Jun 05, 2025

# Mercurius DRUG-seq Library Preparation Kit for 96, 384, and 1536 Samples

DOI
https://dx.doi.org/10.17504/protocols.io.kqdg3wzdqv25/v1

Alithea Genomics[1]

[1]Alithea Genomics SA, Route de la Corniche 8, 1066 Epalinges, Switzerland

Sophia Shilimindri
Alithea Genomics SA





**DOI:** https://dx.doi.org/10.17504/protocols.io.kqdg3wzdqv25/v1

**Protocol Citation:** Alithea Genomics 2025. Mercurius DRUG-seq Library Preparation Kit for 96, 384, and 1536 Samples. **protocols.io** https://dx.doi.org/10.17504/protocols.io.kqdg3wzdqv25/v1



**Protocol status:** Working
**We use this protocol and it's working**

**Created:** June 02, 2025

**Last Modified:** June 05, 2025

**Protocol Integer ID:** 219320

**Keywords:** sequencing library preparation, rna from lysate sample, sequencing library, samples the mercuriustm drug, seq library preparation kit, sequencing data, seq library preparation, seq kit, tedious rna purification step, mercuriustm drug, sequencing, resulting cdna sample, rna, cdna samples from each experimental group, mercurius drug, lysate sample, individual barcodes during the first strand synthesis reaction, seq technology, barcoded mercuriustm oligo, mild cell lysis buffer, seq, preparation of illumina, crude cell lysate, kit, first strand synthesis reaction, lysate, cell lysi

## Abstract

The MERCURIUSTM DRUG-seq kits allow the preparation of Illumina-compatible 3′ RNA sequencing libraries for up to 1,536 lysate samples in a time and cost-efficient manner. The kits include a mild cell lysis buffer to prepare crude cell lysates that can be used directly in the RT reaction, skipping the tedious RNA purification step. The kits are provided in various formats, each containing barcoded MERCURIUSTM Oligo-dT primers designed to tag polyA+ RNA from lysate samples with individual barcodes during the first strand synthesis reaction, allowing the pooling of the resulting cDNA samples from each experimental group into a single tube for streamlined sequencing library preparation. The DRUG-seq technology can be used to generate high-quality sequencing data from 2,000 – 50,000 mammalian cells per well.

## Guidelines

The recommended input range of cells is 15,000-50,000 cells/well of 96WP and 2,000-10,000 cells/well of 384WP. Cells must be seeded a few days in advance for the best results. Depending on the type of cells and experimental design, consider the doubling time of cells after the seeding and the potential effect of the treatment on the cell number during the experiment. To ensure an even distribution of reads after sequencing, the amount of starting material must be as uniform as possible. The cell lysate used in the RT reaction should be at most 50% of the final volume.

## Materials

Reagents supplied include Barcoded Oligo-dT Adapters Set V5 Module, DRUG-seq Library Preparation and UDI Module, and others. Additional required reagents and equipment supplied by the user include plasticware such as 15 mL conical tubes, 0.2 mL 8-Strip Non-Flex PCR Tubes, and reagents like DNA Clean and Concentrator-5 kit, SPRI AMPure Beads, Qubit™ dsDNA HS Assay Kit, and equipment like Benchtop centrifuge for plates, Single and Multichannel pipettes, and others.

# Preparation of Cell Lysate Samples

## 1 PREPARATION OF CELL LYSATE SAMPLES

1.1 Essential considerations for input cells:

- The recommended input range of cells is 15,000-50,000 cells/well of 96WP and 2,000-10,000 cells/well of 384WP.
- Cells must be seeded a few days in advance for the best results.
- Depending on the type of cells and experimental design, consider the doubling time of cells after the seeding and the potential effect of the treatment on the cell number during the experiment.
- To ensure an even distribution of reads after sequencing, the amount of starting material must be as uniform as possible.
- The cell lysate used in the RT reaction should be at most 50% of the final volume.

1.2 ERCC Spike-in Controls (Optional):
To ensure the evaluation of the sequencing reads uniformity across the samples and to assess the impact of sample and library preparation steps on it, we recommend the addition of External RNA Controls Consortium (ERCC) Spike-Ins to the lysate buffer.

1.3 **Cell lysate preparation**

To ensure the evaluation of the sequencing reads uniformity across the samples and to assess the impact of sample and library preparation steps on it, we recommend the addition of External RNA Controls Consortium (ERCC) Spike-Ins to the lysate buffer.

Prepare a working solution of 1x Cell Lysis Buffer with RNase Inhibitor.

| Reagent | 96WP, µL | | 384WP, µL | |
|---|---|---|---|---|
| | Per well | 96 wells +10% | Per well | 384 wells +10% |
| CLB | 6.6 | 700 | 3.3 | 1400 |
| INH | 1.6 | 170 | 0.8 | 340 |
| Water | 11.8 | 1250 | 5.9 | 2500 |
| TOTAL | 20 | 2120 | 10 | 4240 |

Pipette a prepared mix gently a few times, and briefly spin the tube. Keep the mix on ice until further use.

