CRI Bioinformatics Workshop

Bulk-RNAseq Pipeline

Hands-on Training

Center for Research Informatics – Bioinformatics Core

Genomics and proteomics data analysis

BiCF applies appropriate and state-of-the-arts statistical and bioinformatic methodologies to analyze genomics data generated from standard and emerging assays.

Consulting, grant writing and training

BiCF provides consulting services for experimental design or data analysis; grant writing assistance, including bioinformatics development, cost analysis, and documentation of tools to complete the research.

Data management system development

BiCF offers enterprise solutions for project and study management, for data production , sharing and integration .

OUR AWESOME TEAM

Mengjie Chen, PhD

Faculty Director

Wenjun Kang, MS

Technical Director

Yan Li, PhD

Associate Director

Bioinformaticians

Jason Shapiro, PhD

Diana Vera Cruz, PhD

Evan Wu

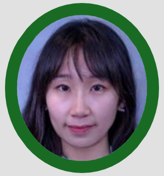
Katie Aracena, PhD

David Tieri, PhD

Houxiang Zhu, PhD

Qiaoshan Lin, PhD

Yildiz Koca, PhD

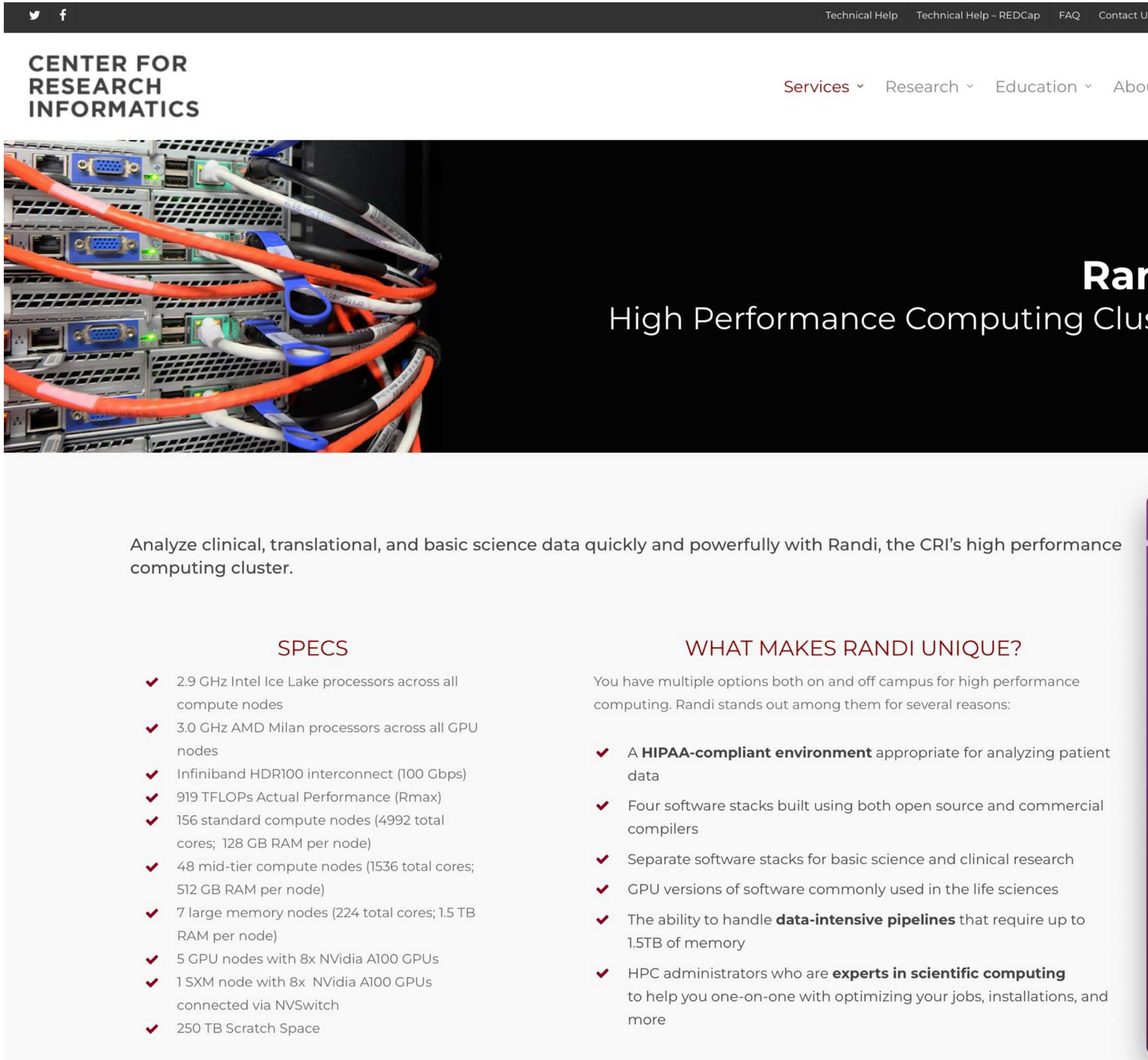


Zhongyu Li, MS

Geetha Priyanka, MS

Contact us: bioinformatics@bsd.uchicago.edu Submit a project request: https://biocore.cri.uchicago.edu/

https://cri.uchicago.edu/hpc/



Randi
High Performance Computing Cluster

Analyze clinical, translational, and basic science data quickly and powerfully with Randi, the CRI's high performance computing cluster.

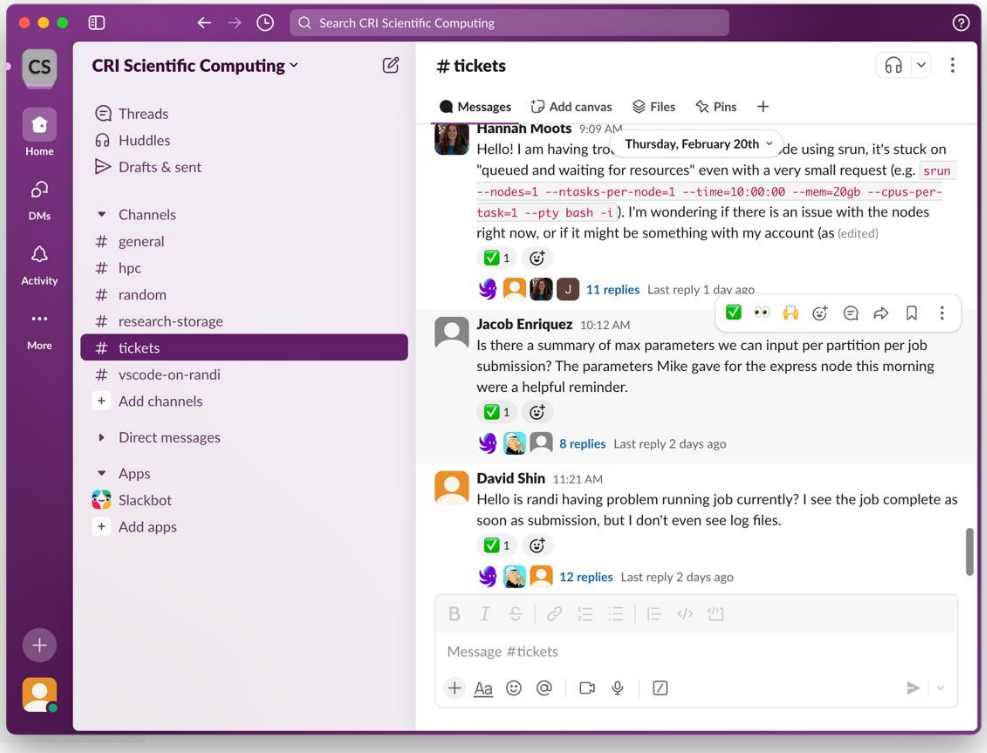
SPECS

- ✓ 2.9 GHz Intel Ice Lake processors across all compute nodes
- ✓ 3.0 GHz AMD Milan processors across all GPU nodes
- ✓ Infiniband HDR100 interconnect (100 Gbps)
- ✓ 919 TFLOPs Actual Performance (Rmax)
- ✓ 156 standard compute nodes (4992 total cores; 128 GB RAM per node)
- ✓ 48 mid-tier compute nodes (1536 total cores; 512 GB RAM per node)
- ✓ 7 large memory nodes (224 total cores; 1.5 TB RAM per node)
- ✓ 5 GPU nodes with 8x NVidia A100 GPUs
- ✓ 1 SXM node with 8x NVidia A100 GPUs connected via NVSwitch
- ✓ 250 TB Scratch Space

WHAT MAKES RANDI UNIQUE?

You have multiple options both on and off campus for high performance computing. Randi stands out among them for several reasons:

- ✓ A **HIPAA-compliant environment** appropriate for analyzing patient data
- ✓ Four software stacks built using both open source and commercial compilers
- ✓ Separate software stacks for basic science and clinical research
- ✓ GPU versions of software commonly used in the life sciences
- ✓ The ability to handle **data-intensive pipelines** that require up to 1.5TB of memory
- ✓ HPC administrators who are **experts in scientific computing** to help you one-on-one with optimizing your jobs, installations, and more



Hannah Moots 9:09 AM
Hello! I am having tro[...]de using srun, it's stuck on "queued and waiting for resources" even with a very small request (e.g. `srun --nodes=1 --ntasks-per-node=1 --time=10:00:00 --mem=20gb --cpus-per-task=1 --pty bash -i`). I'm wondering if there is an issue with the nodes right now, or if it might be something with my account (as (edited)
✓ 1
11 replies Last reply 1 day ago

Jacob Enriquez 10:12 AM
Is there a summary of max parameters we can input per partition per job submission? The parameters Mike gave for the express node this morning were a helpful reminder.
✓ 1
8 replies Last reply 2 days ago

David Shin 11:21 AM
Hello is randi having problem running job currently? I see the job complete as soon as submission, but I don't even see log files.
✓ 1
12 replies Last reply 2 days ago

Center for Research Informatics

Now Offering Live Office Hours for HPC and Bioinformatics Core Services!

- Join us for free live expert support for your high-performance computing (HPC) and bioinformatics research needs.

Office Hours:

Location: Medical Campus Peck Pavillion, N161

High Performance Computing (HPC):

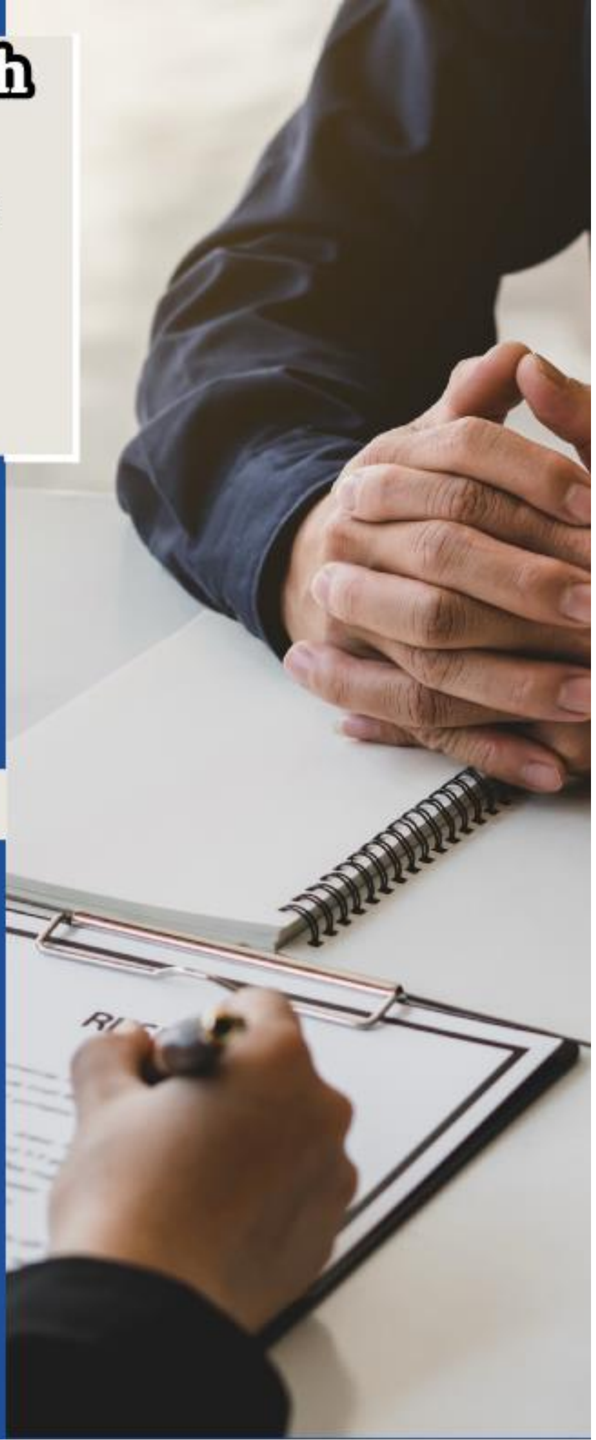
- Every Tuesday | 12:00 PM – 2:30 PM
- Hosted by Michael Jarsulic

Bioinformatics Core:

- Every Tuesday | 12:30 PM – 3:30 PM
- Hosted by Yan Li

Visit our website for more details

cri.uchicago.edu

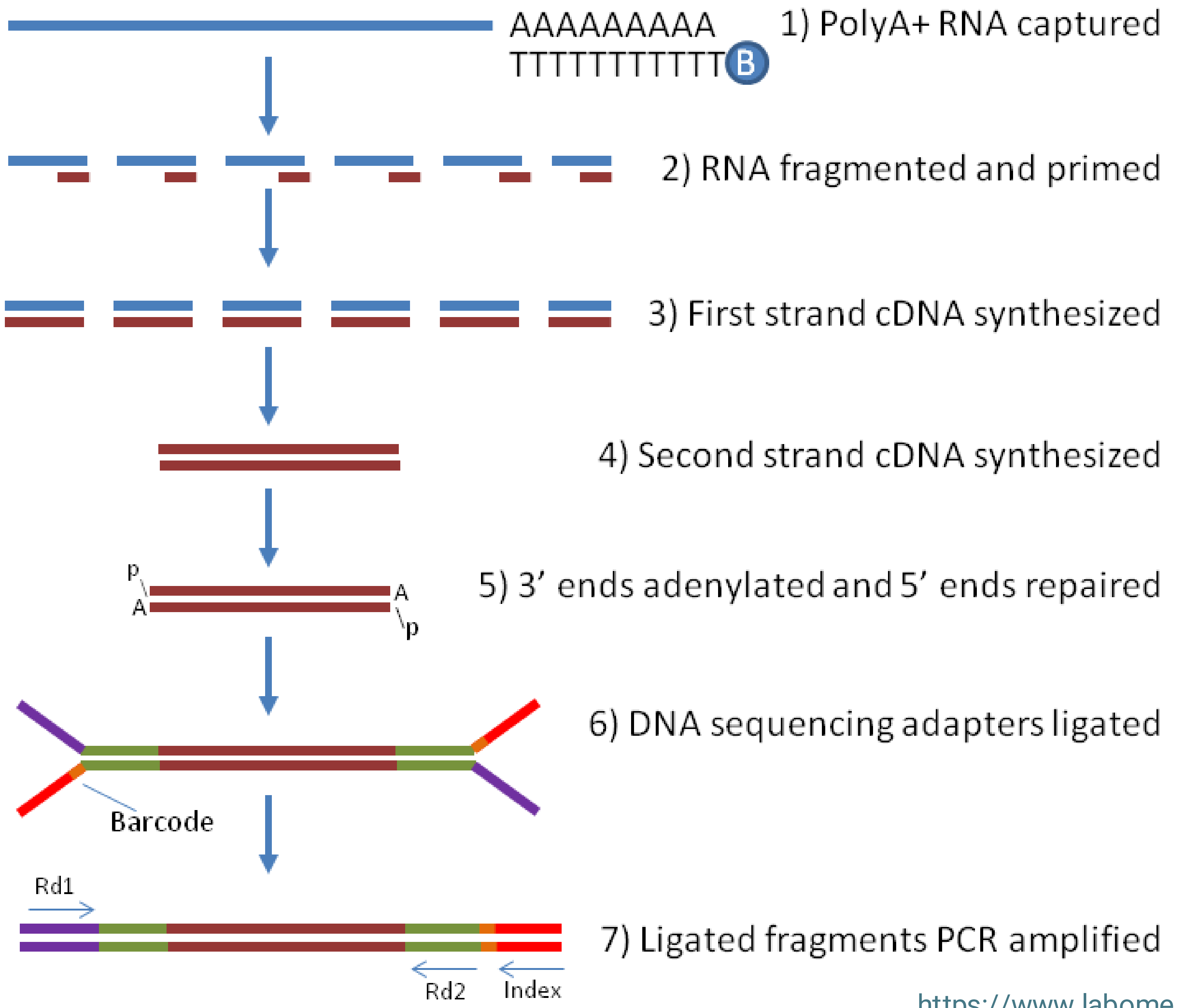


Objectives

- Learn to run Nextflow RNA-Seq pipeline on Randi HPC
- Learn to run our in-house app for differential expression analysis

What is Bulk RNA-Seq?

- Bulk RNA sequencing (**bulk RNA-Seq**) measures **gene expression levels** in a sample by sequencing the **total RNA from a mixture of cells**. Unlike **single-cell RNA-Seq**, bulk RNA-Seq provides an **average expression profile** across all cells in a sample.



1) PolyA+ RNA captured

2) RNA fragmented and primed

3) First strand cDNA synthesized

4) Second strand cDNA synthesized

5) 3’ ends adenylated and 5’ ends repaired

6) DNA sequencing adapters ligated

7) Ligated fragments PCR amplified

https://www.labome.com/method/RNA-seq.html

Biological Questions Bulk RNA-Seq Can Answer

Our focus today

✅ **Differential Gene Expression (DGE)** → Which genes are **upregulated/downregulated** between conditions?

