Globus Galaxies
Platform for Science as a Service

Ravi K Madduri, University of Chicago and Argonne National Laboratory
madduri@uchicago.edu
@madduri
Our vision for a 21st century discovery infrastructure

Provide more capability for more people at lower cost by delivering “Science as a Service”

www.globus.org
Science as a Service
Examples

- Globus Genomics
- PDACS - Portal for data analysis services for cosmological simulations
- CVRG Galaxy – Large-scale ECG Data Analysis
- TBI Analysis – Traumatic Brain Injury Image analysis portal
- Globus Proteomics
- eMatter – Material Science Simulations
- FACE-IT - Framework to Advance Climate, Economic, and Impact Investigations with Information Technology
Our Science Stack

- Galaxy
  - Interactive execution
  - Creation, Execution, Sharing, Discovering Workflows
- Globus
  - Data management
  - Identity Management
- AWS, NERSc, Magellan
  - HTCondor, Chef, EC2, EBS, S3, SNS, NEWT
  - Spot, Route 53, Cloud Formation
Adopting Globus Galaxies For a Science Domain
Winning Combination…

• People and their expertise
  – Computational experts working hand-in-hand with Subject Matter Experts
  – Separation of Concerns (Infrastructure Vs Science)
• Flexible, extensible platform
  – Data types, Tools, Visualization, Data providers
• Leveraging proven best practices in data management, execution management
• Community Engagement
  – Workshops, Demos
Workflow for Illumina RNA-seq

This workflow analyzes Illumina RNA-seq data. It includes steps for mapping reads to reference genomes, quantifying gene expression, and performing differential expression analysis. The results can be visualized using various heatmap and scatter plots. This workflow is particularly useful for analyzing gene expression patterns across different biological conditions.

Workflow for Illumina Exome-seq

This workflow is designed for analyzing Illumina exome sequencing data. It includes steps for variant calling, annotation, and functional analysis. The workflow can help in identifying genetic variants associated with specific traits or diseases. The results can be further explored using tools for genome browser and annotation viewer.

Workflow for Illumina ChIP-seq

ChIP-seq combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. It can be used to map global binding sites for any protein of interest. Analysis includes QCl, mapping to hg19 and identification of peaks.

Globus Genomics

FACE-IT

CVRG

Proteomics
Our work is supported by:

- U.S. DEPARTMENT OF ENERGY
- NSF
- NATIONAL INSTITUTES OF HEALTH
- THE UNIVERSITY OF CHICAGO
- Argonne NATIONAL LABORATORY
- Amazon Web Services
Thank you!

@madduri