**Procedure for the preparation of adherent cells**

1.3.1. Seed the cells in a flat bottom 96WP or 384WP at the density that will enable harvesting:
■ **96WP**: 15’000-50’000 cells per well
■ **384WP**: 2’000-10’000 cells per well

1.3.2. Gently aspirate culture media from the plate and wash cells by adding the following:
■ **96WP:** 80-100 µL DPBS in each well
■ **384WP:** 30-50 µL DPBS in each well

1.3.3. Gently tap the plate and aspirate as much DPBS as possible without disturbing the cell pellet.

1.3.4. Seal the plate and snap-freeze it on dry ice or liquid nitrogen for at least 5 minutes. Alternatively,
the plate can be stored at -80°C for a few weeks.

1.3.5. Proceed to step 1.3.13 for cell lysis.

**Procedure for the preparation of suspension cells**

1.3.6. Seed the cells in a flat bottom or U-shaped **96WP** or **384WP** at the density that will enable
harvesting:
■ **96WP**: 15’000-50’000 cells per well
■ **384WP**: 2’000-10’000 cells per well

1.3.7. Centrifuge the plate at 300x g for 5 minutes.

1.3.8. Gently aspirate culture media from the plate without disturbing the cell pellet and wash cells by
adding:
■ **96WP:** 80-100 µL of DPBS in each well
■ **384WP:** 30-50 µL of DPBS in each well

1.3.9. Centrifuge the plate at 300x g for 5 minutes.

1.3.10. Gently tap the plate and aspirate as much DPBS as possible without disturbing the cell pellet

1.3.11. Seal the plate and snap-freeze it on dry ice or liquid nitrogen for at least 5 minutes. Alternatively,
the plate can be stored at -80°C for a few weeks.

1.3.12. Proceed to step for cell lysis.

**Procedure for cell lysis**

1.3.13. Using a multi dispenser, distribute the prepared CLB in every well.
■ **96WP**: 20 µL per well
■ **384WP**: 10 µL per well

1.3.14. Centrifuge the plate at 300x g for 1 minute to ensure that CLB is uniformly distributed on the
surface of each well.

1.3.15. Incubate the plate at room temperature for 15 min, gently agitating it occasionally.

1.3.16. Transfer the lysate from every well to the corresponding well of the 96- or 384-well PCR plate.
■ Pipette **directly into the plate containing oligos** (follow step 2.1.1); or
■ Pipette **to a new PCR plate** for a future experiment. Ensure proper labeling for easy identification during subsequent use.

1.3.17. Seal the plate with an aluminium seal provided and briefly spin it down.
1.3.18.The cleared lysates can be used directly for library preparation or safely stored at -80°C.

**NOTE:** If several plates must be processed, perform the procedure with each plate individually one by one to avoid
keeping plates at room temperature for a prolonged time.

## Library Prep

### 2 Library Preparation Protocol

### 2.1 Second strand synthesis

At this step, double-stranded full-length cDNA is generated and purified using magnetic beads.
**Preparation**
• Pre-warm the SPRI beads at room temperature for ~30 min.
• Prepare 5 mL of 80% ethanol.

- Thaw the **SSB** reagent at room temperature and mix well before use.
- Keep the **SSE** reagent constantly on ice.
- Prepare Program 3_SSS on the thermocycler (set the lid at 70°C):

| Step | Temperature, °C | Time |
|---|---|---|
| Incubation | 37 | 20 min |
| Incubation | 65 | 30 min |
| Keep | 4 | pause |

**Procedure**

2.4.1. Prepare the SSS reaction mix for the second strand synthesis as follows (with 10% excess):

| Reagent | Per reaction (µL) | |
|---|---|---|
| | For 17 µL elution | For 35 µL elution |
| SSB | 5.0 | 7.0 |
| SSE | 2.0 | 3.0 |
| **TOTAL** | 7.0 | 10.0 |

2.4.2. Transfer **7 µL** or **10 µL** of SSS reaction mix to the tube from step 2.3.7 and mix well by pipetting
up and down 5 times.

2.4.3. Incubate in thermocycler 2.4.4. Proceed immediately to step 2.4.5.

**cDNA clean-up with SPRI beads**

Perform the double-stranded cDNA purification with SPRI magnetic beads using a 0.6x ratio (i.e., 30 µL
of bead slurry plus 50 µL of cDNA).

**NOTE:** Use pre-warmed beads and vortex them vigorously before pipetting.

2.4.5. Complement the final volume to 50 µL with water.

2.4.6. Add 30 µL of beads and mix by pipetting up and down 10 times.

2.4.7. Incubate for 5 min at room temperature.

2.4.8. Place the tube on the magnetic stand, wait 5 min, and carefully remove and discard the
supernatant.

2.4.9. To wash the beads, pipette 200 µL of freshly prepared 80% ethanol into the tube.

2.4.10. Incubate for 30 sec.

2.4.11. Carefully remove the ethanol without touching the bead pellet.
2.4.12. Repeat step 2.4.9 for a total of two washes.
2.4.13. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.
2.4.14. Resuspend the beads in 21 µL of water.
2.4.15. Incubate for 1 min.
2.4.16. Place tubes on the magnetic stand, wait 5 min, and carefully transfer 20 µL of the supernatant
into a new tube to avoid bead carry-over.
2.4.17. Use 2 µL to measure the concentration with Qubit.
**Safe stop:** At this step, the cDNA can be safely kept at -20°C for a few weeks.

