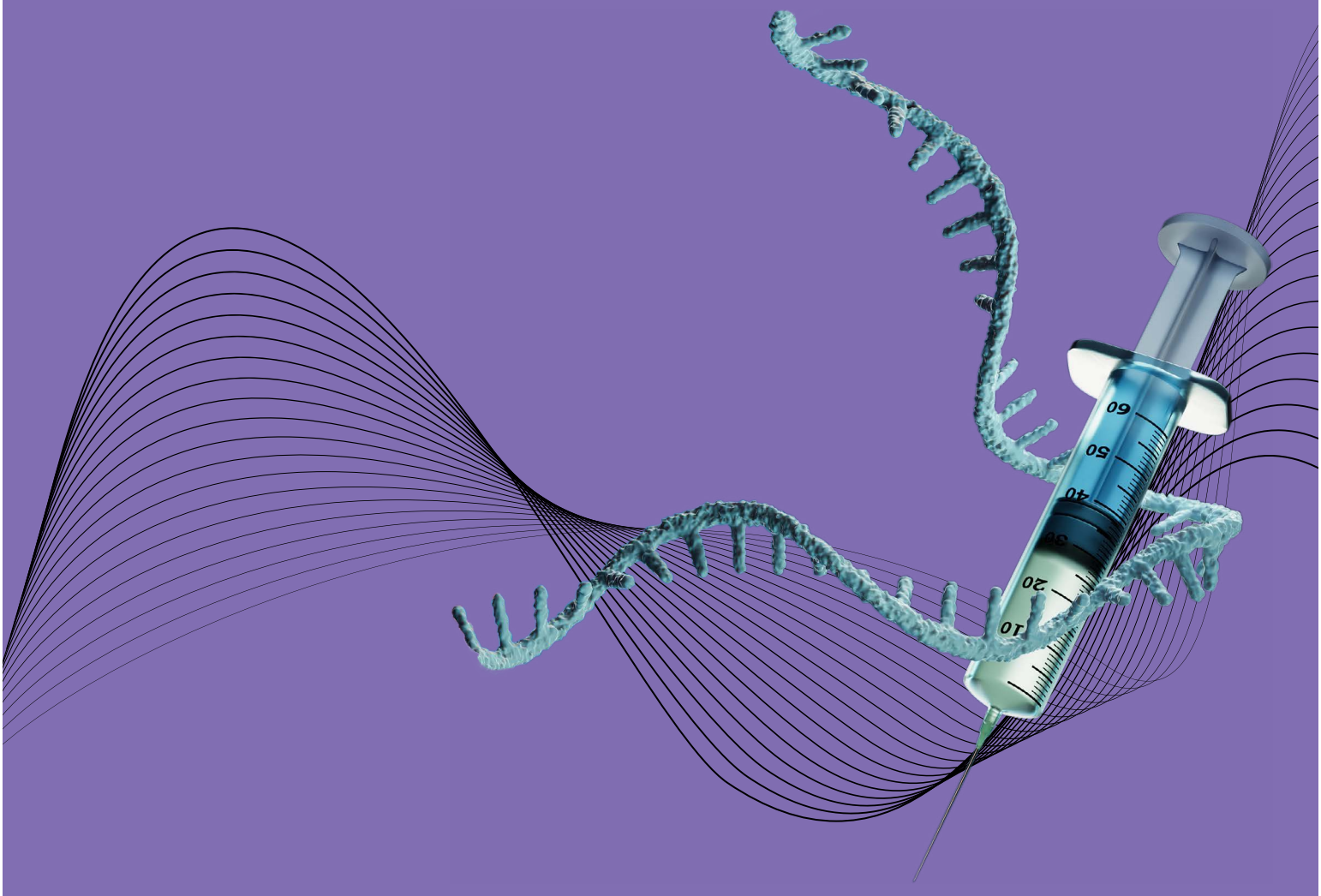
Genomics & Analytical Services

TECH NOTE

Full-Length RNA-Seq: A Novel Method to Assess Sequence Integrity for RNA Therapeutics



