Alithea Genomics

MERCURIUS

Blood BRB-seq kit for 96 samples

User Guide

3'-mRNA-seq UDI Library Preparation with Human Globin Genes Blocker Module

RELATED PRODUCTS:
   Barcoded Oligo-dT Adapters Module for 96 samples (Part # MBV4D4)
   Library Preparation & Indexing Module with Human Globin Blocker for 96 samples (Part # MB096301B, #MB0963-B)

v.0.1.2 September 2021

Table of Contents

Kit Components

Reagents supplied

Barcoded Oligo-dT Adapters Module for 96 samples, Part # MBV4D4

Reagents	Label	Volume	Storage
MERCURIUS Oligo-dT primer plate with 96 barcodes	BUV4 oligo-dT	4 µL / well	-20°C
AluSeal		1 pc	Room temperature

Library Preparation & Indexing Module with Human Globin Blocker for 96 samples, Part # MB096301B

Reagents	Label		Volume	Storage
First Strand (FS) Enzyme	FSE	🔴	45 µL	-20°C
First Strand (FS) Buffer	FSB	🔴	550 µL	-20°C
Exonuclease I Enzyme	EXO	🟣	10 µL	-20°C
Exonuclease Buffer	EXB	🟣	20 µL	-20°C
Second Strand Enzyme	SSE	🟠	20 µL	-20°C
Second Strand Buffer	SSB	🟠	45 µL	-20°C
Tagmentation Enzyme	TE	🔴	10 µL	-20°C
Tagmentation Buffer	TAB	🔴	40 µL	-20°C
Library Amplification Mix	LAM	🟢	200 µL	-20°C
Human Globin Blocker Reagent	GBR		379 µL	-20°C
UDI Adapter Mix 1	MQ.UDI.1		10 µL	-20°C
UDI Adapter Mix 2	MQ.UDI.2		10 µL	-20°C
UDI Adapter Mix 3	MQ.UDI.3		10 µL	-20°C
Nuclease Free Water	ddH2O		1000 µL	-20°C

Additional required reagents and equipment (supplied by the user)

Plasticware	Manufacturer	Product number
Solution reservoir 25ml polystyrene	VWR	613-1175
15 mL conical tubes	Greiner	188271
0.2ml 8-Strip Non-Flex PCR Tubes	Starlab	I1402-3700
Corning® 96 Well Black Polystyrene Microplate	Sigma	CLS3603-48EA

Reagents		
Quant-iT™ RiboGreen™ RNA Assay Kit	Invitrogen	R11490
DNA Clean and Concentrator-5 kit	Zymo	D4014
SPRI AMPure Beads or CleanNA Beads	Beckman Coulter CleanNA	A63881 CNGS0050D
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
High Sensitivity NGS Fragment Analysis Kit	Advanced Analytical	DNF-474
SYBR Green	Thermofisher	S7563
Quant-iT™ RiboGreen™ RNA Assay Kit	Invitrogen	R11490

Equipment		
Benchtop centrifuge for 96 well plates		
Benchtop centrifuge for 1.5 ml tubes		
Single and Multichannel pipettes		
Thermocycler with adjustable lid temperature		
Fragment Analyser / Bioanalyzer / TapeStation		
Qubit™ / 96-plate Reader		
Magnetic stand for 0.2 ml tubes		
Real-Time PCR Instrument (Optional)		

Thermocycler Programs

Program 1_FSS (Lid temp: 80°C)

Step	Temperature, °C	Time
Incubation	42	50 min
Inactivation	70	10 min
Keep	4	pause

Program 2_FPD (Lid temp: 90°C)

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

Program 3_SSS (Lid temp: 80°C)

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

Program 4_TN5AMP (Lid temp: 105°C)

Step	Temperature, °C	Time	Cycles
Initial denaturation	98	1 min	
Denaturation	98	10 sec	
Annealing	63	30 sec	12-17 cycles
Extension	72	1 min	
Final Extension	72	3 min	
Keep	4	pause	

Protocol Overview and Timing

The MERCURIUS Blood BRB-seq kit allows the preparation of Illumina compatible 3`RNA sequencing libraries for up to 96 total RNA isolated from human blood samples in a time and cost-effective manner. The kit provides barcoded MERCURIUS Oligo-dT primers that are designed to tag each sample with individual barcodes during the first strand synthesis reaction. This enables to subsequently pool the samples from each experimental group into a single tube for streamlined sequencing library preparation. Human Globin Blocker Reagent (hGBR) included in the kit contains a set of specific oligonucleotides that anneal to the highly abundant human globin genes transcripts (*HBB*, *HBA1* and *HBA2*) preventing their reverse transcription and therefore ablating them from the final sequencing library. Being seamlessly implemented in the standard BRB-seq protocol the Globin Depletions module enables to use total RNA from human blood samples without any prior preparation or treatment.