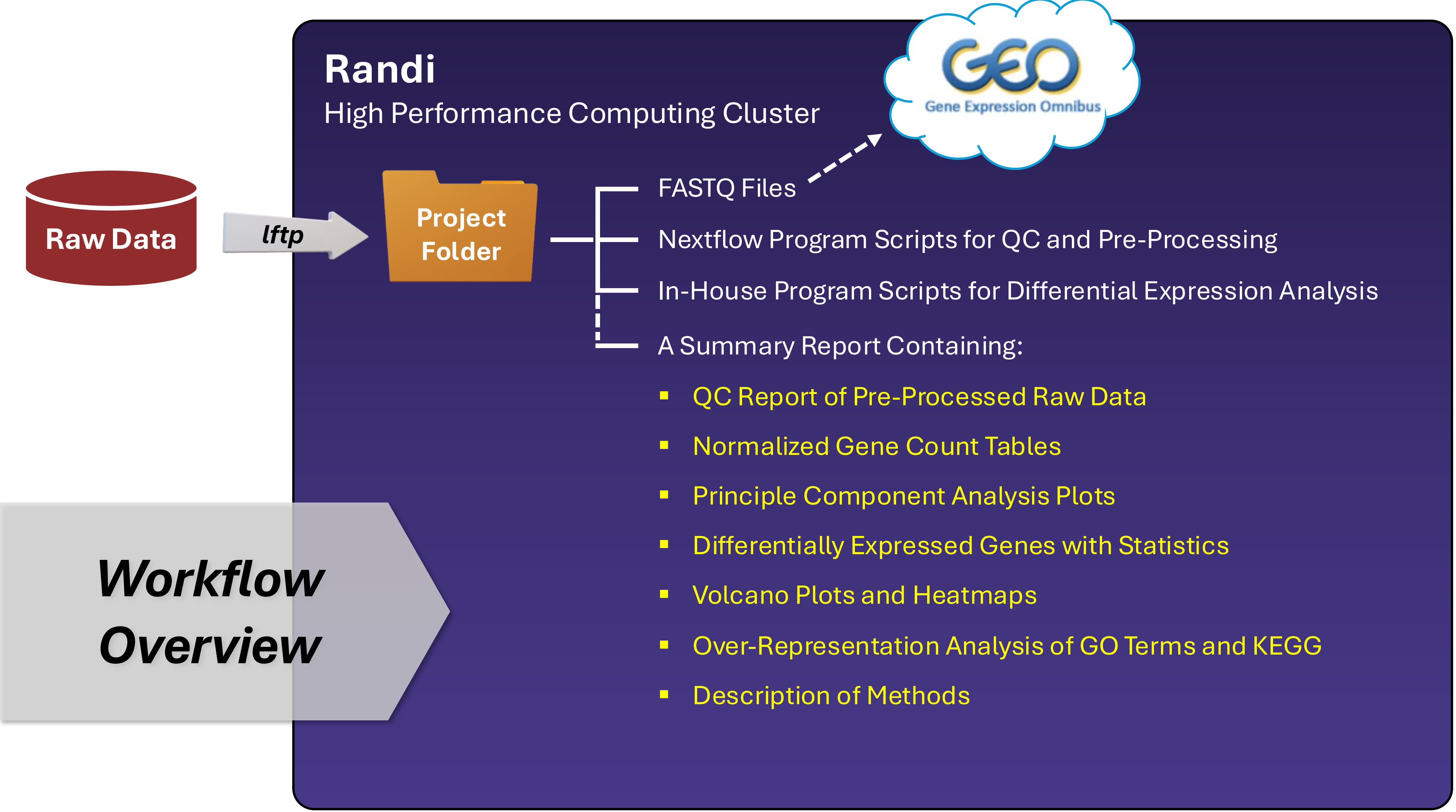
✅ **Pathway & Functional Enrichment** → What **biological processes** are affected? (e.g., gene-set over-representation analysis, GSEA)

✅ **Alternative Splicing & Isoform Analysis** → Are there changes in **splicing patterns**?

✅ **Mutation & Fusion Detection** → Are there **SNPs, RNA editing sites, or fusion transcripts**?

✅ **Cell-Type-Specific Expression (with deconvolution)** → What cell types contribute to gene expression changes?

Workflow Overview



Randi

High Performance Computing Cluster

- FASTQ Files
- Nextflow Program Scripts for QC and Pre-Processing
- In-House Program Scripts for Differential Expression Analysis
- A Summary Report Containing:
  - QC Report of Pre-Processed Raw Data
  - Normalized Gene Count Tables
  - Principle Component Analysis Plots
  - Differentially Expressed Genes with Statistics
  - Volcano Plots and Heatmaps
  - Over-Representation Analysis of GO Terms and KEGG
  - Description of Methods

Agenda & Key Activities

Section 1

- Introduction to the Nextflow RNAseq Pipeline
- Hands-on Practice on Running Nextflow on the Randi Server
- Interpretation of Nextflow Outputs

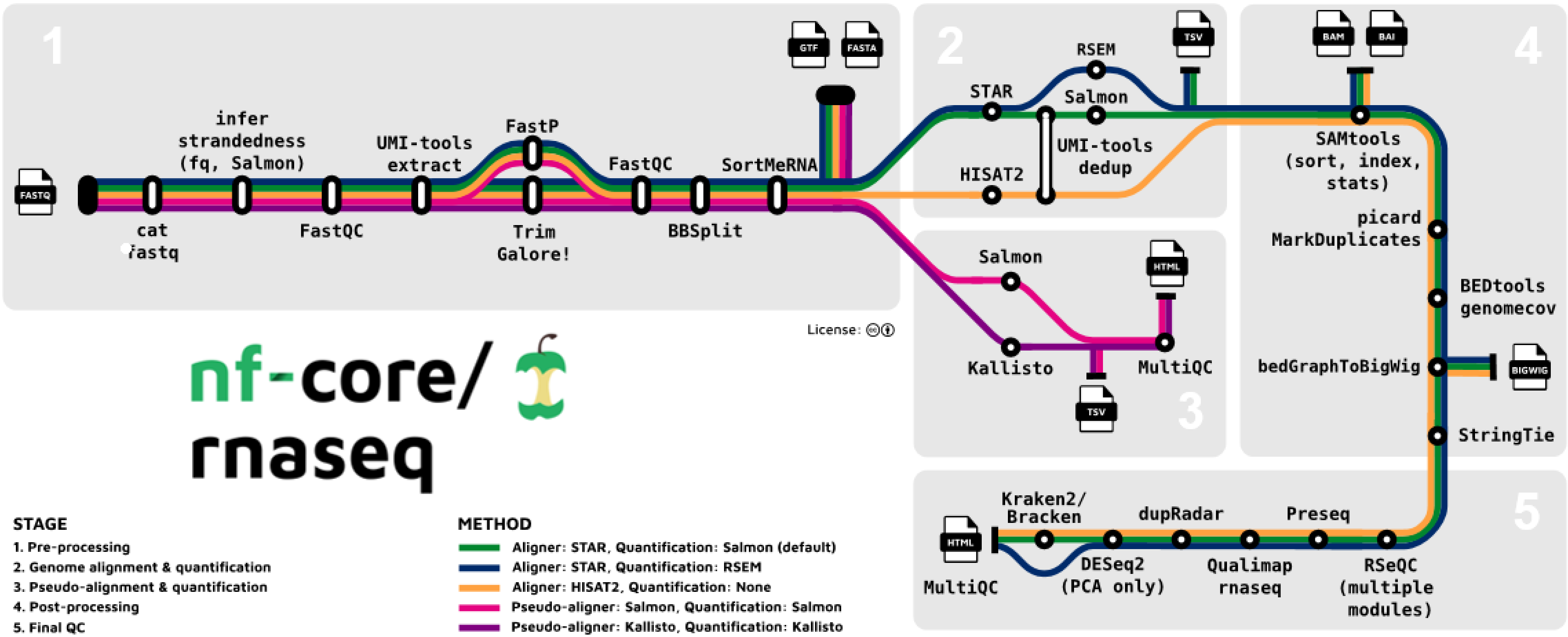
TEN-MINUTES BREAK

Section 2

- Introduction to the DE Analysis Principles
- Demo of the DE Analysis App on Randi
- Hands-on Practice on Running the DE Analysis App
- Interpretation of the DE Analysis Results

Section 1

Nextflow RNAseq Pipeline



nf-core/ rnaseq

STAGE

1. Pre-processing
2. Genome alignment & quantification
3. Pseudo-alignment & quantification
4. Post-processing
5. Final QC

METHOD

- 🟩 Aligner: STAR, Quantification: Salmon (default)
- 🟦 Aligner: STAR, Quantification: RSEM
- 🟧 Aligner: HISAT2, Quantification: None
- 🟥 Pseudo-aligner: Salmon, Quantification: Salmon
- 🟪 Pseudo-aligner: Kallisto, Quantification: Kallisto

1-3 hours

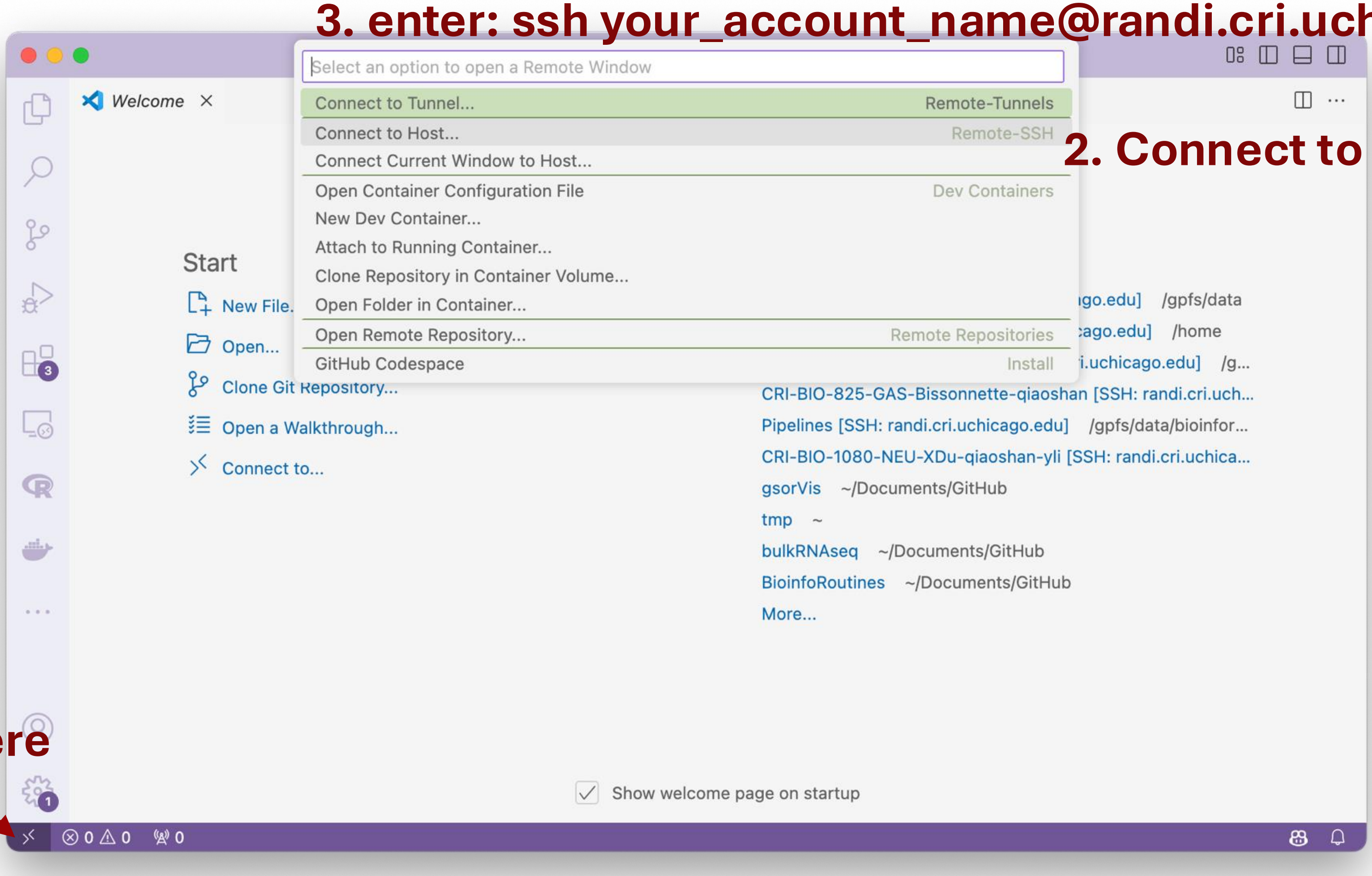
https://nf-co.re/rnaseq/3.18.0/

Run Nextflow RNAseq Pipeline on Randi

Step 1: Log into Randi Terminal / iTerm (MacOS) PuTTY / Xshell (Windows)

Here we do the demo using VSCode since it is very user-friendly and compatible with both systems.

3. enter: ssh your_account_name@randi.cri.uchicago.edu


Welcome
Select an option to open a Remote Window
Connect to Tunnel... Remote-Tunnels
Connect to Host... Remote-SSH
Connect Current Window to Host...
Open Container Configuration File Dev Containers
New Dev Container...
Attach to Running Container...
Clone Repository in Container Volume...
Open Folder in Container...
Open Remote Repository... Remote Repositories
GitHub Codespace Install
Start
New File...
Open...
Clone Git Repository...
Open a Walkthrough...
Connect to...
go.edu] /gpfs/data
ago.edu] /home
i.uchicago.edu] /g...
CRI-BIO-825-GAS-Bissonnette-qiaoshan [SSH: randi.cri.uch...
Pipelines [SSH: randi.cri.uchicago.edu] /gpfs/data/bioinfor...
CRI-BIO-1080-NEU-XDu-qiaoshan-yli [SSH: randi.cri.uchica...
gsorVis ~/Documents/GitHub
tmp ~
bulkRNAseq ~/Documents/GitHub
BioinfoRoutines ~/Documents/GitHub
More...
Show welcome page on startup
0 0 0

