BD® Single-Cell Multiomics Analysis Setup User Guide

For Research Use Only

Doc ID: 47383 Rev. 9.0

23-21333-02
10/2019

Becton, Dickinson and Company
BD Biosciences
2350 Qume Drive
San Jose, CA 95131 USA

bdbiosciences.com
scomix@bdscomix.bd.com

Regulatory information

For Research Use Only. Not for use in diagnostic or therapeutic procedures.

History

Revision	Date	Change made
Doc ID: 47383 Rev. 1.0	09/2017	Initial release.
Doc ID: 47383 Rev. 2.0	11/2017	—Added setup information for multiplex runs. —Rebranded document.
Doc ID: 47383 Rev. 3.0	12/2017	—Updated BD™ Data View content to latest version v1.1. —Moved note to ensure use of correct CWL files under Requirements. —Updated Define App Settings in Seven Bridges Genomics and local installation chapters.
Doc ID: 47383 Rev. 4.0	04/2017	Added chapter on a customer service.
Doc ID: 47383 Rev. 5.0	07/2018	—Removed chapter on a customer service. —Updated to BD™ Data View v1.2. —Added content to set up for analysis of experiments with BD™ AbSeq Ab-Oligos.

Revision	Date	Change made
Doc ID: 47383 Rev. 6.0	10/2018	—In the requirements for local installation, clarified that Microsoft® Windows® is not supported and specified that Python 2.7.15 or later is required. —For CWL-runner on a local installation, added a recommendation of ≥32 GB memory limit. —Clarified that local installation is supported by most Unix-like operating systems.
Doc ID: 47383 Rev. 7.0 23-21333-00	02/2019	Added reference to the BD™ Mouse Immune Single-Cell Multiplexing Kit.
Doc ID: 47383 Rev. 8.0 23-21333-01	07/2019	Added reference to BD Rhapsody™ System Whole Transcriptome Analysis (WTA).
Doc ID: 47383 Rev. 9.0 23-21333-02	10/2019	Added reference to BD Rhapsody™ System Whole Transcriptome Analysis (WTA) and AbSeq

Contents

1

Introduction

About this guide

Introduction

This guide provides detailed instructions on how to set up and run the BD Rhapsody™ Analysis pipelines for sequencing on the Seven Bridges Genomics platform or on a local installation.

For references, including third-party tools, see the *BD Single-Cell Multiomics Bioinformatics Handbook* (Doc ID: 54169).

Genomics technical publications are available for download from the BD Genomics Resource Library at scomix.bd.com/hc/en-us/categories/360000838932-Resource-Library.

2

Requirements

Seven Bridges Genomics platform

Introduction

Create an account only if you will analyze sequencing data on the Seven Bridges Genomics platform.

Seven Bridges Genomics account

1. Go to sbgenomics.com/bdgenomics.
2. Click **Request Access**. In the request access window, enter your email address so that you can receive an email invitation to the Seven Bridges Genomics platform within 24 hours.
3. Click the link in the email invitation, and complete the registration. Seven Bridges Genomics displays the dashboard with the demo projects.

Local installation

Introduction

The system that runs BD Rhapsody™ analyses must meet certain minimum requirements. See Minimum system requirements.

The software applications required for analysis have specific software tools. To ensure that these tools are always available, the analysis is run in a self-contained environment called a docker container. The docker container is obtained by "pulling" or downloading a docker image to your local computer. The docker container has all of the libraries and settings required by the pipeline to run the analysis. In the portable docker container, the analysis can be run reproducibly wherever it is deployed, whether on a local installation or the Seven Bridges Genomics platform. CWL-runner is the tool that manages docker containers to complete the pipeline run. CWL-runner uses two inputs: a CWL workflow file and a YML input specification file. The CWL workflow file describes each step in the pipeline and how each docker container should run to complete the step. The YML file tells CWL-runner where to find the pipeline inputs, such as the sequencer read files and gene panel reference. When the pipeline run is finished, CWL-runner obtains the final outputs in the docker containers and adds them to a designated output folder on your computer.

Minimum system requirements

- Operating system: macOS® or Linux®. Microsoft® Windows® is not supported.
- 8-core processor (>16-core recommended)
- RAM
  - Targeted assays: 32 GB RAM (>128 GB recommended)
  - Whole Transcriptome Analysis (WTA) assays: 96 GB (>192 GB recommended)
- 250 GB free disk space (>1 TB recommended)

Software requirements

Docker

Install the community edition at store.docker.com.

Ensure that docker is running by entering `docker` at the command line.

The docker manual should print to the terminal screen.

Python 2.7.15 or later

1. Check to see if Python 2.7.15 or later is already installed by running at the command line:

   ```
   $ python2 --version
   ```

2. Ensure that you are using a local installation of Python and not a system version. Run:

   ```
   $ which python
   ```

   This should return the path to a local installation and not to a system path (usually /usr/bin/python).

   Using a system installation of python might not give you sufficient permissions to install the required packages.

