



MERCURIUS™

Total Plant BRB-seq Library Preparation Kit for 96 and 384 Samples

PN 10815, 11815

User Guide

July 2026

Related Products

Kit name	Kit PN	Modules	Module PN
Mercurius™ Total Plant BRB-seq Library Preparation 96 kit	10815	Barcoded Oligo-dT Adapters Module 96 samples	10400
		Total Plant BRB-seq Library Preparation and UDI Module 96 samples	10481
Mercurius™ Total Plant BRB-seq Library Preparation 4x 96 kit	11815	Barcoded Oligo-dT Adapters Module 4x 96 samples	10400
		Total Plant BRB-seq Library Preparation and UDI Module 4x 96 samples	10482

Table of Contents

TABLE OF CONTENTS	2
KIT COMPONENTS	3
Reagents supplied	3
Additional required reagents and equipment (supplied by the user).....	4
PROTOCOL OVERVIEW AND TIMING	5
PROTOCOL WORKFLOW	6
PART 1. PREPARATION OF RNA SAMPLES	7
PART 2. LIBRARY PREPARATION PROTOCOL	8
2.1. RNA fragmentation.....	8
2.2. RNA repair, poly (A) tailing, and reverse transcription	9
2.3. Sample pooling and purification	10
2.4. Free primer digestion	12
2.5. Second-strand synthesis and DNA repair.....	12
2.6. cDNA adaptor ligation	13
2.7. Library indexing and amplification.....	15
2.8. Library quality control.....	17
PART 3. LIBRARY SEQUENCING	19
3.1. Sequencing on the Illumina instruments	20
3.2. Sequencing on the Element AVITI instruments	20
PART 4. SEQUENCING DATA PROCESSING	21
4.1. Required software	21
4.2. Merging FASTQ files from individual lanes and/or libraries (Optional)	21
4.3. Sequencing data quality check	22
4.4. Pseudo-alignment and transcriptome quantification	22
4.4.1. Demultiplex FASTQ files	22
4.4.2. Build transcriptome index	23
4.4.3. Quantify transcript abundance.....	23
4.4.4. Assemble transcriptome counts.....	24
4.5. Alignment and gene quantification	24
4.5.1. Preparing the reference genome	25
4.5.2. Aligning to the reference genome and generation of count matrices	25
4.5.3. Generating the count matrix from .mtx file	27
4.5.4. Generating the read count matrix with per-sample stats (Optional)	27
4.5.5. Demultiplexing bam files (Optional)	28
APPENDIX 1. COMPATIBLE ILLUMINA INSTRUMENTS	29

Kit Components

Reagents supplied

Barcoded Oligo-dT Adapters Set V5 Module

Component Name	Label	Amount provided per kit		Storage
		96 samples (PN 10815)	4x 96 samples (PN 11815)	
Plate with 96 barcoded oligo-dT primers, set V5C (PN 10400)	96 V5C OdT	1 plate	4 plates	-20°C
Aluminium Seal	-	1 pc	4 pcs	-20°C/RT

Total Plant BRB-seq Library Preparation and UDI Module

Component Name	Label	Cap colour	Volume, µL		Storage
			96 samples (PN 10481)	4x 96 samples (PN 10482)	
Fragmentation Buffer	FAB	magenta	290	4x 290	-20°C
rRNA Depletion Plant Probes	RPP	magenta	115	4x 115	-20°C
Repair/RT Enzyme	RAE	magenta	140	4x 140	-20°C
Repair/RT Solution	RAX	magenta	930	4x 930	-20°C
Exonuclease I Enzyme	F-EXO	purple	10	10	-20°C
Exonuclease Buffer	F-EXB	purple	20	20	-20°C
Second Strand Enzyme Mix	SSE M	orange	20	20	-20°C
Second Strand Buffer FL	SSB FL	orange	30	30	-20°C
Adapter Ligation Buffer	ALB	blue	100	100	-20°C
BRB-compatible Adapter	BRB.AD	blue	10	10	-20°C
Library Amplification Mix FL	LAM FL	green	200	200	-20°C
UDI Adapter Mix 1	MF.UDI.1	transparent	10	10	-20°C
UDI Adapter Mix 2	MF.UDI.2	transparent	10	10	-20°C
UDI Adapter Mix 3	MF.UDI.3	transparent	10	10	-20°C
UDI Adapter Mix 4	MF.UDI.4	transparent	10	10	-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	A63881
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
High Sensitivity NGS Fragment Analysis Kit	Agilent	DNF-474
Ethanol, 200 proof	-	-
Nuclease-free water	Thermo Fisher	A57775

