

SMARTer[®] smRNA-Seq Kit for Illumina[®]

Cat. Nos. 635029, 635030, 635031

Supplemental product information and tips for success

that's
GOOD
science![®]

Introduction

The purpose of this document is to supplement the user manual for the SMARTer smRNA-Seq Kit for Illumina in order to maximize customer success. This document may be updated periodically, as needed.

Included in this document are the following:

- Example electropherogram profiles of total RNA inputs and sequencing libraries generated with the kit (pg. 3–7)
- Insights and examples concerning the expected sequencing performance of libraries generated with the kit (pg. 8–11)
- Recommendations for data processing and analysis (pg. 12–14)

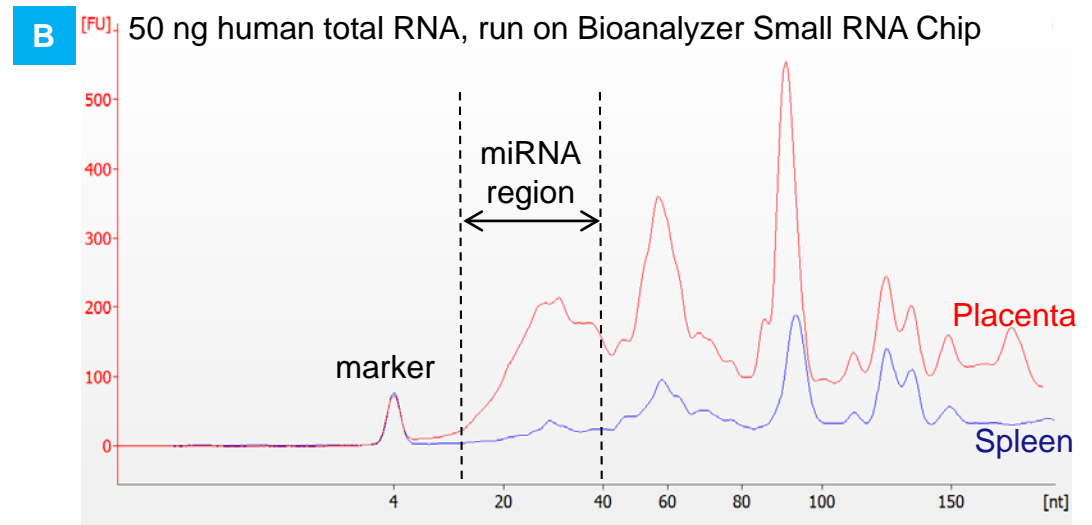
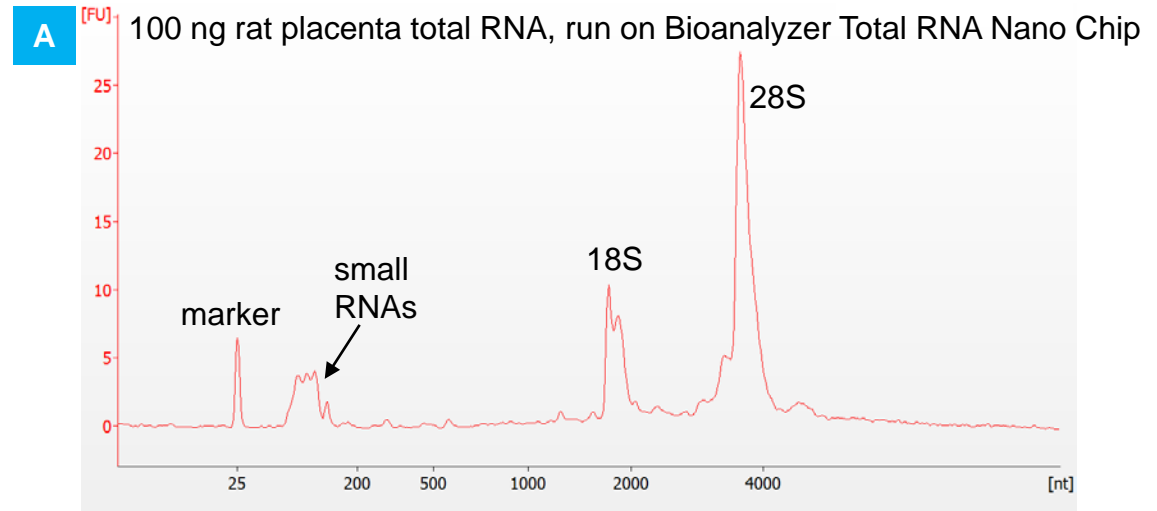
RNA profiles

Example electropherogram results for total RNA inputs. Major features of total RNA profiles generated on the Bioanalyzer Total RNA Nano Chip¹ (**Panel A**) are two main peaks corresponding to ribosomal small and large subunits (18S and 28S in the case of mammalian RNA). However, a fraction of small RNAs—more or less prominent depending on the tissue—can also be detected. This small RNA fraction, comprised of small ribosomal RNAs (5S and 5.8S), tRNAs, snoRNAs, miRNAs, etc., can be more precisely evaluated on the Bioanalyzer Small RNA Chip² (**Panel B**).