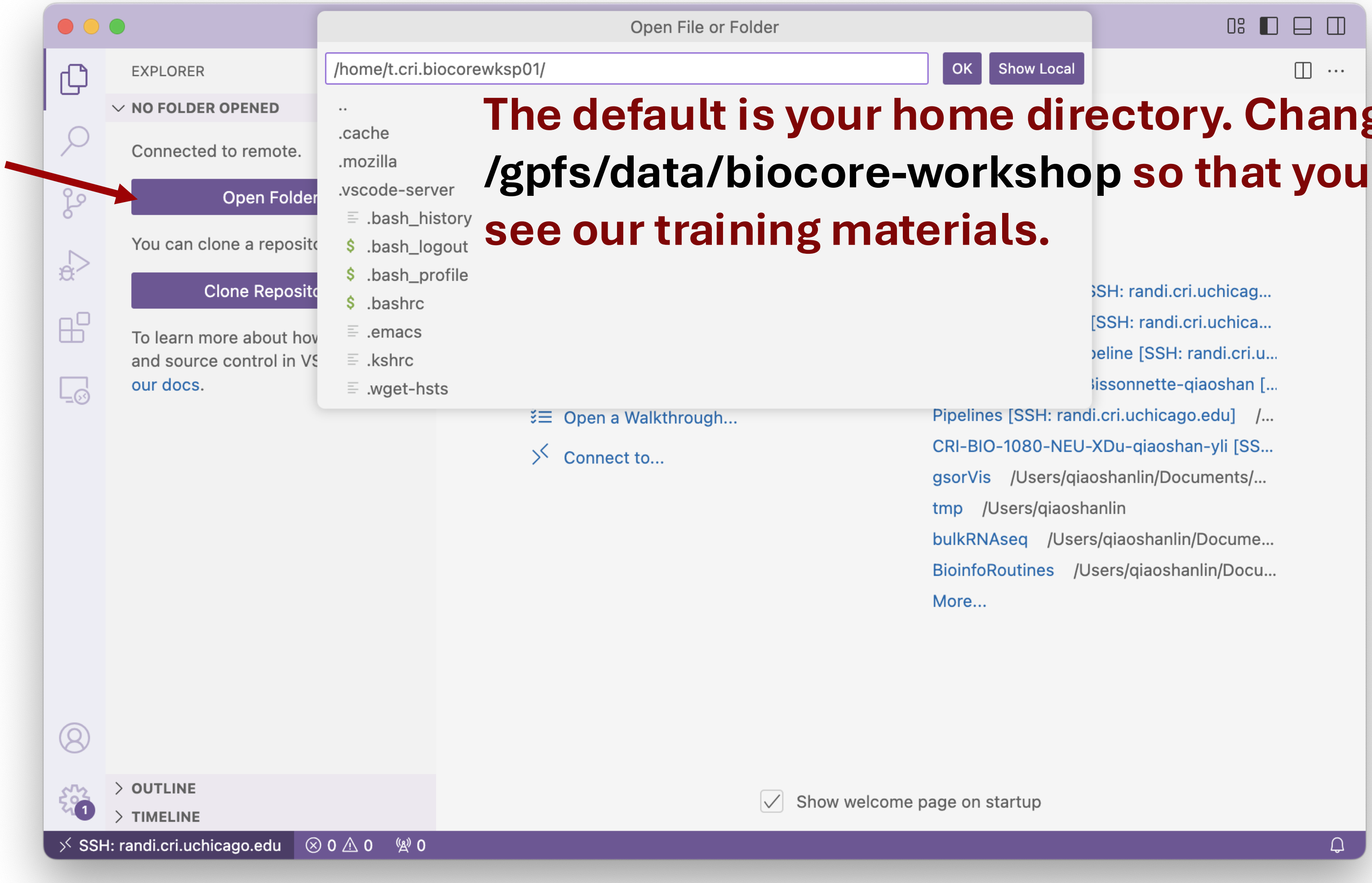
2. Connect to Host

1. Click here

Run Nextflow RNAseq Pipeline on Randi

Step 2: Navigate to the biocore-workshop folder

Open Folder

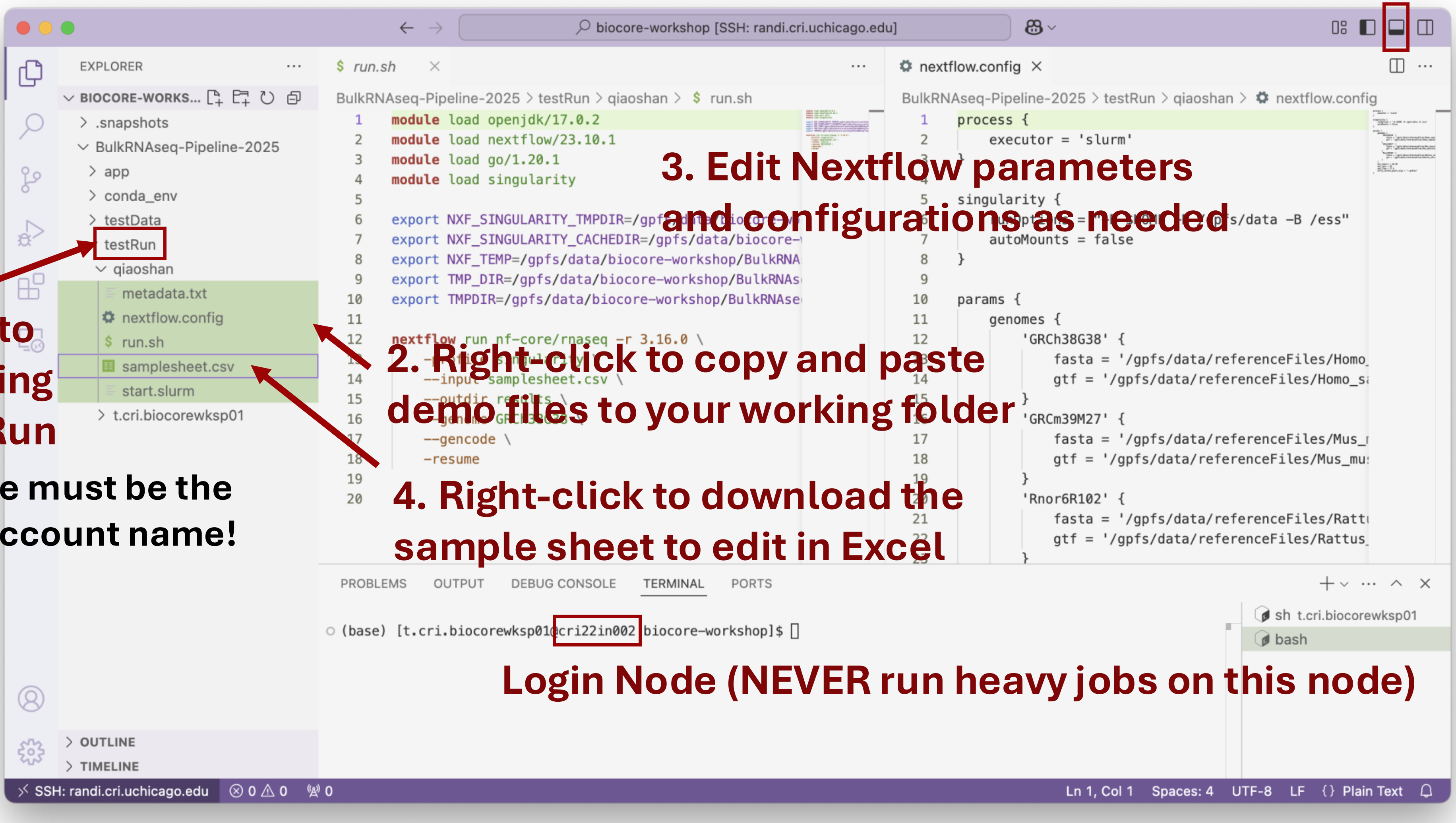


Open File or Folder
/home/t.cri.biocorewksp01/
OK
Show Local
..
.cache
.mozilla
.vscode-server
.bash_history
.bash_logout
.bash_profile
.bashrc
.emacs
.kshrc
.wget-hsts
EXPLORER
NO FOLDER OPENED
Connected to remote.
Open Folder
You can clone a reposito
Clone Reposito
To learn more about hov and source control in VS our docs.
Open a Walkthrough...
Connect to...
SSH: randi.cri.uchicag...
[SSH: randi.cri.uchica...
peline [SSH: randi.cri.u...
bissonnette-qiaoshan [...
Pipelines [SSH: randi.cri.uchicago.edu] /...
CRI-BIO-1080-NEU-XDu-qiaoshan-yli [SS...
gsorVis /Users/qiaoshanlin/Documents/...
tmp /Users/qiaoshanlin
bulkRNAseq /Users/qiaoshanlin/Docume...
BioinfoRoutines /Users/qiaoshanlin/Docu...
More...
Show welcome page on startup
OUTLINE
TIMELINE
SSH: randi.cri.uchicago.edu
0 0 0

The default is your home directory. Change it to /gpfs/data/biocore-workshop so that you can see our training materials.

Run Nextflow RNAseq Pipeline on Randi

Step 3: Set up the Nextflow pipeline



```
module load openjdk/17.0.2
module load nextflow/23.10.1
module load go/1.20.1
module load singularity

export NXF_SINGULARITY_TMPDIR=/gpfs/data/biocore-w
export NXF_SINGULARITY_CACHEDIR=/gpfs/data/biocore-
export NXF_TEMP=/gpfs/data/biocore-workshop/BulkRNA
export TMP_DIR=/gpfs/data/biocore-workshop/BulkRNAs
export TMPDIR=/gpfs/data/biocore-workshop/BulkRNAse

nextflow run nf-core/rnaseq -r 3.16.0 \
    -profile singularity \
    --input samplesheet.csv \
    --outdir results \
    --genome GRCh38G38 \
    --gencode \
    -resume
```

```
process {
    executor = 'slurm'
}

singularity {
    runOptions = "-B $HOME -B /gpfs/data -B /ess"
    autoMounts = false
}

params {
    genomes {
        'GRCh38G38' {
            fasta = '/gpfs/data/referenceFiles/Homo_
            gtf = '/gpfs/data/referenceFiles/Homo_sa
        }
        'GRCm39M27' {
            fasta = '/gpfs/data/referenceFiles/Mus_r
            gtf = '/gpfs/data/referenceFiles/Mus_mus
        }
        'Rnor6R102' {
            fasta = '/gpfs/data/referenceFiles/Rattu
            gtf = '/gpfs/data/referenceFiles/Rattus_
        }
```

```
(base) [t.cri.biocorewksp01@cri22in002 biocore-workshop]$
```

1. Right-click to create a working folder in testRun

The folder name must be the same as your account name!

2. Right-click to copy and paste demo files to your working folder

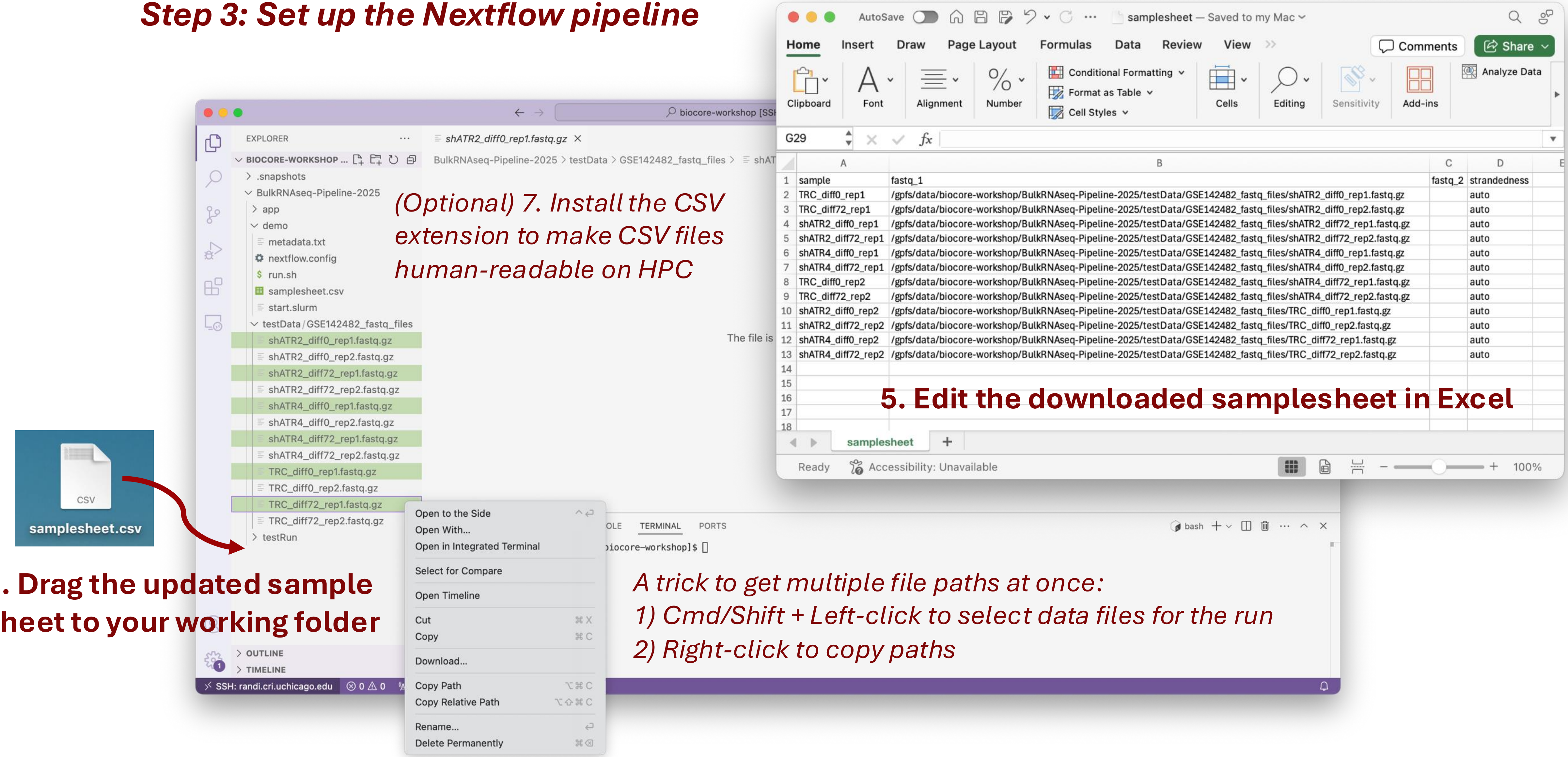
3. Edit Nextflow parameters and configurations as needed

4. Right-click to download the sample sheet to edit in Excel

Login Node (NEVER run heavy jobs on this node)

Run Nextflow RNAseq Pipeline on Randi

Step 3: Set up the Nextflow pipeline



(Optional) 7. Install the CSV extension to make CSV files human-readable on HPC

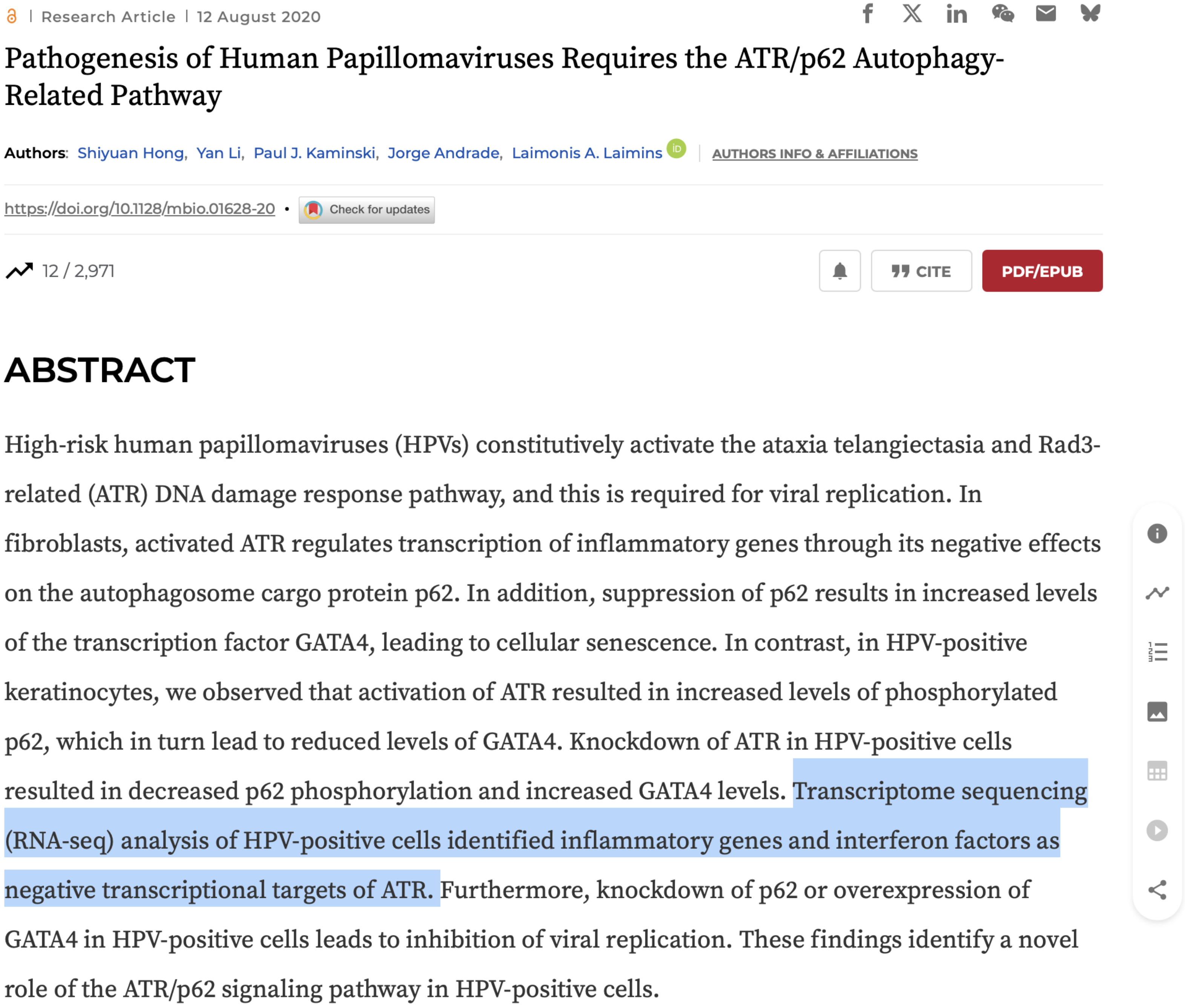
5. Edit the downloaded samplesheet in Excel

sample	fastq_1	fastq_2	strandedness
TRC_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff0_rep1.fastq.gz		auto
TRC_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff0_rep2.fastq.gz		auto
shATR2_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff72_rep1.fastq.gz		auto
shATR2_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff72_rep2.fastq.gz		auto
shATR4_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff0_rep1.fastq.gz		auto
shATR4_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff0_rep2.fastq.gz		auto
TRC_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff72_rep1.fastq.gz		auto
TRC_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff72_rep2.fastq.gz		auto
shATR2_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff0_rep1.fastq.gz		auto
shATR2_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff0_rep2.fastq.gz		auto
shATR4_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff72_rep1.fastq.gz		auto
shATR4_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff72_rep2.fastq.gz		auto

6. Drag the updated sample sheet to your working folder

A trick to get multiple file paths at once:
1) Cmd/Shift + Left-click to select data files for the run
2) Right-click to copy paths

Test Data (GSE142482)



Research Article | 12 August 2020

Pathogenesis of Human Papillomaviruses Requires the ATR/p62 Autophagy-Related Pathway

Authors: Shiyuan Hong, Yan Li, Paul J. Kaminski, Jorge Andrade, Laimonis A. Laimins | AUTHORS INFO & AFFILIATIONS

https://doi.org/10.1128/mbio.01628-20 • Check for updates

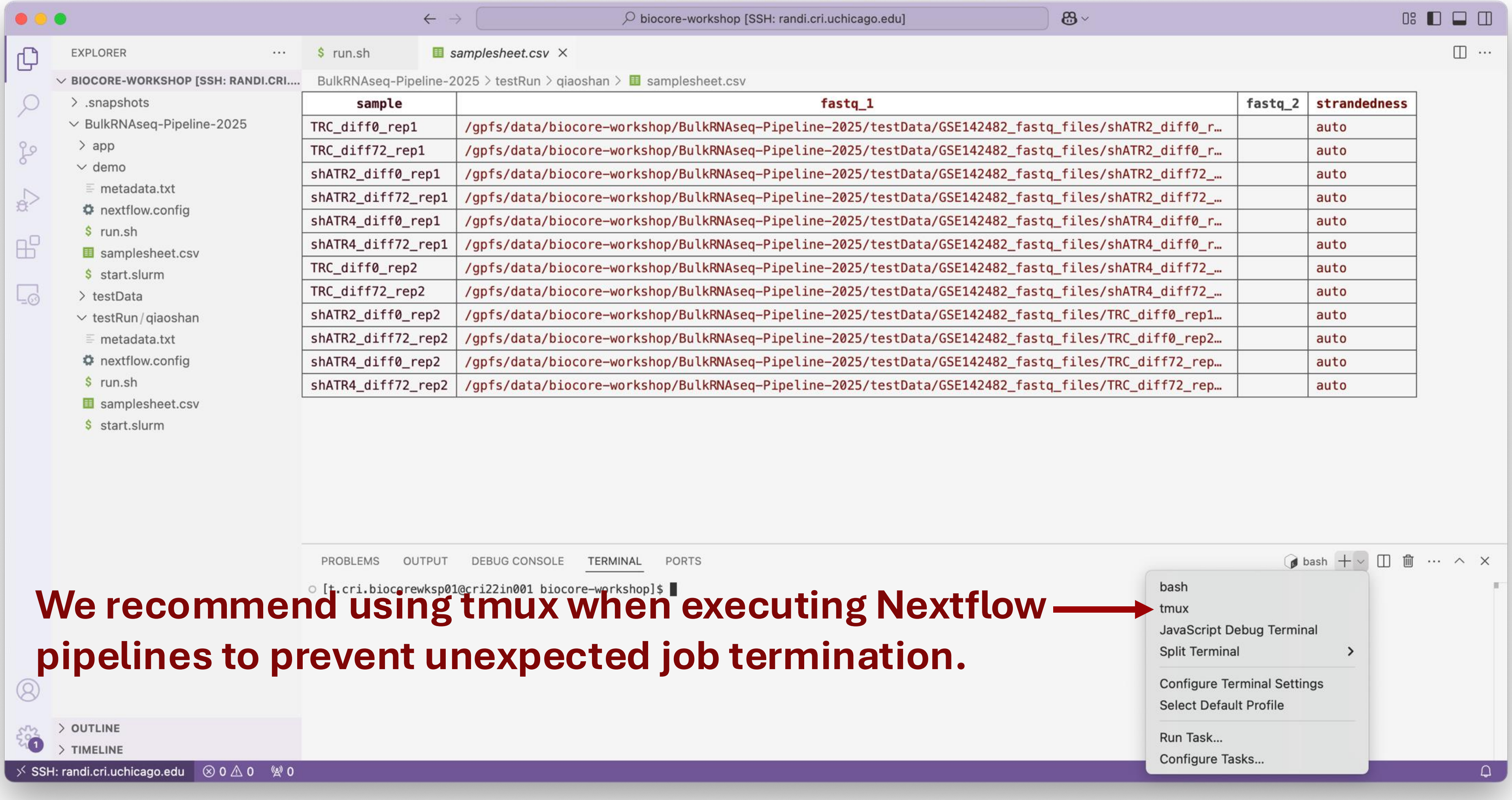
ABSTRACT

High-risk human papillomaviruses (HPVs) constitutively activate the ataxia telangiectasia and Rad3-related (ATR) DNA damage response pathway, and this is required for viral replication. In fibroblasts, activated ATR regulates transcription of inflammatory genes through its negative effects on the autophagosome cargo protein p62. In addition, suppression of p62 results in increased levels of the transcription factor GATA4, leading to cellular senescence. In contrast, in HPV-positive keratinocytes, we observed that activation of ATR resulted in increased levels of phosphorylated p62, which in turn lead to reduced levels of GATA4. Knockdown of ATR in HPV-positive cells resulted in decreased p62 phosphorylation and increased GATA4 levels. Transcriptome sequencing (RNA-seq) analysis of HPV-positive cells identified inflammatory genes and interferon factors as negative transcriptional targets of ATR. Furthermore, knockdown of p62 or overexpression of GATA4 in HPV-positive cells leads to inhibition of viral replication. These findings identify a novel role of the ATR/p62 signaling pathway in HPV-positive cells.