3. If Python 2.7.15 or later is not installed, download and install it from python.org/downloads.

4. If pip is not installed, go to pip.pypa.io/en/stable/installing, and follow the instructions.

5. Update pip before installing cwlref-runner by using the command:

   ```
   $ pip install -U pip
   ```

CWL-runner

1. Install the package from PyPi. Enter:

   ```
   $ pip install cwlref-runner
   ```

2. Ensure that cwl-runner is in your path. Type:

   ```
   $ cwl-runner
   ```

3. If the command is not found, add the install location of the pip packages to $PATH.

   a. Find where cwlref-runner is installed by entering:

      ```
      $ pip show cwlref-runner
      ```

   b. Add the above path to $PATH. For example:

      ```
      $ export PATH=$PATH:/Library/Frameworks/
      Python.framework/Versions/2.7/lib/python2.7
      ```

   c. Restart the command line utility.

CWL and YML files

Ensure that you are using the correct CWL files with your pipeline, or the analysis might fail.

1. If necessary, create a Bitbucket account. Go to bitbucket.org/CRSwDev/cwl.
2. In the left pane, click **Downloads > Download Repository**. The CWL and YML files are downloaded.
3. Unzip the archive. Each folder within the archive is named after the pipeline version it corresponds to.

FASTQ files

Dataset size

BD Biosciences recommends analyzing datasets that are ≤100 GB in size. For datasets (compressed FASTQ FILES from all libraries) >100 GB, contact **BD Biosciences technical support at** scomix@bdscomix.bd.com**.**

Read 1 and Read 2 sequencing files

For the Seven Bridges Genomics platform and local installation, obtain Read 1 and Read 2 sequencing files, and ensure that the FASTQ file names follow these rules:

- An underscore on each side of R1 or R2 (_R1_ and _R2_).
- The <sample> name should be the same for R1 and R2.
- Convert uncompressed files to .gz format.
- Basename of file should end with 001.

  Example:
  <sample>_S1_L001_R1_001.fastq.gz

  <sample>_S1_L001_R2_001.fastq.gz

Do not use special characters or spaces in the filenames, or the analysis might fail. Use only letters, numbers, or hyphens.

Note: If you are downloading the files from BaseSpace, follow these steps:

a. Choose the run to download in BaseSpace.

b. Click the download icon on the main screen.

c. If necessary, install the BaseSpace downloading application.

d. Click **Select all fastq files for this run**.

e. Download the files. This might take several minutes.

For more information, go to help.basespace.illumina.com.

Reference files

Introduction

For targeted assays, separate FASTA reference files are used to store the sequences of gene targets and BD® AbSeq Ab-Oligos (antibody-oligonucleotides) that are used in a BD Rhapsody experiment.

For WTA assays, the reference genome is a compressed tarball that contains the STAR index files for the species of the cells used in the BD WTA experiment. The transcriptome annotation is a GTF file containing gene structure information.

Obtaining pre-designed mRNA panels

Obtain the FASTA panels from the Seven Bridges demo project or by contacting BD Biosciences customer support at scomix@bdscomix.bd.com.

For WTA assays, obtain the reference genome file from the Seven Bridges demo project, downloading from the following link: http://bd-rhapsody-public.s3-website-us-east-1.amazonaws.com/Rhapsody-WTA/ (link cannot be accessed using Internet Explorer), or contact BD Biosciences customer support.

STAR reference/ transcriptome annotation

The GTF file has been preprocessed to contain information for the following gene types: protein_coding, lincRNA, antisense, IG_LV_gene, IG_V_gene, IG_V_pseudogene, IG_D_gene, IG_J_gene, IG_J_pseudogene, IG_C_gene, IG_C_pseudogene, TR_V_gene, TR_V_pseudogene, TR_D_gene,TR_J_gene, TR_J_pseudogene and TR_C_gene.

Designing supplemental or custom mRNA panels

By providing a list of genes to BD Biosciences customer support, we can design custom mRNA targeted panels. Contact BD Biosciences customer support at scomix@bdscomix.bd.com.

For custom reference genome files, contact BD Biosciences customer support at scomix@bdscomix.bd.com.

Downloading, preparing, and saving an AbSeq reference file

If your experiment contains BD AbSeq Ab-Oligos, you are required to have an AbSeq reference file.

1. Download the FASTA file containing all of the BD Ab-Oligo (AbO) sequence. Go to bd-rhapsody-public.s3-website-us-east-1.amazonaws.com/AbSeq-references/BDAbSeq_allReference_latest.fasta.
2. Use a text editor such as Microsoft® Notepad or TextEdit to delete the sequence header and sequence pairs that will not be used in the experiment.

   Do not use a word processor such as Microsoft® Word, which can add unintended special characters to the file.
3. Ensure that the AbSeq reference file follows these rules:
   - File extension is .fa or .fasta
   - Format is:

```
>CD103|ITGAE|AHS0001|pAbO
AAATAGTATCGAGCGTAGTTAAGTTGCGTAGCCGTT
>CD161|KLRB1|AHS0002|pAbO
GTTATGGTTGTCGGTAGAGTATCGTGTTGCGTTAGT
```

   Note: BD Biosciences uses this format for its sequence header: <AntibodyName>|<GeneSymbol>|<SeqID>|pAbO.
4. Save as an .fa or .fasta file locally on your computer.

