ALITHEA GENOMICS

MERCURIUS™

Total DRUG-seq Library Preparation Kit for 96, 384, and 1’536 Samples

PN 10705, 10706, 11661, 11662

User Guide

August 2025
(Early-Access)

Related Products

Kit name	Kit PN	Modules	Module PN
Mercurius™ Total DRUG-seq Library Preparation 96 kit	10705	Barcoded Oligo-dT Adapters Module 96 samples	10513
		Total DRUG-seq Library Preparation and UDI Module 96 samples	10603
Mercurius™ Total DRUG-seq Library Preparation 384 kit	10706	Barcoded Oligo-dT Adapters Module 384 samples	10515
		Total DRUG-seq Library Preparation and UDI Module 384 samples	10604
Mercurius™ Total DRUG-seq Library Preparation 4x 96 kit	11661	Barcoded Oligo-dT Adapters Module 4x 96 samples	10513
		Total DRUG-seq Library Preparation and UDI Module 4x 96 samples	10607
Mercurius™ Total DRUG-seq Library Preparation 4x 384 kit	11662	Barcoded Oligo-dT Adapters Module 4x 384 samples	10513
		Total DRUG-seq Library Preparation and UDI Module 4x 384 samples	10608

Table of Contents

Kit Components

Reagents supplied

Barcoded Oligo-dT Adapters Set V5 Module

Component Name	Label	Amount provided per kit				Storage
		96 samples (PN 10705)	384 samples (PN 10706)	4x 96 samples (PN 11661)	4x 384 samples (PN 11662)	
Plate with 96 barcoded oligo-dT primers, set V5C (PN 10400)	96 V5C OdT	1 plate		4 plates	-	-20°C
Plate with 384 barcoded oligo-dT primers, set V5C (PN 10401)	384 V5C OdT	-	1 plate	-	4 plates	-20°C
Aluminium Seal	-	2 pcs	2 pcs	8 pcs	8 pcs	-20°C/RT

Total DRUG-seq Library Preparation and UDI Module

Component Name	Label	Cap colour	Volume, µL				Storage
			96 samples (PN 10603)	384 samples (PN 10604)	4x 96 samples (PN 10607)	4x 384 samples (PN 10608)	
Fragmentation Buffer	FAB	magenta	290	650	2x 650	4x 650	-20°C
rRNA Depletion Mix	RDP	magenta	150	310	2x 310	4x 310	-20°C
Repair/RT Enzyme	RAE	magenta	140	280	2x 280	4x 280	-20°C
Repair/RT Solution	RAX	magenta	930	1900	2x 1900	4x 1900	-20°C
Exonuclease I Enzyme	F-EXO	purple	10	10	10	10	-20°C
Exonuclease Buffer	F-EXB	purple	20	20	20	20	-20°C
Second Strand Enzyme Mix	SSE M	orange	20	20	20	20	-20°C
Second Strand Buffer FL	SSB FL	orange	30	30	30	30	-20°C
Adapter Ligation Buffer	ALB	blue	100	100	100	100	-20°C
BRB-compatible Adapter	BRB.AD	blue	10	10	10	10	-20°C
Library Amplification Mix FL	LAM FL	green	200	200	200	200	-20°C
UDI Adapter Mix 1	MF.UDI.1	transparent	10	10	10	10	-20°C
UDI Adapter Mix 2	MF.UDI.2	transparent	10	10	10	10	-20°C
UDI Adapter Mix 3	MF.UDI.3	transparent	10	10	10	10	-20°C
UDI Adapter Mix 4	MF.UDI.4	transparent	10	10	10	10	-20°C
Cell Lysis Buffer	CLB	yellow	710	1500	2x 1500	4x 1500	-20°C
RNase Inhibitor	INH	yellow	175	375	2x 375	4x 375	-20°C

Additional required reagents and equipment (supplied by the user)

Plasticware	Manufacturer	Product number
0.2 mL 8-Strip Non-Flex PCR Tubes	Starlab	I1402-3700
Disposable Pipetting Reservoir 25 mL polystyrene	Integra	4382
Disposable Pipetting Reservoir 150 mL polystyrene	Integra	6318

Reagents	Manufacturer	Product number
SPRI AMPure Beads or CleanNA Beads	Beckman Coulter CleanNA	A63881 CNGS0050D
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
High Sensitivity NGS Fragment Analysis Kit	Agilent	DNF-474
Ethanol, 200 proof	-	-
Nuclease-free water	Thermofisher	A57775
DPBS, Cell culture grade	Gibco	10010023
ERCC RNA Spike-In Mix (recommended)	Thermo Fisher	4456740

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
Magnetic stand for 5 mL tubes	Permagen	
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
VIAFLO instrument (optional)	Integra	6001
VIAFLO 96 channel pipetting head, 0.5-12.5 µL (optional)	Integra	6101

Protocol Overview and Timing

The MERCURIUS™ Total DRUG-seq kits allow the time and cost-efficient preparation of Illumina-compatible RNA sequencing libraries for up to 1’536 samples. The kits include a mild cell lysis buffer to prepare crude cell lysates.

The MERCURIUS™ Total DRUG-seq kits are optimized for efficient rRNA depletion in **human, mouse, and rat** cell samples. For use with other species, please contact us at *info@alitheagenomics.com* to discuss compatibility.

This protocol is designed for a total RNA population consisting of various RNA species, including both coding and non-coding transcripts. It involves prior rRNA depletion, fragmentation, subsequent modifications, and library preparation using DRUG-seq technology.

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	10705	96WP	96	96	4
384-sample	10706	384WP	384	384	4
4x 96-sample	11661	96WP	96	384	4
4x 384-sample	11662	384WP	384	1’536	4

Every kit contains barcoded MERCURIUS™ Oligo-dT primers, designed to tag RNA from cell lysate samples with individual barcodes during the first-strand synthesis reaction. This enables the pooling of the resulting cDNA samples from each experimental group into a single tube, facilitating streamlined sequencing library preparation.

