I. Introduction

SMART-Seq® mRNA Long Read (Cat. Nos. 634376 & 634377) allows for multiplexing up to 96 samples using unique barcodes. To demultiplex the resulting sequencing data, you will need to provide Dorado (demultiplexing tool) from Oxford Nanopore Technologies (ONT) with the barcode configuration, barcode and adapter sequences unique to the kit. This information is available for download from takarabio.com (for more information, see Section II.D.2).

Before you can perform demultiplexing, raw sequencing files generated with ONT sequencing devices need to be converted from POD5 format to FASTQ or FASTQ.gz format through basecalling. Once converted, the data should be moved to a Linux server for demultiplexing.

After demultiplexing, we recommend performing strand correction using Restrander. After strand correction, reads will be aligned against a reference genome using Minimap2 or tools with integrated Minimap2 functionality such as Dorado or FLAIR. Downstream analysis after genome alignment will depend on your unique research question and is not supported here.

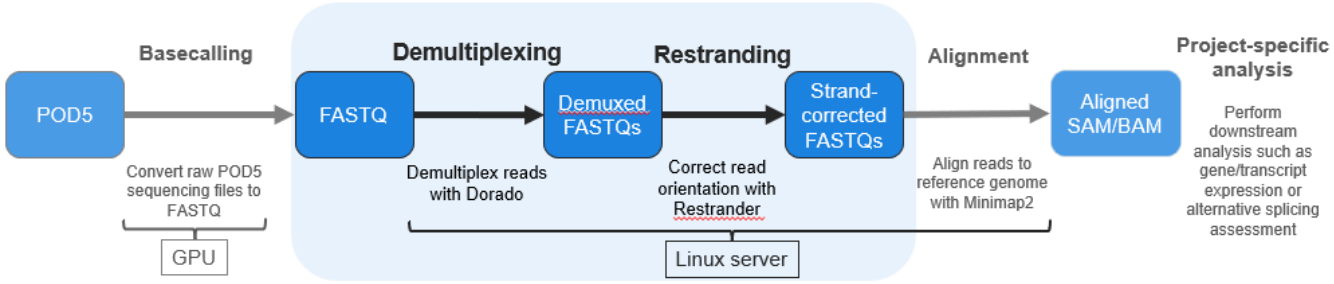

Figure 1. Overview of the bioinformatic workflow after sequencing on ONT sequencers. Starting with FASTQ files converted from POD5 format, the demultiplexing and restranding steps (the second and third arrow in the workflow figure) described by this protocol produce strand-corrected FASTQs ready for downstream analysis.

II. Before You Begin

Please read through all the requirements for use outlined in this subsection to prepare for performing the protocol.

A. Supported Operating Systems & Hardware Guidelines

- These instructions have been tested on Linux RedHat 9.5
- For more information about hardware requirements, refer to the official Dorado documentation

B. User Account Requirements

Read, write, and executable permissions are needed for the account used on the Linux server, both for installing the software, copying the files provided by Takara Bio, and running the third-party software.

C. Third-party Software Dependencies

These programs are not included with the download files supplied by Takara Bio and must be downloaded and installed separately on the Linux server.

- Dorado (verified with version 0.8.3+98456f7, 0.9.0+9dc15a8 and 0.9.1+c8c2c9f) (Github link)
- Optional, but recommended: Restrander (Github link)

D. Required Input Files

The pipeline requires FASTQ files and a custom `config/` folder available for download from Takara Bio to run.

1. FASTQ files

This protocol has been validated to use FASTQ files as generated by the ONT sequencer platform. The files can be stored in any directory on the server or workstation as long as the folder is not private or has read-write user restrictions that would prevent the files from being accessed by Dorado. The files are generally converted from POD5 format to either compressed (.fastq.gz) or decompressed (.fastq) format.

2. `config/` folder from Takara Bio

The protocol requires a set of customized files contained in a `config/` directory to work with the SMART-Seq mRNA Long Read sequencing data. Please sign up to gain access to download the folder.

III. Installation and Configuration Requirements

The instructions below are to confirm Dorado is installed and working properly on the Linux server and to install the required configuration files supplied by Takara Bio for use with Dorado in this protocol.

The latest version of dorado can be downloaded from here: https://dorado-docs.readthedocs.io/en/latest/

A. Confirm Dorado Version

Verify Dorado is installed and meets or exceeds the required version. This can be done in one of two ways:

- From a terminal window and in the directory in which Dorado is installed, type the following:

  ```
  dorado -version
  ```

- If running the command from a different directory from where Dorado is installed, the full path to Dorado (**`$PATH`**) will need to be specified.

  ```
  $PATH/dorado -version
  ```

  Example:

  If Dorado is installed in `/home/bin` and you're running the command from the home folder of a user account (`/home/user`), the command with the path would be:

  ```
  /home/dorado-0.9.1-linux-x64/bin/dorado -version
  ```

B. Download and Confirm the Config Files

1. Sign up to download the zip file containing a config/ folder from our website for the protocol(s) you wish to run (see Section IV). You will also be sent a confirmation email that includes a link directly to the page to download the supporting files, for future reference
2. Download the .zip file and make sure it has completed downloading before proceeding to the next step.
3. (Optional) If the .zip file was not downloaded to the Linux server, copy or transfer it to a location on the server where the demultiplex command will be run.

