

Extending Galaxy's reach: recent progress towards complete multi-omic data analysis workflows



Galaxy Community Conference 2015
July 8, 2015

Tim Griffin
University of Minnesota
tgriffin@umn.edu



Outline

- A mass spectrometrists/protein biochemist's winding road to Galaxy awareness
- The Galaxy-P project and “multi-omics”
- Thoughts about the future



Starting point: an acknowledgement to the Galaxy team

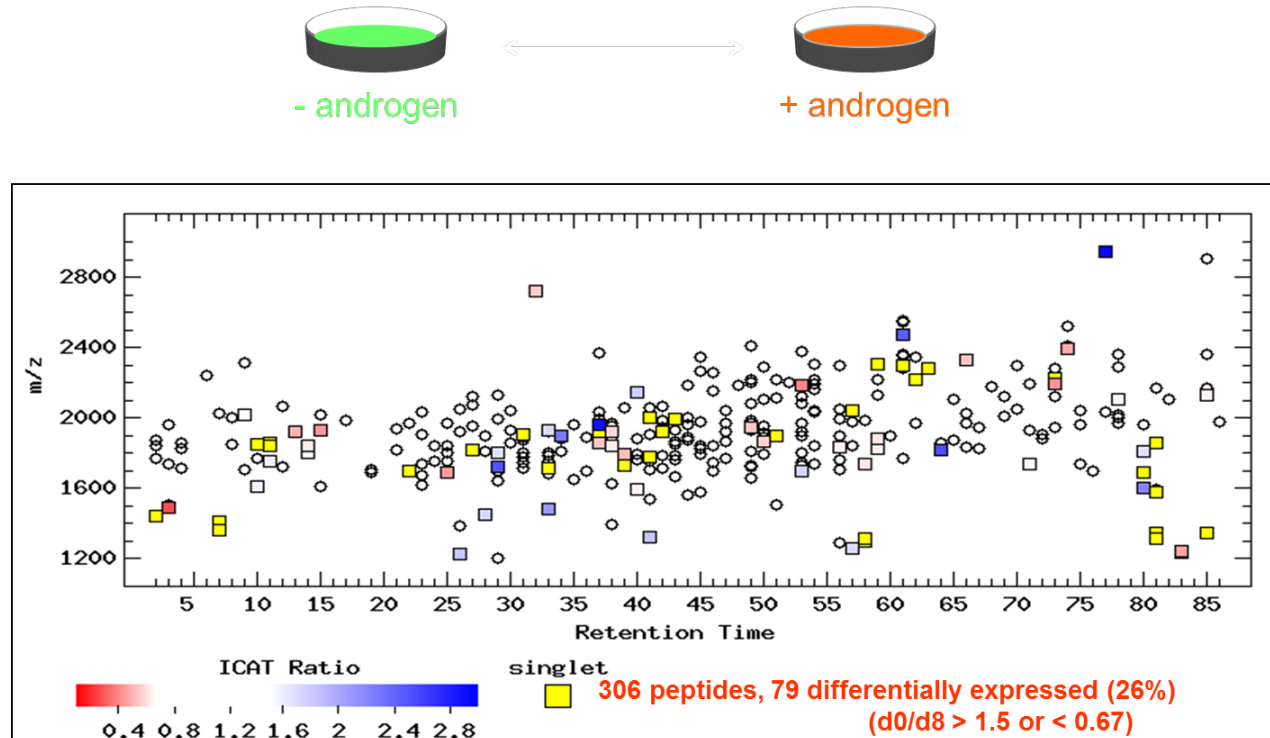


Thank you to the Galaxy creators, core team
and community for your vision and hard work!



A journey towards Galaxy awareness...

Quantitative proteomics circa 2002



Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal Chem.* 75 867-874



The speaker circa 2002



Proteomic data analysis circa 2002



There must be a better way.....

On to the University of Minnesota....



