

ALITHEA
GENOMICS



MERCURIUS™

**DRUG-seq
Library Preparation kit
for 96, 384, and 1536 samples**

PN 10841, 10851, 11041, 11051

User Guide

v.0.5.1.revA August 2023

FOR RESEARCH USE ONLY | ALITHEAGENOMICS.COM

Related Products

Kit name	Kit PN	Modules	Module PN
Mercurius DRUG-seq Library Preparation 96 kit	10841	Barcoded Oligo-dT Adapters Module 96 samples	10513
		DRUG-seq Library Preparation and UDI Module 96 samples	10541
Mercurius DRUG-seq Library Preparation 384 kit	10851	Barcoded Oligo-dT Adapters Module 384 samples	10515
		DRUG-seq Library Preparation and UDI Module 384 samples	10551
Mercurius DRUG-seq Library Preparation 384 (4x96) kit	11041	Barcoded Oligo-dT Adapters Module 4 x 96 samples	10513
		DRUG-seq Library Preparation and UDI Module 4 x 96 samples	10555
Mercurius DRUG-seq Library Preparation 1536 (4x384) kit	11051	Barcoded Oligo-dT Adapters Module 4 x 384 samples	10515
		DRUG-seq Library Preparation and UDI Module 4 x 384 samples	10556

Table of Contents

TABLE OF CONTENTS	2
KIT COMPONENTS	3
Reagents supplied	3
Additional required reagents and equipment (supplied by the user)	4
PROTOCOL OVERVIEW AND TIMING	5
PROTOCOL WORKFLOW	5
PART 1. PREPARATION OF CELL LYSATE SAMPLES	6
1.1. Essential considerations for input cells	6
1.2. Cell lysate preparation	6
PART 2. LIBRARY PREPARATION PROTOCOL	8
2.1. Reverse transcription	8
2.2. Samples pooling and column purification	9
2.3. Free primer digestion	11
2.4. Second strand synthesis	12
2.5. Tagmentation	13
2.6. Library indexing and amplification	14
2.7. Library quality control	15
PART 3. LIBRARY SEQUENCING	17
PART 4. SEQUENCING DATA PROCESSING	18
4.1. OPTION 1: Cloud-based platform for data processing	18
4.2. OPTION 2: Manual data processing	18
4.2.1. Required software	18
4.2.2. Merging <i>fastq</i> files from individual lanes and/or libraries (Optional)	18
4.2.3. Sequencing data quality check	19
4.2.4. Preparing the reference genome	19
4.2.5. Aligning to the reference genome and generation of count matrices	20
4.2.6. Generating the count matrix from .mtx file	21
4.2.7. Demultiplexing bam files (Optional)	21
4.2.8. Generating the read count matrix with per-sample stats (Optional)	22
APPENDIX 1. UNIQUE DUAL INDEXING ADAPTER SEQUENCES	23
APPENDIX 2. LIBRARY QUANTIFICATION WITH QPCR	24
CONTACT INFORMATION	25

Kit Components

Reagents supplied

Barcoded Oligo-dT Adapters Set V5A Module

Component Name	Label	Amount provided per kit				Storage
		96 samples (PN 10841)	384 samples (PN 10851)	4 x 96 samples (PN 11041)	4 x 384 samples (PN 11051)	
Plate with 96 barcoded oligo-dT primers, set V5A (PN 10513)	96 V5A OdT	1 plate	-	4 plates	-	-20°C
Plate with 384 barcoded oligo-dT primers, set V5A (PN 10515)	384 V5A OdT	-	1 plate	-	4 plates	-20°C
Aluminium Seal	-	3 pcs	3 pcs	12 pcs	12 pcs	-20°C/RT

Mercurius DRUG-seq Library Preparation and UDI Module

Component Name	Label	Cap colour	Volume (µL)				Storage
			96 samples (PN 10541)	384 samples (PN 10551)	4 x 96 samples (PN 10555)	4 x 384 samples (PN 10556)	
RT Enzyme	RTE	magenta	11	43	43	4 x 43	-20°C
RT Buffer	RTB	magenta	550	1100	2 x 1100	4 x 1100	-20°C
Exonuclease I Enzyme	EXO	purple	10	10	10	10	-20°C
Exonuclease Buffer	EXB	purple	20	20	20	20	-20°C
Second Strand Enzyme	SSE	orange	20	20	20	20	-20°C
Second Strand Buffer	SSB	orange	45	45	45	45	-20°C
Tagmentation Enzyme	TE	red	10	10	10	10	-20°C
Tagmentation Buffer	TAB	red	40	40	40	40	-20°C
Library Amplification Mix	LAM	green	200	200	200	200	-20°C
UDI Adapter Mix 1	MQ.UDI.1	transparent	10	10	10	10	-20°C
UDI Adapter Mix 2	MQ.UDI.2	transparent	10	10	10	10	-20°C
UDI Adapter Mix 3	MQ.UDI.3	transparent	10	10	10	10	-20°C
UDI Adapter Mix 4	MQ.UDI.4	transparent	10	10	10	10	-20°C
Nuclease-Free Water	NFW	transparent	1800	1800	2 x 1800	8 x 1800	-20°C
Cell Lysis Buffer	CLB	yellow	355	700	2 x 700	4 x 700	-20°C
RNase Inhibitor	INH	yellow	95	170	2 x 170	4 x 170	-20°C

