

Computation Institute

# Globus Galaxies Platform for Science as a Service

Ravi K Madduri, University of Chicago and Argonne  
National Laboratory

[madduri@uchicago.edu](mailto:madduri@uchicago.edu)

@madduri



computationinstitute.org



Our vision for a 21st century  
discovery infrastructure

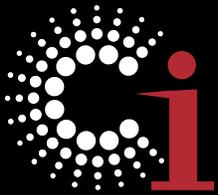
Provide **more** capability for  
**more** people at **lower cost** by  
delivering “**Science as a Service**”

[www.globus.org](http://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

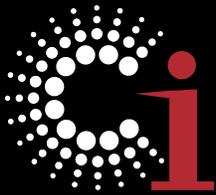
SaaS

PaaS

IaaS



# 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

globus genomics | Galaxy

Tools

search tools

MATERIAL SCIENCE

DATA TRANSFER

NGS APPLICATIONS

DATA MANIPULATION



system status ●

GET STARTED

[Workflow for Illumina RNA-seq »](#)  
Provide information on differential gene expression between NGS samples including alleles and spliced transcripts. This analysis is for paired-end sequences. Includes QC, mapping to hg19 and expression of genes.

[Workflow for Illumina Exome-seq »](#)  
This analysis is an efficient strategy to selectively sequence the coding regions of the genome. The goal of this approach is to identify the functional variations in the exome regions. Analysis for paired-end sequences. Includes QC, mapping to hg19 and variants list.

[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 precisely for any protein of interest. Analysis includes QC, mapping to hg19 and identification of peaks.

Your History

Unnamed history 53.9 MB

- 20: Send: test.vcf
- 19: Variant Filtration on data 14 (log)
- 18: Variant Filtration on data 14 (Variant File)
- 17: Send: test.bam
- 16: Unified Genotyper on data 11 (log)
- 15: Unified Genotyper on data 11 (metrics)
- 14: Unified Genotyper on data 11 (VCF)
- 13: flagstat on data 11
- 12: MarkDups\_deDup.html
- 11: MarkDups\_deDup.bam
- 10: Add or Replace Groups on data 9: bam with read groups replaced
- 9: Paired Read Mate Fixer on data 8: bam with fixed mates
- 8: Reorder SAM/BAM on data 7: reordered bam
- 7: Bowtie2 on data 4 and

FACE-IT | Galaxy

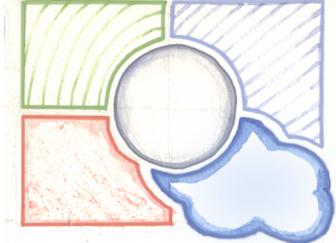
Tools

search tools

Get Data

Workflows

Framework to Advance Climate, Economic, and Impact Investigations with Information Technology



FACE-IT is supported by the NSF cyberSEES program award No.1331782

The Galaxy project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

The Galaxy-ES (Earth System) toolshed is part of the FACE-IT project.

The FACE-IT framework is being developed out of a collaboration between the University of Chicago Computation Institute's center for Robust Decision-making in Climate and Energy Policy (RDCEP) and the University of Florida (ABE/UFPI) to meet the needs of several international communities of researchers working on issues around climate change vulnerabilities, impacts, adaptations, and mitigation.

Your History

Unnamed history 6 shown, 4 deleted 2.2 MB

- 9: Broad2.png
- 8: DayMet Browser
- 7: 193\_370.psims.nc
- 6: DayMet Browser
- 5: Broad2.png
- 2: soil.json

# Globus Genomics

# FACE-IT

globus genomics | Galaxy for CVRG

Tools

search tools

Tool Installer

DATA TRANSFER

CVRG TOOLS AND MODELS

NGS APPLICATIONS

TOOL UPDATES



system status ●

GET STARTED

[Workflow for ECG analysis using Physionet Tools »](#)  
This workflow analyzes ECG datasets using various Physionet apps like wqrs, sqrs, nguess, ann2r

[Calcium Spark »](#)  
The Calcium Spark tools can be used to run and analyze customized simulations of the cardiac calcium release unit model presented in Walker et al. 2014 Biophys J 107(12).

[CellML »](#)  
The CellML tools provide the ability to modify and run computational models encoded in the CellML language.

[Cardiac Myocyte »](#)  
The Cardiac Myocyte tool allows users to customize and simulate an integrative

Your History

project-234-CVRG - sqrs-Mon\_Apr\_29\_2013\_8:56:3 1\_PM 45.1 KB

- 8: Output of ann2r on data 1 and data 6
- 7: Output of ann2r on data 1 and data 5
- 6: Output of nguess on data 1 and data 4
- 5: Output of nguess on data 1 and data 3
- 4: sqrs on data 1 and data 2: Reference annotation file
- 3: wqrs on data 1 and data 2: Reference annotation file
- 2: 234.dat
- 1: 234.hea

BDDS | Galaxy

Tools

search tools

DATA TRANSFER

PROTEOMICS APPLICATIONS

NGS APPLICATIONS

TOOL UPDATES

**BIG DATA for DISCOVERY SCIENCE**

GET STARTED

[Workflow for Illumina RNA-seq »](#)  
Provide information on differential gene expression between NGS samples including alleles and spliced transcripts. This analysis is for paired-end sequences. Includes QC, mapping to hg19 and expression of genes.

[Workflow for Illumina Exome-seq »](#)  
This analysis is an efficient strategy to selectively sequence the coding regions of the genome. The goal of this approach is to identify the functional variations in the exome regions. Analysis for paired-end sequences. Includes QC, mapping to hg19 and variants list.

[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 precisely for any protein of interest. Analysis includes QC, mapping to hg19 and identification of peaks.

[SeattleSeq »](#)

Your History

Unnamed history 93.0 MB

- 18: ASAPRatio Peptide Level Analysis of data 16
- 17: ASAPRatio Protein Level Analysis of data 16
- 16: protein\_prophet.interprophet.peptide\_prophet.XITandem\_vs\_yearst\_0rfs\_all\_REV.20060126.short.fasta.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.pepXML.pep.xml
- 15: interprophet.peptide\_prophet.XITandem\_vs\_yearst\_0rfs\_all\_REV.20060126.short.fasta.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.pepXML.pep.xml
- 14: peptide\_prophet.XITandem\_vs\_yearst\_0rfs\_all\_REV.20060126.short.fasta.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.pepXML.pep.xml
- 13: XITandem\_vs\_yearst\_0rfs\_all\_REV.20060126.short.fasta.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.0R20080317\_S\_SILAC-LH-1-1\_01.mzML.pepXML.pep.xml

# CVRG

# Proteomics

Our work is supported by:

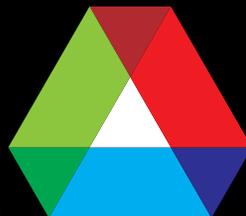


U.S. DEPARTMENT OF  
**ENERGY**

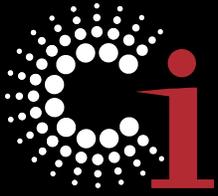


THE UNIVERSITY OF  
**CHICAGO**

**Argonne**  
NATIONAL LABORATORY



[computationinstitute.org](http://computationinstitute.org)



Thank you!

@madduri