

# Running ImmunoVerse on CGC

Guangyuan(Frank) Li, PhD  
NYU Grossman School of Medicine  
Last update: 2024.02.05

# Overview

1. Without the need of writing single line of code, you can simply upload your RNA-Seq fastq.gz files, along with immunopeptidome raw files (if applicable) to the CGC platform, to (a) **profile genetic aberrations**, (b) **generate search space**, (c) **search immunopeptidome**, (d) more to come
2. Following sections (please always read **a** and **b** first):
  - a. [Upload and download](#)
  - b. [Gene expression](#)
  - c. [Alignment](#)
  - d. [Splicing and intron retention](#) (require c, or provide your own bam)
  - e. [Pathogen](#) (require c, or provide your own bam)
  - f. [Transposable element](#) (require c, or provide your own bam)
  - g. [Variants](#) (require c, or provide your own bam)
  - h. [Gene fusion](#)
  - i. [HLA typing](#)
  - j. [summarization](#)
  - k. [Immunopeptidome analysis](#)

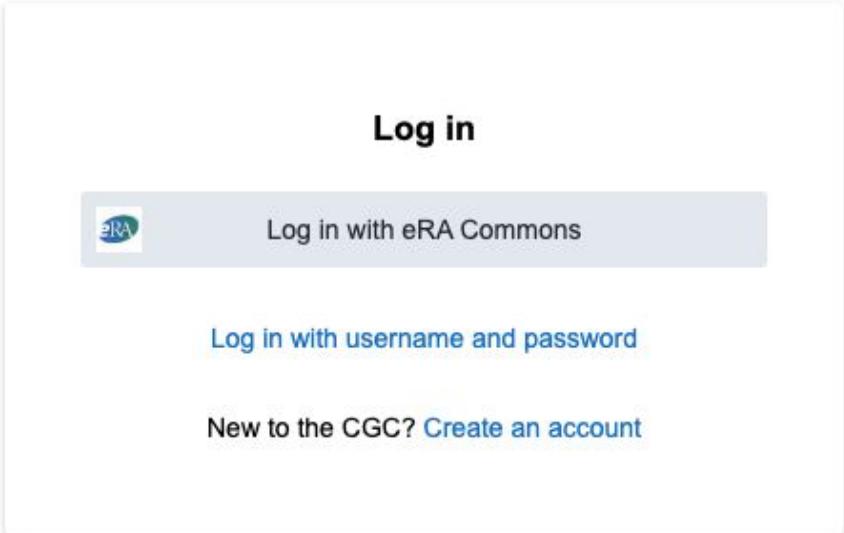
Upload and download

# Step 1: Creating Account

1. Visit <https://cgc.sbggenomics.com/home>
2. eRA common should work just fine, otherwise, please follow the links to create an account



CANCER GENOMICS CLOUD  
SEVEN BRIDGES



The image shows the login interface for the Cancer Genomics Cloud. At the top center is a large "Log in" button. Below it is a "Log in with eRA Commons" button, which features the eRA logo (a blue square with white text) and the text "Log in with eRA Commons". Further down is a "Log in with username and password" link. At the bottom left is a "New to the CGC? Create an account" link.

# Step 2: Create a project

The screenshot shows the CGC interface with a 'Create a project' dialog box open over a list of existing projects.

**Main Dashboard (Left):**

- Projects list:
  - deploy\_pipelines
  - CONTROLLED NeoVerse\_pancancer\_2
  - NeoVerse\_development\_project
  - CONTROLLED NeoVerse\_pancancer
  - Demonstration: Building an App
- Buttons: '+ Create Project' and 'View all projects'
- Section: 'Public Data and Apps'
- Section: 'Analyze'
- Text: '331061 publicly available files'

**Create a project Dialog (Right):**

- Name:** Input field with placeholder 'Name'. A red arrow points to it with the text 'Name it'.
- URL:** Input field with placeholder 'https://cgc.sbggenomics.com/u/ii2g2uc/'.
- Controlled Status:** A checkbox labeled 'CONTROLLED' with a note: 'This project will contain controlled data.' A red arrow points to the checkbox.
- General Information Tab:** Active tab.
- Advanced Settings Tab:** Not active.
- Billing Group:** 'Pilot Funds (ii2g2uc)' dropdown. A red arrow points to it.
- Location:** 'AWS (us-east-1)' dropdown.
- Execution Settings:**
  - Spot Instances:** A checkbox with a note: 'Spot instances can significantly reduce the cost of your task execution if results are not needed urgently.' A red arrow points to the checkbox.
  - Reuse:** A checkbox with a note: 'Automatic reuse of precomputed results can significantly reduce the time and cost of your task execution.'
- Buttons:** 'Cancel' and 'Create'.

**Annotations:**

- 'unclick' is written in red over several project names in the list.
- 'unclick' is written in red over the 'Spot Instances' checkbox in the dialog.
- A large red arrow points from the 'Name it' annotation to the 'Name' input field.
- A red arrow points from the 'Billing account' annotation to the 'Billing group' dropdown.
- A red arrow points from the 'Billing account' annotation to the explanatory text below the 'Billing group' dropdown.

# Step 2: Create a project (cont'd)

deploy\_pipelines  
Created by:li2g2uc · Jan 13, 2025, 10:37

**CONTROLLED** NeoVerse\_pancancer\_2  
Created by:li2g2uc · Sep 28, 2024, 23:01

NeoVerse\_development\_project  
Created by:li2g2uc · May 15, 2024, 12:15

**CONTROLLED** NeoVerse\_pancancer  
NeoVerse  
Created by:li2g2uc · Mar 15, 2024, 16:33

Demonstration: Building an App  
Created by:rowan\_beck\_era · Jul 18, 2023, 13:06

[+ Create Project](#) [View all projects](#)

Public Data and Apps

Analyze

**331061**

### Create a project

Name

**CONTROLLED** This project will contain controlled data.  ⓘ

[General information](#) [Advanced settings](#)

**Network Access settings ⓘ**

Block network access  
Execution within the project won't have network access.

Allow network access Allow network Allow network  
Execution will have unrestricted network access.

**Download Restriction settings ⓘ**

⚠ Download Restriction settings cannot be modified after the project has been created.

Download restriction will be applied to all the files that are imported to the project.