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 Plant BRB-seq Kits** enable rapid and cost-effective preparation of Illumina-compatible RNA-sequencing libraries for up to 384 samples in parallel. The kits are optimized for efficient rRNA depletion from RNA extracted from leaves, stems, and roots. For other plant tissues or sample types, please contact us at info@alitheagenomics.com.

This protocol is designed for total RNA samples containing a broad range of RNA species (see [Part 1](#)), including both coding and non-coding transcripts. The workflow includes RNA fragmentation, rRNA depletion, RNA reverse transcription, and library preparation using Total Plant BRB-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	10815	96WP	96	96	4
4x 96-sample	11815	96WP	96	384	4

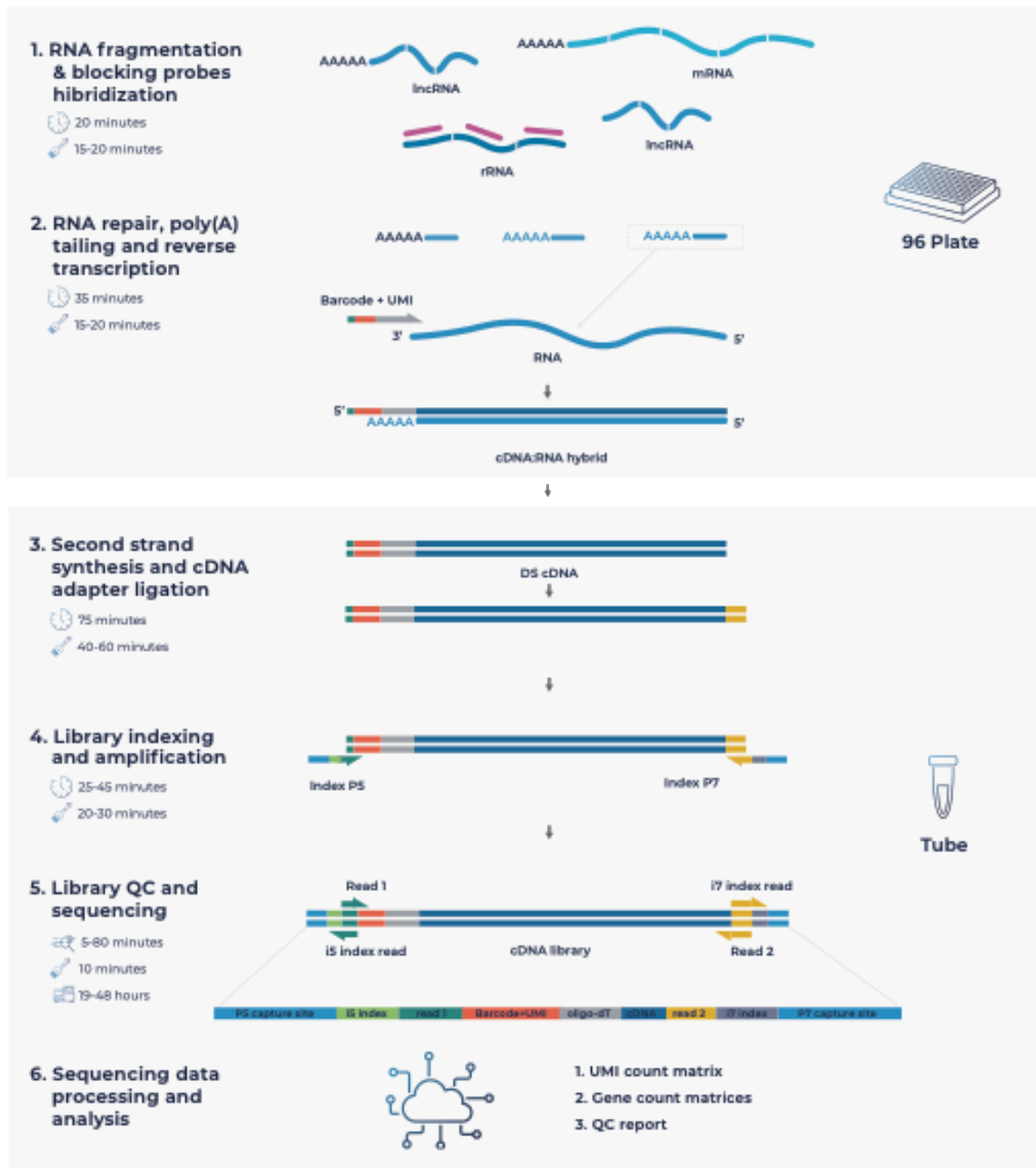
Every kit contains barcoded MERCURIUS™ Oligo-dT primers, designed to tag RNA samples with individual barcodes during the first-strand synthesis reaction. This enables the pooling of the resulting cDNA samples from each experimental group into a single tube, facilitating streamlined sequencing library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