3

Setting up sequencing analysis on Seven Bridges Genomics platform

Introduction

Whether analysis is performed on the Seven Bridges Genomics platform or locally, sequencing uses the BD Rhapsody™ Targeted Analysis Pipeline or BD Rhapsody™ WTA Analysis Pipeline. During the execution of the pipeline, sequencing analysis processes sequencing files to generate molecular counts per cell, read counts per cell, metrics, and an alignment file.

Workflow

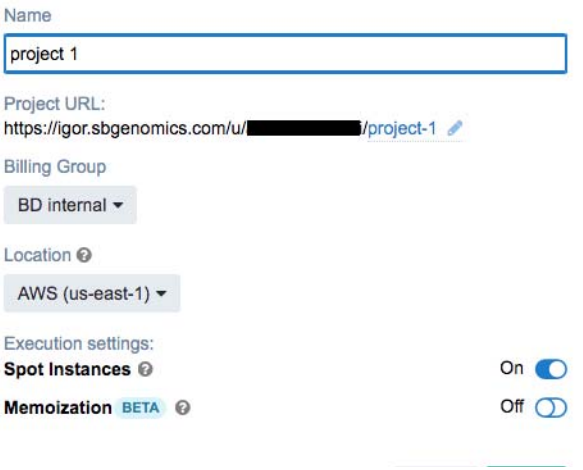
During sequencing analysis, the BD Rhapsody Targeted Analysis Pipeline or BD Rhapsody WTA Analysis Pipeline analyzes only one cartridge per run. To analyze multiple cartridges, create a pipeline run (or task) for each cartridge.

Step	Purpose
1	Create a new project.
2	Import FASTQ files.
3	Import the reference file.
4	Import the BD Rhapsody Targeted Analysis Pipeline or BD Rhapsody WTA Analysis Pipeline.
5	Set up and run the pipeline.
6	Download the output files.

Creating a new project

Procedure

1. At the top of the dashboard, click **Projects > Create a project:**

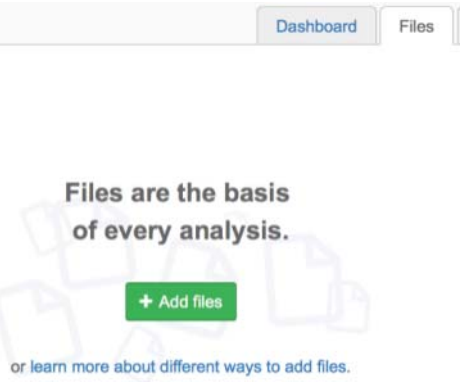


2. On the Create a project dialog, enter the project name, and edit the project URL if necessary.
3. Click **Create**. Seven Bridges Genomics displays the new project dashboard.

Importing FASTQ files

Procedure

1. On the project dashboard, click the **Files** tab, and then click **+Add files**:



2. In the top menu, select the source of the files, such as **Public files**, **Projects**, or **FTP/HTTP**. Seven Bridges Genomics displays instructions on uploading the files. Follow the Seven Bridges Genomics instructions to import your files.

   Use the Desktop Uploader to upload files from BaseSpace. Security permissions on your BaseSpace account prevent FTP/HTTP protocols from working.

3. After import, the files are on the Files tab.

Importing reference files

Importing files

1. On the **Files** tab of the project dashboard, click **+Add files**.
2. Click **Projects**, and then click on **BD Rhapsody Targeted Analysis Pipeline** or **BD Rhapsody WTA Analysis Pipeline** in the left panel.
3. Do one of the following:
   - For Targeted assays: Locate the appropriate FASTA file for your experiment, and click **Copy**.
   - For WTA assays: Locate the appropriate reference genome and transcriptome annotation files for your experiment, and click **Copy**.

Importing supplemental or custom mRNA panels or AbSeq reference files

1. On the project dashboard, click the **Files** tab, and then click **+Add files**.
2. In the top menu, select the source of the files, such as **Public files**, **Projects**, or **FTP/HTTP**. Seven Bridges Genomics displays instructions on uploading the files. Follow the Seven Bridges Genomics instructions to import your files.

   Use the Desktop Uploader to upload files from BaseSpace. Security permissions on your BaseSpace account prevent FTP/HTTP protocols from working.
3. After import, the files are on the Files tab.

Importing the BD Rhapsody pipeline

Importing the pipeline

1. On the project dashboard, click the **Apps** tab, and then click **+Add app**.
2. Click **Public Apps**, and then enter **Rhapsody** to find the appropriate pipeline. Or, copy the workflow from the Demo project.
   - Targeted assays: **BD Rhapsody Targeted Analysis Pipeline**
   - WTA assays: **BD Rhapsody WTA Analysis Pipeline**
3. Click **Copy** on the app window, select the project in the drop-down menu, and then click **Copy** again.
4. Navigate to the Apps tab to confirm that the workflow was copied to the project.

