

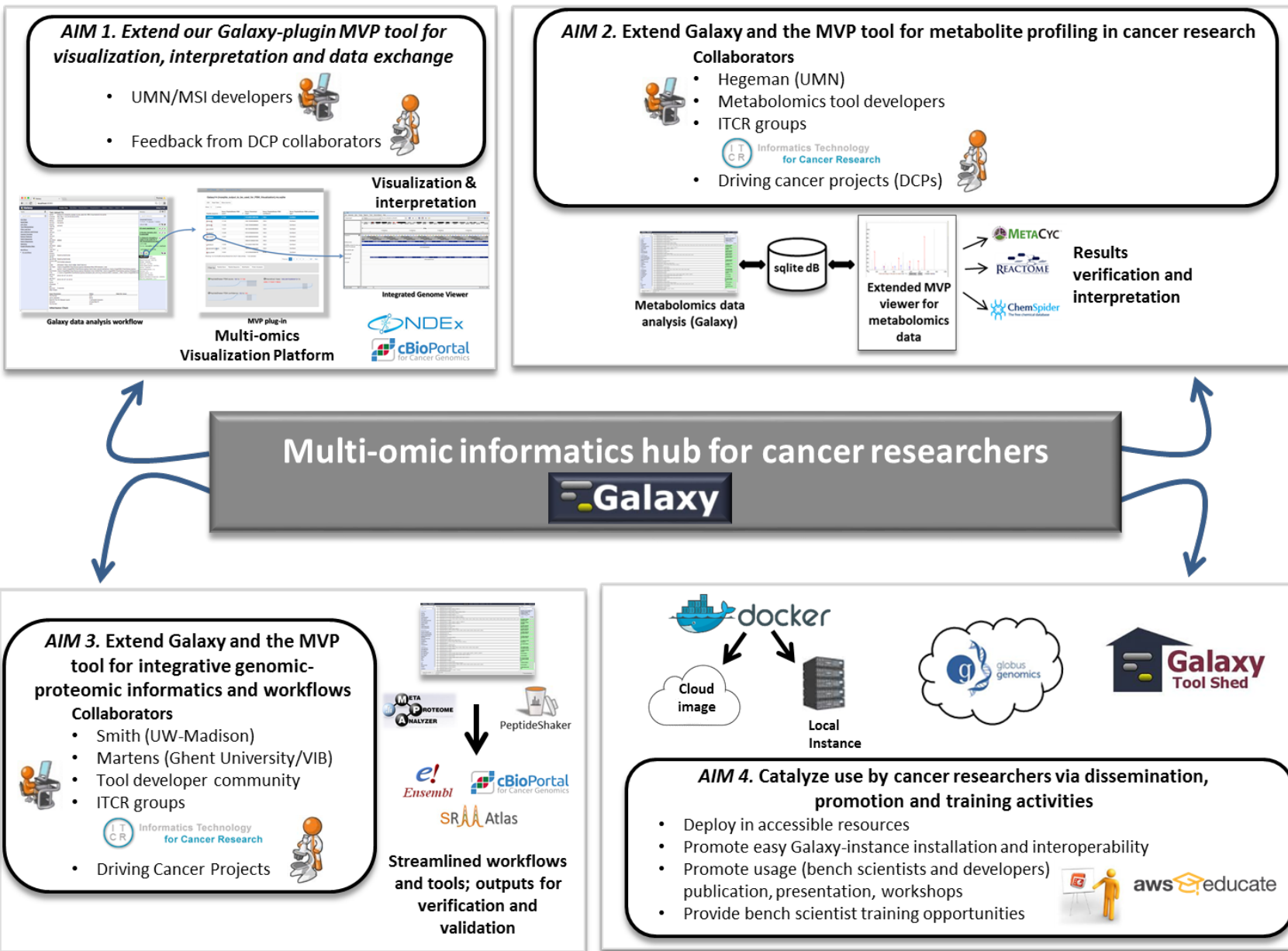
# A Galaxy-Based Multi-Omic Informatics Hub for Cancer Researchers