AZENTA
LIFE SCIENCES

azenta.com

Introduction

RNA therapeutics is a rapidly expanding area of gene therapy; RNA-based therapies can modulate cells in a controlled and targeted manner, effectively acting as gene therapies. RNA can be used to introduce exogenous copies of new transcripts, where it can be translated into a protein that can work as an enzyme or antigen or regulate signaling pathways. Compared to existing methods, such as reverse transcription polymerase chain reaction (RT-PCR) or reverse transcription quantitative polymerase chain reaction (RT-qPCR), next generation sequencing (NGS) offers an effective, high-throughput approach for monitoring full-length RNA therapeutic quality and purity.

The Challenge

As with any therapeutic, RNA therapeutics must undergo strict quality control (QC) to ensure the product is exactly as specified, with no errors or alternate products. In addition to determining the overall fidelity of the RNA therapeutic, the poly(A) tail length is critical for stability and translation of the mRNA product. Existing poly(A) tail length assays have limitations when detecting the length of mixed populations, including intermediates, and can be prone to introducing artifacts that prevent accurate and precise quantitation (Table 1). The ideal QC approach will allow for quantitative and qualitative evaluation of full-length products, as well as any intermediates and byproducts that may be present, allowing for comprehensive resolution of all products present and full-length sequence confirmation.

	Full-Length RNA-Seq	Standard PacBio® Iso-Seq	Illumina® Poly(A) Tail Assays
Single Full-Length Molecule Resolution	✓	✓	—
Poly(A) Tail Length Counting	✓	—	✓
Detection of Truncated RNA Products	✓	—	—

Table 1. Comparison of RNA assays

The Solution

Azenta Life Sciences has developed a novel approach to sequence full-length, manufactured RNA molecules using PacBio long-read technology. Standard PacBio isoform sequencing (Iso-Seq) applies oligo(dT) primer during reverse transcription, resulting in the loss of two aspects of the mRNA template. First, the overall poly(A) length distribution is not captured, and secondly, the standard Iso-Seq process is poly(A)-dependent, thus mRNA products without poly(A) tails are not detected. Our novel approach involves 3' adapter ligation and template switching during reverse transcription to capture all products (Figure 1) within the sample, while maintaining the intact full-length poly(A) tail, if present.