## 2.2 Sample pooling and column purification

After pooling, the barcoded RT samples can be purified using either column-based Zymo Clean & Concentration Kit (Zymo, D4014) or SPRI magnetic beads (Beckman, A63881). Both approaches produce comparable outcomes and can be used interchangeably. Depending on the availability of 3rd party reagents and instruments, the corresponding method should be applied.

**NOTE:** The pool may contain some cell debris, which could block a column membrane during purification leading to a long waiting time. To avoid this, it is recommended to perform a pre-cleaning of the RT pool by passing it through the Zymo column (see below) before mixing it with 7x DNA Binding buffer.

**The procedure of cDNA pre-cleaning and purification using the column-based method**
After the cDNA from each well is pooled in a reservoir, mix it with a 7x volume of DNA binding buffer (Zymo, D4004-1-L). We strongly recommend using a vacuum manifold for the cDNA purification to avoid column membrane damage due to multiple centrifugation rounds. A high-capacity Zymo-Spin IIICG column (Zymo, C1006-50-G) is required to purify large volumes resulting from 384 sample pooling.

| Plasticware | | | Pooling volumes | | | | | |
|---|---|---|---|---|---|---|---|---|
| Plate format | Pipetting strategy | Zymo-Spin column type | First strand cDNA | | DNA Binding Buffer | | TOTAL | |
| | | | Per well, µL | Per plate, mL | Per well, µL | Per plate, mL | Per well, µL | Per plate, mL |
| 96WP | multichannel pipette or pipetting robot | I (#D4014) | 20 | **1.92** | 140 | **13.44** | 160 | **15.36** |
| 384WP | pipetting robot | IIICG (C1006-50-G) | 10 | **3.84** | 70 | **26.88** | 80 | **30.72** |

Table 1: Overview of the recommended pipetting strategy, plasticware, and reagent volumes to be used depending on the number of pooled samples.

**Preparation**

• Make sure Zymo DNA Wash buffer has Ethanol added.

**Procedure**

2.2.1. According to Table 1, use a multichannel pipette or pipetting robot, to transfer the entire RT
volume (20 µL for **96WP**, 10 µL for **384WP**) of each sample into a specific reservoir (25 mL or
100 mL).

2.2.2. Mix the pool well and transfer it to a falcon tube with a pipette.

2.2.3. For RT pool pre-cleaning, place a Zymo column in a new 2 mL tube, add 800 µL of the collected
pool, and briefly centrifuge.

2.2.4. Collect the cleaned flowthrough in a new falcon tube. Repeat step 2.2.3 until all the pool passes
through the Zymo column. Discard the column.

2.2.5. Using a pipette, measure the volume of the pool after cleaning, transfer it to a 50 mL falcon tube
and add 7x DNA Binding buffer accordingly (see Table 1). The color of the mix should turn yellow.

2.2.6. Connect the 25 mL funnel (Zymo, C1039-25) to a Zymo column suitable for purification volume
(Table 1) and place it on a vacuum manifold.
2.2.7. Gently mix the cDNA in the binding buffer mixture and transfer it to a 25 mL funnel using a
2.2.8. 2.2.9. 2.2.10. 2.2.11. 2.2.12. 2.2.13. 2.2.14. 2.2.15. pipetboy.
Turn on the vacuum pump and let the liquid pass through the column.
Transfer any remaining volume to the funnel. Do not let the membrane over dry.
After the entire pool mix has passed through the column, add 200 µL of DNA Wash buffer (with
Ethanol added) directly to the membrane of the column.
Repeat step 2.2.10 once the wash buffer passed through the filter.
Remove the column from the vacuum manifold, put it in a 1.5 mL tube, and centrifuge for 1 min
to remove leftovers from the washing buffer.
Depending on the Zymo-Spin column type used, perform the following:

■ For the type **I** column used with ≤96 samples (**96WP**), add 20 µL of water to the column
matrix and incubate for 1 min.
■ For the type **IIICG** column used with 384 samples (**384WP**), add 38 µL of water to the column matrix and incubate for 1 min.
Transfer the column into a new labeled 1.5 mL tube and centrifuge for 30 sec.
Immediately proceed to step 2.3.

**The procedure of cDNA purification using the SPRI bead-based method**

Perform cDNA purification with SPRI magnetic beads using a 1:1 ratio of cDNA pool and beads slurry. The
purification of large volumes (i.e., 2 mL of the pool from **96WP** and 4 mL from **384WP**) requires three to six 1.5 mL
tubes and a corresponding magnetic stand (Permagen, MSR06).
If the volume of the pool is higher than 750 µL, split it equally in the required number of 1.5 mL tubes and add the
identical volume of beads (i.e., a pool of 1 mL split in 2 tubes with 500 µl per tube and add 500µL of beads per
tube).
**NOTE:** Please use SPRI-beads only if solution is clear and has no visible debris.

