

# Astrocyte: A Scientific Workflow Platform on the BioHPC Cluster

---

Peng Lian, Ph.D.

BioHPC, Lyda Hill Department of Bioinformatics  
UT Southwestern Medical Center

# Agenda

What is Astrocyte?

How does Astrocyte work?

The Astrocyte web interface

The Astrocyte command-line interface (CLI)

Use case: CHIP-seq analysis workflow

Demonstration: Astrocyte AlphaFold workflow

# What is Astrocyte?

UTSouthwestern  
Medical Center  
Lyda Hill Department of Bioinformatics

BioHPC

Astrocyte 1.0.0

Logged in as: s190450

Astrocyte Home My Projects Browse Workflows Documentation



Welcome to Astrocyte!

Astrocyte is BioHPC's Workflow Platform.  
It provides easy access to workflows developed by different groups at UTSW.

▶ Start a Project

☰ Browse Workflows

## HPC Power Made Easy

When you run a workflow from Astrocyte it's processed on the BioHPC *Nucleus* compute cluster. No HPC knowledge is needed. Astrocyte automatically makes use of the 120+ powerful compute nodes and fast storage to get results to you quickly, and handle huge datasets.

## Workflows by Experts

BioHPC works with groups such as the **BICF, CRI, GCRB** so they can deliver data analysis workflows on Astrocyte. Benefit from the experience of a community of bioinformatics experts, sharing their tools with you.

## Reproduce & Understand

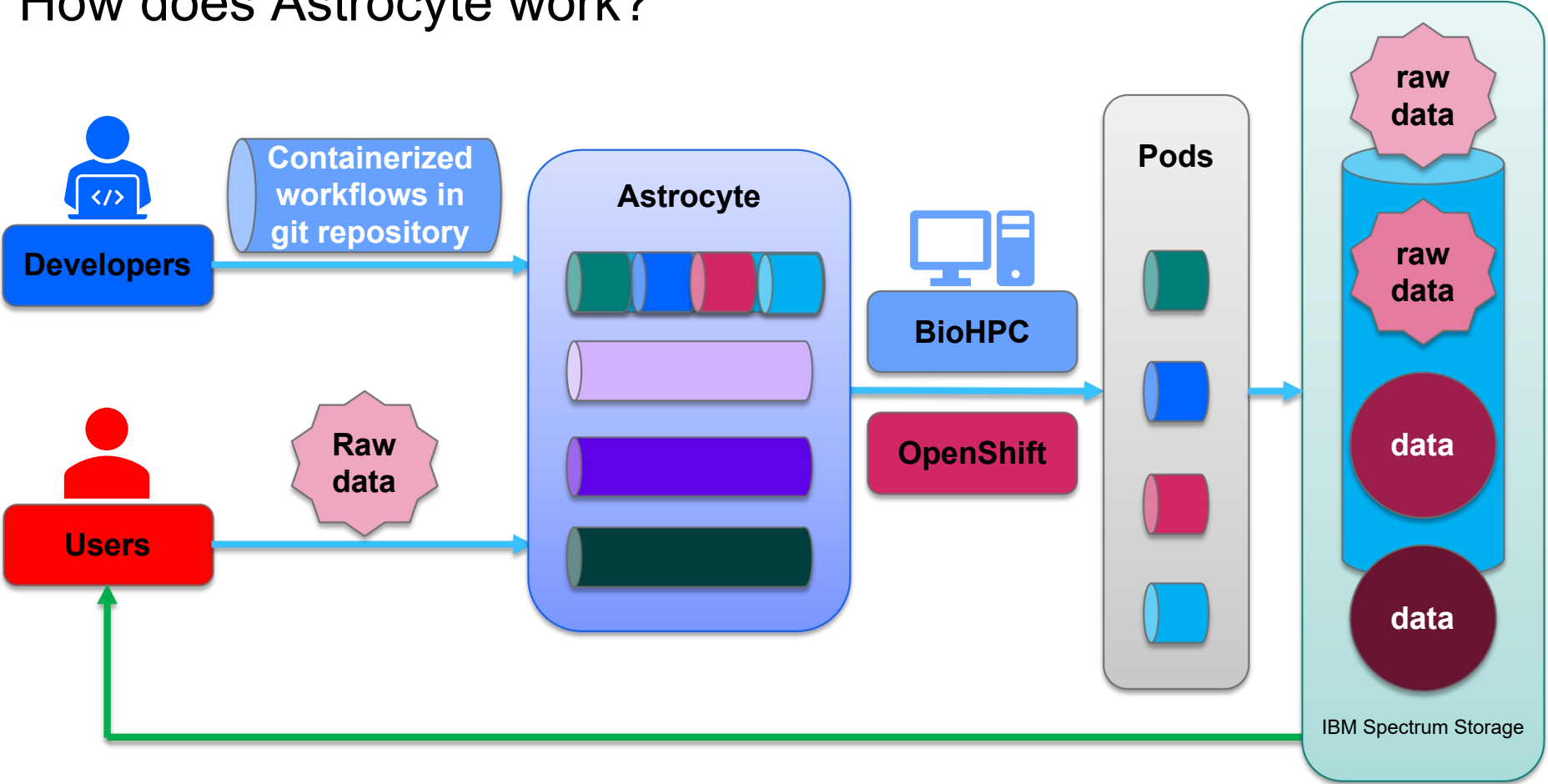
Astrocyte workflows are continually updated, but older versions are always available for long-term studies. Every workflow includes documentation, presented conveniently on the web. Many workflows provide web-based visualization apps, allowing you to easily explore your results.