The protocol can be used to generate high-quality sequencing data starting with 10 - 500 ng of total purified RNA per sample. However, for optimal results, the range of 50 - 200 ng of total RNA and pooling of at least 4 barcodes is recommended. Importantly, the kit can be used to pool any number of samples (from 4 to 96) with two considerations: a) the total RNA amount per pool should be at least 500 ng; b) pooling less than 4 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.

The library indexing is performed using a Unique Dual Indexing (UDI) strategy, which minimizes the risk of barcode misassignment after NGS. This kit provides three UDI pairs, which can be used to prepare and sequence 3 individual BRB-seq libraries in a single flow cell.

Provided below is an estimation of the time required to accomplish each step of the protocol assuming that that RNA samples have been previously quality checked and diluted.

Protocol Workflow

Schematic illustration of protocol workflow
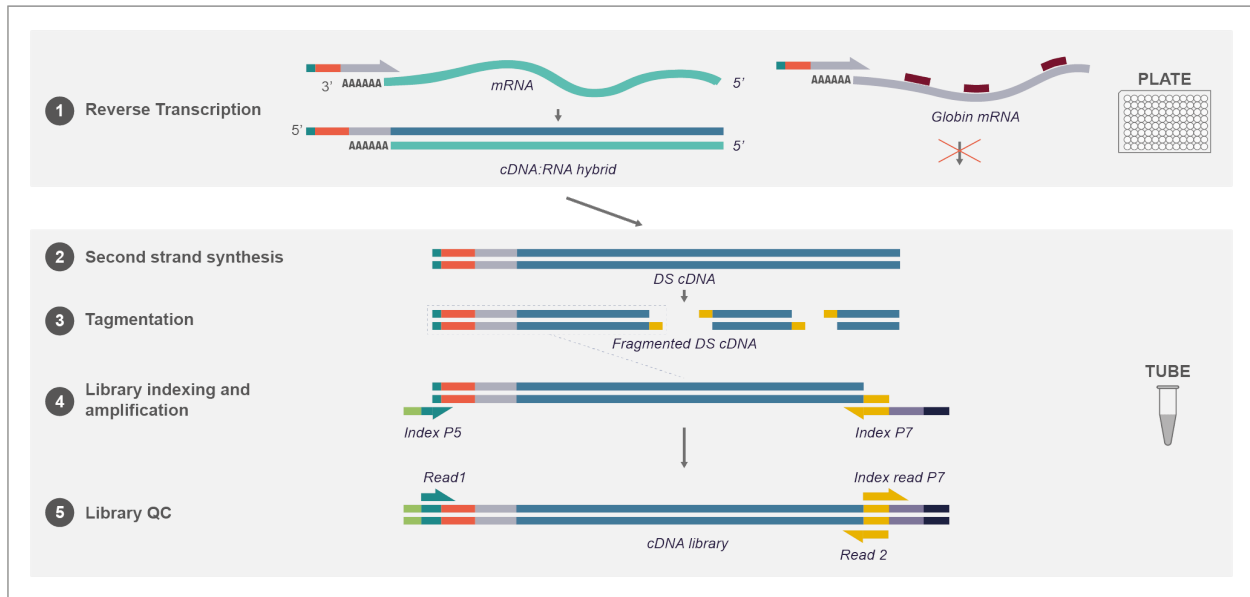
Workflow Timeline

