# 1. Multiplexing

Libraries prepared with the UDI 12 nt Unique Dual Indexing Sets are suitable for sequencing on all Illumina instruments listed below. The complete lists of i5 and i7 index sequences for all UDI 12 nt Sets are available at www.immaginabiotech.com.

Depending on the instrument workflow, flow cell type (paired-end, PE; single-read, SR), and chemistry, i5 indices are sequenced on the forward or the reverse complement strand.

| Illumina Instruments | Flow Cell Type | UDI 12 nt Unique Dual Indexing |
|---|---|---|
| iSeq 100 MiniSeq<br>NextSeq 500-2000<br>HiSeq 3000/4000<br>NovaSeq 6000 (v1.5 reagent kits) | PE | SetB1 (#iMPS#1_21) |

The Index 2 Read (i5) is performed after Read 2 Resynthesis, using the Index 2 (i5) Sequencing Primer.

EXAMPLE: i512_0001 is read as GTCTTTGGCCCT instead of AGGGCCAAAGAC. The read out in reverse complement (GTCTTTGGCCCT) shall be used for demultiplexing and error correction.

## Index Balance

In general, it is important that each nucleotide (A, C, G, and T) is present at each position of the index reads (Index 1 Read, i7; and Index 2 Read, i5), and that the signal intensity of each nucleotide is perfectly balanced to maintain optimal base calling accuracy and read quality. This is particularly critical for instruments that use two-channel detection (e.g., NextSeq, MiniSeq, and NovaSeq) and one-channel detection (iSeq). Using the UDIs in numerical order as the number of libraries to multiplex increases, or column-wise for increasing multiples of 8 samples, will result in optimal nucleotide balance. For smaller numbers of samples we can also suggest the following:

**Four libraries:** Use UDI12B_0001 - 0004 as these contain almost perfect nucleotide balance at each position of the index read.

**Eight libraries:** Use column 1 of the UDI Set A (SetB1 (#iMPS#1_21)).

NOTE: Individual libraries within a lane or run should always be pooled at an equimolar ratio to preserve perfect nucleotide balance at each position of the index read.

## Lane Mix Preparation

Libraries should ideally be pooled in an equimolar ratio for multiplexed sequencing. It is important to ensure accurate quantification of the individual libraries prior to pooling, as well as for the library pool (lane mix). To quantify your libraries:

1. Measure the concentration of each library, using either qPCR or fluorescence-based assays (e.g., QuBit, Thermo Fisher Scientific Inc.)
2. Determine the average library size, using microcapillary electrophoresis analysis (e.g., Bioanalyzer, Agilent Technologies, Inc.). Set the range to 175 - 1,000 bp to include the whole library size distribution, and to exclude any linker-linker artifacts (~175 bp), or overcycling bumps (>1,000 bp).

Molarity is then calculated from the average library size and the concentration (ng/µl) using the following equation:

$$\text{Molarity} = (\text{library concentration (ng/µl)} \times 10^6) / (660 \times \text{average library size (bp)})$$

After pooling the libraries, the prepared lane mix and any dilutions made for denaturing (e.g., 2 nM), should be reanalyzed to ensure the accuracy of the concentration. This can be performed according to steps 1 and 2 as above.

# 2. Sequencing

## General

The amount of library loaded onto the flow cell will greatly influence the number of clusters generated.

All AHARIBO_TRANSLATOME_RNAseq libraries can be sequenced using the standard Illumina Multiplexing Read 1 and Read 2 Sequencing Primers.

A schematic representation of those libraries is shown below.

AHARIBO_TRANSLATOME_RNAseq inserts are on average 150 bp long and libraries can be sequenced in single-read (SR) or paired-end (PE) formats.

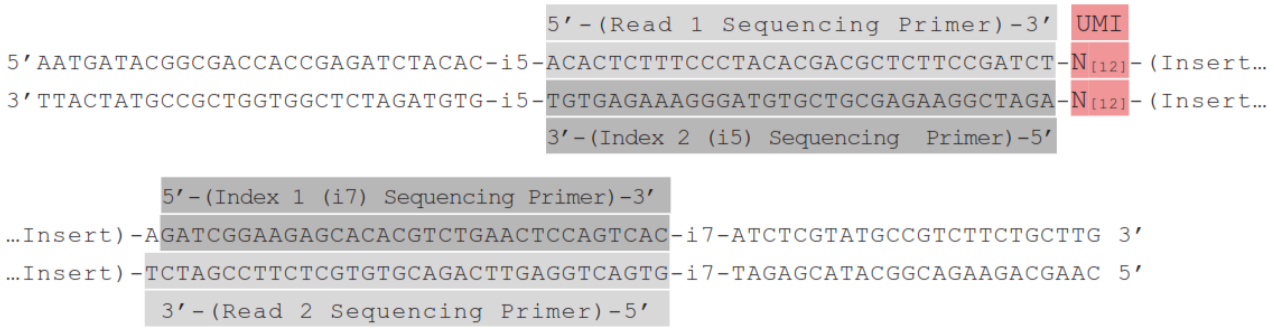
To minimize read overlap and trimming, optimal sequencing modes include SR 150 and PE 75. The required sequencing depth per sample may vary depending on the intended application and sample type. Unique Molecular Identifiers (UMIs) are contained at the start of Read 1 (see below). Therefore, the complete information is conveniently accessible by cost-efficient single-read sequencing.

