

Running ImmunoVerse on CGC

Guangyuan(Frank) Li, PhD
NYU Grossman School of Medicine
Last update: 2025.10.11

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, **d/e/f/g** requires **c** or provide your own bam file, **j** should be after **a-i**, **k** is optional you can use other immunopeptidome workflow):
 - a. [Set things up](#)
 - b. [Gene expression](#)
 - c. [Alignment](#)
 - d. [Splicing and intron retention](#)
 - e. [Pathogen](#)
 - f. [Transposable element](#)
 - g. [Variants](#)
 - h. [Gene fusion](#)
 - i. [HLA typing](#)
 - j. [summarization](#)
 - k. [Immunopeptidome analysis](#)
 - l. [Advanced Usage 1 \(predicting and evaluating spectra\)](#)

Set things up

Step 1: Creating Account

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



Log in



Log in with eRA Commons

[Log in with username and password](#)

New to the CGC? [Create an account](#)

Step 2: Create a project

A project has all the (1) needed files, (2) workflows and (3) tasks being run

The screenshot shows the 'Create a project' dialog in the Cancer Genomics Cloud interface. The dialog is titled 'Create a project' and has a close button (X) in the top right corner. It contains several sections: 'Name' with a text input field; a URL field with the value 'https://cgc.sbgenomics.com/u/li2g2uc/'; a 'CONTROLLED' checkbox which is unchecked, with a red arrow pointing to it and the word 'unchecked' in red text; 'General information' and 'Advanced settings' tabs; 'Billing group' with a dropdown menu showing 'Pilot Funds (li2g2uc)'; 'Location' with a dropdown menu showing 'AWS (us-east-1)'; 'Execution settings' with 'Spot instances' (unchecked) and 'Reuse' (unchecked) options. A red arrow points to the 'Spot instances' checkbox with the word 'unchecked' in red text. A red arrow points to the 'Pilot Funds (li2g2uc)' dropdown with a red text box stating: 'You need a billing account either ideally from your lab, or you can apply \$300 pilot fund (https://www.cancergenomicscloud.org/cgc-apply-for-collaborative-funds)'. At the bottom right are 'Cancel' and 'Create' buttons. In the background, a 'Projects' list is visible with a '+ Create Project' button highlighted by a red arrow.

Annotations:

- Name it**: Red arrow pointing to the 'Name' input field.
- unchecked**: Red text next to the unchecked 'CONTROLLED' checkbox.
- unchecked**: Red text next to the unchecked 'Spot instances' checkbox.
- You need a billing account either ideally from your lab, or you can apply \$300 pilot fund (https://www.cancergenomicscloud.org/cgc-apply-for-collaborative-funds)**: Red text box pointing to the 'Pilot Funds (li2g2uc)' 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

<https://cgc.sbggenomics.com/u/li2g2uc/>

☐ **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
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 displays the ImmunoVerse web application interface. The top navigation bar includes links for Home, Projects, Data, Public Apps, Public Projects, and Developer. Below this, a secondary navigation bar shows Dashboard, Files, PREMIUM, Apps, Tasks, Data Studio, and Interactive Apps. The main content area is divided into two panels. The left panel, titled 'Description', contains an 'Overview' section for 'ImmunoVerse' and a 'Tools available' section listing 14 different pipelines. The right panel, titled 'Members', shows user profiles for 'li2g2uc' (OWNER) and 'sharma28'. Below the members section is an 'Analysis' section with a search bar and a list of tasks. The tasks list shows a sequence of 'rescore_pipeline run' tasks with their status (COMPLETED, ABORTED, FAILED), submission time, and user.

Navigation and Annotations:

- Red arrows point from the text 'When running each app for specific input, you create an instance or task' to the 'Files', 'PREMIUM', 'Apps', and 'Tasks' tabs in the secondary navigation bar.
- Red text annotations highlight 'All the files' near the 'Files' tab and 'Apps built from docker containers' near the 'Apps' tab.

Tools available:

1. Gene Expression Pipeline (all samples keeping pair order, 10GB RAM each done in 30min)
2. Alignment Pipeline (all samples keeping pair order, 30GB RAM each done in 5h)
3. Splicing Intron Pipeline (all samples, 2GB RAM each done in 30min)
4. Transposable Element Pipeline (all samples, 10GB RAM each in 5h)
5. Pathogen Pipeline (all samples, 100GB RAM each done in 10min)
6. Variant Pipeline
 - 6.1 RNA Variant Pipeline (all samples, 2GB RAM each done in 90min)
 - 6.2 VEP pipeline (batch by sample, 5GB RAM each done in 30min)
7. Gene Fusion Pipeline (batch by both sample and pair-end)
8. HLA type pipeline
 - 8.1 decompress pipeline (batch by file, 2GB RAM each done in 5min)
 - 8.2 optotype pipeline (batch by sample and pair-end, 20GB RAM each done in 20min)
9. Summarization Pipeline
10. (optional) bam_to_fastq pipeline
11. (optional) circular RNA pipeline
12. Immunopeptidome Pipeline
 - 12.1 MaxQuant Pipeline
 - 12.2 msconvert pipeline
 - 12.3 Rescore pipeline
 - 12.4 HLA binding pipeline

Analysis Tasks:

Status	Task Name	Submitted by	Time
COMPLETED	rescore_pipeline run - 02-06-25 15:37:21	li2g2uc	Feb 6, 2025 10:37
ABORTED	rescore_pipeline run - 02-06-25 15:09:28	li2g2uc	Feb 6, 2025 10:09
ABORTED	rescore_pipeline run - 02-06-25 14:53:12	li2g2uc	Feb 6, 2025 9:53
ABORTED	rescore_pipeline run - 02-06-25 14:31:20	li2g2uc	Feb 6, 2025 9:31
FAILED	rescore_pipeline run - 02-06-25 14:12:29	li2g2uc	Feb 6, 2025 9:12