**Preparation**

Pre-warm beads at room temperature and vortex them vigorously before pipetting.
2.2.16. Pool the RT samples as described in Table 1.
2.2.17. Transfer the collected pool to a 2 mL or 15 mL tube, depending on the pooled volume. Consider
that the final volume will be twice higher due to the addition of the beads.
2.2.18. Add pre-warmed beads in a 1:1 ratio (i.e., for 960 µL of pooled samples, add 960 µL of beads
slurry), and mix by pipetting up and down ten times.
2.2.19. Incubate for 5 min at room temperature.
2.2.20. Place the tube on the magnetic stand, wait 5 min, and carefully remove and discard the
supernatant.
2.2.21. Incubate for 30 sec.
2.2.22. To wash the beads, pipette 1 mL of freshly prepared 80% ethanol into the tube.
2.2.23. Carefully remove the ethanol without touching the bead pellet.
2.2.24. Repeat step 2.2.21 for a total of two washes.
2.2.25. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.
2.2.26. Resuspend the beads in 37 µL of water and incubate for 1 min.
2.2.27. Place tubes on the magnetic stand, wait 5 min, and carefully transfer 35 µL of supernatant to a
new tube to avoid bead carry-over.
2.2.28. Immediately proceed to step 2.3.

If the RT pool was split into several tubes at step 2.2.16, use one of the following options:
• **[Two tubes only]**, resuspend the beads in **both tubes** in 20 µL/tube, and combine both elutions
in one tube;
• **[Two or more tubes]** resuspend the beads in the **first tube** in 40 µL of water. Keep other tubes
closed to avoid over-drying of the beads. Transfer obtained elution to the next tube and resuspend beads.
Repeat this step for every tube;
• **[Two or more tubes]**, resuspend **every** tube in 37 µL. Combine all elutions in one tube and
perform one additional purification of the pool adding beads slurry accordingly to the pool volume (steps
2.2.18 - 2.2.27). Elute in 37 µL of water and collect 35 µL in a new tube.

2.3 **Reverse transcription**

Each individual cell lysate sample is reverse-transcribed at this step using the barcoded oligo-dT primers
provided in a 96-well (**96WP**) or a 384-well (**384WP**) plate format, depending on the kit type.
Subsequently, all the barcoded samples can be pooled in one tube.

**NOTE:** Barcoded oligo-dT primers are provided lyophilized with the addition of dye. The dye has no impact
on the enzymatic reactions and is used solely for better visualization of reaction preparation and pooling.
Despite variations in appearance caused by the drying process, wells may exhibit traces of dried dye
ranging from dispersed to solid dots on the bottom. The following addition of RT reagents will enable the
visualization of red color, confirming the presence of the oligos in all wells.

**Preparation**

• Thaw the cell lysate samples on ice.
• Thaw the **RTB** reagent at room temperature and mix well before use.
• Briefly spin down the **96WP** or **384WP** plate containing dried oligo-dT primers. This plate will be
referred to as the RT plate.
• Prepare Program 1_RT on the thermocycler (set the lid at 90°C):

| Step | Temperature, °C | Time |
|---|---|---|
| Incubation | 50 | 30 min |
| Inactivation | 85 | 10 min |
| Keep | 4 | pause |

**NOTE:** All the manipulations with cell lysates and RT enzyme should be performed in an RNase-free
environment, with the use of RNase-free consumables and filter tips, on ice, and using gloves.

**Procedure for 96WP and 384WP**

**NOTE:** For the 96-well plate format, if only a portion of the plate is used, we recommend reconstituting
the oligos in 5 µL of water. Pipette 10-15 times and gently transfer 4.6 µL to the new plate where the RT
will be performed. Do not add water to the Master Mix for the RT reaction (as indicated in step 2.1.5
below).
For the 384-well plate format, the volume for oligo reconstitution will depend on the sample volumes, and
no additional water should be introduced into the master mix.
Before opening the seal, it is advisable to centrifuge the oligo plate, and **only the wells intended for use
should be opened**. The seal must remain on the plate to prevent any potential cross-contamination.
2.1.1. Keep the RT plate on ice. Using a multichannel pipette, transfer the following volume of cell lysate
directly to the corresponding wells and pipette 3-5 times to ensure proper reconstitution of dried
oligo-dT:
■ **96WP**: 10 µL
■ **384WP**: 5 µL
2.1.2. The appearance of red color in all wells indicates a proper and uniform reconstitution of oligos.
2.1.3. Carefully re-seal the RT plate and briefly spin it in the centrifuge.
2.1.4. Leave the RT plate on ice for 5 min.
2.1.5. Prepare the Master Mix for the RT reaction (+10%) as follows:

| Reagent | 96WP, µL | | 384WP, µL | |
|---|---|---|---|---|
| | Per well | 96 wells +10% | Per well | 384 wells +10% |
| RTB | 5.0 | 528.0 | 2.5 | 1056 |
| RTE | 0.1 | 10.6 | 0.1 | 43 |
| Water | 4.9 | 517.5 | 2.4 | 1014 |
| TOTAL | 10.0 | 1056.1 | 5.0 | 2113 |

2.1.6. Keep the RT plate on ice and, using a multichannel pipette, transfer the following volume of
Master Mix to each well containing the cell lysate sample:
■ **96WP**: 10 µL
■ **384WP**: 5 µL
2.1.7. Carefully re-seal the RT plate and briefly spin it in the centrifuge.
2.1.8. Transfer the plate to the thermocycler and start Program 1_
RT.
**Safe stop:** After this step, the RT plate can be kept at 4°C overnight.

