

Alternative-splicing detection by NGS

Software & dataset preparation

Wen-Dar Lin

Bioinformatics core, IPMB

wdlin@gate.sinica.edu.tw

Important notice

- This PPT is for people who want to practice at the training course.
- This PPT provides steps that set the dataset and the computing environment.
 - 3 setting steps: dataset, computing environment, and visualization tool
 - Be sure to have about 7GB available space
- The setting may take time so please be sure to perform the setting at least one day before the course.
 - *Please contact me directly if you have any problem on setting up the environment.*

Step 1: dataset

- The only dataset file is available at
 - 1. AS depositar:
<https://data.depositar.io/en/dataset/exampledata-zip-20260922>
 - OR 2. Figshare:
<https://doi.org/10.6084/m9.figshare.33773359>
 - about 750MB
 - Please download and unzip it
 - There should be an “ExampleData” folder. **Remember where this folder is.**
 - If you are using linux system, execute “chmod -R 755 ExampleData” to make sure the folder writable

Step 2: computing environment

- The computing environment requires a number of packages. The current best way to duplicate an exactly-the-same environment is the container solution.
- We provide a docker image
 - so that you can install just *docker desktop*, *docker*, or *singularity* and use the image to create an exactly-the-same environment.
 - *See page 11 if you don't want to use the image.*

Step 2: computing environment

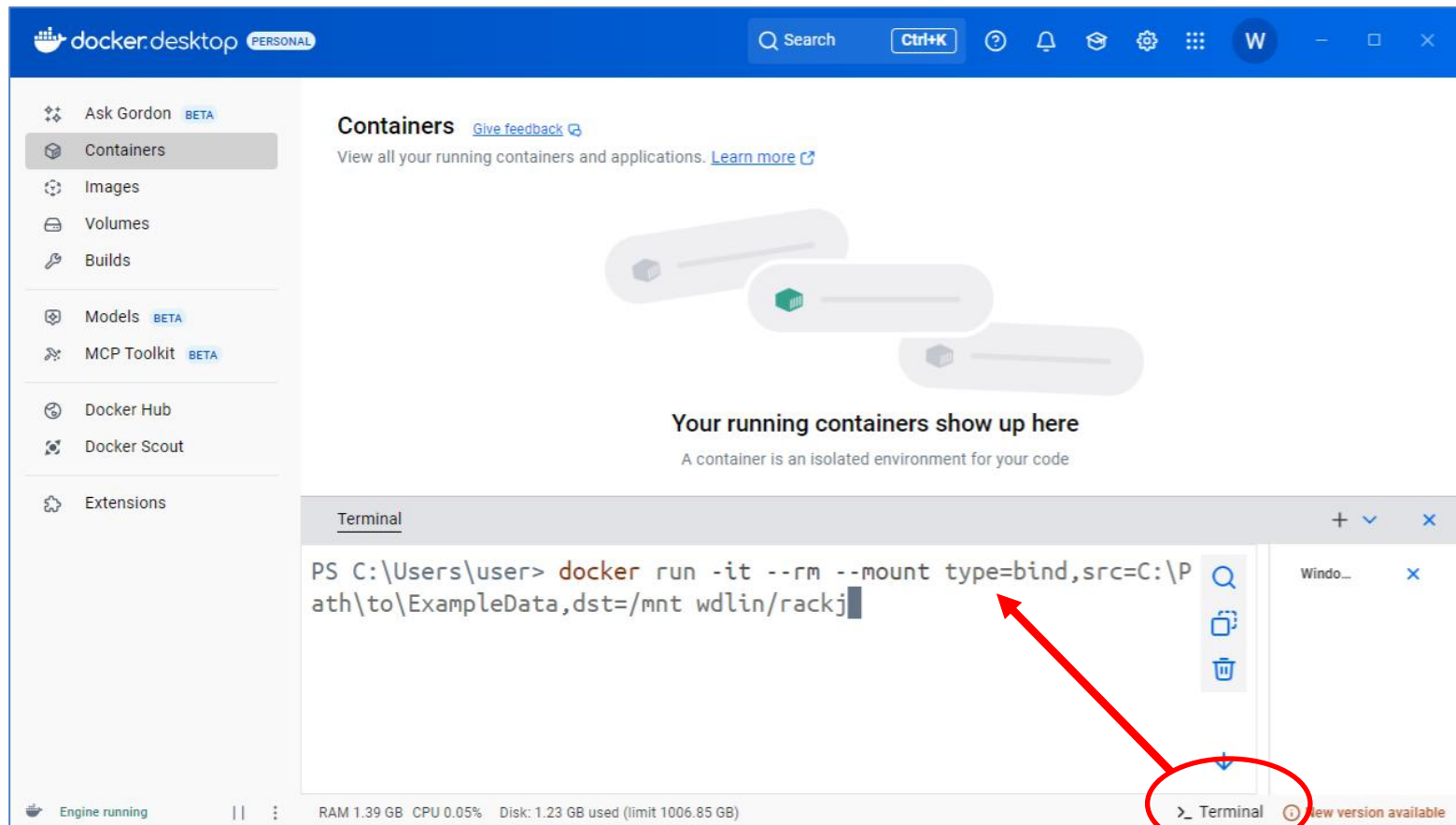
- Docker desktop
 - For computer with a graphical user interface (ex: win11 and macOS)
 - Windows: <https://docs.docker.com/desktop/setup/install/windows-install/>
 - Mac: <https://docs.docker.com/desktop/setup/install/mac-install/>
 - Linux: <https://docs.docker.com/desktop/setup/install/linux/>
- Docker
 - For linux command-line interface
 - <https://docs.docker.com/engine/install/>
- SingularityCE 3.x (or Apptainer 1.x)
 - Usually available in HPC environment
 - Ask your system admin for it