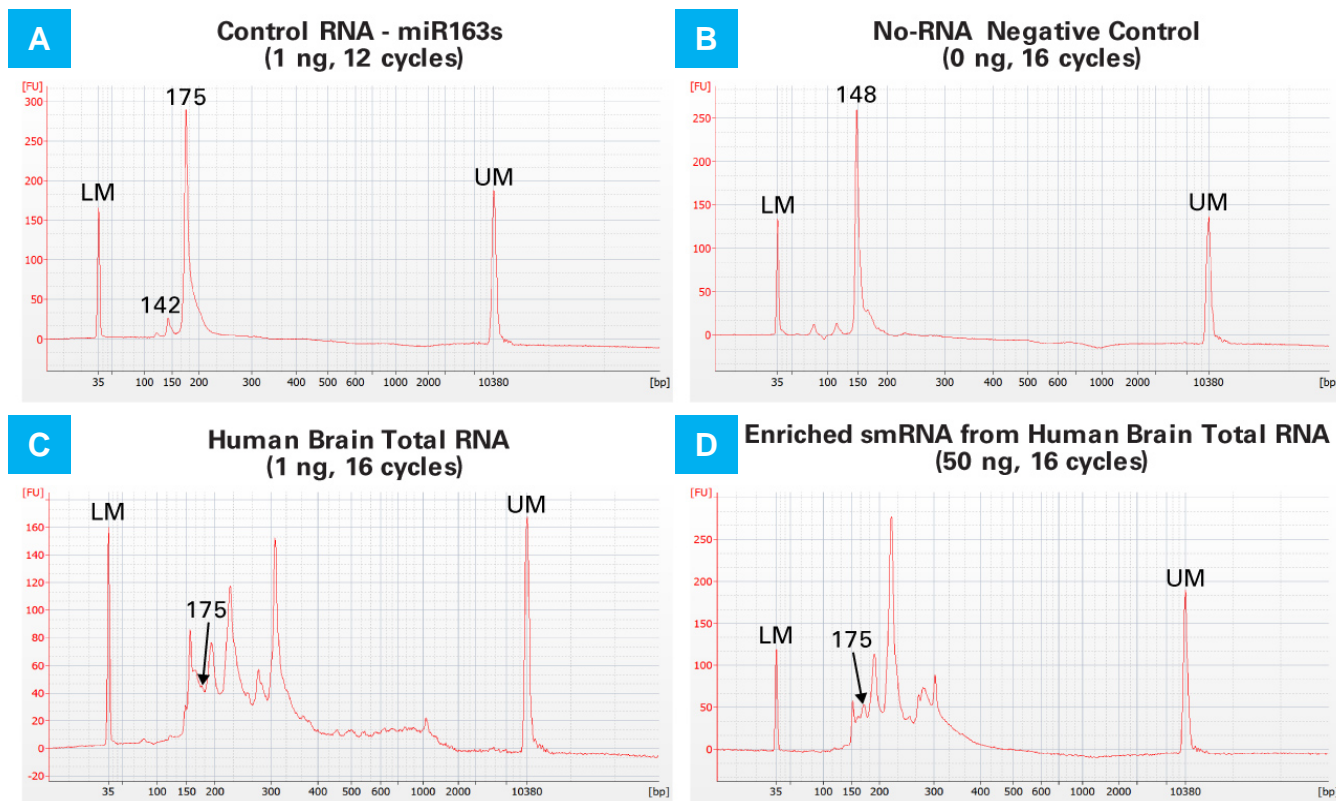
The SMARTer smRNA-Seq Kit for Illumina efficiently captures small RNA fractions in sequencing-ready libraries. See examples of small RNA library profiles in the next three slides.