Protocol Step	Time	Clock (estimated)	Notes
1. Reverse transcription		8 am	
a. RNA samples preparation	20 min		
b. RT Mix preparation	5 min	|	50 min hands on
c. Incubation	60 min	8:30 am	2h 30 min total
d. Sample pooling	20 min	|	
e. Reaction preparation	5 min	|	
f. Incubation	50 min	11 am	**Safe stop +4°C**
2. Second strand synthesis		|	
a. Reaction preparation	5 min	|	35 min hands on
b. Incubation	50 min	12 pm	1h 30 min total
c. Beads cleanup	30 min	|	**Safe stop -20°C**
3. Tagmentation		|	
a. Reaction preparation	5 min	1:45 pm	35 min hands on
b. Incubation	7 min	|	50 min total
c. Beads cleanup	30 min	|	
4. Library amplification		|	
a. Reaction preparation	5 min	2:35 pm	35 min hands on
b. PCR	50 min	|	1h 25 min total
c. Beads cleanup	30 min	|	**Safe stop -20°C**
5. Library QC		|	
a. Qubit quantification	10 min	4 pm	10 min hands on
b. Smear analysis	60 min	5:10 pm	1h 10 min total

Part 1.   PREPARATION OF RNA SAMPLES

1.1.   Important considerations for input RNA

- The recommended range of total RNA amount per well is 50 – 200 ng.
- The recommended RIN number of each sample is >6.
- To ensure an even distribution of reads after sequencing, the RNA amount and integrity (RIN number) of the starting RNA samples must be as uniform as possible, i.e., 50 ng (±10%) of starting material with RIN~6 for every sample. To obtain such uniform amounts, we, therefore, recommend to:
    - Measure the RNA concentration of all samples with a dye-based method (e.g., Quant-iT RiboGreen, see the section below).
    - Dilute samples to obtain the same RNA concentration in all wells (±10%).
    - Re-measure the RNA concentration of all samples to confirm uniform concentrations across the samples.

1.2.   RNA quantification with Quant-IT RiboGreen

- Prepare the required volume of the RiboGreen dye solution in 1x TE buffer including 20% excess volume. The table below provides an example for preparing a quantification solution for 96 samples (+20% excess).
- Distribute 200 µL of RiboGreen solution in a separate well for every sample and include column 12 for the standard RNA dilution: 7 samples + negative control. If you have more than 88 samples, prepare 2 plates with the same layout.
- Transfer 2 µL of RNA into each well and 2 µL of RNA standards in column 12.
- Cover with the lid and briefly spin the plate to remove bubbles.
- Measure the fluorescence with a microplate reader.
- Calculate concentrations based on the fluorescence values of the samples and the RNA standards.

Example RiboGreen solution for 96 samples		
Component	**Stock (x)**	**Volume (mL)**
1x TE buffer	**20x**	**1.152**
1x Dye	**200x**	**0.1152**
Water		**21.7728**

Part 2.   LIBRARY PREPARATION PROTOCOL

2.1.   Reverse transcription

At this step, each individual RNA sample is reverse-transcribed using barcoded oligo-dT primers. Subsequently, all barcoded samples can be pooled in one tube.

Before you start:
- Set the thermocycler to 70°C on hold.
- Thaw **MERCURIUS Oligo-dT primer** plate on the bench at room temperature and briefly spin it down before opening the seal.
- Thaw the **FSB** reagent at room temperature, mix well before use.

NOTE: All the manipulations with RNA and RT enzyme should be performed on ice and using gloves.

- Thaw the RNA samples on ice.
- Transfer 7 µL of normalized RNA (see Part 1. PREPARATION OF RNA SAMPLES) to the corresponding wells of the **MERCURIUS Oligo-dT primer** plate.
- Add 3.6 µL of Globin Blocker Reagent (GBR) to each well containing RNA.
- Seal the plate with the AluSeal and place it in a thermocycler at 70°C for 10 min. From now on the plate will be referred as **RT plate**.
- Immediately put on ice.
- Prepare the Master Mix for RT reaction (+10%) as following:

RT Mix	Per well (µL)	Per well +10% (µL)	Ex: 96 wells (µL)	__ wells (µL)
FSB	5.0	5.50	528.0	
FSE	0.4	0.44	42.2	
TOTAL	5.4	5.94	570.24	

- Distribute **5.4 µL** of Master Mix into each well of the **RT plate**,
- Carefully seal the **RT plate** and briefly spin in the centrifuge
- Incubate the plate in thermocycler **Program 1_FSS.**

Safe stop: After this step the **RT plate** can be kept at 4°C overnight.