## 2.4 Free primer digestion

It is recommended to perform non-incorporated primer digestion immediately after pooling.
**Preparation**
• Label 0.2 mL PCR tubes corresponding to the number of pools prepared.
• Thaw the **EXB** reagent at room temperature.
• Keep the **EXO** reagent on ice.
• Prepare Program 2_FPD on the thermocycler (set the lid at 90°C):

| Step | Temperature, °C | Time |
|---|---|---|
| Incubation | 37 | 30 min |
| Incubation | 80 | 20 min |
| Keep | 4 | pause |

**Procedure**
2.3.1. Depending on the cDNA volume obtained from steps 2.2.15 or 2.2.27, transfer 17 µL or 35 µL of
2.3.2. the eluate from each tube into a new labeled 0.2 mL PCR tube.
Prepare the EXO reaction mix as follows (with 10% excess):

| Reagent | Per reaction (µL) | |
|---|---|---|
| | For 17 µL elution | For 35 µL elution |
| EXB | 2.0 | 4.0 |
| EXO | 1.0 | 1.0 |
| TOTAL | 3.0 | 5.0 |

2.3.3. According to the table, transfer **3 µL** or **5 µL** of EXO reaction mix into each PCR tube with purified cDNA.
2.3.4. Mix by pipetting up and down 5 times.
2.3.5. Briefly spin down in the bench-top centrifuge.
2.3.6. Incubate in thermocycler Program 2_FPD.
2.3.6. Proceed immediately to step 2.4. or keep the tube at 4°C overnight.
**Safe stop**: After this step, the tube(s) can be kept at 4°C overnight.

## Library Sequencing

3 **LIBRARY SEQUENCING**
The libraries prepared with the MERCURIUSTM DRUG-seq kit carry Illumina- and AVITI-compatible adapter sequences. They can be processed on any Illumina instrument (e.g., HiSeq, NextSeq, MiSeq, iSeq, and NovaSeq) or in the Element AVITI System with Adept Workflow.

The MERCURIUS™ DRUG-seq libraries are Unique Dual-Indexed and can potentially be pooled in a sequencing run with other libraries if the sequencing structure is compatible. Please refer to Table 3 for the optimal sequencing structure and Table 4 for the list of i5 and i7 index sequences.
Given the DRUG-seq library structure, the optimal number of cycles for Read 1 is 28 (and 29 for AVITI). The following cycles, 29-60, will cover the homopolymer sequence, which may result in a significant drop of Q30 values reflecting sequencing quality. Therefore, using standard Illumina or AVITI runs setups (e.g., 100 PE or 150 PE) is not recommended.

| Read | Length (cycles) for Illumina | Length (cycles) for AVITI | Comment |
|---|---|---|---|
| Read 1 | 28 | 29 | Sample barcode (14 nt) and UMI (14 nt); +1 extra base for AVITI |
| Index 1 (i7) read | 8 | 8 | Library Index |
| Index 2 (i5) read | 8* | 8* | Library Index (*optional and valid for UDI libraries) |
| Read 2 | 60-90 | 101 | Gene fragment |

**Table 3** Sequencing structure of DRUG-seq libraries

The Unique Dual Indexing (UDI) strategy ensures the highest library sequencing and demultiplexing accuracy and
complies with the best practices for Illumina sequencing platforms. UD-indexed libraries have distinct index
adapters for i7 and i5 index reads (Table 4).

| Name | Type | i7 index sequence | i5 index sequence Forward Workflow | i5 index sequence Reverse Workflow |
|---|---|---|---|---|
| **MQ.UDI.1** | UDI (i7/i5) | TAAGGCGA | TATAGCCT | AGGCTATA |
| **MQ.UDI.2** | UDI (i7/i5) | CGTACTAG | ATAGAGGC | GCCTCTAT |
| **MQ.UDI.3** | UDI (i7/i5) | AGGCAGAA | CCTATCCT | AGGATAGG |
| **MQ.UDI.4** | UDI (i7/i5) | GCGTTGGA | TTGGACTT | AAGTCCAA |

**Table 4** UDI adapter sequences

**NOTE:** Sequencing depth
1. The recommended sequencing depth is 1-5 Mio reads per sample. Deeper sequencing can also be performed
to detect very lowly expressed genes.
2. If only one library is sequenced in a flow cell, the Index reads can be skipped.
3. The loading molarity for the library depends on the type of sequencing instrument (see 3.1 and 3.2) and should
be discussed with the sequencing facility or an experienced person.

## Sequencing Data processing

4 Following Illumina sequencing and standard library index demultiplexing, the user obtains raw read1 and read2 fastq sequencing files (e.g., mylibrary_R1.fastq.gz and mylibrary_R2.fastq.gz). This section explains how to generate ready-for-analysis gene, and UMI read count matrices from raw fastq files. To obtain the data ready for analysis, the user needs to align the sequencing reads to the genome and perform the gene/UMI read count generation, which can be done in parallel with the sample demultiplexing. For manual data processing, the user requires a terminal and a server or powerful computer with an installed set of standard bioinformatic tools.