Upload and Download

Upload files

DashboardFiles PREMIUMAppsTasksData StudioInteractive Apps

deploy_pipelines ⓘ

Interactive BrowsersSettingsNotes

Root

Extension ▾Tags ▾Paired-end ▾+ Clear filters

☐	Name ^	Size	Extension	Task ID
☐	ensembl_protein.fasta reference	11.36 MiB	FASTA	-
☐	gencode.v36.annotation.gtf reference	1.29 GiB	GTF	-
☐	gene_model.txt reference	52.57 MiB	TXT	-
☐	GRCh38.d1.vd1.fa reference	2.94 GiB	FA	-
☐	GRCh38.d1.vd1.gencode.v36.annotation.star-fusion-1.12.0-CTAT-index-archive.tar reference	44.87 GiB	TAR	12f76e64-e489-... Feb. 02, 2025 15:...
☐	hg19_maxquant_combined_txt	-	-	6ea7ebaa-448b-... Feb. 05, 2025 03:...

New Folder+ Add files ▾...

Public Files

Projects

Your Computer

FTP / HTTP

GA4GH Data Repository Service (DRS)

Data Tools

Volumes

Import from a manifest file

Jan. 25, 2025 13:...

Easiest way, just drag the file from your computer

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

Browse files

on your computer containing the links

Add tags

Resolve naming conflicts:

Skip



Import

Download



Root

 New Folder

1 item selected



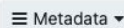
Copy



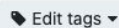
Rename



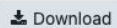
Move



Metadata



Edit tags



Download



Name

Size

Extension

Task ID

Created on



ensembl_protein.fasta

11.36 MiB

FASTA

-

Jan. 13, 2025 10:...

reference



gencode.v36.annotation.gtf

1.29 GiB

GTF

-

Jan. 13, 2025 10:...

reference



gene_model.txt

52.57 MiB

TXT

-

Jan. 13, 2025 10:...

reference

Select

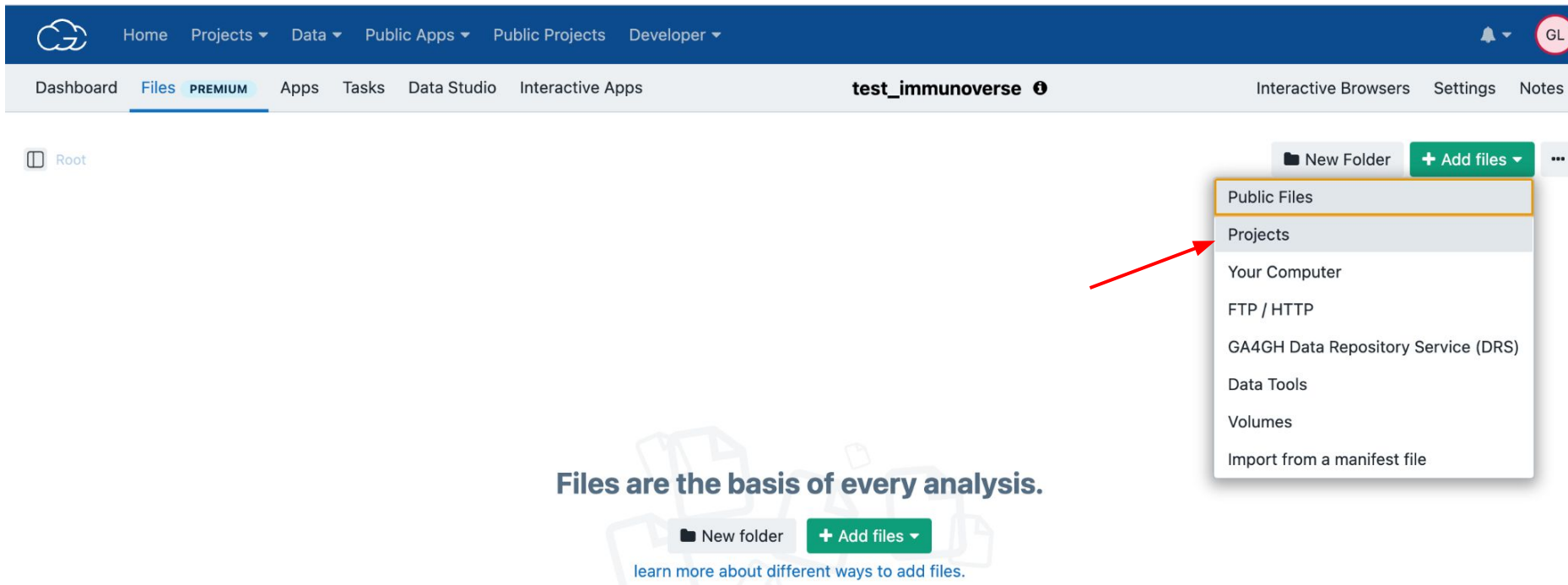
download

Now, Let's copy needed files and workflows from my project to your project (step-by-step after this slide)

- Transfer all files tagged as **reference** to your project
- Within which we will find files that are further tagged by **ImmunoVerse_data**, **star_hg38_index**, **kraken2_db**, please create three folders named **ImmunoVerse_data**, **star_hg38_index**, **kraken2_db** and moved the corresponding files to these three folders, later, the folder will be passed as an argument in respective workflows
- After step 2, You will notice that two files are present twice, please remove the prefix
 - **_1_hg38.fa** -> **hg38.fa**
 - **_1_Ensemble_protein.fasta** -> **Ensemble_protein.fasta**
- Copy all the workflows/dockers from my project