Hong SLi Y, Kaminski PJ, Andrade J, Laimins LA.2020.Pathogenesis of Human Papillomaviruses Requires the ATR/p62 Autophagy-Related Pathway. mBio11:10.1128/mbio.01628-20.https://doi.org/10.1128/mbio.01628-20

Run Nextflow RNAseq Pipeline on Randi

Step 4: Run the Nextflow pipeline



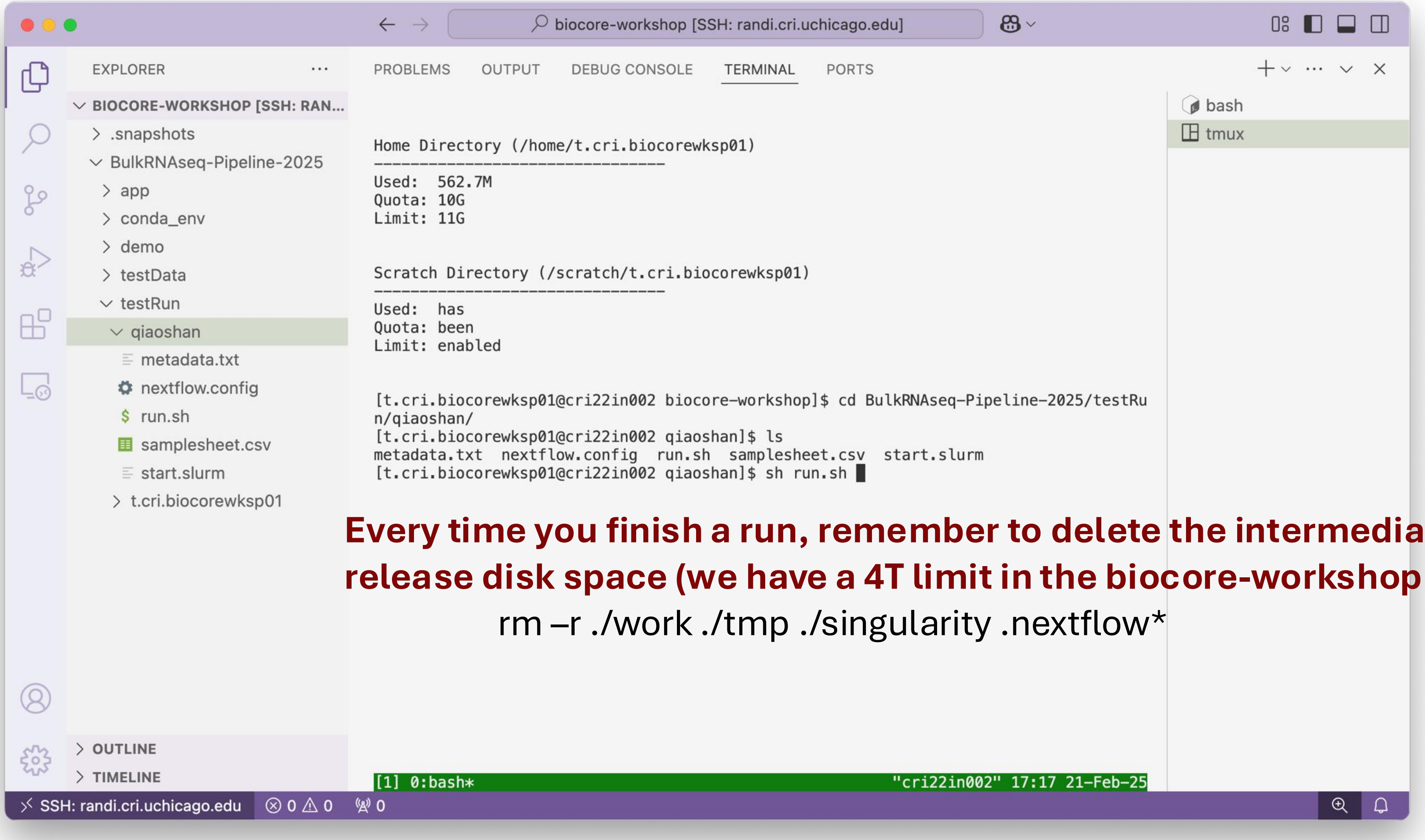
sample	fastq_1	fastq_2	strandedness
TRC_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff0_r…		auto
TRC_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff0_r…		auto
shATR2_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff72_…		auto
shATR2_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR2_diff72_…		auto
shATR4_diff0_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff0_r…		auto
shATR4_diff72_rep1	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff0_r…		auto
TRC_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff72_…		auto
TRC_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/shATR4_diff72_…		auto
shATR2_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff0_rep1…		auto
shATR2_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff0_rep2…		auto
shATR4_diff0_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff72_rep…		auto
shATR4_diff72_rep2	/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testData/GSE142482_fastq_files/TRC_diff72_rep…		auto

```
○ [t.cri.biocorewksp01@cri22in001 biocore-workshop]$ 
```

We recommend using tmux when executing Nextflow pipelines to prevent unexpected job termination.

Run Nextflow RNAseq Pipeline on Randi

Step 4: Run the Nextflow pipeline



```
Home Directory (/home/t.cri.biocorewksp01)
--------------------------------
Used:   562.7M
Quota:  10G
Limit:  11G


Scratch Directory (/scratch/t.cri.biocorewksp01)
--------------------------------
Used:   has
Quota:  been
Limit:  enabled


[t.cri.biocorewksp01@cri22in002 biocore-workshop]$ cd BulkRNAseq-Pipeline-2025/testRu
n/qiaoshan/
[t.cri.biocorewksp01@cri22in002 qiaoshan]$ ls
metadata.txt  nextflow.config  run.sh  samplesheet.csv  start.slurm
[t.cri.biocorewksp01@cri22in002 qiaoshan]$ sh run.sh
[1] 0:bash*                                          "cri22in002" 17:17 21-Feb-25
```

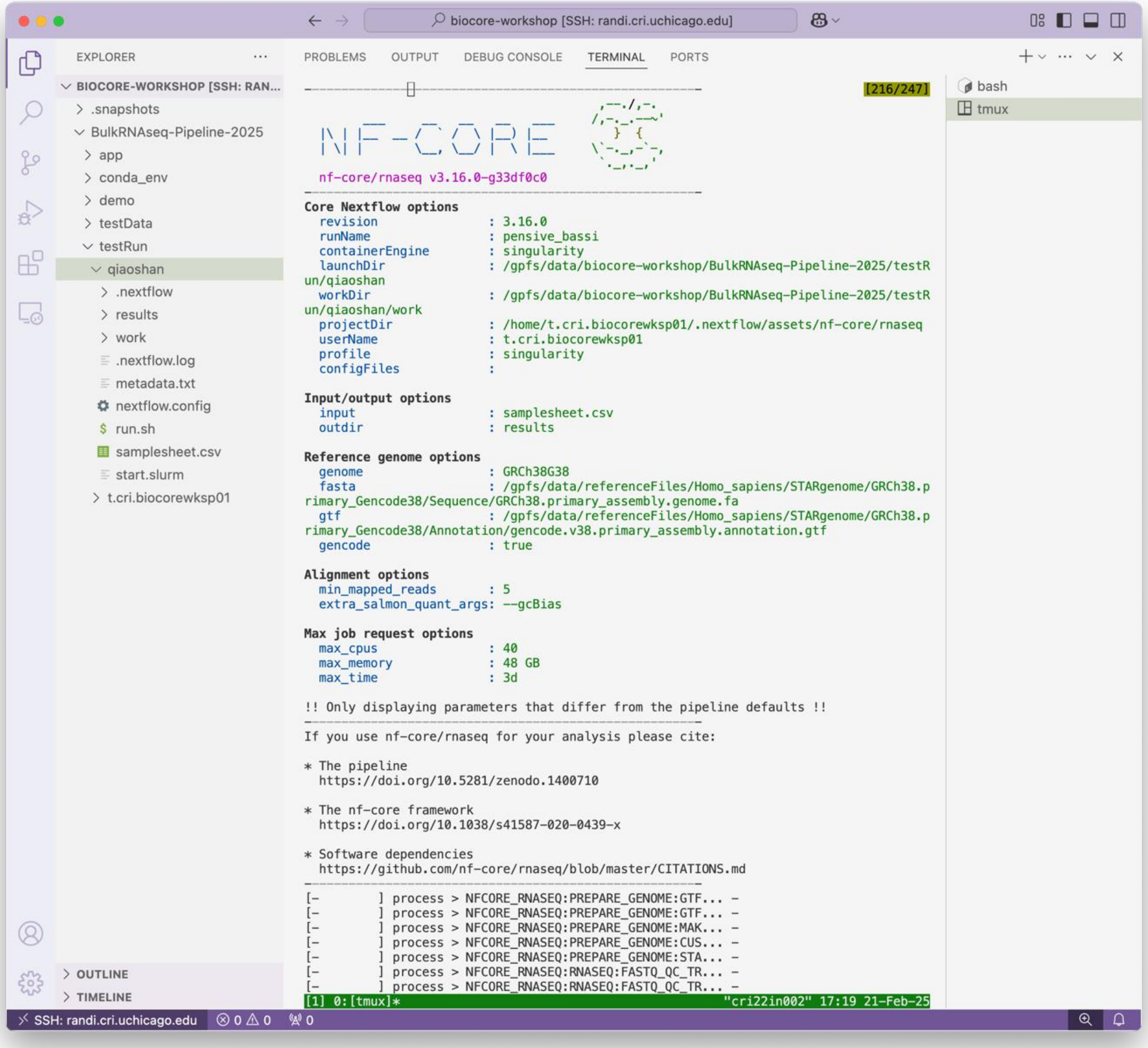
Every time you finish a run, remember to delete the intermediate folders to release disk space (we have a 4T limit in the biocore-workshop folder):