2.2.   Samples pooling

At this step the barcoded RT samples are pooled together and purified using Zymo Clean & Concentration Kit (Zymo, D4014) column. For pooling take 5-20 µL of RT from each well using multichannel pipette and transfer into reservoir containing DNA Binding buffer (Zymo, D4004-1-L).

NOTE: Library normalization

At this point, the volume used for pooling from each well can be adjusted to re-equilibrate the proportion of samples in the pool. This may be required to improve the distribution of sequencing reads in the library specifically if some samples are obtaining too many reads.

Shallow sequencing allows to unequivocally assess the coverage per sample. For this approach, we recommend pooling only a fraction of RT volume from each well (i.e., 5 µL out of 20 µL) for the library preparation. After the library QC by sequencing (see section 2.7.2) the volume used for pooling can be re-adjusted to reduce the variation at the sequencing stage.

- Define the fine volume of pooled RT and transfer 7x of DNA Binding buffer into pooling reservoir (e.g., if pooling 5 µL from 96 wells use 5 x 96 x 7 = 3360 µL of DNA Binding buffer)
- Using a multichannel pipette transfer defined RT volume from each well into a reservoir. Change the tips between pipetting from different rows/columns or wells (when using a single-channel pipette).
- Transfer the DNA Binding buffer into a 15 mL tube and gently mix by inverting the tube.
- If RT plate still contain some volume, carefully seal it using provided AluSeal and store at -20°C.
- Perform the purification following the Zymo Clean & Concentration Kit manual and elute in 20 µL of nuclease-free water.
- Transfer 17 µL of eluate in a new 200 µL PCR tube.

2.3.   Free primer digestion

- Thaw **EXB** at room temperature and **EXO** keep constantly on ice
- Prepare the reaction mix as following:

Reagent	Per reaction (µL)
First-strand cDNA	17.0
EXB	2.0
EXO	1.0
TOTAL	20.0

- Mix by pipetting up and down 5 times
- Briefly spin down in the bench-top centrifuge
- Incubate in thermocycler **Program 2_FPD.**

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

2.4.   Second strand synthesis

At this step, the double-stranded full-length cDNA is generated and purified using magnetic beads.

Before you start:
- Pre-warm SPRI beads at room temperature for ~30 min.
- Thaw the **SSB** reagent at room temperature, mix well before use.

2.4.1. Second strand synthesis reaction setup

- Keep **SSE** reagent constantly on ice.
- Prepare the reaction mix for the second strand synthesis as following:

Reagent	Per reaction (µL)
First-strand cDNA	**20.0**
SSB	**5.0**
SSE	**2.0**
TOTAL	**27.0**

- Incubate in thermocycler **Program 3_SSS.**
- Proceed immediately to step 2.4.2.

2.4.2. Full-length cDNA cleanup with SPRI beads

Purify the double-stranded cDNA with SPRI magnetic beads using a 0.6x ratio (30 µL of bead slurry plus 50 µL cDNA):

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

- Add 23 µL of Nuclease Free Water to second-strand synthesis reaction tube for the final volume of 50 µL
- Add 30 µL of beads and mix by pipetting up and down 10 times
- Incubate 5 min at room temperature
- Place the tube on the magnetic stand, wait 5 min, carefully remove and discard the supernatant
- Wash the beads with freshly prepared 80% ethanol twice:
  - Pipette 200 µL into the tube
  - Incubate 30 sec
  - Carefully remove the ethanol without touching the beads pellet
- Remove the tube from the magnetic stand and let the beads dry for 1-2 min
- Resuspend the beads in 21 µL of Nuclease Free Water
- Incubate for 1 min
- 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
- Measure the concentration with Qubit

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

2.5.   Tagmentation

At this step, the full-length cDNA is tagmented using a Tn5 transposase pre-loaded with adapters for library amplification.

Before you start:
- Pre-warm SPRI beads at room temperature for ~30 min.
- Thaw the **TAB** reagent at room temperature, mix well before use.
- Set the PCR machine at 55°C incubation.

