



# **Globus Genomics – *Powered by Galaxy***

Ravi K Madduri and Globus  
Genomics Team

- Introduction
  - Computation Institute
- Galaxy and Computation Institute
  - Breadth of projects using Galaxy
  - How Globus Genomics came about
- Who Globus Genomics is for?
- Globus Genomics Solution overview
- Example Collaborations
- Benefits to the Galaxy Community
- Discussion



# Computation Institute

- A joint institute of Argonne National Laboratory and the University of Chicago
- Goal: Accelerate discovery by individual researchers and reduce costs for both individuals and institutions by providing robust research data management as a service
- Leverage software-as-a-service to:
  - provide millions of researchers with unprecedented access to powerful tools
  - reduce research IT costs dramatically via economies of scale
- **We are a non-profit service provider to the non-profit research community**

# Breadth of CI Galaxy-related projects

- Genomics
- Cosmology
- Materials Science
- X-Ray Beam Line Science
- Climate Science





# How Globus Genomics Came about

- Started our efforts with CVRG – utilizing Galaxy to implement RNA-Seq analysis pipelines related to cardiovascular related research
- Applied the capabilities to help researchers at the University of Washington investigate neurodevelopmental disorders utilizing Globus to move files from commercial sequencing centers and executing Exome analysis workflows in Galaxy
- Broadened the suite of pipelines to include additional RNA-Seq and ChIP-Seq analysis at WUSTL in support of cancer research
- Enhanced research efforts associated with cancer relapse by extending Exome-Seq pipelines with additional downstream analysis with researchers at the University of Chicago
- Collaborated with leading diabetes researchers at the University of Chicago to apply consensus calling approaches to variant identification
- Engaging with a pediatric oncologist to develop proteomic pipelines in Galaxy
- Continuing to field growing interest in utilizing both Globus and Galaxy to address needs in biomedical research...

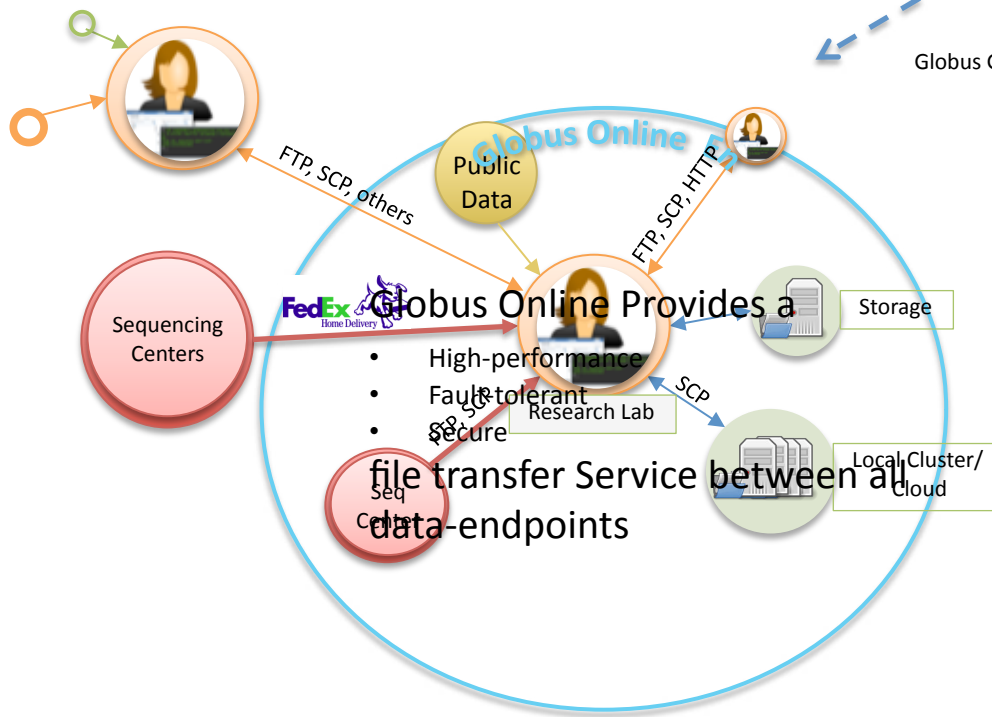
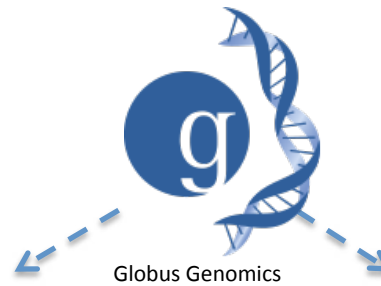
# Who Globus Genomics is For?