rm –r ./work ./tmp ./singularity .nextflow*

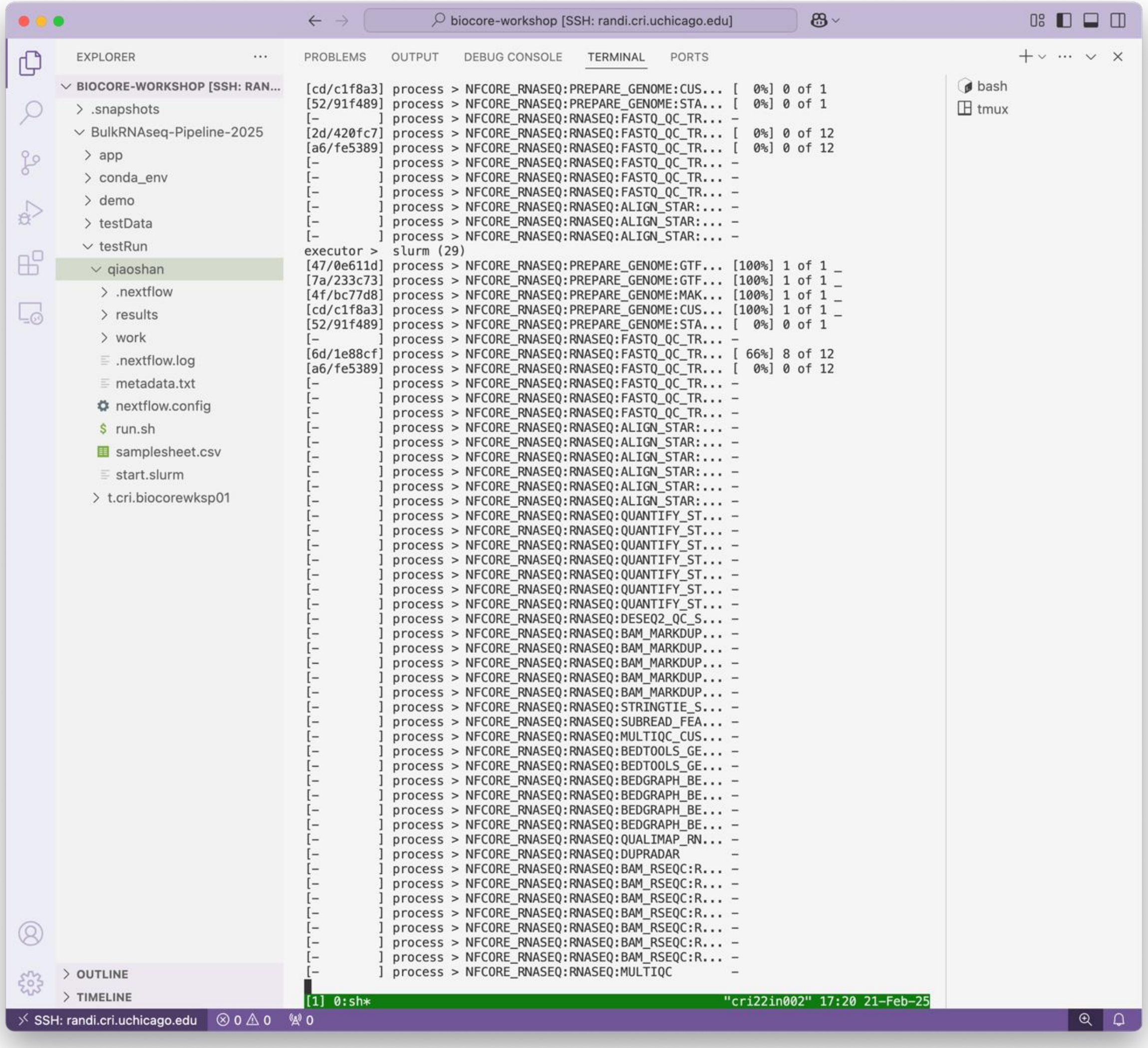
Run Nextflow RNAseq Pipeline on Randi

Step 4: Run the Nextflow pipeline

`control+b; [` to scroll page; press `q` to quit scroll mode



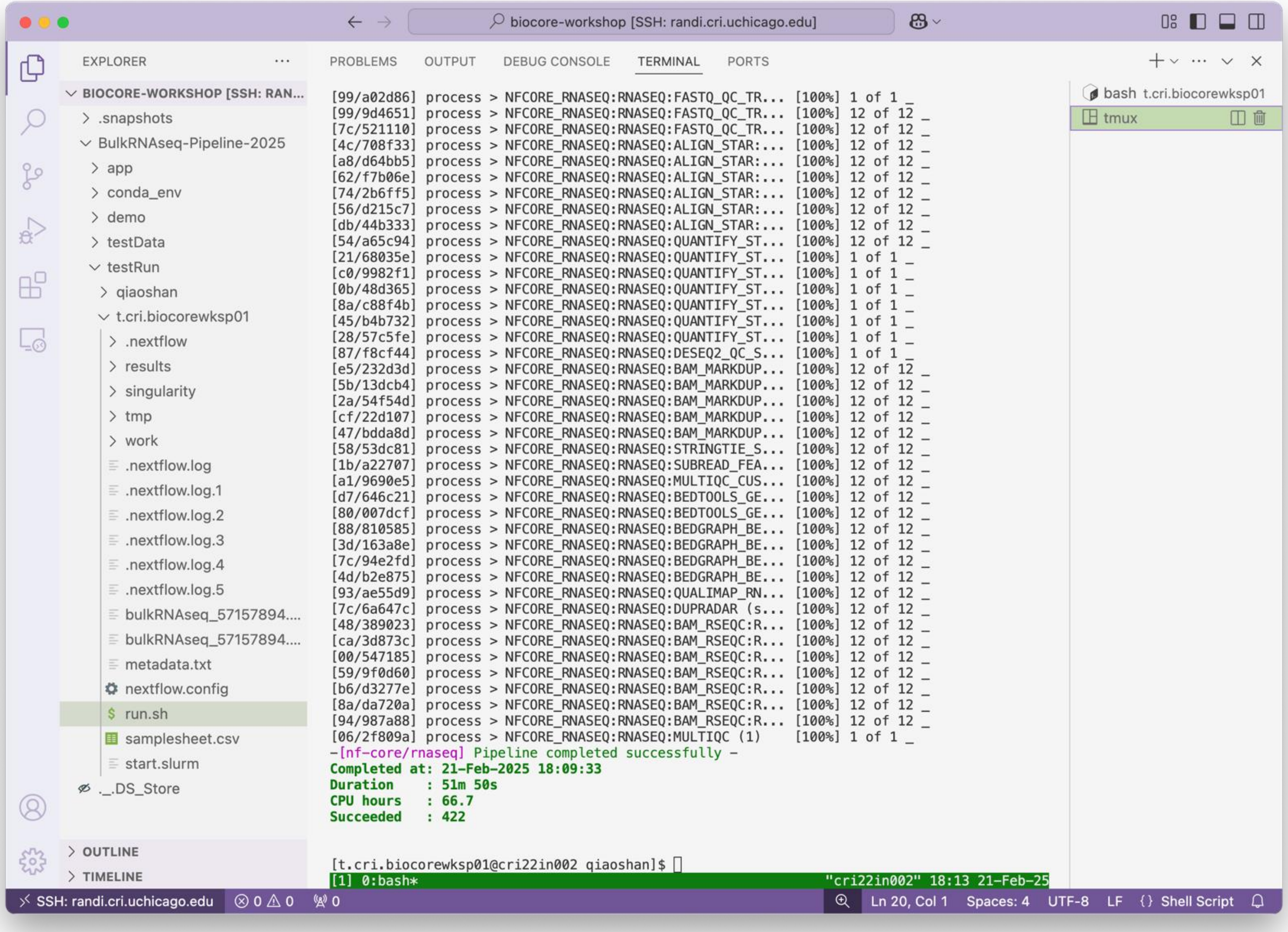
```
-----------------------------------------------------  [216/247]
                                        ,--./,-.
        ___     __   __   __   ___     /,-._.--~'
  |\ | |__  __ /  ` /  \ |__) |__         }  {
  | \| |       \__, \__/ |  \ |___     \`-._,-`-,
                                        `._,._,'
  nf-core/rnaseq v3.16.0-g33df0c0
-----------------------------------------------------
Core Nextflow options
  revision                  : 3.16.0
  runName                   : pensive_bassi
  containerEngine           : singularity
  launchDir                 : /gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testR
un/qiaoshan
  workDir                   : /gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testR
un/qiaoshan/work
  projectDir                : /home/t.cri.biocorewksp01/.nextflow/assets/nf-core/rnaseq
  userName                  : t.cri.biocorewksp01
  profile                   : singularity
  configFiles               :

Input/output options
  input                     : samplesheet.csv
  outdir                    : results

Reference genome options
  genome                    : GRCh38G38
  fasta                     : /gpfs/data/referenceFiles/Homo_sapiens/STARgenome/GRCh38.p
rimary_Gencode38/Sequence/GRCh38.primary_assembly.genome.fa
  gtf                       : /gpfs/data/referenceFiles/Homo_sapiens/STARgenome/GRCh38.p
rimary_Gencode38/Annotation/gencode.v38.primary_assembly.annotation.gtf
  gencode                   : true

Alignment options
  min_mapped_reads          : 5
  extra_salmon_quant_args: --gcBias

Max job request options
  max_cpus                  : 40
  max_memory                : 48 GB
  max_time                  : 3d

!! Only displaying parameters that differ from the pipeline defaults !!
-----------------------------------------------------
If you use nf-core/rnaseq for your analysis please cite:

* The pipeline
  https://doi.org/10.5281/zenodo.1400710

* The nf-core framework
  https://doi.org/10.1038/s41587-020-0439-x

* Software dependencies
  https://github.com/nf-core/rnaseq/blob/master/CITATIONS.md
-----------------------------------------------------
[-        ] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF... -
[-        ] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF... -
[-        ] process > NFCORE_RNASEQ:PREPARE_GENOME:MAK... -
[-        ] process > NFCORE_RNASEQ:PREPARE_GENOME:CUS... -
[-        ] process > NFCORE_RNASEQ:PREPARE_GENOME:STA... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[1] 0:[tmux]*                                 "cri22in002" 17:19 21-Feb-25
```



```
[cd/c1f8a3] process > NFCORE_RNASEQ:PREPARE_GENOME:CUS... [  0%] 0 of 1
[52/91f489] process > NFCORE_RNASEQ:PREPARE_GENOME:STA... [  0%] 0 of 1
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[2d/420fc7] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [  0%] 0 of 12
[a6/fe5389] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [  0%] 0 of 12
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
executor >  slurm (29)
[47/0e611d] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF... [100%] 1 of 1 _
[7a/233c73] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF... [100%] 1 of 1 _
[4f/bc77d8] process > NFCORE_RNASEQ:PREPARE_GENOME:MAK... [100%] 1 of 1 _
[cd/c1f8a3] process > NFCORE_RNASEQ:PREPARE_GENOME:CUS... [100%] 1 of 1 _
[52/91f489] process > NFCORE_RNASEQ:PREPARE_GENOME:STA... [  0%] 0 of 1
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[6d/1e88cf] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [ 66%] 8 of 12
[a6/fe5389] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [  0%] 0 of 12
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:DESEQ2_QC_S... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:STRINGTIE_S... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:SUBREAD_FEA... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:MULTIQC_CUS... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDTOOLS_GE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDTOOLS_GE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUALIMAP_RN... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:DUPRADAR       -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... -
[-        ] process > NFCORE_RNASEQ:RNASEQ:MULTIQC        -

[1] 0:sh*                                     "cri22in002" 17:20 21-Feb-25
```

Run Nextflow RNAseq Pipeline on Randi

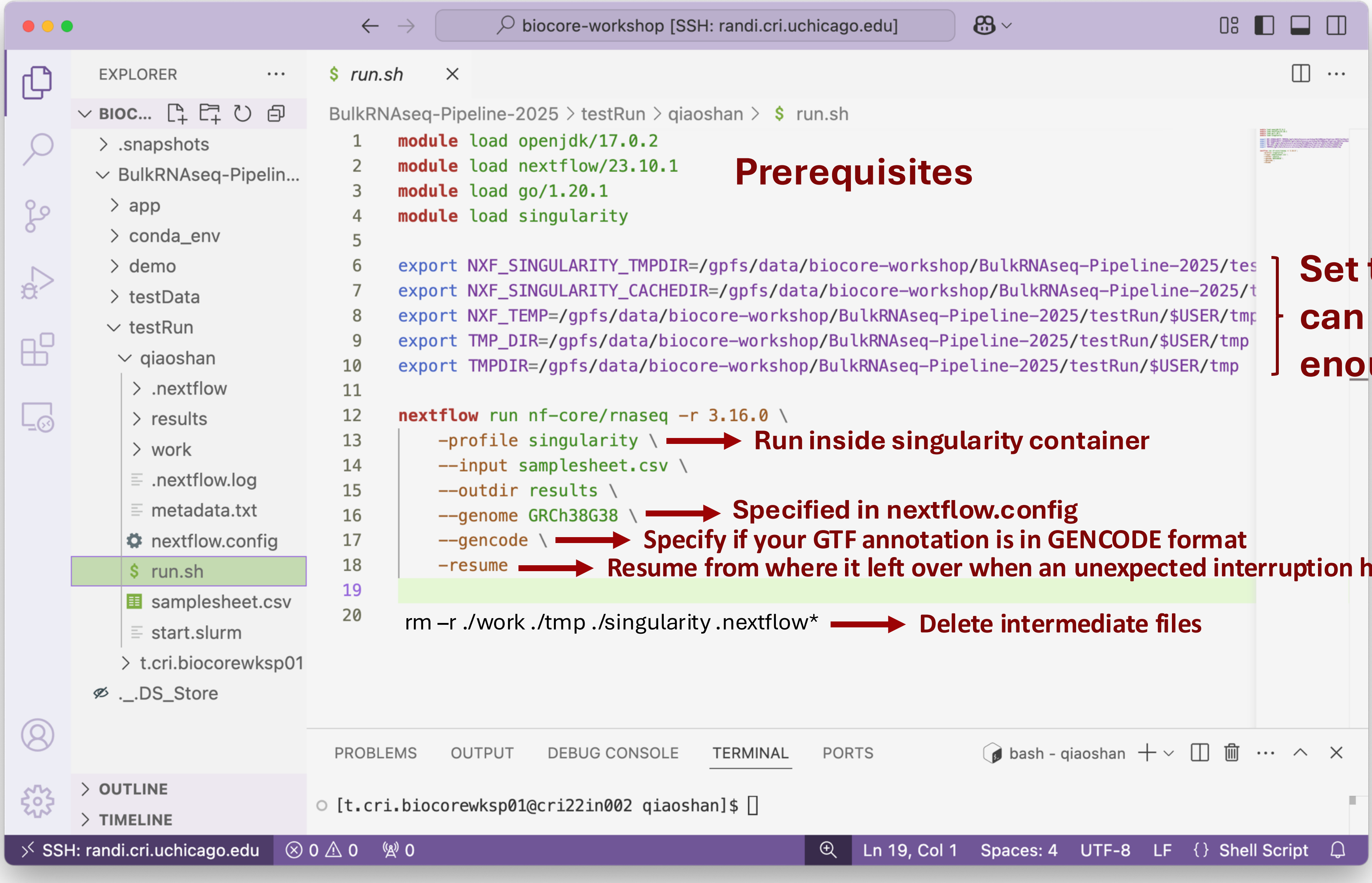
Step 4: Run the Nextflow pipeline



```
[99/a02d86] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [100%] 1 of 1 _
[99/9d4651] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [100%] 12 of 12 _
[7c/521110] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TR... [100%] 12 of 12 _
[4c/708f33] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[a8/d64bb5] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[62/f7b06e] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[74/2b6ff5] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[56/d215c7] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[db/44b333] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:... [100%] 12 of 12 _
[54/a65c94] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 12 of 12 _
[21/68035e] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[c0/9982f1] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[0b/48d365] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[8a/c88f4b] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[45/b4b732] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[28/57c5fe] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_ST... [100%] 1 of 1 _
[87/f8cf44] process > NFCORE_RNASEQ:RNASEQ:DESEQ2_QC_S... [100%] 1 of 1 _
[e5/232d3d] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... [100%] 12 of 12 _
[5b/13dcb4] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... [100%] 12 of 12 _
[2a/54f54d] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... [100%] 12 of 12 _
[cf/22d107] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... [100%] 12 of 12 _
[47/bdda8d] process > NFCORE_RNASEQ:RNASEQ:BAM_MARKDUP... [100%] 12 of 12 _
[58/53dc81] process > NFCORE_RNASEQ:RNASEQ:STRINGTIE_S... [100%] 12 of 12 _
[1b/a22707] process > NFCORE_RNASEQ:RNASEQ:SUBREAD_FEA... [100%] 12 of 12 _
[a1/9690e5] process > NFCORE_RNASEQ:RNASEQ:MULTIQC_CUS... [100%] 12 of 12 _
[d7/646c21] process > NFCORE_RNASEQ:RNASEQ:BEDTOOLS_GE... [100%] 12 of 12 _
[80/007dcf] process > NFCORE_RNASEQ:RNASEQ:BEDTOOLS_GE... [100%] 12 of 12 _
[88/810585] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... [100%] 12 of 12 _
[3d/163a8e] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... [100%] 12 of 12 _
[7c/94e2fd] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... [100%] 12 of 12 _
[4d/b2e875] process > NFCORE_RNASEQ:RNASEQ:BEDGRAPH_BE... [100%] 12 of 12 _
[93/ae55d9] process > NFCORE_RNASEQ:RNASEQ:QUALIMAP_RN... [100%] 12 of 12 _
[7c/6a647c] process > NFCORE_RNASEQ:RNASEQ:DUPRADAR (s... [100%] 12 of 12 _
[48/389023] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[ca/3d873c] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[00/547185] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[59/9f0d60] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[b6/d3277e] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[8a/da720a] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[94/987a88] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:R... [100%] 12 of 12 _
[06/2f809a] process > NFCORE_RNASEQ:RNASEQ:MULTIQC (1)    [100%] 1 of 1 _
-[nf-core/rnaseq] Pipeline completed successfully -
Completed at: 21-Feb-2025 18:09:33
Duration    : 51m 50s
CPU hours   : 66.7
Succeeded   : 422


[t.cri.biocorewksp01@cri22in002 qiaoshan]$ 
[1] 0:bash*                                                "cri22in002" 18:13 21-Feb-25
```

Once completed, return to the bash terminal and run: *tmux kill-session -t [session number]*

Key Parameters



```
module load openjdk/17.0.2
module load nextflow/23.10.1
module load go/1.20.1
module load singularity

export NXF_SINGULARITY_TMPDIR=/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/tes
export NXF_SINGULARITY_CACHEDIR=/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/t
export NXF_TEMP=/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testRun/$USER/tmp
export TMP_DIR=/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testRun/$USER/tmp
export TMPDIR=/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testRun/$USER/tmp

nextflow run nf-core/rnaseq -r 3.16.0 \
    -profile singularity \
    --input samplesheet.csv \
    --outdir results \
    --genome GRCh38G38 \
    --gencode \
    -resume

rm –r ./work ./tmp ./singularity .nextflow*
```

```
[t.cri.biocorewksp01@cri22in002 qiaoshan]$ 
```

Prerequisites

Set to folders that you can access and have enough space

-profile singularity → **Run inside singularity container**

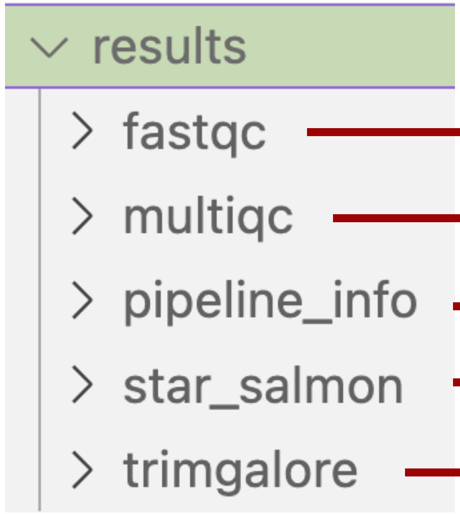
--genome GRCh38G38 → **Specified in nextflow.config**

--gencode → **Specify if your GTF annotation is in GENCODE format**

-resume → **Resume from where it left over when an unexpected interruption happens**

rm –r ./work ./tmp ./singularity .nextflow* → **Delete intermediate files**

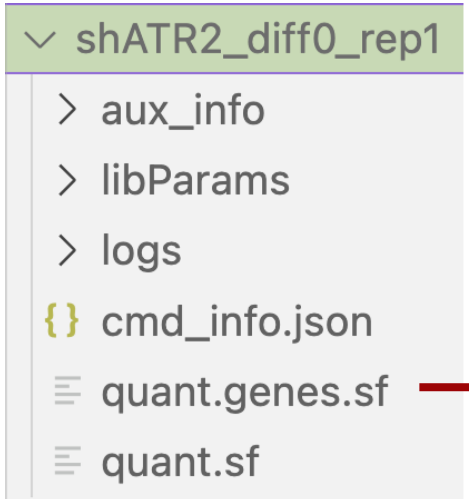
Interpretation of Nextflow Outputs



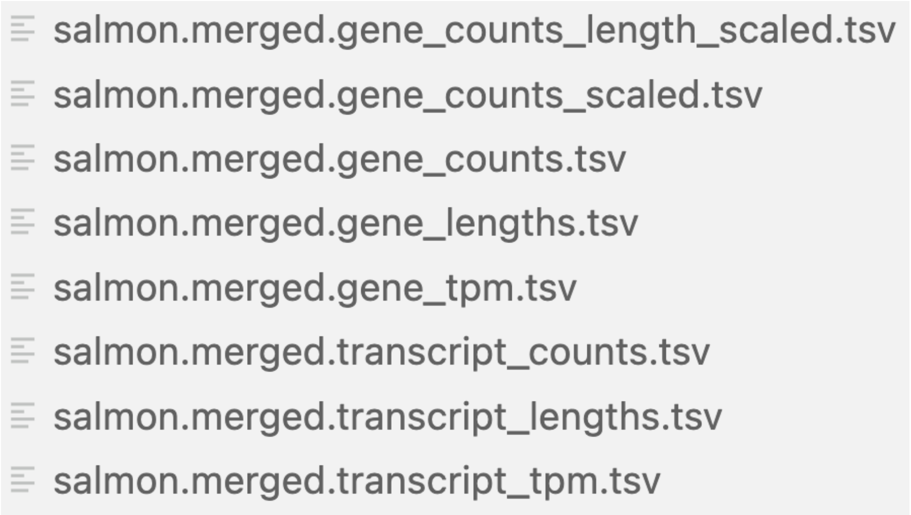
results
- fastqc → FastQC reports of raw reads and trimmed reads
- multiqc → *First to check* One report with QC of each step integrated in one place
- pipeline_info → Commands for each step & Environment setup
- star_salmon → Filtered bam files & **Gene counts**
- trimgalore → Read trimming reports

Each sample has a salmon output folder with *.sf files.

