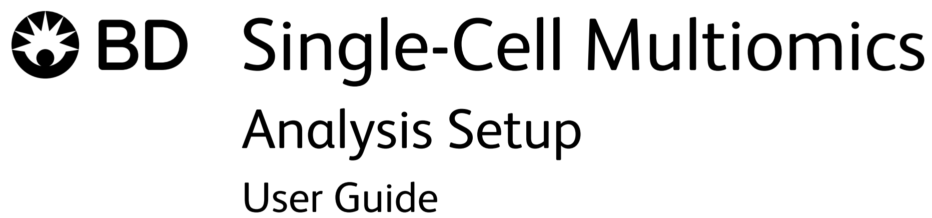


Single-Cell Multiomics

Analysis Setup

User Guide

23-21333(06)
2023-06

For Research Use Only

Copyrights

No part of this publication may be reproduced, transmitted, transcribed, stored in retrieval systems, or translated into any language or computer language, in any form or by any means: electronic, mechanical, magnetic, optical, chemical, manual, or otherwise, without prior written permission from BD.

The information in this guide is subject to change without notice. BD reserves the right to change its products and services at any time. Although this guide has been prepared with every precaution to ensure accuracy, BD assumes no liability for any errors or omissions, nor for any damages resulting from the application or use of this information. BD welcomes customer input on corrections and suggestions for improvement.

Trademarks

BD, the BD Logo and BD Rhapsody are trademarks of Becton, Dickinson and Company or its affiliates. All other trademarks are the property of their respective owners. © 2023 BD. All rights reserved.

For US patents that may apply, see bd.com/patents.

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	2017-09	Initial release.
Doc ID: 47383 Rev. 2.0	2017-11	—Added setup information for multiplex runs. —Rebranded document.
Doc ID: 47383 Rev. 3.0	2017-12	—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	2017-04	Added chapter on a customer service.
Doc ID: 47383 Rev. 5.0	2018-07	—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.
Doc ID: 47383 Rev. 6.0	2018-10	—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	2019-02	Added reference to the BD® Mouse Immune Single-Cell Multiplexing Kit.
Doc ID: 47383 Rev. 8.0 23-21333-01	2019-07	Added reference to BD Rhapsody™ System Whole Transcriptome Analysis (WTA).

Revision	Date	Change made
Doc ID: 47383 Rev. 9.0 23-21333(02)	2019-10	Added reference to BD Rhapsody™ System Whole Transcriptome Analysis (WTA) and AbSeq.
Doc ID: 47383 Rev. 10.0 23-21333(03)	2021-08	—Added updates for Putative Cell Calling and VDJ. —Added additional content describing the running of the pipeline on Seven Bridges and locally.
23-21333(04)	2022-08	—Added updates for new pipeline version.
23-21333(05)	2023-01	—Added support for BD® Flex Single-Cell Multiplexing Kit. —Added support for new cell label structure for BD Rhapsody™ Enhanced Cell Capture Beads v2.0. —Added support for v1.12 Bioinformatics pipeline. —Improved VDJ outputs.
23-21333(06)	2023-06	—Added support for v2.0 Bioinformatics pipeline.

Contents

1. **Introduction** ... 5

2. **Requirements** ... 6

Seven Bridges Genomics platform ... 6
Local installation ... 6
FASTQ files ... 8
Reference files ... 9

3. **Setting up sequencing analysis on Seven Bridges Genomics platform** ... 11

Introduction ... 11
Workflow ... 11
Creating a new project ... 11
Importing FASTQ files ... 12
Importing reference files ... 13
Importing the BD Rhapsody™ Sequence Analysis Pipeline ... 13
Setting up and running the pipeline ... 13
Downloading the output ... 16

4. **Setting up sequencing analysis on a local installation** ... 17

Workflow ... 17
Setting up the input specification file ... 17
Running the pipeline ... 21

5. **Running a pipeline using CWL-runner** ... 22

Running CWL-runner on a local installation ... 22

6. **Reviewing output files** ... 24

Downloading output files on the Seven Bridges Genomics platform ... 24
Sequencing analysis output files ... 25
Reviewing output files ... 26

7. **Troubleshooting** ... 27

Analysis pipeline ... 27

8. **Glossary** ... 29

1. Introduction

This guide provides detailed instructions on how to set up and run the BD Rhapsody™ Sequence Analysis Pipeline 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* (23-21713).

