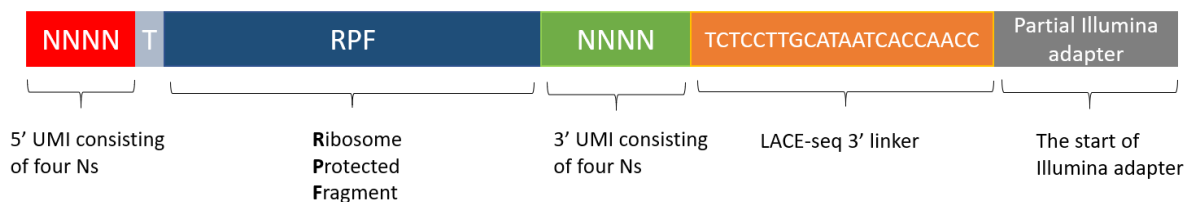


Supplemental product information and tips for success

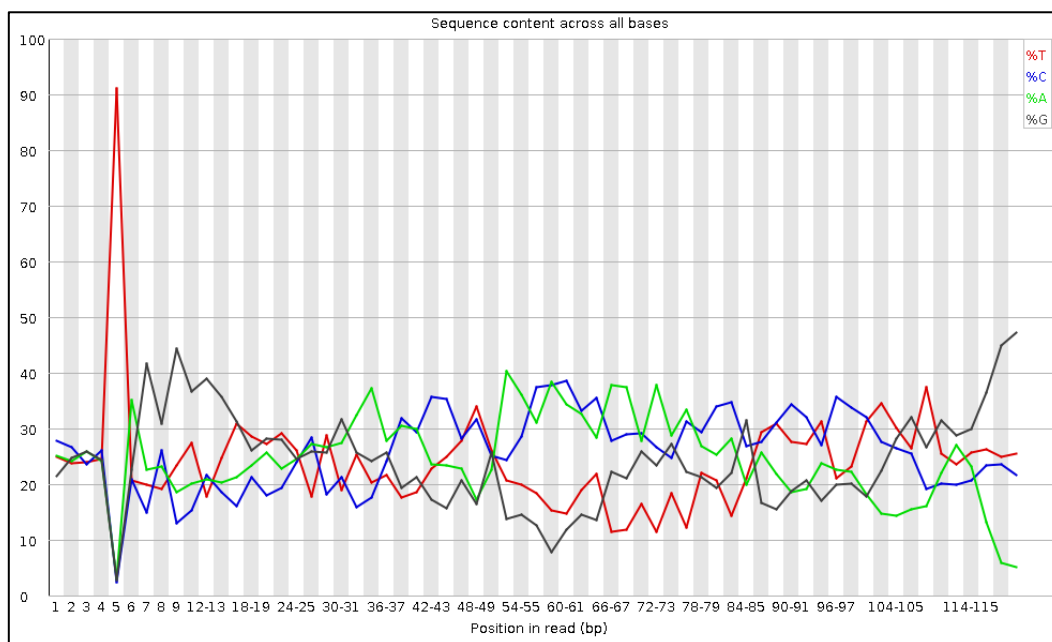
A. Expected Illumina sequencing output

Example of a read generated by LACE-seq for Illumina:



Unique molecular identifiers (UMIs) are strings of random nucleotides that are attached to RPFs prior to PCR amplification and can be used to accurately detect PCR duplicates. The T at position five precedes the start of the RPF.

The sequence content of a high quality LACEseq library has a T peak in position five in 90-100% of the reads (Figure below). The lower this percentage, the higher the adapter-dimer contamination in the library. The adapter-dimer contaminant can be filtered out following section B of this guide.



B. Analysis of sequencing reads

Proper trimming of the reads is important for efficient mapping. Here we provide some guidance on the use of (1) cutadapt (Martin, 2011) to remove LACE-seq 3' linker, (2) UMI-tools `extract` (Smith, 2017) to move the UMI sequence from the read to the read name so that PCR duplicates can be removed after the alignment, (3) cutadapt to remove the T preceding the RPF.

Trimming is done in 3 steps:

>@SEQ:01:0001 read1



1. cutadapt

>@SEQ:01:0001 read1



2. UMI-tools (`extract`)

>@SEQ:01:0001_NNNNNNNN read1



3. cutadapt

>@SEQ:01:0001_NNNNNNNN read1



If cells were treated with CHX and processed with Ribolace module 1, reads between 28 nt and 35 nt are expected to show the best 3nt periodicity.

STEP1: cutadapt

First the LACE-seq linker is trimmed from the 3' end of each read and only reads longer than **X+9** nt are retained, while shorter reads are discarded:

```
cutadapt --minimum-length X+9 -a TCTCCTTGCATAATCACCAACC --discard-untrimmed
-o trim.fastq input.fastq
```

Parameter	Definition
--minimum-length X+9	Reads are retained if they are longer than X+9 nt, where X is the length of the RPF (usually X=20 for ribosome profiling analysis), and 9 is the sum of the lengths of the 5' and 3' UMIs
-a TCTCCTTGCATAATCACCAACC	Removal of the LACE-seq 3' linker and any sequence that may follow
--discard-untrimmed	Reads in which <i>no</i> adapter is found are discarded
-o trim.fastq	The output file name
input.fastq	The input file name

STEP2: UMI-tools (extract)

The sequence of the 5' and 3' UMIs are moved from the read sequence to the read name:

```
umi_tools extract -I trim.fastq --bc-
pattern='^(?P<umi_1>.{4}).+(?P<umi_2>.{4})$' --extract-method=regex -S
extract.fq
```

Parameter	Definition
-I trim.fastq	The input file name must be the same as the output file name in step1
--bc- pattern='^(?P<umi_1>.{4}).+(?P<umi_2>.{4})\$'	extract the first 4 (5'UMI) and the last 4 bases (3'UMI) of each read
-S extract.fq	The output file name

NOTE: UMI-tools dedup can be used after alignment to remove duplicates based on the mapping coordinate and the UMI attached to the read name.

STEP3: cutadapt

The T preceding the RPF is then removed:

```
cutadapt -g ^T --discard-untrimmed -o trim2.fastq extract.fastq
```

Parameter	Definition
-g ^T	Removal of the first T at the start of each read
-o trim2.fastq	The output file name
extract.fastq	The input file name must be the same as the output file name in step2