¹Agilent, Cat. No. 5067-1511

²Agilent, Cat. No. 5067-1548

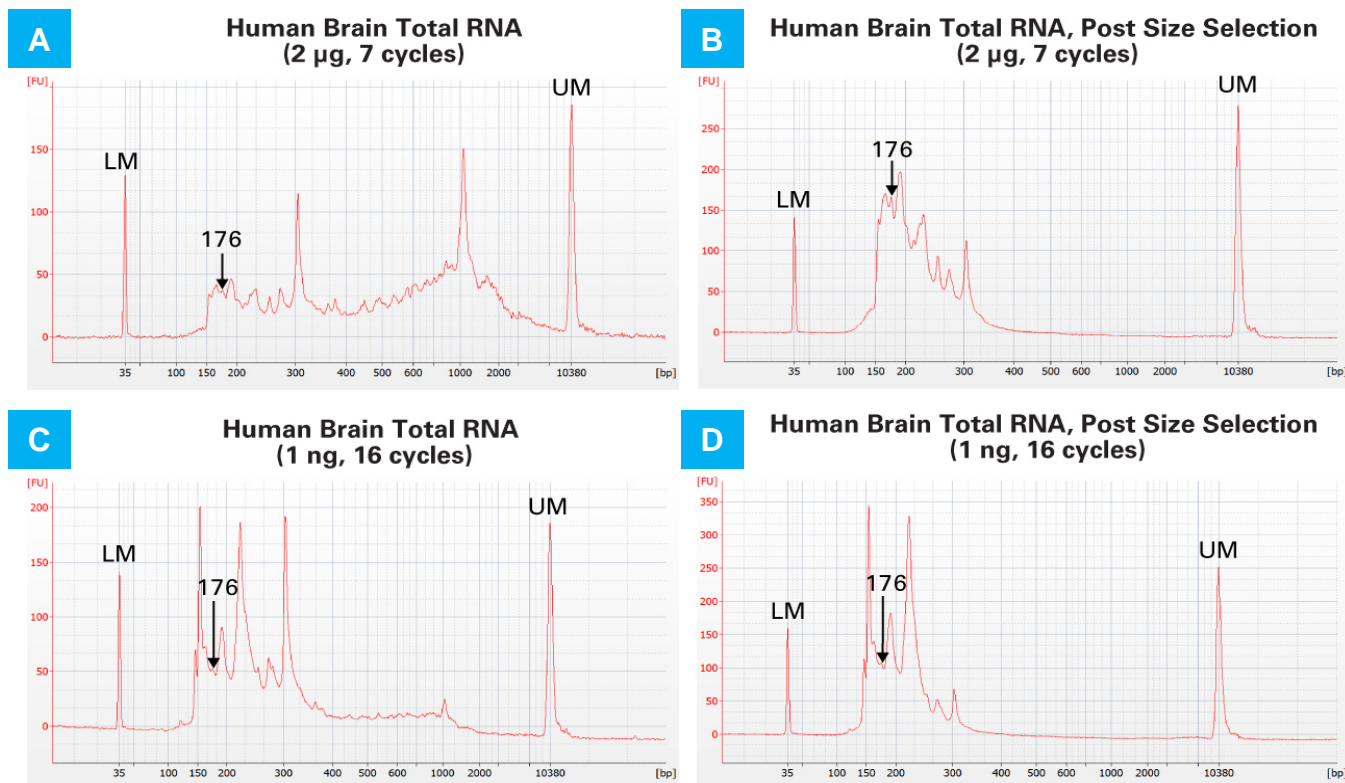


Library profiles – without size selection



Example electropherogram results for libraries generated with the SMARTer smRNA-Seq Kit for Illumina. Libraries were analyzed on an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit (Agilent, Cat. No. 5067-4626). Peaks labeled “LM” and “UM” correspond to DNA reference markers included in each analysis. **Panel A.** Typical result for Control RNA - miR163s. The peak at 175 bp corresponds with the combined size of miR163s plus adapters, while the peak at 142 bp corresponds with the size of primer dimers. **Panel B.** Typical result for a no-RNA negative control. The peak at 148 bp corresponds with the size of adapter dimers. **Panel C.** Example profile for an input consisting of 1 ng of total RNA. **Panel D.** Example profile for an input consisting of 50 ng of enriched smRNA. In **Panels C and D**, the peak at 175 bp corresponds with the combined size of processed miRNAs plus adapters.

Library profiles – gel-free size selection

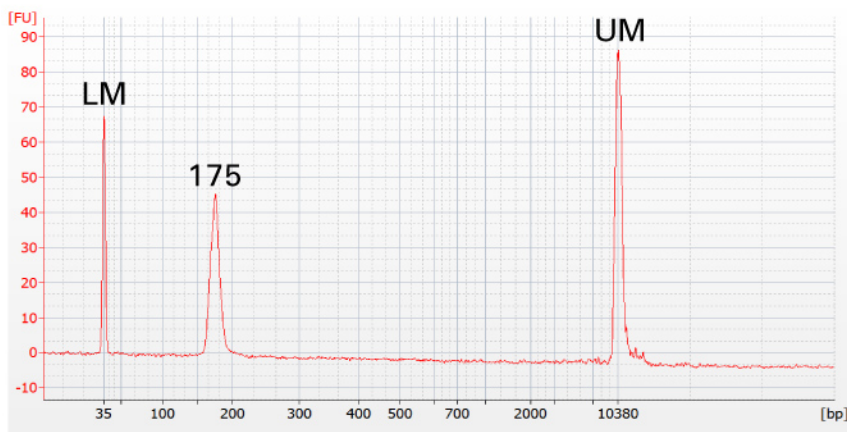


Example electropherogram results for smRNA-seq libraries before and after bead-based size selection. Libraries were analyzed on an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit. Peaks labeled “LM” and “UM” correspond to DNA reference markers included in each analysis. **Panel A.** Library generated from 2 µg of Human Brain Total RNA, prior to size selection. **Panel B.** Library from Panel A following bead-based size selection. **Panel C.** Library generated from 1 ng of Human Brain Total RNA, prior to size selection. **Panel D.** Library from Panel C following bead-based size selection. In all panels, the peak at 176 bp corresponds with the combined size of processed miRNAs plus adapters.