A platform for scientific workflows

## Features:

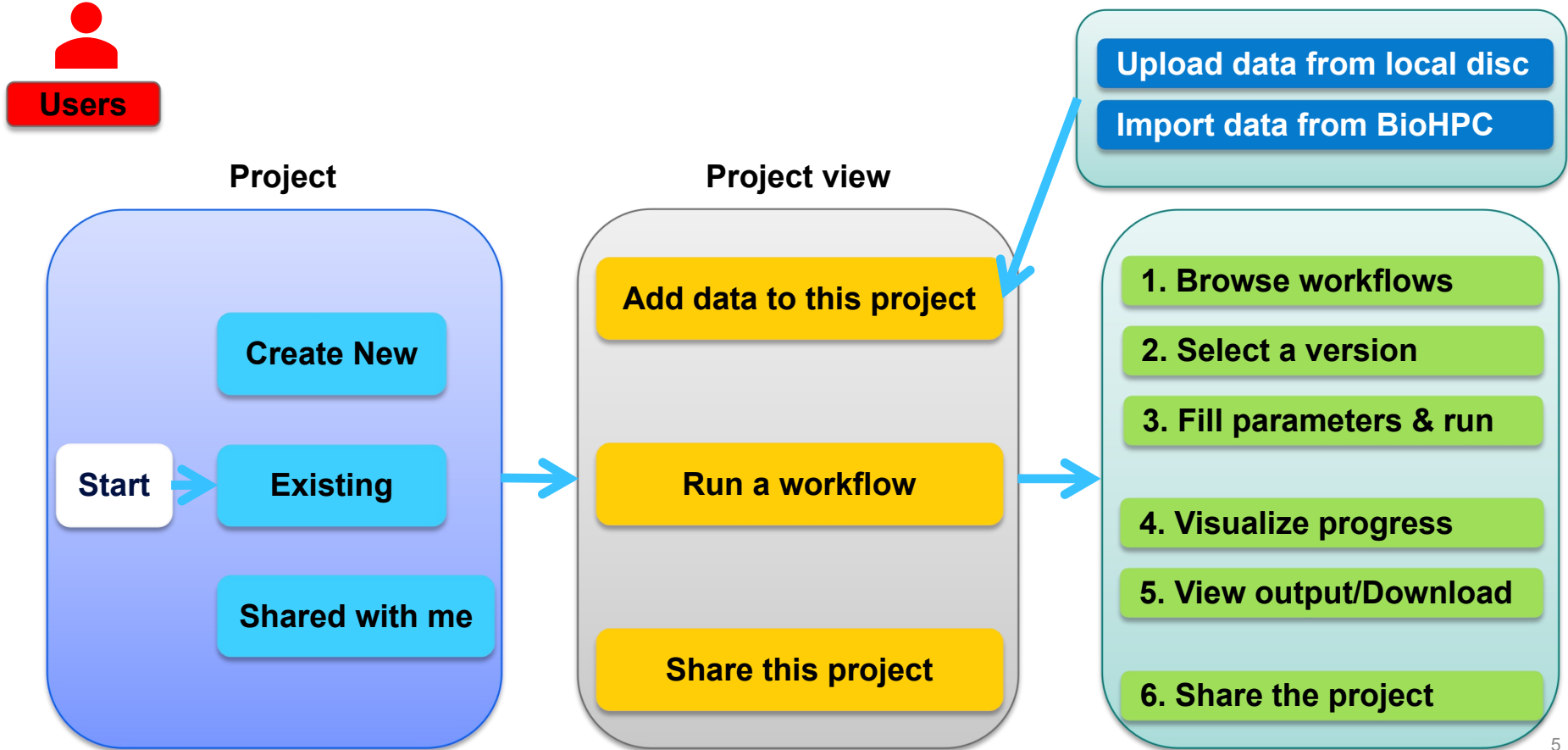
- Standardized workflows
- Good reproducibility
- Easy-to-use web interface
- Online documents
- Result visualization

# How does Astrocyte work?





# The Astrocyte web interface



# The Astrocyte command-line interface (CLI)



Developers

Json Schema

Validate

Convert

YAML

metadata

Validated inputs

check & validate

test run  
run

shiny-prepare  
shiny  
shiny-cluster

ignite-cluster

run on OpenShift

```
workflow_package
├── astrocyte_pkg.yml
├── docs
│   └── index.md
├── test_data
│   └── mobydick.txt
├── vizapp
│   ├── server.R
│   └── ui.R
├── workflow
│   ├── configs
│   └── main.nf
```