Additional required reagents and equipment (supplied by the user)

Plasticware	Manufacturer	Product number
15 mL conical tubes	Greiner	188271
0.2mL 8-Strip Non-Flex PCR Tubes	Starlab	I1402-3700
Zymo-Spin IIICG Columns (optional)	Zymo	C1006-50-G
25 mL reservoir for spin columns	Zymo	C1039-25
Disposable Pipetting Reservoir 25mL polystyrene	Integra	4382
Solution reservoir 150 mL polystyrene (optional)	Integra	6318

Reagents	Manufacturer	Product number
DNA Clean and Concentrator-5 kit	Zymo	D4014
SPRI AMPure Beads	Beckman Coulter	A63881
or		
CleanNA Beads	CleanNA	CNGS0050D
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
High Sensitivity NGS Fragment Analysis Kit	Agilent	DNF-474
SYBR Green	Thermofisher	S7563
Ethanol, 200 proof	-	-
DPBS, Cell culture grade	Gibco	10010023

Equipment	Manufacturer	Product number
Benchtop centrifuge for plates	-	-
Benchtop centrifuge for 1.5 mL tubes	-	-
Single and Multichannel pipettes	-	-
Fragment Analyser / Bioanalyzer / TapeStation	Agilent	M5310AA
Qubit™	Invitrogen	Q33238
Magnetic stand for 0.2 mL tubes	Permagen	MSR812
Magnetic stand for 1.5 mL tubes	Permagen	MSR06
12-channel pipette, 0.5-12.5 µL VIAFLO or similar	Integra	4631
12-channel pipette, 5-125 µL VIAFLO or similar	Integra	4632
8-channel adjustable tip spacing pipette, VOYAGER, 2 – 50 µL	Integra	4726
Pipetboy	Integra	155 000
Vacuum manifold Qiagen Qiavac 24 Plus or similar	Qiagen	19413
Vacuum pump	-	-
Real-Time PCR Instrument (optional)	-	-
VIAFLO instrument (optional)	Integra	6001
VIAFLO 96/384 channel pipetting head, 0.5-12.5 µL (optional)	Integra	6101/6131

Protocol Overview and Timing

The MERCURIUS™ DRUG-seq kits allow the preparation of Illumina-compatible 3'RNA sequencing libraries for up to 1536 RNA 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 RT reaction, skipping the tedious RNA purification step.

The kits are provided in the following formats:

Kit format	PN	PCR plate format	Maximum number of samples in one pool	Maximum number of samples processable	Number of UDI libraries
96-sample	10841	96WP	96	96	4
384-sample	10851	384WP	384	384	4
4x96-sample	11041	96WP	96	384	4
4x384-sample	11051	384WP	384	1536	4

Every kit contains barcoded MERCURIUS™ 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 BRB-seq technology can be used to generate high-quality sequencing data from 2'000 – 50'000 mammalian cells per well. Notably, the kit can be used to pool any number of samples up to the capacity of the provided plate (96 or 384) with two considerations:

- The total cell number per pool should be at least 80'000.
- Pooling less than eight samples may result in low-complexity reads during sequencing, decreasing the overall sequencing quality. If necessary, the latter can be improved by increasing the proportion of PhiX spike-in control during sequencing (see Part 3).

Each library indexing is performed using a Unique Dual Indexing (UDI) strategy, which minimizes the risk of barcode misassignment after NGS. The kits contain four UDI adapters. Every adapter can be used to prepare an individual BRB-seq library. Libraries with different UDI adapters can be pooled and sequenced in a single flow cell.