No

[Cancel](#) [Create](#)

# Step 3: navigate your project

The screenshot shows the ImmunoVerse project interface. On the left, the 'Overview' section displays a list of tools and pipelines:

- All the files
- ImmunoVerse Overview
- Apps built from docker containers
- Users upload the RNA-Seq raw data, and immunopeptidome data if available, and use the following pre-built pipelines to (1) Profile all genetic aberrations, (2) Generate sample-specific search space, (3) Run immunopeptidome search. You don't need to write any code to finish all the analysis (amazing isn't it). We believe separating each module can maximize the flexibility in real-world scenarios, we are also working on connecting all dots into a single workflow in the future.
- Tools available
  - Gene Expression Pipeline (all samples keeping pair order, 10GB RAM each done in 30min)
  - Alignment Pipeline (all samples keeping pair order, 30GB RAM each done in 5h)
  - Splicing Intron Pipeline (all samples, 2GB RAM each done in 30min)
  - Transposable Element Pipeline (all samples, 10GB RAM each in 5h)
  - Pathogen Pipeline (all samples, 100GB RAM each done in 10min)
  - Variant Pipeline
    - RNA Variant Pipeline (all samples, 2GB RAM each done in 90min)
    - VEP pipeline (batch by sample, 5GB RAM each done in 30min)
  - Gene Fusion Pipeline (batch by both sample and pair-end)
  - HLA type pipeline
    - decompress pipeline (batch by file, 2GB RAM each done in 5min)
    - optype pipeline (batch by sample and pair-end, 20GB RAM each done in 20min)
  - Summarization Pipeline
  - (optional) bam\_to\_fastq pipeline
  - (optional) circular RNA pipeline
  - Immunopeptidome Pipeline
    - MaxQuant Pipeline
    - mconvert pipeline
    - Rescore pipeline
    - HLA binding pipeline

# Step 4: upload files

The screenshot shows a user interface for managing files in a cloud storage system. At the top, there's a navigation bar with links for Dashboard, Files (PREMIUM), Apps, Tasks, Data Studio, and Interactive Apps. Below the navigation is a search bar and filter options for Extension, Tags, and Paired-end. A table lists several files with columns for Name, Size, Extension, and Task ID.

Name	Size	Extension	Task ID
ensembl_protein.fasta	11.36 MiB	FASTA	-
gencode.v36.annotation.gtf	1.29 GiB	GTF	-
gene_model.txt	52.57 MiB	TXT	-
GRCh38.d1.vd1.fa	2.94 GiB	FA	-
GRCh38.d1.vd1.gencode.v36.annotation.star-fusion-1.12.0-CTAT-index-archive.tar	44.87 GiB	TAR	12f76e64-e489-...
hg19_maxquant_combined_txt	-	-	6ea7ebaa-448b-...

A red arrow points from the text "Easiest way, just drag the file from your computer" to the "+ Add files" button in a dropdown menu. The menu also includes options like "New Folder", "Your Computer", "FTP / HTTP", "GA4GH Data Repository Service (DRS)", "Data Tools", "Volumes", and "Import from a manifest file".

Easiest way, just drag the  
file from your computer

+ Add files

New Folder

Your Computer

FTP / HTTP

GA4GH Data Repository Service (DRS)

Data Tools

Volumes

Import from a manifest file

# Step 4: upload files (cont'd)

Add files to "deploy\_pipelines"

You can use FTP/HTTP,  
just paste the file URL, no  
folder, paste multiple files'  
URL instead



Import from an FTP or HTTP(S) server: ?

Paste the link of the file(s) you want to import

`https://genome.med.nyu.edu/public/yarmarkovichlab/ImmunoVerse/normal/  
normal_intron.txt`

or  on your computer containing the links

Add tags

Resolve naming conflicts:

Skip

Import

# Step 5: download

Screenshot of a cloud storage interface showing a file list and download options.

The interface includes a top navigation bar with links: Dashboard, Files PREMIUM, Apps, Tasks, Data Studio, Interactive Apps. The title "deploy\_pipelines" is displayed above the file list. On the right, there's an "Interactive Browser" link.

The file list shows three items:

Name	Size	Extension	Task ID	Created on
ensembl_protein.fasta <small>reference</small>	11.36 MiB	FASTA	-	Jan. 13, 2025 10:...
gencode.v36.annotation.gtf <small>reference</small>	1.29 GiB	GTF	-	Jan. 13, 2025 10:...
gene_model.txt <small>reference</small>	52.57 MiB	TXT	-	Jan. 13, 2025 10:...

A red arrow points from the word "Select" to the checkbox next to the first file, "ensembl\_protein.fasta". Another red arrow points from the word "download" to the "Download" button in the toolbar.

Toolbar buttons include: Copy, Rename, Move, Metadata, Edit tags, Download, and a three-dot menu.

# Step 6: copy the needed reference files from my project

The screenshot shows the Immuniverse platform interface. At the top, there is a navigation bar with links: Home, Projects, Data, Public Apps, Public Projects, and Developer. Below the navigation bar is a secondary menu with Dashboard, Files PREMIUM (which is highlighted), Apps, Tasks, Data Studio, and Interactive Apps. In the center, the title "test\_immunoverse" is displayed with an info icon. To the right, there are links for Interactive Browsers, Settings, and Notes. On the far right of the top bar are a bell icon and a user profile icon labeled "GL".

The main content area is titled "Root". On the right side of this area, there is a "New Folder" button and an "Add files" button. A red arrow points to the "Projects" option in a dropdown menu that has opened. The dropdown menu also includes "Public Files", "Your Computer", "FTP / HTTP", "GA4GH Data Repository Service (DRS)", "Data Tools", "Volumes", and "Import from a manifest file".

A large watermark in the background of the main content area reads "Files are the basis of every analysis." There is also a smaller watermark at the bottom left.

At the bottom of the main content area, there is a link: "learn more about different ways to add files."

# Step 6: copy the needed reference files from my project (cont'd)

Add files to "test\_immunoverse"

Please contact me if you don't have access to this project

Search for project: deploy\_pipelines

Copy to Project

Search Extension Tags Paired-end + Clear filters

Name	Size	Ext
ensembl_protein.fasta	11.36 MiB	FASTA
gencode.v36.annotation.gtf	1.29 GiB	GTF
gene_model.txt	52.57 MiB	TXT
GRCh38.d1.vd1.fa	2.94 GiB	FA
GRCh38.d1.vd1.gencode.v36.annotation.s	44.87 GiB	TAR
hg19_maxquant_combined_txt	-	-
hg38.fa	3.05 GiB	FA
hg38.knownGene.gtf	564.57 MiB	GTF
hg38_rmsk_TE.gtf.loclnd	883.82 MiB	LOCIN

Tags filter dialog:

- No value
- ImmunoVerse\_data
- input
- kraken2\_db
- reference
- star\_hg38\_index

Clear selected Apply

## Step 6: copy the needed reference files from my project (cont'd)

Add files to "test\_immuneverse"

Search for project: deploy\_pipelines

Root Exclude subfolders

Search: Extension: Tags: reference Paired-end + Clear filters

Copy to Project

63 items

Name	Path	Size
annot.renamed.txt.gz	Files / ImmunoV...	10.08
reference ImmunoVerse_data		
annot.renamed.txt.gz.tbi	Files / ImmunoV...	2.74 M
reference ImmunoVerse_data		
blacklist_splicing.txt	Files / ImmunoV...	25.59
reference ImmunoVerse_data		
canonical.txt	Files / ImmunoV...	2.50 M
reference ImmunoVerse_data		
chrLength.txt	Files / star_hg3...	13.83
reference star_hg38_index		
chrName.txt	Files / star_hg3...	65.86
reference star_hg38_index		
chrNameLength.txt	Files / star_hg3...	79.69
reference star_hg38_index		
chrStart.txt	Files / star_hg3...	29.85
reference star_hg38_index		
contigs.txt	Files / ImmunoV...	136.68
reference ImmunoVerse_data		
cosmic_prelift.bed.gz	Files / ImmunoV...	1.63 M
reference ImmunoVerse_data		
cosmic_prelift.bed.gz.tbi	Files / ImmunoV...	43.22
reference ImmunoVerse_data		
ensembl_protein.fasta	Files	11.36 M
reference		

Demonstration: Building an App

NeoVerse\_pancancer\_2

NeoVerse\_development\_proj...