2.5.1. Preparation of tagmentation reaction

- Keep **TE** reagent constantly on ice
- Prepare the Tagmentation mix on ice in a PCR tube. If several cDNA samples will be tagmented, prepare a Master mix as follows (with 10% excess):

Reagent	Per reaction (µL)		N +10% (µL)
	For up to 9 ng of cDNA	**For 10-20 ng of cDNA**	
TAB	4.0	4.0	
TE	1.0	2.0	
cDNA	X µL	X µL	
ddH2O	15.0 - X	14.0 - X	
TOTAL	20.0	20.0	

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

2.5.2. Tagmented cDNA cleanup with SPRI beads

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

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

- Add 30 µL of beads to 50 µL of tagmented cDNA and mix by pipetting up and down 10 times.
- Incubate 5 min at room temperature.
- Place the tubes on the magnetic stand, wait 5 min, carefully remove and discard the supernatant
- Wash the beads with freshly prepared 80% ethanol twice:
    - Pipette 200 µL into the tube
    - Incubate 30 sec
    - Carefully remove the ethanol without touching the beads pellet
- Remove tubes from the magnetic stand and let the beads dry for 1-2 minutes
- Re-suspend the beads in 21 µL of Nuclease Free Water
- Incubate for 1 min

- 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.

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

2.6.   Library indexing and amplification

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

Before you start:

- Pre-warm SPRI beads at room temperature for ~30 mins.
- Thaw the **LAM** reagent on ice, mix well before use.
- Thaw the required number of **MQ_UDI Adapters** at room temperature, briefly spin before use.

2.6.1. Library amplification reaction setup

The number of amplification cycles required for library preparation can vary depending on the RNA samples and input cDNA amount used for tagmentation. To determine the optimal number of cycles, follow the Library Quantification Protocol provided in the **Appendix**.

- Prepare the PCR amplification reaction as following:

Reagent	Per reaction (µL)	N +10% (µL)
LAM	**25.0**	
MQ_UDI Adapter	**5**	
Tagmented cDNA	**20.0**	
TOTAL	**50.0**	

- Pipette up and down 5 times.
- Put the tube in a PCR machine and start the **Program 4_TN5AMP** with 15 amplification cycles as a default value.
- Alternatively, run **Program 4_TN5AMP** with 5 cycles and follow Library Quantification Protocol to determine the optimal number of amplification cycles (see **Appendix**).

2.6.2. Indexed cDNA library cleanup with SPRI beads

Purify the final cDNA library with SPRI magnetic beads using a 0.6x ratio (30 µL of bead slurry for 50 µL cDNA library). It is recommended to perform this clean-up twice in order to completely remove the residual adapter dimers.

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

- If 5 µL of library was used for quantification adjust the volume to 50 µL
- Add 30 µL of beads and mix by pipetting up and down 10 times.
- Incubate 5 min at room temperature.

- Place the tubes on the magnetic stand, wait 5 min, carefully remove and discard the supernatant
- Wash the beads with freshly prepared 80% ethanol twice:
  - Pipette 200 µL into the tube.
  - Incubate 30 sec
  - Carefully remove the ethanol without touching the beads pellet.
- Remove tubes from the magnetic stand and let the beads dry for 1-2 min
- Re-suspend the beads in 21 µL of Nuclease Free Water.
- Incubate for 1 min
- Place tubes on the magnetic stand, wait 5 minutes and carefully remove 20 µL of supernatant into a new tube to avoid bead carry-over.

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

2.7.   Library quality control

2.7.1. Pooled library quality control

Before sequencing, the libraries should be subjected to fragment analysis (with Fragment analyzer, Bioanalyzer or TapeStation) and quantification (with Qubit). This information is required to assess the molarity of the libraries and prepare the appropriate library dilution for sequencing. A successful library contains fragments in the range of 300 – 1000 bp (see Figure 1 for an example of a standard BRB-seq library profile).

Importantly, libraries with peaks in the range 250 – 290 bp will likely produce lower quality sequencing data with a reduced proportion of mappable reads (Figure 2). Therefore, it is strongly recommended to remove the peaks in this range bp by performing an additional round of SPRI beads purification with the 0.6x ratio.

Peaks around ~160 bp are adapter dimers and can be tolerated in the final library if they represent less than 10% of the total fragments. Alternatively, the library should be re-purified with SPRI beads (0.6x ratio).