4.1 Required software

• **fastQC** (version v0.11.9 or greater). Software for QC of *fastq* or *bam* files. This software is used to
assess the quality of the sequencing reads, such as the number of duplicates, adapter contamination,
repetitive sequence contamination, and GC content. The software is freely available from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. The website also contains informative
examples of good and poor-quality data.
• **STAR**solo from STAR (version 2.7.9a)**.** Software for read alignment on reference genome (Dobin et
al., 2013). It can be downloaded from Github (https://github.com/alexdobin/STAR). STAR can only be
run on UNIX systems and requires:
• x86-64 compatible processors
• 64-bit Linux or Mac OS X.
• ~30-40Gb of RAM
• **FastReadCounter** (v.1.1 or greater). Software for counting genome-aligned reads for genomic
features. **github.com/DeplanckeLab/FastReadCounter**
• **Picard** (v.2.17.8 or greater) and **Samtools** (v.1.9 or greater). Collections of command-line utilities to
manipulate with BAM files. Note: Picard requires **Java version 8 or higher** to be installed.
• R Software (version 3 or greater).
• (Optional) **BRBseqTools** (version 1.6). The software suite for processing DRUG-seq libraries is
available at https://github.com/DeplanckeLab/DRUG-seqTools.

## 4.2 Data processing

### 4.2.1. Merging fastq files from individual lanes and/or libraries (Optional)

4.2.1.1 Depending on the type of instrument used for sequencing, one or multiple R1/R2 *fastq* files per
library may result from individual lanes of a flow cell. The *fastq* files from individual lanes should
be merged into single *R1.fastq* and single *R2.fastq* files to simplify the following steps. This is an
example of *fastq* files obtained from HiSeq 4 lane sequencing:

> mylibrary_L001_R1.fastq.gz, mylibrary_L002_R1.fastq.gz,
mylibrary_L003_R1.fastq.gz, mylibrary_L004_R1.fastq.gz
> mylibrary_L001_R2.fastq.gz, mylibrary_L002_R2.fastq.gz,
mylibrary_L003_R2.fastq.gz, mylibrary_L004_R2.fastq.gz

4.2.1.2 To merge the *fastq* files from different lanes use a cat command in a terminal. This will generate
two files: *mylibrary_R1.fastq.gz* and *mylibrary_R2.fastq.gz*, containing the information of the
entire library.

> cat mylibrary_L001_R1.fastq.gz mylibrary_L002_R1.fastq.gz
mylibrary_L003_R1.fastq.gz mylibrary_L004_R1.fastq.gz >

**mylibrary_R1.fastq.gz**
> cat mylibrary_L001_R2.fastq.gz mylibrary_L002_R2.fastq.gz
mylibrary_L003_R2.fastq.gz mylibrary_L004_R2.fastq.gz >

**mylibrary_R2.fastq.gz**
4.2.1.3 Move these 2 *fastq* files into a new folder, which will be referenced in this manual as $fastqfolder.

**NOTE:** This step can also be done if you sequenced your library in multiple sequencing runs.
**Warning**: The order of merging files should be kept the same (for e.g., L001, L002, L003, L004, not L002,
L001 ...) to avoid issues when demultiplexing the samples.

**4.2.2. Sequencing data quality check**

4.2.2.1 Run fastQC on both R1 and R2 fastq files. Use –outdir option to indicate the path to the output
directory. This directory will contain HTML reports produced by the software.
> fastqc –-outdir $QCdir/ mylibrary_R1.fastq.gz
> fastqc –-outdir $QCdir/ mylibrary_R2.fastq.gz

4.2.2.2 Check fastQC reports to assess the quality of the samples (see Software and materials).

**NOTES:**
• The report for the R1 *fastq* file may contain some "red flags" because it contains barcodes/UMIs. Still,
it can provide useful information on the sequencing quality of the barcodes/UMIs.
• The main point of this step is to check the R2 *fastq* report. Of note, *per base sequence content* and
*kmer content* are rarely green. If there is some *adapter contamination* or *overrepresented sequence*

detected in the data, it may not be an issue (if the effect is limited to <10~20%). These are lost reads
but most of them will be filtered out during the next step.

**4.2.3. Preparing the reference genome**

The *fastq* files must be aligned (or "mapped") on a reference genome. The STAR (Dobin et al., 20131)
aligner is one of the most efficient tools for RNA-seq reads mapping. It contains a "soft-clipping" tool that
automatically cuts the beginning or the end of reads to improve the mapping efficiency, thus allowing the
user to skip the step of trimming the reads for adapter contamination. Moreover, STAR has a mode called
STARsolo, designed to align multiplexed data (such as DRUG-seq) and directly generate count matrices.
The STAR aligner requires a genome assembly together with a genome index file. The index file
generation is a time-consuming process that is only performed once on a given genome assembly so that
it can be completed in advance and the index files can be stored on the server for subsequent analyses.

4.2.3.1 Download the correct genome assembly fasta file (e.g., Homo_sapiens.GRCh38.dna.primary_assembly.fa) and gene annotation file in gtf format (e.g.,
Homo_sapiens.GRCh38.108.gtf) from Ensembl or UCSC repository. Below is an example of a
human assembly:

```
> wget https://ftp.ensembl.org/pub/release-108/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
> gzip -d Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz # unzip
> wget https://ftp.ensembl.org/pub/release-108/gtf/homo_sapiens/Homo_sapiens.GRCh38.108.gtf.gz
> gzip -d Homo_sapiens.GRCh38.108.gtf.gz # unzip
```

**NOTE:** It's recommended to download the primary_assembly fasta file when possible (without the 'sm' or
'rm' tags). If not available, download the top_level assembly. For the *gtf*, download the one that does not
have the 'chr' or 'abinitio' tags.