Single-Cell Multiomics technical publications are available for download from the BD® Single-Cell Multiomics 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 sevenbridges.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 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 (fastqs) and 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 3

1. Check to see if a version of Python 3 is already installed by running at the command line:

   ```
   $ python3 --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 a version of Python 3 is not installed, download and install it from python.org/downloads.
4. 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/3.6/lib/python3
      ```

   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.

Pipeline image

1. Ensure that docker is running.
2. Download (pull) the docker image by entering:

   ```
   $ docker pull bdgenomics/rhapsody
   ```

   Note: The pull command automatically downloads the most current pipeline version. To download an earlier version, specify the version number. For example:

   ```
   $ docker pull bdgenomics/rhapsody:v1.0
   ```

3. Confirm the pipeline image by entering:

   ```
   $ docker images
   ```

   Note:

   - bdgenomics/rhapsody appears under the repository column.
   - The pipeline version number appears under the tag column.

FASTQ files

Read 1 and Read 2 sequencing files

For the Seven Bridges Genomics platform and local installation, obtain Read 1 and Read 2 sequencing files. Although the FASTQ file names can have any format, we recommend the following:

- Include R1 or R2.
- The <sample> name should be the same for R1 and R2.
- Convert uncompressed files to .gz format.

  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, underscores, 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 files archive is a compressed tarball that contains the STAR (Spliced Transcripts Alignment to a Reference © Alexander Dobin, 2009-2022) index files and the GTF transcriptome annotation corresponding to the species of cells used in the BD® WTA experiment.

The AbSeq Reference is a FASTA file for BD® AbSeq Ab-Oligos used in a BD Rhapsody™ experiment.

If additional transgene sequences are used in the experiment, an additional FASTA file containing the sequences can be used as Supplemental Reference.

Obtaining pre-designed mRNA panels or WTA reference files

Obtain the FASTA references 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 archive file from the Seven Bridges demo project, downloading from the following link: 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, lncRNA, 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. To prepare the AbSeq reference file, you can use the BD AbSeq Panel Generator (abseq-ref-gen.genomics.bd.com) or follow the instructions below.

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.

Building a custom WTA reference archive

The WTA reference archive is a `tar.gz` file with the following internal structure:

```
--Structure of reference archive--

BD_Rhapsody_Reference_Files/ # top level folder

     star_index/ # sub-folder containing STAR index

          [files created with STAR --runMode genomeGenerate]

     GTF for gene-transcript-annotation
     e.g. "gencode.v43.primary_assembly.annotation.gtf"
```

The same docker image used for running the BD Rhapsody™ Sequence Analysis Pipeline can be used for generating a new WTA reference archive with the following steps:

1. Go to bitbucket.org/CRSwDev/cwl and download the file: `make_rhap_reference_<version>.cwl`
2. Gather a matching set of genome sequence in FASTA format and GTF with gene, transcript, and exon annotations, for example, from gencodegenes.org.
3. Run `cwl-runner` like the following example:

```
cwl-runner make_rhap_reference_2.0.cwl --Genome_fasta
GRCh38.primary_assembly.genome.fa --Gtf gencode.v43.primary_
assembly.annotation.gtf --Archive_prefix testrefhuman43
```

The resulting `testrefhuman43.tar.gz` file can be used for the Reference_Archive input of the BD Rhapsody™ Sequence Analysis Pipeline.

3. Setting up sequencing analysis on Seven Bridges Genomics platform

Introduction

Whether analysis is performed on the Seven Bridges Genomics platform or locally, sequencing analysis uses the BD Rhapsody™ Sequence 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™ Sequence 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™ Sequence 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**:

Create a project ×

Name

project 1

Project URL:

https://igor.sbgenomics.com/u/ /project-1

Billing Group

BD internal

Location

AWS (us-east-1)

Execution settings:

Spot Instances On

Memoization BETA Off

Cancel Create

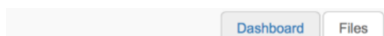
Note: To enable automatic reuse of intermediate files in a rerun, turn **Memoization** on.

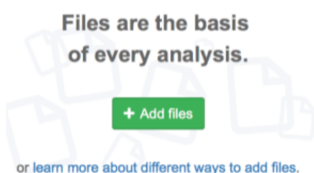
2. On the Create a project dialog, enter the project name, and edit the project URL if necessary.
3. Click **Create**. Seven Bridges Genomics displays the new project dashboard.
4. To change the retention period of intermediate files, click **Settings** in the top right corner. Enter an integer value between 1 and 120 in the Retention period box to specify the number of hours for retention and click **Save**.

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, Your Computer,** or **FTP/HTTP**. Seven Bridges Genomics displays instructions on uploading the files. Follow the Seven Bridges Genomics instructions to import your files.
3. After import, the files are listed 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 **Demo Project** 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 archive 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, Your Computer,** or **FTP/HTTP**. Seven Bridges Genomics displays instructions on uploading the files. Follow the Seven Bridges Genomics instructions to import your files.
3. After import, the files are listed on the Files tab.

Importing the BD Rhapsody™ Sequence Analysis 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, called BD Rhapsody™ Sequence Analysis Pipeline. Or, copy the workflow from the Demo project.
3. Click **Copy** on the app window, select the project in the dropdown 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™ Sequence 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.

Sequence Analysis Pipeline interface:

Inputs

Batching Off

AbSeq Reference — Select file(s)

No files selected

Reads * — Change selection

HumanImmResDemo_S1_L001_R1_001.fastq.gz

HumanImmResDemo_S1_L001_R2_001.fastq.gz

Reference Files Archive — Select file(s)

No files selected

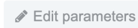
Supplemental Reference — Select file(s)

No files selected

Targeted Reference — Change selection

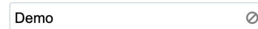
BD_Rhapsody_Immune_Response_Panel_Hs.fasta

App Settings


Edit parameters | Show editable

Name_Settings (#Name_Settings)

Run Name


Demo

Multiplexing_Settings (#Multiplexing_Settings)

Sample Tag Names

Sample Tags Version

Single-Cell Multiplex Kit - Human

VDJ_Settings (#VDJ_Settings)

VDJ Species Version

No value

Putative_Cell_Calling_Settings (#Putative_Cell_Calling_Settings)

Enable Refined Putative Cell Calling

No value

Exact Cell Count

No value

Expected Cell Count

No value

Putative Cell Calling

No value

Intronic_Reads_Settings (#Intronic_Reads_Settings)

Exclude Intronic Reads

No value

Bam_Settings (#Bam_Settings)

Generate Bam Output

No value

Complete all required fields, which appear in red.

Pipeline parameters

Input field	Input	Required?
AbSeq Reference	FASTA AbSeq reference file generated from Importing supplemental or custom mRNA panels or AbSeq reference files on page 13. **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
Enable Refined Putative Cell Calling	Enable use of refined putative cell calling algorithm for cell calling if set to True. By default, putative cells are determined using only the basic algorithm (minimum second derivative along the cumulative reads curve).	Optional
Exact Cell Count	Set a specific number (>=1) of cells as putative, based on those with the highest error-corrected read count.	Optional
Expected Cell Count	Guide the basic putative cell calling algorithm by providing an estimate of the number of cells expected. Usually this can be the number of cells loaded into the Rhapsody cartridge.	Optional
Generate Bam Output	A Bam read alignment file contains reads from all the input libraries, but creating it can consume a lot of computing and disk resources. By setting this field to True, the Bam file will be created.	Optional
Exclude Intronic Reads	By default, reads aligned to exons and introns are considered and represented in molecule counts. Including intronic reads may increase sensitivity, resulting in an increase in molecule counts and the number of genes per cell for both cellular and nuclei samples. Intronic reads may indicate unspliced mRNAs and are also useful, for example, in the study of nuclei and RNA velocity. When set to True, intronic reads will be excluded.	Optional