1U24CA199347

Tim Griffin, PI

Pratik Jagtap, Project Co-Leader

# PROJECT OVERVIEW





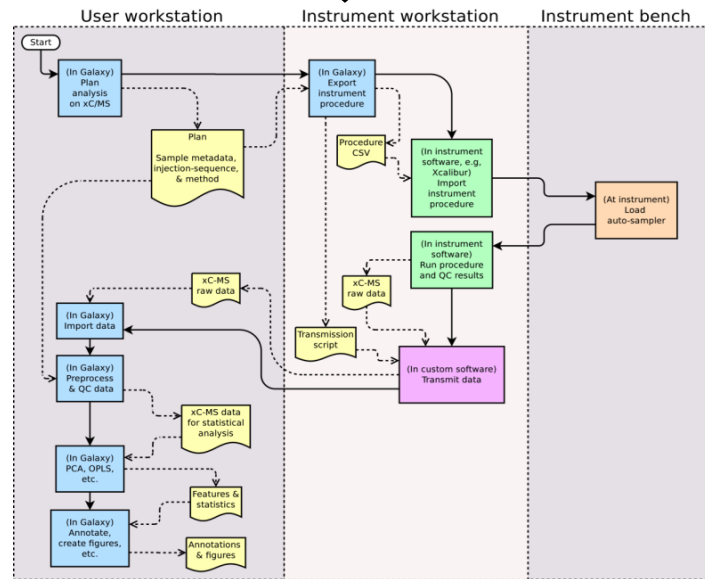
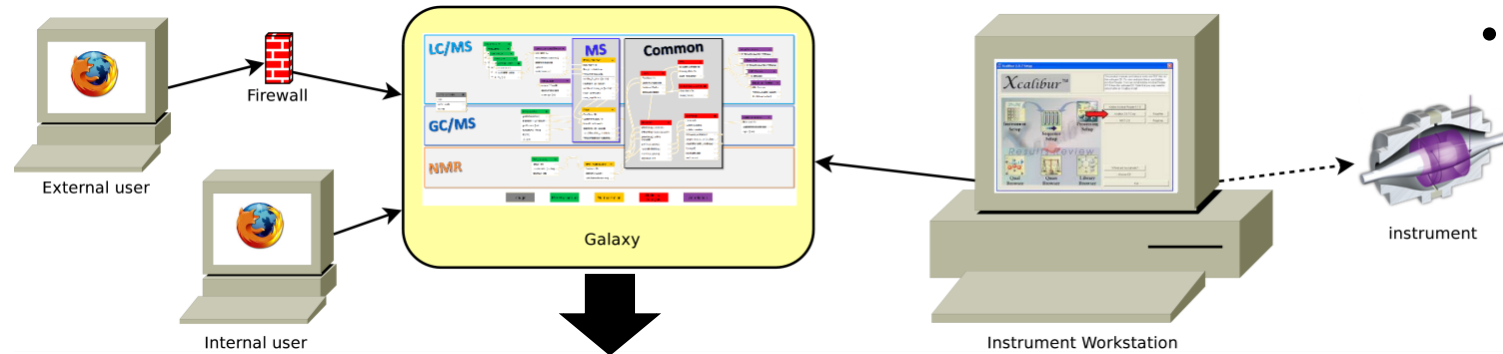
Art Eschenlauer



Adrian Hegeman  
(Co-I)

# METABOLOMICS (AIM 2)

- **GOAL: reproducible, scalable and bench researcher-friendly workflow for MS-based metabolite quant and annotation (ID)**
- User-driven resource, defining requirements for reproducibility, throughput, and statistical analysis across datasets:
  - Galaxy promotes usability, reproducibility and scalability (and portability)
  - Direct upload from instrument and other automated data processing steps promotes throughput
  - Multivariate statistical tools and customized tool for null values promotes comparisons across hundreds of datasets
  - “Dockerized” to promote portability



