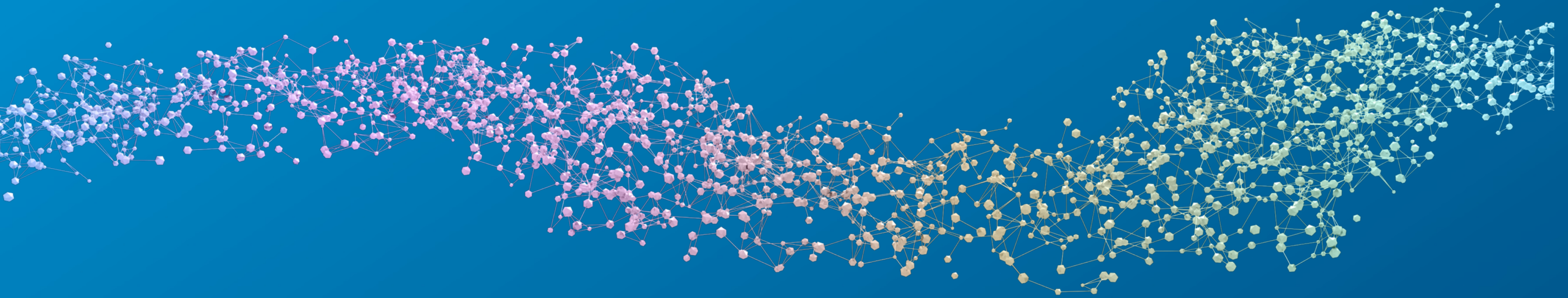
ALITHEA
GENOMICS



MERCURIUS™

DRUG-seq Library Preparation Kit for 1536-Well Plates

PN 11091

User Guide

Early Access

June 2026

FOR RESEARCH USE ONLY | ALITHEAGENOMICS.COM

Related Products

Kit Name	Kit PN	Module	Module PN
MERCURIUS™ DRUG-seq Library Preparation Kit for 1536 Well-Plates	11091	Barcoded Oligo-dT Adapters Module 1 (384)	10430-1
		Barcoded Oligo-dT Adapters Module 2 (384)	10430-2
		Barcoded Oligo-dT Adapters Module 3 (384)	10430-3
		Barcoded Oligo-dT Adapters Module 4 (384)	10430-4
		DRUG-seq Library Preparation and UDI Module 1536 Samples	10791

Table of Contents

TABLE OF CONTENTS **2**
KIT COMPONENTS **3**
PROTOCOL OVERVIEW **6**
PROTOCOL WORKFLOW **7**
PART 1. PREPARATION OF CELL PELLET SAMPLES **8**
1.1. Essential considerations for input cells 8
1.2. Essential considerations for cell culture plates 8
1.3. Cell pellet preparation 8
PART 2. LIBRARY PREPARATION PROTOCOL **9**
2.1. Lysis and reverse transcription mix distribution 9
2.2. RT incubation 11
2.3. Sample pooling 12
2.4. Sample Purification 12
2.5. Free primer digestion 14
2.6. Second-strand synthesis 15
2.7. Tagmentation 16
2.8. Library indexing and amplification 17
2.9. Library quality control 19
PART 3. LIBRARY SEQUENCING **22**
3.1. Sequencing on the Illumina instruments 23
3.2. Sequencing on the Element AVITI instruments 23
PART 4. SEQUENCING DATA PROCESSING **24**
4.1. Required software 24
4.2. Data processing 24
APPENDIX 1. ERCC SPIKE-IN CONTROL **29**
APPENDIX 2. COMPATIBLE ILLUMINA INSTRUMENTS **30**

Kit Components

Reagents supplied

Barcoded Oligo-dT Adapters Set V6 Module (10430)

Component Name	Label	Provided	Storage
Plate 1 with 384 barcoded oligo-dT primers, set V6A (PN 10430-1)	1536 V6 384 P1	1 plate	-20°C
Plate 2 with 384 barcoded oligo-dT primers, set V6A (PN 10430-2)	1536 V6 384 P2	1 plate	-20°C
Plate 3 with 384 barcoded oligo-dT primers, set V6A (PN 10430-3)	1536 V6 384 P3	1 plate	-20°C
Plate 4 with 384 barcoded oligo-dT primers, set V6A (PN 10430-4)	1536 V6 384 P4	1 plate	-20°C

DRUG-seq Library Preparation and UDI Module (10791)

Component Name	Label	Cap color	Volume, µL	Storage
Cell Lysis Buffer	CLB	yellow	150	-20°C
RNase Inhibitor	INH	yellow	470	-20°C
Oil Blue Dye	OBD	transparent	800	-20°C/RT
RT Enzyme	RTE	magenta	55	-20°C
RT Buffer	RTB	magenta	2 x 1400	-20°C
Exonuclease I Enzyme	EXO	purple	10	-20°C
Exonuclease Buffer	EXB	purple	20	-20°C
Second Strand Enzyme	SSE	orange	20	-20°C
Second Strand Buffer	SSB	orange	45	-20°C
Tagmentation Enzyme	TE	red	30	-20°C
Tagmentation Buffer	TAB	red	40	-20°C
Library Amplification Mix	LAM	green	200	-20°C
UDI Adapter Mix 1	MQ.UDI.1	transparent	10	-20°C
UDI Adapter Mix 2	MQ.UDI.2	transparent	10	-20°C
UDI Adapter Mix 3	MQ.UDI.3	transparent	10	-20°C
UDI Adapter Mix 4	MQ.UDI.4	transparent	10	-20°C

Additional required reagents and equipment (supplied by the user)

Plasticware	Manufacturer	Product number
1536 well-plate (validated options):		
1536 Well Cell Culture Microplate CELLSTAR®, PS, HiBase	Greiner Bio-One	782080
1536 Well SCREENSTAR Microplate, Cycloolefin, TC, Sterile	Greiner Bio-One	789866
E8 Blk, 1536, SQ Well, ULowBase, 188um, TC Treated CO2	Aurora Microplates	EBC201200A
15 mL conical tubes	Corning	352096
twin.tec PCR Plate 384	Eppendorf	951020702
Adhesive PCR sealing foil sheets	Thermo Fisher	AB-0626
VBLOK200 reservoir	ClickBio	VBLOK200-1
50 mL conical tube (pooling method dependent)	Corning	352070
Zymo-Spin IIICG Columns (pooling method dependent)	Zymo	C1006-50-G