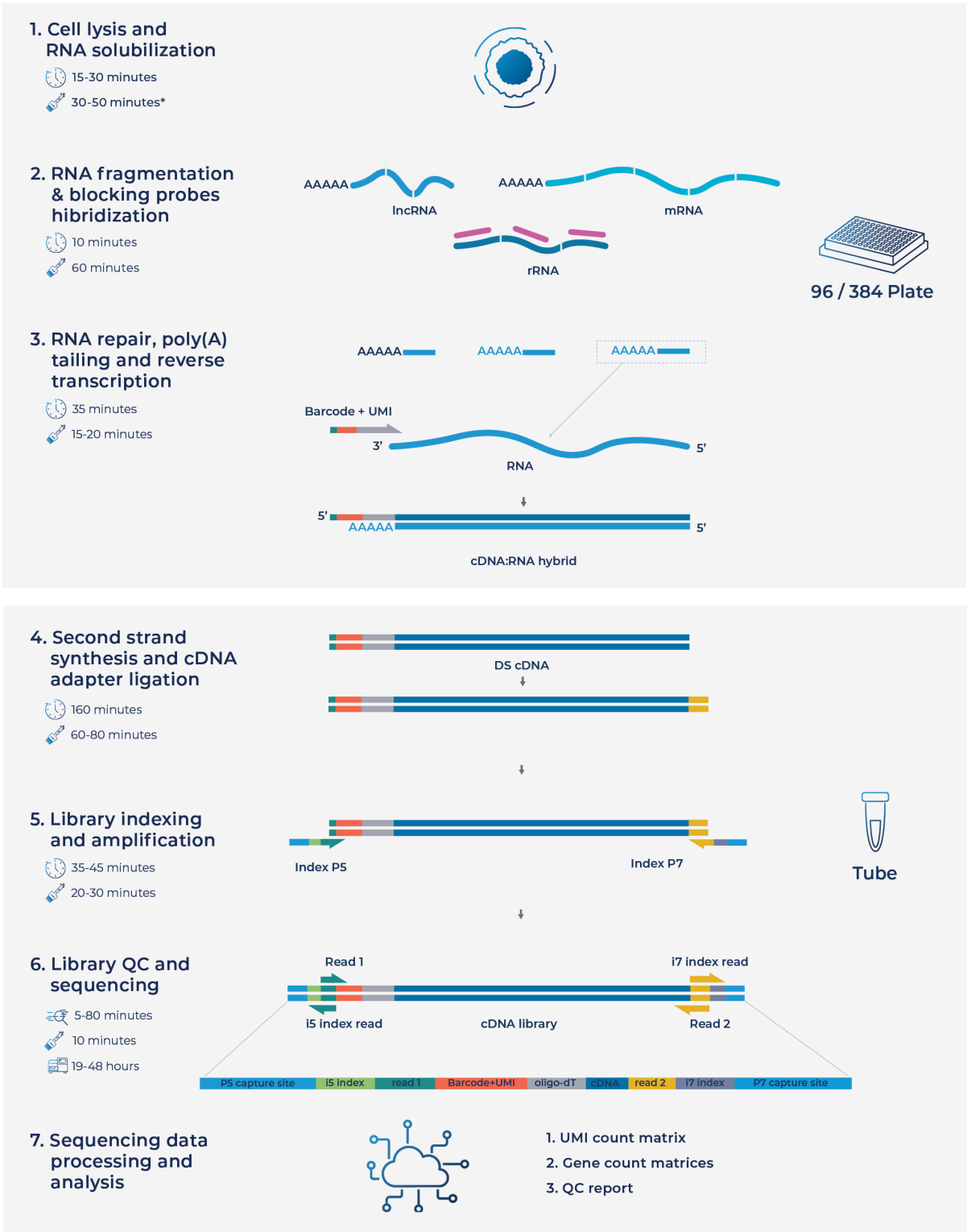
The DRUG-seq technology can be used to generate high-quality sequencing data starting with 2’000 – 25’000 mammalian cells per well. Notably, the kit can be used to pool any number of samples up to 384, with two considerations:

- The total cell number per pool should be at least 45’000.
- 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. Every adapter can be used to prepare an individual library. Libraries with different UDI adapters can be pooled and sequenced in a single flow cell.

Figure 1 provides an estimation of the time required to accomplish each step of the protocol.

Protocol Workflow



Overall time

Incubation time: 4h20-6h.
Hands-on time: 3h15-4h10.
QC time: 5min-1h20 (depending on the instrument used: Qubit or Fragment Analyzer).
Sequencing time: 19-48 hours (depending on instrument and sequencing run settings).

Figure 1 Schematic illustration of the protocol workflow

Part 1. PREPARATION OF CELL LYSATE SAMPLES

1.1. Essential considerations for input cells

- The recommended input range of cells is 5'000-25'000 cells/well for **96WP** and 2'000-10'000 cells/well for **384WP** (on the day of sample preparation). It is advisable to use a cell number closer to the lower end of the recommended range rather than the upper limit.
- **NOTE:** Do not use more than 25'000 cells per well, as it will result in a high rRNA content.
- Cells should be seeded a few days in advance for optimal results.
- To obtain the best result prior to the experiment, ensure that cell viability is >70% (e.g., trypan blue, propidium iodide).
- 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 the potential effect of the treatment on the cell number during the experiment.
- To ensure an even distribution of reads after sequencing, the amount of starting material must be as uniform as possible. We suggest automating cell seeding instruments or verifying cell counts twice.

1.2. ERCC Spike-in Controls (Optional)

To ensure the uniformity of sequencing reads across samples and to assess the impact of sample and library preparation steps on this, we recommend adding External RNA Controls Consortium (ERCC) Spike-Ins to the lysis buffer (Thermo Fisher, 4456740). Please refer to Appendix 1 for detailed information before proceeding with the lysis step.

1.3. Cell pellet preparation

At this step, plated cells are washed with DPBS and frozen at -80°C for at least 5 min. If possible, snap-freeze the plate with dry ice or liquid nitrogen beforehand.

NOTE: The freezing step is required to achieve a higher exon mapping.

Procedure for the preparation of adherent cells

1.3.1. Seed the cells in a flat-bottom **96WP** or **384WP** at the density that will enable harvesting:
- **96WP:** 5'000-25'000 cells per well
- **384WP:** 2'000-10'000 cells per well

1.3.2. Gently aspirate culture media from the plate and wash cells by adding the following:
- **96WP:** 80-120 µL DPBS in each well
- **384WP:** 30-50 µL DPBS in each well

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

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

1.3.5. Proceed to step 1.4.1 for cell lysis.

Procedure for the preparation of suspension cells

1.3.6. Seed the cells in a flat-bottom or U-shaped **96WP** or **384WP** at the density that will enable harvesting:
- **96WP:** 5'000-25'000 cells per well
- **384WP:** 2'000-10'000 cells per well

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

1.3.8. Gently aspirate culture media from the plate and wash cells by adding the following:
- **96WP:** 80-120 µL DPBS in each well
- **384WP:** 30-50 µL DPBS in each well

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

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

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

1.3.12. Proceed to step 1.4.1 for cell lysis.

1.4. Cell lysate preparation

At this step, frozen cells are lysed directly in a cell plate by adding 1x Cell Lysis Buffer to the wells. The lysates can be used directly for the RNA fragmentation step.

Preparation

- Thaw the CLB and INH tubes on ice.
- Mix well and briefly spin down before use
- Prepare a working solution of 1x Cell Lysis Buffer with RNase Inhibitor:

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

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

Procedure for cell lysis

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

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

1.4.3. Incubate the plate at room temperature for 15 min. Do not exceed or shorten the incubation time. Put the plate back on ice after the RT incubation.

1.4.4. Pipette the lysate **into a new PCR plate** (96- or 384-well) for a future experiment. Ensure proper labeling for easy identification during subsequent use.

1.4.5. Seal the plate with the aluminum seal provided and briefly spin it down.

1.4.6. The lysates can be used directly for RNA fragmentation (see step 2.1) or safely stored at -80°C for a few weeks.

NOTE: If multiple plates must be processed, perform the procedure with each plate individually, one at a time, to avoid keeping plates at room temperature for an extended period.

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 or plate.

2.1. RNA fragmentation

The RNA molecules are fragmented during this step, while rRNA species are selectively inhibited to prevent their downstream amplification.

Preparation

- Thaw the cell lysate samples on ice.
- Thaw the **FAB** and **RDP** reagents and mix well before use.
- Prepare Program 1_Fragmentation on the thermocycler (set the lid at 100°C) and preheat it to 94°C.