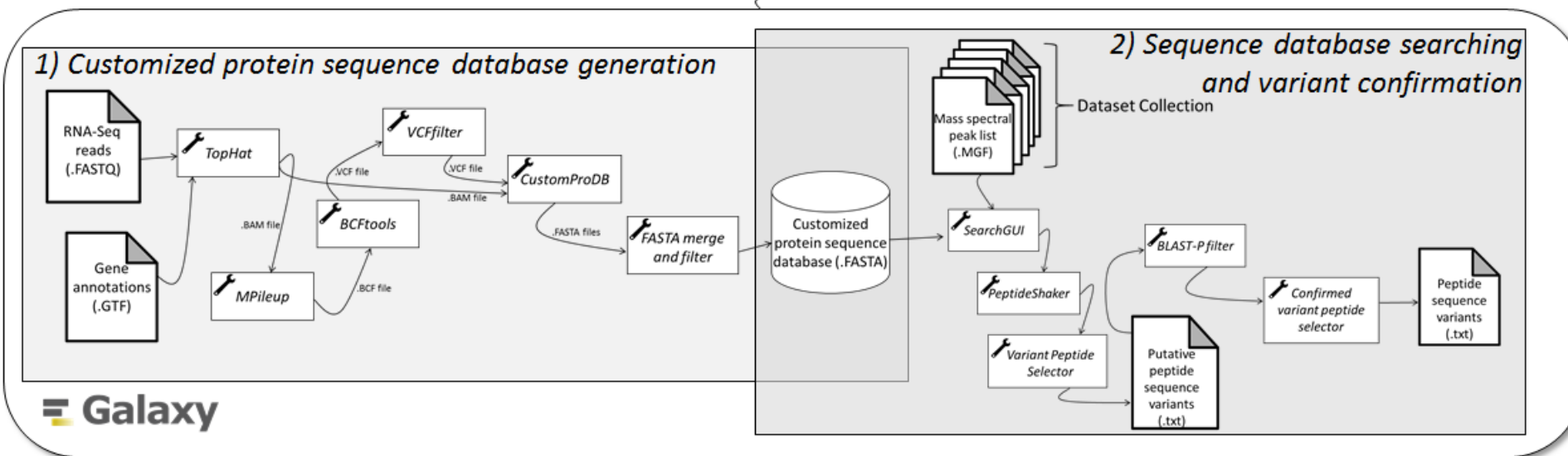
*Galaxy-based workflow for high-throughput metabolite quant and annotation (leveraging Workflow4metabolomics, xCMS-based)*

# PROTEOGENOMICS (AIM 3)

*An accessible proteogenomics informatics resource for cancer researchers  
(Cancer Research, Special Computer Resources issue)*



- *Workflows and example input data*
- *Larger capacity instance*



- *Training materials and workflows*

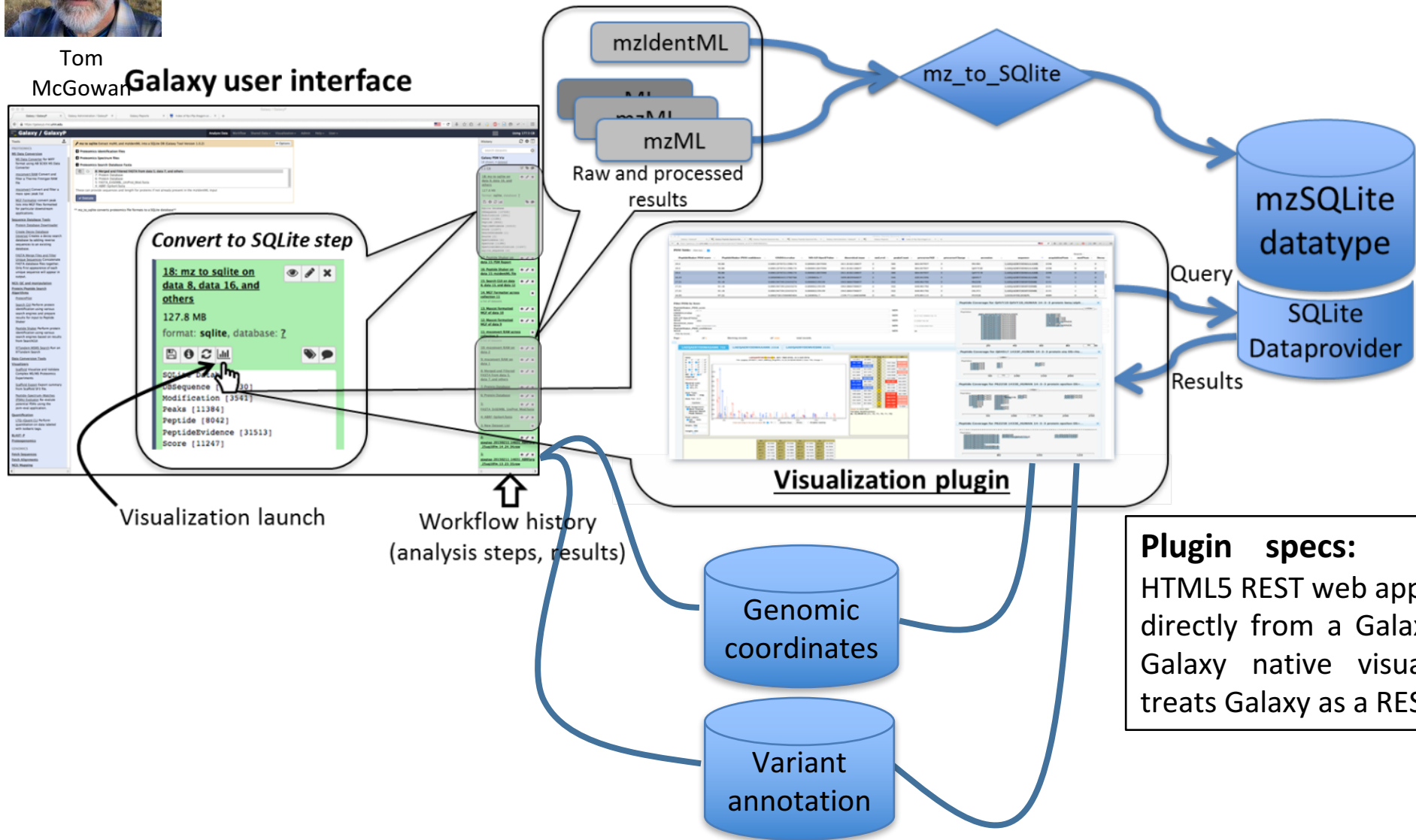
# MULTI-OMICS VISUALIZATION PLATFORM (MVP) (AIM 1)