Protocol Workflow

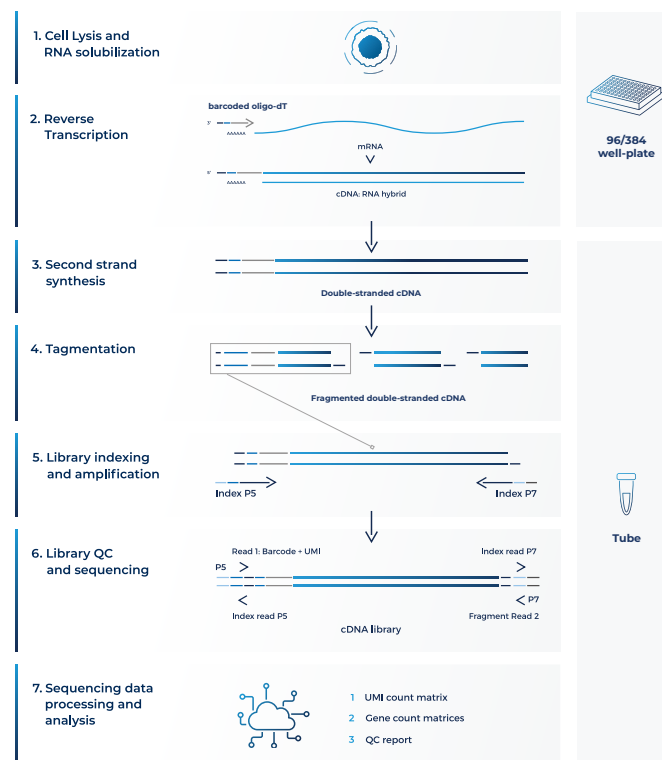


Figure 1 Schematic illustration of the protocol workflow

Part 1. PREPARATION OF CELL LYSATE SAMPLES

NOTE: Before starting every step, briefly spin down tubes and plates before opening them to ensure that all liquid or particles are collected at the bottom of the tube/plate.

1.1. Essential considerations for input cells

- The recommended input range of cells is 10'000-50'000 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 (human, mouse, cancer, or primary cells) and experimental design (e.g., induction of differentiation or apoptosis, cell cycle arrest, etc.), consider the doubling time of cells after the seeding and potential effect of the treatment on a 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. For this, we suggest using automatized cell seeding instruments or double-verified cell count.
- The cell lysate used in the RT reaction should be at most 50% of the final volume.

1.2. Cell lysate preparation

At this step, plated cells are lysed directly in a 96- or 384-well plate by adding a 1x Cell Lysis Buffer to pre-washed cells. The lysates can be directly used for the reverse transcription reaction.

Preparation

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	3.3	350	1.65	700
INH	0.8	85	0.4	170
NFW	5.9	664	2.95	1251
TOTAL	10	1099	5	2120

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

Procedure for adherent cells preparation

- 1.2.1. Seed the cells in a flat bottom 96WP or 384WP at the density that will enable harvest:
 - **96WP:** 10'000-50'000 cells per well
 - **384WP:** 2'000-10'000 cells per well
- 1.2.2. 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
- 1.2.3. Gently tap the plate and aspirate as much DPBS as possible without disturbing the cell pellet.
- 1.2.4. Immediately proceed to step [1.2.11](#).

Procedure for suspension cells preparation

- 1.2.5. Seed the cells in a flat bottom or U-shaped **96WP** or **384WP** at the density that will enable harvest:
 - **96WP:** 10'000-50'000 cells per well
 - **384WP:** 2'000-10'000 cells per well

- 1.2.6. Centrifuge the plate at 300x g for 5 minutes.
- 1.2.7. Gently aspirate culture media from the plate without disturbing the cell pellet and wash cells by adding:
 - **96WP**: 80-100 μ L of DPBS in each well
 - **384WP**: 30-50 μ L of DPBS in each well
- 1.2.8. Centrifuge the plate at 300x g for 5 minutes.
- 1.2.9. Gently tap the plate and aspirate as much DPBS as possible without disturbing the cell pellet.
- 1.2.10. Immediately proceed to step [1.2.11](#).

Procedure for cell lysis

- 1.2.11. Using a multi dispenser, distribute prepared CLB in every well.
 - **96WP**: 10 μ L per well
 - **384WP**: 5 μ L per well
- 1.2.12. Gently tap the plate to ensure that CLB is uniformly distributed on the surface of each well.
- 1.2.13. Incubate the plate on ice for 15 min, slightly agitating it from time to time.
- 1.2.14. Label a new 96- or 384-well PCR plate and transfer the entire volume of cell lysates from every well to a corresponding well of the new 96- or 384-well PCR plate. Preferably use a multichannel pipette and avoid intensive pipetting to reduce bubbles.
- 1.2.15. To avoid contamination of the lysate with floating debris and nuclei and to prevent a gDNA contamination of the samples, we strongly recommend centrifuging the plate with lysates at 600x g for 5 min (pre-cooled, +4°C).
- 1.2.16. Transfer the supernatant from every well to the corresponding well of the new 96- or 384-well PCR plate. Preferably use a multichannel pipette and avoid transferring any cell debris.
- 1.2.17. Seal the plate with an Aluseal provided and briefly spin it down.
- 1.2.18. 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.