Putative cell calling	Specify the data to be used for putative cell calling: mRNA or AbSeq. mRNA is the default selected option.	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 or WTA RNA library, and, if applicable, the Sample Tag, TCR, BCR, and BD® AbSeq libraries.	Yes
Targeted 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
Reference Genome Archive (WTA only)	This is a STAR indexed reference genome file, along with a GTF gene annotation file, compressed in the .tar.gz format.	Yes
Run Name	Specify a run name to be used as the base output filename. Use only letters, numbers, hyphens, or underscores. If any other special characters are included, they will be corrected to hyphens.	Optional

Pipeline parameters (continued)

Input field	Input	Required?
Sample Tags Version	For a multiplexed samples run only. Specifies the Sample Tags used: Single-Cell Multiplex Kit—Human Single-Cell Multiplex Kit—Mouse Single-Cell Multiplex Kit—Flex	Required for multiplexed samples
Supplemental Reference	This a FASTA file that contains additional transgene sequences.	Optional
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
VDJ Species Version	For experiments with VDJ libraries. Specify the species and chain types: Human VDJ - BCR and TCR Human VDJ - BCR only Human VDJ - TCR only Mouse VDJ - BCR and TCR Mouse VDJ - BCR only Mouse VDJ - TCR only	Required for VDJ experiment

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, Mouse, or Flex) from the dropdown 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 on page 24.

4. Setting up sequencing analysis on a local installation

Workflow

During sequencing analysis, the BD Rhapsody™ Sequence 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 `pipeline_inputs_template.yml` is downloaded from the CWL folder.

1. Obtain the FASTQ files. See Read 1 and Read 2 sequencing files on page 8.
2. Obtain the targeted mRNA reference file or reference genome archive file 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 on page 9.
4. Specify the desired file paths in the YML file for Reads and Reference with the exact input field listed in the table.
   - The required input fields for Targeted assays are Reads and Targeted Reference.
   - The required input fields for WTA assays are Reads and Reference Genome Archive.
   - The required input fields for AbSeq-only assays are Reads and AbSeq Reference.

Input specification file

Input field	Input	Required?
AbSeq_Reference	FASTA AbSeq reference file generated from Importing supplemental or custom mRNA panels or AbSeq reference files on page 13. **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
Enable_Refined_Cell_Call	Enable use of refined putative cell calling algorithm for cell calling if set to True. By default, putative cells are determined using only the basic algorithm (minimum second derivative along the cumulative reads curve).	Optional

Input specification file (continued)

Input field	Input	Required?
Exact_Cell_Count	Set a specific number (>=1) of cells as putative, based on those with the highest error-corrected read count.	Optional
Expected_Cell_Count	Guide the basic putative cell calling algorithm by providing an estimate of the number of cells expected. Usually this can be the number of cells loaded into the Rhapsody cartridge.	Optional
Generate_Bam	A Bam read alignment file contains reads from all the input libraries, but creating it can consume a lot of compute and disk resources. By setting this field to True, the Bam file will be created.	Optional
Exclude_Intronic_Reads	By default, reads aligned to exons and introns are considered and represented in molecule counts. Including intronic reads may increase sensitivity, resulting in an increase in molecule counts and the number of genes per cell for both cellular and nuclei samples. Intronic reads may indicate unspliced mRNAs and are also useful, for example, in the study of nuclei and RNA velocity. When set to True, intronic reads will be excluded.	Optional
Putative_Cell_Call	Specify the data to be used for putative cell calling: mRNA or AbSeq. mRNA is the default selected option.	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, TCR, BCR, and BD® AbSeq libraries.	Yes
Targeted_Reference (Targeted only)	Select the FASTA reference file. This is a pre-designed, supplemental, or custom panel.	Yes
Reference_Archive (WTA only)	This is a STAR indexed reference genome file, along with a GTF gene annotation file, compressed in the `.tar.gz` format.	Yes
Run_Name	Specify a run name to be used as the base output filename. Use only letters, numbers, hyphens, or underscores. If any other special characters are included, they will be corrected to hyphens.	Optional
Sample_Tags_Version	For a multiplexed samples run only. Specifies the Sample Tags used: human (hs), mouse (mm), or flex.	Required for multiplexed samples
Supplemental_Reference	This is a FASTA file that contains additional transgene sequences.	Optional
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

Input specification file (continued)

Input field	Input	Required?
VDJ_Version	For experiments with VDJ libraries. Specify the species and chain types: human mouse humanBCR humanTCR mouseBCR mouseTCR	Required for VDJ experiment

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 reference files 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

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

Reference_Archive:
  class: File
  location: test/RhapRef_Human_WTA_2023-02.tar.gz
```