James Johnson (JJ)



Tom McGowan



**Plugin specs:** Pure JavaScript, HTML5 REST web application; accessed directly from a Galaxy history via the Galaxy native visualization registry; treats Galaxy as a REST endpoint

# MVP: Protein-centric variant visualization

MVP Viewer   Defaults   Reset   Save to Galaxy   Verified Scans

Clear Map  
P22531 : SPR2E\_HUMAN Small proline-rich protein 2E OS=Homo sapiens GN=SPRR2E PE=2 SV=2

Chr: 1

Galaxy206-[mz\_to\_sqlite\_on\_data\_42\_data\_38\_and\_others].mz.sqlite

PSM Overview

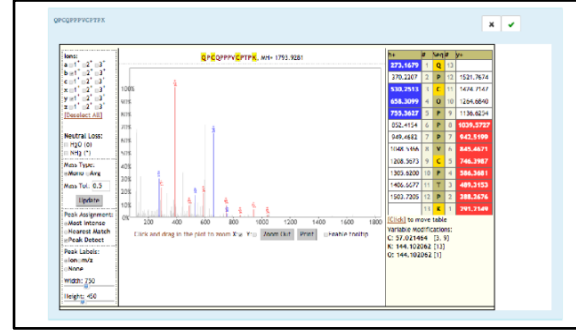
[View in Protein](#)   [Load from Galaxy](#)

ELVIS, DRINKS; MILK  

Peptide Sequence

- > @SQADYETR
- > QECQPPVVCPTPK
- > FTPDHVVYAR
- > SFVALSQDIQK
- > QSESTHGQSAPSTGGR
- > VVGSELIQK
- > VDLVAHMASGE
- > LLLEIASR
- > DRTLDEINFLR
- > IPDGCQVQ

**View annotated MS/MS spectra**



**Detailed viewing in IGV**

Human (hg38)   chr1   chr1:153,093,532-153,093,751   Go

220 bp

153,093,540 bp   153,093,560 bp   153,093,580 bp   153,093,600 bp   153,093,620 bp   153,093,640 bp   153,093,660 bp   153,093,680 bp   153,093,700 bp   153,093,720 bp   153,093,740 bp

Sequence

V T C S W D E L A G R Y K V K S D E B T F A Q T A E V G R A M D L V V R A Q G L S G V M A L G L A W T I B E V A G R A A G I A A M D K T L L A E G W T E W L A G W G R C H R T F E L A L L R W A G L W T E W W A G L R A L R G W T L W L W R G A H R R W L A G L L A L L E I T R H S Y L L E C G H F G W G G E G Y T G C H F E W H C G G E C C G H F G C G S G R F G C G H S G H F G V C R T C C G W G C L H C C C D

Gene

SPRR2E

chr1:153,093,602   99M of 326M

# METAPROTEOMICS (AIM

## 3) **Microbiome: Microbial genetic potential and**

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

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

**DNA** from samples. **16S rRNA** (economical) or **Shotgun sequencing** (expensive).

Multiple studies that **correlate taxonomy with observed phenotype.**

- **TAXONOMIC COMPOSITION**

## **Metaproteomics:**

The large-scale characterization of the entire protein complement of environmental microbiota at a given point in time.