The Total Plant BRB-seq technology can be used to generate high-quality sequencing data starting from 100 pg to 100 ng of RNA per well. Notably, the kit can be used to pool up to 96 samples, but pooling fewer than 8 samples may result in low-complexity reads during sequencing, decreasing 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 estimate of the time required for each step of the protocol, assuming that RNA samples have been previously quality-checked and diluted.

Protocol Workflow



Overall time

- Incubation time: 2h35-2h55.
- Hands-on time: 1h30-2h10.
- 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 RNA SAMPLES

Guidelines for RNA samples

MERCURIUS™ Total Plant BRB-seq is based on early sample multiplexing; therefore, it is crucial to ensure that the quantity, purity, and quality of RNA are uniform for all samples before initiating library preparation. Individual sample quality checks and adjustments will not be possible after pooling.

Quantity

It is recommended to use 100 pg to 100 ng of total RNA input per well. Library construction should be initiated with 10–100 ng of pooled cDNA, generated by combining material from at least eight RT wells. Higher RNA input in the RT reaction generally results in increased library complexity following sequencing.

Purity

RNA samples extracted using TRIzol™, phenol, or chloroform compounds are prone to residual contamination with organic solvents that may inhibit the reverse transcription reaction. This can result in low cDNA library yields, loss of sequencing reads for a fraction of pooled samples, and uneven read distributions. To ensure the high purity of RNA, assess the 260/230 ratio of at least a few samples from the same RNA isolation batch using a spectrophotometer (i.e., Nanodrop). The 260/230 ratio values should be >1.5.

Uniformity

To ensure an even distribution of reads after sequencing, the RNA amount, integrity (RIN), and 260/230 values of the starting RNA samples must be as uniform as possible, i.e., 50 ng ($\pm 10\%$) of starting material, with preferably a RIN > 7 for every sample. To obtain such uniform amounts, we therefore recommend:

- Measuring the RNA concentration of all samples using a dye-based method (e.g., Qubit Quant-iT or RiboGreen for a large number of samples).
- Diluting the samples to obtain the same RNA concentration in all wells ($\pm 10\%$), ensuring a minimum of 8 μL /well, which will be used in the protocol.
- Re-measuring the RNA concentration of all samples to confirm uniformity across samples.
- Ensuring that the 260/230 ratios are >1.5 and similar across the samples.
- Ensuring that the RIN values are similar across the samples, preferably > 7. Highly degraded RNA (RIN <2) are not supported with this protocol.

Validated Plant Species

MERCURIUS™ Total Plant BRB-seq is currently validated on the following species:

Name	Species
Thale Cress	<i>Arabidopsis thaliana</i>
Maize	<i>Zea mays</i>
Lettuce	<i>Actuca sativa</i>
Melon	<i>Cucumis melon</i>
Pepper	<i>Capsicum annuum</i>
Tomato	<i>Solanum lycopersicum</i>
Watermelon	<i>Citrullus lanatus</i>
Spinach	<i>Spinacia oleracea</i>

The rRNA depletion system is designed for a large panel of plant species and is likely compatible with a broader range of plants (e.g. rice, potato, soybean, wheat, ...).