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

Pre-sequencing library QC:

- Measure the library concentration with Qubit
- Assess the library profile with the fragment analyzer
- If necessary, re-purify the libraries to remove the peaks <300 bp.

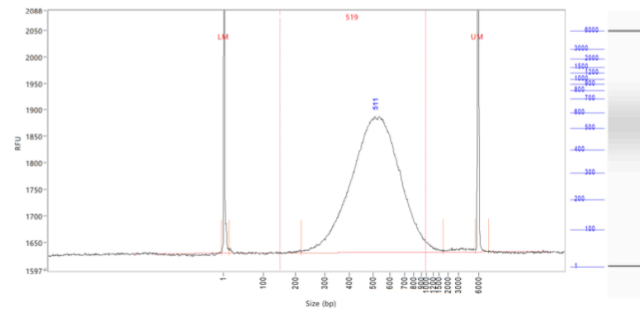Figure 1. The successful library profile with fragments 300-1000 bp

Figure 2. The example of an over-tagmented library profile with a peak at 290 bp and an adapter peak at 160 bp
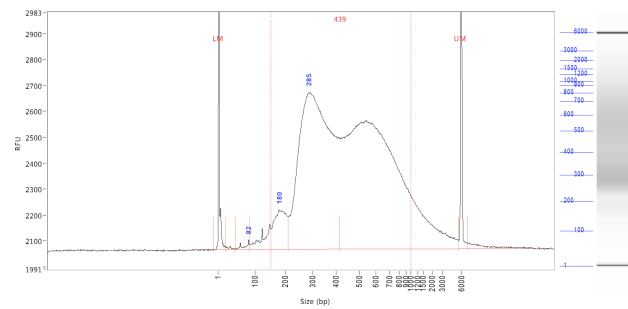
Figure 3. The example of an under-tagmented library profile with a major peak > 1000 bp
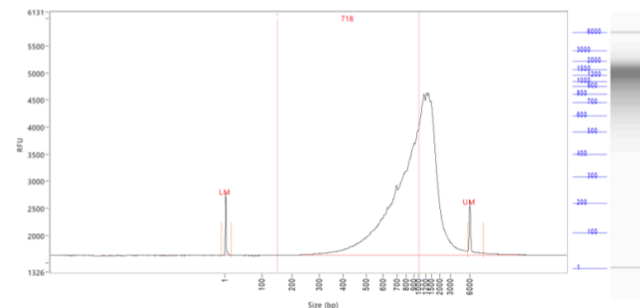
2.7.2. Assessing uniformity of reads distribution across the samples

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

Part 3.   LIBRARY SEQUENCING

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

The MERCURIUS BRB-seq libraries are Dual Indexed and potentially can be pooled in a sequencing run with other libraries if the sequencing structure is compatible.

The libraries require paired-end sequencing with the following structure:

Read	Length (cycles)	Comment
Read1	25-28	Default is 25. Sample barcode and UMI
i7 index	8	Library Index
i5 index	8*	Library Index (*optional and valid for UDI libraries)
Read2	50-90	Gene fragment

NOTES:

- The recommended sequencing depth is 4-5 M reads per sample but deeper sequencing can also be performed in order to enable the detection of very low expressed genes.

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

- The loading molarity for the library depends on the sequencing instrument and should be discussed with the sequencing facility or responsible person.

Part 4.   SEQUENCING DATA PROCESSING

4.1.   Software

fastQC, version v0.11.7. Software for QC of *.fastq* or *.bam* files. This software is used to assess the quality of the sequencing reads such as 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 bad quality data.

STAR, version 2.7.9a. Software for read alignments on reference genome (Dobin et al., 2013). Can be downloaded from Github (https://github.com/alexdobin/STAR).

STAR can only be run on UNIX systems and requires:

- x86-64 compatible processors
- 64 bit Linux or Mac OS X.
- ~30-40Gb of RAM

BRBseqTools, version 1.6. Software suite for processing BRB-seq libraries, is available from: https://github.com/DeplanckeLab/BRB-seqTools.

R Software, version 3 or greater.

4.2.   Data processing pipeline

This pipeline requires a pair of *.fastq* files (*mylibrary*_R1.fastq.gz and *mylibrary*_R2.fastq.gz) normally obtained for each library after Illumina sequencing and standard index demultiplexing. The following guidelines will help you to generate a gene count matrix ready for analysis.