# Version control of the workflows



Developers

## workflow repositories

BioHPC Git



### Astrocyte

BICF

CRI

Radiology

Green Center

## Tags for publishing

publish\_2.1.3

test\_1.0.2

**PURPOSE:**  
For publishing/testing

**MAJOR:**  
changes that break the  
parameter structure of  
previous versions

**MINOR:**  
new functionality

**PATCH:**  
small bug fixes

# Astrocyte CLI: the YAML file

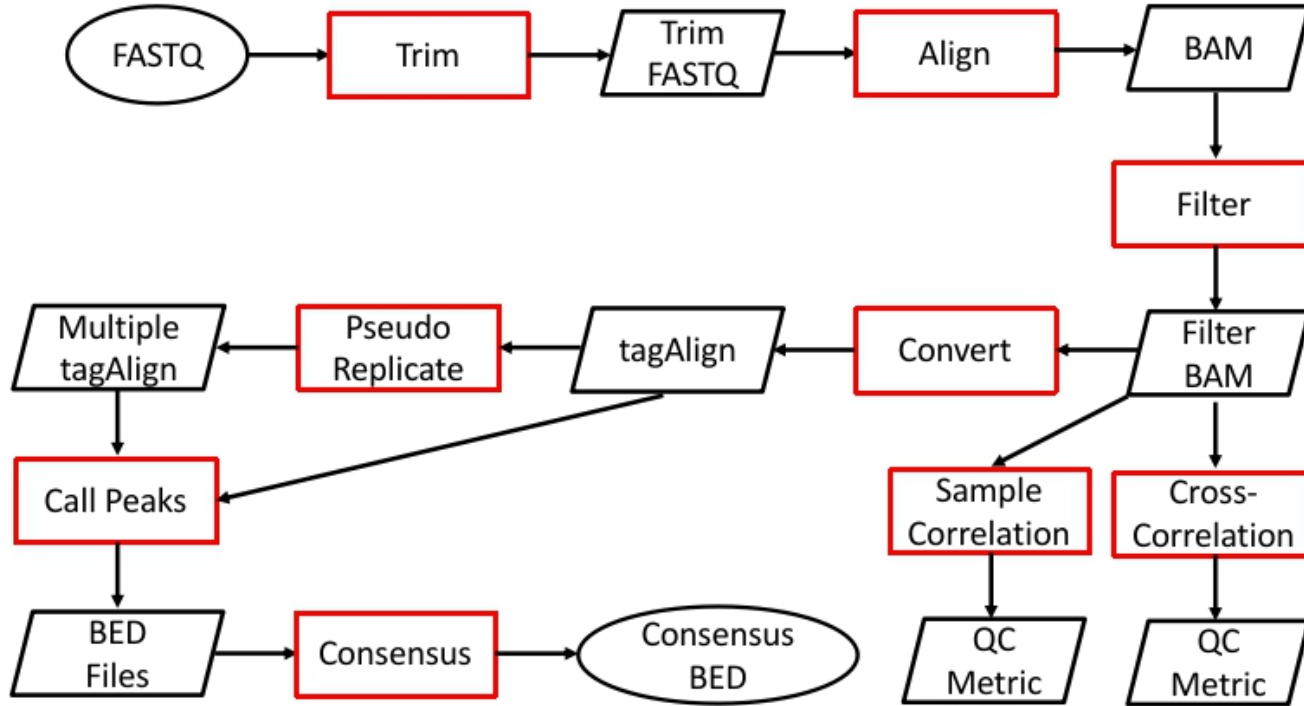


Developers

```
#####  
# BASIC INFORMATION  
#####  
#  
# A unique identifier for the workflow package, text/underscores only  
name: 'example wordcount'  
# Who wrote this?  
author: 'David Trudgian, Peng Lian'  
# A contact email address for questions  
email: 'biohpc@helixpursouthwestern.edu'  
# A more informative title for the workflow package  
title: 'Example Wordcount Workflow'  
# A summary of the workflow package in plain text  
description: |  
  This is a minimal test workflow package that counts the occurrences of words  
  in a test file. It can be used as a template to develop workflows, and as to  
  test the astrocyte platform. Please note that start from publish version  
  0.8.0 and test version 0.6.2, this workflow will only support Astrocyte 0.4.0  
  and above.  
  
#### New Features in Astrocyte 0.4.0 and above ####  
citation: | | You, 12 days ago · Add citation label  
  Please cite individual programs and versions of pipeline  
  used, and the overall pipeline doi: 12.3456/zenodo.9876543.  
  Please cite in publications:  
  Pipeline was developed by BioHPC.  
  
# The minimum Astrocyte version that requires to run this workflow. For old pipelines, which do not have this label  
# a default value of 0.3.1 will be assigned automatically. A request of minimum version less than 0.4.0 will be ignored.  
minimum_astrocyte_version: '0.4.0'  
# The Nextflow version that requires to run this workflow. For old pipelines, which do not have this label  
# a default value of 0.31.0 will be assigned automatically. Please make sure the requested nextflow version is available  
# in the module list.  
nextflow_version: '20.01.0'  
# (Optional) The Nextflow config file to use for this workflow. If provided, the file should exist in workflow/configs  
nextflow_config: 'biohpc.config'  
# The container to use for this workflow, none/singularity. If omitted, the default value 'none' will be used.  
containers: 'singularity'  
# The version of singularity to use. This is required if container == 'singularity'  
singularity_version: '3.5.3'  
  
# -----  
# DOCUMENTATION  
# -----  
#  
# A list of documentation file in .md format that should be viewable from the  
# web interface. These files are in the 'docs' subdirectory. The first file  
# listed will be used as a documentation index and is index.md by convention  
# To supply a title for documentation use a pair of [ 'filename', 'title' ]  
  
documentation_files:  
- [ 'index.md', 'Wordcount Help' ]  
  
# -----  
# NEXTFLOW WORKFLOW CONFIGURATION  
# -----  
# Remember - The workflow file is always named 'workflow/main.f'  
# The workflow must publish all final output into $baseDir  
  
# A list of cluster environment modules that this workflow requires to run.  
# Specify versioned module names to ensure reproducibility.  
workflow_modules:  
- Test
```

```