Step	Temperature, °C	Time
Incubation	94	3 min
Incubation	75	2 min
Incubation	70	2 min
Incubation	65	2 min
Incubation	60	2 min
Incubation	55	2 min
Incubation	37	2 min
Incubation	25	2 min
Keep	4	pause

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

Procedure

2.1.1. Prepare the Fragmentation Master mix (+10%) as follows:

Reagent	96WP, µL		384WP, µL	
	Per well	96 wells +10%	Per well	384 wells +10%
FAB	**2.6**	**286**	**1.5**	**642**
RDP	**1.3**	**143**	**0.7**	**300**
TOTAL	**3.9**	**429**	**2.2**	**942**

2.1.2. Using a multichannel pipette or robot, transfer the following volume of cell lysates to the new 96- or 384-well PCR plate:

- **96WP:** 9 µL per well
- **384WP:** 5 µL per well

2.1.3. Using a multichannel pipette or robot, transfer the following volumes of Fragmentation Master mix to each well:

- **96WP:** 3.9 µL per well
- **384WP:** 2.2 µL per well

2.1.4. Carefully seal the plate, vortex it well, and briefly spin it in the centrifuge.

2.1.5. Incubate in a thermocycler Program 1_Fragmentation.

2.1.6. Keep the plate on ice.

2.1.7. Proceed immediately to step 2.2.

2.2. RNA repair, poly (A) tailing, and reverse transcription

At this step, fragmented RNA molecules are repaired, poly-adenylated, and reverse-transcribed using the barcoded oligo-dT primers provided in a 96- or 384-well plate. Subsequently, all the barcoded samples can be pooled into 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 to enhance the visualization of reaction preparation and pooling.

Despite variations in appearance caused by the drying process, wells may exhibit traces of dried dye ranging from dispersed to solid dots on the bottom. The following addition of the reagents will enable the visualization of red color, confirming the presence of the oligos in all wells.

Preparation

- Thaw all tubes on ice and mix well before use.
- Spin down the plate with barcoded oligo-dT to ensure that the pellet is at the bottom of the wells.
- Prepare Program 2_Repair/RT on the thermocycler (set the lid at 90°C) and pre-heat it to 37°C:

Step	Temperature, °C	Time
Incubation	37	30 min
Inactivation	75	5 min
Keep	4	pause

Procedure

2.2.1. Keep the plate with RNA and oligo-dT on ice.

2.2.2. Using a multichannel pipette or robot, pipette the cell lysate with fragmented RNA to the oligo-dT plate, with corresponding samples to the barcodes:

- **96WP:** 10 µL per well
- **384WP:** 5 µL per well

2.2.3. Depending on the number of samples, prepare the following Repair/RT Master mix (+10%) as follows:

Reagent	96WP, µL		384WP, µL	
	Per well	96 wells +10%	Per well	384 wells +10%
RAX	8.75	927.5	4.4	1855
RAE	1.25	132.5	0.6	265
TOTAL	10.0	1060	5	2120

2.2.4. Using a multichannel pipette or robot, pipette the Repairing/RT Master mix to each well:

- **96WP:** 10 µL per well
- **384WP:** 5 µL per well

2.2.5. Carefully seal the plate, vortex well, and briefly spin it in the centrifuge.

2.2.6. Incubate in a thermocycler Program 2_Repair/RT.

2.2.7. Proceed to step 2.3.

Safe stop: After this step, the RT plate can be kept at 4°C overnight or at -20°C for a few days.

2.3. Sample pooling and purification

After pooling, the barcoded RT samples can be purified using either column-based Zymo Clean & Concentration Kit (Zymo, D4014) or SPRI magnetic beads (Beckman, A63881). While both approaches produce similar results, Zymo columns should be favoured when working with large pools (96 or 384 samples). Depending on the availability of 3rd party reagents and instruments, the corresponding method should be applied.

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

The procedure of cDNA pre-cleaning and purification using the column-based method

After the cDNA from each well is pooled in a reservoir, mix it with a 7x volume of DNA binding buffer (Zymo, D4004-1-L). We strongly recommend using a vacuum manifold to purify cDNA, as it helps avoid

column membrane damage that can occur with 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.

Plate format	Pipetting strategy	Zymo-Spin column type	First strand cDNA		DNA Binding Buffer		TOTAL	
			Per well, µL	Per plate, mL	Per well, µL	Per plate, mL	Per well, µL	Per plate, mL
96WP	multichannel pipette or pipetting robot	I (#D4014)	10	**0.96**	70	**6.72**	80	**7.68**
384WP	pipetting robot	IIICG (C1006-50-G)	5	**1.92**	35	**13.44**	40	**15.36**

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

Preparation

- Ensure that Zymo DNA Wash buffer has Ethanol added.

Procedure

2.3.1. According to Table 1, use a multichannel pipette or pipetting robot to transfer RT volume (10 µL for **96WP**, 5 µL for **384WP**) of each sample into a specific reservoir (25 mL or 100 mL).

2.3.2. Mix the pool well and transfer it to a Falcon tube with a pipette.

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

2.3.4. Collect the cleaned flow-through in a new Falcon tube. Repeat step 2.3.3 until the entire pool has passed through the Zymo column. Discard the column.

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

2.3.6. Connect the 25 mL funnel (Zymo, C1039-25) to a Zymo column suitable for the purification volume (Table 1) and place it on a vacuum manifold.

2.3.7. Gently mix the cDNA in the binding buffer mixture and transfer it to a 25 mL funnel using a pipetboy.

2.3.8. Turn on the vacuum pump and let the liquid pass through the column.

2.3.9. Transfer any remaining volume to the funnel. Do not let the membrane over-dry.

2.3.10. After the entire pool mix has passed through the column, add 200 µL of DNA Wash buffer (with Ethanol added) directly to the membrane of the column.

2.3.11. Repeat step 2.3.10 once the wash buffer has passed through the filter.

2.3.12. Remove the column from the vacuum manifold, place it in a 1.5 mL tube, and centrifuge for 1 min to remove leftovers from the washing buffer.

2.3.13. Depending on the Zymo-Spin column type used, perform the following:

- For the type **I** column used with ≤96 samples (**96WP**), add 18 µL of water to the column matrix and incubate for 1 min.
- For the type **IIICG** column used with 384 samples (**384WP**), add 25 µL of water to the column matrix and incubate for 1 min. We recommend repeating this step with the eluted material to ensure that all cDNA is recovered from the column matrix.

2.3.14. Transfer the column into a new labeled 1.5 mL tube and centrifuge for 30 sec.

2.3.15. Immediately proceed to step 2.4.

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

Perform cDNA purification using SPRI magnetic beads with a 1:0.9 ratio of cDNA pool to beads slurry. The purification of large volumes (i.e., 1 mL from **96WP** and 4 mL from **384WP**) may require a few 1.5 mL tubes and a corresponding magnetic stand (e.g., Permagen, MSR06).

If the pool's volume is higher than 500 µL, split it equally among the required number of 1.5-2 mL tubes and add the identical volume of beads (i.e., a pool of 1 mL split into 2 tubes with 500 µL per tube, and add 900 µL of beads per tube).

NOTE: Please use SPRI-beads only if the solution is clear and has no visible debris.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 5 mL of 80% ethanol.