4.2.1. Merging *.fastq* files from individual lanes and/or libraries

Depending on the type of the instrument used for sequencing, you can obtain one or multiple R1/R2 *.fastq* files per library resulting from individual lanes of a flow cell. For example:

```
>   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
```

The .fastq files from individual lanes should be merged together into a single *R1.fastq* and a single *R2.fastq* files to simplify the following steps.

This is done with a `cat` command in a terminal:

```
>   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
```

This generates two files: *mylibrary_R1.fastq.gz* and *mylibrary_R2.fastq.gz*, contains the information for entire library.

Put these 2 files in a folder named $fastqfolder

NOTE: This step can also be done if you sequenced your library in multiple sequencing runs.

Warning: The order of merging files should be kept the same (for e.g., L001, L002, L003, L004, not L002, L001 ...) to avoid issues when demultiplexing the samples.

4.2.2. Sequencing quality check

Run fastQC on both *.fastq* files. The output folder is $QCdir, it will contain an *html* file and the report from fastQC.

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

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

NOTES:

- Report for R1 fastq file contains some "red flags" because it contains barcodes/UMIs. Stil it can provide useful information on the sequencing quality of the barcodes/UMIs.

- The main point of this step is to check the R2 fastq report. Of note, *per base sequence content* and *kmer content* are rarely green. If there is some *adapter contamination* or *overrepresented sequence* detected in the data, it may not be an issue (if the effect is limited to <10~20%). These are lost reads but most of them will be filtered out during the next step.

4.2.3. Preparing the reference genome

The *.fastq* files then require to be aligned on a reference genome. We use [STAR](#) (Dobin et al., 2013) aligner for aligning BRB-seq libraries, which is currently one of most efficient tools for this applicaiton. It contains a "soft-clipping" tool that automatically cuts the beginning or the end of the reads to improve the mapping efficiency, thus allowing the user to skip the step of trimming the reads for adapter contamination. Moreover, STAR has a mode called STARsolo, which is specifically designed for aligning multiplexed data (such as BRB-seq) and directly generating count matrices.

In order to use STAR, you need to download the genome assembly of the species you want to align on. Then, you will need to index the genome, before it's ready for being used as a reference for alignment. However, this download/indexing step has to be done only once. If you need to align on the same genome in the future, you may use the index created before.

NOTE: Try to download the primary_assembly fasta file if available (without the 'sm' or 'rm' tags). If not available, download the top_level assembly. For the gtf, download the one that does not have the 'chr' or 'abinitio' tags.

Download the proper genome assembly (*.fasta* file, here called $fastafile) and gene annotation (*.gtf* file, here called $gtffile) from Ensembl or UCSC. Below is an example for a human assembly:

```
>   wget ftp://ftp.ensembl.org/pub/release-102/fasta/homo_sapiens/dna/$fastafile
>   gzip -d $fastafile # Because it is gzipped
```

```
>   wget ftp://ftp.ensembl.org/pub/release-102/gtf/homo_sapiens/$gtffile

>   gzip -d $gtffile # Because it is gzipped
```

Once these files are downloaded, an index for the genome assembly can be created using STAR (the output folder containing the index files will be called $genomedir).

```
>   STAR --runMode genomeGenerate --genomeDir $genomedir --genomeFastaFiles $fastafile --
    sjdbGTFfile $gtffile --runThreadN 8
```

NOTES:

- The --runThreadN parameter can be modified depending on the number of cores available on your machine. The larger is this number, the more parallelized/fast will be the indexing.

- Depending on the genome assembly, STAR can use up to 32-40Gb of RAM. So, you should use a machine that has this RAM capacity.

4.2.4. Aligning to the reference genome

After the genome index is created, both R1 and R2 *.fastq* files can be aligned to this reference genome. For this step, use the "solo" mode of STAR, which not only aligns the reads to the reference genome but also creates the UMI count matrices.