Setting up and running the pipeline

Procedure

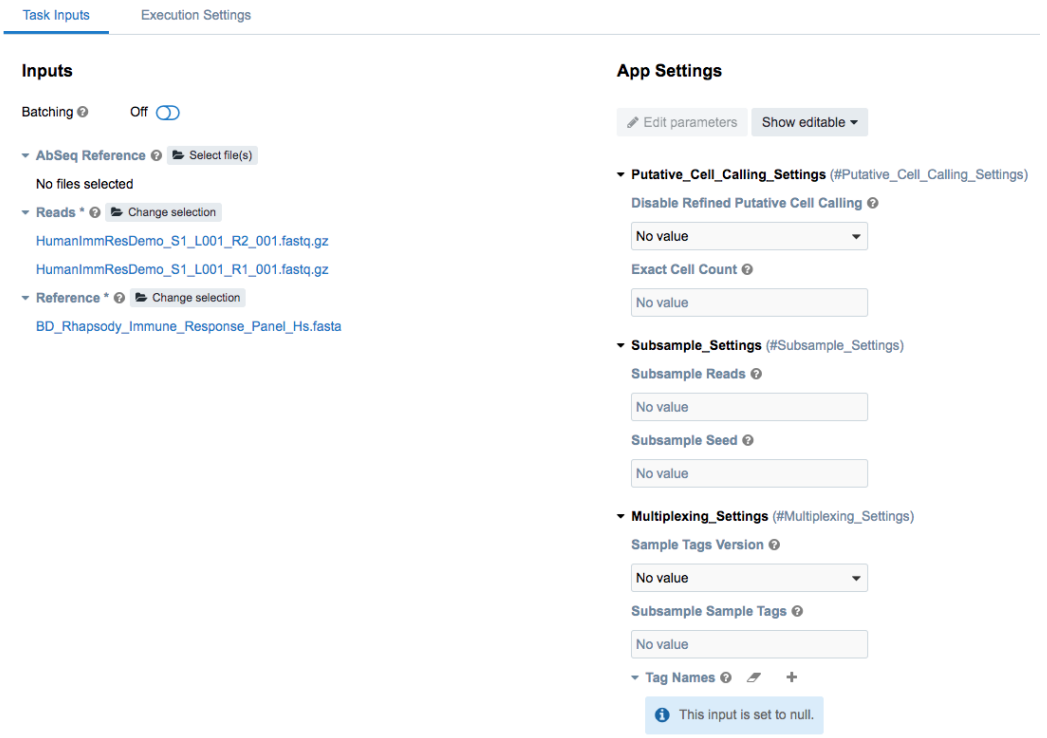
1. Click the **Apps** tab to view the apps.

   Note: If the app is highlighted in yellow, an update is available. Select the refresh icon to get the latest app version.

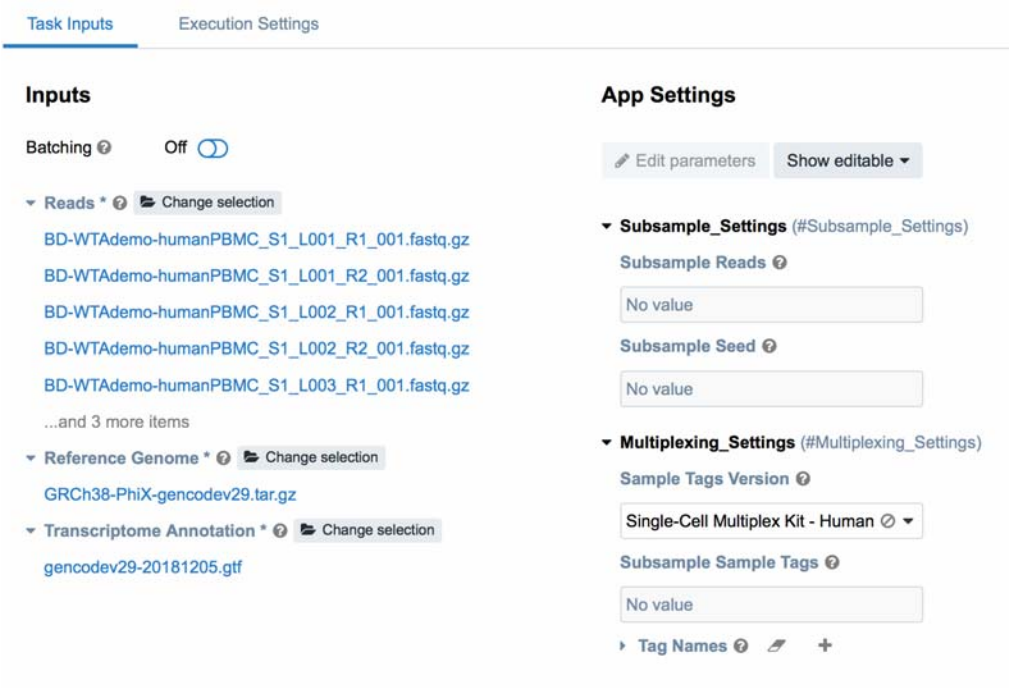
2. By the **BD Rhapsody Targeted Analysis Pipeline** or **BD Rhapsody WTA Analysis Pipeline**, click the green play button under Actions.