**Proteins** from samples.

Potential to unravel the **mechanistic details of microbial interactions with host / environment** by analyzing the **functional dynamics of the microbiome.**

- **TAXONOMIC COMPOSITION**
- **FUNCTIONAL EXPRESSION**

## DATABASE GENERATION

## DATABASE SEARCH & STRATEGIES

## FUNCTIONAL ANALYSIS

**DECEMBER 2016:**

*Metaproteomics  
Contribution-fest*

<http://z.umn.edu/mphack2016>

Metagenomic  
database

*In silico  
Translation*

Translated  
metaproteomic  
database

Database  
Generation

Target-Decoy database

Peak list  
(MGF)

PSM Report from the  
search with microbial peptides

Microbial  
Peptides

PeptideShaker

searchgui

UniPept Analysis (Pept2Prot)

UniPept	x
Peptide Input	
output_json (json)	<input type="radio"/>
output_tsv (tabular)	<input type="radio"/>
output_csv (csv)	<input type="radio"/>
output_unmatched (tabular)	<input type="radio"/>

- Peptides to PSMs
- Biological Processes and PSMs
- Molecular Function and PSMs
- Cellular localization and PSMs

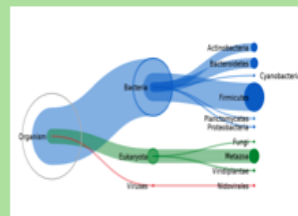
GO Term  
Mapping

Functional analysis using GO Term analysis

UniPept Analysis (Pept2LCA)

UniPept	x
Peptide Input	
output_json (json)	<input type="radio"/>
output_tsv (tabular)	<input type="radio"/>
output_csv (csv)	<input type="radio"/>
output_unmatched (tabular)	<input type="radio"/>

- Genera and PSMs



## TAXONOMY ANALYSIS

**APRIL 2017:**

*Metaproteomics  
Gateway*

[z.umn.edu/metaproteomicsgateway](http://z.umn.edu/metaproteomicsgateway)

*Jetstream*



**NATIONAL CANCER INSTITUTE**  
Informatics Technology for  
Cancer Research





# PROMOTION AND DISSEMINATION (AIM 4)

WEBSITE: [galaxyp.org](http://galaxyp.org)

## WORKSHOPS AT ANNUAL CONFERENCES:

**American Society of Mass Spectrometry (ASMS)**

Minneapolis 2013 . Baltimore 2014 .  
St. Louis 2015 . San Antonio 2016 .  
Indianapolis 2017

**Galaxy Community Conference (GCC)**  
Norwich 2015 . Bloomington 2016  
Montpellier 2017

**Association of Biomolecular Research Facilities (ABRF)**  
Fort Lauderdale 2016 . San Diego 2017

## MANUSCRIPTS:

being actively developed by the Griffin research group at the University of Minnesota Supercomputing Institute. The project is funded by the National Science Foundation and the National Cancer Institute's Informatics Technology for Cancer Research program.

[z.umn.edu/galaxypreferences](http://z.umn.edu/galaxypreferences)

## ORAL PRESENTATIONS AND POSTERS:

**ASMS**

2013 . 2017  
**GCC**  
Baltimore 2014 . St. Louis 2015 . San Antonio 2016

**ABRF**

Albuquerque 2014 . St. Louis 2015

**US Human Proteomics Organization**  
Seattle 2014 . Tempe 2015

**American Thoracic Society**  
2017 Washington DC

**International Metaproteomics Symposium**  
2017 Alghero



## TWITTER:

[twitter.com/usegalaxyp](https://twitter.com/usegalaxyp)



# FUTURE PLANS

- **MULTI-VISUALIZATION PLATFORM**

- Extend to more variant types: define requirements for input data types
- Integration with Javascript IGV (ITCR)  NATIONAL CANCER INSTITUTE  
Informatics Technology for  
Cancer Research
- Proteogenomics output compatibility with CRAVAT for interpretation (ITCR)  NATIONAL CANCER INSTITUTE  
Informatics Technology for  
Cancer Research

- **METABOLOMICS**

- Automate for assistance in experimental design; continued testing and feedback
- Portability: Dissemination via cloud-based gateway (Jetstream) and Tool Shed

- **PROTEOGENOMICS**

- Automate generation of genomic annotation for all proteins and variant type
- Extend with new methods for sequential variant-specific database searching
- Extend with improved FDR assignment for putative variants

