SMART-Seq® DE3 Demultiplexer User Guide

I. Introduction

The SMART-Seq DE3 Demultiplexer is bioinformatics software for use only with sequencing data generated from SMART-Seq mRNA 3′ DE. This software uses in-line indexes incorporated in paired reads to demultiplex your samples and generate an output file with sorted read data. These in-line indexes are hard-coded in the software.

More specifically, SMART-Seq mRNA 3′ DE generates a library in which unique in-line indexes are added to the cDNA to serve as cell barcodes and enable sample pooling prior to library preparation and sequencing. The library produces paired reads where the in-line index is encoded in the first six nucleotides of read 2 and the sequence of interest is contained in paired read 1. For more information on SMART-Seq mRNA 3′ DE, please visit the product page.

NOTE: If you know you have Java 1.7 or later installed and are only looking for the command arguments, please see Section IV.B.

II. Before You Begin

A. Input Files

Input files are paired reads, where read 1 contains a sequence of interest and read 2 contains a corresponding barcode sequence. SMART-Seq DE3 Demultiplexer works with either gzipped files (`.gz` extension) or uncompressed files (commonly with `.fq` or `.fastq` extensions).

B. Verifying Java and Confirming Operating System

SMART-Seq DE3 Demultiplexer is written in Java and is compatible with any operating system where Java 1.7 or newer is installed. We have tested SMART-Seq DE3 Demultiplexer on Linux (Ubuntu 16.40), Windows (7 and 10) and Mac (OS X El Capitan).

After opening a terminal window, verify that you have the correct version of Java installed by typing the following command into your terminal window (see section IV.A for instructions on opening a terminal window):

```
java -version
```

```
$ java -version
java version "1.8.0_91"
Java(TM) SE Runtime Environment (build 1.8.0_91-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.91-b14, mixed mode)
$
```

Figure 1. After typing `java -version` into your terminal, you should see a version number displayed. Verify that the first two numbers are 1.7 or higher.

C. Example Data

We have provided sample data in the zip archive that you can use to confirm operation of the demultiplexing software. The paired-end sample data contains a pooled library of 2,500 reads and all 12 in-line indexes. This representative dataset was acquired with a library synthesized with SMART-Seq mRNA 3′ DE and sequenced on an Illumina® MiSeq® instrument. The data are contained in two files named:

- `Example_Data_R1_.fastq.gz` (from read 1)
- `Example_Data_R2_.fastq.gz` (from read 2)

All examples below refer to this sample data.

III. Installing SMART-Seq DE3 Demultiplexer

The software is available to download as a ZIP file from https://www.takarabio.com/products/next-generation-sequencing/bioinformatics-tools/smart-seq-de3-demultiplexer.

Unzip the archive in a convenient directory. The software is packaged as an executable Java JAR file. No additional steps are necessary to install the software.

IV. Running SMART-Seq DE3 Demultiplexer

SMART-Seq DE3 Demultiplexer is written as a command-line-interface tool and is executed from within a terminal window.

A. Opening a Terminal Window

- **Mac:** The Terminal application is typically found under **Applications > Utilities > Terminal**
- **Linux (Ubuntu):** Use the keyboard shortcut **[Ctrl][Alt][T]**.

  Alternatively, you can find the terminal by opening the [Dash] (upper left on most desktops), typing `terminal` and selecting the Terminal application.
- **Windows:** Click the [Start] button and type `cmd` in the run box.

  Alternatively, select “Command Prompt” from **Start > Program Files > Accessories**.

B. Running the Demultiplexing Command

The command to demultiplex your sequencing run is (all in one command):

```
java -jar \
/path/to/SMART_Seq_DE3_Demultiplexer.jar \
/path/to/read1 \
/path/to/read2
```

NOTE: on Mac and Linux systems `~/` is not expanded automatically by Java.

- Use the full path, i.e., `/home/tbusa/SMART_Seq_DE3_Demuliplexer/Example_Data_R1_.fastq.gz`
- Do not use `~/SMART_Seq_DE3_Demuliplexer/Example_Data_R1_.fastq.gz`

Table 1 provides a brief explanation of the command:

Table 1. SMART_Seq_DE3_Demultiplexer command breakdown.

Column 1	Column 2
`java -jar`	System command to execute a Java jar file
`/path/to/SMART_Seq_DE3_Demultiplexer.jar`	Full system path to the SMART-Seq DE3 Demultiplexer jar file
`/path/to/read1`	Full system path to read 1 FASTQ file
`/path/to/read2`	Full system path to read 2 FASTQ file

As an alternative to typing the demultiplex command, you can drag and drop files into the command window. If you'd like to avoid typing out the full file paths, you can drag a file onto the terminal window to automatically add each file path to the end of the current command line. In order to use this method to fill out the full command, use the following steps:

1. Open the terminal window and the folder containing the SMART_Seq_DE3_Demultiplexer.jar file and your read data.
2. Type: `java -jar`
3. Press **[Space bar].**
4. Drag and drop the `SMART_Seq_DE3_Demultiplexer.jar` file into the terminal window (you should now see a full path to the executable file).
5. Press **[Space bar].**
6. Drag and drop the read 1 file into the terminal window (you should see the full path for the FASTQ file).
7. Press **[Space bar].**
8. Drag and drop the read 2 file into the terminal window (you should see the full path for the FASTQ file).