   For both targeted and WTA assays, the Task Inputs table displays the Inputs and App Settings.

Targeted Analysis Pipeline interface:



WTA Analysis Pipeline interface:



Complete all required fields, which appear in red.

Input field	Input	Required?
AbSeq Reference	FASTA AbSeq reference file generated from Importing supplemental or custom mRNA panels or AbSeq reference files (page 21). **Ensure that the AbSeq reference file contains the BD AbSeq Ab-Oligos that were used in the experiment. Otherwise, the read mapping will be incorrect.**	Optional
Reads	R1 reads and R2 reads. Ensure to include all FASTQ sequencing data from the experiment, including R1 and R2 files for the targeted RNA library, and, if applicable, the Sample Tag and BD® AbSeq libraries.	Yes
Reference (Targeted only)	This is an mRNA reference file. Select the FASTA reference file. This is a pre-designed, supplemental, or custom panel. **Ensure that the reference matches the species and panel used for the experiment. Otherwise, read mapping will not be correctly aligned.**	Yes
Sample Tags Version	For a multiplexed samples run only. Specifies the Sample Tags used: Single-Cell Multiplex Kit—Human Single-Cell Multiplex Kit—Mouse	Required for multiplexed samples
Subsample Sample Tags	For a multiplexed samples run only. Any number of reads >1 or a fraction of reads between 0<n<1 to indicate the percentage of reads to subsample per Sample Tag.	Optional for multiplexed samples

Input field (continued)	Input	Required?
Tag Names	For a multiplexed samples run only. To enter a new sample, click **+** to add a row. Enter one tag name per row. Use the following format, using a hyphen—**no spaces or forward slashes allowed**: **Sample Tag number-sample name** Example: 3-Ramos **Note:** Until the tag name is in the correct format, a red *expected type* warning message is displayed.	Optional for multiplexed samples
Subsample Reads	Any number of reads >1 or a fraction between 0< n<1 to indicate the percentage of reads to subsample.	Optional
Subsample Seed	For use when replicating a previous subsampling run only. Obtain the seed generated from the log file for the SplitFastQ node. To obtain the log file, see Downloading the log file from Seven Bridges Genomics (page 48). **Entering the seed ensures that the same reads are subsampled to reproduce the results. If no seed is needed, leave blank and the pipeline will generate one randomly.**	Optional
Reference Genome (WTA only)	This is a STAR indexed reference genome file ending in .tar.gz.	Yes
Transcriptome Annotation (GTF) (WTA only)	This is a file that describes gene structures and ends with .gtf.	Yes

1. On the Set Input Data tab, import your files for analysis according to these requirements:
   - For every R1 .fastq.gz file, import the paired R2 .fastq.gz file.

  - Multiple R1 and R2 reads can be run together as long as they are from the same library, but the files can be generated from different sequencer runs.

2. If necessary, set the options on the Define App Settings tab. For example:

   When using a BD® Single-Cell Multiplexing Kit, be sure to select the Sample_Tags_Version (Single-Cell Multiplex Kit - Human or Mouse) from the drop-down menu.

3. Click **Run**. Seven Bridges Genomics displays the app running on the Tasks tab.
4. If you enabled email notifications, look for notification of the completed run.

Downloading the output

Procedure

See Downloading output files on the Seven Bridges Genomics platform (page 42).

This page intentionally left blank

4

Setting up sequencing analysis on a local installation

Workflow

During sequencing analysis, the BD Rhapsody Targeted Analysis Pipeline or BD Rhapsody WTA Analysis Pipeline analyzes only one cartridge per run. To analyze multiple cartridges, create a pipeline run (or task) for each cartridge. During clustering analysis, multiple cartridges can be merged and analyzed together.

Step	Purpose
1	Set up the input specification file.
2	Run the pipeline using CWL-runner at the command line.

Setting up the input specification file

Procedure

The input specification file template.yml is downloaded from the CWL folder.

1. Obtain the FASTQ files. See Read 1 and Read 2 sequencing files (page 14).
2. Obtain the mRNA reference file or reference genome and transcriptome annotation files from BD Biosciences technical support at scomix@bdscomix.bd.com.
3. If your experiment contains BD AbSeq Ab-Oligos, obtain the AbSeq Reference file. See Downloading, preparing, and saving an AbSeq reference file (page 16).
4. Specify the desired file paths in the YML file for Reads and Reference with the exact input field listed in the table. (Optional) Define BAM input, subsample, and subsample seed input fields.
   - The required input fields for Targeted assays are Reads and Reference.
   - The required input fields for WTA assays are Reads, Reference_Genome, and Transcriptome_Annotation.