workflow parameters:  
  
- id: story  
  type: files  
  required: true  
  description: |  
    A text file containing a story  
  regex: '-(txt)$'  
  min: 1  
  
- id: test_int  
  type: integer  
  required: true  
  default: 123  
  min: 1  
  max: 1000  
  description: |  
    This is an example integer field, it is not used by the workflow  
  
- id: test_real  
  type: real  
  required: true  
  default: 123.456  
  min: 1  
  max: 1000  
  description: |  
    This is an example real field, it is not used by the workflow  
  
- id: test_string  
  type: string  
  required: true  
  default: 'Example'  
  description: |  
    This is an example string field, it is not used by the workflow  
  
- id: test_select  
  type: select  
  required: true  
  default: 'Dog'  
  description: |  
    This is an example select field, it is not used by the workflow  
  choices:  
  - [ 'Cat', 'Cat, Kitten' ]  
  - [ 'Dog', 'Dog, Puppy' ]  
  - [ 'Horse', 'Horse, Foal' ]  
  
- id: test_multiselect  
  type: multiselect  
  required: true  
  default: 'Banana'  
  description: |  
    This is an example multiselect field, it is not used by the workflow  
  choices:  
  - [ 'Apple', 'Green Apple' ]  
  - [ 'Banana', 'Yellow Banana' ]  
  - [ 'Strawberry', 'Red Strawberry' ]  
  
# -----  
# SHINY APP CONFIGURATION  
# -----  
# Remember - The vizapp is always 'vizapp/server.R' 'vizapp/ui.R'  
# The workflow must publish all final output into $baseDir  
  
# Name of the R module that the vizapp will run against  
vizapp_r_module: 'R/3.4.1-gcckl'  
  
# List of any CRAN packages, not provided by the modules, that must be made  
# available to the vizapp  
vizapp_cran_packages:  
- shiny  
- shinyFiles
```

# Use case I: ChIP-seq analysis workflow



# Astrocyte web: create a project

## My Projects

In Astrocyte **projects** are used to organize your work. You upload **input data** into a project, and can then run **workflows** against this input data. Try to separate your work into natural projects, so that you can easily share them with other users if required.

+ Start a New Project

Create New Project

Existing Projects

ID	Name	Created	Workflows Run	Input Files	Size	Actions
1956	<a href="#">biohpc_chipseq</a>	Feb. 22, 2022, 12:41 p.m.	2	1	775.7 MB	 Delete
1905	<a href="#">alphafold</a>	Oct. 3, 2021, 10:45 p.m.	3	3	394.8 MB	 Delete
1885	<a href="#">chip_seq</a>	Aug. 3, 2021, 11:06 p.m.	1	10	389.4 GB	 Delete
1882	<a href="#">rshiny</a>	July 14, 2021, 11:19 a.m.	3	4	484.0 MB	 Delete
1716	<a href="#">wordcount</a>	June 2, 2020, 2:26 p.m.	12	6	347.0 MB	 Delete

Projects Shared with Me

ID	Name	Created	Workflows Run	Input Files	Size	Actions
PRJ1909	<a href="#">Liver</a>	Oct. 18, 2021, 2:49 p.m. by Xun WANG	8	9	11.6 GB	
PRJ1779	<a href="#">ERX_41_SUM159_WT_and_KO_4h</a>	Sept. 23, 2020, 10:08 a.m. by Xihui Liu	23	9	1.8 TB	



# Astrocyte web: add data to the project



**Users**

## Project 1956 - biohpc\_chipseq

Owner: Peng Lian (s190450)

Created: Feb. 22, 2022, 12:41 p.m. by Peng Lian (s190450)

### Input data in this project

To run a workflow against input data you need to upload it into this project. Click the button below to add new files from your web browser or the BioHPC cluster. You can also download or delete existing files from the project in the list below.