Part 2. LIBRARY PREPARATION PROTOCOL

2.1. Reverse transcription

Each individual cell lysate sample is reverse-transcribed at this step using barcoded oligo-dT primers provided in 96-well (**96WP**) or 384-well (**384WP**) formats, depending on the kit type. Subsequently, all 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 a 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 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(s) containing dried oligo-dT primers. This plate will be referred to as the RT plate(s).
- 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**

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. Carefully re-seal the RT plate(s) and briefly spin it in the centrifuge.

2.1.3. Leave the RT plate(s) on ice for 5 min.

2.1.4. Prepare Master Mix for 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
NFW	4.9	517.5	2.4	1014
TOTAL	10.0	1056.1	5.0	2113

2.1.5. Keep the RT plate(s) 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

2.1.6. Carefully re-seal the RT plate(s) and briefly spin it in the centrifuge.

2.1.7. Transfer the plate to the thermocycler and start Program 1_RT.

Safe stop: After this step, the RT plate(s) can be kept at 4°C overnight.

2.2. Samples pooling and column purification

After pooling, the barcoded RT samples can be purified with 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: A 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, μ L	Per plate, mL	Per well, μ L	Per plate, mL	Per well, μ L	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 reagents 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 μ L for **96WP**, 10 μ L for **384WP**) of each sample into a specified reservoir.
- 2.2.2. Mix the pool well and transfer it to a falcon tube with a pipette.
- 2.2.3. For the RT pool pre-cleaning, place a Zymo column in a new 2 mL tube, add 800 μ L of the collected pool, and briefly centrifuge.
- 2.2.4. Collect a cleaned flowthrough in a new falcon tube. Repeat step [2.2.3](#) until all 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 a 7x DNA Binding buffer accordingly (see [Table 1](#)).
- 2.2.6. Add a volume of the 7x DNA Binding buffer according to the combined volume of the RT ([Table 1](#)). The color of the mix should turn yellow.
- 2.2.7. 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.
- 2.2.8. Gently mix the cDNA in the binding buffer mixture in a reservoir and, using a pipetboy, transfer it to a 25 mL funnel.
- 2.2.9. Turn on the vacuum pump and let the liquid pass through the column.
- 2.2.10. Transfer any remaining volume to the funnel, do not let the membrane overdry.
- 2.2.11. After the entire pool mix passed through the column, add 200 μ L of DNA Wash buffer (with Ethanol added) directly to the membrane of the column.
- 2.2.12. Repeat step [2.2.11](#) once the wash buffer passed through the filter.

- 2.2.13. Remove the column from a vacuum manifold, put it in a 1.5ml tube, and centrifuge for 1 min to remove leftovers from the washing buffer.
- 2.2.14. 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 NFW to the column matrix and incubate for 1 min.
 - For the type **III**CG column used with 384 samples (**384WP**), add 38 µL of NFW to the column matrix and incubate for 1 min.
- 2.2.15. Transfer the column into a new labeled 1.5 mL tube and centrifuge for 30 sec.
- 2.2.16. Immediately proceed to step **2.3**.

The procedure of cDNA purification using the SPRI beads-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 amount 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: However, we still recommend performing a pool pre-cleaning (see 2.2.1-2.2.4) before cDNA purification.

Preparation

Pre-warm beads at room temperature and vortex them vigorously before pipetting.

- 2.2.17. Pool the RT samples as described in **Table 1**.
- 2.2.18. 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 adding the beads.
- 2.2.19. 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.
- 2.2.20. Incubate for 5 min at room temperature.
- 2.2.21. Place the tube on the magnetic stand, wait 5 min, and carefully remove and discard the supernatant.
- 2.2.22. To wash the beads, pipette 1 mL of freshly prepared 80% ethanol into the tube.
- 2.2.23. Incubate for 30 sec.
- 2.2.24. Carefully remove the ethanol without touching the beads pellet.
- 2.2.25. Repeat step **2.2.22** for a total of two washes.
- 2.2.26. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.
- 2.2.27. Resuspend the beads in 37 µL of NFW and incubate for 1 min.
- 2.2.28. Place tubes on the magnetic stand, wait 5 min, and carefully transfer 35 µL of supernatant into a new tube to avoid bead carry-over.
- 2.2.29. Immediately proceed to step **2.3**.

If the RT pool was split into several tubes at step 2.2.17, 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 and more tubes]** resuspend the beads in the **first tube** in 40 µL of NFW. 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 and 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 2.2.19 - 2.2.28). Elute in 37 µL of NFW and collect 35 µL in a new tube.

2.3. Free primer digestion

It is recommended to perform the non-incorporated primer digestion immediately after pooling.

