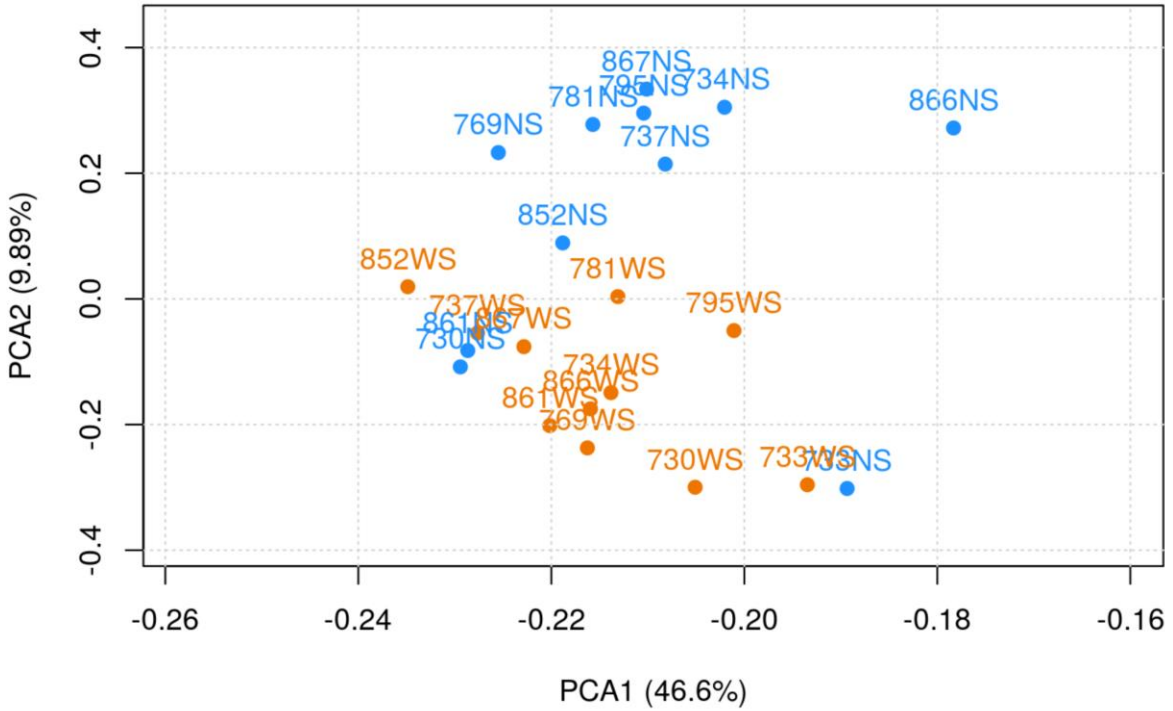


TAXONOMY

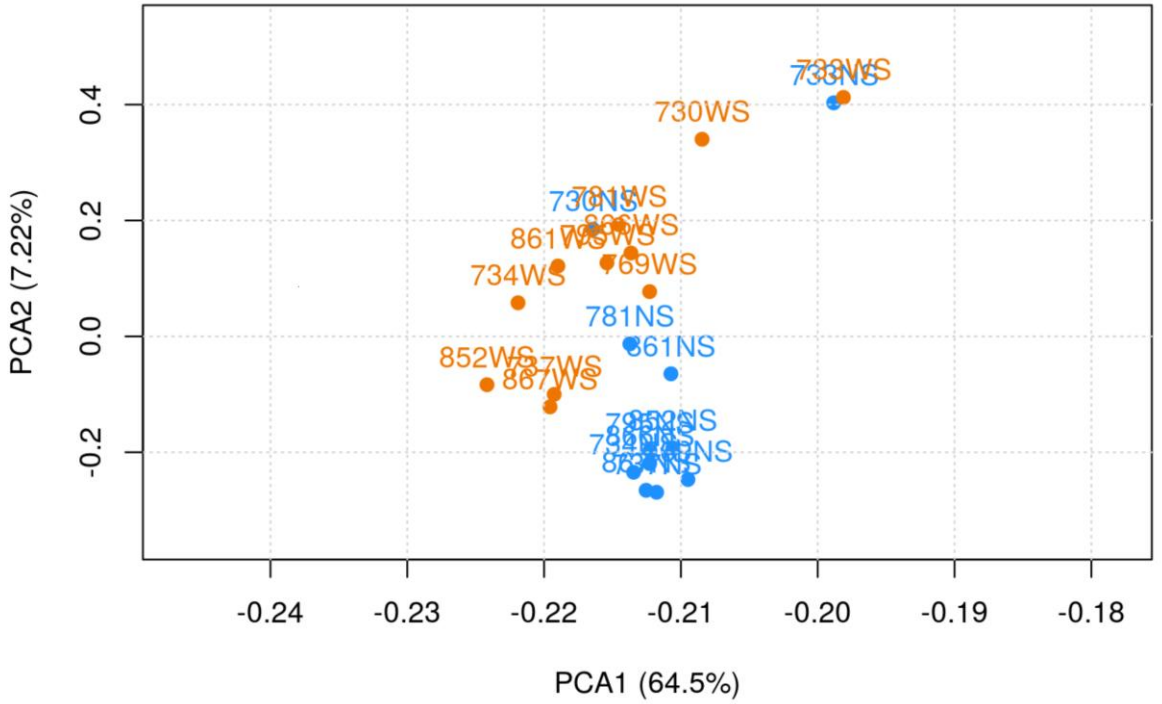


FUNCTION

Oral dysbiosis results: PCA plots

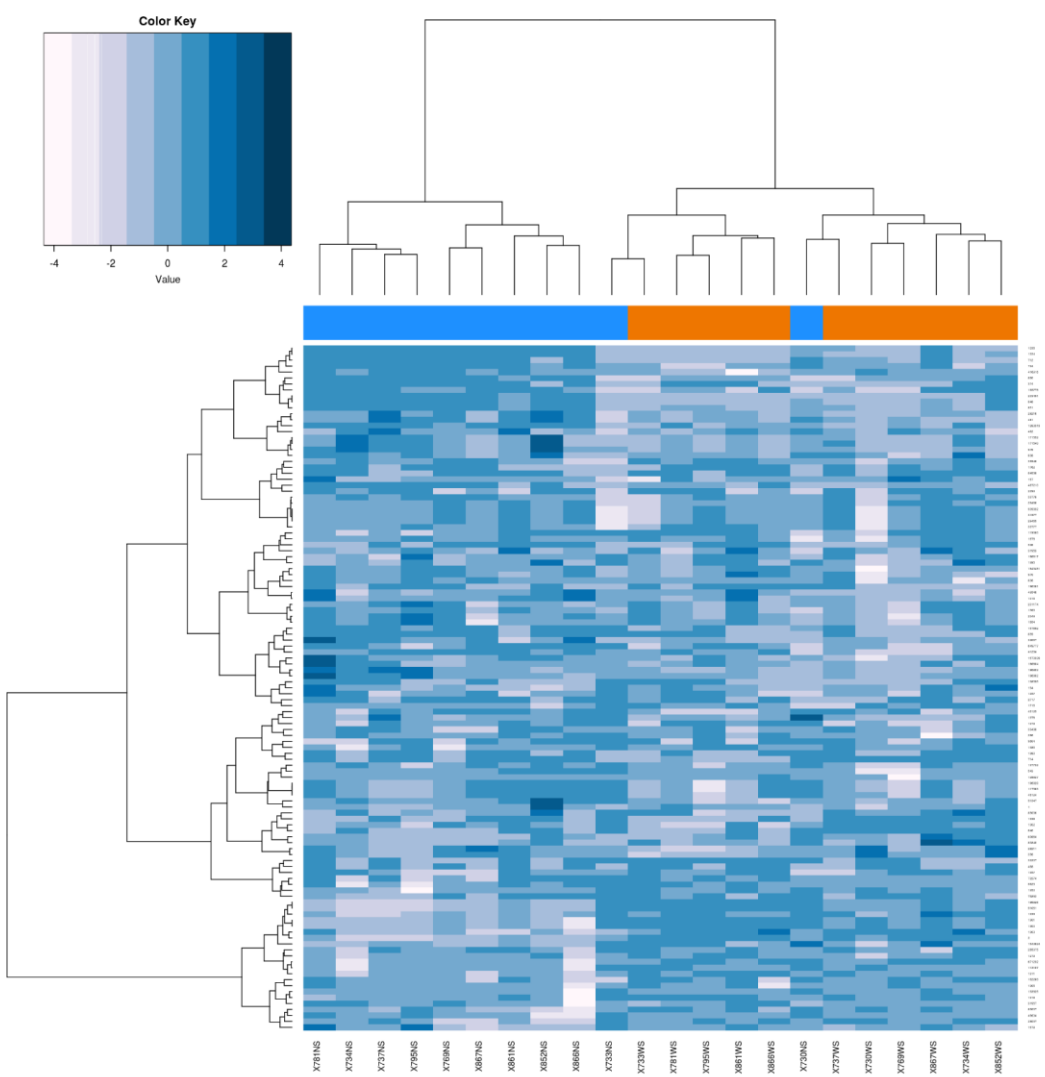


TAXONOMY

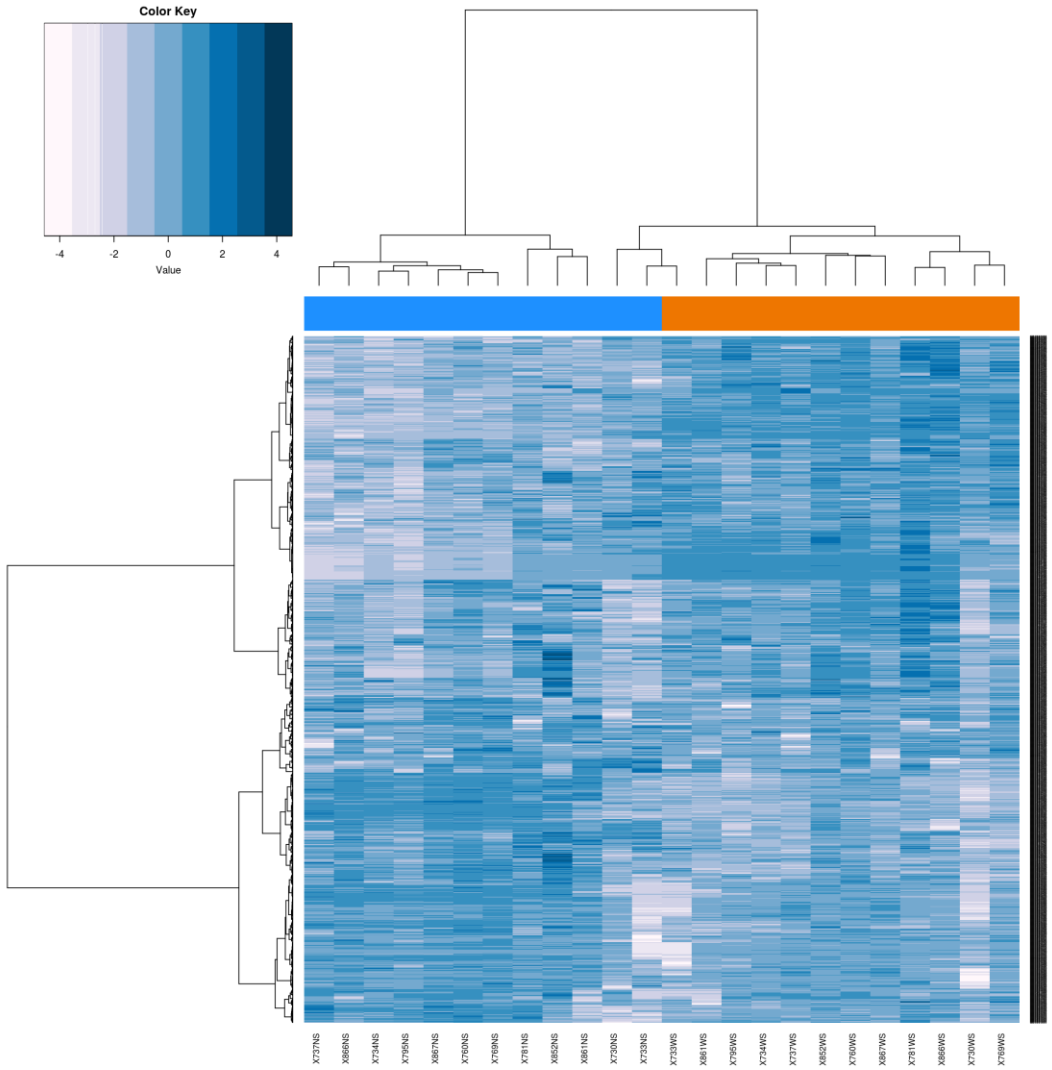


FUNCTION

Oral dysbiosis results: Heatmaps

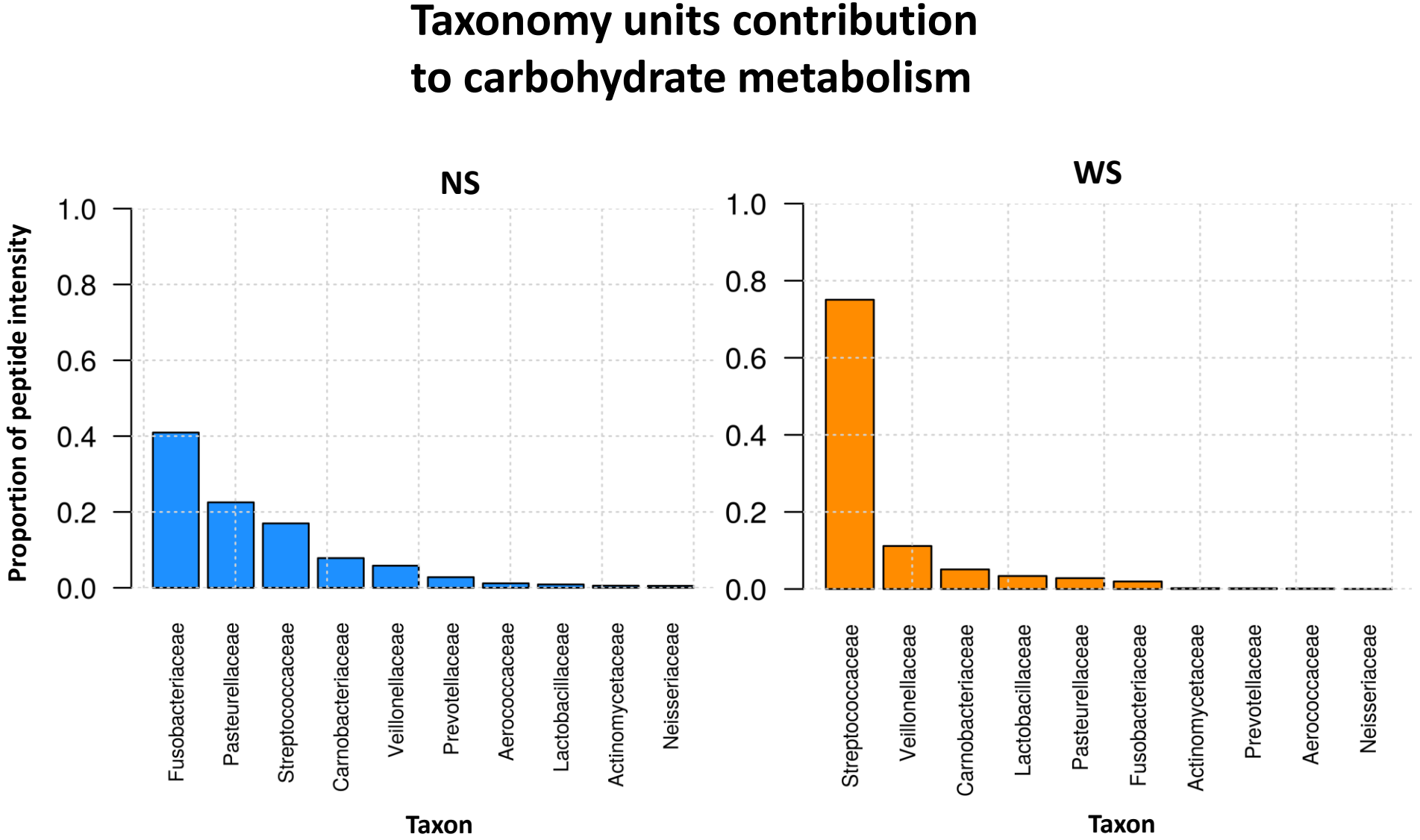


TAXONOMY



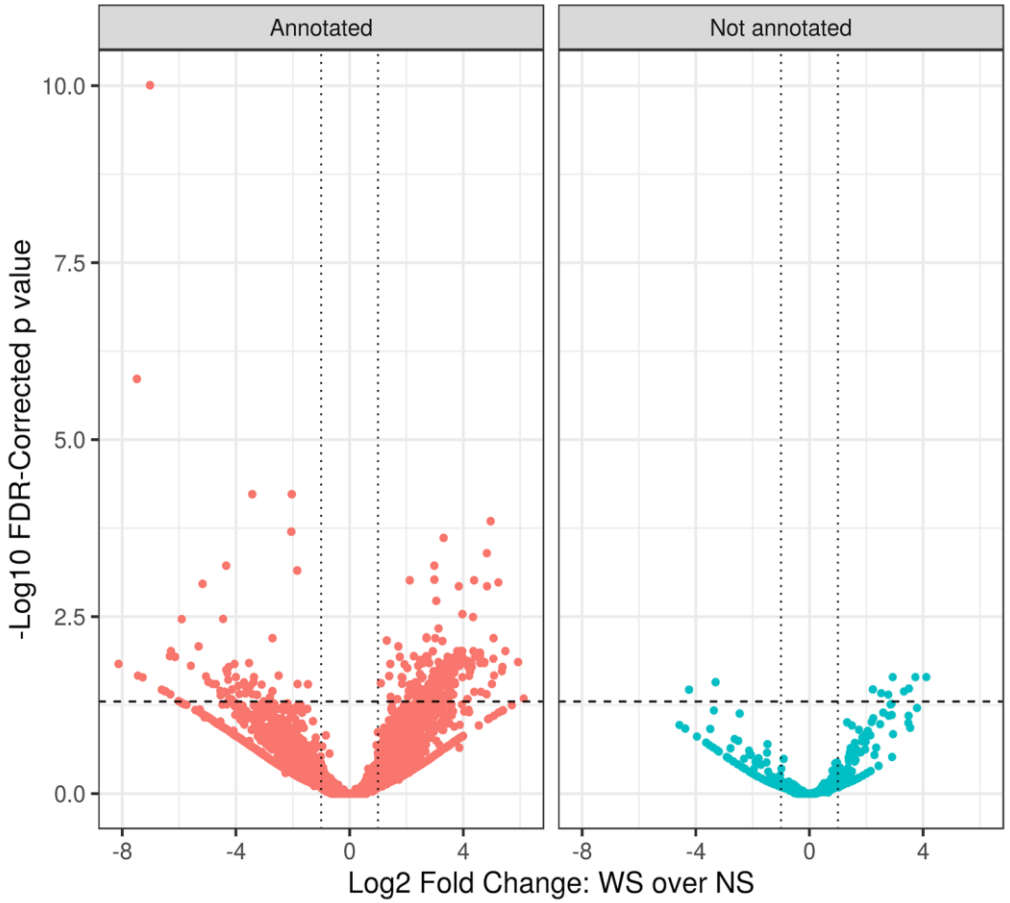
FUNCTION

Oral dysbiosis results: Function-Taxonomy