[Add Data To This Project](#)

Filename	Size	Actions
G1E_ER4_CTCT_chr9.fastq	29.2 MB	<a href="#">Download</a> <a href="#">Delete</a>

### Workflows run in this project

Astrocyte provides many workflow created by different groups at UTSW for you to run against your data. To begin, make sure you have added input data into your project and then click the 'Run a workflow' button to choose a workflow to run.

[Run a workflow in this project](#)

ID	Name	Date	Workflow	Status	Size	Actions
3804	test_vizapp	Feb. 22, 2022, 1:16 p.m.	Astrocyte Example ChIPSeq Workflow Version test_0.0.3	Completed successfully	366.4 MB	<a href="#">Delete</a>
3808	test_vizapp	Feb. 22, 2022, 1:42 p.m.	Astrocyte Example ChIPSeq Workflow Version publish_0.0.8	Completed successfully	380.2 MB	<a href="#">Delete</a>

### Sharing











You can share access to this project with another user. Anyone you share the project with will be able to upload/remove data, run and remove workflows, but cannot delete the project itself. You can only share with a user that has logged into Astrocyte at least once.

[Share With User](#)

# Astrocyte web: choose a workflow to run



## Available Workflows

Sort by:		Popularity	Name (a-z)	Name (z-a)
	BICF	<b>BICF RNAseq Analysis Workflow</b> This is a workflow package for the BioHPC/BICF RNAseq workflow system. It implements differential expression analysis, gene set enrichment analysis, gene fusion analysis and variant identification using RNAseq data.	<b>Current Version:</b> rnaseq_bicf - 2.0.0 <b>Author:</b> Brandi Cantarel <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BICF	<b>BICF ChIP-seq Analysis Workflow</b> This is a workflow package for the BioHPC/BICF ChIP-seq workflow system. It implements ChIP-seq analysis workflow and visualization application.	<b>Current Version:</b> chipseq_analysis_bicf - 1.1.3 <b>Author:</b> Holly Ruess, Spencer D. Barnes, Jeremy A. Mathews, Beibei Chen and Venkat Malladi <b>Contact:</b> <a href="mailto:bicf@utsouthwestern.edu">bicf@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	WBMF	<b>aMAP Pipeline execution Workflow</b> This workflow automates various steps involved in Image registration and segmentation process. The input is stitched tif image files from tissuecyte microscopes and probability maps from ilastik training. The workflow performs the task of converting tif to nifty files and then registering them to the reference atlas. The output nii files are converted back to tif stack.	<b>Current Version:</b> aMAP_Workflow - 0.0.8 <b>Author:</b> Apoorva Ajay <b>Contact:</b> <a href="mailto:Apoorva.Ajay@utsouthwestern.edu">Apoorva.Ajay@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BioHPC	<b>Example Wordcount Workflow</b> This is a minimal test workflow package that counts the occurrences of words in a test file. It can be used as a template to develop workflows, and as to test the astrocyte platform. Please note that start from publish version 0.0.8 and test version 0.0.2, this workflow will only support Astrocyte 0.4.1 and above.	<b>Current Version:</b> example_wordcount - 0.0.9 <b>Author:</b> David Trudgian, Peng Lian, Devin O'Kelly <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BICF	<b>BICF SRA Download Pipeline</b> This is a workflow package for the BioHPC/BICF SRA download workflow system. It implements the SRA Download workflow.	<b>Current Version:</b> sra_pipeline_bicf - 1.1.0 <b>Author:</b> Jon Gesell, Jeremy Mathews <b>Contact:</b> <a href="mailto:bicf@utsouthwestern.edu">bicf@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BioHPC	<b>Astrocyte Example ChIPseq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPseq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.	<b>Current Version:</b> astrocyte_example - 0.0.7 <b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
		<b>Normalized data profiling and visualization</b> This workflow uses R and Cluster to visualize the profile of a tabular dataset such as RNA-seq data, microarray data, proteomic data, and so on. We assume that the data have been normalized. If not, the workflow can be used to visualize some features of the original data but functions such as clustering and heatmap might be much less meaningful.	<b>Current Version:</b> Normalized_data_profiling - 0.0.64 <b>Author:</b> Liang Shi <b>Contact:</b> <a href="mailto:liang.shi@utsouthwestern.edu">liang.shi@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BICF	<b>BICF Germline Variant Analysis Workflow</b> This is a workflow package for the BioHPC/BICF Germline Variant workflow system. It implements a simple germline variant analysis workflow using TrimGalore, BWA, Speedseq, GATK, Samtools and Pindel. SNPs and Indels are integrated using BAYSIC, then annotated using SnpEff and SnpSift.	<b>Current Version:</b> germline_bicf - 1.0.0 <b>Author:</b> Brandi Cantarel <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BICF	<b>BICF ATAC-seq Analysis Workflow</b> This is a workflow package for the BICF ATAC-seq workflow system. It implements ATAC-seq analysis workflow and visualization application.	<b>Current Version:</b> atacseq_analysis_bicf - 2.0.4 <b>Author:</b> Holly Ruess, Spencer Barnes, and Venkat Malladi <b>Contact:</b> <a href="mailto:bicf@utsouthwestern.edu">bicf@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>
	BICF	<b>BICF CellRanger count Workflow</b> This is a workflow package for the BICF/Strand Lab CellRanger count workflow system. It implements 10x CellRanger count analysis workflow application.	<b>Current Version:</b> cellranger_count - 2.2.0 <b>Author:</b> Gervaise H. Henry, Jeremy Mathews, Jon Gesell, and Venkat Malladi <b>Contact:</b> <a href="mailto:bicf@utsouthwestern.edu">bicf@utsouthwestern.edu</a>	<a href="#">▶ Run Workflow</a> <a href="#">📄 Documentation</a> <a href="#">🔄 All Versions</a>