- Mass spectrometry-based proteomics data volume and complexity continues to grow and so does needs for better computational tools
- **Dabbling in developing our own software:**

LTD-iQuant: A freely available software pipeline for automated and accurate protein quantification of isobaric tagged peptide data from LTD instruments

Proteomics 2010, 10, 3533–3538

Getiria Onsongo¹, Matthew D. Stone², Susan K. Van Riper³, John Chilton⁴, Baolin Wu⁵, LeeAnn Higgins², Troy C. Lund⁶, John V. Carlis¹ and Timothy J. Griffin²

- **Sought out resources on campus:**

UNIVERSITY OF MINNESOTA
**SUPERCOMPUTING
INSTITUTE**



Our in-house workflow engine, circa 2008



John Chilton



Pratik Jagtap

The TINT portal is a web 2.0 GUI for interacting with a pipeline providing 3rd party tools for analysis of raw mass spectrometry data

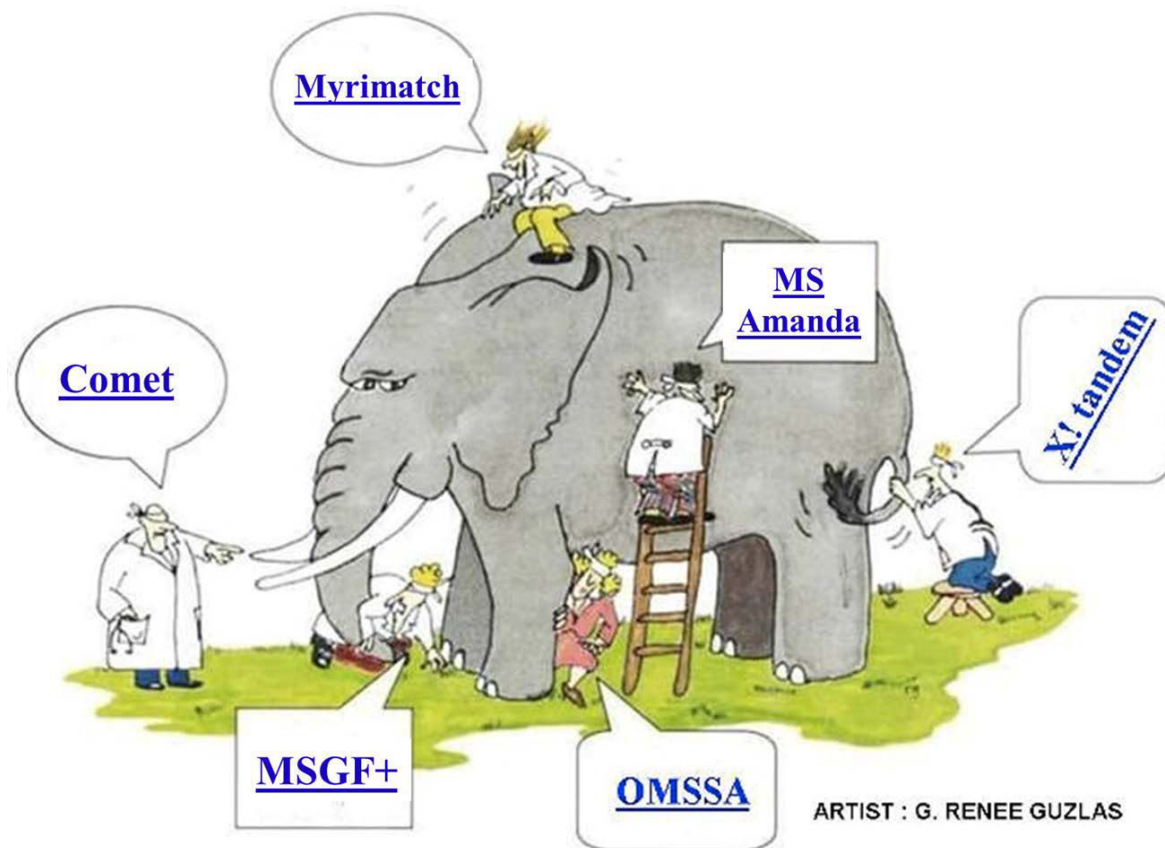
"[TINT] has provided essential and enabling infrastructure for numerous projects using mass spectrometry-based proteomics in the Griffin lab"

-- Dr. Tim Griffin, Department of Biochemistry, Molecular Biology, and Biophysics