Procedure

2.3.16. Using a multichannel pipette or robot, pool the RT samples in the reservoir (Integra, 4382 or 6318).
   - **96WP:** 3 µL per well
   - **384WP:** 2 µL per well

2.3.17. **CRITICAL:** Pooling more volume may result in lower gene detection and exon mapping accuracy.

2.3.18. Transfer the collected pool into a 2 mL or 5 mL tube, depending on the pooled volume. The final volume will be almost two times higher due to the addition of the beads.

2.3.19. Add pre-warmed beads in a 1:0.9 ratio (i.e., for 768 µL of pooled samples, add 681 µL of beads slurry), and mix by pipetting up and down ten times.

2.3.20. Incubate for 5 min at room temperature.

2.3.21. If necessary, split the volume into several tubes.

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

2.3.23. To wash the beads, pipette 1000 µL of freshly prepared 80% ethanol into the tube.

2.3.24. Incubate for 30 sec.

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

2.3.26. Repeat step 2.3.23 for a total of two washes.

2.3.27. Remove the tube from the magnetic stand and let the beads dry for 1-5 min, or until all visible ethanol has fully evaporated.

2.3.28. Resuspend the beads in 18 µL of water and incubate for 1 min.

2.3.29. Place the tubes on the magnetic stand, wait 5 min, and carefully transfer 17 µL of supernatant to a new tube to avoid bead carry-over.

2.3.30. Immediately proceed to step 2.4

NOTE: If the RT pool was split into several tubes at step 2.3.21, resuspend the beads in the **first tube** in 20 µL of water. Keep other tubes closed to avoid over-drying of the beads. Transfer the obtained elution to the next tube and resuspend the beads. Repeat this step for every tube.

2.4. Free primer digestion

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

Preparation

- Label 0.2 mL PCR tubes corresponding to the number of pools prepared.
- Thaw the **F-EXB** reagent at room temperature.
- Keep the **F-EXO** reagent constantly on ice.
- Prepare Program 3_FPD on the thermocycler (set the lid at 90°C), pre-heat the thermocycler to 37°C:

Step	Temperature, °C	Time
Incubation	37	6 min
Incubation	80	1 min
Keep	4	pause

Procedure

2.4.1. Transfer 17 µL of the eluate from step 2.3.14 (if column was used) or 2.3.29 (if SPRI beads) into a new labeled 0.2 mL PCR tube.

2.4.2. Prepare the F-EXO reaction Master mix as follows (with 10% excess):

Reagent	Per reaction + 10%, µL
F-EXB	2.2
F-EXO	1.1
TOTAL	**3.3**

2.4.3. According to the table, transfer 3 µL of the F-EXO reaction mix into each PCR tube with purified cDNA.

2.4.4. Mix by pipetting up and down 5 times.

2.4.5. Briefly spin down in the bench-top centrifuge.

2.4.6. Incubate in the thermocycler Program 3_FPD.

2.4.7. Proceed immediately to step 2.5.1.

2.5. Second-strand synthesis and DNA repair

At this step, double-stranded full-length cDNA is generated and repaired.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 5 mL of 80% ethanol.
- Thaw the **SSB FL** reagent at room temperature and mix well before use.
- Keep the **SSE M** reagent constantly on ice.
- Prepare Program 4_SSS on the thermocycler (set the lid at 90°C), pre-heat the thermocycler to 42°C:

Step	Temperature, °C	Time
Incubation	42	10 min
Incubation	62	20 min
Keep	4	pause

Procedure

2.5.1. Prepare the SSS FL Master mix for the second-strand synthesis as follows (with 10% excess):

Reagent	Per reaction, µL
SSB FL	5
SSE M	3.2
Water	11.8
TOTAL	20

2.5.2. Transfer **20 µL** of the SSS reaction mix to the tube from step 2.4.7 and mix well by pipetting up and down 5 times.

2.5.3. Briefly spin down in the bench-top centrifuge.

2.5.4. Incubate in a thermocycler Program 4_SSS.

2.5.5. Proceed immediately to step 2.5.6.

cDNA clean-up with SPRI beads

Perform the cDNA purification using SPRI magnetic beads with a 1:0.9 ratio of cDNA pool to beads slurry.

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

2.5.6. Complement the volume of the SSS sample to 50 µL with nuclease-free water.

2.5.7. Add 45 µL of beads (50 µL sample) and mix by pipetting 10 times.

2.5.8. Incubate for 5 min at room temperature.

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

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

2.5.11. Incubate for 30 sec.

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

2.5.13. Repeat step 2.5.10 for a total of two washes.

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

2.5.15. Resuspend the beads in 21 µL of water.

2.5.16. Incubate for 2 min.

2.5.17. Place tubes on the magnetic stand, wait 2 min, and carefully remove 20 µL of supernatant into a new tube to avoid bead carryover.

2.5.18. Use 2 µL to measure the concentration with Qubit (recommended).

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

2.6. cDNA adaptor ligation

At this step, the BRB-compatible adaptor is ligated to the cDNA fragments to facilitate the following amplification of the library with Unique Dual Indexing (UDI) primers.

NOTE: We highly recommend using only 50% of the cDNA obtained in 2.5.17, not exceeding 200 ng. Using the entire cDNA yield or more than 200 ng does not improve the reaction and should be avoided. Additionally, using less than 25% of the cDNA may result in low library complexity.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 mins.
- Prepare 5 mL of 80% ethanol.
- Thaw the ALB and BRB.AD reagents on ice and mix well before use.
- Prepare Program 5_ADL on the thermocycler (set the lid at 90°C):

Step	Temperature, °C	Time
Incubation	20	15 min
Keep	4	pause

Procedure

2.6.1. Complement every sample from step 2.5.17 to **18.75 µL** with water (if necessary). Then pipette **BRB.AD** and then add **ALB** **one by one** as indicated in the table below. It is **not recommended** to prepare a master mix for all samples.

Reagent	Per reaction, µL
cDNA	18.75
BRB.AD	1.25
ALB	20
TOTAL	**40**

2.6.2. **CRITICAL**: Mix well by pipetting up and down 10 times. This is essential to ensure a sufficient ligation. The presence of small bubbles will not interfere with performance.

2.6.3. Briefly spin down the tube in the bench-top centrifuge.

2.6.4. Incubate in the thermocycler Program 5_ADL.

2.6.5. **CRITICAL**: Proceed immediately to step 2.6.6

cDNA clean-up with SPRI beads

Perform the cDNA purification using SPRI magnetic beads with a 1:0.9 ratio of cDNA pool to beads slurry (i.e., 50 µL of cDNA plus 45 µL of bead slurry).

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

2.6.6. Complement the final reaction volume to 50 µL with water.

2.6.7. Add 45 µL of beads and mix by pipetting 10 times.

2.6.8. Incubate for 5 min at room temperature.

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

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

2.6.11. Incubate for 30 sec.

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

2.6.13. Repeat step 2.6.10 for a total of two washes.

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

2.6.15. Resuspend the beads in 21 µL of water.

2.6.16. Incubate for 1 min.

2.6.17. Place the tubes on the magnetic stand, wait 5 min, and carefully remove 20 µL of supernatant into a new tube to avoid bead carryover.