Library profiles – BluePippin size selection

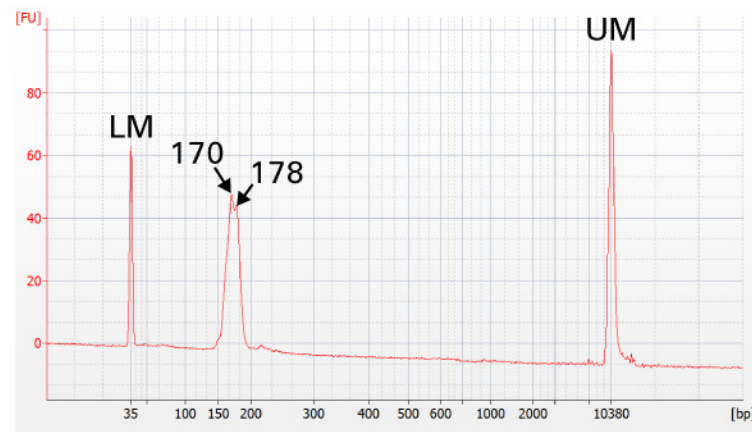
A

Human Brain Total RNA, Post Size Selection
(1 ng, 16 cycles)



B

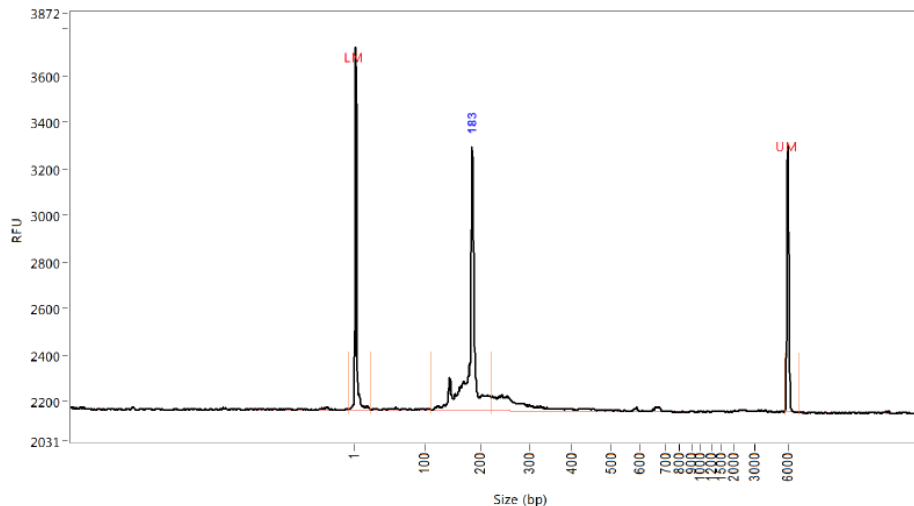
Human Brain Total RNA, Post Size Selection
(1 ng, 16 cycles)



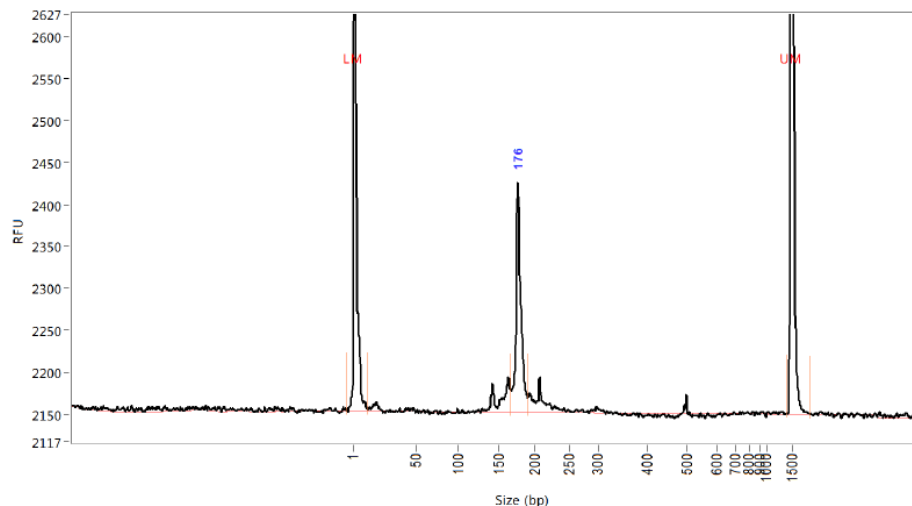
Example electropherogram results for smRNA-seq libraries after size selection with the BluePippin system. Libraries were generated using the SMARTer smRNA-Seq Kit for Illumina with the indicated inputs and cycling parameters, and following size selection with the BluePippin system, were analyzed on an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit. Peaks labeled “LM” and “UM” correspond to DNA reference markers included in each analysis. When libraries are selected using the specified range (148–185 bp), profiles exhibit a narrow peak extending from 150 bp to 200 bp, and either a single peak (**Panel A**), or a double peak (**Panel B**) at ~175 bp due to minor variation in size-selection efficiency.