Sounds tedious, but you only need to set things up once :)

Copy the needed reference files from my project



The screenshot displays the 'test_immunoverse' project interface. The top navigation bar includes links for Home, Projects, Data, Public Apps, Public Projects, and Developer. Below this, a secondary bar shows Dashboard, Files (highlighted as PREMIUM), Apps, Tasks, Data Studio, and Interactive Apps. The main content area shows a 'Root' directory with a 'New Folder' button and a '+ Add files' dropdown menu. The dropdown menu is open, showing options: Public Files, Projects (highlighted with a red arrow), Your Computer, FTP / HTTP, GA4GH Data Repository Service (DRS), Data Tools, Volumes, and Import from a manifest file. At the bottom, a message states 'Files are the basis of every analysis.' with a 'New folder' button and an 'Add files' button, followed by a link to 'learn more about different ways to add files.'

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

Dashboard Files PREMIUM Apps Tasks Data Studio Interactive Apps test_immunoverse ⓘ Interactive Browsers Settings Notes

Root

New Folder + Add files ▾

- Public Files
- Projects
- Your Computer
- FTP / HTTP
- GA4GH Data Repository Service (DRS)
- Data Tools
- Volumes
- Import from a manifest file

Files are the basis of every analysis.

New folder + Add files ▾

[learn more about different ways to add files.](#)

Copy the needed reference files from my project (cont'd)

Add files to "test_immunoverse"

Search for project

deploy_pipelines

NeoVerse_pancancer_2

NeoVerse_development_proj...

NeoVerse_pancancer

Demonstration: Building an App

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

Search

Extension

Tags

Paired-end

+

Clear filters

Copy to Project

☐

Name

Size

Exten

☐

ensembl_protein.fasta

reference

11.36 MiB

FASTA

☐

gencode.v36.annotation.gtf

reference

1.29 GiB

GTF

☐

gene_model.txt

reference

52.57 MiB

TXT

☐

GRCh38.d1.vd1.fa

reference

2.94 GiB

FA

☐

GRCh38.d1.vd1.gencode.v36.annotation.s

reference

44.87 GiB

TAR

☐

hg19_maxquant_combined_txt

-

-

☐

hg38.fa

reference

3.05 GiB

FA

☐

hg38.knownGene.gtf

reference

564.57 MiB

GTF

☐

hg38_rmsk_TE.gtf.locInd

reference

883.82 MiB

LOCIN

Search

☐ No value

☐ ImmunoVerse_data

☐ input

☐ kraken2_db

☒ reference

☐ star_hg38_index

Clear selected

Apply

Copy the needed reference files from my project (cont'd)

Add files to "test_immunoverse" x

Search for project

deploy_pipelines

NeoVerse_pancancer_2

NeoVerse_development_proj...

NeoVerse_pancancer

Demonstration: Building an App

Root

Exclude subfolders ☐

Search

Extension

Tags: reference

Paired-end

+

Clear filters

☒

Name ^

	Path	Size
☒ reference Immunoverse_data	Files / Immunoverse_data	10.08
☒ reference Immunoverse_data	Files / Immunoverse_data	2.74 M
☒ reference Immunoverse_data	Files / Immunoverse_data	25.59
☒ reference Immunoverse_data	Files / Immunoverse_data	2.50 M
☒ reference star_hg38_index	Files / star_hg38_index	13.83
☒ reference star_hg38_index	Files / star_hg38_index	65.86
☒ reference star_hg38_index	Files / star_hg38_index	79.69
☒ reference star_hg38_index	Files / star_hg38_index	29.85
☒ reference Immunoverse_data	Files / Immunoverse_data	136.6
☒ reference Immunoverse_data	Files / Immunoverse_data	1.63 M
☒ reference Immunoverse_data	Files / Immunoverse_data	43.22
☒ reference	Files	11.36

63 items

Copy to Project

Organize a bit (create a folder named ImmunoVerse_data and move all files tagged as ImmunoVerse_data to this folder)

Dashboard Files PREMIUM Apps Tasks Data Studio Interactive Apps test_immunoverse Interactive Browsers Settings Notes

Root

Search Extension Tags + Clear filters

Create New Folder + Add files ...

Create New Folder

Folders can't be subsequently renamed.

Name

ImmunoVerse_data

Path

Files /

Cancel Create

Name	Extension	Task ID	Create
_2_ensembl_protein.fasta	FASTA	-	Feb. 06
_1_ensembl_protein.fasta	FASTA	-	Feb. 06
_1_hg38.fa	FA	-	Feb. 06
opts.k2d	K2D	-	Feb. 06
normal_erv.txt	TXT	-	Feb. 06
normal_intron.txt	TXT	-	Feb. 06
SA	-	-	Feb. 06
SAindex	-	-	Feb. 06
splice_erv_db.txt	156 bytes	TXT	Feb. 06
tcga_tmp_prelift.bed	323.92 MiB	BED	Feb. 06
ensembl_protein.fasta	11.36 MiB	FASTA	Feb. 06

63 items

Organize a bit (create a folder named ImmunoVerse_data and move all files tagged as ImmunoVerse_data to this folder)

The screenshot shows the Google Cloud Storage interface. The top navigation bar includes links for Home, Projects, Data, Public Apps, Public Projects, and Developer. The main header shows the project name 'test_immunoverse'. The left sidebar displays a file tree with a folder named 'ImmunoVerse_data' and several files, each with a 'reference' tag and an 'ImmunoVerse_data' tag. A modal window is open in the center, showing a search bar and a list of tags. The 'ImmunoVerse_data' tag is selected, and the 'Apply' button is highlighted. Red arrows point to the 'ImmunoVerse_data' tag in the file list and the 'Apply' button in the modal.