Preparation

- Label 0.2 mL PCR tubes corresponding to the number of pools prepared.
- Thaw **EXB** at room temperature.
- Keep **EXO** 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 2.2.16 or 2.2.28, transfer 17 μL or 35 μL of the eluate from each tube into a new labeled 0.2 mL PCR tube.
- 2.3.2. 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.
- 2.3.7. 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.

2.4. Second strand synthesis

The double-stranded full-length cDNA is generated and purified at this step using magnetic beads.

Preparation

- Pre-warm 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 **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

Second strand synthesis reaction setup

- 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 Program 3_SSS.
- 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 NFW.
- 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 beads pellet.
- 2.4.12. Repeat step 2.4.9 for the 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 NFW.
- 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 up to a few weeks.

2.5. Tagmentation

At this step, the full-length cDNA is tagmented using a Tn5 transposase pre-loaded with adapters for library amplification. It is recommended to use 20 ng of cDNA for tagmentation to obtain a higher library complexity with less PCR amplification.

Preparation

- Pre-warm SPRI beads at room temperature for ~30 min.
- If needed, prepare fresh 5 mL of 80% ethanol.
- Thaw the **TAB** reagent at room temperature and mix well before use.
- Keep the **TE** reagent constantly on ice.
- Set the PCR machine at 55°C incubation.

Procedure

- 2.5.1. Prepare the Tagmentation mix on ice in a PCR tube. If several cDNA samples are tagmented, prepare the Master mix as follows (with 10% excess):

Reagent	Per reaction (µL), for cDNA inputs		
	≤ 9 ng	10-14 ng	15-20 ng
TAB	4.0	4.0	4.0
TE	1.0	2.0	3.0
cDNA	X µL	X µL	X µL
NFW	15.0 – X	14.0 – X	13.0 – X
TOTAL	20.0	20.0	20.0

- 2.5.2. Keep the mix on ice and pipette up and down 10 times with the pipette set at 5 µL. Pay attention to mixing the reaction volume thoroughly.
- 2.5.3. Incubate for 7 min at 55°C in the PCR machine.
- 2.5.4. Immediately put the tube on ice and add 30 µL of water for the final volume of 50 µL.

Tagmented cDNA clean-up with SPRI beads

Purify the tagmented cDNA with SPRI magnetic beads using a 0.6x ratio.

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

- 2.5.5. Add 30 µL of beads to 50 µL of tagmented cDNA and mix by pipetting up and down 10 times.
- 2.5.6. Incubate for 5 min at room temperature.
- 2.5.7. Place the tubes on the magnetic stand, wait 5 min, and carefully remove and discard the supernatant.
- 2.5.8. To wash the beads, pipette 200 µL of freshly prepared 80% ethanol into the tube.
- 2.5.9. Incubate for 30 sec.
- 2.5.10. Carefully remove the ethanol without touching the beads pellet.
- 2.5.11. Repeat steps 2.5.8 - 2.5.10 for a total of two washes.
- 2.5.12. Remove tubes from the magnetic stand and let the beads dry for 1-2 minutes.
- 2.5.13. Resuspend the beads in 21 µL of NFW.
- 2.5.14. Incubate for 1 min.
- 2.5.15. 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.
- 2.5.16. Proceed immediately to step 2.6.

2.6. Library indexing and amplification

At this step, the 5' terminal fragments are amplified using Unique Dual Indexing (UDI) adapter primers. The kit contains 4 Illumina-compatible primer pairs to generate UDI libraries. The index sequences are indicated in [Table 3](#).

The number of amplification cycles required for library preparation is usually in the range of 12-17. The actual number may depend on the RNA samples and the input cDNA amount used for fragmentation. To determine the optimal number of cycles, follow the Library Quantification Protocol in [Appendix 2](#).

It is strongly recommended to perform the final library beads clean-up twice to remove the primer dimer fragments.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 mins.
- Prepare 10 mL of 80% ethanol.
- Thaw the **LAM** reagent on ice and mix well before use.
- Thaw the required number of **MQ.UDI Adapters** at room temperature, briefly spin before use.
- Prepare **Program 4_TN5AMP** (set the lid at 100°C) on the thermocycler (* The exact number of PCR cycles should be determined following [Appendix 2](#)):

Step	Temperature, °C	Time	Cycles
Initial denaturation	98	1 min	
Denaturation	98	10 sec	
Annealing	63	30 sec	15 or TBD*
Extension	72	1 min	
Final extension	72	3 min	
Keep	4	pause	

Library amplification reaction setup

2.6.1. Prepare the PCR amplification reaction as follows:

Reagent	Per reaction (µL)
LAM	25.0
MQ.UDI Adapter	5.0
Tagmented cDNA	20.0
TOTAL	50.0

2.6.2. Pipette up and down 5 times.

2.6.3. Put the tube in the PCR machine and start **Program 4_TN5AMP** with the default 15 PCR cycles or follow [Appendix 2](#) to determine the optimal number of amplification cycles.