2.6.18. Proceed immediately to step 2.7.10.

2.7. Library indexing and amplification

At this step, the library undergoes amplification using Unique Dual Indexing (UDI) primers. The kit contains four Illumina-compatible primer pairs, allowing for the generation of up to four UDI libraries. The index sequences are indicated in Table 3.

The number of amplification cycles required for library preparation typically ranges from 8 to 19, depending on the number of cells and sample quantity.

It is strongly recommended that the final library bead clean-up be performed twice to remove primer dimer fragments.

Preparation

- Pre-warm the SPRI beads at room temperature for ~30 min.
- Prepare 10 mL of 80% ethanol.
- Thaw the **LAM FL** reagents on ice and mix well before use.
- Thaw the required number of **MF.UDI Adapters** at room temperature; briefly spin them before use.
- Prepare Program 6_AMP (set the lid at 105°C) on the thermocycler:

Step	Temperature, °C	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	8-19*
Annealing and Extension	65	75 sec	
Final extension	65	5 min	1
Keep	4	pause	

*The required number of PCR cycles can be estimated based on the amount of cDNA used for adapter ligation (see below). Follow the guidelines below.

Library amplification reaction setup

2.7.1. Prepare the PCR amplification reaction as follows:

Reagent	Per reaction, µL
LAM FL	25
MF.UDI Adapter	5
Ligated cDNA	20
TOTAL	**50**

2.7.2. Pipette up and down 5 times.

2.7.3. Put the tube in the ice and do one of the following:

- **Highly recommended**: Determine the optimal number of amplification cycles using real-time PCR (follow steps 2.7.4 - 2.7.10); or
- Start Program 6_AMP and set the required number of PCR cycles based on the amount of cDNA used for adaptor ligation (step 2.5.18) (less preferable).

cDNA used for library prep, ng	Number of cycles
175	8
80	11
20	12
10	13
5	14
4	15
1.5	16
0.9*	17
0.5	19

*If the amount of cDNA is less than 0.9 ng, we strongly recommend doing qPCR.

2.7.4. For **cycle number optimization**, transfer a 5 µL aliquot from the PCR reaction, prepared in 2.7.3 into a separate PCR tube. Use the remaining 45 µL to perform 5 cycles of library preamplification using Program 6_AMP and set the PCR machine to pause at 4°C afterward.

2.7.5. Use 5 µL from the PCR reaction (kept separately in 2.7.4) to set up a qPCR reaction in a suitable PCR tube or plate as follows:

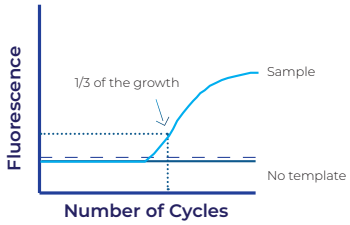
Reagent	Per reaction, µL
Library	5
SYBR 100x*	0.1
LAM FL	2.5
Water	2.4
TOTAL	**10**

*Prepare 100x dilution with nuclease-free water from 10'000x stock

2.7.6. Run the qPCR reaction using Program 6_AMP for 30 cycles.

2.7.7. Analyze the multicomponent plot (as shown in Figure 2) and determine the optimal number of amplification cycles based on the growth curve.

Figure 2 Determination of the additional number of amplification cycles with qPCR



2.7.8. While the qPCR is running, the main sample will complete 5 preamplification cycles and remain paused at 4°C.

2.7.9. In the same PCR machine, resume the reaction by setting the remaining number of amplification cycles, depending on qPCR results (determined in 2.7.7). *For instance, if the optimal number of cycles determined in step 2.7.7 is 10, subtract the 5 cycles already completed. Set the machine to run 5 additional cycles*.

2.7.10. After PCR is complete, proceed to library clean-up (2.7.11)

Indexed cDNA library clean-up with SPRI beads

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

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

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

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

2.7.13. Incubate for 5 min at room temperature.

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

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

2.7.16. Incubate for 30 sec.

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

2.7.18. Repeat step 2.7.15 for a total of two washes.

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

2.7.20. Resuspend the beads in 21 µL of water.

2.7.21. Incubate for 1 min.

2.7.22. Place tubes on the magnetic stand, wait 5 min, and carefully remove 20 µL of supernatant into a new tube to avoid bead carry-over.

2.7.23. Perform the bead clean-up once again by repeating the procedure from step 2.7.11.

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

2.8. Library quality control

Pooled library quality control

Before sequencing, the libraries should be subjected to fragment analysis using a Fragment Analyzer, Bioanalyzer, or TapeStation, and quantification using Qubit. This information is required to assess the library molarity and prepare the appropriate library dilution for sequencing. A successful library contains fragments in the 300 – 700 bp range with a peak at 400-500 bp; see Figure 3 for an example of a standard Total DRUG-seq library profile.

Sometimes library can show a few sharp peaks, which mainly represent specific or highly abundant transcripts (see Figure 4). This pattern typically has no impact on the library quality overall.

Notably, the bead clean-up must be performed twice to remove primer dimer fragments, which can likely result in lower-quality sequencing data with reduced mappable reads. Additionally, the presence of high-molecular-weight fragments can impact library quality (see Figure 5). Therefore, it is strongly recommended that those peaks be removed by performing an additional round of SPRI bead purification with the 0.7x ratio (see step 2.7.11).

Library quantification can also be performed unbiasedly using qPCR with standard Illumina library quantification kits (e.g., 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 2.7.11 – 2.7.22 to remove the peaks <300 bp.

Figure 3 A successful library profile with fragments between 300-700 bp

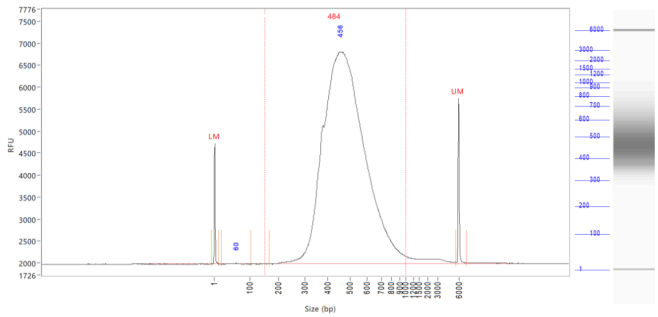


Figure 4 An example of a library profile demonstrating some sharp peaks.

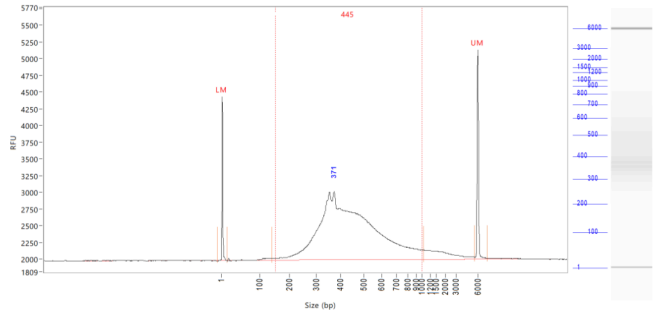
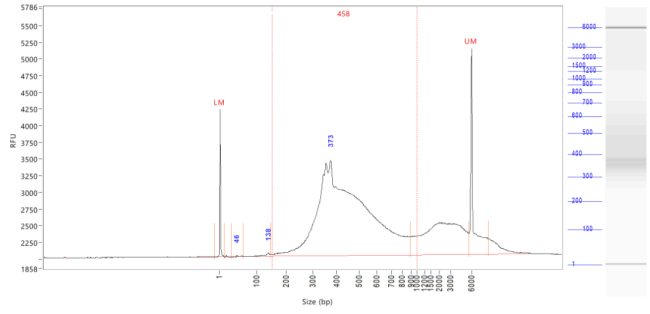


Figure 5 An example of a library profile with some large fragments (>1kbp)



Assessing uniformity of read distribution across the samples

For projects involving highly heterogeneous samples, it is recommended to validate the uniformity of read coverage across the samples by shallow library sequencing (see step 2.3). This approach ensures that every sample will obtain enough reads required for the analysis. DRUG-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 are sufficient to assess the fraction of reads attributed to every sample.