YML file example showing WTA with Sample multiplexing and TCR/BCR (VDJ) analysis

```
#!/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_Archive:
   class: File
   location: test/RhapRef_Human_WTA_2023-02.tar.gz

Sample_Tags_Version: flex
VDJ_Version: human
```

6. Save the modified template YML file.

Running the pipeline

Procedure

See Running a pipeline using CWL-runner on page 22.

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 on page 6.

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. We recommend ≥32 GB for Targeted and ≥64 GB for WTA. See Local installation on page 6. 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 on page 17.
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_pipeline_<version>.cwl`, and the input specification file is the `pipeline_inputs_template_<version>.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. Jobs in some steps can run in parallel. To enable this, use the flag `--parallel`

   An example command:

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

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

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

6. 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 on page 24.

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 the output file to view it, and click **Download** to download and save the output file. To download more than one output file at a time, click the Folder icon to the right of Outputs. Click the check boxes by files to download, or click the gray check box at the top to select all files, and then click **Download**.
5. View the output files. See Sequencing analysis output files on page 25.

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 files

Output	File	Content
Metrics summary	<sample_name>_Metrics_Summary.csv	Report containing sequencing, molecules, and cell metrics
Pipeline report HTML	<sample_name>_Pipeline_Report.html	Summary report containing the results from the sequencing analysis pipeline run
BAM and BAM Index	<sample_name>.BAM <sample_name>.BAM.bai	Alignment file of R2 and associated R1 annotations
Data tables[a]	<sample_name>_RSEC_MolsPerCell_MEX.zip <sample_name>_DBEC_MolsPerCell_MEX.zip	Molecules per bioproduct per cell, based on RSEC or DBEC
	<sample_name>_RSEC_MolsPerCell_Unfiltered_MEX.zip	Unfiltered tables containing all cell labels with ≥10 reads
Bioproduct Statistics	<sample_name>_Bioproduct_Stats.csv	Metrics from RSEC and DBEC Unique Molecular Identifier adjustment algorithms on a per-bioproduct basis
Single-cell analysis tool inputs	<sample_name>_Seurat.rds	Seurat (.rds) input file containing RSEC molecules data table and all cell annotation metadata.
	<sample_name>.h5ad	Scanpy input file containing RSEC molecules data table and all cell annotation metadata.

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:

Outputs when multiplex option selected

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, metric summary, 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.

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

Outputs when VDJ option selected

Output	File	Content
VDJ metrics	<sample_name>_VDJ_Metrics.csv	Overall metrics from the VDJ analysis
VDJ per Cell metrics	<sample_name>_VDJ_perCell.csv <sample_name>_VDJ_perCell_uncorrected.csv	Cell specific read and molecule counts, VDJ gene segments, CDR3 sequences, paired chains, and cell type
VDJ Dominant Contigs	<sample_name>_VDJ_Dominant_Contigs_AIRR.csv	Dominant contig for each cell label chain type combination (putative cells only)
VDJ Unfiltered Contigs	<sample_name>_VDJ_Unfiltered_Contigs_AIRR.csv	All contigs that were assembled and annotated successfully (all cells)

Reviewing output files

See the *BD® Single-Cell Multiomics Bioinformatics Handbook* (23-21713).

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

7. Troubleshooting

Analysis pipeline

Introduction

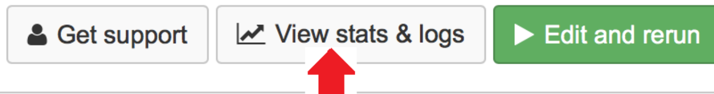
This topic describes how to respond to a task failure while running the BD Rhapsody™ Sequence Analysis Pipeline .

Arranging BD Biosciences to join the project on Seven Bridges Genomics

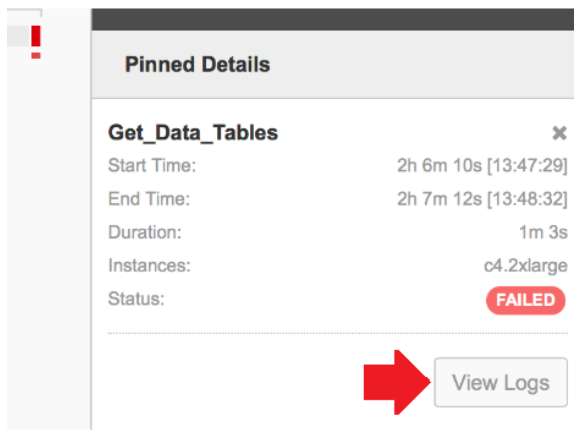
If a task fails on the Seven Bridges Genomics platform, contact BD Biosciences technical support at scomix@bdscomix.bd.com to troubleshoot the issue. Tech support will provide you with instructions on inviting a support team member to your project. To troubleshoot the issue yourself, access the log files.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

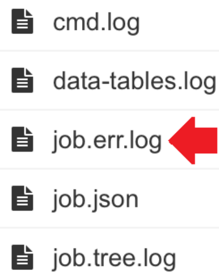


2. Locate the failed node in your pipeline run. Completed nodes are in green, and the failed node is in red. Click the failed node, and on the right, click **View Logs** for that node:



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.

8. Glossary

A

AIRR Adaptive Immune Receptor Repertoire.

B

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

Bioproduct Identifiers for biologically-derived products such as mRNA and protein. Examples of identifiers are gene name for mRNA or AbSeq identifier for AbSeq.

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™ Sequence Analysis Pipeline containing read count or molecule count per bioproduct.

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 bioproduct.
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.

Y

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

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

bdbiosciences.com
scomix@bdscomix.bd.com