Step 2: computing environment

- Commands to create the environment
 - Docker desktop
 - Open *terminal* (bottom-right corner of Docker Desktop) and enter:
docker run -it --rm --mount
type=bind,src=<path_to_ExampleData>,dst=/mnt
wdlin/rackj
 - Docker (command line)
 - sudo docker run -it --rm --mount
type=bind,src=<path_to_ExampleData>,dst=/mnt
wdlin/rackj
 - Singularity
 - singularity run --bind "<path_to_ExampleData>:/mnt"
docker://wdlin/rackj

Step 2: computing environment

(example of using Docker Desktop, Win11)



Step 2: computing environment

- No matter which approach you pick from the last page, must run the command at least one time to enter the computing environment
 - so that the image would be cached in your system.

Step 2: computing environment

- If you entered the environment, it should be a linux environment and our data files are at /mnt
 - (example of using Singularity)

```
wdlin@comp01:/RAID2/R418/20260922_AS$ ml Singularity/3.8.5-GCCcore-11.2.0
wdlin@comp01:/RAID2/R418/20260922_AS$ singularity run --bind
"/RAID2/R418/20260922_AS/ExampleData:/mnt" docker://wdlin/rackj
INFO:      Using cached SIF image
Singularity> cd /mnt/
Singularity> ls -l
total 358193
-rwxr-xr-x+ 1 wdlin R418          86 Oct  4  2021 README.txt
-rwxr-xr-x+ 1 wdlin R418  49811410 Oct  4  2021 TAIR10_GFF3_genes_transposons.gff
-rwxr-xr-x+ 1 wdlin R418 121182577 Oct  4  2021 TAIR10_chr_all.fas
-rwxr-xr-x+ 1 wdlin R418  37814246 Sep 16 15:41 control_rep1.merged.bam
(... deleted)
drwxr-sr-x+ 2 wdlin R418          3 Sep 16 15:43 treatment_rep5_tophat2
-rwxr-xr-x+ 1 wdlin R418  38590467 Sep 16 15:41 treatment_rep7.merged.bam
drwxr-sr-x+ 2 wdlin R418          3 Sep 16 15:43 treatment_rep7_tophat2
-rwxr-xr-x+ 1 wdlin R418  37492552 Sep 16 15:41 treatment_rep9.merged.bam
drwxr-sr-x+ 2 wdlin R418          3 Sep 16 15:43 treatment_rep9_tophat2
Singularity> exit
```

Step 3: visualization tool

- We will use IGV for visualization of read alignments with genome annotation
 - Please install IGV Desktop Application for your laptop
 - <https://igv.org/doc/desktop/#DownloadPage/>

If you don't want to use the docker image

- Please refer <https://sourceforge.net/p/rackj/code/HEAD/tree/tags/trunk/Dockerfile> for package installation steps for a Ubuntu24.04 machine
 - The file is for building the docker image. Adjust some steps if needed.
 - You may skip the step of *rackj perl lib*. It is not related with the course.