Input field	Input	Required?
Reads	R1 reads and R2 reads. Ensure to include all FASTQ sequencing data from the experiment, including R1 and R2 files for the targeted RNA library, and, if applicable, the Sample Tag and BD AbSeq libraries.	Yes
Reference (Targeted only)	Select the FASTA reference file. This is a pre-designed, supplemental, or custom panel.	Yes
Reference_ Genome (WTA only)	Select STAR index (tar.gz). This is a pre-built index or a custom index.	Yes

Input field (continued)	Input	Required?
Transcriptome_ Annotation (WTA only)	Select the GTF file.	Yes
AbSeq_Reference	FASTA AbSeq reference file generated from Importing supplemental or custom mRNA panels or AbSeq reference files (page 21). **Ensure that the AbSeq reference file contains the BD AbSeq Ab-Oligos that were used in the experiment. Otherwise, the read mapping will be incorrect.**	Optional
Subsample	Any number of reads >1 or a fraction between 0< n<1 to indicate the percentage of reads to subsample.	Optional
Subsample_seed	For use when replicating a previous subsampling run only. Obtain the seed generated from the log file for the SplitAndSubsample node. To obtain the log file, see Downloading the log file from Seven Bridges Genomics (page 48). **Entering the seed ensures that the same reads are subsampled to reproduce the results. If no seed is needed, leave blank, and the pipeline will generate one randomly.**	Optional
Sample_Tags _Version	For a multiplexed samples run only. Specifies the Sample Tags used: human (hs), mouse (mm).	Required for multiplexed samples

Input field (continued)	Input	Required?
Subsample_Tags	For a multiplexed samples run only. Any number of reads >1 or a fraction of reads between 0<n<1 to indicate the percentage of reads to subsample per Sample Tag.	Optional for multiplexed samples
Tag_Names	For a multiplexed samples run only. Associate a name with each Sample Tag, which will appear in the output files. Within square brackets, enter a comma-separated list of Sample Tag numbers and associated names. For each sample, use the following format, using a hyphen—**no spaces or forward slashes allowed**: **Sample Tag number-sample name** Example: Tag_Names: [3-Ramos, 4-BT549]	Optional for multiplexed samples

5. If necessary, specify multiple R1 and R2 reads under **Reads** by including additional file objects and following the nomenclature for each file. For example:

```
-class: File
location: "path/to/additional_R1_fastq.gz"
```

For example:

YML file example showing a pair of FASTQ files and a panel reference file as input

Targeted:

```
#!/usr/bin/env cwl-runner
cwl:tool: Rhapsody

Reads:
 - class: File
   location: path/to/mySample_R1_.fastq.gz
 - class: File
   location: path/to/mySample_R2_.fastq.gz

Reference:
 - class: File
   location: path/to/reference.fasta
AbSeq_Reference:
 - class: File
   location: path/to/abseq_reference.fasta
```

WTA:

```
#!/usr/bin/env cwl-runner
cwl:tool: Rhapsody

Reads:
 - class: File
   location: path/to/sample_S1_L001_R1_001.fastq.gz
 - class: File
   location: path/to/sample_S1_L001_R2_001.fastq.gz
Reference_Genome:
   class: File
   location: path/to/reference.tar.gz
Transcriptome_Annotation:
   class: File
   location: path/to/annotation.gtf
```

YML file example showing 50% subsampling of the reads

Targeted:

```
#!/usr/bin/env cwl-runner
cwl:tool: Rhapsody

Reads:
 - class: File
   location: "test/mySample2_R2_.fastq.gz"
 - class: File
   location: "test/mySample2_R1_.fastq.gz"


Reference:
 - class: File
   location: "test/Immune_Response_Panel_Hs_with_Phix.fasta


Subsample: 0.5
```

WTA:

```
#!/usr/bin/env cwl-runner
cwl:tool: Rhapsody

Reads:
 - class: File
   location: path/to/sample_S1_L001_R1_001.fastq.gz
 - class: File
   location: path/to/sample_S1_L001_R2_001.fastq.gz
Reference_Genome:
   class: File
   location: path/to/reference.tar.gz
Transcriptome_Annotation:
   class: File
   location: path/to/annotation.gtf
Subsample: 0.5
```

YML file example showing choice of human Sample Tags, 50% subsampling of reads per Sample Tag, and Sample Tag naming

Targeted:

```
#!/usr/bin/env cwl-runner
cwl:tool: mist

Reads:
 - class: File
   location: /path/to/mySample_R1_.fastq.gz
 - class: File
   location: /path/to/mySample_R2_.fastq.gz
 - class: File
   location: /path/to/mySampleTag_R1_.fastq.gz
 - class: File
   location: /path/to/mySampleTag_R2_.fastq.gz

Reference:
 - class: File
   location: /path/to/targeted_sampleTags.fasta

Sample_Tags_Version: human

Subsample_Tags: 0.5

Tag_Names: [4-mySample, 9-myOtherSample, 6-alsoThisSample]
```