NeoVerse\_pancancer

ImmunoVerse

# Step 7: organize a bit for ImmunoVerse\_data

The screenshot shows a cloud storage interface with a dark theme. At the top, there's a navigation bar with links like Home, Projects, Data, Public Apps, Public Projects, and Developer. Below the navigation bar is a header for the project "test\_immunoverse". The main area shows a list of files and folders on the left and a detailed view of a selected folder on the right.

**Left Panel (File List):**

- Root
- Search bar
- Extension filter dropdown
- Tags filter dropdown
- Add files button
- More options button

**Right Panel (Selected Folder View):**

Extension	Task ID	Create
FASTA	-	Feb. 06
FASTA	-	Feb. 06
FA	-	Feb. 06
K2D	-	Feb. 06
TXT	-	Feb. 06
TXT	-	Feb. 06
-	-	Feb. 06
-	-	Feb. 06
TXT	-	Feb. 06
156 bytes	TXT	Feb. 06
323.92 MiB	BED	Feb. 06
11.36 MiB	FASTA	Feb. 06
63 items		

**Modal Dialog (Create New Folder):**

**Instructions:** Folders can't be subsequently renamed.

**Fields:**

- Name: ImmunoVerse\_data (highlighted by a red arrow)
- Path: Files /

**Buttons:**

- Cancel
- Create (highlighted by a red arrow)

# Step 7: organize a bit for ImmunoVerse\_data

The screenshot shows the Nextflow interface with a search filter overlay. The filter is set to 'ImmunoVerse\_data' and has a green 'Apply' button at the bottom.

Search results:

Name	Size	Extension	Task ID	Create
ImmunoVerse_data	-	-	-	Feb. 06
_2_ensembl_protein.fasta	11.36 MiB	FASTA	-	Feb. 06
reference ImmunoVerse_data	11.36 MiB	FASTA	-	Feb. 06
_1_ensembl_protein.fasta	3.05 GiB	FA	-	Feb. 06
reference	64 bytes	K2D	-	Feb. 06
_1_hg38.fa	1.37 GiB	TXT	-	Feb. 06
reference ImmunoVerse_data	96.94 MiB	TXT	-	Feb. 06
normal_intron.txt	23.20 GiB	-	-	Feb. 06
reference ImmunoVerse_data	1.46 GiB	-	-	Feb. 06
SA	156 bytes	TXT	-	Feb. 06
reference star_hg38_index	323.92 MiB	BED	-	Feb. 06
SAindex	reference star_hg38_index			
splice_erv_db.txt	tcga_tmp_prelift.bed			
reference ImmunoVerse_data				
reference ImmunoVerse_data				

64 items

# Step 7: organize a bit for ImmunoVerse\_data

The screenshot shows a file management interface with a 'Move' dialog box open. The dialog box is titled 'Move' and asks 'Move 31 selected items?'. It contains a tree view under 'Files' with a single item selected: 'ImmunoVerse\_data'. A red arrow points to this selected item. Below the tree view are sections for 'Tags' (with a checked checkbox for 'Keep preexisting tags') and 'Add new tags by separating them with enter key' (an empty input field). At the bottom left is a 'New folder' button, and at the bottom right are 'Cancel' and 'Move' buttons, with the 'Move' button highlighted by a red arrow. The background shows a list of 31 selected items, including various FASTA, TXT, and BED files, along with their sizes and extensions. The top navigation bar includes 'Home', 'Projects', 'Data', 'Public Apps', 'Public Projects', 'Developer', 'Dashboard', 'Files PREMIUM', 'Apps', 'Tasks', 'Data Studio', 'Interactive Apps', 'test\_immuneVerse', 'Interactive Browsers', 'Settings', and 'Notes'.

Size	Extension	Task
11.36 MiB	FASTA	-
1.37 GiB	TXT	-
96.94 MiB	TXT	-
156 bytes	TXT	-
323.92 MiB	BED	-
16.84 MiB	BED	-
73 bytes	TXT	-
43.22 KiB	TBI	-
75 bytes	TXT	-
10.08 GiB	TXT.GZ	-
136.68 KiB	TXT	-

# Step 7: organize a bit for ImmunoVerse\_data

The screenshot shows the ImmunoVerse Data Management interface. At the top, there's a navigation bar with links for Home, Projects, Data, Public Apps, Public Projects, and Developer. A user icon 'GL' is in the top right. Below the navigation is a secondary header with links for Dashboard, Files, Apps, Tasks, Data Studio, and Interactive Apps. The main title 'test\_immunoverse' is displayed. On the left, a sidebar shows 'Files > ImmunoVerse\_data'. In the center, a file named '\_2\_ensembl\_protein.fasta' is selected. A red box highlights the file name, and two red arrows point to the dropdown arrow and the close ('X') button next to it. Below the file name, there are tabs for REFERENCE and IMMUNOVERSE DATA. A status bar at the bottom indicates the file size (11.4 MB), creation date (February 6, 2025), modification date (February 6, 2025), and host location (AWS us-east-1). A red annotation on the right side of the screen reads: 'Click the \_2\_ensembl\_protein.fasta, you can see a rename icon, click it and rename it a bit'. The interface includes sections for File (Experimental strategy, Library ID, Platform, Platform unit ID, File segment number, Quality scale, Paired-end, Reference genome) and General (Investigation).

Click the \_2\_ensembl\_protein.fasta, you  
can see a rename icon, click it and  
rename it a bit

# Step 8: organize a bit for star\_hg38\_index

The screenshot shows a file management interface with a 'Move' dialog box open. The dialog box is titled 'Move' and asks 'Move 16 selected items?'. It contains a tree view under 'Files' with several items, including 'ImmunoVerse\_data' and 'star\_hg38\_index'. A red arrow points to the 'star\_hg38\_index' item. Below the tree view is a 'Tags' section with a checked checkbox for 'Keep preexisting tags' and a text input field for adding new tags. At the bottom of the dialog box are 'New folder', 'Cancel', and 'Move' buttons. The 'Move' button is highlighted with a red arrow. In the background, the main interface shows a list of files with columns for 'Size', 'Extension', and 'Task II'. A red arrow also points to the 'star\_hg38\_index' file in this list.

Move 16 selected items?

