Cogent™ NGS Analysis Pipeline v1.0 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 v1.0 User Manual.

Before you begin

A. Supported operating systems
- Linux: CentOS 6.9 or higher, RedHat 7.5 or higher, Ubuntu 17 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
- Miniconda3 4.8.2 or higher
- Bash UNIX shell

D. Required input files
- FASTQ files generated by an Illumina® sequencing platform on results from:
  - ICELL8® cx Single-Cell System workflow
  - ICELL8 Single-Cell System workflow
  - SMARTer® Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian
- Well-list text file from the ICELL8 system or similar TDT/CSV format file. For more about this file, please refer to the CogentAP User Manual.

Confirm Miniconda3 version

- Verify Miniconda3 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 4.8.2
  ```

- Verify the base conda environment can be activated by typing:

  `source activate` or `conda activate`

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

  ```
  $source activate
  (base) $
  ```

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

  `source deactivate` or `conda deactivate`

  ```
  (base) $source deactivate
  $
  ```

- Verify the install location of Miniconda3 is configured in the file `.bash_profile`

  a. For an individual user account, type:

  `more ~/.bash_profile`

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

  ```
  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.

NOTE: For more information on any command in this section, please refer to Section IV.A of the CogentAP User Manual.

Installation

1. Sign up to download the installation script from our website. You will receive an email automatically after signing up that will contain a password you will need in Step 4.

   - If the Takara Bio mappa™ software was previously installed on the same computer the Cogent NGS Analysis Pipeline Software will be installed, proceed to Step 2
   - If this is a new install, skip to Step 3

2. (Uninstall the previous version)
   a. If you want to save any files stored in the `mappa/` directory, make sure to move them to another folder location before continuing.
   b. Delete the `mappa/` directory and all subfolders.

(continued on page 2)

3. (Install the new version): Move or copy the installation script onto the Linux server into the directory location where you want to install.
4. From the same directory location in Step 3, run the command:

```
bash takarabio_Linux64_installer.sh
CogentAP <PASSWORD>
```

`<PASSWORD>` will be replaced by the unique password included in the sign-up confirmation email.

NOTE: If genomes other than human need to be loaded into the software prior to use, please refer to Section IV.E of the CogentAP User Manual.

Generate raw FASTQ files

1. Log into 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. 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

NOTE: For instructions to set up the `$COGENT_AP_HOME` environment variable referenced in the command, please see Section IV.C of the CogentAP User Manual.

4. Move the raw FASTQ files to your preferred storage location. They are typically named `Undetermined_<TEXT>.fastq.gz` and generated in the `<RUN_ID>` folder in the working folder.

To run the Cogent NGS Analysis Pipeline

UI:

1. Double-click on the executable (`CogentAP_launcher`) to run it.
2. Fill out the fields in the interface.
3. Click [Start].

Command line:

- To demultiplex (demux)

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

- To analyze

```
%COGENT_AP_HOME%/cogent analyze \
  -i <DEMUXED_FASTQ_R1> \
  -p <DEMUXED_FASTQ_R2> \
  -d <DEMUX_RESULT> \
  -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 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
- `<DEMUXED_FASTQ_R1>` and `<DEMUXED_FASTQ_R2>` are the full paths to the respective FASTQ files generated by the demux command.
- `<DEMUX_RESULT>` is the full path of the demultiplex results file named like `<TEXT>_counts_all.csv`
- `<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

06.20 US