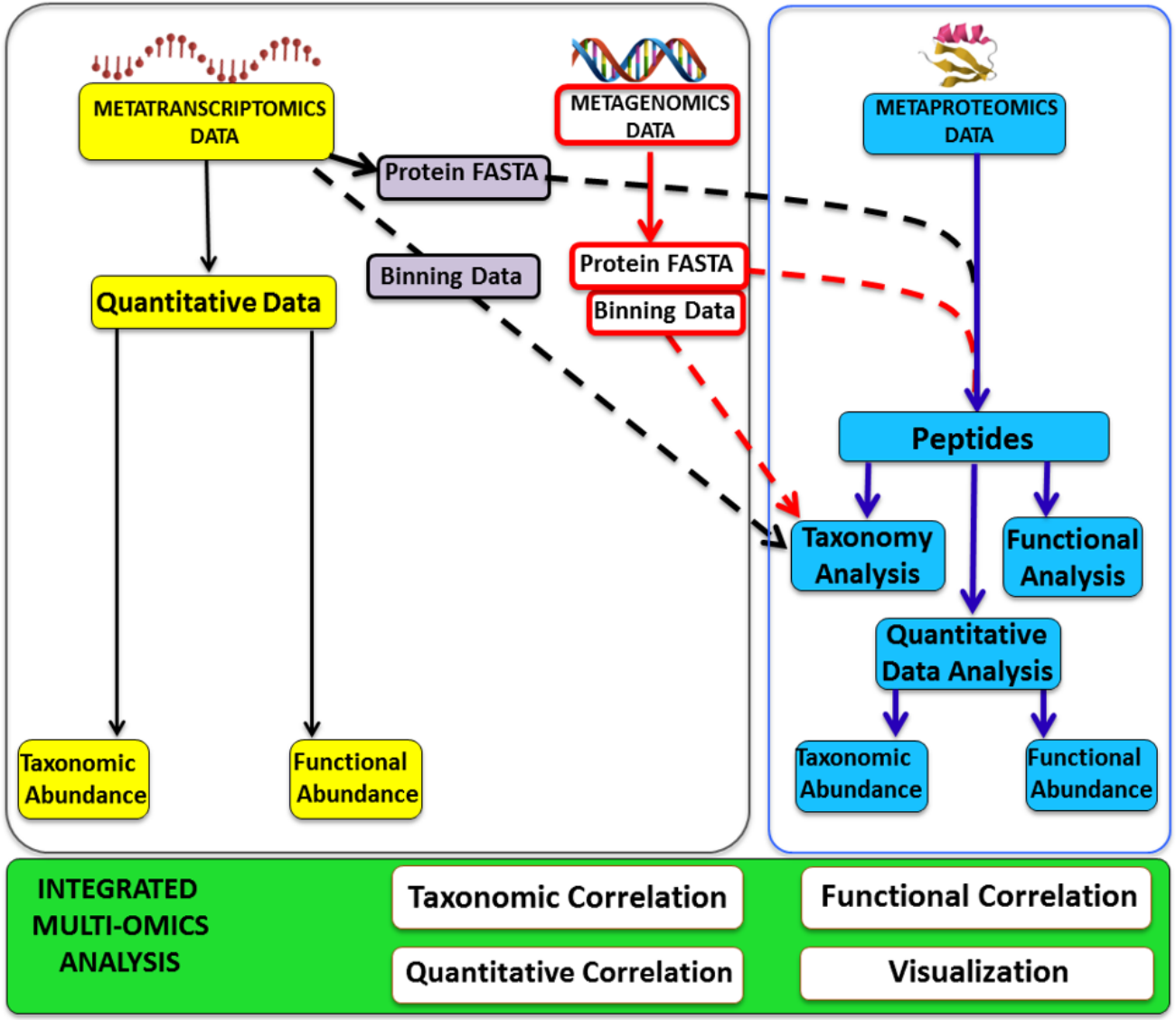
Future Directions

- Analyze more datasets (clinical and environmental)
- Alternative tools for quantitation, taxonomy & function.
- Investigate peptides/proteins of unknown function/taxonomy
- Integrate the metaproteomics workflow with an existing metatranscriptomics quantitative analysis & visualization workflow (ASaiM) within Galaxy.



Differential expression analysis: proteins of known (L) and unknown (R) function