Contact info@alitheagenomics.com for technical support.

Note: Many plant genomes remain incompletely annotated. As a result, annotation gaps or errors may decrease mapping efficiency and produce abnormal gene-body coverage patterns in total RNA-seq data.

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 RNA samples on ice.
- Thaw the **FAB** and **RPP** 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 manipulations with RNA samples and the RT enzyme should be performed in an RNase-free environment, using RNase-free consumables and filter tips, on ice, and with gloves.

Procedure

- 2.1.1. Prepare the Fragmentation Master mix (with 15% extra) as follows:

Reagent	96WP, μ L	
	Per well	96 wells + 15%
FAB	2	220.8
RPP	1	110.4
TOTAL	3	331.2

- 2.1.2. Using a multichannel pipette or robot, transfer 7.5 μ L of the RNA samples to a 96-well PCR plate on ice.
- 2.1.3. Using a multichannel pipette or robot, add 3 μ L of Fragmentation Master mix into each well, containing the RNA sample.
- 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-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 solely 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 colour, 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 plates with RNA or oligo-dT on ice.
- 2.2.2. Using a multichannel pipette or robot, add 10 µL of fragmented RNA per well into the oligo-dT plate, preserving the original well order to maintain correct barcode assignment.
- 2.2.3. Depending on the number of samples, prepare the following Repair/RT Master mix (with 15% excess) as follows:

Reagent	96WP, µL	
	Per well	96 wells + 15%
RAX	8.75	966
RAE	1.25	138
TOTAL	10	1104

- 2.2.4. Using a multichannel pipette or robot, pipette 10 µL of the Repairing/RT Master mix into each well.
- 2.2.5. Carefully seal the plate, vortex well, and briefly spin it in the centrifuge.
- 2.2.6. Incubate in the 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 for the higher input range, we recommend using SPRI-beads for samples below 10 ng.

We highly recommend pooling only half of the reaction volume when pooling more than 20 samples (i.e., 10 μ L out of 20 μ L).

The procedure of cDNA 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). Use a vacuum manifold to prevent column membrane damage that can occur with multiple centrifugation rounds.

# Samples	Pipetting strategy	Zymo-Spin column type	First strand cDNA		DNA Binding Buffer		TOTAL	
			Per well, μ L	Per plate, mL	Per well, μ L	Per plate, mL	Per well, μ L	Per plate, mL
\leq20 samples	multichannel pipette or pipetting robot	I (#D4014)	10	0.20	70	1.4	80	1.6
>20 samples		IIICG (#C1006-50-G)	10	0.96	70	6.72	80	7.68

Table 1 Overview of the recommended pipetting strategy, plasticware, and reagent volumes to be used

Preparation

- Ensure the Zymo DNA Wash buffer contains ethanol.

Procedure

- 2.3.1. Following [Table 1](#), use a multichannel pipette or pipetting robot to transfer the RT volume of each sample into its designated reservoir (25 mL).
- 2.3.2. Add 7x DNA Binding buffer accordingly (see [Table 1](#)). The mix should turn yellow.
- 2.3.3. 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. Alternatively, a few rounds of spin down can be performed in a single column, if no funnel is used (proceed to step [2.3.7](#) afterwards).
- 2.3.4. Gently mix the cDNA in the binding buffer mixture and transfer it to a 25 mL funnel using a Pipetboy.
- 2.3.5. Turn on the vacuum pump and let the liquid pass through the column.
- 2.3.6. Transfer any remaining volume to the funnel. Do not let the membrane over-dry.
- 2.3.7. After the entire pool mix has passed through the column, add 200 μ L of DNA Wash buffer (with Ethanol added) directly to the membrane of the column.
- 2.3.8. Repeat step [2.3.7](#) once the wash buffer has passed through the filter.
- 2.3.9. 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.
- 2.3.10. Add 18 μ L of water to the column matrix and incubate for 1 min.
- 2.3.11. Transfer the column into a new labelled 1.5 mL tube and centrifuge for 30 sec.
- 2.3.12. Immediately proceed to step [2.4](#).