Profiling smRNA libraries using a Fragment Analyzer

A High Sensitivity NGS Fragment Analysis Kit



B High Sensitivity Small Fragment Analysis Kit



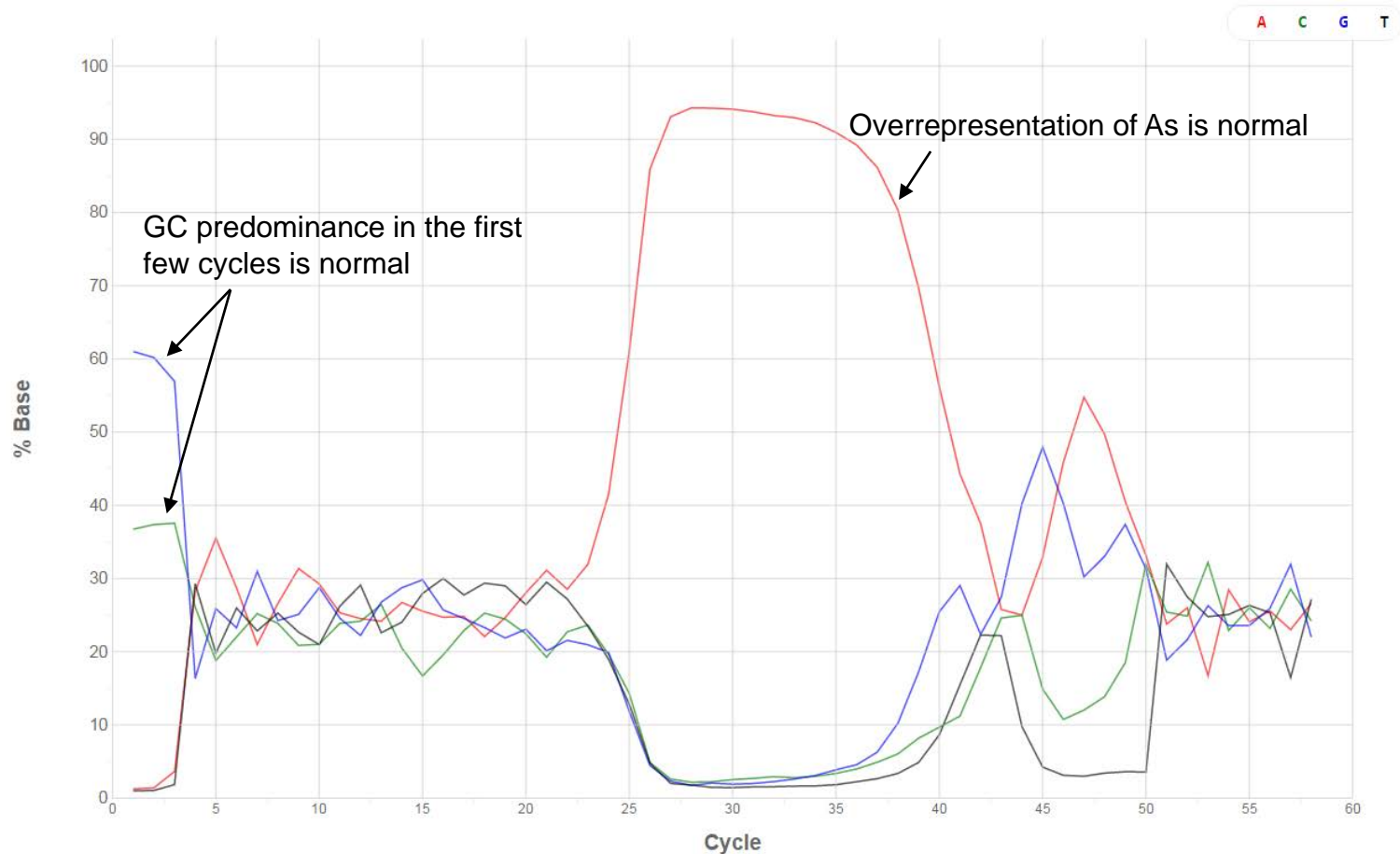
Control RNA – miR163s (1 ng, 12 cycles)

Example electropherogram results for smRNA-seq libraries profiled on a Fragment Analyzer. For profiling sequencing libraries on the Fragment Analyzer (Advanced Analytical), we only recommend the High Sensitivity Small Fragment Analysis Kit (Advanced Analytical, Cat. No. DNF-477), as it provides the most accurate size information for libraries generated with the SMARTer smRNA-Seq Kit for Illumina. **Panel A.** Typical result for library generated from Control RNA - miR163s, analyzed with the High Sensitivity NGS Fragment Analysis Kit. While the expected size of miR163s plus adapters is 175 bp, analysis with this kit typically yields a major peak that is ~10 bp larger than expected (in this example the peak is at 183 bp). **Panel B.** Analysis of the library from Panel A using the High Sensitivity Small Fragment Analysis Kit yields a major peak at 176 bp, which is much closer to the expected size of miR163s plus adapters (175 bp).