Files

- ImmunoVerse\_data
- star\_hg38\_index

Tags

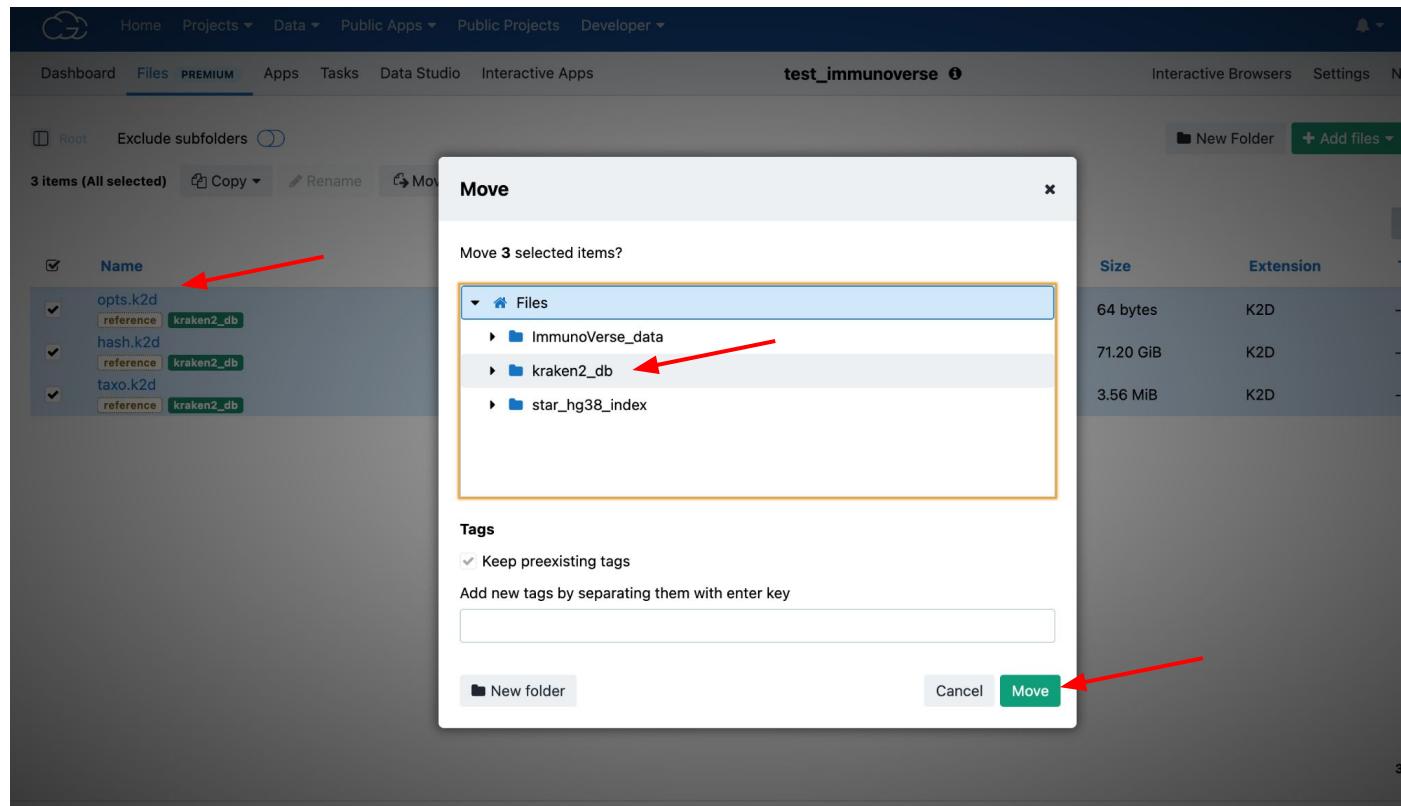
Keep preexisting tags

Add new tags by separating them with enter key

New folder Cancel Move

Size	Extension	Task II
23.20 GiB	-	-
1.46 GiB	-	-
29.85 KiB	TXT	-
14.88 MiB	TAB	-
10.83 MiB	TXT	-
79.69 KiB	TXT	-
972 bytes	TXT	-
3.63 GiB	-	-
467.08 KiB	OUT	-
65.86 KiB	TXT	-
48.57 MiB	TAB	-

# Step 9: organize a bit for kraken2\_db



# Step 10: final clean of files

Screenshot of a file management interface showing a list of selected items for deletion.

Selected items:

- kraken2\_db
- star\_hg38\_index
- ImmunoVerse\_data
- \_1\_ensembl\_protein.fasta** (marked for delete)
- \_1\_hg38.fa** (marked for rename)
- ensembl\_protein.fasta
- gencode.v36.annotation.gtf
- GRCh38.d1.vd1.fa

Action buttons at the top:

- Root
- New Folder
- Add files
- ...

Table of selected items:

Name	Size	Extension	Task ID	Create
kraken2_db	-	-	-	Feb. 06
star_hg38_index	-	-	-	Feb. 06
ImmunoVerse_data	-	-	-	Feb. 06
<b>_1_ensembl_protein.fasta</b>	11.36 MiB	FASTA	-	Feb. 06
<b>_1_hg38.fa</b>	3.05 GiB	FA	-	Feb. 06
ensembl_protein.fasta	11.36 MiB	FASTA	-	Feb. 06
gencode.v36.annotation.gtf	1.29 GiB	GTF	-	Feb. 06
GRCh38.d1.vd1.fa	2.94 GiB	FA	-	Feb. 06

Annotations:

- A red arrow points to the row for **\_1\_ensembl\_protein.fasta** with the text "delete".
- A red arrow points to the row for **\_1\_hg38.fa** with the text "Rename to hg38.fa".

# Step 10: final clean of files (done)

Dashboard Files PREMIUM Apps Tasks Data Studio Interactive Apps test\_immunoverse ⓘ Interactive Browsers Settings Notes

Root

15 items (All selected) Copy Rename Move Metadata Edit tags Download ...

New Folder + Add files ...

<input checked="" type="checkbox"/>	Name	Size	Extension	Task ID	Create
<input checked="" type="checkbox"/>	kraken2_db	-	-	-	Feb. 06
<input checked="" type="checkbox"/>	star_hg38_index	-	-	-	Feb. 06
<input checked="" type="checkbox"/>	ImmunoVerse_data	-	-	-	Feb. 06
<input checked="" type="checkbox"/>	hg38.fa	3.05 GiB	FA	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	ensembl_protein.fasta	11.36 MiB	FASTA	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	gencode.v36.annotation.gtf	1.29 GiB	GTF	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	GRCh38.d1.vd1.fa	2.94 GiB	FA	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	GRCh38.d1.vd1.gencode.v36.annotation.star-fusion-1.12.0-CTAT-index-archive.tar	44.87 GiB	TAR	12f76e64-e489...	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	uniprot_reviewed_curated_addition.fasta	14.20 MiB	FASTA	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	homo_sapiens_vep_112_GRCh38.tar.gz	25.46 GiB	TAR.GZ	-	Feb. 06
<input checked="" type="checkbox"/>	reference				
<input checked="" type="checkbox"/>	hg38_rmsk_TE.gtf.locind	883.82 MiB	LOCIND	-	Feb. 06
<input checked="" type="checkbox"/>	reference				

15 items

# Gene expression

Home Projects ▾ Data ▾ Public Apps ▾ Public Projects Developer ▾

GL

Dashboard Files PREMIUM Apps Tasks Data Studio Interactive Apps

test\_immunoverse ⓘ

Interactive Browsers Settings Notes

Search names and description Category: All ▾ Toolkit: All ▾ Language and version: All ▾ Status: Available ▾ Cost Estimator: All ▾

Create app + Add apps ▾

No apps

No apps with the given search term can be found.