## Dual-Indexed Library Sequencing Workflows

The workflow for dual-indexed library sequencing differs, depending on the Illumina instrument and flow cell type. Dual indexing can be performed on single-read (SR) and paired-end (PE) flow cells. All HiSeq systems support SR and PE flow cells. NextSeq, MiniSeq, MiSeq, iSeq, and NovaSeq systems use use PE flow cells only, which can also be used in single-read mode. Illumina defines Forward Strand (A) and Reverse Complement (B) Workflows for dual indexing read-out, which refer to the order of Index 2 read-out in relation to Read 2 Resynthesis.

The example below shows the sequencing setup for dual-indexed AHARIBO_TRANSLATOME_RNAseq libraries sequenced with the Reverse Complement Workflow (B) on a paired-end flow cell.

EXAMPLE: MiniSeq, iSeq, HiSeq 3000 / 4000 (PE), NovaSeq 6000 (v1.5 chemistry), and NextSeq instruments use the Reverse Complement Workflow (B) with the Multiplexing Index 2 (i5) Sequencing Primer. A minimum of eight cycles with imaging are required for i5 Index read-out, provided the correct UDI set is used for the correct i5 Index read out workflow. 12 nt, 10 nt, or 8 nt can be read out optionally.




## Sequencing Primers

Standard Illumina sequencing primers are used for all dual-indexed libraries. The Multiplexing Read 1 Sequencing Primer is always used for Read 1 sequencing, and the Index 1 (i7) Sequencing Primer is always used for Index 1 Read (i7) sequencing. The Index 2 Read (i5) is initiated using different sequencing primers specific to the instrument and flow cell type.

**Read 1 for AHARIBO_TRANSLATOME_RNAseq libraries:**

Multiplexing Read 1 Sequencing Primer (not supplied):

5’ ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3’

**Index 1 Read (i7):**

Multiplexing Index 1 (i7) Sequencing Primer (not supplied):

5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCAC 3’

**Index 2 Read (i5):**

Multiplexing Index 2 (i5) Sequencing Primer (not supplied):

5’ AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT 3’

**Read 2:**

Multiplexing Read 2 Sequencing Primer (not supplied):

5’ GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3’

# 3. Data Analysis

This section describes a basic bioinformatics workflow for the analysis of AHARIBO_TRANSLATOME_RNAseq NGS data and is kept as general as possible for integration with your standard pipeline. In contrast to most other library preparation protocols, AHARIBO_TRANSLATOME_RNAseq libraries generate reads in forward orientation, thus mapping should be performed to the corresponding strand of the genome.

## Demultiplexing

Demultiplexing can be carried out by the standard Illumina pipeline. i7 and i5 12 nt index sequences are available for download at www.immaginabiotech.com.

## Processing Raw Reads

We recommend the use of a general fastq quality control tool such as FastQC or NGS QC Toolkit to examine the quality of the sequencing run. These tools can also identify over-represented sequences, which may optionally be removed from the data set.

## Trimming

As AHARIBO_TRANSLATOME_RNAseq libraries are based on random priming the first 9 nucleotides of Read 2 may have an increased error rate. As random priming may also occur at the junction between the ultimate exon and the poly(A) tail, mapping rates can be increased by trimming of poly(A) sequences at the 3' end of Read 1 and poly(T) sequences the 5' end of Read 2, when analyzing data from paired end runs. Further, AHARIBO_TRANSLATOME_RNAseq libraries contain $N_{12}$ Unique Molecular Identifiers (UMIs) at the start of Read 1. Hence, the first 12 nucleotides of Read 1 can be trimmed before proceeding to alignment. Alternatively, a less stringent aligner could be used with relaxed settings. Low quality sequences and adapter sequences should be trimmed. In case an adapter sequence is detected at the 3’ end of Read 2, an additional 12 nucleotides upstream of the adapter can also be trimmed (i.e., the UMI sequence).

## Alignment

After trimming, filtered and trimmed reads can be aligned with a short read aligner to the reference genome or assembled de novo. Please note, that Read 1 reflects the RNA transcript sequence not the cDNA sequence. This is important for downstream applications. If data from paired-end runs with read length >100 nucleotides is analyzed, ensure that the aligner used can handle overlaps (e.g., use relaxed settings).