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

Perform cDNA purification using SPRI magnetic beads at a 1:0.9 (cDNA:beads) ratio. Use an appropriate magnetic stand for tube-based separation (e.g., Permagen MSR06).

For larger volumes (e.g., ~1 mL from a 96-well plate), divide the cDNA evenly across the required number of 1.5–2 mL tubes. Add an equivalent proportion of SPRI beads to each tube (e.g., a 960 µL pool split into two tubes of 480 µL each, followed by 432 µL of beads per tube).

Preparation

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

Procedure

2.3.13. Using a multichannel pipette or robot, pool the RT samples in a reservoir (Integra, 4382):

- **>20 samples:** 10 µL
- **≤20 samples:** 20 µL

CRITICAL: It is not recommended to pool more than 10 µL per well when pooling >20 samples.

2.3.14. Transfer the collected pool into a 1.5mL or 2 mL tube, depending on the pooled volume. The final volume will be almost twice as high due to the addition of the beads.

If necessary, split the volume into several tubes.

2.3.15. Add pre-warmed beads in a 1:0.9 ratio (i.e., for 960 µL of pooled samples, add 864 µL of beads slurry), and mix by pipetting up and down ten times.

2.3.16. Incubate for 5 min at room temperature.

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

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

2.3.19. Incubate for 30 sec.

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

2.3.21. Repeat step 2.3.18 for a total of two washes.

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

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

Note: When splitting the RT cleanup into several tubes at step 2.3.14, 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.

2.3.24. 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.

2.3.25. Immediately proceed to step 2.4

2.4. Free primer digestion

Digestion of the left-over barcoded oligo-dT.

CRITICAL: Perform the 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 2.3.11 (if column was used) or 2.3.24 (if SPRI beads) into a new labelled 0.2 mL PCR tube.
- 2.4.2. Prepare the F-EXO reaction Master mix as follows:

Reagent	Per reaction, µL
F-EXB	2
F-EXO	1
TOTAL	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	5 min
Incubation	62	10 min
Keep	4	pause

Procedure

2.5.1. Prepare the SSS FL Master mix for the second-strand synthesis as follows:

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 at 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 SSS sample volume to 50 μL with nuclease-free water.

2.5.7. Add 45 μL of beads 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 transfer 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 provide an anchor for the library amplification and indexing.

NOTE: Use 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 min.
- Prepare 5 mL of 80% ethanol.
- Thaw the ALB and BRB.AD reagents on ice and mix well before use.
- Prepare Program 5_AD_L on the thermocycler (lid off):

Step	Temperature, $^{\circ}\text{C}$	Time
Incubation	20	15 min
Keep	4	pause

Procedure

- 2.6.1. Transfer the cDNA quantity from step 2.5.17 to be used for library preparation into a new tube and adjust the volume 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

CRITICAL: Mix well by pipetting up and down 15 times. This is essential to ensure a sufficient ligation. The presence of small bubbles will not interfere with performance.

- 2.6.2. Briefly spin down the tube in the bench-top centrifuge.
 2.6.3. Incubate in the thermocycler Program 5_ADL.
 2.6.4. Proceed immediately to step 2.6.5.

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.

- 2.6.5. Complement the final reaction volume to 50 µL with water.
 2.6.6. Add 45 µL of beads and mix by pipetting 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. Repeat step 2.6.9 for a total of two washes.
 2.6.13. Remove the tube from the magnetic stand and let the beads dry for 1-2 min.
 2.6.14. Resuspend the beads in 21 µL of water.
 2.6.15. Incubate for 1 min.
 2.6.16. Place the tubes on the magnetic stand, wait 5 min, and carefully transfer 20 µL of supernatant into a new tube to avoid bead carryover.
 2.6.17. Proceed immediately to step 2.7.1.

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 21, depending on the quality and quantity of RNA samples.

It is recommended to perform the final library bead clean-up 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** reagent 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	8-21*
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. Follow the guidelines below.