Clear filters

- Public apps
- Projects
- My created apps

# Gene expression

Add apps to test\_immunoverse

Search Search names and description Category: All ▾ Toolkit: All ▾ Language and version: All ▾

Name	Type	Modified by	Modified on	⋮
deploy_pipelines	Tool	li2g2uc	Feb 06, 2025 13:31	
NeoVerse_pancancer_2	↳ hla_binding_pipeline	Tool	li2g2uc	
NeoVerse_development_project	↳ rescore_pipeline	Tool	li2g2uc	
NeoVerse_pancancer	↳ msconvert_pipeline	Tool	li2g2uc	
Demonstration: Building an App	↳ maxquant_pipeline	Tool	li2g2uc	
	↳ summarization_pipeline	Tool	li2g2uc	
	↳ STAR-Fusion Build Fusio	Tool	li2g2uc	
	↳ SBG Decompressor CWL	Tool	li2g2uc	
	↳ variant_pipeline	Tool	li2g2uc	
	↳ Variant Effect Predictor	Tool	li2g2uc	
	↳ OptiType	Tool	li2g2uc	
	↳ STAR-Fusion	Tool	li2g2uc	
	↳ alignment_pipeline	Tool	li2g2uc	
	↳ gene_pipeline	Tool	li2g2uc	
	↳ splicing_intron_pipeline	Tool	li2g2uc	
	↳ telocal_pipeline	Tool	li2g2uc	

# Gene expression

Dashboard Files PREMIUM Apps Tasks Data Studio Interactive Apps test\_immunoverse ⓘ Interactive Browsers Settings Notes

Search names and description Category: All Toolkit: All Language and version: All Status: Available Cost Estimator: All Create app + Add apps

Name	Type	Source	Workflow L...	Modified ...	Modified ...	⋮
> <a href="#">COPY</a> gene_pipeline	Tool	deploy_pipelines	CWL	li2g2uc	Feb 06, 2025...	 Run 

Showing 1 of 1 < >

# Gene expression

DRAFT gene\_pipeline run - 02-06-25 18:49:10 ⚙

Last update by li2g2uc on Feb. 6, 2025 13:49

App: gene\_pipeline - Revision: 6

Task Inputs Execution Settings

Inputs

- Batching Off
- ensembl\_protein Change selection
  - ensembl\_protein.fasta
- fastq\_gz\_files Change selection
  - HN19-9674\_R1.fastq.gz
  - HN19-9674\_R2.fastq.gz
  - HN20-9844\_R1.fastq.gz
  - HN20-9844\_R2.fastq.gz
  - HN21-10181\_R1.fastq.gz
  - ...and 1 more item
- kallisto\_index Change selection
  - kallisto\_index
- nuorf Change selection
  - nuorf.fasta
- uniprot\_isoform Change selection
  - uniprot\_reviewed\_curated\_addition.fasta

App Settings

cores 3

outdir .

strand no

Output Settings

gene_fasta	No value
isoform_fasta	No value
nuorf.fasta	No value
tpm_result	No value

Consistent with number of samples

Remember, select file in the order, sample1\_R1, sample1\_R2, sample2\_R1, sample2\_R2...

# Gene expression

Task Inputs    Execution Settings 

**Spot Instances** Off 

Spot instances can significantly reduce the cost of your task execution if results are not needed urgently.

[Learn more](#)

---

**Memoization (WorkReuse)** Off 

Automatic reuse of precomputed results can significantly reduce the time and cost of your task execution.

[Learn more](#)

---

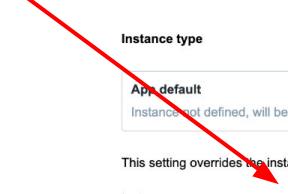
**Elastic Disk BETA** Off 

Automatic extension of attached disk space will allow task execution to continue if the original disk capacity becomes insufficient.

[Learn more](#)

---

**Instance type**

**App default** Instance not defined, will be automatically selected 

**Custom** Select an instance from the list 

This setting overrides the instance set by the app developer and the instance selection from any previous run of this task. [Learn more](#).

Instance: r5.4xlarge (16vCPUs, 128GB RAM)   4096 

Attached storage (GB) 

Price: \$1.008 per hour

**1024GB should be sufficient, but the max you can go to 4096GB**

# Gene expression

Dashboard Files Apps Tasks Data Studio Interactive Apps **deploy\_pipelines** ⓘ Interactive Browsers Settings Notes

DRAFT gene\_pipeline run - 02-06-25 18:49:10 ⚙️ Last update by li2g2uc on Feb. 6, 2025 13:49 App: gene\_pipeline - Revision: 6

Task Inputs Execution Settings

Inputs

- ensembl\_protein
  - ensembl\_protein.fasta
- fastq\_gz\_files
  - HN19-9674\_R1.fastq.gz
  - HN19-9674\_R2.fastq.gz
  - HN20-9844\_R1.fastq.gz
  - HN20-9844\_R2.fastq.gz
  - HN21-10181\_R1.fastq.gz
  - HN21-10181\_R2.fastq.gz
  - ...and 1 more item
- kallisto\_index
  - kallisto\_index
- nuorf
  - nuorf.fasta
- uniprot\_isoform
  - uniprot\_reviewed\_curated\_addition.fasta

App Settings

- cores
- outdir
- strand

Show non-default ⓘ Output Settings

- 3
- gene\_fasta
  - HN19-9674\_gene.fasta
  - HN20-9844\_gene.fasta
  - HN21-10181\_gene.fasta
- isoform\_fasta
  - HN19-9674\_isoform.fasta
  - HN20-9844\_isoform.fasta
  - HN21-10181\_isoform.fasta
- nuorf\_fasta
  - HN19-9674\_nuorf.fasta
  - HN20-9844\_nuorf.fasta
  - HN21-10181\_nuorf.fasta
- tpm\_result
  - HN19-9674\_gene\_tpm.txt
  - HN20-9844\_gene\_tpm.txt
  - HN21-10181\_gene\_tpm.txt

Once finishing setup, click run

Once done, you will have fasta (canonical protein sequences with expressed gene), isoform fasta (isoform protein that are expressed), nuorf fasta (cryptic orfs), tpm (gene to tpm in each sample)

# Alignment

Copy this pipeline

Completed alignment\_pipeline run - 01-25-25 18:19:27

Executed on Jan. 25, 2025 13:20 by li2g2uc

Spot Instances: Off | Memoization: Off | Quality: All | Processor: 3 | Duration: 1 hour 49 minutes

App: alignment\_pipeline - Revision: 3

Inputs

- fastq\_files
  - HN19-9674\_R1.fastq.gz
  - HN19-9674\_R2.fastq.gz
  - HN20-9844\_R1.fastq.gz
  - HN20-9844\_R2.fastq.gz
  - HN21-10181\_R1.fastq.gz
  - ...and 1 more item
- sequence
  - GRCh38.d1.vd1.fa
- star\_index
  - star\_hg38\_index

App Settings

- cores
- outdir

Output Settings

- 3 bai
  - HN19-9674\_secondAligned.sortedByCoord.out.bam.bai
  - HN20-9844\_secondAligned.sortedByCoord.out.bam.bai
  - HN21-10181\_secondAligned.sortedByCoord.out.bam.bai
- bam
  - HN19-9674\_secondAligned.sortedByCoord.out.bam
  - HN20-9844\_secondAligned.sortedByCoord.out.bam
  - HN21-10181\_secondAligned.sortedByCoord.out.bam

Remember, select file in the order, sample1\_R1, sample1\_R2, sample2\_R1, sample2\_R2...