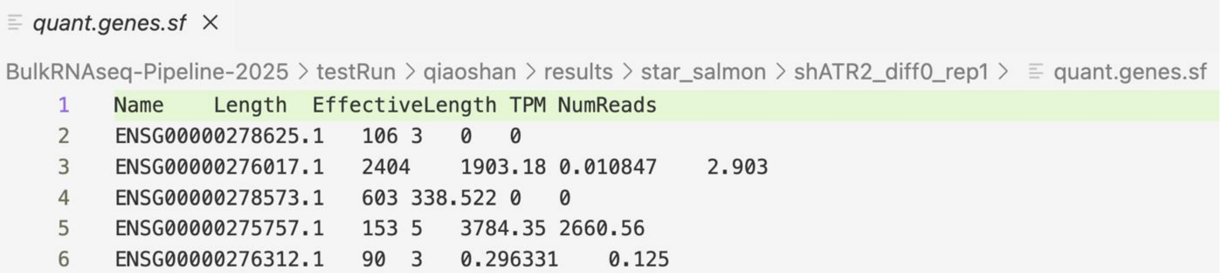
I recommend using *.sf files for the downstream to take advantage of the bias correction by Salmon



shATR2_diff0_rep1
- aux_info
- libParams
- logs
- cmd_info.json
- quant.genes.sf
- quant.sf



- salmon.merged.gene_counts_length_scaled.tsv
- salmon.merged.gene_counts_scaled.tsv
- salmon.merged.gene_counts.tsv
- salmon.merged.gene_lengths.tsv
- salmon.merged.gene_tpm.tsv
- salmon.merged.transcript_counts.tsv
- salmon.merged.transcript_lengths.tsv
- salmon.merged.transcript_tpm.tsv



BulkRNAseq-Pipeline-2025 > testRun > qiaoshan > results > star_salmon > shATR2_diff0_rep1 > quant.genes.sf

```
Name	Length	EffectiveLength	TPM	NumReads
ENSG00000278625.1	106	3	0	0
ENSG00000276017.1	2404	1903.18	0.010847	2.903
ENSG00000278573.1	603	338.522	0	0
ENSG00000275757.1	153	5	3784.35	2660.56
ENSG00000276312.1	90	3	0.296331	0.125
```

Agenda & Key Activities

Section 1

- Introduction to the Nextflow RNAseq Pipeline
- Hands-on Practice on Running Nextflow on the Randi Server
- Interpretation of Nextflow Outputs

TEN-MINUTES BREAK

Section 2

- Introduction to the DE Analysis Principles
- Demo of the DE Analysis App on Randi
- Hands-on Practice on Running the DE Analysis App
- Interpretation of the DE Analysis Results

Section 2

DE Analysis

Key Considerations in DE Analysis

⚠ **Batch Effects** → Use PCA/MDS to check for unwanted variation

⚠ **Replicates Matter** → More replicates = higher statistical power (highly recommend >=3 rep)

⚠ **Read Depth** → Sufficient sequencing depth (25M)

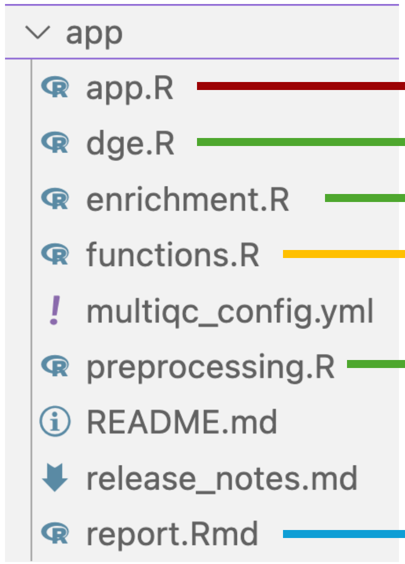
⚠ **Data Distribution** → RNA-seq data is often over-dispersed (use a model like Negative Binomial)

⚠ **Normalization** → Correct for library size & sequencing depth

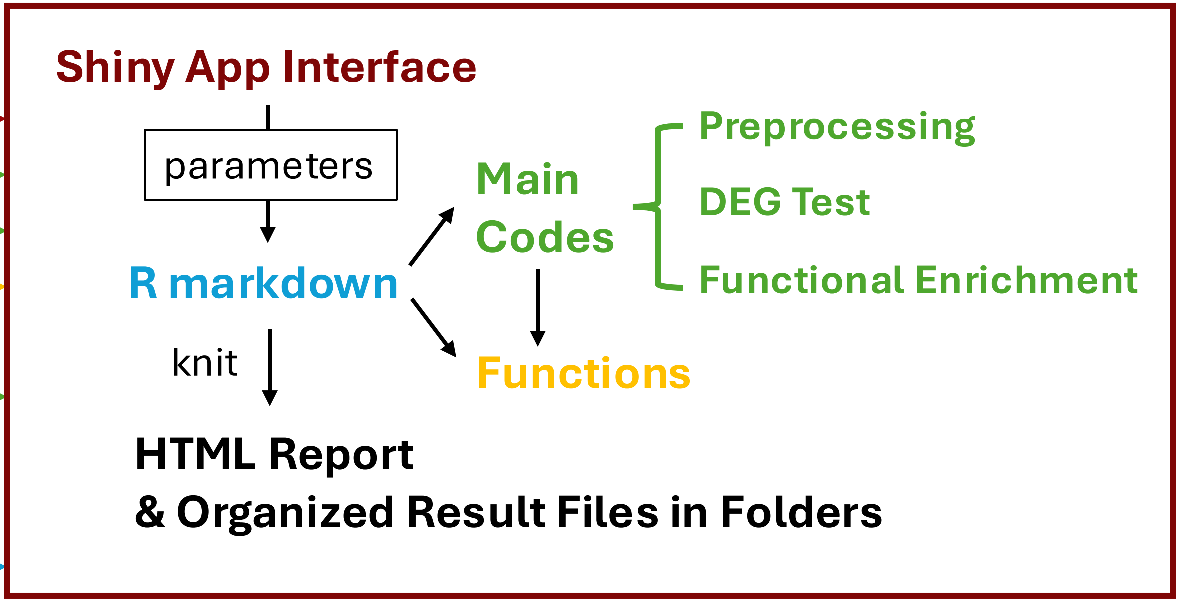
⚠ **Multiple Testing Correction** → Use adjusted p-values to control false positives

In-House Downstream Analysis

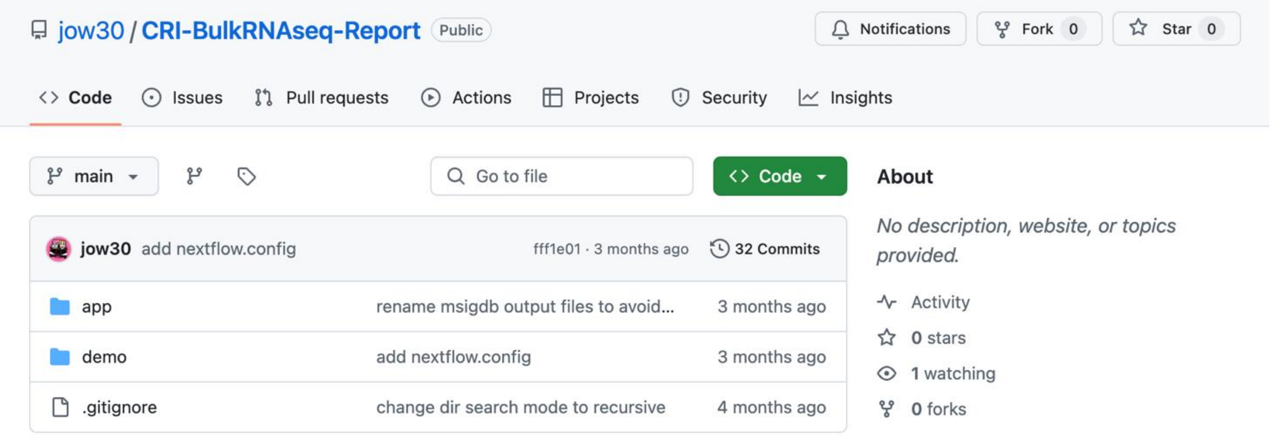
Architecture



- app
  - app.R
  - dge.R
  - enrichment.R
  - functions.R
  - multiqc_config.yml
  - preprocessing.R
  - README.md
  - release_notes.md
  - report.Rmd



Shiny App Interface



jow30 / CRI-BulkRNAseq-Report Public

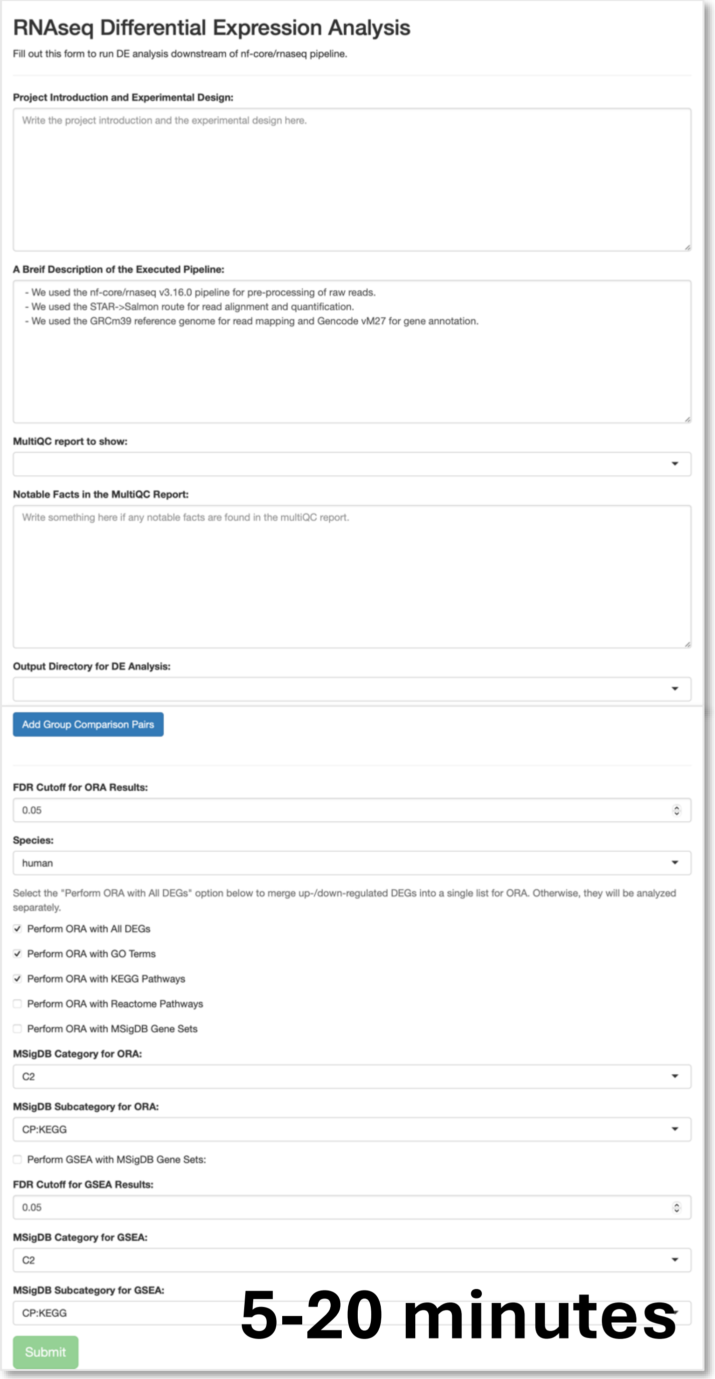
jow30 add nextflow.config fff1e01 · 3 months ago 32 Commits

app	rename msigdb output files to avoid...	3 months ago
demo	add nextflow.config	3 months ago
.gitignore	change dir search mode to recursive	4 months ago

About

No description, website, or topics provided.

- Activity
- 0 stars
- 1 watching
- 0 forks



RNAseq Differential Expression Analysis

Fill out this form to run DE analysis downstream of nf-core/rnaseq pipeline.

Project Introduction and Experimental Design:

Write the project introduction and the experimental design here.

A Breif Description of the Executed Pipeline:

- We used the nf-core/rnaseq v3.16.0 pipeline for pre-processing of raw reads.
- We used the STAR->Salmon route for read alignment and quantification.
- We used the GRCm39 reference genome for read mapping and Gencode vM27 for gene annotation.

MultiQC report to show:

Notable Facts in the MultiQC Report:

Write something here if any notable facts are found in the multiQC report.

Output Directory for DE Analysis:

FDR Cutoff for ORA Results: 0.05

Species: human

Select the "Perform ORA with All DEGs" option below to merge up-/down-regulated DEGs into a single list for ORA. Otherwise, they will be analyzed separately.

☑ Perform ORA with All DEGs

☑ Perform ORA with GO Terms

☑ Perform ORA with KEGG Pathways

☐ Perform ORA with Reactome Pathways

☐ Perform ORA with MSigDB Gene Sets

MSigDB Category for ORA: C2

MSigDB Subcategory for ORA: CP:KEGG

☐ Perform GSEA with MSigDB Gene Sets:

FDR Cutoff for GSEA Results: 0.05

MSigDB Category for GSEA: C2

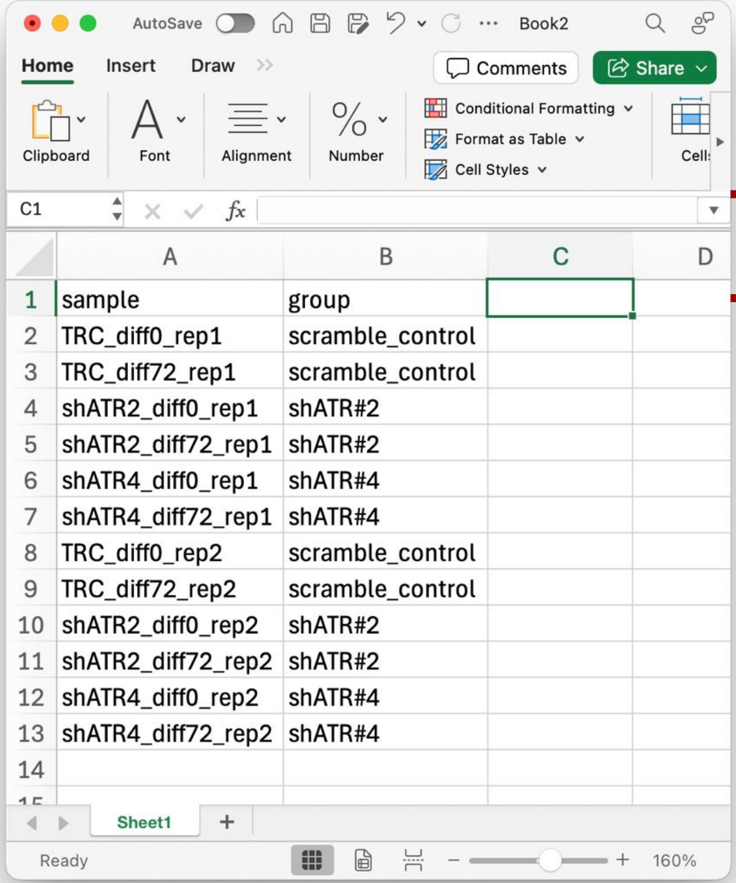
MSigDB Subcategory for GSEA: CP:KEGG

5-20 minutes

In-House Downstream Analysis

Step 1: Setup experimental groups

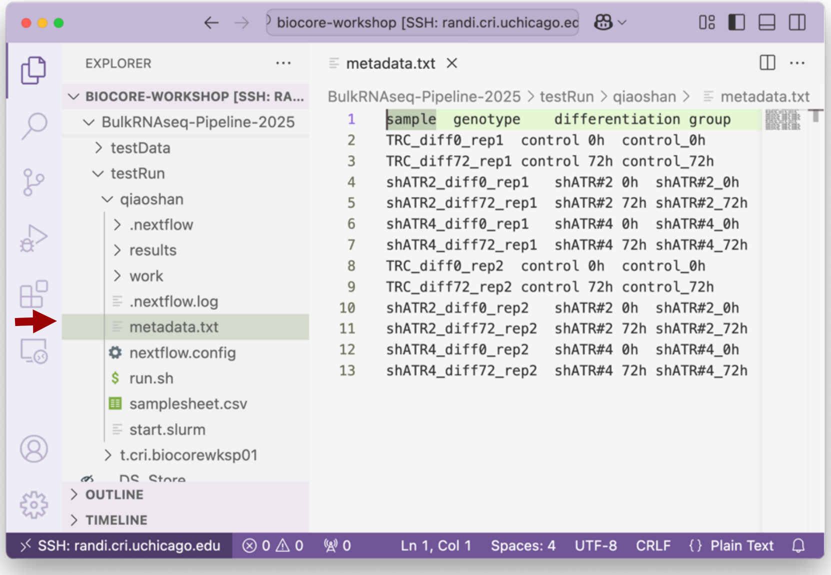
1. Edit the table in Excel



	A	B	C	D
1	sample	group		
2	TRC_diff0_rep1	scramble_control		
3	TRC_diff72_rep1	scramble_control		
4	shATR2_diff0_rep1	shATR#2		
5	shATR2_diff72_rep1	shATR#2		
6	shATR4_diff0_rep1	shATR#4		
7	shATR4_diff72_rep1	shATR#4		
8	TRC_diff0_rep2	scramble_control		
9	TRC_diff72_rep2	scramble_control		
10	shATR2_diff0_rep2	shATR#2		
11	shATR2_diff72_rep2	shATR#2		
12	shATR4_diff0_rep2	shATR#4		
13	shATR4_diff72_rep2	shATR#4		
14				

2. Save the table in the txt or csv format

3. Drag to your working folder



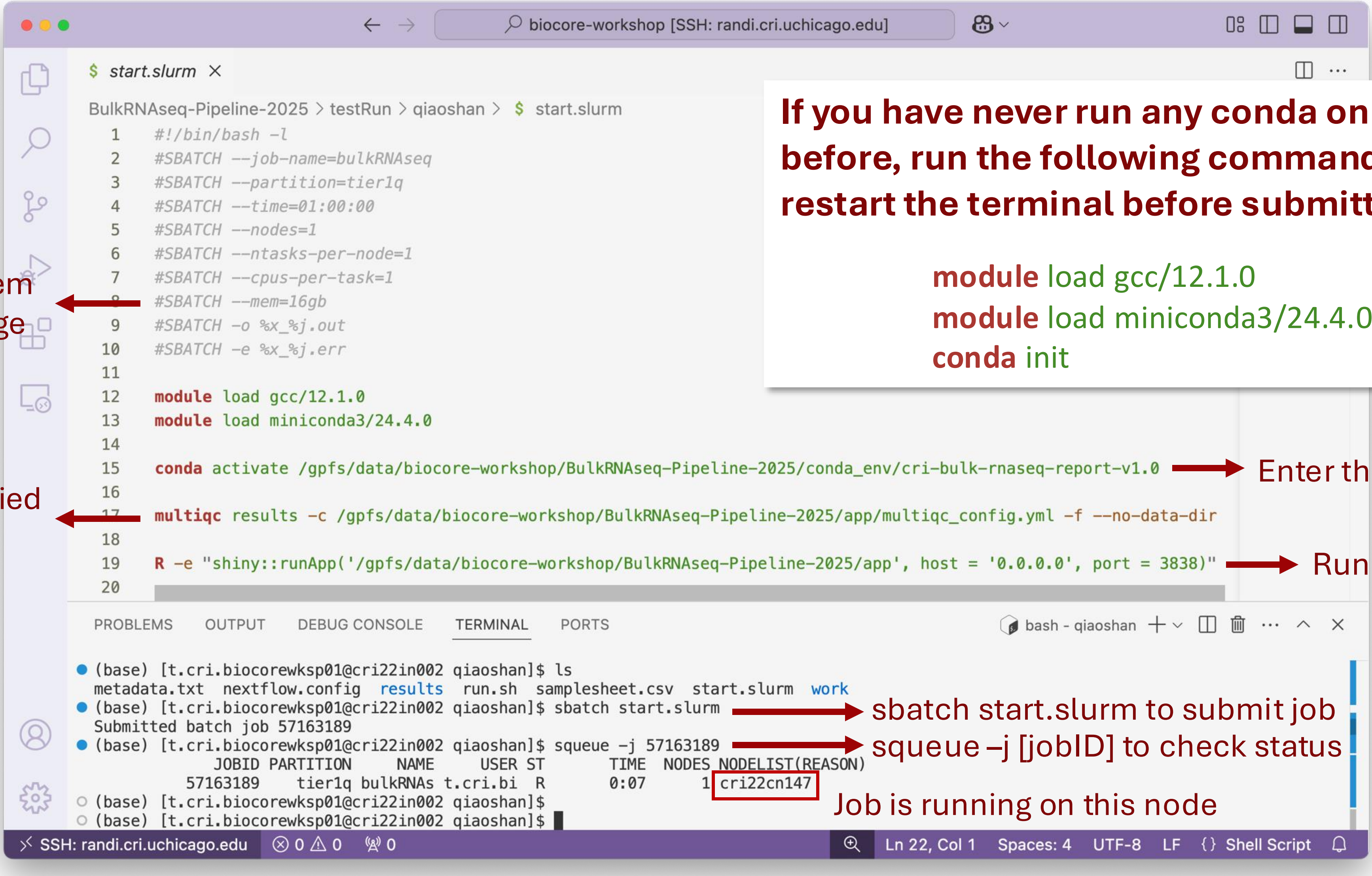
```
sample  genotype    differentiation group
TRC_diff0_rep1  control 0h  control_0h
TRC_diff72_rep1 control 72h control_72h
shATR2_diff0_rep1   shATR#2 0h  shATR#2_0h
shATR2_diff72_rep1  shATR#2 72h shATR#2_72h
shATR4_diff0_rep1   shATR#4 0h  shATR#4_0h
shATR4_diff72_rep1  shATR#4 72h shATR#4_72h
TRC_diff0_rep2  control 0h  control_0h
TRC_diff72_rep2 control 72h control_72h
shATR2_diff0_rep2   shATR#2 0h  shATR#2_0h
shATR2_diff72_rep2  shATR#2 72h shATR#2_72h
shATR4_diff0_rep2   shATR#4 0h  shATR#4_0h
shATR4_diff72_rep2  shATR#4 72h shATR#4_72h
```

The header must contain *sample* and *group*. You can add as many experimental factors as you want to columns.