Library amplification reaction setup

2.7.1. Assemble the PCR amplification reaction as follows:

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

2.7.2. Pipette up and down 5 times.

2.7.3. Place the tube on ice and continue with:

- Step [2.7.8 \(Recommended\)](#) – define the number of PCR cycles by qPCR
- Step [2.7.4.](#) – define the number of PCR cycles based on cDNA or library inputs

2.7.4. 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](#)) or RNA input (step [2.1.2](#)), see both tables below, then continue to library clean-up ([2.7.12](#)).

PCR cycles - cDNA Input

cDNA used for library prep, ng	Number of PCR cycles
150-200	8
50-100	9
20-40	10
10	11
4-5	12
2-3	13
1	14
0.5	15

PCR cycles - Library Input

Total RNA used for library prep	Number of PCR cycles
8x 100ng	10
8x 10 ng	14
8x 1ng	17
8x 100pg	21

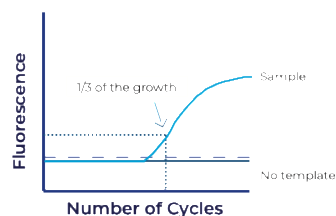
- 2.7.5. If the cDNA amount is not measurable or less than 0.5 ng, we recommend performing qPCR to determine the optimal number of amplification cycles (follow steps 2.7.6 - 2.7.11).
- 2.7.6. For **cycle number optimization**, transfer a 5 μ L aliquot from the PCR reaction, prepared in 2.7.1, into a separate PCR tube. Use the remaining 45 μ L to perform 5 cycles of library preamplification using Program 6_AMP and set the PCR machine to pause at +4°C afterward.
- 2.7.7. Use 5 μ L from the PCR reaction (kept separately in 2.7.6) to set up a qPCR reaction in a suitable PCR tube or plate as follows:

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

*Prepare 100x dilution with nuclease-free water from 10000x stock

- 2.7.8. Run the qPCR reaction using Program 6_AMP for 30 cycles.
- 2.7.9. Analyze the multicomponent plot (as shown in figure below) and determine the optimal number of amplification cycles based on the growth curve

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



- 2.7.10. Resume the reaction by setting the remaining number of amplification cycles, according to the qPCR results (determined in 2.7.8).

For instance, if the optimal number of cycles determined in step 2.7.8 is 10, subtract the 5 cycles already completed. Set the machine to run 5 additional cycles.

- 2.7.11. After amplification is complete, proceed to library clean-up below.

Indexed cDNA library clean-up with SPRI beads

Purify the final cDNA library using SPRI magnetic beads at a 1:0.7 ratio (35 μ L of bead slurry for 50 μ L of cDNA library).

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

- 2.7.12. Adjust the library volume to 50 μ L with water.
- 2.7.13. Add 35 μ L of beads and mix by pipetting up and down 10 times.
- 2.7.14. Incubate for 5 min at room temperature.
- 2.7.15. Place the tubes on the magnetic stand, wait 5 min, carefully remove, and discard the supernatant.
- 2.7.16. To wash the beads, pipette 200 μ L of freshly prepared 80% ethanol into the tube.
- 2.7.17. Incubate for 30 sec.
- 2.7.18. Carefully remove the ethanol without touching the bead pellet.

- 2.7.19. Repeat step 2.7.16 for a total of two washes.
- 2.7.20. Remove tubes from the magnetic stand and let the beads dry for 1-2 min.
- 2.7.21. Resuspend the beads in 21 μ L of water.
- 2.7.22. Incubate for 1 min.
- 2.7.23. Place tubes on the magnetic stand, wait 5 min, and carefully transfer 20 μ L of supernatant into a new tube to avoid bead carry-over.
- 2.7.24. Perform the bead clean-up once again by repeating the procedure from step 2.7.12.

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. Library quantification can also be performed unbiasedly using qPCR with standard Illumina library quantification kits (e.g., KAPA HiFi, Roche). This information is required to assess the libraries' molarity and to 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 Plant BRB-seq library profile.

It's preferable to perform the bead clean-up twice to remove primer dimer fragments, which can likely result in lower-quality sequencing data with reduced mappable reads (see Figure 4). Additionally, high-molecular-weight fragments can sometimes be observed in the final library (see Figure 5). They generally do not impact the quality of the results if most of the library is within the 300-700 bp range.