Consistent with number of samples

# Splicing and intron retention

COMPLETED splicing\_intron\_pipeline run - 01-25-25 22:32:13

Get support

View stats & logs

Edit and rerun

Executed on Jan. 25, 2025 17:35 by li2g2uc

Spot Instances: Off

Memoization (WorkReuse): Off

Price: \$1.97

Duration: 41 minutes

App: splicing\_intron\_pipeline - Revision: 0

Each bam take 2GB RAM, you don't need 40 cores,  
in this case, three files I will select 3 cpus, almost  
any instance should work, done in 30 min

## Inputs

### bam\_files

HN19-9674\_secondAligned.sortedByCoord.out.bam  
HN20-9844\_secondAligned.sortedByCoord.out.bam  
HN21-10181\_secondAligned.sortedByCoord.out.bam

### gene\_model

gene\_model.txt

### reference

gencode.v36.annotation.gtf

### sequence

hg38.fa

## App Settings

Show non-default

cores

40

outdir

.

strand

no

## Output Settings

### intron\_peptide

HN19-9674\_secondAligned.sortedByCoord.out\_intron\_peptid...  
HN20-9844\_secondAligned.sortedByCoord.out\_intron\_peptid...  
HN21-10181\_secondAligned.sortedByCoord.out\_intron\_pepti...

### intron\_result

HN19-9674\_secondAligned.sortedByCoord.out\_intron.txt  
HN20-9844\_secondAligned.sortedByCoord.out\_intron.txt  
HN21-10181\_secondAligned.sortedByCoord.out\_intron.txt

### splicing\_result

HN19-9674\_secondAligned.sortedByCoord.out\_splicing.txt  
HN20-9844\_secondAligned.sortedByCoord.out\_splicing.txt  
HN21-10181\_secondAligned.sortedByCoord.out\_splicing.txt



# Pathogen

COMPLETED **pathogen\_pipeline run - 01-25-25 22:38:00** ↗

Get support

View stats & logs

Edit and rerun

Executed on Jan. 25, 2025 17:39 by li2g2uc

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$1.13 | Duration: 40 minutes

▼ App: pathogen\_pipeline - Revision: 0

Each bam take 100GB RAM, please always set  
cores=1, each sample takes 10min

## Inputs

### bam\_files

- HN19-9674\_secondAligned.sortedByCoord.out.bam
- HN20-9844\_secondAligned.sortedByCoord.out.bam
- HN21-10181\_secondAligned.sortedByCoord.out.bam

### kraken2\_db\_dir

#### kraken2\_db

## App Settings

Show non-default ▾

cores

1

mode

pair

outdir

## Output Settings

### test\_report

- HN19-9674\_secondAligned.sortedByCoord.out\_test\_report.txt
- HN20-9844\_secondAligned.sortedByCoord.out\_test\_report.txt
- HN21-10181\_secondAligned.sortedByCoord.out\_test\_report.txt



# Transposable element

COMPLETED telocal\_pipeline run - 01-25-25 22:35:57

Executed on Jan. 25, 2025 17:37 by li2g2uc

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$10.79 | Duration: 4 hours, 12 minutes

App: telocal\_pipeline - Revision: 0

Each bam take 10GB RAM, if you set cores=3, then you need 30GB RAM, each finishes in 5h

Inputs	App Settings	Output Settings
bam_files	cores: 20	output: HN19-9674_secondAligned.sortedByCoord.out_TElocal_out.c...
	outdir: .	HN20-9844_secondAligned.sortedByCoord.out_TElocal_out.c...
	strand: no	HN21-10181_secondAligned.sortedByCoord.out_TElocal_out....
te_local_gene		
hg38.knownGene.gtf		
te_local_te		
hg38_rmsk_TE.gtf.locInd		

# Gene fusion

BATCH 3 STAR-Fusion run - 02-06-25 19:11:07 Last update by li2g2uc on Feb. 6, 2025 14:11 App: STAR-Fusion - Revision: 0

A bit different, you still should select your fastq.gz file, but you need to properly label about these files (show you in next slide) so you can batch run them

Task Inputs Execution Settings

**Inputs**

Batching  On

Input files \*

Batch by: File metadata

This task will be batched by file metadata (Sample ID) and this will create 3 groups.

Hn19 (2 items)   
Hn20 (2 items)   
Hn21 (2 items)

CTAT genome lib archive

Batch by: None

GRCh38.d1.vd1.gencode.v36.annotation.star-fusion-1.1...

**Execution Settings**

For app settings, you only need to select coding effect as True, other left as default

Please always choose c5.9xlarge, take about 1h to run

App Settings

Fusion predictions

Fusion predictions abridged

FusionInspector HTML fusions summary

FusionInspector fusion predictions

STAR-Fusion output archive

Chimeric read filtering parameters: Min non-specific multimapping read percentage

Downstream analysis of fusion candidates: Denovo reconstruct

Downstream analysis of fusion candidates: Examine coding effect

Downstream analysis of fusion candidates: Extract

Output Settings

# Gene Fusion (label them by sample ID and pair-end)

Root Exclude subfolders

2 items selected Copy Rename Move Metadata Edit tags Download ...

New Folder Add files ...

Name	Path	Size	Extension	Task II
<input checked="" type="checkbox"/> HN19-9674_R1.fastq.gz Input	Files	1.92 GiB	FASTQ.GZ	-
<input checked="" type="checkbox"/> HN19-9674_R2.fastq.gz Input	Files	1.97 GiB	FASTQ.GZ	-
<input type="checkbox"/> HN20-9844_R1.fastq.gz Input	Files	1.60 GiB	FASTQ.GZ	-
<input type="checkbox"/> HN20-9844_R2.fastq.gz Input	Files	1.64 GiB	FASTQ.GZ	-
<input type="checkbox"/> HN21-10181_R1.fastq.gz Input	Files	1.79 GiB	FASTQ.GZ	-
<input type="checkbox"/> HN21-10181_R2.fastq.gz Input	Files	1.83 GiB	FASTQ.GZ	-

# Gene Fusion (label them by sample ID and pair-end)

Dashboard   Files   PREMIUM   Apps   Tasks   Data Studio

Root   Exclude subfolders

2 items selected   Copy   Rename   Move

Name

- HN19-9674\_R1.fastq.gz   input
- HN19-9674\_R2.fastq.gz   input
- HN20-9844\_R1.fastq.gz   input
- HN20-9844\_R2.fastq.gz   input
- HN21-10181\_R1.fastq.gz   input
- HN21-10181\_R2.fastq.gz   input

**Update metadata values**

Luminary  Enter value

Primary site  Enter value

Disease type  Enter value

Age at diagnosis  Enter value

Vital status  Enter value

Days to death  Enter value

Sample ID  HN19

Sample UUID  Enter value

Sample type  Enter value

Aliquot ID  Enter value

Aliquot UUID  Enter value

**Custom metadata**

sbg\_public\_files\_category  Enter value

species  Enter value

HN19 HN19

Cancel   Save

# Gene Fusion (label them by sample ID and pair-end)

The screenshot shows a file manager interface with a dark theme. On the left, a list of files is displayed under the 'Root' folder. Several files are selected, indicated by a checked checkbox next to each file name. Red arrows point from the text 'By doing that, when you supply all fastq.gz files, the program will figure out how to pair, and how to parallelize' to the selected files. In the center, a modal window titled 'Update metadata values' is open. The window contains a warning message: '⚠ You can edit metadata only on files. Read more'. Below this is a 'Metadata schema' section with various fields and dropdown menus. One dropdown menu for 'Paired-end' has the value '1' selected, which is also highlighted with a red arrow. At the bottom right of the modal are 'Cancel' and 'Save' buttons.