Name	Size	Extension	Task ID	Create
ImmunoVerse_data	-	-	-	Feb. 06
_2_ensembl_protein.fasta	11.36 MiB	FASTA	-	Feb. 06
_1_ensembl_protein.fasta	11.36 MiB	FASTA	-	Feb. 06
_1_hg38.fa	3.05 GiB	FA	-	Feb. 06
opts.k2d	64 bytes	K2D	-	Feb. 06
normal_erv.txt	1.37 GiB	TXT	-	Feb. 06
normal_intron.txt	96.94 MiB	TXT	-	Feb. 06
SA	23.20 GiB	-	-	Feb. 06
SAindex	1.46 GiB	-	-	Feb. 06
splice_erv_db.txt	156 bytes	TXT	-	Feb. 06
tcga_tmp_prelift.bed	323.92 MiB	BED	-	Feb. 06

64 items

Organize a bit (create a folder named ImmunoVerse_data and move all files tagged as ImmunoVerse_data to this folder)

The screenshot shows the ImmunoVerse web interface with a 'Move' dialog box open. The dialog box is titled 'Move' and contains a list of files to be moved. A red arrow points to the 'ImmunoVerse_data' folder in the list. Another red arrow points to the 'Move' button at the bottom right of the dialog box.

Move 31 selected items?

Files

- ImmunoVerse_data

Tags

☒ Keep preexisting tags

Add new tags by separating them with enter key

Name	Size	Extension	Task II
2_ensembl_protein.fasta	11.36 MiB	FASTA	-
normal_erv.txt	1.37 GiB	TXT	-
normal_intron.txt	96.94 MiB	TXT	-
splice_erv_db.txt	156 bytes	TXT	-
tcga_tmp_prelift.bed	323.92 MiB	BED	-
snaf_prelift.bed	16.84 MiB	BED	-
hdr_cosmic.txt	73 bytes	TXT	-
cosmic_prelift.bed.gz.tbi	43.22 KiB	TBI	-
hdr_redi.txt	75 bytes	TXT	-
annot.renamed.txt.gz	10.08 GiB	TXT.GZ	-
contigs.txt	136.68 KiB	TXT	-

Organize a bit (create another folder named star_hg_index and move all files tagged as star_hg38_index to this folder)

The screenshot shows the 'test_immunoverse' interface with a 'Move' dialog box open. The dialog prompts to move 16 selected items. The destination folder 'star_hg38_index' is selected under 'ImmunoVerse_data'. The 'Keep preexisting tags' checkbox is checked. A red arrow points to the 'Move' button.

Move

Move 16 selected items?