If a batch effect needs to be corrected, add a *batch* column so that the batch-effect-removal option can be enabled.

In-House Downstream Analysis

Step 2: Run the slurm script


biocore-workshop [SSH: randi.cri.uchicago.edu]
$ start.slurm
BulkRNAseq-Pipeline-2025 > testRun > qiaoshan > $ start.slurm
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
bash - qiaoshan
SSH: randi.cri.uchicago.edu
⊗ 0 ⚠ 0
0
Ln 22, Col 1
Spaces: 4
UTF-8
LF
{ } Shell Script

```
#!/bin/bash -l
#SBATCH --job-name=bulkRNAseq
#SBATCH --partition=tier1q
#SBATCH --time=01:00:00
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=1
#SBATCH --mem=16gb
#SBATCH -o %x_%j.out
#SBATCH -e %x_%j.err

module load gcc/12.1.0
module load miniconda3/24.4.0

conda activate /gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/conda_env/cri-bulk-rnaseq-report-v1.0

multiqc results -c /gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/app/multiqc_config.yml -f --no-data-dir

R -e "shiny::runApp('/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/app', host = '0.0.0.0', port = 3838)"

```

```
● (base) [t.cri.biocorewksp01@cri22in002 qiaoshan]$ ls
  metadata.txt  nextflow.config  results  run.sh  samplesheet.csv  start.slurm  work
● (base) [t.cri.biocorewksp01@cri22in002 qiaoshan]$ sbatch start.slurm
  Submitted batch job 57163189
● (base) [t.cri.biocorewksp01@cri22in002 qiaoshan]$ squeue -j 57163189
               JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
            57163189    tier1q bulkRNAs t.cri.bi  R       0:07      1 cri22cn147
○ (base) [t.cri.biocorewksp01@cri22in002 qiaoshan]$
○ (base) [t.cri.biocorewksp01@cri22in002 qiaoshan]$ 
```

Increase mem if data is large

Produce a simplified multiQC report.

Enter the conda environment

Run R shiny app

sbatch start.slurm to submit job

squeue –j [jobID] to check status

Job is running on this node

If you have never run any conda on Randi before, run the following commands and restart the terminal before submitting job:

```
module load gcc/12.1.0
module load miniconda3/24.4.0
conda init
```

In-House Downstream Analysis

Step 3: Fill out the form

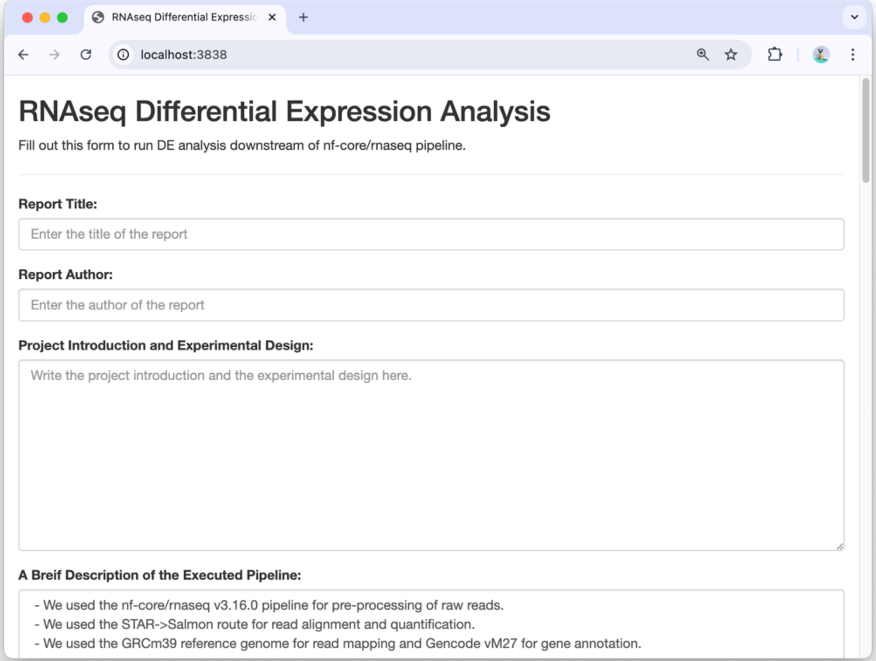
1. Open a local terminal and run:

```
ssh -N -f -L 3838:shinyApp_running_node:3838 your_account@randi.cri.uchicago.edu
```

   1. -L XXXX:cri22cnYYY:3838 → **Forward local port XXXX to remote port 3838**.
   2. -N → **No interactive shell**.
   3. -f → **Run in the background** (not supported in Windows OpenSSH).

2. Open http://localhost:3838/ in a browser (need to wait until multiqc is done)

3. Follow the hints to fill out the form and submit



RNAseq Differential Expression Analysis

Fill out this form to run DE analysis downstream of nf-core/rnaseq pipeline.

Report Title:

Enter the title of the report

Report Author:

Enter the author of the report

Project Introduction and Experimental Design:

Write the project introduction and the experimental design here.

A Breif Description of the Executed Pipeline:

- We used the nf-core/rnaseq v3.16.0 pipeline for pre-processing of raw reads.
- We used the STAR->Salmon route for read alignment and quantification.
- We used the GRCm39 reference genome for read mapping and Gencode vM27 for gene annotation.

Interpretation of DE Analysis Results

PCA Plot

- A dimension reduction approach
- To show similar samples clustering together
- To infer which gene (or variable) is the most valuable for clustering the data
- To reveal batch effects

Volcano Plot

- To highlight significant DEGs

Heatmap

- To visualize and validate expression changes across samples

Over Representation Analysis (ORA)

- Which gene set is enriched with DEGs (whether a gene set contains more DEGs than expected by chance)

Gene Set Enrichment Analysis (GSEA)

- Identifies pathways that are globally up- or down-regulated (by using continuous gene rankings instead of requiring an arbitrary DEG cutoff)

Reminders

- Please remember to delete your folder after practice to save space for other people.

- We will hold the BulkRNAseq-Pipeline-2025 folder for you until Mar. 25. All data will be removed to make room for our next workshop.

Upcoming Workshop: Spatial Transcriptome

Diana Vera Cruz, PhD

Jason Shapiro, PhD

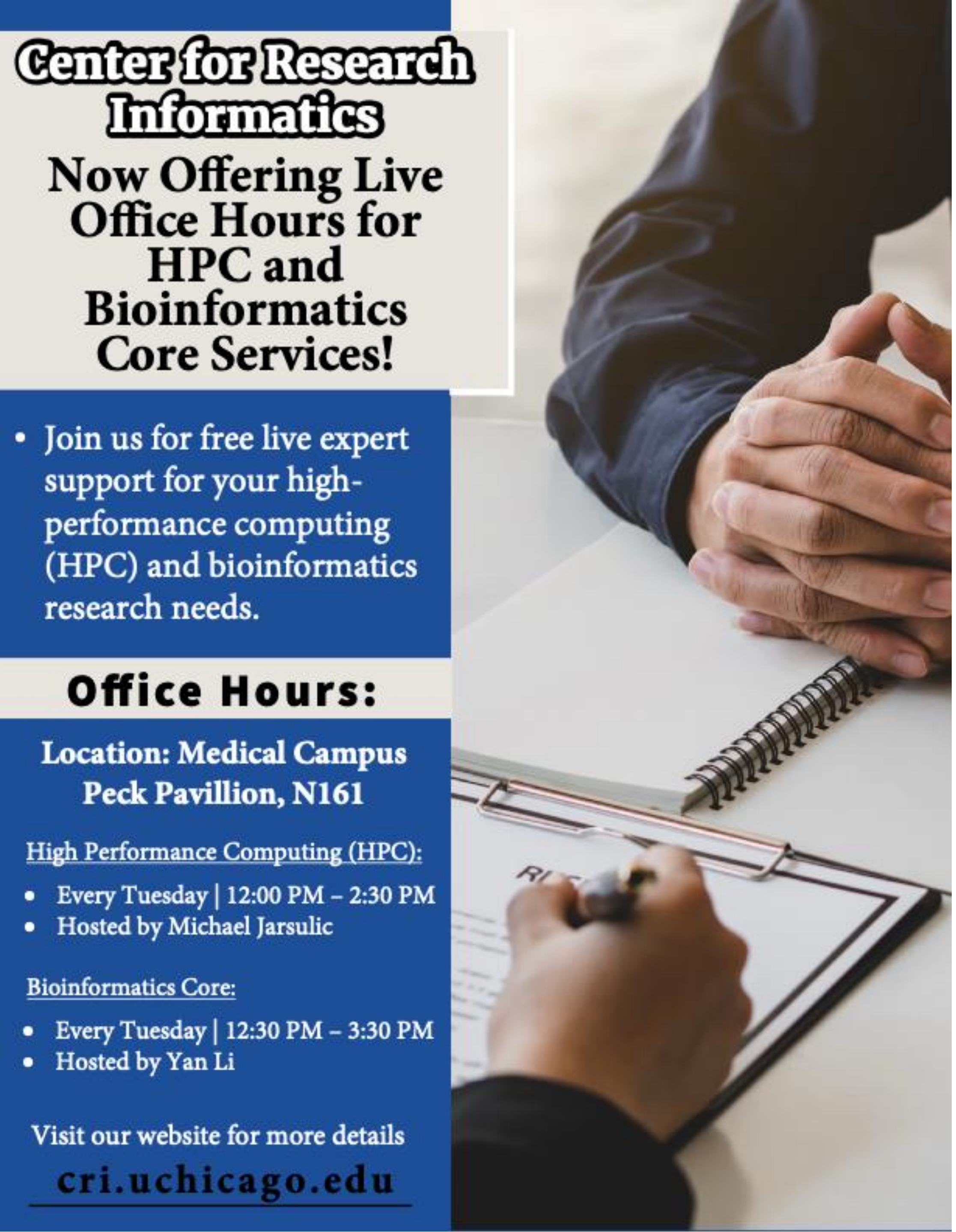
Date: Beginning of April

Overview of methods for analyzing 10X Visium and Nanostring GeoMX data. Including:

-Data pre-processing and quality control
-Worked examples of different analytical workflows
-Introduction to common R packages

Thanks for attending! 💐

Q & A



Center for Research Informatics

Now Offering Live Office Hours for HPC and Bioinformatics Core Services!

- Join us for free live expert support for your high-performance computing (HPC) and bioinformatics research needs.

Office Hours:

Location: Medical Campus Peck Pavillion, N161

High Performance Computing (HPC):

- Every Tuesday | 12:00 PM – 2:30 PM
- Hosted by Michael Jarsulic

Bioinformatics Core:

- Every Tuesday | 12:30 PM – 3:30 PM
- Hosted by Yan Li

Visit our website for more details

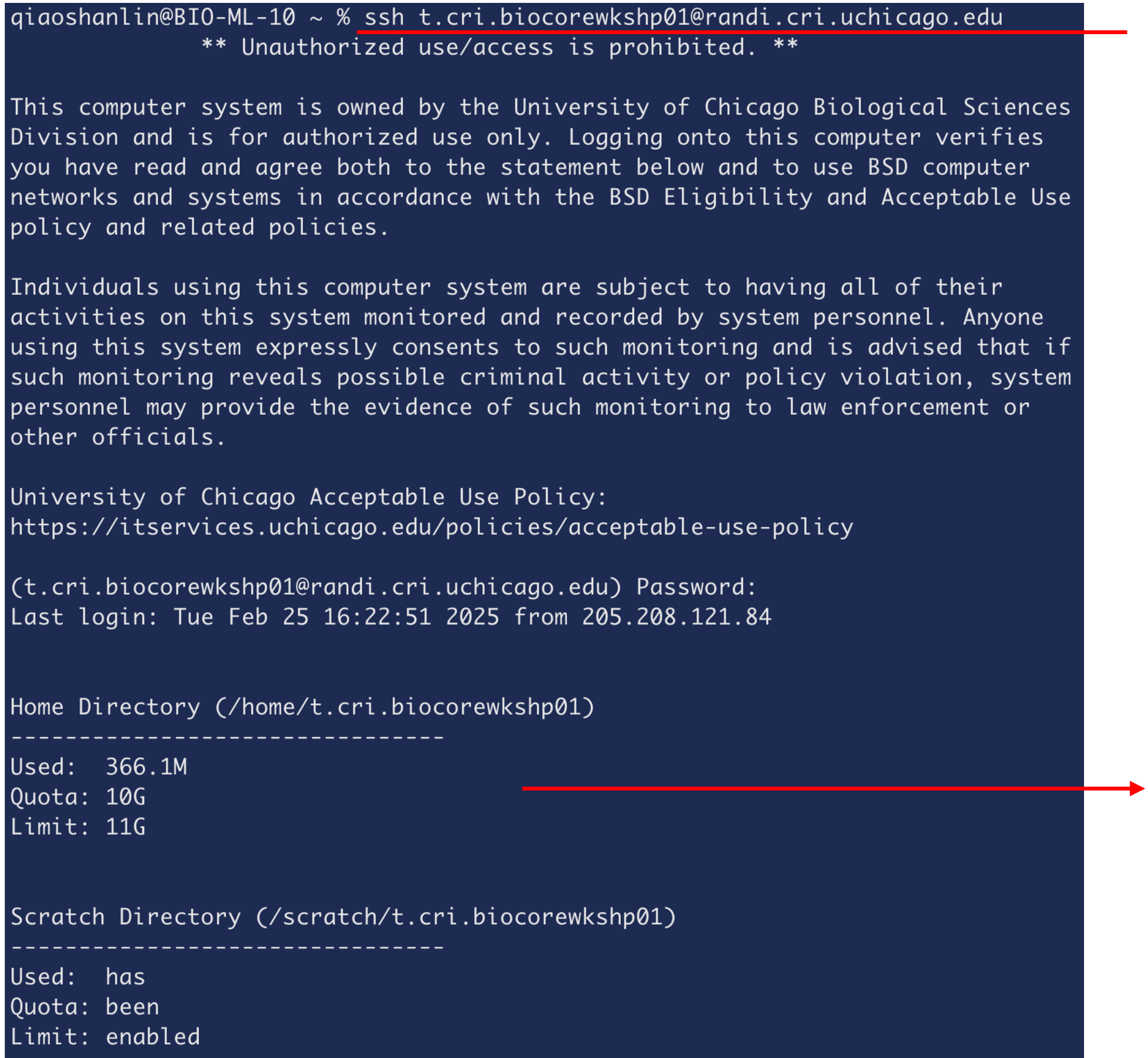
cri.uchicago.edu

Please give us some feedback!

https://mycri.cri.uchicago.edu/educations/trainings/75/survey/

Run Nextflow RNAseq Pipeline on Randi

If using Terminal:



```
qiaoshanlin@BIO-ML-10 ~ % ssh t.cri.biocorewkshp01@randi.cri.uchicago.edu
              ** Unauthorized use/access is prohibited. **

This computer system is owned by the University of Chicago Biological Sciences
Division and is for authorized use only. Logging onto this computer verifies
you have read and agree both to the statement below and to use BSD computer
networks and systems in accordance with the BSD Eligibility and Acceptable Use
policy and related policies.

Individuals using this computer system are subject to having all of their
activities on this system monitored and recorded by system personnel. Anyone
using this system expressly consents to such monitoring and is advised that if
such monitoring reveals possible criminal activity or policy violation, system
personnel may provide the evidence of such monitoring to law enforcement or
other officials.

University of Chicago Acceptable Use Policy:
https://itservices.uchicago.edu/policies/acceptable-use-policy

(t.cri.biocorewkshp01@randi.cri.uchicago.edu) Password:
Last login: Tue Feb 25 16:22:51 2025 from 205.208.121.84


Home Directory (/home/t.cri.biocorewkshp01)
--------------------------------
Used:  366.1M
Quota: 10G
Limit: 11G


Scratch Directory (/scratch/t.cri.biocorewkshp01)
--------------------------------
Used:  has
Quota: been
Limit: enabled
```