- **Remember Dr. Karen Reddy's Talk from GCC 2012?**
  - Our goals of accessibility, democratization, sustainability are already motivated by the keynote and BioTeam's presentations yesterday
- **Cloudman and Biocloudcentral are great and serve a lot of users and increase adoption of Galaxy**
  - Usage may still require considerable IT expertise
- **With Globus Genomics we wish to serve**
  - Individual PIs and small labs who don't have the IT expertise to setup and run Galaxy
  - Core labs at Universities who wish to focus time working with researchers on downstream analysis
- **We are Non-Profit**
  - Subscription charged for amazon and operations, continued development



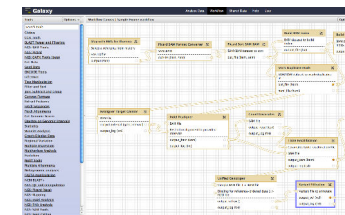
# Globus Genomics – Solution overview

- Integrated Globus-Galaxy platform
- Provides access to scalable and elastic compute resources via AWS
- Provide technical support and bioinformatics consulting
- Provide access to pre-integrated end-points for reliable and high-performance data transfer (Broad, Perkin Elmer, UC FGF, etc.)



Data Management

## Galaxy Based Workflow Management System



Galaxy  
Data Libraries

- Globus Online Integrated within Galaxy
- Web-based UI
- Drag-Drop workflow creations
- Easily modify Workflows with new tools



Globus Genomics on  
Amazon EC2

- Analytical tools are automatically run on the scalable compute resources when possible

Data Analysis





# Example Collaborations

## Onel-Skol Lab



**Background:** Cancer researchers sequencing normal and relapse genomes from cancer patients to investigate genetic factors in cancer relapse

**Approach:** Replaced outsourced analysis with Globus Genomics

**Results:** Achieved greater than 10X speed-up in analysis of NGS data and 10X cost savings compared to alternative solutions

**Future Plans:** Leverage flexibility in Globus Genomics to extend analysis pipelines and compare results utilizing recently added algorithms



# Example Collaborations

## Dobyns Lab



**Background:** Investigate the nature and causes of a wide range of human developmental brain disorders

**Approach:** Replaced manual analysis with Globus Genomics

**Results:** Achieved greater than 10X speed-up in analysis of exome data

**Future Plans:** Leverage scale-out capability of Globus Genomics by running increasingly larger data sets



# Example Collaborations

## Georgetown Medical Center



**Background:** Innovation Center for Biomedical Informatics is an academic hub for innovative research in the field of biomedical informatics.

**Approach:** Augment current team and tools with a NGS analysis platform to support standard and best-practice pipelines while leveraging elastic cloud-based resources.

**Results:** Pilot effort is just getting underway

**Future Plans:** Provide Globus Genomics as a well-managed platform-as-a-service for ICBI collaborators and users

- **Increase adoption of Galaxy**
  - Both Genomics and non-Genomics
- **Contributing tools, tool wrappers to the toolshed**
  - This is already happening, you should see more of it in coming weeks
- **Add support for different computational modalities**
- **Further integration of Globus Online services for data management**
- **Help Accelerate research and lower time to science**



# Questions?

- **For more information on Globus Online and Globus Genomics**
  - [www.globusonline.org](http://www.globusonline.org)
  - [www.globus.org/genomics](http://www.globus.org/genomics)
- **Acknowledgements**
  - The Cardiovascular Research Grid (R24HL085343)
  - U.S. Department of Energy under contract DE-AC02-06CH11357
  - Amazon Web Services
  - The Galaxy Community



Thank you!