Indexed cDNA library clean-up with SPRI beads

Purify the final cDNA library with SPRI magnetic beads using a 0.7x ratio (35 µL of bead slurry for 50 µL cDNA library).

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

2.6.4. Adjust the library volume to 50 µL with NFW.

2.6.5. Add 35 µL of beads and mix by pipetting up and down 10 times.

2.6.6. Incubate for 5 min at room temperature.

2.6.7. Place the tubes on the magnetic stand, wait 5 min, carefully remove, and discard the supernatant.

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

2.6.9. Incubate for 30 sec.

2.6.10. Carefully remove the ethanol without touching the beads pellet.

2.6.11. Repeat step [2.6.8](#) for a total of two washes.

2.6.12. Remove tubes from the magnetic stand and let the beads dry for 1-2 min.

- 2.6.13. Resuspend the beads in 21 μL of NFW.
- 2.6.14. Incubate for 1 min.
- 2.6.15. Place tubes on the magnetic stand, wait 5 minutes, and carefully remove 20 μL of supernatant into a new tube to avoid bead carry-over.
- 2.6.16. Perform the bead clean-up once again by repeating the procedure from step 2.6.4.

Safe stop: At this step, the cDNA libraries can be safely kept at -20°C for up to a few weeks.

2.7. Library quality control

Pooled library quality control

Before sequencing, the libraries should be subjected to fragment analysis (with Fragment analyzer, Bioanalyzer, or TapeStation) and quantification (with Qubit). This information is required to assess the molarity of the libraries and prepare the appropriate library dilution for sequencing. A successful library contains fragments in the range of 300 – 1000 bp with a peak at 400-700 bp; see Figure 2 for an example of a standard BRB-seq library profile.

Importantly, libraries with primer dimer peaks at 180 bp and ranging at 250 – 290 bp will likely produce lower-quality sequencing data with a reduced proportion of mappable reads (Figure 3). Therefore, it is strongly recommended to remove those peaks by performing an additional round of SPRI beads purification with the 0.7x ratio (see step 2.6.4).

Undertagmented libraries have a broader fragment range distribution with a peak at 1000-1200 bp (Figure 4). Only a fraction of such libraries contains fragments that can be efficiently sequenced; therefore, it is recommended to re-tagment the cDNA for the best results.

Library quantification can also be done unbiasedly by qPCR using standard Illumina library quantification kits (i.e., KAPA HiFi, Roche).

Pre-sequencing library QC:

- Use 2 μL of the library to measure the concentration with Qubit.
- Use 2 μL of the library to assess the profile with the Fragment Analyzer instrument or similar.
- If necessary, re-purify the libraries by following steps 2.6.4 – 2.6.15 to remove the peaks <300 bp.

Figure 2 The successful library profile with fragments 300-1000 bp

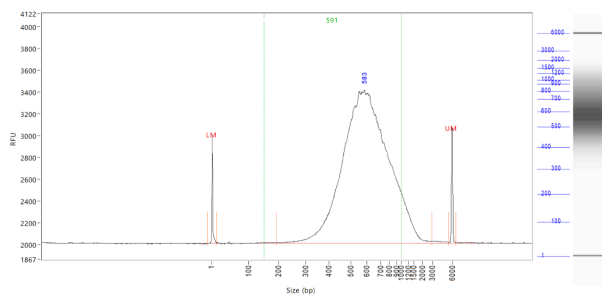


Figure 3 The example of an over-tagmented library profile with a peak at 290 bp and an adapter peak at 160 bp

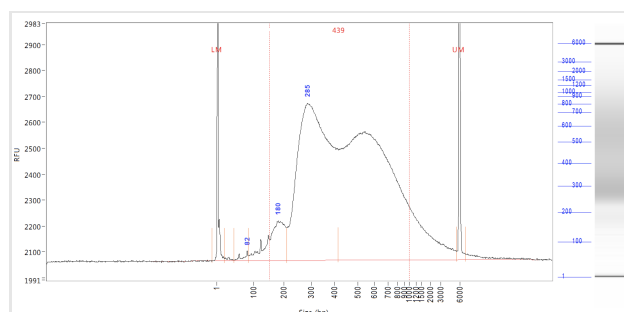
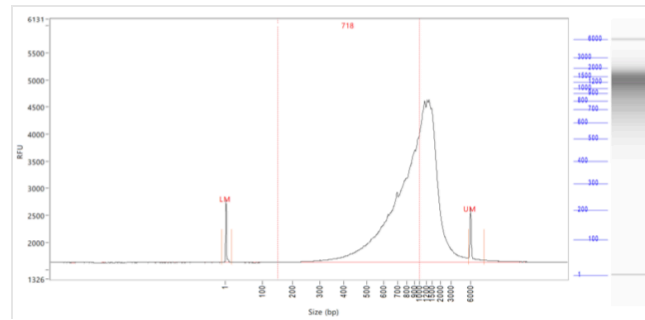


Figure 4 The example of an under-tagmented library profile with a major peak > 1000 bp



Assessing uniformity of reads distribution across the samples.

