

GLBRC 2014 Retreat Informatics Tutorials

Tutorial 1. The GLBRC RNA-seq Analysis Pipelines

Tuesday 7:00-8:00 PM. Century Center Suites 8-9. *Tutors:* Nick Thrower, Oleg Moskvina

What they are, why they are, how to access them, and more.

Part A. RNA-Seq analysis can be run using the CLC Genomics Workbench. This tool allows filtering, mapping, assembly and quantification of NGS data among several other features. Many labs in Area 1 rely on the CLC pipeline.

Part B. Overview of custom GLBRC RNA-Seq processing pipelines, including low-level processing (GLSeq) and Downstream Analysis Assistant (DAA) pipelines.

You will learn:

1. Differences and details for each pipeline
2. What to expect from a finished analysis
3. How to get your data analyzed
4. What's in the different output files

Tutorial 2. Using the GLBRC Hub for Effective Collaboration

Tuesday 7:00-8:00 PM. Century Center Suites 10-11. *Tutor:* Dirk Norman

The Hub is a web-based content management system that can be used for document management, task management, process automation, or just provide a central place where a group of people can share "stuff" knowing it's protected.

You will learn:

1. The basic layout of the system
2. Document and task management
3. How to share content

Tutorial 3. Cytoscape: Exploring Systems Biology Dimensions of Your Project

Wednesday 12:45-1:30 PM. Century Center Suites 8-9. *Tutor:* Oleg Moskvina

Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data.

You will learn:

1. Strategies: ways to use Cytoscape in your project
2. Tactics: overview of user interface and useful apps for Cytoscape 3.X

Tutorial 4. The GLBRC Cell Wall Facility LIMS

Wednesday 12:45-1:30 PM. Century Center Suites 10-11. *Tutors:* Nick Santoro, Cliff Foster

The Cell Wall Facility accepts samples for high throughput digestibility and sugar composition analysis.

You will learn:

1. Protocols and assays available from the facility
2. How to create experiments and submit new samples
3. How to retrieve analysis results

Tutorial 5. Using the GLBRC's Compute and Big Data Storage Infrastructure

Wednesday 1:30-2:30 PM. Century Center Suites 8-9. *Tutors:* Dirk Norman, Branden Timm

The GLBRC's central data repository automatically syncs and stores GLBRC sequencing data. This data is then accessible from internal GLBRC compute servers.

You will learn:

1. What data is automatically synced to GLBRC storage and how you can access it
2. Available compute resources within the GLBRC and how to access them
3. Tips and tricks for working in a Linux environment and basic Condor usage

Tutorial 6. Finding your Samples, Experiments, and Analytical Information in STARLIMS

Wednesday 1:30-2:30 PM. Century Center Suites 10-11. *Tutor:* Ying Gao

The GLBRC Central LIMS collects and manages GLBRC production chain materials and all QC data. The LIMS also manages fermentation experiments and multiomic samples tracked for analytical analyses. New technology has been applied to develop new applications that can access the LIMS from any web browser on any operating system. The first new applications have been developed for occasional LIMS users who only need to order materials and submit samples.

You will learn:

1. How to use the new ordering system (HTML5 version of StarLIMS: Online Ordering, Receive Orders and Inventory Manager)
2. How to find your analytical samples and look up sample status (a new LIMS application)
3. How to find experiments and related information in the “REDiME Data & File Center”

Tutorial 7. R/Bioconductor

Wednesday 2:30-3:30 PM. Century Center Suites 8-9. *Tutor:* Yury Bukhman

R is a widely used tool and programming environment for data analysis. Bioconductor is a suite of R packages specifically designed for bioinformatics applications.

You will learn: Basics of R and Bioconductor usage

Tutorial 8. Using GLOW (Great Lakes Omics Warehouse) to Track Experiments, Samples, Computational Workflows, and Data

Wednesday 2:30-3:30 PM. Century Center Suites 10-11.

Tutors: David Benton, Darin Kalisak, Oleg Moskvina

GLOW is a database for associating samples, computational analyses, and files for next-generation sequencing experiments (currently RNA-seq and genome resequencing). It is automatically populated with GLBRC sequence data from JGI and the results of in-house transcript mapping.

You will learn:

1. How to register your samples in GLOW
2. How to find experiments, samples, and files in GLOW
3. How to use experimental and derived data from GLOW

Tutorial 9 The GLBRC Genome Suite

Wednesday 7:00-8:00 PM. Century Center Suites 8-9. *Tutor:* Nick Thrower

The Genome Suite is a custom genomic data analysis tool designed around RNA-Seq data analysis. It integrates genome sequence with functional annotation and expression profiling tools.

You will learn:

1. How to open and inspect expression datasets
2. How to view and modify functional annotations
3. How to navigate genomic sequence

Tutorial 10. Pathway Tools

Wednesday 7:00-8:00 PM. Century Center Suites 10-11. *Tutor:* Yury Bukhman

SRI Pathway Tools provide a graphical interface to EcoCyc and other BioCyc databases, enable visualization of metabolic pathways and regulatory networks. They also make it possible to overlay omics data on pathways and networks. Pathway Tools are available through a web browser and as a stand-alone desktop application

You will learn to:

1. Use the GLBRC Pathway Tools server
2. Query a database to find out gene functions
3. Display pathways
4. Visualize omics data
5. Get the desktop version of the software