Reagents	Manufacturer	Product number
DPBS, Cell culture grade	Gibco	10010023
Nuclease-free water	Thermo Fisher	A57775
ERCC RNA Spike-In Mix (optional)	Thermo Fisher	4456740
Oil (validated options):		
• Mineral oil	Sigma-Aldrich	69794
• Silicone oil	Sigma-Aldrich	317667
• Vapor-Lock	Qiagen	981611
DNA Clean and Concentrator-5 kit (pooling method dependent)	Zymo	D4014
SPRI Magnetic Beads (recommended):		
• AMPure XP Beads for DNA Cleanup	Beckman Coulter	A63881
• CleanNGS magnetic beads	CleanNA	CNGS0050
• Mag – Bind® TotalPure NGS	Omega Bio-tek	M1378
• MagFlo™ NGS	Integra	7000
• MagMax™ PureBind	Applied biosystems	A58521
Ethanol, 200 proof	-	-
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
High Sensitivity NGS Fragment Analysis Kit (optional)	Agilent	DNF-474
SYBR Green (optional)	Thermo Fisher	S7563

Equipment	Manufacturer	Product number
Centrifuge for plates	-	-
Benchtop centrifuge for 0.2 mL tubes	-	-
Centrifuge for 1.5 mL tubes (pooling method dependent)	-	-
Centrifuge for 15 mL tubes	-	-
Single and Multichannel pipettes	-	-
Magnetic stand for 0.2 mL tubes	Permagen	MSR812
Magnetic stand for 5 mL tubes (pooling method dependent)	Permagen	MSR8X5
Vacuum manifold (pooling method dependent)	Macherey-Nagel	730150N
Vacuum pump (pooling method dependent)	-	-
Pipetboy (pooling method dependent)	Integra	155 000
Vortex	-	-
Qubit™	Invitrogen	Q33238
Fragment Analyser (or Bioanalyzer or TapeStation)	Agilent	M5310AA
Real-Time PCR instrument (optional)	-	-
384-channel liquid handler (recommended options): • Firefly • VIAFLO 384 • Microlab STARlet • Bravo Automated Liquid Handling Platform	 SPT Labtech Integra Hamilton Agilent	 3276-00006 6031 - G5563AA
Microplate dispenser (recommended options): • WellJet • Multidrop Combi+	 Integra Thermo Fisher	 5000 5840330
Incubator (validated options): • BD23 • IF55	 Binder Memmert	 9010-0187 IF55
VBLOK200 reservoir compatible centrifugation system options: • Multifuge X4 Pro with HIGHPlate™ 6000 Microplate Rotor • Thermo Fisher Multifuge / Megafuge with M-20 Microplate Rotor (optional – with Sealed Microplate Carriers) • Sorvall Legend T (or RT) Plus with HIGHplate® Windshielded Microplate Rotor • ROTANTA 460 with Swing-out rotor, 2-place	 Thermo Fisher Thermo Fisher Thermo Fisher Hettich	 75009900 / 75003606 75003624 (75003625) 75004363 (750004373) / 75006444 5650 / 5622

Protocol Overview

MERCURIUS™ DRUG-seq kit for 1536-well plates allows for the preparation of Illumina-compatible 3' RNA sequencing libraries from up to 1536 frozen cell samples in a time and cost-efficient manner. The workflow largely follows the standard MERCURIUS™ DRUG-seq procedure (10841, 10851, 11041, 11051), with the initial steps optimized for 1536-well plates. A single lysis and reverse transcription (RT) master mix is added directly to the frozen cells in the culture plate. The RT reaction is then incubated in the same plate using a standard incubator.

The kit is provided in the following format:

Kit format	PN	Plate format	Maximum number of samples in one pool	Maximum number of samples	Number of UDI libraries
1536 samples	11091	1536	1536	1536	4

Each kit contains four plates of 384 barcoded MERCURIUS™ Oligo-dT primers, for a total 1536 barcodes, designed to tag polyA+ RNA from lysates with individual barcodes during the first-strand synthesis reaction. This allows the pooling of the resulting cDNA samples from each experimental group into a single tube, streamlining sequencing library preparation.

The proposed DRUG-seq technology enables the generation of high-quality sequencing data from 400 to 1,000 mammalian cells per well.

Library indexing is performed using Unique Dual Indexes (UDIs), minimizing the risk of barcode misassignment during demultiplexing. The kit includes four UDI adapters, each enabling the preparation of a distinct library. Libraries prepared with different UDIs can be pooled and sequenced together within a single flow cell.