Expected sequencing performance

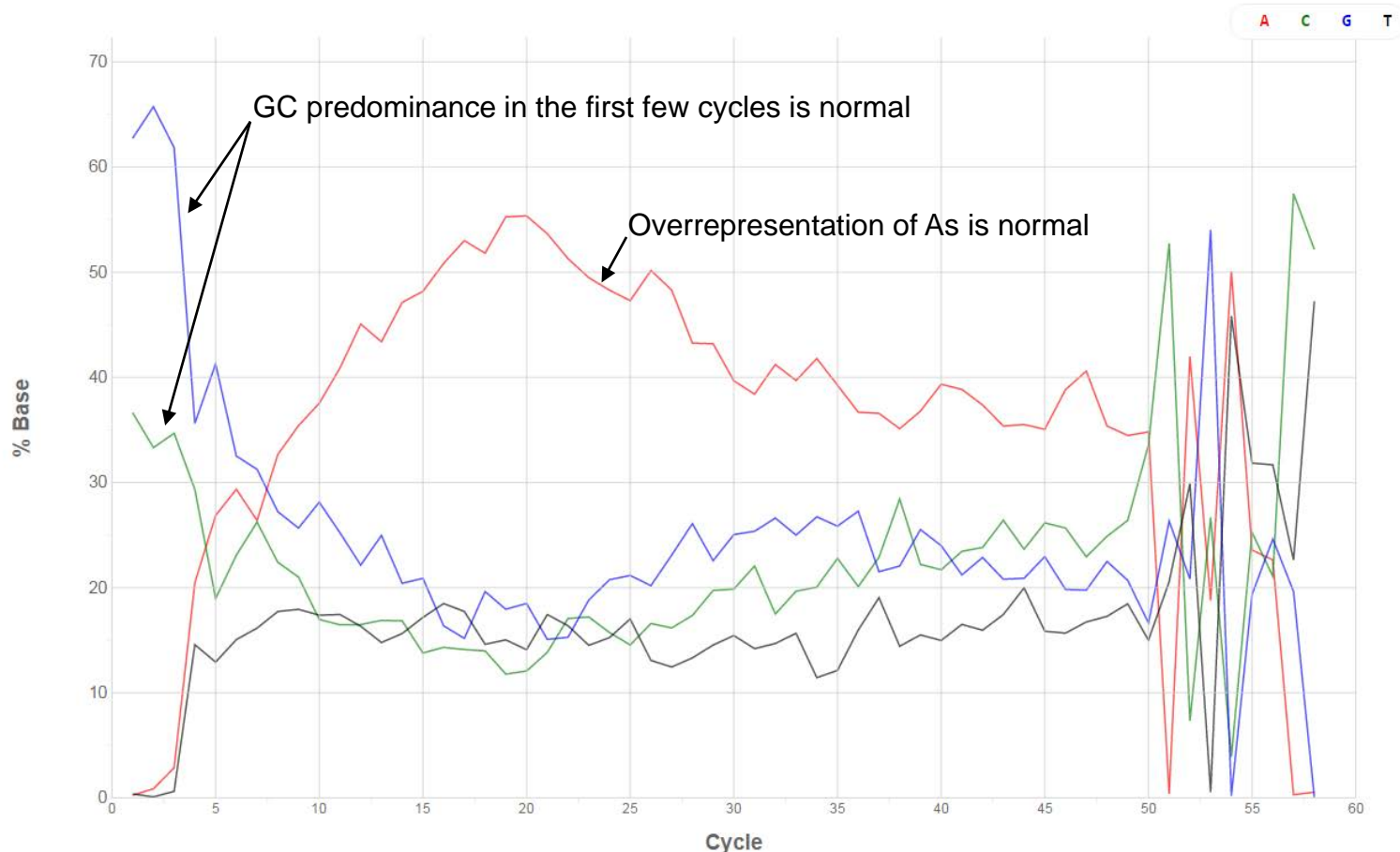
- Kit validation was mostly performed on a MiSeq® instrument using v2 chemistry and 1 x 50 bp reads.
- Due to the disproportionate amount of As introduced during library preparation, the expected Passing Filter (PF) rate for SMARTer smRNA-Seq libraries sequenced on a MiSeq instrument is typically between 70% and 80% (up to 85% has been observed).
- On a MiSeq instrument we have found that inclusion of a PhiX spike-in at concentrations of 10–20% does not improve sequencing performance. However, it is still good practice to include ~5% PhiX in each sequencing run.
- We have found that aiming for a lower than recommended clustering does not improve the PF rate. In our hands, a slightly higher than recommended clustering (e.g. ~1,000–1,100 k/mm² on a MiSeq instrument using v2 chemistry) typically yields the expected 12–15 million reads PF.
- The elevated frequency of As may also affect the error rate, sometimes leading to lower %≥Q30 scores than are typically obtained with well-balanced libraries.
- The following three slides include examples of sequencing runs performed on a MiSeq instrument.
- For a well-defined and high-quality artificial RNA sample like the miRXplore library (next slide), all RNA molecules are similarly sized (~21–23 nt), resulting in an overrepresentation of As starting around cycle 24. However, real-world samples (slides 10 and 11) typically contain RNA molecules of diverse sizes, including some <15 nt, which results in overrepresentation of As starting around cycles 8–10.