Part 3. LIBRARY SEQUENCING

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

The MERCURIUS™ Total 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 2 for the optimal sequencing structure and Table 3 for the list of i5 and i7 index sequences.

Given the DRUG-seq library structure, the optimal number of cycles for Read 1 is 28 (and 29 for AVITI). The following cycles, 29-60, will cover the homopolymer sequence, which may result in a significant drop of Q30 values reflecting sequencing quality. Standard paired-end run setups on Illumina platforms (e.g., 100 PE or 150 PE) are not suitable due to the low performance of the sequencing machine for 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 2 Sequencing structure of Total DRUG-seq libraries

The Unique Dual Indexing (UDI) strategy ensures the highest library sequencing and demultiplexing accuracy and complies with the best practices for Illumina sequencing platforms. UD-indexed libraries have distinct index adapters for i7 and i5 index reads (Table 3).

Name	Type	i7 index sequence	i5 index sequence Forward Workflow	i5 index sequence Reverse Workflow
MF.UDI.1	UDI (i7/i5)	GCTTGTCA	AGGCGAAG	CTTCGCCT
MF.UDI.2	UDI (i7/i5)	CAAGCTAG	TAATCTTA	TAAGATTA
MF.UDI.3	UDI (i7/i5)	AGTTCAGG	CAGGACGT	ACGTCCTG
MF.UDI.4	UDI (i7/i5)	GACCTGAA	GTACTGAC	GTCAGTAC

Table 3 UDI adapter sequences

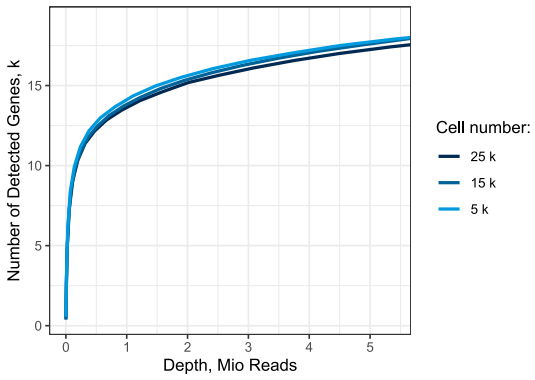
NOTE: Sequencing depth

1. The recommended sequencing depth is 2-5 Mio reads per sample (see Figure 6 and Figure 7). We recommend getting 15-20 Mio reads per sample to detect very lowly expressed genes.

2. If only one library is sequenced in a flow cell, the Index reads can be skipped.

3. The library's loading molarity depends on the sequencing instrument (see 3.1 and 3.2) and should be discussed with the sequencing facility or an experienced person.

Figure 6 Number of detected genes as a function of the sequencing depth for different numbers of seeded 293 cells per well (4 to 8 wells pooled)



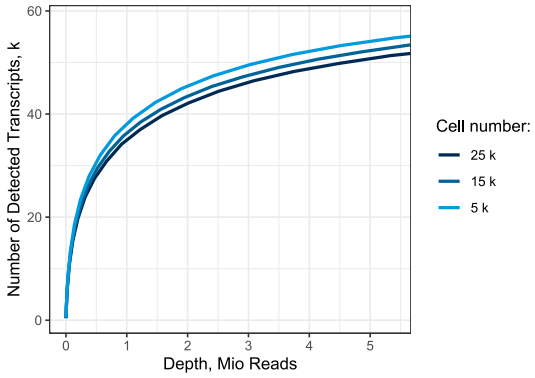


Figure 7 Number of detected transcripts as a function of the sequencing depth for different numbers of seeded 293 cells per well (4 to 8 wells pooled)

3.1. Sequencing on the Illumina instruments

Table 4 indicates the loading concentration for the Illumina instruments. For the list of Illumina instruments with forward or reverse workflow, please refer to Appendix 2.

Instrument	Final loading concentration	PhiX
MiSeq	20 pM	1 %
iSeq	100 pM	1 %
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 a prior dilution of the libraries to 0.8 nM before denaturation.

Table 4 Reference loading concentrations for various Illumina instruments

3.2. Sequencing on the Element AVITI instrument

For the most optimal results, the MERCURIUS™ Total DRUG-seq libraries can be sequenced with the Element Biosciences AVITI System using 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 5).

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

* - requires 4nM of library before conversion

Table 5 Loading concentration for Cloudbreak AVITI 2x75 High Output sequencing kit

NOTE: Sequencing depth

Please note that the **Cloudbreak AVITI 2x75 High Output sequencing** yields 1 Mio reads. Therefore, for the 384-sample library, the sequence depths would be limited to a maximum of 2.5 Mio reads/sample.

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 transcriptome-level count matrices from raw FASTQ files.

To obtain the data ready for analysis, the user needs to demultiplex the sequencing reads by sample barcodes and perform transcriptome pseudo-alignment and quantification, as well as perform an alignment of the sequencing reads to the genome, and perform the gene/UMI read count generation.

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

4.1. Required software

Tool	Description	Version
fastQC	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.	>0.11.9
STARsolo from STAR	Software for read alignment on reference genome (Dobin et al., 2013). It can be downloaded from Github (https://github.com/alexdobin/STAR)	2.7.9
FastReadCounter	Software for counting genome-aligned reads for genomic features	>1.1
Picard	Collections of command-line utilities to manipulate with BAM files. Used in this user guide for demultiplexing of BAM files.Java version 8 or higher	>2.17.8
Samtools	Collections of command-line utilities to manipulate with BAM files. Used in this user guide for sorting and indexing of BAM files.	>1.9
fqtk	Demultiplexes pooled FASTQs based on inline barcodes	0.3.1
kallisto	Pseudoaligns reads to a transcriptome for quantification	0.48.0
R	Data processing and visualization. Required packages installed: data.table and Matrix. Please refer to the official R documentation for instructions on installing R packages: https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages	>3.0.0

IMPORTANT: Please refer to the official webpages of each software tool mentioned to review system requirements before installation.

4.2. Merging FASTQ files from individual lanes and/or libraries (Optional)

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
```

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
```

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 (i.e., L001, L002, L003, L004, not L002, L001 ...) to avoid issues when demultiplexing the samples.

4.3. Sequencing data quality check

Perform basic quality control checks on raw sequencing reads to assess read quality, GC content, duplication levels, adapter contamination, and other key metrics before downstream processing. Run FastqQC 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.