Protocol Workflow

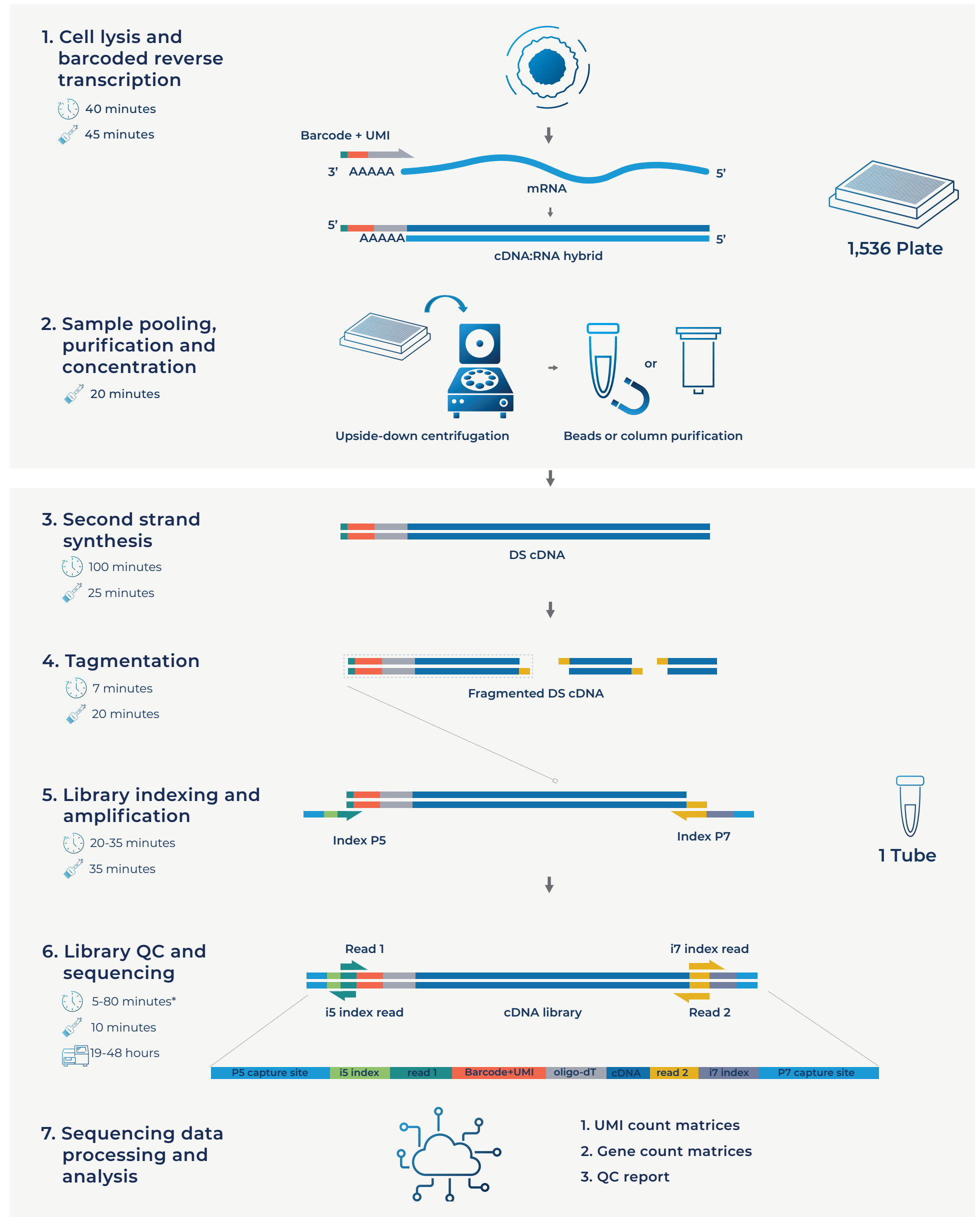


Figure 1 Schematic illustration of the protocol workflow.

Part 1. PREPARATION OF CELL PELLET SAMPLES

1.1. Essential considerations for input cells

- **We recommend seeding 400–1,000 cells per well**. Because 1536-well plates have a limited surface area, it is advisable to perform a cell-titration experiment and visually assess confluence to determine the optimal seeding density for your specific cell type.
- **Seed cells a ≥1 day in advance** to ensure they reach the appropriate growth phase for optimal assay performance.
- **Consider the biology of your model system** - including species (e.g., human, mouse), cell origin (e.g., cancer, primary), and experimental conditions (e.g., induction of differentiation, apoptosis, or cell-cycle arrest). Account for cell doubling time after seeding as well as any expected treatment-related effects on cell number during the assay.
- **Ensure uniform starting material** to achieve an even distribution of sequencing reads. To improve consistency, we recommend using automated cell-seeding systems or validating cell counts through duplicate measurements.

1.2. Essential considerations for cell culture plates

The RT and lysis steps of the MERCURIUS™ 1536-well plate DRUG-seq workflow are performed directly in the 1536-well cell culture plates rather than in standard PCR plates. This decreases the number of pipetting steps in and out of these plates and increases sensitivity but also introduces two important considerations:

1. Surface coatings (e.g., collagen or other ECM coatings)

Many users employ cell culture plates with specialized surface treatments or biological coatings designed to promote cell adhesion. Some coatings could interact with reagents in the Lysis/RT mix, potentially affecting the enzymatic reactions.

2. Hydrophilic surface treatment and hydrophobic forces

Unlike PCR plates, which are deliberately hydrophobic, cell culture plates are treated to be hydrophilic, to improve cell attachment. As a result, capillary action can occur at the liquid–wall interface, leading to the formation of a meniscus and the upward creep of the reaction mixture along the edges of the wells. Given the small well volumes, this effect can become pronounced; in extreme cases, liquid may accumulate at the rim and even spill out of the well.

An oil overlay, in addition to preventing evaporation, suppresses this capillary-driven redistribution by reducing the liquid–air interface and stabilizing the reaction volume.

Practical recommendation

Because plate composition, coatings, and surface treatments can vary widely across vendors and product lines, we strongly recommend validating any new 1536-well plate before use. If plate-related interference is suspected, or if you need guidance on plate selection, please contact us at *info@alitheagenomics.com*.

1.3. Cell pellet preparation

1.3.1. Seed the cells in a flat-bottom 1536-well plate.

1.3.2. Gently aspirate the culture medium from the plate.

1.3.3. Wash the cells by adding 5-7 µL of DPBS to each well. Aspirate **as much DPBS as possible** without disturbing the cell pellet.

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

1.3.5. Proceed to step 2.1 for the combined cell lysis and RT.

Part 2. LIBRARY PREPARATION PROTOCOL

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

All manipulations with samples and RT enzyme should be performed in an RNase-free environment, using RNase-free consumables and filter tips, on ice, and with gloves.

2.1. Lysis and reverse transcription mix distribution

Each cell sample is lysed and reverse-transcribed directly in the 1536-well cell culture plate using barcoded oligo-dT primers provided in 4 x 384-well plates. The current protocol uses an **oil overlay strategy** to prevent evaporation during RT incubation. Subsequently, all barcoded samples can be pooled into one single tube as described in step 2.3.

The Lysis/RT master mix distribution to the 1536-well plate is performed in three steps:

- **Step 1 –** Mixing the Lysis/RT master mix with barcodes in 4 x 384-well plates (referred as **384-RT** plates);
- **Step 2 –** Dispensing the **oil overlay** on top of the cells;
- **Step 3 –** Transferring the Lysis/RT master mix with barcodes from each **384-RT** plate into the **1536-well plate with cells,** directly inside/below the oil layer.

ERCC Spike-in Controls (Optional)

To enable uniform evaluation of sequencing reads across samples and to monitor the effects of sample- and library-preparation steps, we recommend adding External RNA Controls Consortium (ERCC) spike-ins (Thermo Fisher, 4456740) directly to the Lysis/RT master mix.

Instruments

The following steps require automation instruments. We recommend using a combination of microplate dispensers and liquid handling robots for the RT/lysis plate preparation, oil dispensing in 1536-well plates, and transfer of the master in the 1536-well plates. A list of suggested instruments in the reagent table.

Preparation

- Thaw the **CLB**, **RTB** and **OBD** (optional) reagents at room temperature. Mix well and spin down.
- Thaw the four **barcoded oligo-dT** plates at room temperature and spin down (300 x g, 30 sec).
- Keep the **INH** and **RTE** reagents on ice. Spin down.
- Thaw the cells in the 1536-well plate on ice (only when the **384-RT** plates are ready).
- For **ERCC** addition, follow Appendix 1 for step 2.1.1.