4.2.3.2 Use STAR to create an index for the genome assembly. Indicate the output folder name

containing the index files using *--genomeDir* option:
> STAR --runMode genomeGenerate --genomeDir /path/to/genomeDir --genomeFastaFiles Homo_sapiens.GRCh38.dna.primary_assembly.fa --sjdbGTFfile Homo_sapiens.GRCh38.108.gtf --runThreadN 8
**NOTES:**
• The --runThreadN parameter can be modified depending on the number of cores available on your
machine. The larger this number is, the more parallelized/fast the indexing will be.
• STAR can use up to 32-40Gb of RAM depending on the genome assembly. So, you should use a
machine that has this RAM capacity.

**4.2.4. Aligning to the reference genome and generation of count matrices**

After the genome index is created, both R1 and R2 *fastq* files can be aligned to this reference genome.
For this step, use the "solo" mode of STAR, which not only aligns the reads to the reference genome but
also creates the gene read count and UMI (unique molecular identifier) count matrices.
The following parameters should be adjusted according to the sequencing information:
• *--soloCBwhitelist*: a text file with the list of barcodes (one barcode sequence per lane) which is used
by STAR for demultiplexing. This file is provided according to version of the MERCURIUS kit used.
Example of "barcodes_96_V5D_star.txt":
> TACGTTATTCCGAA
> AACAGGATAACTCC
> ACTCAGGCACCTCC
> ACGAGCAGATGCAG
• *--soloCBstart*: Start position of the barcode in the R1 fastq file, equal to 1.
• *--soloCBlen*: Length of the barcode. This value should match the length of the barcode sequence in
the file specified by –soloCBwhitelist. The barcode length depends on the version of the oligo-dT
barcodes provided in the kit. For the barcode plate set V5, the default value is 14.
• *--soloUMIstart*: Start position of the UMI, it's *soloCBlen* + 1 since the UMI starts right after the barcode
sequence.
• *--soloUMIlen*: The length of UMI. This parameter depends on the version of the oligo-dT barcodes in
the kit and the number of sequencing cycles performed for Read1. For the barcode plate set V5 the
default value is 14.

• --readFilesIn: name and path to the input *fastq* files.

The order of the fastq files provided in the script is important. The first *fastq* must contain genomic
information, while the second the barcode and UMI content. Thus, files should be provided for STARsolo
in the following order: --readFilesIn mylibray_R2 mylibrary_R1.
This step will output *bam* files and count matrices in the folder $bamdir.
• --genomeDir: a path to the genome indices directory generated before ($genomeDir).
Output count matrix parameters:
By default, STARsolo produces a UMI count matrix, i.e., containing unique non-duplicated reads per
sample for each gene. This type of count data is a standard for single-cell RNA-seq analysis. For bulk
RNA-seq analysis, a gene read count matrix is usually used. The following parameters will enable the
generation of the output of interest.
--*soloUMIdedup* **NoDedup**, will generate a read count matrix output
--*soloUMIdedup* **NoDedup 1MM_Directional,** will generate both UMI and read count matrices in mtx
format.

> STAR --runMode alignReads --outSAMmapqUnique 60 --runThreadN 8 --outSAMunmapped Within --soloStrand Forward --quantMode GeneCounts --outBAMsortingThreadN 8 --genomeDir $genomeDir --soloType CB_UMI_Simple --soloCBstart 1 --soloCBlen 14 --soloUMIstart 15 --soloUMIlen 14 --soloUMIdedup NoDedup 1MM_Directional --soloCellFilter None --soloCBwhitelist barcodes.txt --soloBarcodeReadLength 0 --soloFeatures Gene --outSAMattributes NH HI nM AS CR UR CB UB GX GN sS sQ sM --outFilterMultimapNmax 1 --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --outFileNamePrefix $bamdir --readFilesIn mylibrary_R2.fastq.gz mylibrary_R1.fastq.gz
This step will output *bam* files and count matrices in the folder $bamdir.

The demultiplexing statistics can be found in the "*bamdir/Solo.out/Barcodes.stats*" file.
The alignment quality and performance metrics can be found in the "*bamdir/Log.final.out*" file.

**NOTE**: The most important statistic at this step is the proportion of "Uniquely mapped reads" which is
expected to be greater than 70% (for human, mouse or drosophila).

#### 4.2.5. Generating the count matrix from .mtx file

STARsolo will generate a count matrix (*matrix.mtx* file) located in the *bamdir/Solo.out/Gene/raw* folder.
This file is a sparse matrix format that can be transformed into a standard count matrix using an R script
provided below:

```
> #Myscript.R
> library(data.table)
> library(Matrix)
> matrix_dir ← "$bamdir/Solo.out/Gene/raw"
> f ← file(paste0(matrix_dir, "matrix.mtx"), "r")
> mat ← as.data.frame(as.matrix(readMM(f)))
> close(f)
> feature.names = fread(paste0(matrix_dir, "features.tsv"), header = FALSE,
stringsAsFactors = FALSE, data.table = F)
> barcode.names = fread(paste0(matrix_dir, "barcodes.tsv"), header = FALSE,
stringsAsFactors = FALSE, data.table = F)
> colnames(mat) ← barcode.names$V1
> rownames(mat) ← feature.names$V1
> fwrite(mat, file = umi.counts.txt, sep = "\t", quote = F, row.names = T,
col.names = T)
```

The resulting UMI/gene count matrix can be used for a standard expression analysis following
conventional bioinformatic tools.