4. Unzip the file. You should see a directory `config/` where it is unzipped that contains the files shown in Figure 2. Table 1 (next page) outlines a brief description of each of the files.

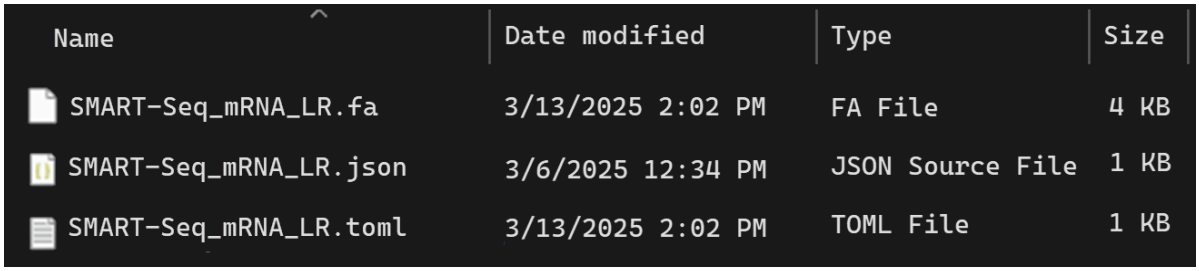

Figure 2. File contents of the `config/` directory provided by Takara Bio.

Table 1. Description of the files in the config/ package from Takara Bio.

Filename	Usage
`SMART-Seq_mRNA_LR.toml`	Barcode arrangement information required by Dorado (Section IV.A)
`SMART-Seq_mRNA_LR.fa`	Sequence information of barcodes and adapters required by Dorado
`SMART-Seq_mRNA_LR.json`	Configuration file required by Restrander (Section IV.B)

C. Uninstall Procedure

Delete the .zip file and `config/` directory from the server. If desired, Dorado can be uninstalled per the directions for that software.

IV. Protocol

A. Protocol: Convert the POD5 Files to FASTQ format

1. Basecall raw sequencing data to convert the POD5 files output from the sequencer to FASTQ or FASTQ.gz files.

   Refer to the Oxford Nanopore documentation at https://nanoporetech.com/document/data-analysis for the procedure and follow the instructions specific to the sequencer you are using.

2. Once converted, make sure the FASTQ/FASTQ.gz files are stored in a single directory either on the Linux server that will be used for demultiplexing or a network drive reachable by the Linux server.

B. Protocol: Demultiplex the Data with Dorado

The command to demultiplex the data has the following syntax:

```
$PATH1/dorado demux --kit-name="SMART-Seq" --emit-fastq --emit-summary \
  --threads <THREADS> \
  --output-dir $PATH2/<OUTPUT> \
  --barcode-arrangement $PATH3/config/SMART-Seq_mRNA_LR.toml \
  --barcode-sequences $PATH3/config/SMART-Seq_mRNA_LR.fa \
  $PATH4/<FASTQ_DIR>
```

where:

- **<THREADS>** is the number of threads to use for Dorado. The number of threads should not exceed the number of CPUs on the server (e.g., ≤ 4 on a quad-core server). The default value is '1'.

- **<OUTPUT>** should be the name of the output folder (including the full path, if necessary) to be created by the analysis to store the output files.
- **<FASTQ_DIR>** is the name of the directory containing (Section 1).
- **$PATH1–$PATH4** represents the full path (directory tree) to where the file following it is located on the server.
  - **$PATH1** : where Dorado is installed
  - **$PATH2** : where the output folder generated by Dorado will be saved
  - **$PATH3** : where the unzipped `config/` directory is located
  - **$PATH4** : where the directory of FASTQ files are located

1. Files in the Local Working Directory (Example #1)

The example below is for an account `user`, with a home directory of `/home/user`.

- Running as four (4) threads
- The Dorado program, `config/` directory, and FASTQ file directory `data/` are also in the home directory. Dorado will be called by adding **./** to the beginning of the executable filename
- The output will be written to a directory, `demux-output/`, in the home directory.