Procedure

2.1.1. On ice, in a 15 mL tube, prepare the Lysis/RT master mix +20% as follows:

Reagent	Volume, µL	
	Per well (384 WP)	4 x 384 WP +20%
CLB	0.08	147
INH	0.25	461
RTB	1.5	2765
RTE	0.03	55
Water	1.84	3391
TOTAL	**3.7**	**6820**

2.1.2. Vortex well and keep on ice.

Step 1 – Preparation of 384-RT plates with Lysis/RT master mix and barcodes

2.1.3. Prepare and label four 384-well plates (e.g., 384_P1, 384_P2, *etc*) with the corresponding barcoded oligo-dT set names. These plates will be referred as 384-RT plates.

2.1.4. Dispense 3.7 µL of the Lysis/RT master mix into each well of every **384-RT** plate. Keep the plates on ice.

2.1.5. Open the plates containing the barcoded oligo-dT and transfer 1.3 µL of barcodes to the corresponding **384-RT** plate (384_P1 to 384_P1, 384_P2 to 384_P2, *etc*).

CRITICAL: Avoid cross-contamination between barcoded oligo-dT during transfer.

2.1.6. When all four **384-RT** plates are prepared, seal well the plates and vortex them (500 rpm, 10 sec). Briefly spin down (300 x g, 30 sec). Keep the plates on ice.

NOTE: **384-RT** plates can be kept at 4°C for 4h.

Step 2 – Oil dispensing

2.1.7. Thaw the cells in the 1536-well plate on ice when all four **384-RT** plates are ready.

Facultative: For improved visual inspection, the oil can be dyed blue using the **20x Oil Blue Dye** (OBD) at a 1:20 ratio of dye to oil (silicon or mineral).

2.1.8. Using a nanodispenser or liquid handling robot, dispense 2 µL of oil overlay into each well of the 1536-well plate, directly onto the cell layer.

CRITICAL: Do not use oil in combination with an adhesive seal. If available, cover the plate with its original lid during RT incubation to facilitate handling.

Step 3 – Distribution of the prepared mixes into a 1536-well cell plate

From the **384-RT** plates P1-P4 prepared in step 2.1.8, transfer 3 µL of Lysis/RT/barcodes master mix into the corresponding quadrants of the 1536-well plate containing the cell pellets and oil, following the 384-well layout below (see Figure 2). Dispense the master mix beneath or inside the oil layer, avoiding deposition on the oil surface (see Figure 3). This plate will be referred as the **RT-Plate**.

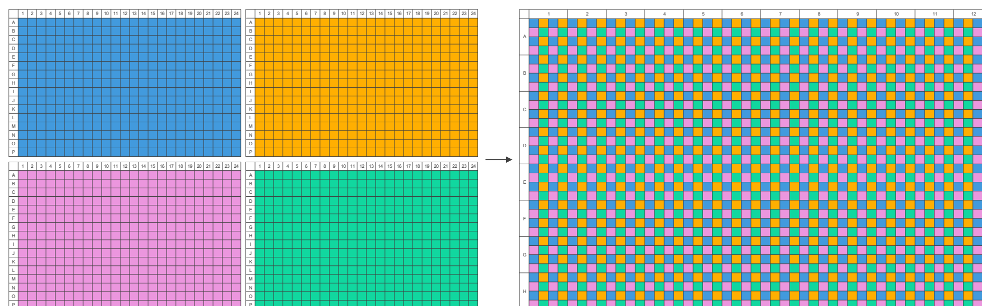


Figure 2 Assembly of 4 x 384-well plates into one 1536-well plate.

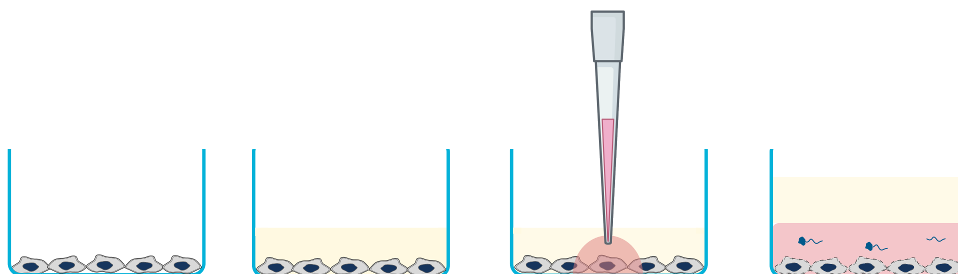


Figure 3 Correct master mix dispensing below the oil layer.

2.1.9. If available, cover the **RT-Plate** with its original lid.

2.1.10. Immediately spin down the plate at 1,000 x g for 30 sec.

2.1.11. Proceed to step 2.2.

2.2. RT incubation

The RT incubation step can be carried out **in a cell incubator or oven**. The oil overlay prevents evaporation, thereby limiting the risk of cross-contamination between samples.

Note: Alternatively, the **RT-Plate** can be sealed and clamped between two aluminum components during RT incubation. This approach requires a custom clamp. Please contact us at *info@alitheagenomics.com* for the details.

Preparation

- Preheat the incubator (or oven) to 37°C.

Procedure

2.2.1. If available, cover the **RT-Plate** with its original lid.

2.2.2. Place the **RT-Plate** in the pre-heated incubator (or oven) and incubate at 37°C for 30 min.

2.2.3. After incubation, cool down the plate on ice for 10 min.

Safe stop: After this step, the plate can be sealed and kept at 4°C overnight.

CRITICAL: To prevent bias from pre-pooling storage, ensure that all RT plates from the same batch for joint analysis are kept under identical conditions, including storage at 4°C.

2.3. Sample pooling

The procedure for pooling samples is at the user's discretion. While manual pipetting is possible, we recommend using VBLOK200 reservoirs (ClickBio) and centrifuging the 1536-well plate upside-down for optimal recovery.

Procedure

2.3.1. Remove the lid from the **RT-Plate**.

2.3.2. Place the plate upside down onto the VBLOK200 reservoir, making sure it is properly seated.

2.3.3. Centrifuge the plate-reservoir at 200 x g for 30 sec to collect all the liquid from the wells into the reservoir.

2.3.4. Gently remove the plate from the reservoir.

2.3.5. Transfer the entire content of the reservoir into a 15 mL tube. The tube should contain ~4.5 mL of pooled cDNA samples and ~3 mL of oil. Note that the total volume may be higher if residual PBS remains from the cell wash.

2.3.6. To separate the cDNA samples from oil, centrifuge the 15 mL tube at 1,000 x g for 30 sec. The cDNA samples represent the lower aqueous phase (**pink layer**), while the oil forms the upper phase (transparent or **blue** if dyed).