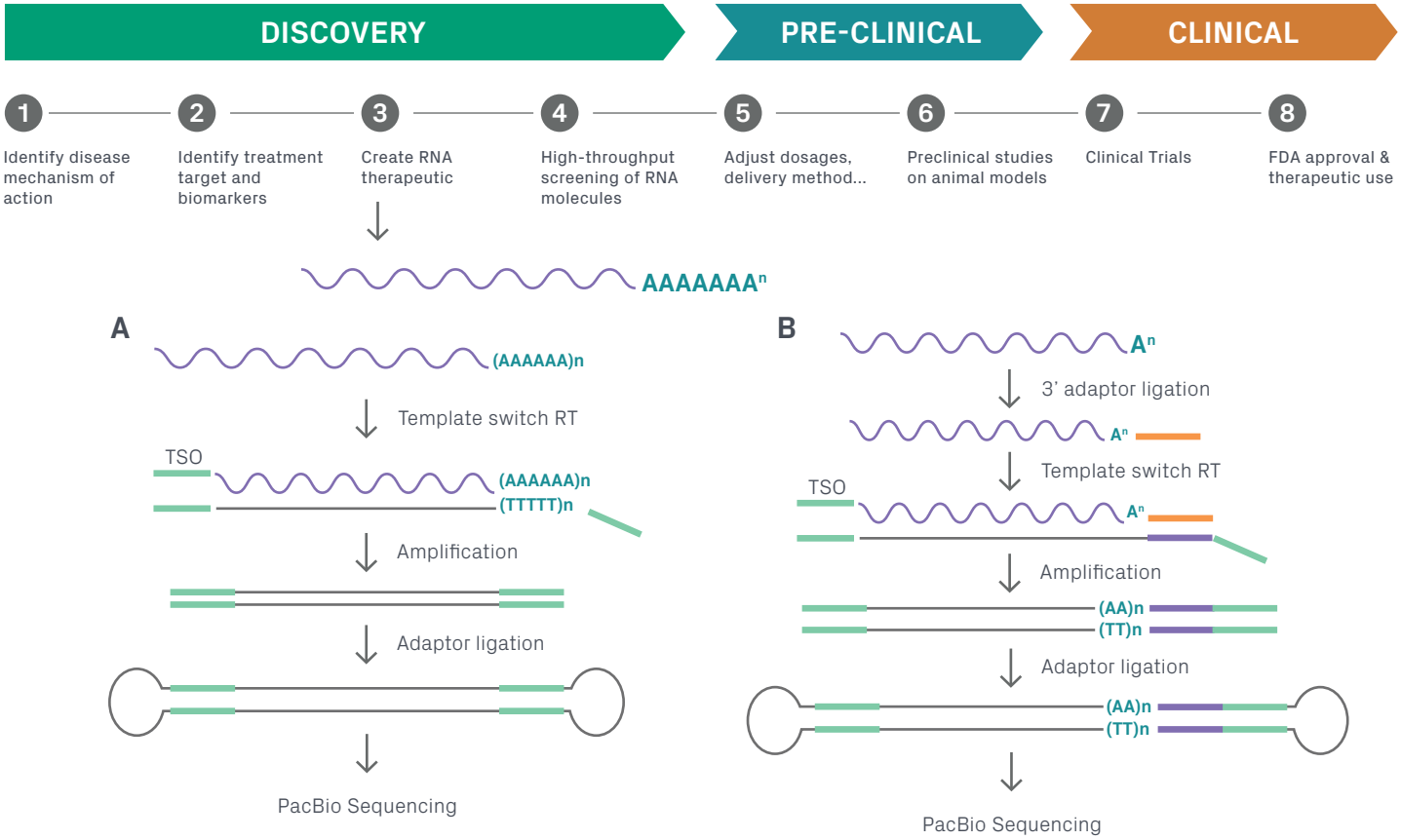


Figure 1. Validate and confirm RNA therapeutic product using a modification of Iso-Seq. A. Workflow for Iso-Seq. B. Workflow for full-length RNA-Seq. Full-length sequencing of RNA molecules determines the exact length of the poly(A) tail by deep sequencing to validate therapeutic purity and accuracy. This method also allows for sequencing of intermediate and alternative products.

Following amplification and adapter ligation, the library is sequenced using PacBio Single-Molecule, Real-Time (SMRT®) long-read sequencing technology. This method allows for full, end-to-end sequencing of products up to and beyond 10kb without limitations or bias due to repetitive sequence or C/G content. As the poly(A) tail is not required for adapter ligation, any partial or non-polyadenylated products will also be detected. Crucially, this new method preserves the entire length of the poly(A) tail, enabling sensitive and specific counting of the poly(A) tail length.

The Results

Here, we describe our novel full-length RNA-Seq approach and present final QC results using this methodology. In addition to sequence fidelity, this NGS approach offers read depth sufficient to identify and confirm overall sample purity, including intermediate and alternative products present within the sample. Further, the entire molecule including the full poly(A) tail is sequenced end-to-end to ensure complete fidelity of the manufactured RNA therapeutic.