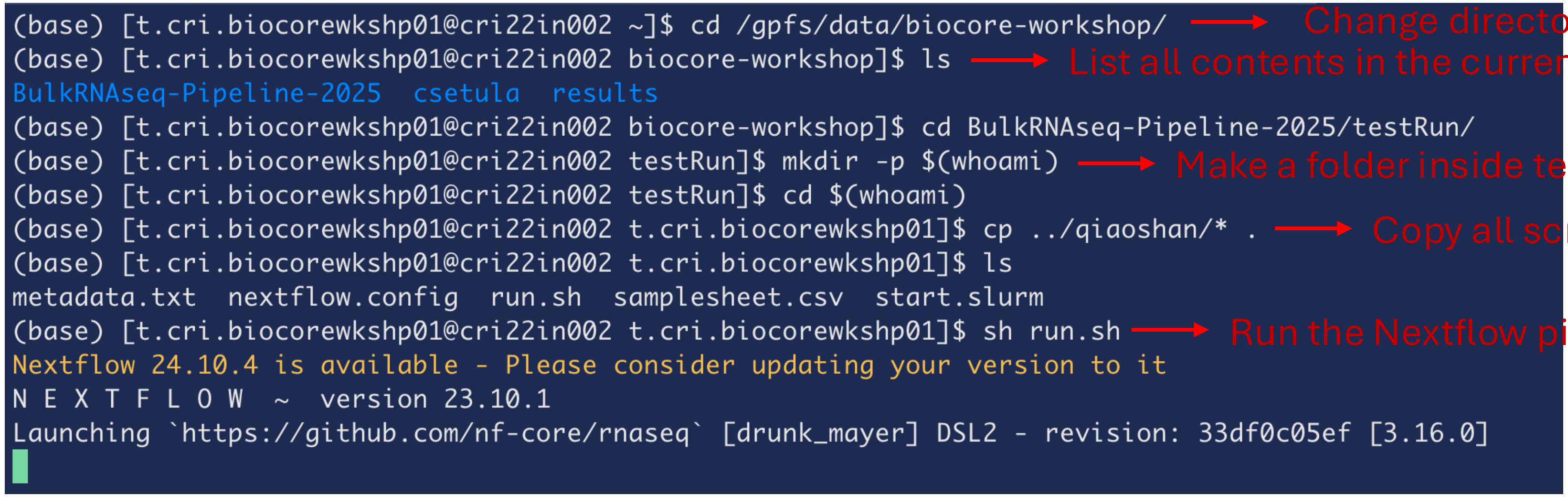
ssh your_account_name@randi.cri.uchicago.edu

Enter password and you will be logged in

Every account has a 11G limit in the home directory

Run Nextflow RNAseq Pipeline on Randi

If using Terminal:



```
(base) [t.cri.biocorewkshp01@cri22in002 ~]$ cd /gpfs/data/biocore-workshop/
(base) [t.cri.biocorewkshp01@cri22in002 biocore-workshop]$ ls
BulkRNAseq-Pipeline-2025  csetula  results
(base) [t.cri.biocorewkshp01@cri22in002 biocore-workshop]$ cd BulkRNAseq-Pipeline-2025/testRun/
(base) [t.cri.biocorewkshp01@cri22in002 testRun]$ mkdir -p $(whoami)
(base) [t.cri.biocorewkshp01@cri22in002 testRun]$ cd $(whoami)
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ cp ../qiaoshan/* .
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ ls
metadata.txt  nextflow.config  run.sh  samplesheet.csv  start.slurm
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ sh run.sh
Nextflow 24.10.4 is available - Please consider updating your version to it
N E X T F L O W  ~  version 23.10.1
Launching `https://github.com/nf-core/rnaseq` [drunk_mayer] DSL2 - revision: 33df0c05ef [3.16.0]
```

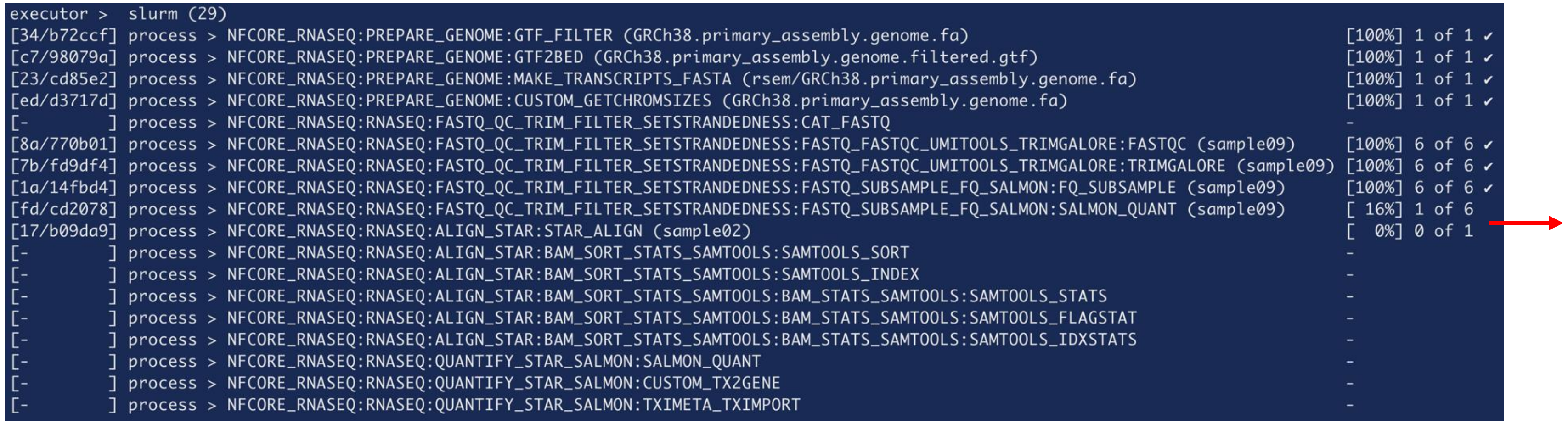
→ Change directory to biocore-workshop

→ List all contents in the current directory

→ Make a folder inside testRun with your account name

→ Copy all scripts from qiaoshan

→ Run the Nextflow pipeline

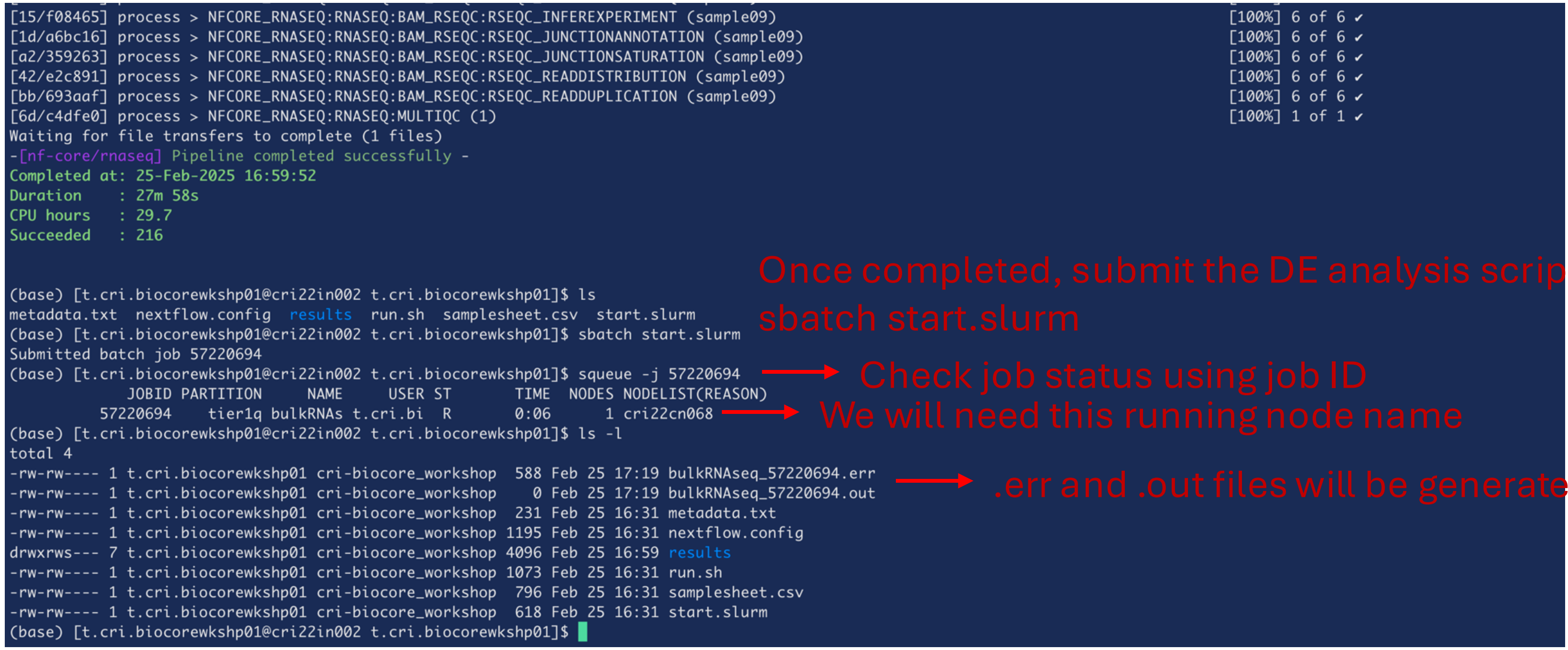


```
executor >  slurm (29)
[34/b72ccf] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF_FILTER (GRCh38.primary_assembly.genome.fa)                                      [100%] 1 of 1 ✔
[c7/98079a] process > NFCORE_RNASEQ:PREPARE_GENOME:GTF2BED (GRCh38.primary_assembly.genome.filtered.gtf)                               [100%] 1 of 1 ✔
[23/cd85e2] process > NFCORE_RNASEQ:PREPARE_GENOME:MAKE_TRANSCRIPTS_FASTA (rsem/GRCh38.primary_assembly.genome.fa)                     [100%] 1 of 1 ✔
[ed/d3717d] process > NFCORE_RNASEQ:PREPARE_GENOME:CUSTOM_GETCHROMSIZES (GRCh38.primary_assembly.genome.fa)                            [100%] 1 of 1 ✔
[-        ] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TRIM_FILTER_SETSTRANDEDNESS:CAT_FASTQ                                              -
[8a/770b01] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TRIM_FILTER_SETSTRANDEDNESS:FASTQ_FASTQC_UMITOOLS_TRIMGALORE:FASTQC (sample09)     [100%] 6 of 6 ✔
[7b/fd9df4] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TRIM_FILTER_SETSTRANDEDNESS:FASTQ_FASTQC_UMITOOLS_TRIMGALORE:TRIMGALORE (sample09) [100%] 6 of 6 ✔
[1a/14fbd4] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TRIM_FILTER_SETSTRANDEDNESS:FASTQ_SUBSAMPLE_FQ_SALMON:FQ_SUBSAMPLE (sample09)      [100%] 6 of 6 ✔
[fd/cd2078] process > NFCORE_RNASEQ:RNASEQ:FASTQ_QC_TRIM_FILTER_SETSTRANDEDNESS:FASTQ_SUBSAMPLE_FQ_SALMON:SALMON_QUANT (sample09)      [ 16%] 1 of 6
[17/b09da9] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:STAR_ALIGN (sample02)                                                            [  0%] 0 of 1
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS:SAMTOOLS_SORT                                            -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS:SAMTOOLS_INDEX                                           -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_STATS                        -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_FLAGSTAT                     -
[-        ] process > NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_IDXSTATS                     -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_STAR_SALMON:SALMON_QUANT                                                           -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_STAR_SALMON:CUSTOM_TX2GENE                                                         -
[-        ] process > NFCORE_RNASEQ:RNASEQ:QUANTIFY_STAR_SALMON:TXIMETA_TXIMPORT                                                       -
```

→ Nextflow pipeline is running

Run Nextflow RNAseq Pipeline on Randi

If using Terminal:



```
[15/f08465] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:RSEQC_INFEREXPERIMENT (sample09)                           [100%] 6 of 6 ✔
[1d/a6bc16] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:RSEQC_JUNCTIONANNOTATION (sample09)                        [100%] 6 of 6 ✔
[a2/359263] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:RSEQC_JUNCTIONSATURATION (sample09)                        [100%] 6 of 6 ✔
[42/e2c891] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:RSEQC_READDISTRIBUTION (sample09)                          [100%] 6 of 6 ✔
[bb/693aaf] process > NFCORE_RNASEQ:RNASEQ:BAM_RSEQC:RSEQC_READDUPLICATION (sample09)                           [100%] 6 of 6 ✔
[6d/c4dfe0] process > NFCORE_RNASEQ:RNASEQ:MULTIQC (1)                                                          [100%] 1 of 1 ✔
Waiting for file transfers to complete (1 files)
-[nf-core/rnaseq] Pipeline completed successfully -
Completed at: 25-Feb-2025 16:59:52
Duration    : 27m 58s
CPU hours   : 29.7
Succeeded   : 216


(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ ls
metadata.txt  nextflow.config  results  run.sh  samplesheet.csv  start.slurm
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ sbatch start.slurm
Submitted batch job 57220694
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ squeue -j 57220694
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
          57220694    tier1q bulkRNAs t.cri.bi  R       0:06      1 cri22cn068
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$ ls -l
total 4
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop  588 Feb 25 17:19 bulkRNAseq_57220694.err
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop    0 Feb 25 17:19 bulkRNAseq_57220694.out
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop  231 Feb 25 16:31 metadata.txt
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop 1195 Feb 25 16:31 nextflow.config
drwxrws--- 7 t.cri.biocorewkshp01 cri-biocore_workshop 4096 Feb 25 16:59 results
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop 1073 Feb 25 16:31 run.sh
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop  796 Feb 25 16:31 samplesheet.csv
-rw-rw---- 1 t.cri.biocorewkshp01 cri-biocore_workshop  618 Feb 25 16:31 start.slurm
(base) [t.cri.biocorewkshp01@cri22in002 t.cri.biocorewkshp01]$
```

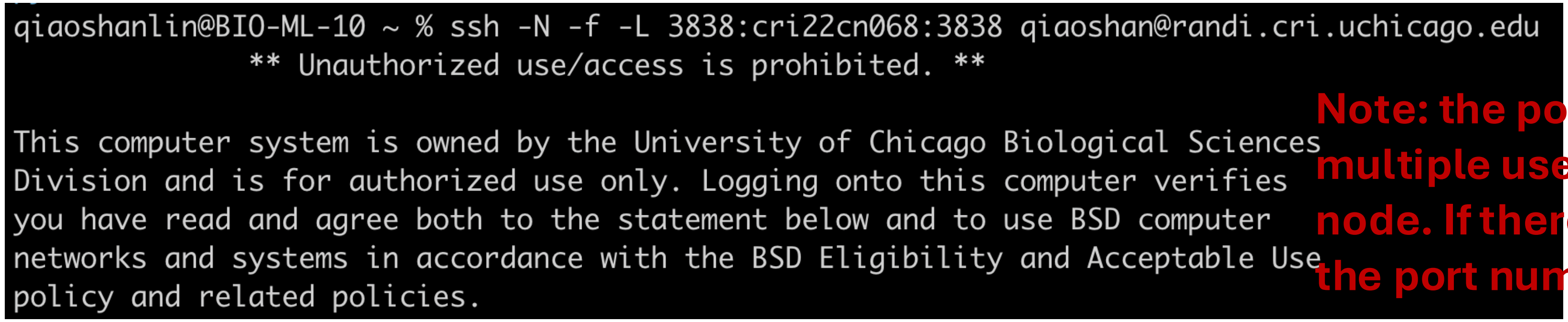
Once completed, submit the DE analysis script by sbatch start.slurm

Check job status using job ID

We will need this running node name

.err and .out files will be generated when the job starts running

Open a new terminal:



```
qiaoshanlin@BIO-ML-10 ~ % ssh -N -f -L 3838:cri22cn068:3838 qiaoshan@randi.cri.uchicago.edu
                 ** Unauthorized use/access is prohibited. **

This computer system is owned by the University of Chicago Biological Sciences
Division and is for authorized use only. Logging onto this computer verifies
you have read and agree both to the statement below and to use BSD computer
networks and systems in accordance with the BSD Eligibility and Acceptable Use
policy and related policies.
```

Forward local port to remote port

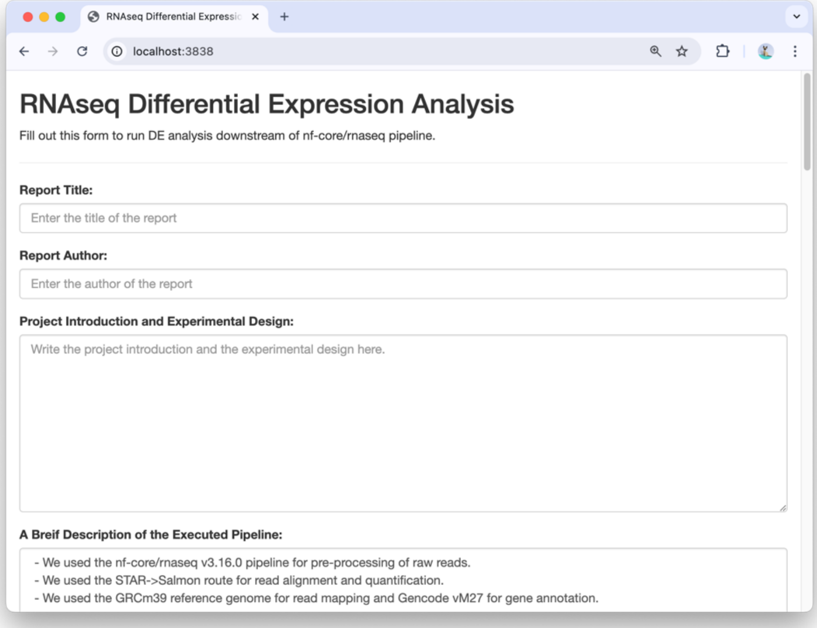
Note: the port might be conflicted when multiple users are running on the same node. If there shows an error, try to change the port number and restart the job.

Run Nextflow RNAseq Pipeline on Randi

If using Terminal:

Finally, open http://localhost:3838/ in a browser (need to wait until multiqc is done)

And you should be able to see an interface like :



RNAseq Differential Expression Analysis

Fill out this form to run DE analysis downstream of nf-core/rnaseq pipeline.

Report Title:

Enter the title of the report

Report Author:

Enter the author of the report

Project Introduction and Experimental Design:

Write the project introduction and the experimental design here.

A Breif Description of the Executed Pipeline:

- We used the nf-core/rnaseq v3.16.0 pipeline for pre-processing of raw reads.
- We used the STAR->Salmon route for read alignment and quantification.
- We used the GRCm39 reference genome for read mapping and Gencode vM27 for gene annotation.

To see the results, you can follow this webpage:
https://uchicago.service-now.com/kb_view.do?sys_kb_id=27ffbebb97fc025423087cf11153af5b
or run the following command in your local terminal to download everything:
scp -r XXX@randi.cri.uchicago.edu:/gpfs/data/biocore-workshop/BulkRNAseq-Pipeline-2025/testRun/**XXX** **~/Document/bulkRNAseq_test**
(remember to replace **XXX** with your account name; you can also change **~/Document/bulkRNAseq_test** to some other folders on your computer)

For more troubleshooting, please come to our office hour