2.3.7. Proceed immediately to step 2.4.1 or to step 2.4.19.

2.4. Sample Purification

After sample pooling, the barcoded cDNA can be purified using one of the following methods:

- **Column-based**, e.*g., Zymo Clean & Concentration Kit*.
- **Using SPRI magnetic beads**, *e.g., Beckman*.

Recommended input for second-strand synthesis

Given the enhanced sensitivity of the MERCURIUS™ 1536-Well Plate DRUG-seq workflow, it is not necessary to purify the entire first-strand cDNA prior to second-strand synthesis. A titration experiment can be used to define the optimal pooling volume.

We generally recommend using 2 µL per well (66% of the reaction). For highly active cells or high cell input, pooling 1 µL per well (33% of the reaction) may be sufficient. When cell input falls below the recommended range, we advise purifying the full reaction volume (3 µL per well, 100%) to preserve sensitivity.

Option #1 - cDNA purification using the column-based method

We recommend using a vacuum manifold for purification of the pooled, barcoded single-stranded cDNA samples. A high-capacity Zymo-Spin IIICG column (Zymo, C1006-50-G) and a 25 mL Zymo funnel (Zymo, C1039-25) are required to purify the large volume generated by 1536-sample pooling.

Preparation

- Make sure the Zymo DNA Wash Buffer is reconstituted with ethanol.

Procedure

2.4.1. Transfer a fraction of pooled cDNA samples (33%, 66% or 100%, lower **pink phase** from step 2.3.6) into a new 50 mL tube.

CRITICAL: Minimize oil carryover. Having more than 10% of oil in the transferred volume will interfere with the purification step and negatively impact sequencing data.

2.4.2. Add the corresponding volume of Zymo DNA Binding Buffer to the 50 mL tube from step 2.4.1, following the table below:

Reagent	Volume, mL		
	For 1 µL/well (33%)	For 2 µL/well (66%)	For ~3 µL/well (100%)
Pooled cDNA samples	1.5	3	4
Zymo DNA Binding Buffer	10.5	21	28
TOTAL	**12**	**24**	**32**

2.4.3. Mix gently.

2.4.4. Connect the 25 mL Zymo funnel to a Zymo-Spin IIICG column and place it on a vacuum manifold.

2.4.5. Transfer up to 24 mL of the mix from step 2.4.3 to the 25 mL funnel.

2.4.6. Turn on the vacuum pump.

2.4.7. Allow all the liquid to pass through the membrane. Do not let the membrane over-dry.

2.4.8. If necessary, repeat steps 2.4.5 and 2.4.7 to purify the entirety of the pooled cDNA samples from step 2.4.3.

2.4.9. Remove the funnel.

2.4.10. Add 200 µL of Zymo DNA Wash Buffer (with ethanol) to the column.

2.4.11. Let the wash buffer pass through the membrane.

2.4.12. Repeat steps 2.4.10 and 2.4.11 for a total of two washes.

2.4.13. Turn off the vacuum pump.

2.4.14. Remove the column from the vacuum manifold, place it in a collection tube, and centrifuge at 13,000 x g for 1 min to remove any remaining wash buffer.

2.4.15. Place the column in a new 1.5 mL tube.

2.4.16. Add 38 µL of water onto the column membrane without touching it, and incubate at room temperature for 1 min.

2.4.17. Centrifuge at 13,000 x g for 30 sec for elution. Keep the tube with eluted cDNA on ice.

2.4.18. Immediately proceed to step 2.5.

Option #2 - cDNA purification using the SPRI bead-based method

Purify the pooled, barcoded single-stranded cDNA samples with SPRI magnetic beads at a **1x bead-to-sample ratio**.

CRITICAL: SPRI bead cleanup is currently limited to pooling **1 µL per well** (33% of the reaction volume). This limitation arises because pooling larger volumes would require more than 40 µL of elution buffer which is incompatible with the downstream primer digestion and second-strand synthesis reactions.

To enable pooling at the recommended **2 µL per well**, additional enzyme volumes are required in the subsequent steps. For details on the modified protocol, please contact us at info@alitheagenomics.com.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 6 mL of 80% ethanol.

Procedure

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

2.4.19. Transfer 1.5 mL of pooled cDNA samples (lower **pink phase** from step 2.3.6) into a new 5 mL tube, limiting the transfer of oil to <10%.

2.4.20. Add 1.5 mL of beads.

2.4.21. Mix by pipetting up and down at least 10 times.

2.4.22. Incubate for 5 min at room temperature.

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

2.4.24. To wash the beads, pipette 5 mL of freshly prepared 80% ethanol into the tube.

2.4.25. Incubate for 30 sec.

2.4.26. Carefully remove the ethanol without touching the bead pellet.

2.4.27. Remove the tube from the magnetic stand and let the beads dry for ~5 min. Extend the drying step if necessary, ensuring that all ethanol has fully evaporated.

2.4.28. Resuspend the beads in 40 µL of water.

2.4.29. Incubate for 1 min.

2.4.30. Place the tube on the magnetic stand, wait 5 min, then carefully transfer 35 µL of the supernatant to a new tube. Avoid bead carryovers.

2.4.31. Keep the tube with eluted cDNA on ice.

2.4.32. Immediately proceed to step 2.5.

2.5. Free primer digestion

NOTE: It is recommended to perform non-incorporated primer digestion immediately after the clean-up step.

Preparation

- Thaw the **EXB** reagent at room temperature. Mix well and spin down.
- Keep the **EXO** reagent on ice. Spin down.
- Prepare Program 1_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

2.5.1. On ice, in a new 0.2 mL PCR tube, prepare the EXO reaction mix as follows:

Reagent	Volume, µL
EXB	4
EXO	1
TOTAL	**5**

2.5.2. Transfer 35 µL of purified cDNA samples to the EXO reaction mix.

2.5.3. Mix the EXO reaction by pipetting up and down at least 10 times.

2.5.4. Briefly spin down.

2.5.5. Incubate in the thermocycler Program 1_FPD.

Safe stop: After this step, the tube can be kept at 4°C overnight.

2.6. 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 500 µL of 80% ethanol.
- Thaw the **SSB** reagent at room temperature. Mix well and spin down.
- Keep the **SSE** reagent on ice. Spin down.
- Prepare Program 2_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

2.6.1. On ice, to the tube from step 2.5.5, add the following:

Reagent	Volume, µL
SSB	7
SSE	3
TOTAL	**10**

2.6.2. Mix the SSS reaction by pipetting up and down at least 10 times.

2.6.3. Briefly spin down.

2.6.4. Incubate in the thermocycler Program 2_SSS.

Safe stop: After this step, the tube can be kept at 4°C overnight.