To confirm consistency and accuracy of full-length RNA sequencing, we examined poly(A) length distribution in plasmid DNA and RNA spike-ins (Figure 2). As shown in Figure 2A, these *in vitro* transcribed RNAs had variable poly(A) tail lengths ranging from no tail to 120 nucleotide poly(A) tail. DNA plasmids with the same structure were subjected to PacBio library preparation and sequencing to serve as a baseline reference (Figure 2B). The RNA spike-in molecules were subjected to 3' adapter ligation and template switching during reverse transcription. The resulting cDNA was then used as template for PacBio library preparation and SMRT sequencing on the Sequel® IIe platform. The normalized PacBio sequencing results of the RNA spike-ins (Figure 2C) are consistent with the DNA plasmids (Figure 2B).

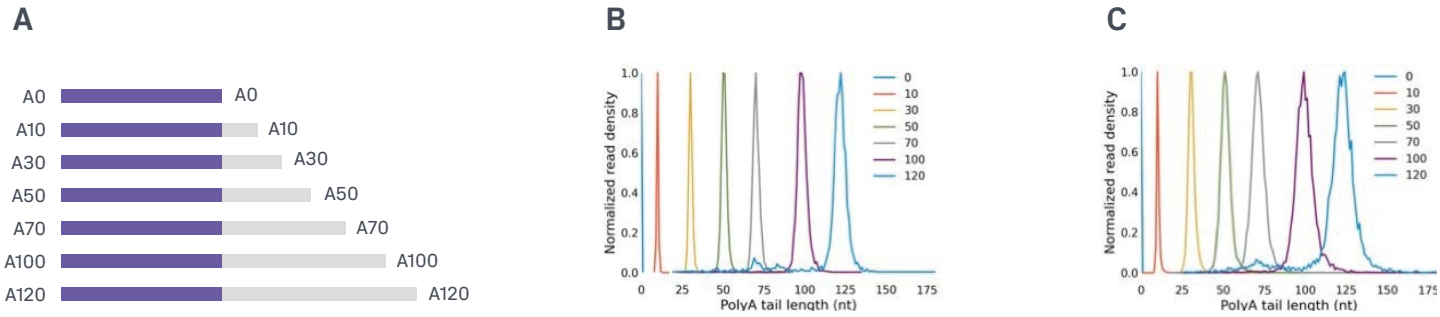


Figure 2. Poly(A) length distribution is consistent between RNA spike-ins and plasmid DNA. RNA spike-ins were produced by *in vitro* transcription using DNA templates with different poly(A) lengths. A. Structures of the RNA spike-ins. B, C. Poly(A) length distribution was accurately determined on both DNA templates (B) and RNA (C).

Next, to demonstrate the ability of full-length RNA-Seq to quantitate the percentage of template molecules with a mixed population of poly(A) tail lengths, libraries were prepared from synthetic RNA molecules with a range of poly(A) tail lengths mixed at pre-defined ratios. Figure 3A indicates the length of the poly(A) tail, the expected amount of this product in the sample pool, and the actual observed following sequencing. The observed percentage of each template closely matches with the expected (Figure 3B, R2=0.98), indicating this method offers highly accurate quantitation of product ratios.

A

Sample	Expected	Observed
10A	10.0%	8.8%
30A	5.0%	8.1%
50A	30.0%	35.7%
70A	20.0%	23.1%
100A	60.0%	55.5%
120A	75.0%	68.8%

B

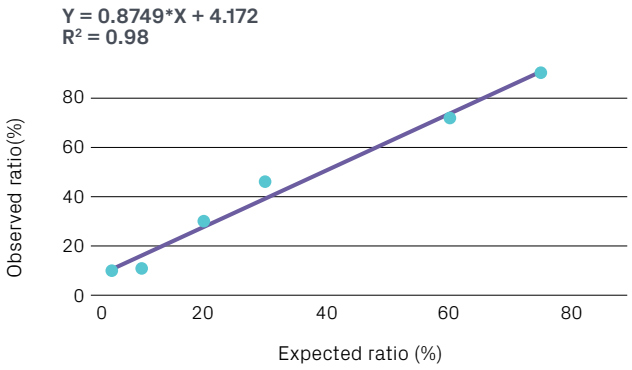
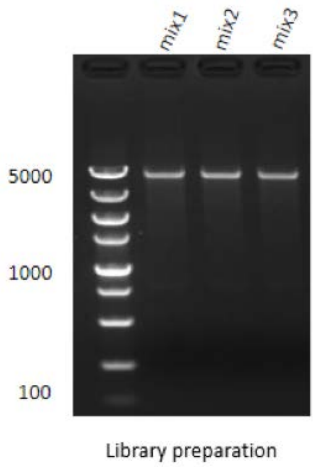