Input:

- Raw FASTQ files from sequencing (e.g., mylibrary_R1.fastq.gz, mylibrary_R2.fastq.gz)

Output:

- HTML and .zip QC reports in the specified output directory
  (e.g., fastqc_out_dir/mylibrary_R1_fastqc.html, fastqc_out_dir/mylibrary_R2_fastqc.html)

Command line:

```
> fastqc --outdir fastqc_out_dir/ mylibrary_R1.fastq.gz
> fastqc --outdir fastqc_out_dir/ 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.

4.4. Pseudo-alignment and transcriptome quantification

This section explains how to process Multiplexed Total DRUG-seq sequencing data to obtain transcript-level expression estimates. This approach uses **pseudo-alignment,** which provides a much faster and resource-efficient method for determining transcript abundances from sequencing reads.

4.4.1. Demultiplex FASTQ files

BRB-seq libraries are generated by pooling barcoded samples prior to sequencing. The first step is to **split this pooled data into individual samples** using the inline barcodes embedded in Read 1 or Read 2. `fqtk` — a tool designed for barcode and UMI extraction and demultiplexing.

Input:

- Raw paired-end FASTQ files (e.g., mylibrary_R1.fastq.gz, mylibrary_R2.fastq.gz)
- Barcode reference file (tab-separated: sample_id <TAB> barcode sequence)
- CIGAR strings that specify barcode/UMI layout in reads

Output:

- Demultiplexed paired FASTQ files per sample (e.g., sample1_R1.fq.gz, sample1_R2.fq.gz)

Command line:

```
>  fqtk demux -i mylibrary_R1.fastq.gz mylibrary_R2.fastq.gz \
>             -r 14B14M 90T \
>             -o dmx_fastq \
>             -s barcode_ref.txt
```

NOTE: 14B14M means a 14-nt cell barcode followed by a 14-nt UMI in Read 1, and 90T is a fixed 90-nt length of genomic Read 2.

4.4.2. Build transcriptome index

Transcript quantification tools like Kallisto require a **prebuilt index** of the transcriptome. This index maps k-mers to known transcripts and is used to **efficiently pseudoalign reads.**

Input:

- cDNA FASTA file (transcript sequences)
- (Optional) Corresponding GTF annotation (not needed for index, but useful for downstream analysis)

Output:

- Kallisto index file (e.g., Homo_sapiens.GRCh38.idx)

Command Example:

```
>  kallisto index -i Homo_sapiens.GRCh38.idx
   Homo_sapiens.GRCh38.cdna.all.fa
```

How to Obtain the FASTA File:

Download the cDNA reference FASTA from Ensembl:

```
> wget ftp://ftp.ensembl.org/pub/release-
104/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh38.cdna.all.fa.gz

> gunzip Homo_sapiens.GRCh38.cdna.all.fa.gz
```

4.4.3. Quantify transcript abundance

This step estimates **transcript-level expression** for each sample using Kallisto's fast and pseudoalignment algorithm. Outputs include abundance estimates, effective lengths, TPM expression values, and bootstrapping results.

Input:

- Sample-specific FASTQ files from demultiplexing step 4.4.1.
- Kallisto index file

Output:

- Abundance files for each sample (e.g., abundance.tsv, run_info.json)

Command Example:

```
>  kallisto quant -i Homo_sapiens.GRCh38.idx  -o quant/sample1 -l 550 -
   s 150 -b 5 sample1_R1.fq.gz sample1_R1.fq.gz -t 30
```

NOTE:

`-l` and `-s` indicate estimated fragment length and estimated standard deviation of fragment length in a **single-end** run — only one input FASTQ should be listed.

`-b 5` adds bootstrapping for variance estimation.

`-t 30` uses 30 threads for performance.

4.4.4. Assemble transcriptome counts

To facilitate downstream analysis (e.g., differential expression, clustering), you can **combine quantification results across all samples** into a single matrix of estimated transcript counts.

Input:

- Quant/ folder containing all abundance.tsv files

Output:

- Count matrix (e.g., transcript_count_matrix.csv)

Example R script:

```
    # R script for collecting transcriptome data

     cbind_vec2matrix <- function(list_vectors, row_names,
col_names){
         df_ = data.frame(do.call(cbind, list_vectors))
         colnames(df_) = col_names
         rownames(df_) = row_names
         return(df_)
     }

     list_dirs = list.dirs("quant/", recursive = F)
     est_counts_l = list()
     tmp_l = list()
     sample_name_l = list()
     for(i in 1:length(list_dirs)){
         this_dir = list_dirs[[i]]
         abundance_file = paste0(this_dir, '/abundance.tsv')
         if(file.exists(abundance_file)){
             abundance_tab = read.table(abundance_file, header=T)
             est_counts_l[[i]] = abundance_tab[['est_counts']]
             tmp_l[[i]] = abundance_tab[['tpm']]
             sample_name_l[[i]] = gsub("^\\\\/", "", gsub("$in_dir",
'', this_dir))
         }
     }

     df_counts = cbind_vec2matrix(est_counts_l, row_names =
abundance_tab[["target_id"]], col_names = unlist(sample_name_l))
     df_tpm    = cbind_vec2matrix(tmp_l,        row_names =
abundance_tab[["target_id"]], col_names = unlist(sample_name_l))

     lib_name = gsub("_kallisto_out","","$in_dir")
     write.csv(df_counts, paste0(lib_name,".counts.txt"), quote=F)
     write.csv(df_tpm, paste0(lib_name,".tpm.counts.txt"), quote=F)
```

NOTE: Ensure that you are running the script from the same directory that contains the `quant/` folder.

4.5. Alignment and gene quantification

While transcriptome pseudo-alignment provides fast quantification at the transcript level, full-length data also benefits from traditional genome alignment to support additional analyses—such as quality control metrics, gene body coverage, and gene-level quantification. This section explains how to align sequencing reads to a reference genome using the **STAR** aligner, and how to use the aligned reads to generate gene-level counts.

4.5.1. Preparing the reference genome

Before aligning sequencing reads, a genome index must be generated from a reference genome and a corresponding gene annotation file. This is a one-time step per genome version. STAR uses this index to efficiently map RNA-seq reads, including for multiplexed libraries (e.g., DRUG-seq), using its STARsolo mode, which also generates count matrices.

Input:

- Reference genome FASTA file (e.g., Homo_sapiens.GRCh38.dna.primary assembly.fa)
- Gene annotation GTF file (e.g., Homo_sapiens.GRCh38.108.gtf).

Output:

- A directory containing STAR genome index files, including:
  - SA
  - Genome
  - sjdbList.fromGTF.out.tab
  - and other supporting files

Download reference files (Ensembl, Human GRCh38):