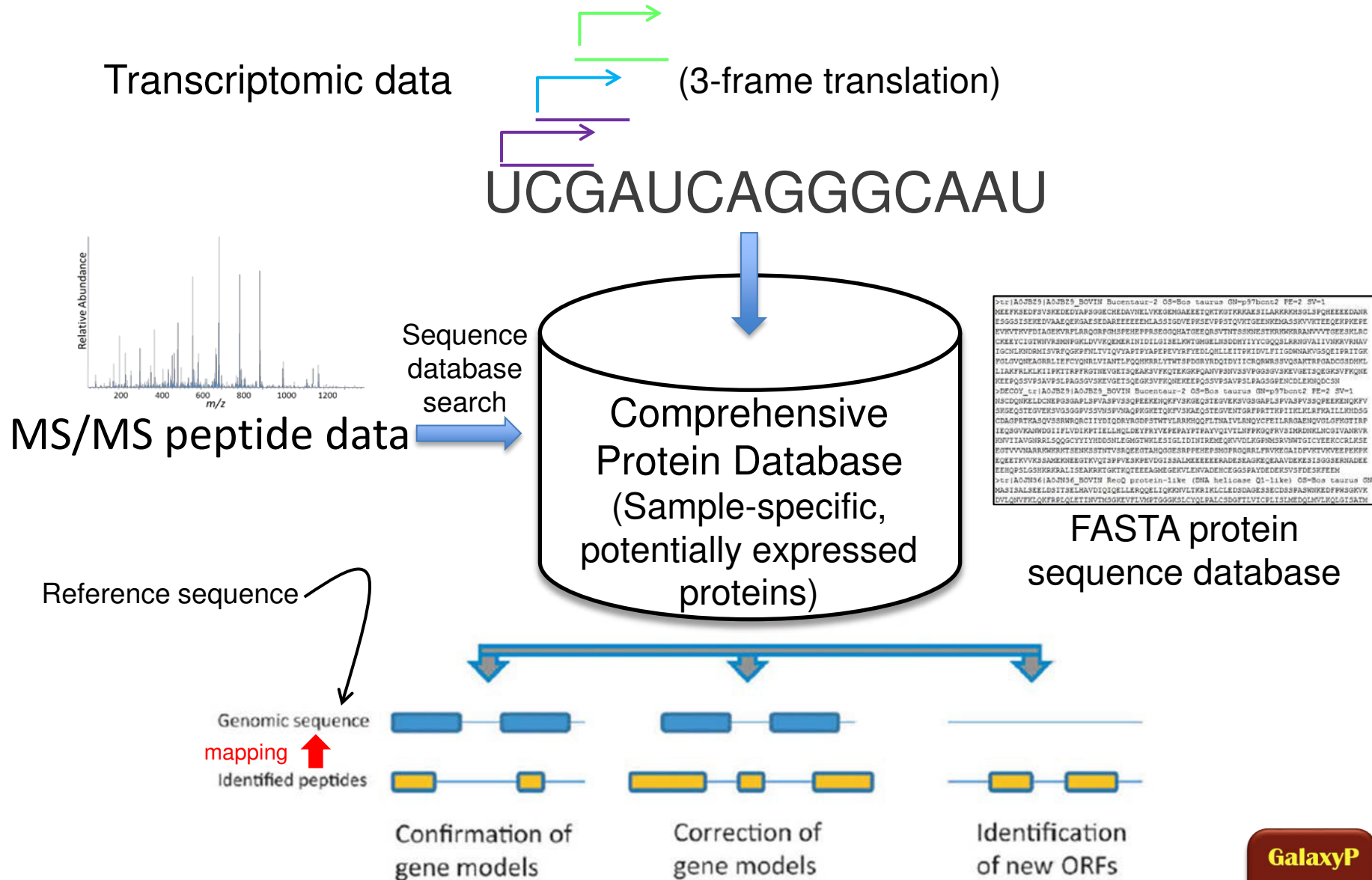
All our problems are solved, right?

Not quite....



- New software for protein identification and quantification continued to emerge...how to we integrate it into our pipeline?

And..."multi-omic" approaches started to catch our eye



So how do we solve our “big data” problems?

Extend TINT?
Something new?



How about Galaxy?

Huh?!?