Workflows per page: 10 25 50 100



# Astrocyte web: select a version of the workflow



Users

Workflow: biohpc/astrocyte\_example\_chipseq.git

Latest Published Version:	0.0.8
Git Repository	<a href="https://git.biohpc.swmed.edu/biohpc/astrocyte_example_chipseq.git">https://git.biohpc.swmed.edu/biohpc/astrocyte_example_chipseq.git</a> (Link)
Last Synchronized	Feb. 22, 2022, 1:34 p.m.
Created	May 15, 2016, 6:28 p.m.
Updated	Feb. 22, 2022, 1:34 p.m.
<input type="checkbox"/> Check for Updates	Synchronization status: SUCCESS Version <b>publish_0.0.8</b> status: READY

BioHPC

This workflow provided by:

[UTSW BioHPC](#)

## Published Versions

Version	Git Tag
astrocyte_example - 0.0.8	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application. Please note that start from publish version 0.0.8 and test version 0.0.1, this workflow will only support Astrocyte 0.4.1 and above.</p> <p><b>Author:</b> David Trudgian, Peng Lian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.7	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.</p> <p><b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.6	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.</p> <p><b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>

astrocyte_example - 0.0.3	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.</p> <p><b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.2	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.</p> <p><b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.1	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application.</p> <p><b>Author:</b> David Trudgian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>

## Test Versions

Version	Git Tag
astrocyte_example - 0.0.3 (test)	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application. Please note that start from publish version 0.0.8 and test version 0.0.1, this workflow will only support Astrocyte 0.4.1 and above.</p> <p><b>Author:</b> David Trudgian, Peng Lian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.2 (test)	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application. Please note that start from publish version 0.0.8 and test version 0.0.1, this workflow will only support Astrocyte 0.4.1 and above.</p> <p><b>Author:</b> David Trudgian, Peng Lian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>
astrocyte_example - 0.0.1 (test)	<p><b>Astrocyte Example ChIPSeq Workflow</b> This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application. Please note that start from publish version 0.0.8 and test version 0.0.1, this workflow will only support Astrocyte 0.4.1 and above.</p> <p><b>Author:</b> David Trudgian, Peng Lian <b>Contact:</b> <a href="mailto:biohpc-help@utsouthwestern.edu">biohpc-help@utsouthwestern.edu</a></p> <p><a href="#">▶ Run this Version</a> <a href="#">📄 Documentation</a> <a href="#">⚙ Developer Information</a></p>

# Astrocyte web: fill out parameter forms



Users

## Running Workflow biohpc/astrocyte\_example\_chipseq.git (0.0.3)

### Astrocyte Example ChIPSeq Workflow

This is an example workflow package for the BioHPC astrocyte workflow system. It implements a simple ChIPSeq analysis workflow using BWA and MACS, plus a simple R Shiny visualization application. Please note that start from publish version 0.0.8 and test version 0.0.1, this workflow will only support Astrocyte 0.4.1 and above.

 BioHPC

This workflow provided by:  


 This workflow will run using Nextflow and SLURM batch jobs.

### Parameters

Project  
Project 1882: rshiny

Name for this run  
test\_chipseq

One or more input FASTQ files from a ChIPSeq experiment (required)  
G1E\_ER4\_CTCT\_chr9.fastq

Reference genome for BWA alignment (required)  
UCSC hg38

Run Workflow

### Documentation

#### ChIPSeq Example

### Astrocyte ChIPSeq Example

This workflow carries out a simple ChIPSeq alignment and peak calling using BWA and MACS 1.4. One or more FASTQ files containing reads from a ChIPSeq experiment can be selected as input. For each file this workflow:

1. Aligns the reads to a selected genomic reference using BWA aln.
2. Converts BWA's native output into SAM format.
3. Sorts and indexes the SAM file, and converts into binary BAM format using Picard.
4. Performs ChIPSeq peak calling using MACS 1.4, with simple `--no-model` and `--single-profile` options. Wig files are produced as well as standard spreadsheet output.

### Workflow Parameters

- **fastq** - Choose one or more ChIPSeq read files to process. All should be CHIP files - i.e. there is no control. Each file will be processed as an independent sample.
- **index** - Choose a genomic index to use as a reference for alignment of ChIPSeq reads. A variety of options are available for human and murine samples.