For projects involving highly heterogeneous samples, it is recommended to validate the uniformity of reads coverage across the samples by shallow library sequencing (see step 2.2). This approach ensures that every sample will obtain enough reads required for the analysis. BRB-seq libraries can be added as spike-ins to the compatible sequencing run (see Part 3). For this validation, 0.5-1M sequencing reads per library is sufficient to assess the fraction of reads attributed to every sample.

Part 3. LIBRARY SEQUENCING

The libraries prepared with the MERCURIUS™ BRB-seq kit carry Illumina-compatible adapter sequences. They can be processed on any Illumina instrument (i.e., HiSeq, NextSeq, MiSeq, iSeq, and NovaSeq).

The MERCURIUS™ BRB-seq libraries are Unique Dual Indexed and potentially can be pooled in a sequencing run with other libraries if the sequencing structure is compatible. Please refer to [Table 2](#) for the optimal sequencing structure and [Table 3](#) for the list of i5 and i7 index sequences.

Given the BRB-seq library structure, the optimal number of cycles for Read 1 is 28. The following cycles, 29-60, will cover the homopolymer sequence, which may significantly drop Q30 values reflecting sequencing quality. Therefore, using standard Illumina runs setups (e.g., 100 PE or 150 PE) is not recommended.

Read	Length (cycles)	Comment
Read 1	28	Sample barcode (14 nt) and UMI (14 nt)
Index 1 (i7) read	8	Library Index
Index 2 (i5) read	8*	Library Index (*optional and valid for UDI libraries)
Read 2	60-90	Gene fragment

Table 2 Sequencing structure of BRB-seq libraries

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 both i7 and i5 index reads ([Table 3](#)). Please refer to [Appendix 1](#) for the list of Illumina instruments with forward or reverse workflow.

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 3 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 enable the detection of very low 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 sequencing instrument (see [Table 4](#)) and should be discussed with the sequencing facility or an experienced person.

Instrument	Final loading concentration	PhiX
MiSeq	20 pM	1 %
iSeq	100 pM	1 %
NextSeq Standard Workflow	2.2 pM	1 %
NovaSeq XP Workflow*	160 pM	1 %
HiSeq4000	270 pM	1 %

* - adjusted molarity for BRB-libraries sequencing. We recommend a prior dilution of the libraries to 0.8nM before denaturation.

Table 4 Reference loading concentrations for various Illumina instruments

Part 4. SEQUENCING DATA PROCESSING

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.

4.1. OPTION 1: Cloud-based platform for data processing

The users of BRB-seq kits are offered free access to the devoted cloud-based BRB-seq data processing pipeline. This easy-to-use workflow enables the data processing to generate gene and UMI count matrices for the downstream RNA-seq analysis from the raw *fastq* files. The pipeline also produces a QC report and can be operated by a user without bioinformatics or coding skills.

The pipeline, still in the developer beta testing mode, is accessible via the following [link \(https://gcw-portal-ali.genomsys.com/\)](https://gcw-portal-ali.genomsys.com/). Please request credentials to access the pipeline by email at info@alitheagenomics.com

Currently, only the human, mouse, and fly genomes are supported.

4.2. OPTION 2: Manual data processing

To obtain the data read 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.

The user requires a terminal and a server or powerful computer with an installed set of standard bioinformatic tools for manual data processing.

4.2.1. Required software

- [fastQC](#) (version v0.11.7 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](#) (version 2.7.10a). Software for reads alignment on reference genome (Dobin et al., 2013). It can be downloaded from [Github](https://github.com/alexdobin/STAR) (<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
- [FeatureCounts](#) (v.1.6.0 or greater). Software for counting genome-aligned reads for genomic features, such as genes and exons.
- [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 BRB-seq libraries is available from: <https://github.com/DeplanckeLab/BRB-seqTools>.

4.2.2. Merging *fastq* files from individual lanes and/or libraries (Optional)

4.2.2.1 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 a single *R1.fastq* and a 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
```

4.2.2.2 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
```

4.2.2.3 Move these 2 *fastq* files in 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 (e.g., L001, L002, L003, L004, not L002, L001 ...) to avoid issues when demultiplexing the samples.