By doing that, when you supply all fastq.gz files, the program will figure out how to pair, and how to parallelize

# Variants (step 1 is to get vcf files)

COMPLETED **variant\_pipeline run - 01-26-25 16:14:21** ↗

Executed on Jan. 26, 2025 11:15 by li2g2uc

Each file only takes 2GB ram, so based on this to select the instance, will finish in 2h

Spot Instances: Off ⓘ | Memoization (WorkReuse): Off ⓘ | Price: \$2.60 ⓘ | Duration: 1 hour, 40 minutes ⓘ

App: variant\_pipeline - Revision: 2

**Inputs** ↗

- bam\_files ↗
  - HN19-9674\_secondAligned.sortedByCoord.out.bam
  - HN20-9844\_secondAligned.sortedByCoord.out.bam
  - HN21-10181\_secondAligned.sortedByCoord.out.bam
- sequence ↗
  - GRCh38.d1.vd1.fa

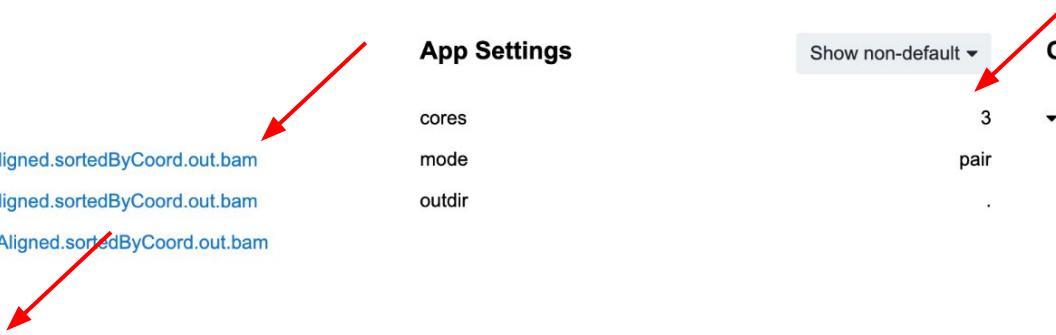
**App Settings**

Show non-default ▾

cores	3
mode	pair
outdir	.

**Output Settings** ↗

- vcf ↗
  - HN19-9674\_variants.vcf
  - HN20-9844\_variants.vcf
  - HN21-10181\_variants.vcf



# Variants (step 2 is to run variant effect predictor)

COMPLETED Variant Effect Predictor run - 01-26-25 20:40:30

Executed on Jan. 26, 2025 15:40 by ll2g2uc

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$0.43 | Duration: 27 minutes

App: Variant Effect Predictor - Revision: 0

Inputs	App Settings	Output Settings
Chromosome synonyms	Add 1000 genomes phase 3 global allele frequency	False Compressed (bgzip/gzip) output
No files selected	Add APPRIS identifiers	False Optional file with VEP warnings and errors
Custom annotation - BigWig sources only	Add CCDS transcript identifiers	False HN21-10181_variants.vep.vcf_warnings.txt
No files selected	Add Ensembl protein identifiers	False Output summary stats file
Custom annotation sources	Add GA4GH Variation Representation Specification	False HN21-10181_variants.vep.vcf_summary.html
No files selected	Add HGVS identifiers	False VEP output file
Fasta file(s) to use to look up reference sequence	Add MANE Select identifiers	False HN21-10181_variants.vep.vcf
No files selected	Add MANE Select or MANE Plus Clinical identifiers	False
GFF annotation file	Add UniProt-associated database identifiers	False
No files selected	Add a flag indicating if the transcript is canonical	False
GTF annotation file	Add allele frequency from continental 1000 genomes populations	False
No files selected	Add biotype of transcript or regulatory feature	False
Input VCF	Add cDNA, CDS and protein positions (position/length)	False
HN21-10181_variants.vcf	Add gene symbols where available	False
NCBI BAM file for correcting transcript models	Add genomic HGVS identifiers	False
No files selected	Add gnomAD allele frequencies	False
Optional config file	Add gnomAD allele frequencies from genome populations	False
No files selected	Add miRNA report	False
Species cache file	Add reference allele in the output	False
homo_sapiens_vep_112_GRCh38.tar.gz	Add transcript support level	False
dbNSFP database file	Add transcript version	False

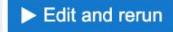
You can either just run one vcf at a time or batch by file

Using c4.2xlarge or c5.9xlarge, seems that each vcf takes about 5GB RAM

Red arrows point to the "Input VCF" and "Species cache file" sections.

# HLA typing (step 1 is decompress fastq.gz to fastq)

t..

**COMPLETED** **SBG Decompressor CWL1.0 run - 01-27-25 21:29:01: file: HN21-10181\_R1.f...** 

Executed on Jan. 27, 2025 16:32 by [li2g2uc](#)

Spot Instances: **Off** | Memoization (WorkReuse): **Off** | Price: **\$0.05** | Duration: **5 minutes**

App: SBG Decompressor CWL1.0 - Revision: 0

Inputs	App Settings	Show non-default	Output Settings
 <b>Input Archive File</b>   HN21-10181_R1.fastq.gz	 Flatten Outputs 	False	 <b>Output Files</b>   HN21-10181_R1.fastq

You should batch by file so multiple run can be parallelized, c4.2xlarge should work, will be done in 10 mins



# HLA typing (step 1 is decompress fastq.gz to fastq)

..

**COMPLETED OptiType run - 01-27-25 22:09:29: sample\_id: HN19**

Executed on Jan. 27, 2025 17:10 by [li2g2uc](#)

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$0.12 | Duration: 13 minutes

App: OptiType - Revision: 0

Inputs

Input file(s) [HN19-9674\\_R2.fastq](#) [HN19-9674\\_R1.fastq](#)

Type of data

Will be done in 20mins

Either c4.x2large or c5.4xlarge or r5.16xlarge, you can specify the pair and sample and batch by file metadata

Show non-default

Output Settings

rna

Config output [HN19\\_config.ini](#)

Coverage plot [HN19.coverage\\_plot.pdf](#)

HLA 4-digits results [HN19.result.tsv](#)

HLA 8-digits results [HN19.result\\_type.tsv](#)

HLA Types

HI ΔΔ\*02:01

Summarization (step 1 is to put all the outputs generated before into a dedicated folder, if we call it /result like below)

Dashboard   Files PREMIUM   Apps   Tasks   Data Studio   Interactive Apps

deploy\_pipelines ⓘ   Interactive Browsers   Settings   Notes

Root / result   New Folder   + Add files ...