Example run: miRXplore libraries



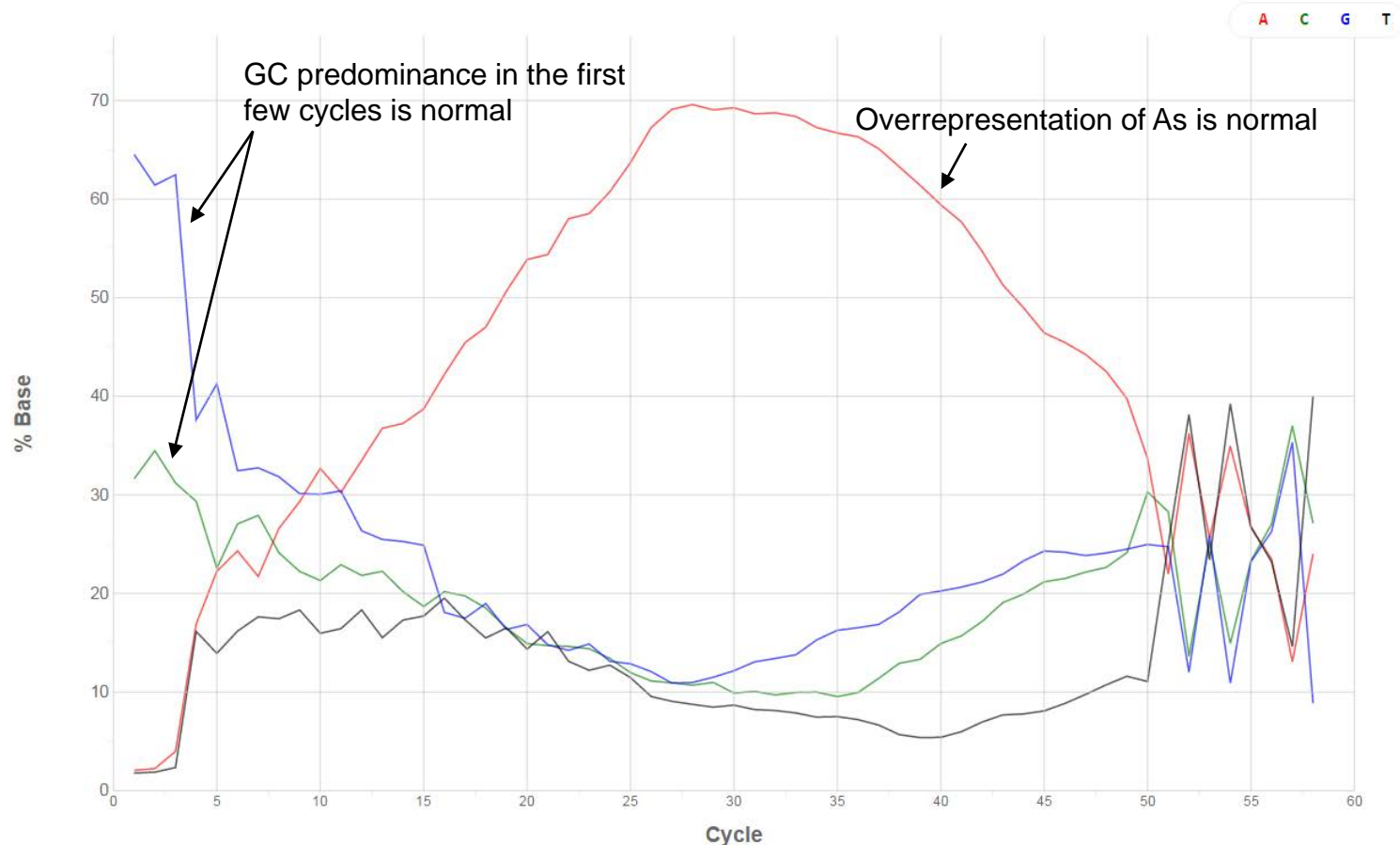
Percentage base calling from a typical MiSeq run performed on libraries generated from the miRXplore Universal Reference (pool of artificial miRNAs). Libraries were prepared without size selection. The proportions of A:T:G:C for each cycle are presented in graph form using software included on Illumina's BaseSpace platform. Cycles 1–50 correspond to Read1, while cycles 51–58 correspond to the library barcode. Note that overrepresentation of As begins around cycle 24 because input RNAs are all ~21–23 nt in length.

Example run: bead-selected libraries



Percentage base calling from a typical MiSeq run following gel-free (AMPure beads) size selection. Size selection was performed as described in Section V.E (Size Selection Using Agencourt AMPure XP Beads) of the SMARTer smRNA-Seq Kit for Illumina User Manual. The proportions of A:T:G:C for each cycle are presented in graph form using software included on Illumina's BaseSpace platform. Cycles 1–50 correspond to Read1, while cycles 51–58 correspond to the library barcode. Note that overrepresentation of As begins around cycles 8–10 due to the inclusion of input RNAs <15 nt in length.

Example run: BluePippin-selected libraries



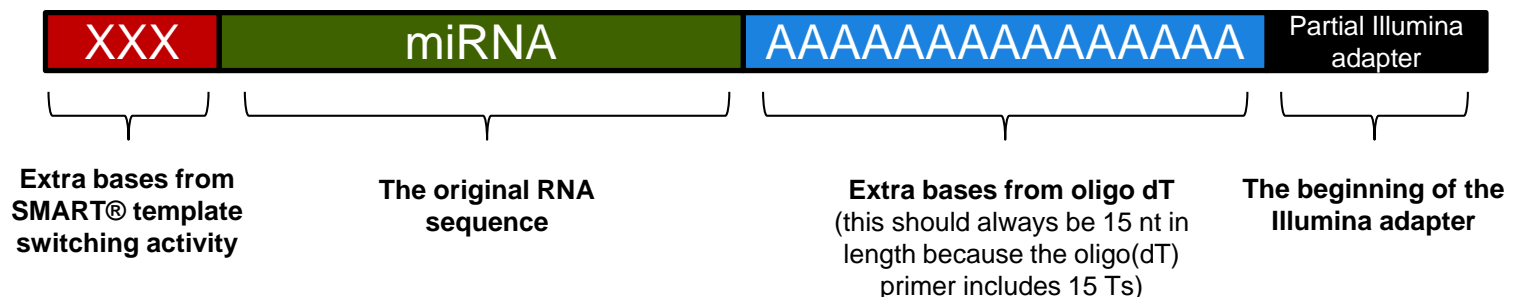
Percentage base calling from a typical MiSeq run following BluePippin size selection. Size selection was performed as described in Section V.F (Size Selection Using the BluePippin System) of the SMARTer smRNA-Seq Kit for Illumina User Manual. The proportions of A:T:G:C for each cycle are presented in graph form using software included on Illumina's BaseSpace platform. Cycles 1–50 correspond to Read1, while cycles 51–58 correspond to the library barcode. Note that overrepresentation of As begins around cycles 8–10 due to the inclusion of input RNAs <15 nt in length.