Anne-Françoise Lamblin

UNIVERSITY OF MINNESOTA
SUPERCOMPUTING
INSTITUTE



Galaxy-P: A new community-based informatics paradigm for MS-based proteomics

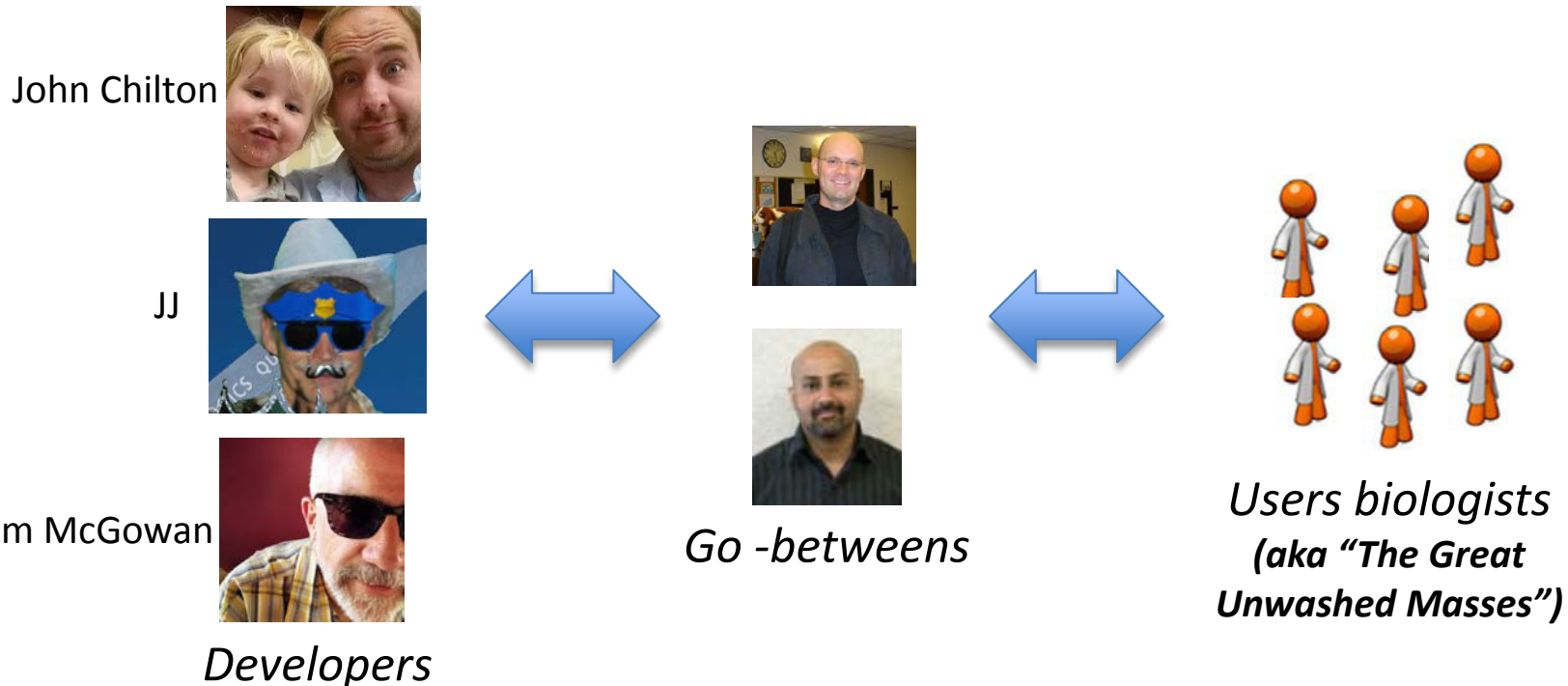


The Galaxy-P project

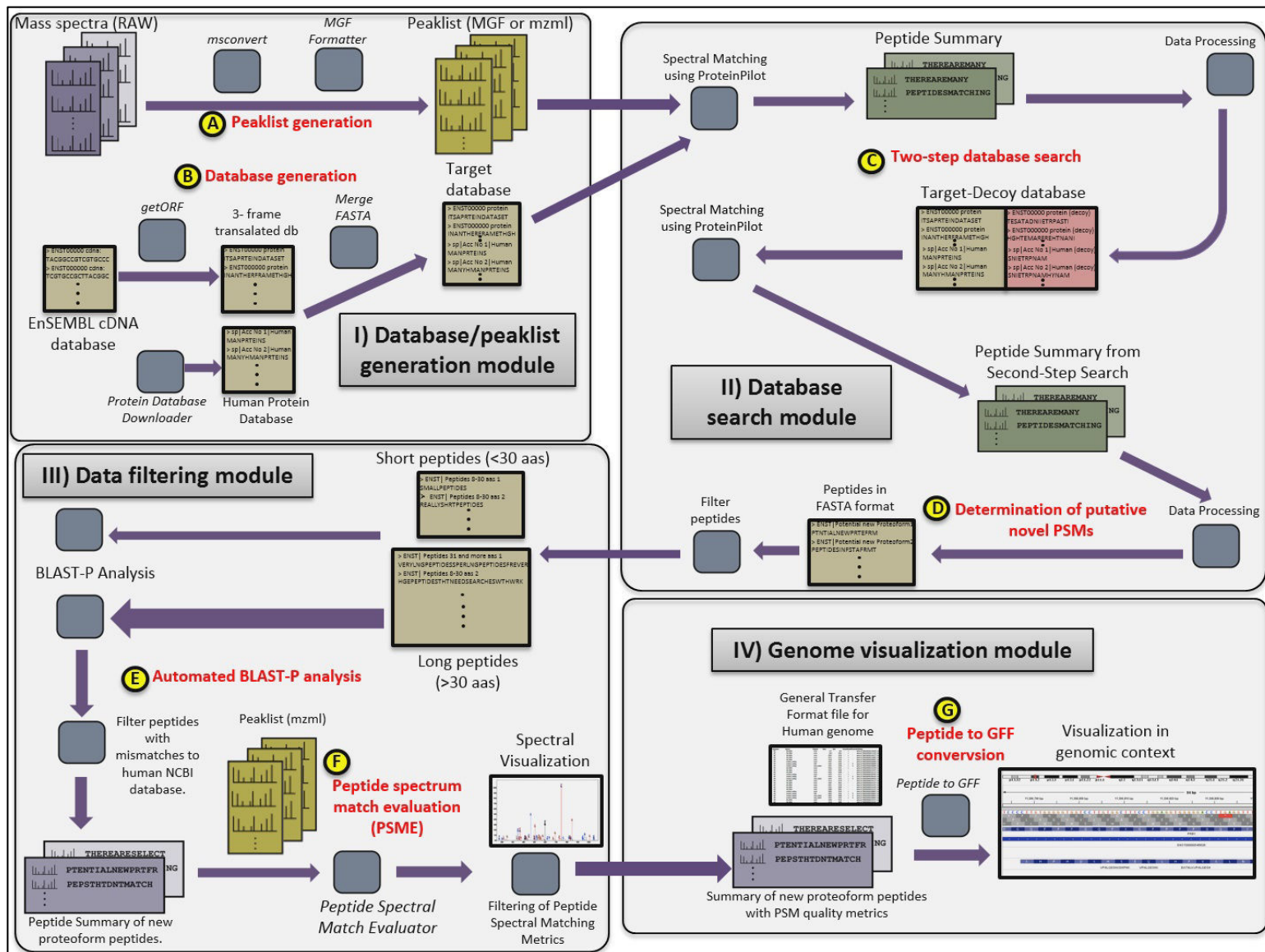


- Define data types and operations and deploy basic tools for mass spectrometry-based proteomics data analysis
- Focus on multi-omic approaches in proteogenomics and metaproteomics
- Promote the use of Galaxy to a new user community