As everything is stored in, the command will be run in, and the output will be written to the home directory, the command might look like:

```
./dorado demux --kit-name="SMART-Seq" --emit-fastq --emit-summary \
  --threads 4 \
  --output-dir demux-output \
  --barcode-arrangement config/SMART-Seq_mRNA_LR.toml \
  --barcode-sequences config/SMART-Seq_mRNA_LR.fa \
  data
```

2. Files are in Remote Directories (Example #2)

The example below is for an account `user`, with a home directory of `/home/user`.

- The Dorado executable is installed in the home directory, at `/home/user/dorado-0.9.1-linux-x64/bin/` ($PATH1)
- Dorado will be run with four (4) threads
- The output of the command will be written to a directory, `dorado_output/`, in `/home/user/`($PATH2)
- The `config/` directory was unzipped in `/home/user/bin/`($PATH3)
- The FASTQ file directory, `data`, is located in `/home/user/fastq_output/`($PATH4)
- Since all paths are specified explicitly, the command can be run from anywhere on the server

The command would look like:

```
/home/user/dorado-0.9.1-linux-x64/bin/dorado demux
  --kit-name="SMART-Seq" --emit-fastq --emit-summary \
  --threads 4 \
  --output-dir /home/user/dorado_output/ \
  --barcode-arrangement /home/user/bin/config/SMART-Seq_mRNA_LR.toml \
  --barcode-sequences /home/user/bin/config/SMART-Seq_mRNA_LR.fa \
  /home/user/fastq_output/data
```

C. Protocol: Process the Dorado Output with Restrander (Optional)

Restrander (Schuster, Ritchie, and Gouil 2023) should be used after basecalling/demultiplexing and before mapping. Restranding reads is optional but highly recommended and will produce a fq.gz file containing strand-corrected reads.

Restrander can be downloaded from its Github repository: https://github.com/mritchielab/restrander

Restrander command syntax:

```
$PATH5/restrander \
  $PATH6/<input>.fastq\
  $PATH7/<output>.fastq\
  $PATH8/config/SMART-Seq_mRNA_LR.json
```

where:

- **<input>.fastq** should be the name of an output file from the demux analysis (Dorado)
- **<output>.fastq** is the name you want to give the file output from Restrander.
- `SMART-Seq_mRNA_LR.json` is the name of the configuration file for Restrander and is provided in the `config/` folder from Takara Bio.
- **$PATH5–$PATH8** represents the full path (directory tree) to where the file following it is located on the server.
  - **$PATH5** : where Restrander is installed. This will point to the Restrander executable within the Restrander folder. If the Restrander tool is in the same location as where the command is being run from, replace `$PATH5` with a dot (`./restrander`)
  - **$PATH6** : the directory specified as **$PATH2/<OUTPUT>** in the demux command.
  - **$PATH7** : where the output file from Restrander should be placed.
  - **$PATH8** : where the unzipped `config/` directory is located.

Using similar file locations as Example #2 in Section B, the command might look like:

```
/home/user/restrander/restrander \
  /home/user/dorado_output/sample1-demux.fastq \
  /home/user/restrander_output/sample1-restrand.fastq \
  /home/user/bin/config/SMART-Seq_mRNA_LR.json
```

For any questions about Restrander, please refer to the instructions on the Github page.

NOTE: Takara Bio does not provide support for using Restrander.

D. Follow-Up Steps/Tools

These steps outlined below are only provided as examples of what you might do after demultiplexing (or demultiplexing and running Restrander on) your data. You will need to decide what additional bioinformatics processing should be done to your demultiplexed data, based on your experimental needs.

1. **Genome alignment**: We recommend using Minimap2 with the following settings but encourage users to make adjustments based on their specific sample.

   ```
   minimap2 -ax splice ref.fa nanopore-cdna.fastq > aln.sam
   ```

   Minimap2 download and documentation can be found at github.com/lh3/minimap2.

   Alternatively, genome alignment can be performed with tools which have Minimap2 integrated, such as Dorado or FLAIR (github.com/BrooksLabUCSC/flair). Please refer to the specific instructions of the tool used.

2. **Gene expression** analysis may be performed with any long-read compatible tool (e.g., FeatureCounts, subread.sourceforge.net/featureCounts.html)

3. **Transcript analysis** and **alternative splicing** may be assessed with tools such as FLAIR (including FLAIR splice)(https://flair.readthedocs.io/en/latest/), SQANTI3 (github.com/ConesaLab/SQANTI3), salmon (github.com/COMBINE-lab/salmon), or any other long-read compatible transcriptome analysis tool.

NOTE: Takara Bio does not support or endorse any specific tool for follow-up data analysis.

V. References

Schuster, J., Ritchie, M. E. & Gouil, Q. Restrander: rapid orientation and artefact removal for long-read cDNA data. *NAR Genomics Bioinforma.* **5,** lqad108 (2023).

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.

This document has been reviewed and approved by the Quality Department.