If the library contains mainly longer fragments (>1000 bp, Figure 6), it may be an indication of a failed fragmentation step. Such a library needs to be re-prepared, starting from the RNA.

Pre-sequencing library QC:

- Use 2 μ L of the library to measure the concentration with Qubit.
- Use 2 μ L of the library to assess the profile with the Fragment Analyzer instrument or similar.
- If necessary, re-purify the libraries by following steps 2.7.12 – 2.7.23 to remove the peaks <300 bp.

Figure 3 Successful library profile with fragments between 300-700 bp

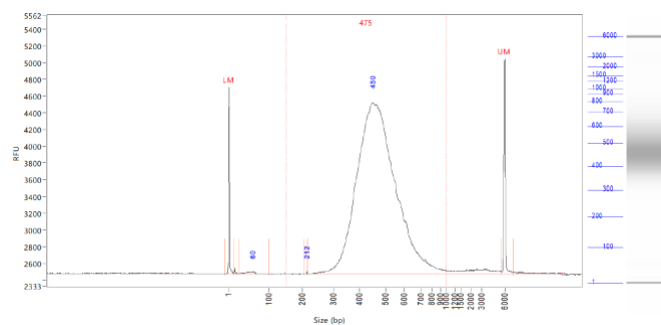


Figure 4 Library profile showing a primer peak around 50bp

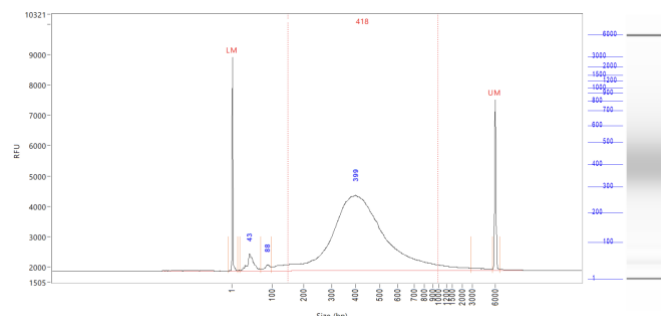


Figure 5 Library profile showing a primer peak around 50bp with some large fragments (>1kbp)

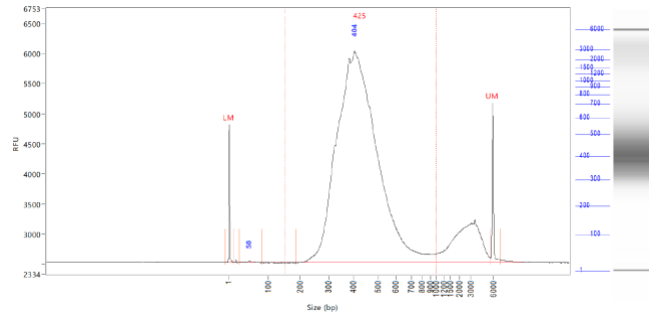
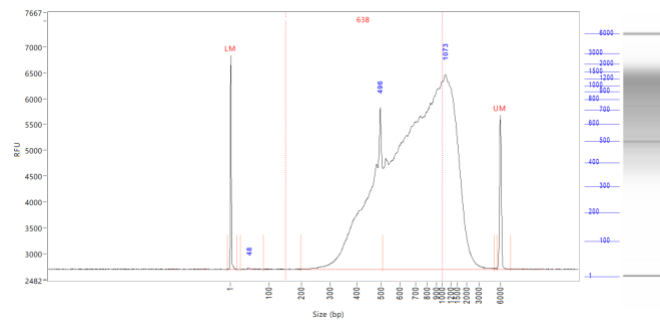


Figure 6 Library profile obtained from the failed RNA fragmentation step



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. This approach ensures that every sample will obtain enough reads required for the analysis. Total Plant BRB-seq libraries can be added as spike-ins to the compatible sequencing run (see Part 3). For this validation, 0.5-1M sequencing reads per library are sufficient to assess the fraction of reads attributed to every sample.

Part 3. LIBRARY SEQUENCING

The libraries prepared with the MERCURIUS™ Total Plant BRB-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.