There are two important parameters to check/set in this step. First, you need to have a $barcodefile with the list of barcodes that you used for multiplexing your samples. One barcode per line. Second, you need to change the following parameters according to your sequencing:

- *--soloCBstart*: Beginning of the barcode in the R1 fastq file. Equals to **1,** since the R1 file starts with the barcode sequence, which is followed by the UMI.
- *--soloCBlen*: Length of your barcodes (this should match the length of the barcode sequence in $barcodefile). In current version of **MERCURIUS Oligo-dT primers** (V4) the barcodes are **12** nt long.
- *--soloUMIstart*: Usually, the UMI starts right after the barcode. In current version of **MERCURIUS Oligo-dT primers** (V4) the UMIs are 16 nt long so this will be set to **13**
- *--soloUMIlen*: The length of UMI depends on the number of sequencing cycles performed for the Read1. The sum of *CBlen + UMIlen* is equal to the Read1 length in the *R1.fastq* file, therefore *UMIlen* will equal to Read1_length - 12. For example: if the Read1 length was 21 cycle (21 nt) the UMI will be 21 - 12 = 9, therefore, use the option *soloUMIlen* **9**.

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

```
>   STAR --runMode alignReads --outSAMmapqUnique 60 --runThreadN 8 --outSAMunmapped
    Within --soloStrand Forward --quantMode GeneCounts --outBAMsortingThreadN 8 --
    genomeDir $genomedir --soloType CB_UMI_Simple --soloCBstart 1 --soloCBlen 12 --
    soloUMIstart 13 --soloUMIlen 9 --soloCellFilter None --soloCBwhitelist $barcodefile -
    -soloFeatures Gene --outSAMattributes NH HI nM AS CR UR CB UB GX GN sS sQ sM --
    outFilterMultimapNmax 1 --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate -
    -outFileNamePrefix $bamdir --readFilesIn $fastqfolder/mylibrary_R2.fastq.gz
    $fastqfolder/mylibrary_R1.fastq.gz
```

The alignment quality and performance metrics are in the "$bamdir/Log.final.out" file.

> **NOTE**: The most important statistic is the proportion of "Uniquely mapped reads" which is expected to be greater than 60% (for human, mouse or drosophila).

4.2.5. Generating the UMI and Gene Count matrices

STARsolo will generate the UMI count matrix (matrix.mtx file) located in $bamdir/Solo.out/Gene/raw folder. This file is sparse matrix format that can be transformed into standard Count Matrix using a small R script (provided below) that will produce the output in the $count_matrix folder.

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

Alternatively, raw read Gene Count matrix can be generated instead of the UMI Count matrix using the BRBseqTools software (see **Software and materials**).

For this run the following command on the *.bam* file is generated by STAR, that should be located: $bamdir/Aligned.sortedByCoord.out.bam

```
>   java -jar /software/BRBSeqTools-1.6.jar ExtractReadCountMatrix -b
    $bamdir/Aligned.sortedByCoord.out.bam -c $barcodefile -gtf $gtffile -o $count_matrix/
```

The resulting UMI and Gene Count matrices can be used for subsequent analysis of gene expression using established pipelines and tools.

Appendix

Unique Dual Indexing Adapters sequences

Name	Type	i7 index sequence (any sequencer)	i5 index sequence (NextSeq)	i5 index sequence (NovaSeq)
MQ.UDI.1	UDI (i7/i5)	TAAGGCGA	AGGCTATA	TATAGCCT
MQ.UDI.2	UDI (i7/i5)	CGTACTAG	GCCTCTAT	ATAGAGGC
MQ.UDI.3	UDI (i7/i5)	AGGCAGAA	AGGATAGG	CCTATCCT

Library quantification with qPCR

To determine the optimal number of cycles the following library quantification procedure using real-time PCR is recommended at the **step 2.6.1.**

- After purification of the tagmented library setup the PCR amplification reaction and perform 5 cycles of library preamplification (**Program 4.TN5AMP**).
- Put the tube on ice
- Use 5 µL aliquot from the PCR reaction to prepare a qPCR reaction mix in the appropriate PCR tube or plate as following:

Reagent	Per reaction, µL	N +10%, µL
5 cycles pre-amplified library	5.0	
SYBR 100X	0.1	
LAM	2.5	
ddH2O	2.4	
TOTAL	10.0	

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

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

- Determine the cycle number depending on the growth curve in multicomponent plot as in the **Figure 4**
- Perform PCR with determined additional number of cycles with the tube containing the remaining 45 µL of pre-amplified library.

Figure 4. Determination of
additional number of amplification
cycles with qPCR.