Procedure for cDNA purification with SPRI beads

Purify the double-stranded cDNA with SPRI magnetic beads at a **0.7x bead-to-sample ratio**.

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

2.6.5. To the tube from step 2.6.3, add 35 µL of beads.

2.6.6. Mix by pipetting up and down at least 10 times.

2.6.7. Incubate for 5 min at room temperature.

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

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

2.6.10. Incubate for 30 sec.

2.6.11. Carefully remove the ethanol without touching the bead pellet.

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

2.6.13. Resuspend the beads in 22 µL of water.

2.6.14. Incubate for 1 min.

2.6.15. Place the tube on the magnetic stand, wait 5 min, then carefully transfer 20 µL of the supernatant to a new tube. Avoid bead carryovers.

2.6.16. Measure the concentration with Qubit. Use 2 µL of purified cDNA.

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

2.7. Tagmentation

At this step, the full-length double-stranded cDNA is tagmented with Tn5 transposase preloaded with adapters for library amplification.

NOTE: We recommend using at least 20 ng of cDNA for tagmentation to achieve higher library complexity with fewer PCR amplification cycles. Higher cDNA inputs for tagmentation usually yield better results.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 500 µL of 80% ethanol.
- Thaw the **TAB** reagent at room temperature. Mix well and spin down.
- Keep the **TE** reagent on ice. Spin down.
- Set the thermocycler to 55°C (incubation).

Procedure

2.7.1. On ice, in a new 0.2 mL PCR tube, prepare the appropriate **Tagmentation reaction mix** according to the cDNA input:

Reagent	Volume, µL, for the following cDNA inputs				
	≤ 9 ng	10-14 ng	15-20 ng	50 ng	100 ng
TAB	4	4	4	4	8
TE	1	2	3	7	14
cDNA	X µL	X µL	X µL	X µL	X µL
Water	15 – X	14 – X	13 – X	9 – X	18 – X
TOTAL	**20**	**20**	**20**	**20**	**40**

2.7.2. Thoroughly mix the Tagmentation reaction by pipetting up and down at least 10 times. Keep the tube on ice.

2.7.3. Briefly spin down.

2.7.4. Incubate in the thermocycler for 7 min at 55°C.

2.7.5. Immediately put the tube on ice and proceed to step 2.7.6.

Procedure for tagmented cDNA purification with SPRI beads

Purify the tagmented cDNA using SPRI magnetic beads at a **0.6x bead-to-sample ratio**.

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

2.7.6. To the tube from step 2.7.5, adjust the volume to 50 µL with water.

2.7.7. Add 30 µL of beads.

2.7.8. Mix by pipetting up and down at least 10 times.

2.7.9. Incubate for 5 min at room temperature.

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

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

2.7.12. Incubate for 30 sec.

2.7.13. Carefully remove the ethanol without touching the bead pellet.

2.7.14. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.

2.7.15. Resuspend the beads in 22 µL of water.

2.7.16. Incubate for 1 min.

2.7.17. Place the tube on the magnetic stand, wait 5 min, then carefully transfer 20 µL of the supernatant to a new 0.2 mL PCR tube. Avoid bead carryovers.

2.7.18. Keep the tube on ice.

2.7.19. Proceed immediately to step 2.8.

2.8. Library indexing and amplification

At this step, 5' terminal fragments are amplified using the Unique Dual Indexing (UDI) adapter primers. The kit contains 4 Illumina-compatible primer pairs for generating UDI libraries. The index sequences are indicated in Table 2.

The number of amplification cycles required for library preparation typically ranges from 8 to 15. The precise number may depend on the samples and the cDNA input for tagmentation. To determine the optimal number of cycles, follow the library quantification protocol below.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 500 µL of 80% ethanol.
- Thaw the LAM reagent on ice. Mix well and spin down.
- Thaw the required number of **MQ.UDI Adapters** at room temperature. Mix well and spin down.
- Prepare Program 3_AMP on the thermocycler (set the lid at 100°C):

Step	Temperature, °C	Time	Cycles
Initial extension	72	3 min	1
Initial denaturation	98	1 min	1
Denaturation	98	10 sec	5 + TBD*
Annealing	63	30 sec	
Extension	72	1 min	
Final extension	72	3 min	1
Keep	4	pause	

*The exact number of PCR cycles can be determined following the library quantification protocol below.

Procedure

2.8.1. On ice, to the tube from step 2.7.19, add the following:

Reagent	Volume, µL
LAM	25
MQ.UDI Adapter	5
TOTAL	**30**

2.8.2. Mix the Library Amplification reaction by pipetting up and down at least 10 times.

2.8.3. Briefly spin down.

2.8.4. For cycle number optimization, transfer a 5 µL aliquot of the Library Amplification reaction from step 2.8.3 into a new tube. Use the remaining 45 µL to perform **5 cycles** of library pre-amplification with Program 3_AMP and set the thermocycler to pause at 4°C.

2.8.5. To the tube with 5 µL aliquot of Library Amplification reaction from step 2.8.4, add the following:

Reagent	Volume, µL
SYBR (1:100)*	0.1
LAM	2.5
Water	2.4
TOTAL	**5**

*Prepare a 1:100 dilution with nuclease-free water from the 10,000x stock.

2.8.6. Mix the qPCR reaction by pipetting up and down at least 10 times.

2.8.7. Briefly spin down.

2.8.8. Place the tube in the qPCR machine and run Program 3_AMP for 30 cycles.

2.8.9. Determine the number of PCR cycles from the growth curve in the multicomponent plot, as shown below.

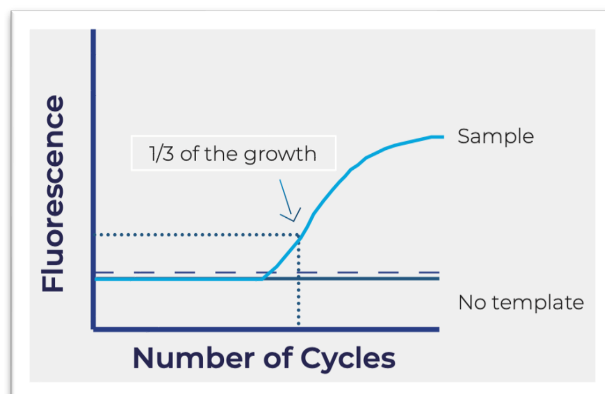


Figure 4 Determination of the optimal number of amplification cycles with qPCR.