MERCURIUS™ Total Plant BRB-seq libraries are Unique Dual-Indexed and can be pooled in a sequencing run with other non-MERCURIUS™ libraries if the sequencing structure is compatible and the indexes do not overlap. Please refer to [Table 2](#) for the optimal sequencing structure and [Table 3](#) for the list of i5 and i7 index sequences.

Given the structure of the Total Plant BRB-seq libraries, the optimal read length for Read 1 is 28 cycles (29 cycles on AVITI). Cycles 29–60 correspond to a homopolymeric poly(T) stretch, which can significantly reduce sequencing quality, often reflected by decreased Q30 scores. For this reason, standard Illumina paired-end configurations (e.g., 100 or 150 bp PE runs) are not recommended, as sequencing performance degrades across homopolymer regions.

On the AVITI platform, however, a customized configuration is suitable: Read 1 at 200 bp allows progression through the oligo-dT region into the cDNA, while Read 2 at 100 bp is recommended and fully 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 Plant BRB-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](#)).

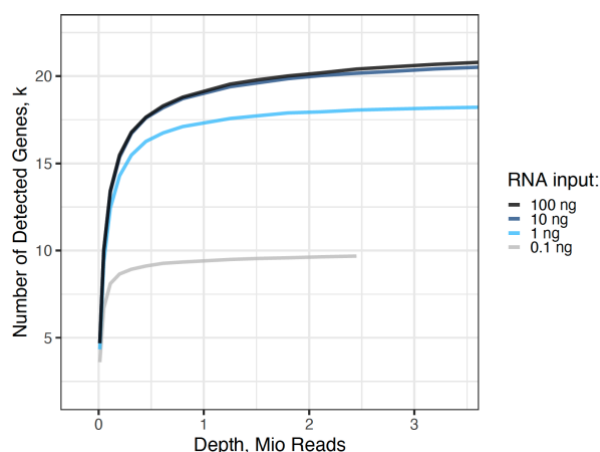
Name	Type	i7 index sequence	i5 index sequence Forward Workflow	i5 index sequence Reverse Workflow
MF.UDI.1	UDI (i7/i5)	GCTTGTCA	AGGCGAAG	CTTCGCCT
MF.UDI.2	UDI (i7/i5)	CAAGCTAG	TAATCTTA	TAAGATTA
MF.UDI.3	UDI (i7/i5)	AGTTCAGG	CAGGACGT	ACGTCCTG
MF.UDI.4	UDI (i7/i5)	GACCTGAA	GTA CTGAC	GTCAGTAC

Table 3 UDI adapter sequences

NOTE: Sequencing depth

1. The recommended sequencing depth is 2-5 Mio reads per sample (see [Figure 7](#)). For extremely lowly expressed genes, 15-20 Mio reads per sample may be required.
2. 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 7 Number of detected genes as a function of the sequencing depth for different quantities of starting RNA per well (*Arabidopsis Thaliana*), SE sequencing



3.1. Sequencing on the Illumina instruments

Table 4 indicates the loading concentration for the Illumina instruments. Please refer to Appendix 1 for a list of Illumina instruments with forward or reverse workflows.

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 the Total Plant BRB-seq libraries sequencing. We recommend diluting the libraries to 0.8 nM before denaturation.

Table 4 Reference loading concentrations for various Illumina instruments

3.2. Sequencing on the Element AVITI instruments

MERCURIUS™ Total Plant BRB-seq libraries can be sequenced with 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 5).

We recommend using PMG shift for the run such as:

- Advanced Run Settings > PMG shift > 2

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 5 Loading concentration for Cloudbreak AVITI and AVITI24 2x75 High Output sequencing kits

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

4.1. Required software

Tool	Description	Version
fastQC	Software for QC of <i>fastq</i> or <i>bam</i> 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 <i>good</i> and <i>poor-quality</i> 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 Plant BRB-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. [fgtk](#) — 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 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., BRB-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 Plant BRB-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 mylibrary_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. Compatible Illumina instruments

Illumina instruments can use two workflows for sequencing the i5 index (see the details in the [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



Alithe Genomics SA

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



Alithe Genomics Inc.

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