Files

- ImmunoVerse_data
 - star_hg38_index

Tags

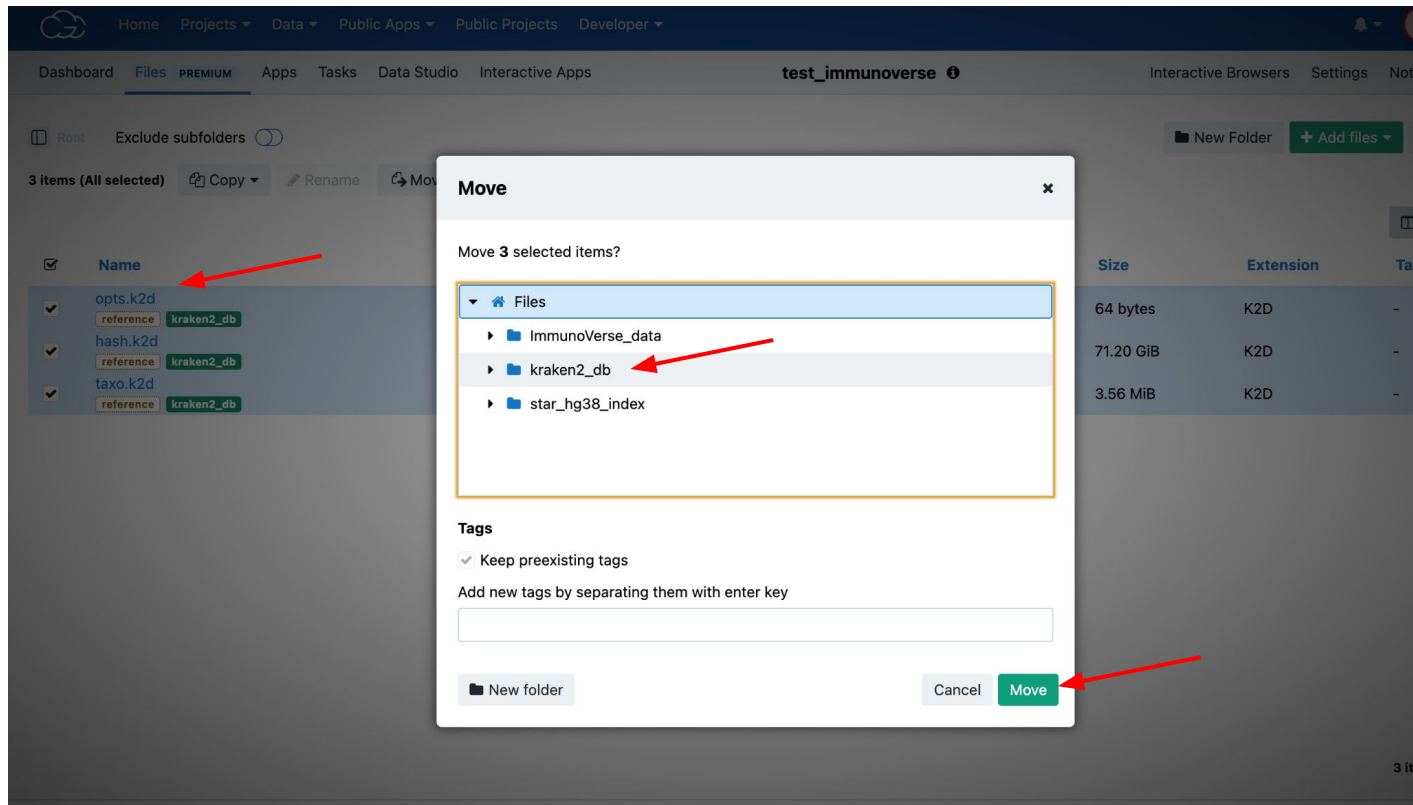
☒ Keep preexisting tags

Add new tags by separating them with enter key

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	-

16 items

Organize a bit (create a folder named kraken2_db, and move all files tagged as kraken2_db to this folder)



Final clean (if applied to you, may not need)

Click the
file and
rename

Root Exclude subfolders

ensembl

Extension Tags + Clear filters

Name

_1_ensembl_protein.fasta

reference ImmunoVerse_data

Root Exclude subfolders

hg38

Extension Tags + Clear filters

Name

star_hg38_index

_1_hg38.fa

reference

Dashboard Files Apps Tasks Data Studio Interactive Apps

Files

Rename

_1_hg38.fa

REFERENCE

3.0 GiB (3,273,481,150 bytes) · Created on May 8, 2025 15:58 (Eastern Day)

Metadata Raw View

Copy my app/apps

The screenshot shows the Immuniverse web interface. The top navigation bar includes links for Home, Projects, Data, Public Apps, Public Projects, and Developer. A red arrow points to the 'Apps' link in the secondary navigation bar. The user's profile 'test_immunoverse' is visible. Below the navigation, there is a search bar and several filter buttons: 'Category: All', 'Toolkit: All', 'Language and version: All', 'Status: Available', and 'Cost Estimator: All'. A large 'No apps' message is displayed in the center, stating 'No apps with the given search term can be found.' and providing a 'Clear filters' button. On the right side, a 'Create app' button is next to a green '+ Add apps' button. A red arrow points to the '+ Add apps' button, which has a dropdown menu open showing three options: 'Public apps' (highlighted with an orange border), 'Projects', and 'My created apps'. Another red arrow points to the 'Projects' option in the dropdown menu.

Home Projects Data Public Apps Public Projects Developer

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

No apps
No apps with the given search term can be found.
Clear filters

Create app + Add apps

- Public apps
- Projects
- My created apps

Copy my app/apps

Add apps to test_immunoverse

Category: All Toolkit: All Language and version: All

deploy_pipelines	Name	Type	Modified by	Modified on	
NeoVerse_pancancer_2	hla_binding_pipeline	Tool	li2g2uc	Feb 06, 2025 13:31	Copy
NeoVerse_development_project	rescore_pipeline	Tool	li2g2uc	Feb 06, 2025 10:37	Copy
NeoVerse_pancancer	msconvert_pipeline	Tool	li2g2uc	Feb 05, 2025 18:28	Copy
Demonstration: Building an App	maxquant_pipeline	Tool	li2g2uc	Feb 04, 2025 18:22	Copy
	summarization_pipeline	Tool	li2g2uc	Feb 04, 2025 12:03	Copy
	STAR-Fusion Build Fusio	Tool	li2g2uc	Jan 31, 2025 11:33	Copy
	SBG Decompressor CWI	Tool	li2g2uc	Jan 27, 2025 16:28	Copy
	variant_pipeline	Tool	li2g2uc	Jan 26, 2025 11:14	Copy
	Variant Effect Predictor	Tool	li2g2uc	Jan 24, 2025 16:17	Copy
	OptiType	Tool	li2g2uc	Jan 24, 2025 15:47	Copy
	STAR-Fusion	Tool	li2g2uc	Jan 24, 2025 15:23	Copy
	alignment_pipeline	Tool	li2g2uc	Jan 24, 2025 14:52	Copy
	gene_pipeline	Tool	li2g2uc	Jan 14, 2025 14:02	Copy
	splicing_intron_pipeline	Tool	li2g2uc	Jan 13, 2025 12:12	Copy
	telocal_pipeline	Tool	li2g2uc	Jan 13, 2025 12:12	Copy

Finished setting things up!!!

Dashboard

Files **PREMIUM**

Apps

Tasks

Data Studio

Interactive Apps

test_immunoverse ⓘ

Interactive Browsers

Settings

Notes

Root

New Folder

+ Add files

...

15 items (All selected)

Copy

Rename

Move

Metadata

Edit tags

Download

...

15 items

Gene Expression

- For any workflow
 - Select inputs
 - Select reference files
 - Specify parameters
 - Specify running instances
- Each workflow will have slightly different requirement, which I will showcase step by step

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](#) 

[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](#)

...and 1 more item

▼ [kallisto_index](#) 

[kallisto_index](#)

▼ [nuorf](#) 

[nuorf.fasta](#)

▼ [uniprot_isoform](#) 

[uniprot_reviewed_curated_addition.fasta](#)

App Settings

 [Show editable](#) ▼

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

Keep it

Whether your library is stranded or not, no is always safe if you don't know

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)

Price: \$1.008 per hour

Attached storage (GB)

4096

Each sample take
~10GB RAM, so
three cores will take
30GB, so try to select
one that has more
than 3 cores and
30GB, any instance
would work

Take 30mins to finish

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

Gene expression

DashboardFilesAppsTasksData StudioInteractive Apps

deploy_pipelines ⓘ

Interactive BrowsersSettingsNotes

DRAFT

gene_pipeline run - 02-06-25 18:49:10 ⓘ

Get support

Discard

Run

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

...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 ⓘ

gene_fasta ⓘ

3

HN19-9674_gene.fasta

HN20-9844_gene.fasta

HN21-10181_gene.fasta

isoform_fasta ⓘ

no

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 [Get support](#) [View stats & logs](#) [Edit and rerun](#)

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

Spot Instances: **Off** | Memoization: **Off** | Caching: **Off** | Parallelism: **Off** | Priority: **Low** | Capacity: **On-demand**

App: alignment_pipeline - Revision: 1

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

- 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

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: 45 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

cores 1
mode pair
outdir .