Limitations for data analysis

Unambiguous determination of input RNA termini is not possible:

- 3': Not possible to tell if RNA ends with an "A" (due to the use of dT priming—the oligo(dT) primer cannot distinguish between naturally occurring and artificially added As at RNA termini)
- 5': Not possible to identify with 100% confidence the beginning of the 5' sequence (at low frequencies, template switching can add more than 3 nt to cDNA 5' ends)

Example of a 50 bp read :



Trimming of reads

Proper trimming of the reads is important for efficient mapping.

One of the tools that can be used with a command-line interface is *cutadapt* (Martin, 2011):

<https://cutadapt.readthedocs.org/en/stable/index.html>

Here we provide some guidance on the use of this particular tool. The following command works well for the SMARTer smRNA-Seq Kit for Illumina:

```
cutadapt -m 15 -u 3 -a AAAAAAAAAA input.fastq > output.fastq
```

Command explanations:

-m 15: after trimming, only reads 15 nt or longer will be retained, while shorter ones will be discarded.

Use **-m 18** instead if a more stringent cutoff is desired.

-u 3: trim the first 3 nt of all reads—this takes care of removing extra nt inserted due to the SMART template-switching mechanism

-a AAAAAAAAAA: adapter sequence to be identified and removed. The adapter sequence and everything 3' of the adapter (including any Illumina adapter sequence) will be removed. By default, cutadapt will accept an error rate of 10%.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.*, 17(1):10–12.

<http://dx.doi.org/10.14806/ej.17.1.200>. See also <https://cutadapt.readthedocs.org/en/stable/index.html>

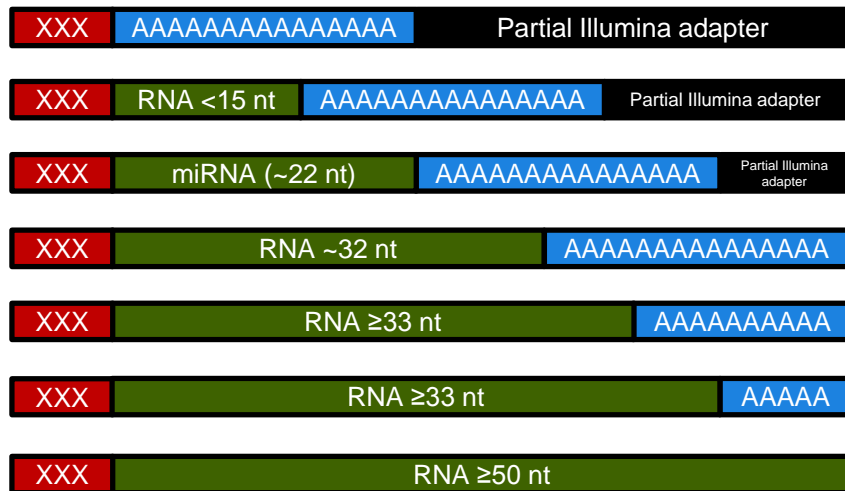
Example reads before and after trimming

Because reverse transcription involves dT-priming on artificial poly(A) tails, a good proportion of reads will include stretches of As (particularly when 20–25 bp-sized library inserts are selected with the BluePippin system). Below is a schematic representation of what the reads may look like, before and after trimming. It is common practice to discard reads shorter than 15 nt after trimming.

Before trimming (range of possibilities)

Position within read:

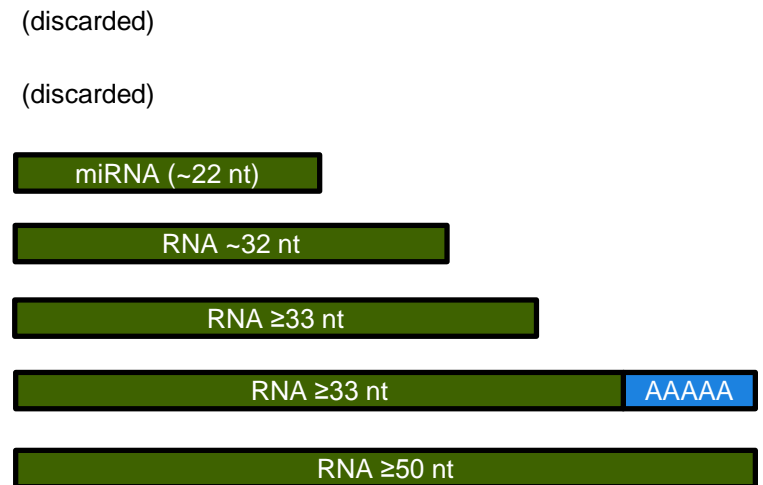
1 4 50



After trimming

Position within read:

1 4 50



that's
GOOD
science!®