#### 4.2.6. Generating the read count matrix with per-sample stats (Optional)

Given a multiplex BAM file obtained with STARsolo and a set of barcodes, the software FastReadCounter
produces a read count matrix with per-sample statistics with the following code:

```
> #!/bin/bash
>
> gtf_file=homo_sapience.gtf ### GTF genome annotation file
> output_folder=counts/ ### Name of the final count output file
> bam_dir=mypath/bam_demult ### Directory with demultiplexed bam
files
> barcode_file=V5D_96_frc.txt ### Barcode reference file
```

>
> FastReadCounter-1.0.jar" --bam ${bam_dir}/${bam_dir}.bam \
> --gtf "${gtf_file}" \
> --umi-dedup none \
> --barcodeFile \${barcode_file} \
> -o ${output_folder}

The resulting read count matrices can be used for subsequent gene expression analysis using established
pipelines and tools.

**NOTE:** Please contact us at info@alitheagenomics.com in case you don't have the barcode sequences
(in your email, please indicate the name of the barcode set and the PN of the barcode module).

**4.2.7. Demultiplexing bam files (Optional)**

Generation of demultiplexed bam files, i.e., individual bam files for each sample, might be needed in some
cases, for example, for submitting the raw data to an online repository that does not accept multiplexed
data (for example, GEO or ArrayExpress), or for running an established bulk RNA-seq data analysis
pipeline.
For this purpose, the Picard tool can be used with the following parameters:
• $out_dir, The output directory for demultiplexed bam files
• $path_to_bam, the path to multiplexed single bam file
• $barcode_brb.txt, tab-delimited file containing 2 columns: sample_id and barcode seq. Example of
barcode_96_V5D_brb.txt:
> Sample1 TACGTTATTCCGAA
> Sample2 AACAGGATAACTCC
> Sample3 ACTCAGGCACCTCC
> Sample4 ACGAGCAGATGCAG$

**NOTE:** This file is different from the list of barcode files provided to STAR.

Run the following Picard script:

> #!/bin/bash
> demultiplexed_bam_out_dir=$out_dir
> input_bam=$path_to_bam

```
> barcode_info=$barcode_brb.txt
>
> while IFS=$'\t' read -r -a line
> do
> sample_id="${line[0]}"
> tag_value="${line[1]}"
>
> java -jar /path/to/picard.jar FilterSamReads \
> I=${input_bam} \
> O=${demultiplexed_bam_out_dir}/${sample_id}.bam \
> TAG=CR TAG_VALUE=${tag_value} \
> FILTER=includeTagValues
> done < "$barcode_info"
```

**NOTE:** Please contact us at info@alitheagenomics.com in case you don't have the barcode sequences
(in your email, please indicate the name of the barcode set and the PN of the barcode module).

## Appendices

5 **Appendix 1. ERCC Spike-In Control**

The current protocol includes the addition of External RNA Controls Consortium (ERCC) Spike-Ins to the lysate buffer.
Prepare a 1:100 dilution of the ERCC RNA Spike-In mix in nuclease-free water. Mix 990 µL of pre-chilled water with 10 µL of ERCC. Pipette well and aliquot the dilution into 50 µL aliquots, keeping them at -20°C.
The working solution of 1x Cell Lysis Buffer with ERCC Spike-In controls consists of the following:

| Reagent | 96WP, µL | | 384WP, µL | |
|---|---|---|---|---|
| | Per well | 96 wells +10% | Per well | 384 wells +10% |
| CLB | 6.6 | 700 | 3.3 | 1400 |
| INH | 1.6 | 169.6 | 0.8 | 340 |
| Water | 11 | 1166 | 5.8 | 2457.6 |
| ERCC* (1:100) | 0.8 | 84.4 | 0.1 | 42.4 |
| TOTAL | 20 | 2120 | 10 | 4240 |

**The final ERCC is 1:1000 in a 384-type well (equal to 50 ng of RNA/well) and 1:250 in a 96-type well (150-200 ng of RNA/well)*

**Cell Lysis Buffer (CLB) preparation with ERCC**

1. Thaw the CLB and ERCC tubes on ice and avoid their long-term storage.
2. Keep the nuclease-free water on ice to maintain a cold temperature.
3. Spin down all the tubes before pipetting.
4. Add the water to a 15 mL falcon tube first, then the CLB, INH, and the ERCC (in this particular
order).
5. Pipette the prepared mix a few times and briefly spin the tube. Keep it on ice until further use.
6. Follow the main protocol for cell lysis procedure (step 1.3.13)

6 **Appendix 2. Compatible Illumina instruments**

Illumina instruments can use two workflows for sequencing i5 index (see the details in Indexed Sequencing Overview Guide on Illumina's website).

**Forward strand workflow instruments:**

- NovaSeq 6000 with v1.0 reagents
- MiSeq with Rapid reagents
- HiSeq 2500, HiSeq 2000

Reverse strand workflow instruments:

- NovaSeq 6000 with v1.5 reagents
- iSeq 100
- MiniSeq with Standard reagents
- NextSeq
- HiSeq X, HiSeq 4000, HiSeq 3000