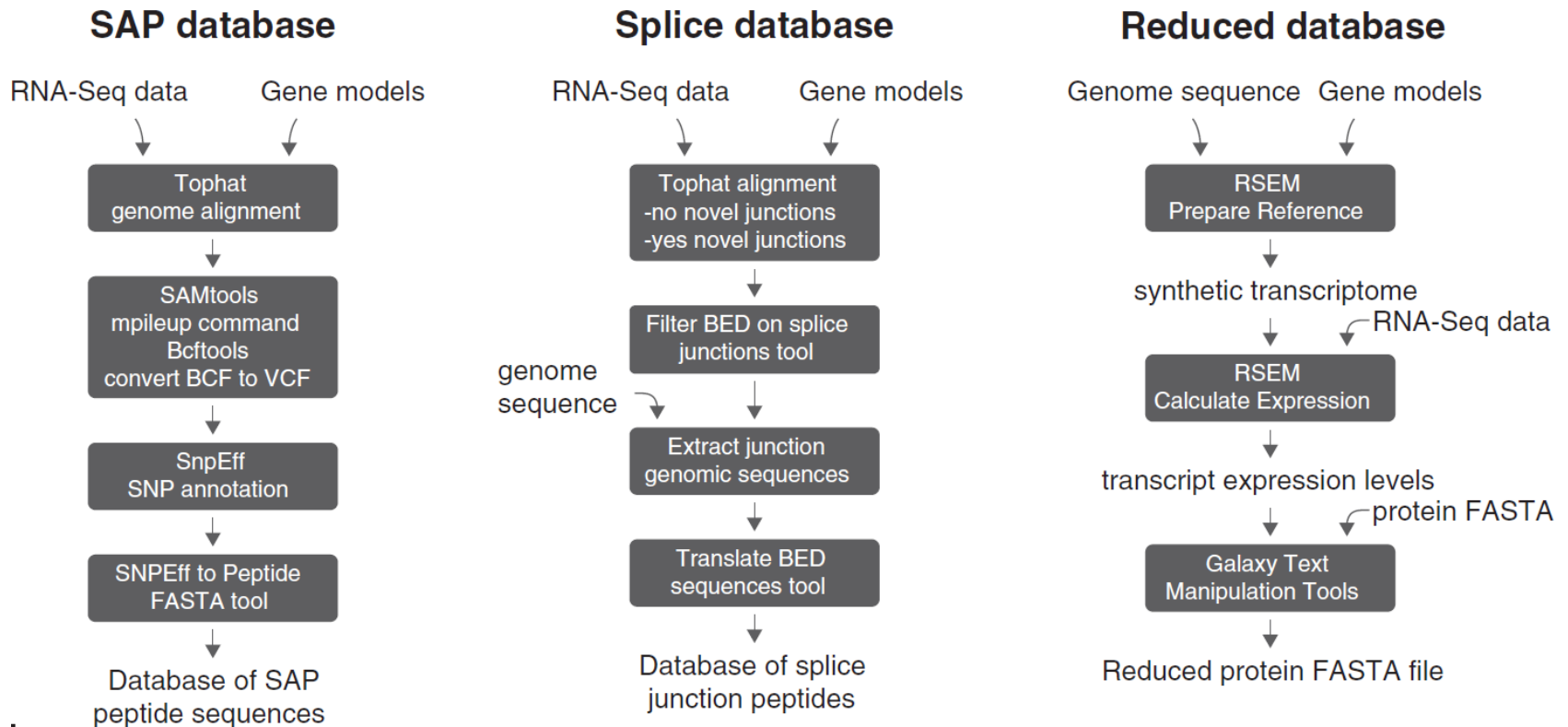
Galaxy-P working model



Putting Galaxy to work in proteogenomics



Module 1 – Database generation : Playing to Galaxy's strengths



Sheynkman *et al. BMC Genomics* 2014, **15**:703
<http://www.biomedcentral.com/1471-2164/15/703>



METHODOLOGY ARTICLE

Open Access

Using Galaxy-P to leverage RNA-Seq for the discovery of novel protein variations

Gloria M Sheynkman¹, James E Johnson², Pratik D Jagtap^{3,4}, Michael R Shortreed¹, Getiria Onsongo², Brian L Frey¹, Timothy J Griffin^{3,4} and Lloyd M Smith^{1,5*}



Module 2 – Sequence database searching: global development



SearchGUI
(multiple search programs)