Show non-default ▼

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** 

 Get support

 View stats & logs

 Edit and rerun

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

▼ bam_files

[HN19-9674_secondAligned.sortedByCoord.out.bam](#)
[HN20-9844_secondAligned.sortedByCoord.out.bam](#)
[HN21-10181_secondAligned.sortedByCoord.out.bam](#)

▼ te_local_gene

[hg38.knownGene.gtf](#)

▼ te_local_te

[hg38_rmsk_TE.gtf.locInd](#)

App Settings

cores 20
outdir .
strand no

Show non-default ▼

Output Settings

▼ output

[HN19-9674_secondAligned.sortedByCoord.out_TElocal_out.c...](#)
[HN20-9844_secondAligned.sortedByCoord.out_TElocal_out.c...](#)
[HN21-10181_secondAligned.sortedByCoord.out_TElocal_out...](#)



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

Get support Discard Run

Task Inputs Execution Settings

Inputs

Batching [?](#) On ☒

Input files * [?](#) [Change selection](#)

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 [?](#) [Change selection](#)

Batch by: [None](#)

[GRCh38.d1.vd1.gencode.v36.annotation.star-fusion-1.1...](#)

App Settings

[✎ Edit parameters](#) [Show editable](#)

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

[?](#)

[No value](#)

Downstream analysis of fusion candidates: Denovo reconstruct

[?](#)

[No value](#)

Downstream analysis of fusion candidates: Examine coding effect

[?](#)

[True](#) [✎](#)

Downstream analysis of fusion candidates: Extract

Output Settings

[Fusion predictions](#) [?](#) [No value](#)

[Fusion predictions abridged](#) [?](#) [No value](#)

[FusionInspector HTML fusions summary](#) [?](#) [No value](#)

[FusionInspector fusion predictions](#) [?](#) [No value](#)

[STAR-Fusion output archive](#) [?](#) [No value](#)

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

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

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

Root Exclude subfolders

2 items selected

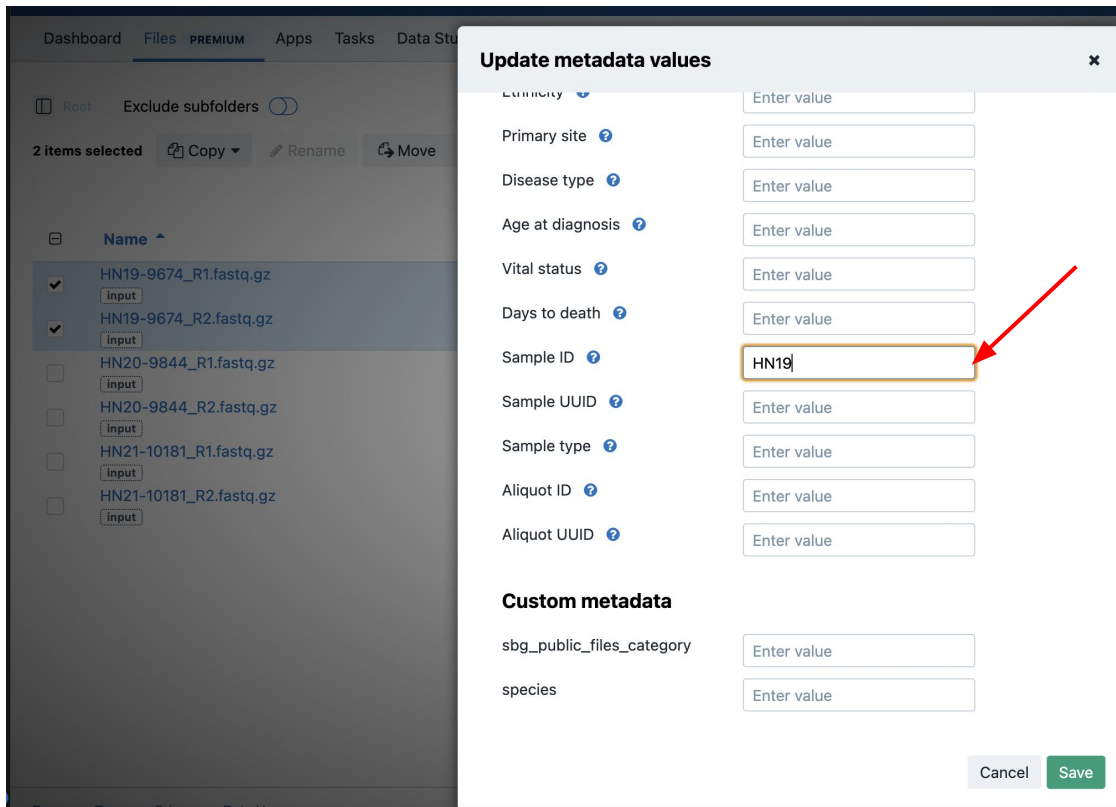
Copy Rename Move Metadata Edit tags Download