C. Example Command

Here, we provide an example command run on Linux with the zip archive extracted into our home folder (`/home/tbusa/`). The following command will demultiplex the example data included in the archive.

NOTE: The following command is a single line.

```
java -jar \
/home/tbusa/SMART_Seq_DE3_Demultiplexer/SMART_Seq_DE3_Demultiplexer.jar \
/home/tbusa/SMART_Seq_DE3_Demultiplexer/Example_Data_R1_.fastq.gz \
/home/tbusa/SMART_Seq_DE3_Demultiplexer/Example_Data_R2_.fastq.gz
```

V. Examining the Demultiplexing Command Output

A. Output Files

The demultiplexing command always produces 26 files, even if no sequences are placed in a file (e.g. if you are analyzing data from 6 pooled samples, there will still be 26 files produced by the software). The original data files are maintained in their original directory. Three kinds of files are produced by the demultiplexing command:

- **Read 1:** There are 12 "read 1" output files that are named with the in-line index added to the beginning of the original file name and the extension `.fastq` added to the end (e.g., `IL01_Example_Data_R1_.fastq.gz.fastq`). These files now contain sequencing reads that are usable for downstream analyses.
- **Read 2:** There are 12 "read 2" output files that are named similarly to the read 1 output files (e.g., `IL01_Example_Data_R2_.fastq.gz.fastq`). These files are provided to the user to verify that the demultiplexing command performed as expected. For instance, one might verify that the sequences in the `IL12_*` file all start with `CTTGTA`.; this sequence corresponds to in-line index 12.
- **Undetermined files:** There are two `UNDETERMINED` files, one associated with read 1 and one associated with read 2. When the software cannot match the first six nucleotides in read 2 to one of the twelve expected barcodes, it places those reads in the `UNDETERMINED_*` file-pair.

Figure 2. This image displays all of the output files created after successfully running the demultiplexing software on the provided sample data. Twenty-six new files are generated and added to the directory with the original data file.

B. Command Summary

A summary of the demultiplexing results is written to the screen as a table with three columns:

- In-line index
- Output file path
- Number of sequencing reads that correspond to each index

The table is sorted by the name of the in-line index, with the `UNDETERMINED` sequences listed at the end. In the example shown in Figure 3 (next page), notice that each file is prefixed with the in-line index sequence and appended with the FASTQ extension, as discussed in Section V.A above.

```
$ java -jar /home/clontech/SMART_Seq_DE3_Demuliplexer/SMART_Seq_DE3_Demultiplexer.jar /home/clontech/SMART_Seq_DE3_
Demuliplexer/Example_Data_R1_.fastq.gz /home/clontech/SMART_Seq_DE3_Demuliplexer/Example_Data_R2_.fastq.gz
Summary:
Barcode File    Number_of_Records
CTAGCT  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL01_Example_Data_R1_.fastq.gz.fastq  221
CGATGT  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL02_Example_Data_R1_.fastq.gz.fastq  202
TTAGGC  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL03_Example_Data_R1_.fastq.gz.fastq  214
TGACCA  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL04_Example_Data_R1_.fastq.gz.fastq  204
ACAGTG  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL05_Example_Data_R1_.fastq.gz.fastq  222
TATAAT  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL06_Example_Data_R1_.fastq.gz.fastq  180
CAGATC  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL07_Example_Data_R1_.fastq.gz.fastq  221
ACTTGA  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL08_Example_Data_R1_.fastq.gz.fastq  190
GATCAG  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL09_Example_Data_R1_.fastq.gz.fastq  195
TAGCTT  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL10_Example_Data_R1_.fastq.gz.fastq  223
GGCTAC  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL11_Example_Data_R1_.fastq.gz.fastq  214
CTTGTA  /home/clontech/SMART_Seq_DE3_Demuliplexer/IL12_Example_Data_R1_.fastq.gz.fastq  150
UNDETERMINED    /home/clontech/SMART_Seq_DE3_Demuliplexer/UNDETERMINED_Example_Data_R1_.fastq.gz.fastq  64
$
```

Figure 3. After running the command, the software will write a summary to the terminal window. The summary will display each in-line index (or UNDETERMINED for sequences that can't be matched), the output file for read 1 associated with each inline index, and the number of sequences that were sorted into each file.

VI. Technical Support

For any technical support regarding this protocol, please visit: takarabio.com/support.

Products

Cat. #	Product	Size
635040	SMART-Seq mRNA 3′ DE	96 Rxns
635041		192 Rxns

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.
2560 Orchard Parkway, San Jose, CA 95131, USA
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

05.24 US