How does sequencing RNA work?

Researchers and drug developers use RNA sequencing to generate digital data from RNA molecules in a biological sample, revealing which RNAs are present. This data is analyzed using bioinformatics tools to understand gene activity and RNA behavior, providing insights that are essential for studying disease biology and designing effective RNA-based therapeutics.

There are a number of sequencing-based assays developed to study RNA, and they can be roughly divided into two approaches: short-read and long-read.

Each type of sequencing has different analytical strengths, and when used together, they can offer a comprehensive view of RNA biology.

Short-read sequencing

Short-read sequencing requires researchers to fragment RNA into smaller pieces to sequence. This type of sequencing is commonly used by Illumina sequencers and is also called fragment-based sequencing.

To perform short-read RNA sequencing, the RNA needs to be isolated and broken into short, 50 – 500 nucleotide-long fragments.

Next, reverse transcription is performed to create complimentary DNA (cDNA) of the fragments. The cDNA is then amplified, and the different fragments are analyzed by a sequencer to generate sequencing reads, where each read lists the bases identified in that fragment.

The fragments are then compared to a reference sequence, such as the sequence of an in vitro transcribed (IVT) RNA or the human genome, to determine where each fragment maps within the original RNA sample.

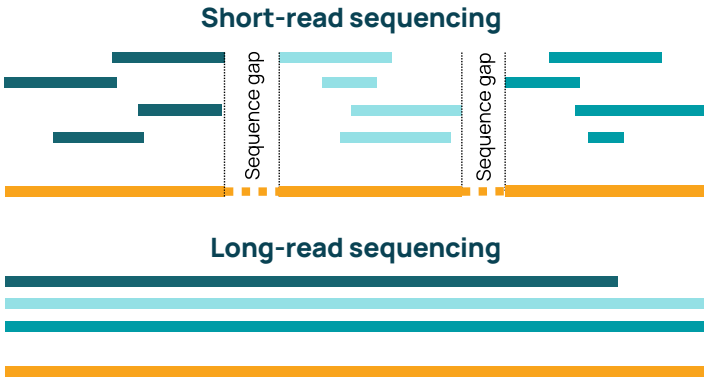
Short-read sequencing is typically highly accurate for single-nucleotide level analysis, but it requires the computational assembly of fragments together since the full length of RNA isn't sequenced in one read. In addition, the small sizes can make it difficult to map a specific fragment uniquely to larger repetitive regions of RNA.

Long-read sequencing

Long-read sequencing overcomes these limitations by allowing scientists to directly sequence an entire RNA in one read. Unlike short-read sequencing, scientists do not have to fragment the RNA strand, and reverse transcription into cDNA is optional.

One example of long-read sequencing was developed by Oxford Nanopore Technology, where RNA molecules are fed through a nanopore that has an electric current flowing across it. Each nucleotide base of the RNA produces a distinct signal that is recorded and interpreted by a basecalling algorithm to identify the correct base.

Since the RNA remains full-length, scientists can gather information about the entire strand, including features that are difficult to measure with short-read sequencing. For example, they can identify the whole length of the poly(A) tail, repetitive regions, and different isoforms of genes with long-read sequencing.



Sequencing at Eclipsebio

At Eclipsebio, we use sequencing across our portfolio of capabilities, including both short- and long-read sequencing.

With sequencing, our partners can deeply characterize RNA therapeutics, including the detection of dsRNA impurities and translation dynamics. Our academic partners can use sequencing to gather data on the full spectrum of RNA biology, including where and how miRNAs and regulating proteins bind.

Interested in using sequencing to gain deeper insights into RNA biology? Reach out to us to learn more.