2.8.10. After qPCR completion, resume Program 3_AMP. Set the remaining number of amplification cycles based on the qPCR results from step 2.8.9.

For instance, if the optimal number of cycles determined in step 2.8.9 is 10, subtract the 5 cycles already completed. Set the program to run 5 additional cycles.

NOTE: Typically, 20 ng of cDNA input for tagmentation yields at least 40 ng of DNA library after 11 PCR cycles. The table below shows the expected number of PCR cycles and library yield for 20-100 ng of tagmented cDNA.

Tagmented cDNA, ng	# PCR cycles	Expected library yield
20-40	11-12	>40
50-70	10-11	>40
80-100	9-10	>40

Safe stop: After this step, the tube can be kept at 4°C overnight.

Procedure for final library purification with SPRI beads

Purify **twice** the final library with SPRI magnetic beads at a **0.7x bead-to-sample ratio**.

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

2.8.11. Adjust the library volume to 50 µL with water.

2.8.12. Add 35 µL of beads.

2.8.13. Mix by pipetting up and down at least 10 times.

2.8.14. Incubate for 5 min at room temperature.

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

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

2.8.17. Incubate for 30 sec.

2.8.18. Carefully remove the ethanol without touching the bead pellet.

2.8.19. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.

2.8.20. Resuspend the beads in 22 µL of water.

2.8.21. Incubate for 1 min.

2.8.22. Place the tube on the magnetic stand, wait 5 min, then carefully transfer 20 µL of the supernatant to a new tube. Avoid bead carryovers.

2.8.23. Perform the bead clean-up once again by repeating the procedure from step 2.8.11.

Safe stop: At this step, the final library can be safely kept at -20°C for a few weeks.

2.9. 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 library's molarity and dilute the libraries for sequencing. A successful library contains fragments in the range of 300–1,000 bp with a peak at 400–700 bp; see Figure 5 for an example of a standard DRUG-seq library profile.

Libraries with a broader peak (Figure 6) may indicate a light under-tagmentation but still yield high-quality data.

Libraries with a well-defined second sharp peak (>1,000 bp, Figure 7) are also acceptable for sequencing and can produce reliable data.

Importantly, libraries with primer-dimer peaks at 180 bp and between 250–290 bp will likely yield lower-quality sequencing data with a reduced proportion of mapped reads (Figure 8). Therefore, it is strongly recommended

to remove those peaks by performing an additional round of SPRI bead purification of the final library at a 0.7x bead-to-sample ratio (see 2.8.11). According to the final library yield, purify only once or twice.

Under-tagmented libraries with most molecules exceeding 1,000 bp (Figure 9) contain only a fraction of fragments suitable for efficient sequencing. If possible, re-tagmentation of the cDNA is recommended for optimal results (see 2.7).

Procedure

NOTE: 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 quantify its concentration (e.g. Qubit Fluorometer, Thermofisher).
- Use 2 μL of the library to assess the profile (e.g. Fragment Analyzer, Agilent Technology).
- If necessary, re-purify the library (see 2.8.11) to remove the peaks <300 bp.

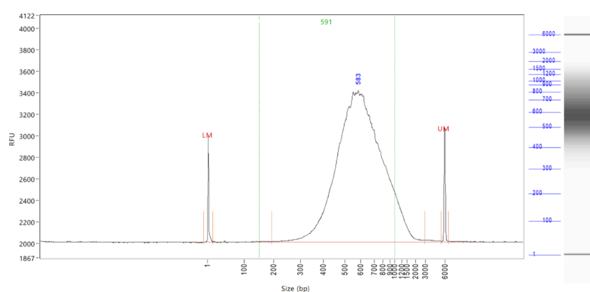


Figure 5 A successful library profile with fragments between 300–1,000 bp.

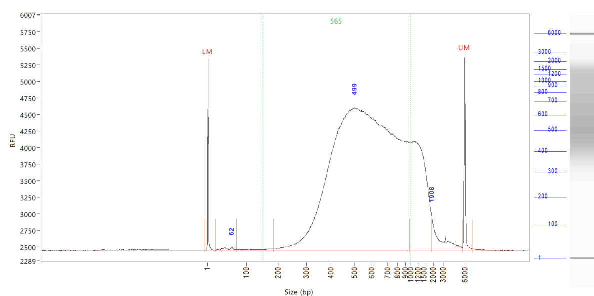


Figure 6 An acceptable library profile with fragments between 300–2,000 bp.

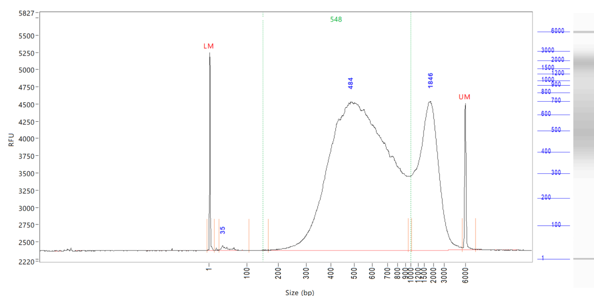


Figure 7 An example of a library profile with a secondary peak >1,000 bp.

Figure 8 An example of an over-tagmented library profile with a peak at 290 bp and an adapter peak at 160 bp.

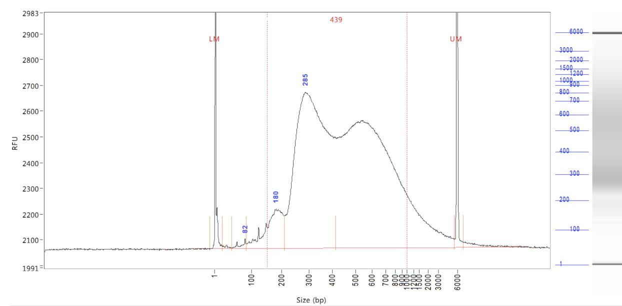
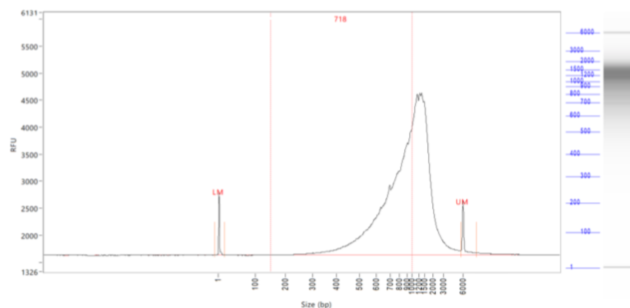


Figure 9 An example of an under-tagmented library profile with a major peak >1,000 bp.