### Visualization App

The example visualization app demonstrates integration of Shiny into astrocyte by implementing a simple file chooser that access the output of the workflow.

### Test Data

The test data directory of this workflow package includes a subset of reads from Chr19 for a CTCF ChIP in a G1E cell line.

Originally made available as example data for the Galaxy ChIP-Seq exercises at <https://usegalaxy.org/u/james/p/exercise-chip-seq>

### Credits

This example workflow is derived from original scripts kindly contributed by the Xu lab, Children's Research Institute at UT Southwestern.

# Astrocyte web: submit and wait for the result of a run



Users

Project 1956 - biohpc\_chipseq → Run 3808 - test\_vizapp

### Run Information

Running Workflow	Astrocyte Example ChIPSeq Workflow biohpc/astrocyte_example_chipseq/g1 / 0.0.8
Status	COMPLETE
Created	Feb. 22, 2022, 1:42 p.m. by s190450
Size	380.2 MB <a href="#">Delete working files</a>

### Parameters

Parameter	Value
index	/project/apps/database/Genomes/Homo_sapiens/UCSC/hg38/Sequence/BWA/index/genome.fa
fastq	G1E_ER4_CTCT_chr9.fastq

### Input Files

Filename	Size
G1E_ER4_CTCT_chr9.fastq	29.2 MB

### Workflow Output / Visualization

You can **download** an archive file containing all output of the workflow, or **export** it directly to a location on the BioHPC cluster storage for further work. If you wish to use the output file(s) as input for other runs, select to output to the **incoming** directory.

*Note - Mac OSX cannot extract zip files >4GB. A tar file download will be added shortly.*

Download Workflow Output:

Export output:

### Output Browser

Click the "Generate Direct Link" button to obtain a direct web link you can use with external tools, such as the UCSC Browser, that need to access the file directly. These links are valid for 24 hours.

Current Directory: (/)

G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_MACS_wiggle	(5.8 KB)	<a href="#">Generate Direct Link</a>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_summits.bed	(7.0 KB)	<a href="#">Generate Direct Link</a>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_peaks.xls	(5.9 KB)	<a href="#">Generate Direct Link</a>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_peaks.bed	(10.5 MB)	<a href="#">Generate Direct Link</a>

### Workflow Progress

Task Number	Task Name	SLURM Job	Status	Submitted	Duration
1	bwaaln (1)	2897298	COMPLETED	2021-09-02 00:36:56.132	13.7s
2	bwasamse (1)	2897299	COMPLETED	2021-09-02 00:37:09.929	14.9s
3	sam2bam (1)	2897301	COMPLETED	2021-09-02 00:37:24.932	9.9s
4	macs14 (1)	2897302	COMPLETED	2021-09-02 00:37:34.905	2m 50s

### Advanced / Troubleshooting

This section contains links intended for use by BioHPC and workflow creators when troubleshooting problems.

# Astrocyte web: launch the vizapp for online analysis



Workflow Output / Visualization

You can **download** an archive file containing all output of the workflow, or **export** it directly to a location on the BioHPC cluster storage for further work.

If you wish to use the output file(s) as input for other runs, select to output to the **incoming** directory.

*Note - Mac OSX cannot extract zip files >4GB. A tar file download will be added shortly.*

Download Workflow Output:

Export Output:

The **Visualization App** (vizapp) allows you to explore the results of your workflow on the web. Use the buttons below to start/stop and connect to a vizapp session. It takes 30s for the vizapp to start, or longer if there is a queue on the BioHPC cluster. Please stop the vizapp when you are finished using it, as it occupies a slot on the BioHPC cluster.

Vizapp Status: RUNNING   (<http://vnc.biohpc.swmed.edu:4453>)

Output Browser

Click the 'Generate Direct Link' button to obtain a direct web link you can use with external tools, such as the UCSC Browser, that need to access the file directly. These links are valid for 24 hours.

Current Directory: (/)

File Name	Size	Action
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_MACS_wiggle		
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_summits.bed	(5.8 KB)	<input type="button" value="Generate Direct Link"/>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_peaks.xls	(7.0 KB)	<input type="button" value="Generate Direct Link"/>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam_bwa_nomodel_peaks.bed	(5.9 KB)	<input type="button" value="Generate Direct Link"/>
G1E_ER4_CTCT_chr9.fastq.sam.gz.bam	(10.5 MB)	<input type="button" value="Generate Direct Link"/>