WTA:

```
#!/usr/bin/env cwl-runner
cwl:tool: rhapsody

Reads:
 - class: File
   location: /path/to/Sample_S1_L001_R1_001.fastq.gz
 - class: File
   location: /path/to/Sample_S1_L001_R2_001.fastq.gz
 - class: File
   location: /path/to/SampleTag_S1_L001_R1_001.fastq.gz
 - class: File
   location: /path/to/SampleTag_S1_L001_R2_001.fastq.gz
Reference_Genome:
   class: File
   location: /path/to/reference.tar.gz
Transcriptome_Annotation:
   class: File
   location: /path/to/annotation.gtf
Sample_Tags_Version: human
Subsample_Tags: 0.5
Tag_Names: [4-mySample, 9-myOtherSample, 6-alsoThisSample]
```

6. Save the modified template YML file.

Running the pipeline

Procedure

See Running a pipeline using CWL-runner (page 37).

5

Running a pipeline using CWL-runner

Running CWL-runner on a local installation

Procedure

Local installation is supported by most Unix-like operating systems such as macOS or Linux. Minimum system requirements must be met. See Local installation (page 11).

To run the pipeline on macOS, perform these additional configuration steps:

1. To enable CWL-runner to set up volumes, run the command:

   ```
   $ export TMPDIR=/tmp/docker_tmp
   ```

2. To increase the memory available to docker:

   a. Click the docker icon in the menu bar to open the docker menu.

   b. Click **Preferences**, and navigate to the Advanced tab.

   c. Use the slider to increase the memory limit. BD Biosciences recommends ≥32 GB for Targeted and ≥64 GB for WTA. See Local installation (page 11). Lower limits are sufficient for smaller datasets.

   d. Click **Apply & Restart** at the bottom of the window.

Running CWL-runner

1. In the terminal, ensure that you are in a directory that contains the CWL files that were downloaded from the Bitbucket repository. The edited YML file for input specifications must also be present in this directory. See Setting up sequencing analysis on a local installation (page 29).

2. Run the pipeline by entering the command:

   ```
   $ cwl-runner workflow.cwl input.yml
   ```

   If running the sequencing analysis pipeline, the workflow is the file rhapsody.cwl, and the input specification file is the edited template.yml.

   If running the clustering analysis pipeline, the workflow is the file ClusteringAnalysis.cwl, and the input specification file is the edited ClusteringAnalysis-template.yml.

3. If desired, you can specify the output directory for the analysis using the flag `--outdir`

   An example command:

   ```
   $ cwl-runner --outdir
   /path/to/results_folder rhapsody.cwl my_sample.yml
   ```

   Note: The output directory must be an existing directory. If no output directory is specified, files are output to the working directory.

4. Confirm that the following message displays after the pipeline is completed:

   ```
   Final process status is success.
   ```

5. Access the output files. All output files are found in the output directory specified in the CWL-runner command. If no output directory is specified, the files are output to the directory from which the command was called. See Reviewing output files (page 41).

This page intentionally left blank

6

Reviewing output files

Downloading output files on the Seven Bridges Genomics platform

Procedure

1. Select the project from the Projects drop-down menu to view output files.
2. Click the **Tasks** tab to view the list of tasks.
3. Click the name of the completed task to view Outputs on the right of the screen.
4. Click **Download** to download and save the output file. To download all files at once, click the **Files** tab, click the checkboxes by the files to download, and then click **Download**.
5. View the output files. See Sequencing analysis output files (page 43).

Sequencing analysis output files

Most output files contain a header summarizing the pipeline run. Headers contain all of the information needed to rerun the pipeline with the same settings.

Output	File	Content
Metrics summary	<sample_name>_Metrics_Summary.csv	Report containing sequencing, molecules, and cell metrics
BAM and BAM Index	<sample_name>.final.BAM <sample_name>.final.BAM.bai	Alignment file of R2 and associated R1 annotations
Data tables[a]	<sample_name>_RSEC_MolsPerCell.csv <sample_name>_RSEC_ReadsPerCell.csv <sample_name>_DBEC_MolsPerCell.csv <sample_name>_DBEC_ReadsPerCell.csv	Reads per gene per cell and molecules per gene per cell, based on RSEC or DBEC
	<sample_name>_RSEC_MolsPerCell_Unfiltered.csv.gz <sample_name>_RSEC_ReadsPerCell_Unfiltered.csv.gz <sample_name>_DBEC_MolsPerCell_Unfiltered.csv.gz <sample_name>_DBEC_ReadsPerCell_Unfiltered.csv.gz	Unfiltered tables containing all cell labels of ≥10 reads
Expression data[a]	<sample_name>_Expression_Data.st	The expression sparse matrix, a table of counts in sparse format
	<sample_name>_Expression_Data_Unfiltered.st.gz	Compressed file containing all cell labels of ≥10 reads

Output (continued)	File	Content
Cell label filtering	<sample_name>_Cell_Label_Filter.png	Visualization of cell label filtering results
Second derivative curve	<sample_name>_Cell_Label _Second_Derivative_Curve.png	
Putative cells origin	<sample_name>_Putative_Cells _Origin.csv	Algorithm that found the putative cell: basic or refined
Unique Molecular Identifier (UMI) metrics	<sample_name>_UMI_Adjusted_Stats.csv	Metrics from RSEC and DBEC Unique Molecular Identifier adjustment algorithms on a per-gene basis

a. For a multiplexed samples run, the tables contain counts for putative cells from all samples combined.