75 items (All selected)   Copy   Rename   Move   Metadata   Edit tags   Download   ...

<input checked="" type="checkbox"/>	Name	Size	Extension	Task ID	Created
<input checked="" type="checkbox"/>	HN19-9674_gene.fasta	10.75 MiB	FASTA	add7309e-434...	Jan. 14, 2024
<input checked="" type="checkbox"/>	HN19-9674_gene_tpm.txt	1.38 MiB	TXT	add7309e-434...	Jan. 14, 2024
<input checked="" type="checkbox"/>	HN19-9674_isoform.fasta	5.69 MiB	FASTA	add7309e-434...	Jan. 14, 2024
<input checked="" type="checkbox"/>	HN19-9674_nuorf.fasta	16.18 MiB	FASTA	add7309e-434...	Jan. 14, 2024
<input checked="" type="checkbox"/>	HN19-9674_R1.fastq	9.39 GiB	FASTQ	88fddae5-1d1e...	Jan. 27, 2024
<input checked="" type="checkbox"/>	HN19-9674_R2.fastq	9.39 GiB	FASTQ	56da2aa1-53fc...	Jan. 27, 2024
<input checked="" type="checkbox"/>	HN19-9674_secondAligned.sortedByCoord.out.bam	3.94 GiB	BAM	a198756c-09ce...	Jan. 25, 2024
<input checked="" type="checkbox"/>	HN19-9674_secondAligned.sortedByCoord.out.bam.bai	3.96 MiB	BAI	a198756c-09ce...	Jan. 25, 2024
<input checked="" type="checkbox"/>	HN19-9674_secondAligned.sortedByCoord.out_intron.txt	295.28 KiB	TXT	ba97093b-cc2...	Jan. 25, 2024
<input checked="" type="checkbox"/>	HN19-9674_secondAligned.sortedByCoord.out_intron_peptide.txt	989.38 KiB	TXT	ba97093b-cc2...	Jan. 25, 2024
<input checked="" type="checkbox"/>	HN19-9674_secondAligned.sortedByCoord.out_splicing.txt	7.65 MiB	TXT	ba97093b-cc2...	Jan. 25, 2024

## Summarization (step 2 is to run summarization pipeline to get all the search space)

COMPLETED **summarization\_pipeline run - 02-04-25 19:22:35** ↗

Executed on Feb. 4, 2025 14:22 by li2g2uc

Spot Instances: Off ⓘ | Memoization (WorkReuse): Off ⓘ | Price: \$0.03 ⓘ | Duration: 48 minutes ⓘ

1 core, 100GB should be safe

App: summarization\_pipeline - Revision: 6

Inputs ↗

- db ↗
- ImmunoVerse\_data ↗
- intdir ↗
- result ↗

Take 1h to finish

App Settings

Show non-default ↗

outdir

Output Settings ↗

fastas ↗

- HN19-9674\_Abelson\_murine\_leukemia\_virus\_UP000147198.fasta
- HN19-9674\_Kirsten\_murine\_sarcoma\_virus\_UP000242176.fasta
- HN19-9674\_Mus\_musculus\_mobilized\_endogenous\_polytropic...
- HN19-9674\_TE\_self\_translate.fasta
- HN19-9674\_intron.fasta

...and 20 more items

Immunopeptidome analysis (Step 1 is maxquant, you need a folder of all raw files, and a folder of all fasta as search space)

COMPLETED **maxquant\_pipeline run - 02-04-25 23:22:42**

Executed on Feb. 4, 2025 18:23 by li2g24

Spot Instances: Off | Memoization (Off) | Duration: 8 hours, 57 minutes

App: maxquant\_pipeline - Revision: 4

Take about 5-24 hours

**Inputs**   
fasta\_dir   
  test.fasta  
immuno\_dir   
  test\_immuno

**App Settings**

- hla\_class
- outdir
- peptide\_fdr
- sample\_run\_name
- technology

Show non-default **Output Settings**   
1 output\_same\_immuno\_dir   
  hg19\_maxquant\_combined.txt  
1 hg19  
orbitrap

Always use instance with more than 20 cores and 100GB RAM

A folder with all maxquant tabular results

You can use specific fdr like 0.01 or 0.05, or if you want to do rescore later, it requires whole PSM lists (so fdr=1)

Immunopeptidome analysis (Step 2 is to use msconvert from proteowizard to convert raw to mzml, it is required for rescoring and visualization)

COMPLETED **msconvert\_pipeline run - 02-05-25 14:51:27** 

Executed on Feb. 5, 2025 09:51 by li2g2uc

Spot Instances: Off  | Memoization (WorkReuse): Off  | Price: \$0.24  | Duration: 8 minutes 

App: msconvert\_pipeline - Revision: 2

C4.2xlarge should be fine, 10mins

**Inputs**   
raw\_file   
[20240110\\_E\\_OdinLC\\_IC\\_PDX\\_HD\\_19.raw](#)

**App Settings**  
outdir 

Show non-default 

**Output Settings**   
mzml   
[20240110\\_E\\_OdinLC\\_IC\\_PDX\\_HD\\_19.mzML](#)



# Immunopeptidome analysis (Step 3 is the rescoring step using ms2rescore)

COMPLETED **rescore\_pipeline run - 02-06-25 15:37:21**

Executed on Feb. 6, 2025 10:37 by [li2g2uc](#)

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$0.07 | Duration: 3 minutes

App: [rescore\\_pipeline](#) - Revision: 18

Inputs	App Settings	Show non-default	Output Settings
<ul style="list-style-type: none"><li>▼ maxquant_dir <ul style="list-style-type: none"><li><a href="#">hg19_maxquant_combined_txt</a></li></ul></li><li>▼ mzml_dir <ul style="list-style-type: none"><li><a href="#">test_mzml</a></li></ul></li></ul>	<ul style="list-style-type: none"><li>outdir .</li><li>peptide_fdr 0.05</li><li>sample_run_name test</li><li>technology orbitrap</li></ul>	Show non-default	<ul style="list-style-type: none"><li>▼ output <ul style="list-style-type: none"><li><a href="#">test_msmsScans_new.txt</a></li></ul></li></ul>

Please put the generated mzML files from last step to a folder

# Immunopeptidome analysis (Step 4 is the HLA binding prediction)

COMPLETED **hla\_binding\_pipeline run - 02-06-25 19:08:42**

Executed on Feb. 6, 2025 14:09 by li2g2uc

Spot Instances: Off | Memoization (WorkReuse): Off | Price: \$0.08 | Duration: 2 minutes

App: hla\_binding\_pipeline - Revision: 3

**Inputs**   
hla\_type   
test\_hla\_type.txt  
rescored.txt   
test\_msmsScans\_new.txt

**App Settings**

Setting	Value
hla_class	1
outdir	.
sample_run_name	test

**Output Settings** Unknown file name

Just a tab-delimited txt file with two column, raw and hla,  
raw should be the raw file name, hla format is like below

	A	B	C	D	E	F
1	raw		hla			
2	20240110_E_OdinLC_IC_PDX_HD_19		A*02:01,A*03:01,B*18:01,B*44:02,C*12:03,C*07:04			
3						