- **METAPROTEOMICS**

- Identify cancer-specific projects and implement tools and workflows
- Develop tools and workflows via developers and users collaborative contribution-fest
- Disseminate workflows via publications and JetStream instance

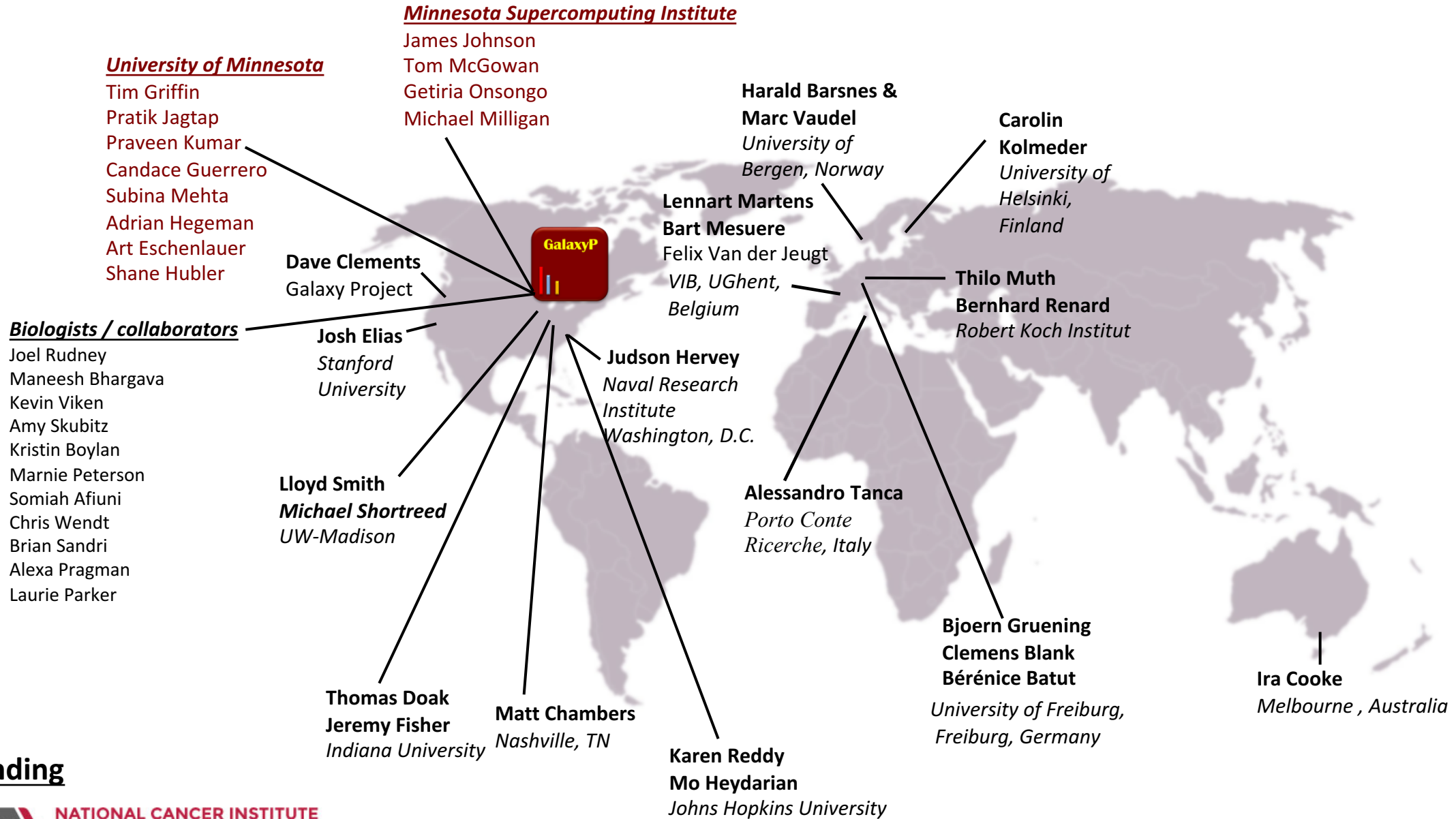
- **PROMOTION AND TRAINING**

- Continued workshops and presentations (ABRF, ASMS, GCC....AACR)
- Webinars, hackathons
- Expand Jetstream resources 

- **NEW ADDITION!**

Supplement IMAT-ITCR project for DIA tool deployment (with Parker lab, UMN)

# ACKNOWLEDGEMENTS



## Funding