Workflow Progress

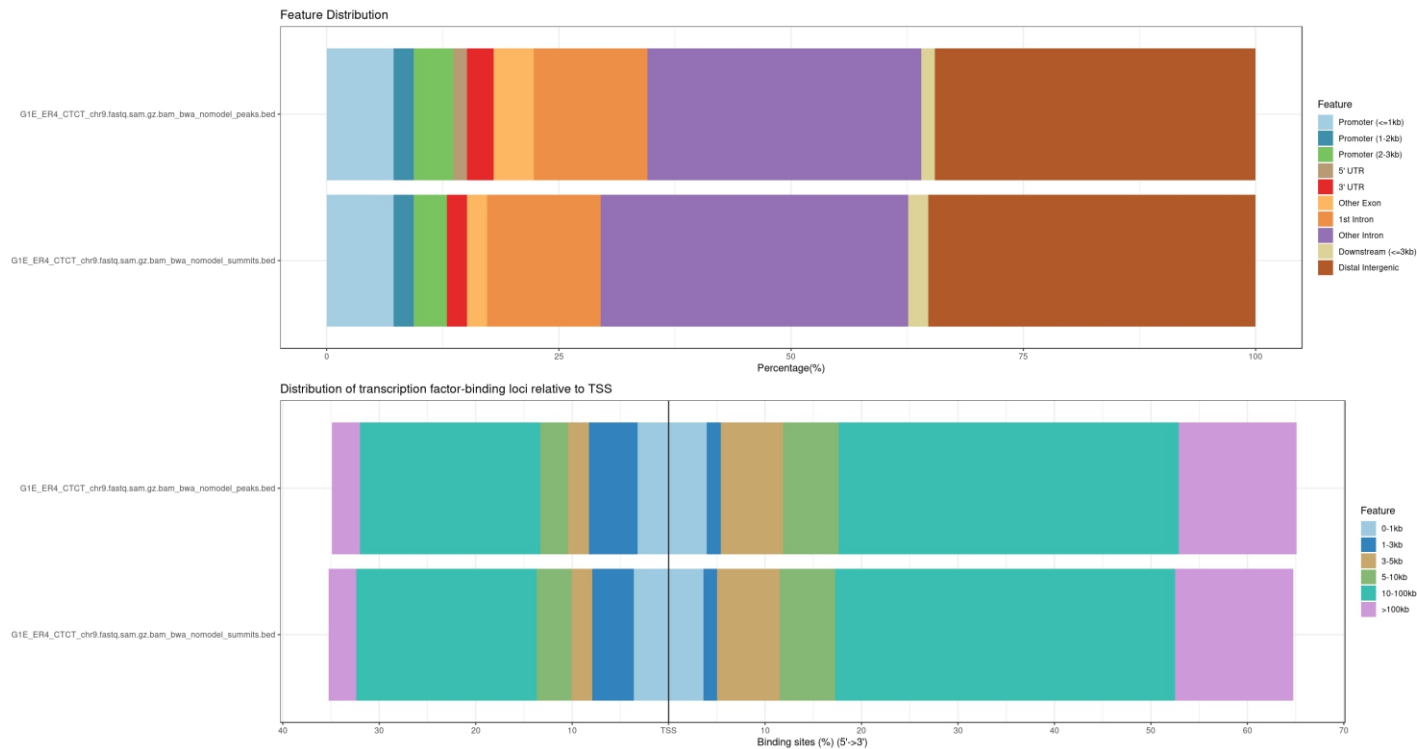
Task Number	Task Name	SLURM Job	Status	Submitted	Duration
1	bwaaln (1)	3099508	COMPLETED	2022-02-22 13:42:45.705	14.7s
2	bwaaln (1)	3099509	COMPLETED	2022-02-22 13:42:45.705	14.7s

# Astrocyte web: visualization of the result



Users

## Astrocyte ChIP-seq Example



# Demonstration: Astrocyte AlphaFold workflow

BioHPC protein viewer

load file <

structure <

surface <

ligand <

selection <

label <

contact <

stage <

snapshot <

**Animation**

None Spin Rock

Sequence

Fullscreen

Sequence +



Showing file: biohpc

The image shows a screenshot of the BioHPC protein viewer interface. The main window displays a 3D protein structure of the letters 'BIOHPC' in a stylized font. The structure is composed of many small, colorful spheres (atoms) connected by thin lines (bonds). The colors range from blue and green to yellow and orange. The letters are arranged in a row, with 'BIOHPC' being the most prominent. The background is dark. On the left side, there is a dark sidebar with various controls. At the top of the sidebar, there is a 'load file' button. Below it are several icons representing different views: 'structure', 'surface', 'ligand', 'selection', 'label', 'contact', 'stage', and 'snapshot'. Under the 'Animation' section, there are three buttons: 'None', 'Spin', and 'Rock'. Below that is a 'Sequence' toggle switch, which is currently turned on. At the bottom of the sidebar is a 'Fullscreen' button. In the top right corner of the main window, there is a 'Sequence' label and a '+' icon. At the bottom of the main window, it says 'Showing file: biohpc'.

# Summary

As a platform for scientific workflows, Astrocyte bridges the knowledge gap between scientists from different fields.

The web interface and the online documentations enable the user who has little HPC knowledge to run scientific workflows easily.

The Astrocyte CLI provides a toolkit for the workflow developers to compose and test their workflows.

There are already many scientific workflows from different institutes of UT Southwestern Medical Center are running on the Astrocyte platform.

# Questions

All questions are welcome!