Figure 3. Accurate quantification of poly(A) lengths in an RNA mixture. A. The synthetic RNA molecules shown in Figure 2 were mixed in the ratios indicated in the table. B. The observed percentage of each template closely matches with the expected (R2= 0.98).

Finally, the full-length RNA-Seq method was tested using a multivalent mRNA vaccine mixture. As evident in the gel (Figure 4A), all three of these RNA therapeutics are approximately the same overall length and appear to be pure based on gel imaging. Three different mRNA therapeutic products were pooled in different proportions; the resulting pools were then processed independently through the full-length RNA-Seq workflow. The consequent PacBio sequencing results (Figure 4B) indicate the expected vs. observed ratios of each product. Again, the observed percentage of each template closely matches with the expected (Figure 4C; R2=0.99), indicating this method offers highly accurate quantitation of tail length and product ratios with real-world samples.

A



B

Mix A	Expected	Observed
Spike A	33.3%	33.0%
Spike B	33.3%	33.0%
Spike C	33.3%	34.0%

Mix B	Expected	Observed
Spike A	10.0%	10.0%
Spike B	30.0%	30.0%
Spike C	60.0%	60.0%

Mix C	Expected	Observed
Spike A	25.0%	24.0%
Spike B	25.0%	25.0%
Spike C	50.0%	51.0%

C

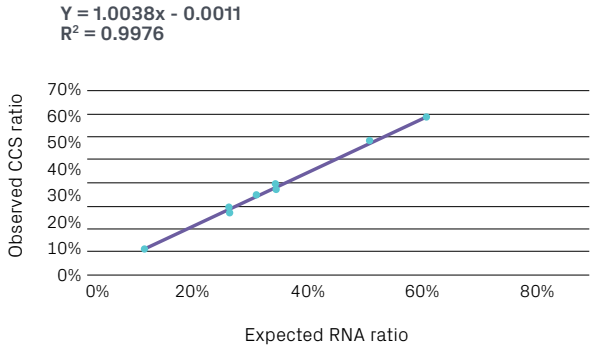


Figure 4. Accurate detection of proportion of each mRNA in pooled multivalent mix. RNA spike-ins were produced by *in vitro* transcription using DNA templates with different poly(A) lengths. A. Structures of the RNA spike-ins. B, C. Poly(A) length distribution was accurately determined on both RNA (B) and DNA templates (C).

The Conclusion

Compared to existing methods, this novel approach to sequence full-length, manufactured RNA molecules using PacBio long-read technology offers an effective, high-throughput approach for monitoring full-length RNA therapeutic quality and purity. It is ideal for quantitative and qualitative evaluation of primary, full-length products, as well as detection and quantification of any intermediates and byproducts that may be present. This stringency is required for mRNA sequence identity, integrity, and purity QC of mRNA drug products and substances. This method can be used to determine the proportion and full sequence identity of each mRNA, including poly(A) tail length, and is optimal for QC of monovalent and multivalent, mRNA-based vaccines or drugs.

To learn more about full-length RNA-Seq and how it can help you ensure the integrity of your mRNA products for RNA therapeutics, submit an inquiry to speak with a technical expert.

SPEAK WITH AN EXPERT

AZENTA
LIFE SCIENCES

azenta.com