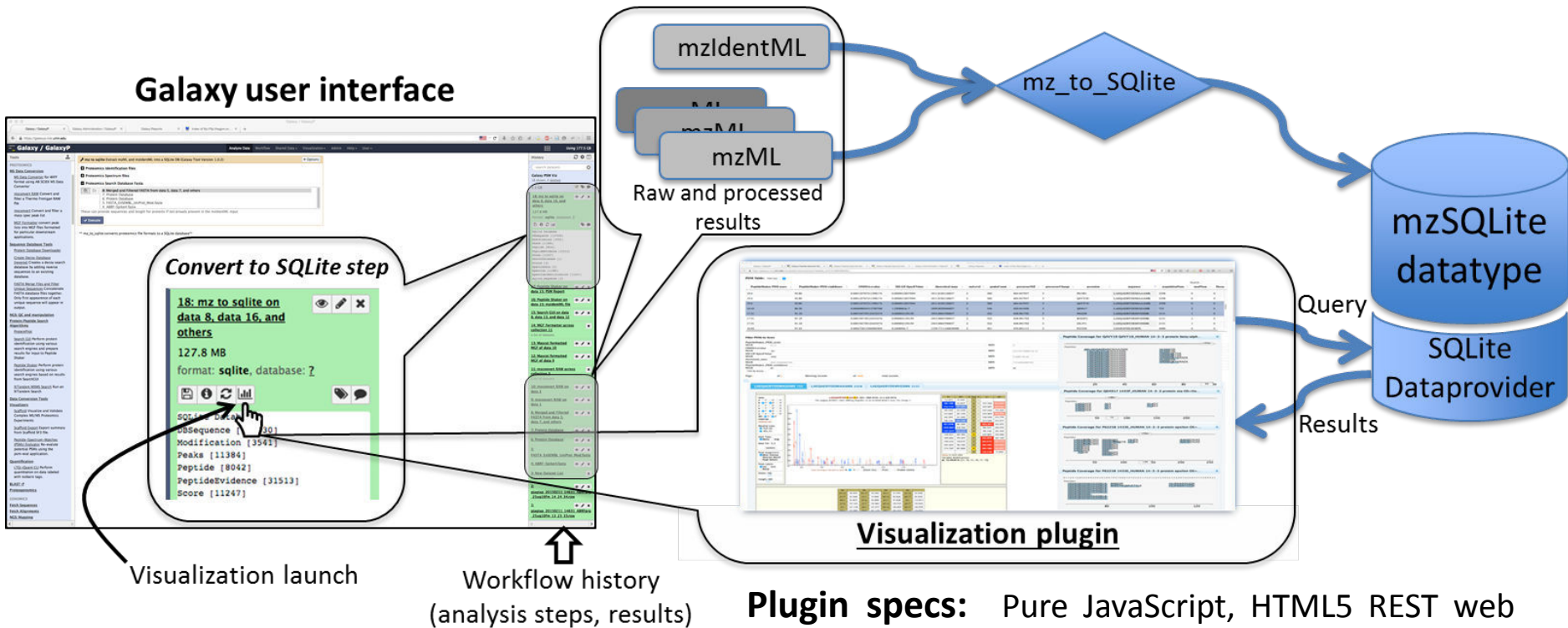


PeptideShaker
Nat Biotechnol. 2015 33:22-4



Module 3 – Data Filtering/QC/Visualization

Galaxy user interface



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



JJ



Tom McGowan



Training and promotion activities

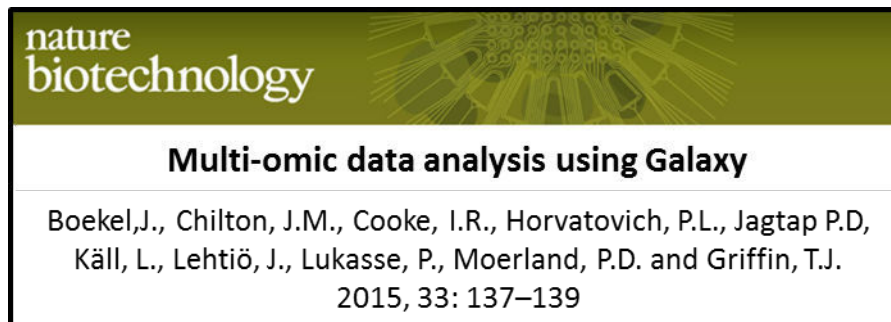
- A tool for training undergraduates to postdocs in bioinformatics:
 - Students with computational or biological interests and those in-between



- Promoting Galaxy-P to target user-communities:



- Publication:



