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.

Step 1: dataset

- The only dataset file is at https://data.depositar.io/en/dataset/exampledata-zip-20250930
  - It is about 500MB
  - 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 last page 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:

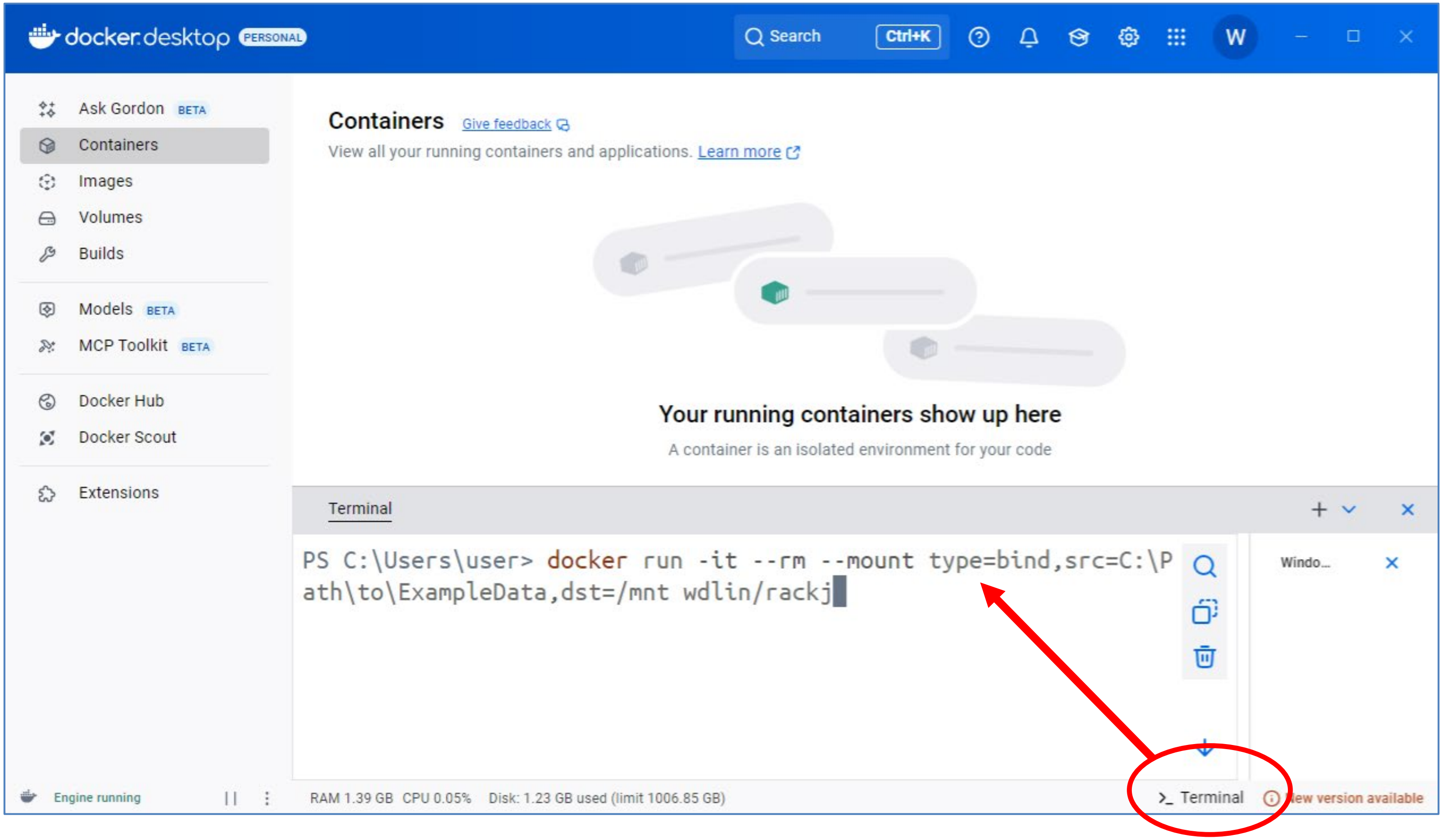
      ```
      docker run -it --rm --mount type=bind,src=<path_to_ExampleData>,dst=/mnt wdlin/rackj
      ```
  - Docker (command line)

    -
      ```
      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)



Containers

View all your running containers and applications. Learn more

Your running containers show up here

A container is an isolated environment for your code

```
PS C:\Users\user> docker run -it --rm --mount type=bind,src=C:\P
ath\to\ExampleData,dst=/mnt wdlin/rackj
```

RAM 1.39 GB CPU 0.05% Disk: 1.23 GB used (limit 1006.85 GB)

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@comp13:/RAID2/R418/20250930_AS$ singularity run --bind "/path/to/ExampleData:/mnt"
docker://wdlin/rackj
INFO:    Using cached SIF image
Singularity> cd /mnt/
Singularity> ls -l
total 358142
-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 22 11:14 control_rep1.merged.bam
-rwxr-xr-x+ 1 wdlin R418  38837864 Sep 22 11:14 control_rep2.merged.bam
-rwxr-xr-x+ 1 wdlin R418  36550185 Sep 22 11:14 control_rep4.merged.bam
drwxr-sr-x+ 2 wdlin R418        14 Sep 22 11:14 src
-rwxr-xr-x+ 1 wdlin R418   3806606 Oct  4  2021 tair10.strand.cgff
-rwxr-xr-x+ 1 wdlin R418   5238153 Oct  4  2021 tair10.strand.model
-rwxr-xr-x+ 1 wdlin R418  36394270 Sep 22 11:14 treatment_rep5.merged.bam
-rwxr-xr-x+ 1 wdlin R418  38590467 Sep 22 11:14 treatment_rep7.merged.bam
-rwxr-xr-x+ 1 wdlin R418  37492552 Sep 22 11:14 treatment_rep9.merged.bam
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.