If the multiplex option was selected, the following outputs are generated:

Output	File	Content
Sample Tags metrics	<sample_name>_Sample_Tag_Metrics.csv	Metrics from the sample determination algorithm
Sample Tag calls	<sample_name>_Sample_Tag_Calls.csv	Assigned Sample Tag for each putative cell
Per-sample folder	<sample_name>_Sample_Tag<number>.zip <sample_name>_Multiplet_and _Undetermined.zip	Data tables and expression matrix for a particular sample. **Note:** For putative cells that could not be assigned a specific Sample Tag, a Multiplet_and _Undetermined.zip file is also output.

Reviewing output files

See the *BD Single-Cell Multiomics Bioinformatics Handbook* (Doc ID: 54169).

Genomics technical publications are available for download from the BD Genomics Resource Library at scomix.bd.com/hc/en-us/categories/360000838932-Resource-Library.

This page intentionally left blank

7

Troubleshooting

Analysis pipeline

Introduction

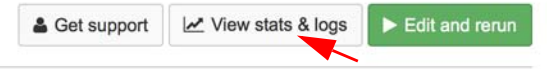
This topic describes how to respond to a task failure while running the BD Rhapsody Analysis Pipeline.

Arranging BD Biosciences to join the project on Seven Bridges Genomics

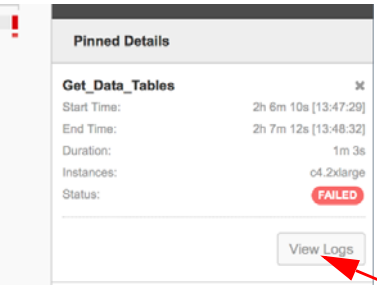
If a task fails on the Seven Bridges Genomics platform, contact BD Biosciences technical support at scomix@bdscomix.bd.com to troubleshoot the issue. Tech support will provide you with instructions on inviting a support team member to your project. To troubleshoot the issue yourself, access the log files. See Downloading the log file from Seven Bridges Genomics.

Downloading the log file from Seven Bridges Genomics

1. From within a failed task, click **View Stats & Logs** in the upper right corner:

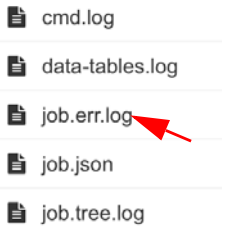


2. Locate the failed node in your pipeline run. Completed nodes are in green, and the failed node is in red. Click the failed node, and on the right, click **View Logs** for that node:



A list of files contained in the failed node are displayed.

3. Click **job.err.log** to display the log content and download it:



Accessing the log file in a local installation

If a pipeline run completed successfully, all logs are collected in a Logs folder in your output directory. But if a pipeline run fails, the Logs folder is absent from the directory. You need to navigate to the *tmp* directory containing the intermediate files for that node to obtain the log files:

1. In the terminal STDOUT, find the failed node command call from CWL-runner. This is the most recent command call.
2. Locate the tmp folder name, which is in the format:

   ```
   [job Name_of_failed_node] /tmp/tmpb0kyIg $
   ```

3. Navigate to that directory. The log file will have the .log extension.
4. Send the log file to scomix@bdscomix.bd.com, or contact BD Biosciences technical support without it.

This page intentionally left blank

8

Glossary

B

BAM An alignment file in binary format. A binary SAM file.

C

called cell A putative cell that has been assigned a Sample Tag.

CWL Common workflow language. A way to describe commands and to connect them to create workflows.

D

data tables Output of BD Rhapsody Targeted Analysis Pipeline containing read count or molecule count per gene.

DBEC Distribution-based error correction.

F

FASTA Text-based format that contains one or more DNA or RNA sequences.

FASTQ A file in standardized, text-based format that contains the output of base reads and per-base quality values from a sequencer.

L

library A sequencing library derived through amplification of genomic material that had been captured by a collection Cell Capture Beads from a BD Rhapsody™ kit.

P

putative cell A single cell determined to be putative by the cell label filtering algorithm.

R

R1 reads Contains information about the cell label and molecular identifier.

R2 reads Contains information about the gene.

RSEC Recursive substitution error correction.

S

SAM Tab-delimited text file with sequence alignment data.

Sample Tag Antibody-oligo tag that identifies a sample in a multiplexed run.

T

t-SNE t-distributed stochastic neighbor embedding (t-SNE). An algorithm for dimensionality reduction. It allows for the representation of high-dimensional data (multiple genes across multiple cells) into a two-dimensional space, which can then be visualized in a scatter plot.

Y

YML YAML: “Yaml ain't markup language.” A data serialization language used for configuration files to various applications.

This page intentionally left blank