```
>  # Download and decompress reference genome (FASTA)
>  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
>
>  # Download and decompress annotation file (GTF)
>  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
```

Recommendation:

- Use the `primary_assembly` FASTA file when available (avoid 'sm' or 'rm' tags).
- For GTF, select the version without `chr` prefixes or `abinitio` tags for compatibility.

Index Generation Command:

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

Parameter details:

- `--genomeDir`: Output directory where STAR will write the index files.
- `--genomeFastaFiles`: Full path to the decompressed FASTA file.
- `--sjdbGTFfile`: Full path to the decompressed GTF annotation file.
- `--runThreadN`: Number of CPU cores to use for parallel processing.

NOTE:

- Adjust `--runThreadN` based on available CPU cores (higher values = faster indexing).
- STAR requires **~32–40 GB RAM** depending on genome size — ensure sufficient memory.
- The generated index can be reused for all analyses with the same genome/annotation.

4.5.2. Aligning to the reference genome and generation of count matrices

After the reference genome index is prepared, sequencing reads (FASTQ files) can be aligned to the genome using STAR. In this context, the STARsolo mode is used, which not only performs the alignment but also generates gene and UMI (unique molecular identifier) count matrices directly. This step is tailored for multiplexed libraries such as those used in the Total DRUG-seq full-length protocol.

Input:

- Paired-end FASTQ files (e.g., mylibrary_R1.fastq.gz and mylibrary_R2.fastq.gz)
- STAR genome index (from Step 1.5.1)
- Barcode whitelist file (text file with one barcode sequence per line). Example:

```
TACGTTATTCCGAA
AACAGGATAACTCC
ACTCAGGCACCTCC
ACGAGCAGATGCAG
```

Output:

- Aligned BAM files (sorted by coordinate)
- Gene and UMI count matrices in Matrix Market (MTX) format
- Demultiplexing statistics (Solo.out/Barcodes.stats)
- Alignment summary (Log.final.out)

Command example:

```
>  STAR --runMode alignReads \
>       --outSAMmapqUnique 60 \
>       --runThreadN 8 \
>       --outSAMunmapped Within \
>       --soloStrand Forward \
>       --quantMode GeneCounts \
>       --outBAMsortingThreadN 8 \
>       --genomeDir /path/to/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 /path/to/bamdir/libraryname/ \
>       --readFilesIn mylibrary_R2.fastq.gz mylibrary_R1.fastq.gz
```

Parameter details:

- `--genomeDir`: Path to the STAR genome index directory (e.g., `/path/to/genomeDir`).
- `--readFilesIn`: Order is important. R2 (genomic reads) first, R1 (barcodes and UMI) second.
- `--soloCBwhitelist`: Text file with one barcode per line; use the version appropriate for your kit (e.g., Mercurius V5).
- `--soloCBstart, --soloCBlen`: Barcode position and length in R1 (e.g., start = 1, length = 14).
- `--soloUMIstart, --soloUMIlen`: UMI starts after barcode (e.g., start = 15, length = 14).
- `--soloUMIdedup: NoDedup`: produces read count matrix; `1MM_Directional`: adds UMI count matrix
- `--outFileNamePrefix`: Output prefix directory, e.g., `/path/to/bamdir/libraryname/`

Additional NOTE:

- **Output folder**: BAM files and matrices will be saved in `/path/to/bamdir/libraryname/`
- **Demultiplexing statistics**: Found in `/path/to/bamdir/libraryname/Solo.out/Barcodes.stats`
- **Alignment summary**: Found in `/path/to/bamdir/libraryname/Log.final.out`

IMPORTANT: The order of the fastq files provided in the script is important. The first fastq must contain genomic information, while the second the barcode and UMI content. Thus, files should be provided for STARsolo in the following order: `--readFilesIn` mylibray_R2 mylibrary_R1.

4.5.3. 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:

```
> library(data.table)
> library(Matrix)
> matrix_dir <- "/path/to/bamdir/libraryname/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.

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

Once you obtain a multiplexed BAM file from STARsolo, you can use **FastReadCounter** to extract gene-level read counts for each sample based on the associated barcodes, ready for downstream differential expression analysis.

Input:

- **BAM file**: A multiplexed alignment file
  (e.g., /path/to/bamdir/libraryname/Aligned.sortedByCoord.out.bam)
- **GTF file**: Genome annotation file (e.g., Homo_sapiens.GRCh38.108.gtf)
- **Barcode file**: A tab-separated file listing the expected barcodes. Example of expected format:

```
TACGTTATTCCGAA sample_1
AACAGGATAACTCC sample_2
ACTCAGGCACCTCC sample_3
ACGAGCAGATGCAG sample_4
```

Output:

- **Gene read count matrix**: Located in the specified output folder (counts/)
  - One file per barcode/sample
  - Summary statistics per sample
  - Global matrix across all barcoded samples

Command line example:

```
> #!/bin/bash
>
> gtf_file=homo_sapience.gtf   # Genome annotation file in GTF format
> output_folder=counts/        # Output directory for final results
> bam_path=/path/to/bamdir/libraryname/Aligned.sortedByCoord.out.bam    #
  Path to directory and prefix of the BAM file
> barcode_file=barcode_frc.txt # File listing expected barcodes
>
> FastReadCounter-1.0.jar \
>   --bam ${bam_path} \
>   --gtf ${gtf_file} \
>   --umi-dedup none \
>   --barcodeFile ${barcode_file} \
>   -o ${output_folder}
```

NOTE:

- The BAM file should be sorted and indexed if required by your downstream tools.
- The barcode file must match the barcodes used during BRB-seq library preparation.
- If you do **not have the barcode sequences**, please contact info@alitheagenomics.com, including the barcode set name and the Product Number (PN) of your barcode module

4.5.5. 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 (e.g., GEO or ArrayExpress), or for running a bulk RNA-seq analysis pipeline. This can be done using the **Picard** tool.

Input:

- Aligned.sortedByCoord.out.bam — multiplexed BAM file generated by STARsolo
- barcode_brb.txt — tab-separated file with two columns: sample_id and barcode sequence. Example:

```
Sample1 TACGTTATTCCGAA
Sample2 AACAGGATAACTCC
Sample3 ACTCAGGCACCTCC
Sample4 ACGAGCAGATGCAG
```

Output:

- A BAM file for each sample in the specified output directory.

Command line example:

```
>  #!/bin/bash
>
>  demultiplexed_bam_out_dir=/path/to/output_bams
>  input_bam=/path/to/bamdir/libraryname/Aligned.sortedByCoord.out.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"
```

Appendix 1. ERCC Spike-In Control

The current protocol includes the addition of External RNA Controls Consortium (ERCC) Spike-Ins to the lysate buffer.

Prepare a 1:100 dilution of the ERCC RNA Spike-In mix in nuclease-free water. Mix 990 µL of pre-chilled water with 10 µL of ERCC. Pipette well and aliquot the dilution into 50 µL aliquots, keeping them at -20°C.

The working solution of 1x Cell Lysis Buffer with ERCC Spike-In controls consists of the following:

Reagent	96WP, µL		384WP, µL	
	Per well	96 wells +10%	Per well	384 wells +10%
CLB	6.6	700	3.3	1400
INH	1.6	170	0.8	340
Water	11.6	1229	5.8	2457.6
ERCC* (1:100)	0.2	21	0.1	42.4
TOTAL	20	2120	10	4240

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

Cell Lysis Buffer (CLB) preparation with ERCC

1. Thaw the CLB and ERCC tubes on ice and avoid their long-term storage.
2. Keep the nuclease-free water on ice to maintain a cold temperature.
3. Spin down all the tubes before pipetting.
4. First, add the water to a 15 mL Falcon tube, then the CLB, INH, and the ERCC (in this particular order).
5. Pipette the prepared mix a few times and briefly spin the tube. Keep it on ice until further use.
6. Follow the main protocol for cell lysis procedure (step 1.4.1)

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