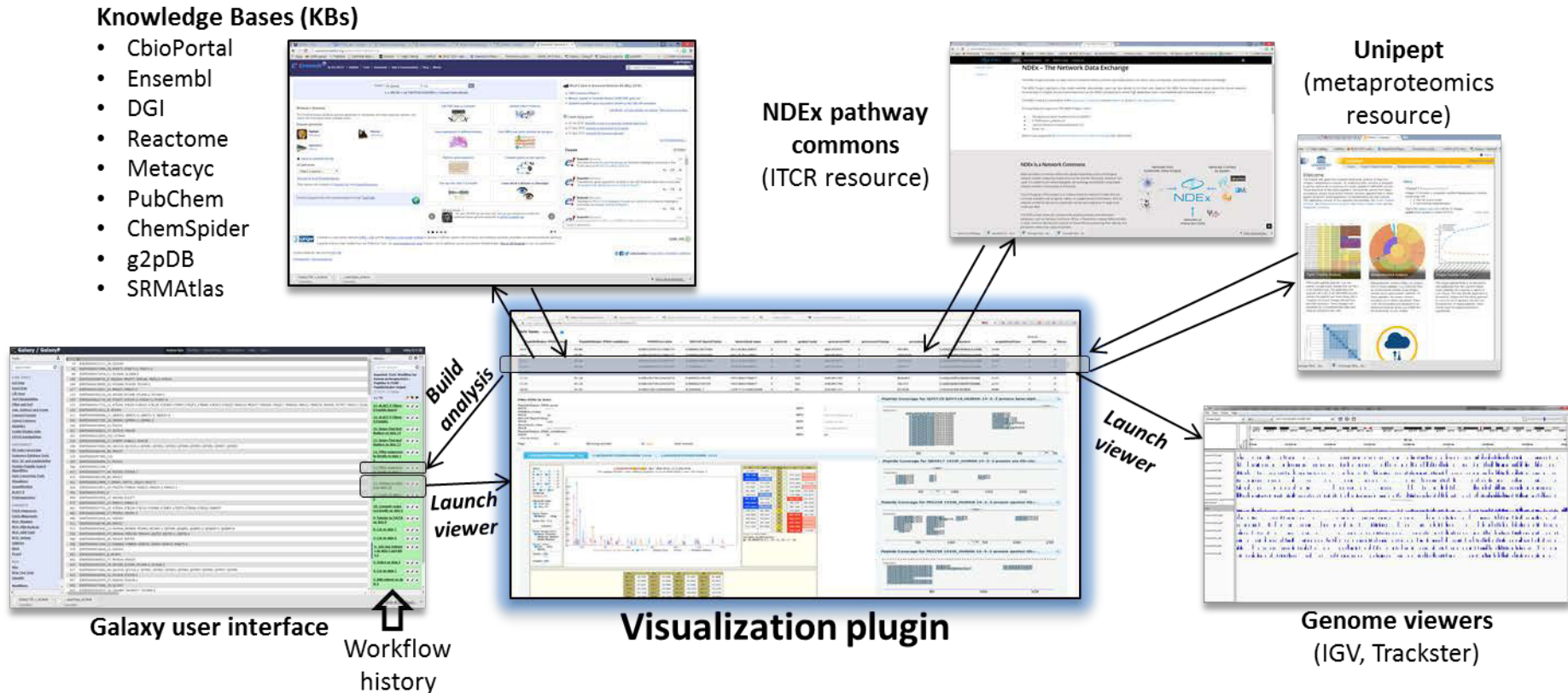
Where are we going?

- **Show biological value!!**
 - Lung disease proteogenomics and metaproteomics
 - Ground squirrel proteogenomics
 - Periodontal metaproteomics
 - Ovarian cancer proteogenomics and metaproteomics
 - mRNA regulation in cancer (proteogenomics)
 - ***And always looking for more.....!!***
- **New avenues in proteomics:** Data-independent acquisition
- **Galaxy-M:** metabolomics tools (in process)
- **Easier install for the non-expert:** Docker containers etc.
- **Implementation of Galaxy-P on cloud resources:** Globus Genomics, Chorus
- Extending the power of **visualization** and **interpretation** tools



Some musings on extensions to our visualization plugin

- Automating biological interpretation and hypothesis-generation





UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Biochemistry, Molecular Biology & Biophysics

Kevin Murray
Ray Sajulga
Candace Guerrero

Center for Mass Spectrometry and Proteomics

Pratik Jagtap
Ebbing de Jong
LeeAnn Higgins
Todd Markowski

UNIVERSITY OF MINNESOTA
**SUPERCOMPUTING
INSTITUTE**

James Johnson
Tom McGowan
Trevor Wennblom
Getiria Onsongo
Bill Gallip
Ben Lynch
Anne Lamblin
Bart Gottschalk



COMMUNITY BASED SOFTWARE DEVELOPMENT

Harald Barsnes and Marc Vaudel
University of Bergen, Bergen, Norway
Bjoern Gruening
University of Freiburg, Freiburg, Germany
Lennart Martens
VIB Department of Medical Protein Research, Ugent, Belgium
Ira Cooke
La Trobe University, Melbourne, Australia
John Chilton
Galaxy Team
Penn State University



Dave Clements

Funding
NSF, NIH