Update values

Clear values

Name	Path	Size	Extension	Task II
☒ HN19-9674_R1.fastq.gz input	Files	1.92 GiB	FASTQ.GZ	-
☒ HN19-9674_R2.fastq.gz input	Files	1.97 GiB	FASTQ.GZ	-
☐ HN20-9844_R1.fastq.gz input	Files	1.60 GiB	FASTQ.GZ	-
☐ HN20-9844_R2.fastq.gz input	Files	1.64 GiB	FASTQ.GZ	-
☐ HN21-10181_R1.fastq.gz input	Files	1.79 GiB	FASTQ.GZ	-
☐ HN21-10181_R2.fastq.gz input	Files	1.83 GiB	FASTQ.GZ	-

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



The screenshot shows a file management interface on the left and a modal window titled "Update metadata values" on the right. The modal contains a list of metadata fields, each with an "input" button. A red arrow points to the "Sample ID" field, which is highlighted with a yellow border and contains the text "HN19".

Update metadata values ✕

Country	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

Cancel Save

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

The screenshot shows a file management interface with a list of files on the left and a modal dialog titled "Update metadata values" on the right. The file list includes:

- 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

Three red arrows point from the file names to the "Update metadata values" dialog. The dialog contains a warning: "You can edit metadata only on files. Read more". Below this is a "Metadata schema" section with the following fields:

- Experimental strategy: Enter value
- Library ID: Enter value
- Platform: Enter value
- Platform unit ID: Enter value
- File segment number: Enter value
- Quality scale: -
- Paired-end: -
- Reference genome: 1 (selected)
- Investigation: 2
- Case ID: Enter value
- Case UUID: Enter value
- Gender: Enter value

At the bottom of the dialog are "Cancel" and "Save" buttons. A red arrow points from the "Paired-end" dropdown menu to the "Reference genome" dropdown menu, indicating a relationship between the two settings.

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 

 Get support

 View stats & logs

 Edit and rerun

Executed on Jan. 26, 2025 11:15 by [li2g2uc](#)

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

App: [variant_pipeline](#) - Revision: 2

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

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

cores
mode
outdir

Show non-default ▼

Output Settings

▼ vcf

3
pair
.
 [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

Get support

View stats & logs

Edit and rerun

Executed on Jan. 26, 2025 15:40 by li2g2uc

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

App: Variant Effect Predictor - Revision: 0

Inputs

Chromosome synonyms

No files selected

Custom annotation - BigWig sources only

No files selected

Custom annotation sources

No files selected

Fasta file(s) to use to look up reference sequence

No files selected

GFF annotation file

No files selected

GTF annotation file

No files selected

Input VCF

[HN21-10181_variants.vcf](#)

NCBI BAM file for correcting transcript models

No files selected

Optional config file

No files selected

Species cache file

[homo_sapiens_vep_112_GRCh38.tar.gz](#)

dbNSFP database file

App Settings

Add 1000 genomes phase 3 global allele frequency

Add APPRIS identifiers

Add CCDS transcript identifiers

Add Ensembl protein identifiers

Add GA4GH Variation Representation Specification

Add HGVS identifiers

Add MANE Select identifiers

Add MANE Select or MANE Plus Clinical identifiers

Add UniProt-associated database identifiers

Add a flag indicating if the transcript is canonical

Add allele frequency from 1000 genomes populations

Add biotype of transcript or regulatory feature

Add cDNA, CDS and protein positions (position/length)

Add gene symbols where available

Add genomic HGVS identifiers

Add gnomAD allele frequencies

Add gnomAD allele frequencies from genome populations

Add miRNA report

Add reference allele in the output

Add transcript support level

Add transcript version

Show all

Output Settings

Compressed (bgzip/gzip) output No value

Optional file with VEP warnings and errors

[HN21-10181_variants.vep.vcf_warnings.txt](#)

Output summary stats file

[HN21-10181_variants.vep.vcf_summary.html](#)

VEP output file

[HN21-10181_variants.vep.vcf](#)

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

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...

Get support

View stats & logs

Edit and rerun

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

Input Archive File

HN21-10181_R1.fastq.gz

App Settings

Flatten Outputs

Show non-default

False

Output Settings

Output Files

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

App Settings

Type of data: rna

Output Settings

- 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

Get support | View stats & logs | Edit and rerun

Show non-default

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

Will be done in 20mins

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

DashboardFiles PREMIUMAppsTasksData StudioInteractive Apps

deploy_pipelines ⓘ

Interactive BrowsersSettingsNotes

Root / result

New FolderAdd files ...

75 items (All selected)Copy ▾RenameMoveMetadata ▾Edit tags ▾Download...

☒

Name ^	Size	Extension	Task ID	Created
☒ HN19-9674_gene.fasta	10.75 MiB	FASTA	add7309e-434...	Jan. 14, 20
☒ HN19-9674_gene_tpm.txt	1.38 MiB	TXT	add7309e-434...	Jan. 14, 20
☒ HN19-9674_isoform.fasta	5.69 MiB	FASTA	add7309e-434...	Jan. 14, 20
☒ HN19-9674_nuorf.fasta	16.18 MiB	FASTA	add7309e-434...	Jan. 14, 20
☒ HN19-9674_R1.fastq	9.39 GiB	FASTQ	88fddae5-1d1e...	Jan. 27, 20
☒ HN19-9674_R2.fastq	9.39 GiB	FASTQ	56da2aa1-53fc-...	Jan. 27, 20
☒ HN19-9674_secondAligned.sortedByCoord.out.bam	3.94 GiB	BAM	a198756c-09ce...	Jan. 25, 20
☒ HN19-9674_secondAligned.sortedByCoord.out.bam.bai	3.96 MiB	BAI	a198756c-09ce...	Jan. 25, 20
☒ HN19-9674_secondAligned.sortedByCoord.out_intron.txt	295.28 KiB	TXT	ba97093b-cc2...	Jan. 25, 20
☒ HN19-9674_secondAligned.sortedByCoord.out_intron_peptide.txt	989.38 KiB	TXT	ba97093b-cc2...	Jan. 25, 20
☒ HN19-9674_secondAligned.sortedByCoord.out_splicing.txt	7.65 MiB	TXT	ba97093b-cc2...	Jan. 25, 20