MULTI-OMICS APPROACH FOR FUNCTIONAL MICROBIOME ANALYSIS



- Multi-omic approaches (metatranscriptomics & metaproteomics) characterize the functional molecules that may contribute to microbial responses.
- We are implementing a metatranscriptomics and metaproteomics quantitative analysis pipeline within Galaxy-P.

GalaxyP
galaxyp.org

ACCESSING THE MULTI-OMIC WORKFLOWS

Proteogenomics Gateway: z.umn.edu/proteogenomicsgateway

Step-by-step instructions for Galaxy instance usage: z.umn.edu/pginnov18

Metaproteomics Gateway: z.umn.edu/metaproteomicsgateway

Step-by-step instructions for Galaxy instance usage: z.umn.edu/suppS1

Tools also available on : https://proteomics.usegalaxy.eu/

Publications: z.umn.edu/galaxypreferences

Contact Us: http://galaxyp.org/contact

twitter.com/usegalaxyp

ACKNOWLEDGMENTS

Funding

University of Minnesota

Timothy Griffin PI

Praveen Kumar
Candace Guerrero
Subina Mehta
Adrian Hegeman (Co-I)
Art Eschenlauer
Ray Sajulga
Caleb Easterly
Andrew Rajczewski

Minnesota Supercomputing Institute

James Johnson
Thomas McGowan
Michael Milligan

Brook Nunn
U of Washington

Biologists / collaborators

Laurie Parker
Joel Rudney
Maneesh Bhargava
Amy Skubitz
Chris Wendt
Brian Crooker
Steven Friedenberg
Kevin Viken
Kristin Boylan
Marnie Peterson
Somiah Afiuni
Brian Sandri
Alexa Pragman
Wanda Weber
Amy Treeful

Josh Elias
Stanford University

Lloyd Smith (Co-I)
Michael Shortreed
UW-Madison

Thomas Doak
Jeremy Fisher
Haixu Tang
Sujun Li
Indiana University

Matt Chambers
Nashville, TN

Stephan Kang
Intero Life Sciences

Judson Hervey
Naval Research Institute
Washington, D.C.

Harald Barsnes
Marc Vaudel
University of Bergen, Norway

Magnus Øverlie Arntzen
Francesco Delogu
NMBU, Oslo, Norway

Carolin Kolmeder
University of Helsinki, Finland

Lennart Martens (Co-I)
Bart Mesuere
Robbert G Singh
VIB, UGhent, Belgium

Thilo Muth
Bernhard Renard
Robert Koch Institut

Alessandro Tanca
Porto Conte Ricerche, Italy

Anamika Krishanpal
Priyabrata Panigrahi
Persistent Systems Limited

Bjoern Gruening
Bérénice Batut
University of Freiburg, Freiburg, Germany

Ira Cooke and Maria Doyle
Townsville , Australia

Maria Doyle
Melbourne , Australia

galaxyp.org

twitter.com/usegalaxyp

GalaxyP
galaxyp.org