Part 3. LIBRARY SEQUENCING

The libraries prepared with the MERCURIUS™ DRUG-seq kit carry Illumina- and AVITI-compatible adapter sequences. They can be processed on any Illumina instrument (e.g., HiSeq, NextSeq, and NovaSeq) in the Element AVITI System with Adept Workflow. The libraries can also be sequenced on Ultima Genomics instruments, following a conversion step using custom primers.

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 1 for the optimal sequencing structure and Table 2 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. Standard paired-end run setups on Illumina platforms (e.g., 100 PE or 150 PE) are not suitable because the sequencing machine performs poorly on homopolymer sequences.

However, on the AVITI platform, a custom setup with Read1 at 200 bp would be sufficient to read through the oligo-dT sequence and into the cDNA, and Read2 at 100 bp is recommended and compatible.

Read	Length (cycles)		Comment
	for Illumina	for AVITI	
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 1 Sequencing structure of DRUG-seq libraries.

The Unique Dual Indexing (UDI) strategy ensures the highest accuracy in library sequencing and demultiplexing and complies with best practices for Illumina sequencing platforms. UD-indexed libraries have distinct index adapters for i7 and i5 index reads (Table 2).

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 2 UDI adapter sequences.

NOTE: Sequencing depth

1. The recommended sequencing depth is **1 Mio reads per sample**. Deeper sequencing can also be performed to detect 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).

3.1. Sequencing on the Illumina instruments

The loading concentration for the Illumina instruments is indicated in Table 3. Refer to Appendix 2 for the list of Illumina instruments that support forward or reverse workflows.

Instrument	Final loading concentration	PhiX
NextSeq 500/550/550Dx	2.2 pM	1 %
NextSeq 2000, manual denature	85 pM	1%
NextSeq 2000, onboard denature	850 pM	1%
NovaSeq Standard Workflow*	160 pM	1 %
NovaSeq XP Workflow	100 pM	1 %
HiSeq4000	270 pM	1 %

* - adjusted molarity for DRUG-seq libraries sequencing. We recommend diluting the libraries to 0.8 nM before denaturation.

Table 3 Reference loading concentrations for various Illumina instruments.

3.2. Sequencing on the Element AVITI instruments

The MERCURIUS™ DRUG-seq libraries can be sequenced on the Element Biosciences AVITI and AVITI24 Systems using the Cloudbreak AVITI 2x75 High Output sequencing kits (#860-00004).

Libraries must be converted with the Adept PCR-Plus module (#830-00018) for linear loading (Table 4).

Type	Loading molarity, pM	Library starting amount for denaturation, nM	PhiX control	PhiX, %
Cloudbreak (AVITI)	14	1*	PhiX Control Library, Adept	2 %
Cloudbreak (AVITI24)	28	1*	PhiX Control Library, Adept	2 %

* - requires 2 nM of library before conversion

Table 4 Loading concentration for Cloudbreak AVITI and AVITI24 2x75 High Output sequencing kits.

NOTE: Sequencing depth

Please note that the **Cloudbreak AVITI** yields 1 B reads, and **Cloudbreak AVITI24** yields 1.5 B reads. Therefore, for the 1536-sample library, the sequencing depth is limited to a maximum of 0.65 Mio reads/sample (AVITI) or ~1 Mio reads/sample (AVITI24).

3.3. Sequencing on the Ultima Genomics Instruments

MERCURIUS™ DRUG-seq libraries can be sequenced on the Ultima Genomics Systems after a conversion step requiring custom NR-X2 and Tr-X1 primers (Solaris V1).

Contact your local Ultima Genomics vendor or us at info@alitheagenomics.com for more information.

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.

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 a powerful computer with a set of standard bioinformatics tools installed.

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 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.

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

```
> fastqc --outdir $QCdir/ mylibrary_R1.fastq.gz
> fastqc --outdir $QCdir/ mylibrary_R2.fastq.gz
```

NOTE: 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., 2013[1]) 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:

[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

```
> STAR --runMode genomeGenerate --genomeDir /path/to/genomeDir --genomeFastaFiles
  Homo_sapiens.GRCh38.dna.primary_assembly.fa --sjdbGTFfile
  Homo_sapiens.GRCh38.108.gtf --runThreadN 8
```

NOTE: 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.

NOTE: 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 mylibray_R1.

- --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.

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 --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
```

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). The command described above will retain only uniquely mapped reads to ensure high quality results. Consequently, unmapped reads include both truly unmapped fragments as well as reads that map equally well to more than one position (i.e., multi-mapped reads).

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).

Appendix 1. ERCC Spike-In Control

External RNA Controls Consortium (ERCC) Spike-Ins can be added to the Lysis/RT Master mix, used in step 2.1.1.

Preparation

On ice, prepare a 1:100 dilution of the ERCC RNA Spike-In mix in nuclease-free water:

- Add 990 µL of pre-chilled water to 10 µL of ERCC.
- On ice, mix by pipetting up and down at least 10 times and keep.
- Briefly spin down.
- Prepare 65 µL aliquots and store at -20°C.

Procedure for Lysis/RT master mix preparation with ERCC

1. Thaw the **CLB** and **RTB** reagents on ice. Mix well and spin down.
2. Keep the **INH** and **RTE** reagents on ice. Spin down.
3. Thaw the **ERCC (1:100)** on ice. Mix well and spin down.
4. Keep the nuclease-free water on ice.
5. On ice, in a 15 mL tube, prepare the Lysis/RT master mix for 4 x 384-well plates +20% as follows:

Reagent	Volume, µL	
	Per well (384 WP)	4 x 384 WP +20%
CLB	0.08	147
INH	0.25	461
RTB	1.5	2765
RTE	0.03	55
ERCC (1:100)	0.02	37
Water	1.82	3355
TOTAL	**3.7**	**6820**

NOTE: The ERCC content should represent ~1% of the total mapped reads maximum.

6. Vortex well and keep on ice.
7. Follow the main protocol for the Lysis/RT master mix distribution (step 2.1.3).

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

ALITHEA GENOMICS

info@alitheagenomics.com

www.alitheagenomics.com

Alithea Genomics SA
Batiment Serine
Rue de la Corniche 8, level -2
1066 Epalinges
Switzerland
Tel: +41 78 830 3139

Alithea Genomics Inc.
321 Ballenger Center Drive
Suite 125 C/O Alithea Genomics Inc.
Frederick, MD 21703
USA
Tel: +1 240 656 3142