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

COMPLETED **summarization_pipeline run - 02-04-25 19:22:35**

Get support

View stats & logs

Edit and rerun

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

Spot Instances: Off | Memoization (Work/Pause): Off | Price: \$0.93 | Duration: 48 minutes

App: summarization_pipeline - Revision: 6

Inputs

db

ImmunoVerse_data

intdir

result

Take 1h to finish

App Settings

outdir

Keep it

Show non-default

Output Settings

fastas

HN19-9674_Abelson_murine_leukemia_virus_UP000147198.f...
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

Till now, you had sample-specific
aberrations and search space (fastas)

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**

Get support

View stats & logs

Edit and rerun

Executed on Feb. 4, 2025 18:23 by li2g2k

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

App: maxquant_pipeline - Revision: 4

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

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

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** 

 Get support

 View stats & logs

 Edit and rerun

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 - 06-15-25 19:39:29**  Find an instance with more than 64GB RAM should be safe

Executed on June 15, 2025 15:40 by [li2g2uc](#)

Spot Instances: **Off**  | Memoization (WorkReuse): **Off**  | Price: **\$0.06**  | Duration: **3 minutes** 

App: [rescore_pipeline](#) - Revision: 20

Inputs 

- ▼ **input1**  
 -  [hg19_maxquant_combined.txt](#)
- ▼ **input2**  
 -  [test_mzml](#)

App Settings

mode 	rescore
run_rescore 	False
technology 	orbitrap

Show non-default ▼

Output Settings 

output_mirror_plot	No value
output_ms2pip_prediction	No value
▼ output_rescore 	
msmsScans_new.txt	

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

Get support

View stats & logs

Edit and rerun

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

Show non-default

- hla_class: 1
- outdir: .
- sample_run_name: test

Output Settings

- output
 - 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						

Advanced Usage 1

In the manuscript, we showcased that ImmunoVerse wraps **MS2PIP spectrum prediction function to evaluate the neoantigen spectra confidence on the fly**, this function can be achieved through CGC as well, demonstrated in the following slides

Step 1: predict in silico spectrum for multiple peptides

COMPLETED **rescore_ms2pip run - 06-15-25 20:01:33**

Get support

View stats & logs

Edit and rerun

Executed on June 15, 2025 16:02 by [li2g2uc](#)

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

Any instance could work, lightweight

App: [rescore_ms2pip](#) - Revision: 2

Inputs

input1

[peptides.csv](#)

input2

No files selected

	A	B	C
1	seq	charge	
2	NEVTTEIRF	2	
3	VTYNYPVHY	2	
4			
5			

csv file

App Settings

mode

Show non-default

ms2pip_intensity

Output Settings

output_mirror

No value

output_ms2pip_prediction

[ms2pip_prediction_NEVTTEIRF_2.png](#)

[ms2pip_prediction_NEVTTEIRF_2.tsv](#)

[ms2pip_prediction_VTYNYPVHY_2.png](#)

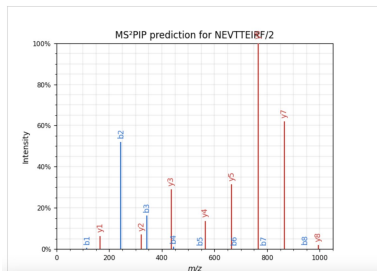
[ms2pip_prediction_VTYNYPVHY_2.tsv](#)

Files

[ms2pip_prediction_NEVTTEIRF_2.png](#)

25.2 KIB (25,774 bytes) - Produced on June 15, 2025 16:07 (Eastern Daylight Time), by [rescore_ms2pip run](#)

Metadata Raw View Preview



Step 2: Compare experimental and predicted spectra

COMPLETED **rescore_ms2pip run - 06-15-25 20:08:31**

Get support

View stats & logs

Edit and rerun

Executed on June 15, 2025 16:09 by li2g2uc

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

App: rescore_ms2pip - Revision: 2

From last step

Inputs

input1

ms2pip_prediction_NEVTTEIRF_2.tsv

input2

20240110_E_OdinLC_IC_PDX_HD_19.mzML

App Settings

mass_analyzer

mode

scan

Show non-default

Output Settings

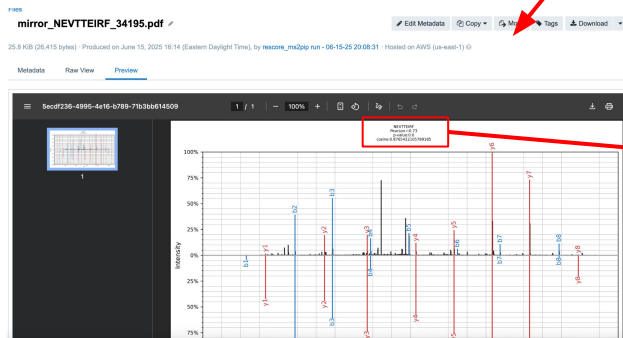
output_mirror

mirror_NEVTTEIRF_34195.pdf

output_ms2pip_prediction

No value

You should have
mzml file for each
raw file



NEVTTEIRF
Pearson r:0.73
p-value:0.0
cosine:0.8765422105789185