4.2.3. Sequencing data quality check

4.2.3.1 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
```

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

NOTES:

- Report for 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 some *adapter contamination* or *overrepresented sequence* is 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.

4.2.4. 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 the 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 BRB-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.

4.2.4.1 Download the right 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 for 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
```

¹ Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, Thomas R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, Volume 29, Issue 1, January 2013, Pages 15–21, <https://doi.org/10.1093/bioinformatics/bts635>

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.

4.2.4.2 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.

4.2.5. 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 the version of the MERCURIUS kit used. Example of "[barcodes_96_V5A_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 the 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 mylibrary_R2 mylibrary_R1`.

- `--genomeDir`: a path to the genome indices directory generated before (`$genomeDir`).

Output count matrix parameters:

By default, STARsolo produces 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. Following parameters will enable generating the output of interest.

`--soloUMI dedup NoDedup`, will generate a read count matrix output

`--soloUMI dedup NoDedup 1MM_All`, will generate both UMI and read count matrices in *mtx* format.

This step will output *bam* files and count matrices in the folder `$bamdir`.

```
> 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 --
soloUMI dedup NoDedup 1MM_All --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
```

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).

4.2.6. 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.

4.2.7. Demultiplexing bam files (Optional)

Generation of demultiplexed bam files, i.e., individual bam file 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:

- `$out_dir`, The output dir for demultiplexed bam files
- `$path_to_bam`, the path to multiplexed single bam file
- `$barcode_brb.txt`, the tab-delimited file containing 2 columns: `sample_id` and barcode seq. Example of [barcode_96_V5A_brb.txt](#):

```
> Sample1      TACGTTATTCGAA
> Sample2      AACAGGATAACTCC
> Sample3      ACTCAGGCACCTCC
> Sample4      ACGAGCAGATGCAG
```

NOTE: This file differs from the list of barcodes 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)

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

After splitting (demultiplexing) the *bam* file into files for each sample, produce a read counting matrix with per-sample statistics using the *featureCounts* tool with the following code:

```
> #!/bin/bash  
>  
> gtf=homo_sapience.gtf          ### GTF genome annotation file  
> output_name=read_counts.txt    ### Name of the final count  
> output file  
> bam_dir=myspath/bam_demult    ### Directory with demultiplexed bam  
> files  
>  
> featureCounts [options] -t exon -g gene_id -s 1 -a $gtf -o $output_name  
> ${bam_dir}/*.bam
```

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

Appendix 1. Unique Dual Indexing adapter sequences

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

Appendix 2. Library quantification with qPCR

To determine the optimal number of cycles, the following library quantification procedure using real-time PCR is recommended at step 2.6.3.

- After purification of the tagmented library, set up the PCR amplification reaction and perform 5 cycles of library preamplification (Program 4_TN5AMP).
- Put the tube on ice.
- Use 5 μ L aliquot from the PCR reaction to prepare a qPCR reaction mix in the appropriate PCR tube or plate as follows:

Reagent	Per reaction, μ L	N +10%, μ L
5 cycles pre-amplified library	5.0	
SYBR 100X*	0.1	
LAM	2.5	
NFW	2.4	
TOTAL	10.0	

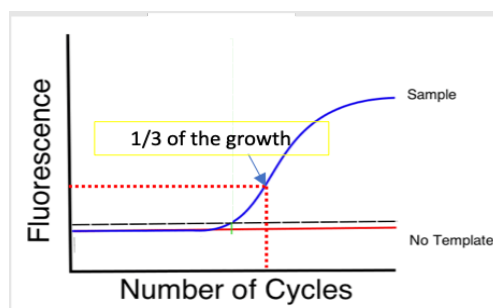
* Prepare 100X dilution with NFW from 10'000X stock

- Put it in the qPCR machine and run it with the following program:

Step	Temperature	Time	Cycles
Initial denaturation	98°C	1 min	
Denaturation	98°C	10 sec	25
Annealing	63°C	30 sec	
Extension + acquisition	72°C	1 min	

- Determine the cycle number depending on the growth curve in the multicomponent plot, as in Figure 5.
- Perform PCR with a determined additional number of cycles with the tube containing the remaining 45 μ L of the pre-amplified library.

Figure 5 Determination of the additional number of amplification cycles with qPCR



Typically, 10 ng of cDNA is sufficient to obtain 20-40 ng of DNA library after 13-14 cycles of amplification. Table 5 shows the approximate amount of amplification cycles and the expected library yield.

Tagmented cDNA, ng	# PCR cycles	Expected Library yield, ng
5	14-15	
10	13-14	20-40
20	11-12	

Table 5 Expected yield of BRB-seq libraries from different amounts of tagmented cDNA

Contact information

Alithea Genomics SA

Route de la Corniche 5
1066 Épalinges
Switzerland
Tel: +41 78 830 31 39

www.alitheagenomics.com

info@alitheagenomics.com