Cogent™ NGS Analysis Pipeline v3.0.19 (beta) Quick Start Guide

The following information is provided as a high-level introduction to the software, also referred to as CogentAP. For more detailed information, please see the Cogent NGS Analysis Pipeline v2.0 User Manual.

Before you begin

A. Supported operating systems
- Linux: CentOS 8 or higher, RedHat 8 or higher, Ubuntu 18.04 or higher

B. Hardware requirements
- CPU: 24 cores
- Memory: 64 GB RAM
- Free disk space: at least 500 GB

C. Additional dependencies
- Internet connectivity on the server
- Conda 24.4.0 or higher
- Bash UNIX shell

D. Required input files
- FASTQ files generated by an Illumina® sequencing platform.

  NOTE: For a list of supported Takara Bio chemistries, please refer to our bioinformatics portal web page.
- A well-list text file, Illumina sample sheet, or similar TDT/CSV format file

Confirm Conda version

1. Verify conda is installed and meets or exceeds the required version by typing the following into a terminal window:

   ```
   conda -V
   ```

   If conda is successfully installed, it should return text with the version number.
   Example:

   ```
   conda 24.4.0
   ```

2. Verify the base conda environment can be activated by typing:

   ```
   conda activate
   ```

   If conda is successfully installed, the terminal will look similar to this:

   ```
   $ conda activate
   (base) $
   ```

   Type the following command to return to the default Linux prompt.

   ```
   conda deactivate
   ```

   ```
   (base) $ conda dectivate
   $
   ```

3. Verify the install location of Miniconda3 is configured in the file `.bash_profile`
   a. For an individual user account, type:

      ```
      more ~/.bash_profile
      ```

   b. Confirm something similar to the following is showing in the file (all on one line):

      ```
      export
      PATH="/home/<USERNAME>/miniconda3/
      bin:$PATH"
      ```

      where `<USERNAME>` is replaced by the username of the account that installed conda,

      If no `.bash_profile` file exists or the line isn't displaying, it will need to be manually created and populated.

Installation

1. Sign up to download the installation package from our website.
2. Move or copy the ZIP file downloaded from Step 1 onto the Linux server into the directory location where you want to install.
3. Unzip the installation package (all on one line):

```
unzip Cogent_NGS_Analysis_Pipeline_v3.0.19.zip \
&& mv Cogent_NGS_Analysis_Pipeline_v3.0.19.zip \
CogentAP
```

4. Run these commands sequentially to install the packages required to run CogentAP:

```
cd CogentAP
bash CogentAP_setup.sh install
```

5. After the conda environment packages install step completes, install the human genome (for the mouse genome, use mm10 rather than hg38):

```
bash CogentAP_setup.sh genome_install
hg38
```

Generation of raw FASTQ files

1. Log in to a server that stores the run folder from Illumina sequencing and has the bcl2fastq program installed.
2. Change to a working folder where you want the raw FASTQ files to be located after being generated.
3. To convert BCL files to FASTQ files using bcl2fastq, *go to Step #4*. If using BCL Convert, *go to Step #5*.
4. Run bcl2fastq with the following syntax template:

```
bcl2fastq -R <RUN_FOLDER> -o <RUN_ID>
--no-lane-splitting --sample-sheet
$COGENT_AP_HOME/config/SampleSheet_dum
my.csv > <RUN_ID>.stdout 2 >
<RUN_ID>.stderr
```

   where:

   - `<RUN_FOLDER>` is the path to the sequencing run folder
   - `<RUN_ID>` is the ID number automatically generated by the Illumina sequencer

   The file `SampleSheet_dummy.csv` is stored in the CogentAP `config` folder

   Go to Step #6.

   NOTE: See Section IV.D of the Cogent Analysis Pipeline v2.0 User Manual for how to set up the `$COGENT_AP_HOME` variable.

5. Run BCL Convert with the following syntax template:

```
bcl-convert --bcl-input-directory
<RUN_FOLDER> --output-directory
<RUN_ID> --no-lane-splitting --sample-
sheet=DummySampleSheet >
<RUN_ID>.stdout 2 > <RUN_ID>.stderr
```

   Templates for the `DummySampleSheet` for BCL Convert are stored in the CogentAP `config` folder.

   Go to Step #6.

6. Move the raw FASTQ files to your preferred storage location. They are typically generated in the `<RUN_ID>` folder and named similar to:

```
Undetermined_S0_R1_001.fastq.gz
Undetermined_S0_R2_001.fastq.gz
```

Running Cogent NGS Analysis Pipeline Software

- To demultiplex (demux)

```
<COGENT_AP_HOME>/cogent rna demux \
  -f <FASTQ_R1> \
  -p <FASTQ_R2> \
  -b <WELL-LIST> \
  -t <EXP_TYPE> \
  -o <OUTPUT>
```

- To analyze

```
<COGENT_AP_HOME>/cogent rna analyze \
  -i <DEMUX_RESULT_DIR> \
  -g <GENOME> \
  -t <EXP_TYPE> \
  -o <OUTPUT>
```

where:

- `<COGENT_AP_HOME>` is the path to the directory where CogentAP is installed
- `<FASTQ_R1>` and `<FASTQ_R2>` are the full paths to the FASTQ files generated by an Illumina sequencing platform.
- `<WELL-LIST>` is the full path to the ICELL8 system WellList, Illumina's sample sheet or TDT/CSV format file
- `<EXP_TYPE>` is the experiment type used (e.g., `icell8_fla`, refer to the CogentAP User Manual for more options)
- `<OUTPUT>` is a string; it will be the name of the output folder created by the analysis AND the prefix of all the results files
- `<DEMUX_RESULT_DIR>` is the full path of the demultiplex result directory you specified in `demux` command
- `<GENOME>` is a name of genome build (e.g., `hg38`)

Contact Us	
Customer Service/Ordering	**Technical Support**
tel: 800.662.2566 (toll-free)	tel: 800.662.2566 (toll-free)
fax: 800.424.1350 (toll-free)	fax: 800.424.1350 (toll-free)
web: takarabio.com/service	web: takarabio.com/support
e-mail: ordersUS@takarabio.com	e-mail: technical_support@takarabio.com

Notice to Purchaser

Our products are to be used for **Research Use Only**. They may not be used for any other purpose, including, but not limited to, use in humans, therapeutic or diagnostic use, or commercial use of any kind. Our products may not be transferred to third parties, resold, modified for resale, or used to manufacture commercial products or to provide a service to third parties without our prior written approval.

Your use of this product is also subject to compliance with any applicable licensing requirements described on the product's web page at takarabio.com. It is your responsibility to review, understand and adhere to any restrictions imposed by such statements.

Takara Bio USA, Inc.

United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

10.24 US