Tercer Ejercicio Fase Oposición TECNICOS SUPERIORES ESPECIALIZADOS OEP/2018-19.
TRIBUNAL nº 4. GENOMICA
Traducir todo el texto excepto Figura 1, pie de figura, Tabla 1 y pie de tabla.

Best practices for read trimming for Illumina Stranded mRNA and Total RNA workflows

Explore the impact of the T-overhang on sequence read quality and options for read trimming.

Introduction

RNA sequencing (RNA-Seq) with next-generation sequencing (NGS) is a powerful method for discovering, profiling, and quantifying RNA transcripts. Advances in the Illumina portfolio of RNA library preparation kits deliver the high-quality data researchers require, with a streamlined workflow. Illumina offers three RNA library prep kits:

- Illumina Stranded mRNA Prep, Ligation provides a cost-efficient option for coding RNA-focused analyses.
- Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero™ Plus enables whole-transcriptome analysis, capturing coding and multiple forms of noncoding RNA.
- Illumina RNA Prep with Enrichment brings bead-linked transposome (BLT) technology to RNA enrichment.

Illumina Stranded mRNA and Total RNA Prep kits feature innovations to streamline the ligation-based library preparation chemistry, supporting increased throughput by multiplexing up to 384 unique dual indexes (UDIs) in a single reaction. After cDNA synthesis, double-stranded DNA fragments undergo "A-tailing", in which a deoxyadenosine nucleotide is added to the 3´ end. This enables rapid ligation of sequencing adapters designed with 3´ T-overhangs, not present in previous ligation-based library prep kits (Figure 1A).

A by-product of this approach is that the first cycle of sequencing Read 1 and Read 2 will be derived from the T-overhang in the adapter and detected as a "T" (Figure 1B), and not from the DNA being sequenced. This may pose a challenge for Real-Time Analysis (RTA) software, as the first base for every cluster on the flow cell will be a "T", resembling a low-diversity sequence. The presence of this low-diversity sequence within the first six cycles of a read may make it more difficult to define monoclonal clusters during image analysis, particularly for two-channel sequencing by synthesis (SBS) chemistry, which is used by the NextSeq™ 550 and NovaSeq™ 6000 Systems.

This technical note describes the effect of the T-overhang on sequencing data quality and recommends best practices for trimming the first base from sequencing reads to minimize the potential impact on downstream RNA-Seq data analysis.*

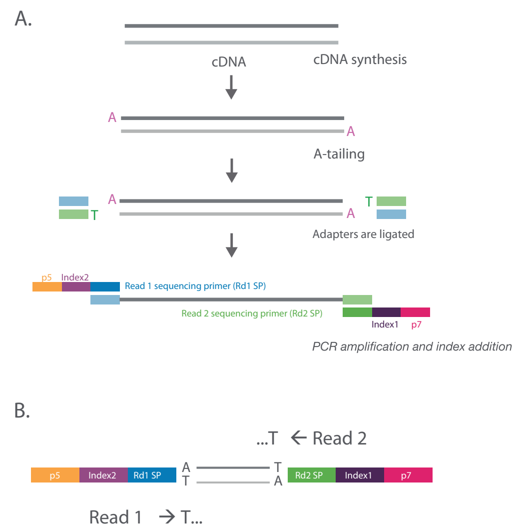

Figure 1: Illumina Stranded RNA library preparation—(A) After cDNA synthesis is complete, ligation of adapters and PCR amplification produces high-quality libraries. (B) The use of T-overhangs in sequencing adapters to facilitate rapid ligation results in all reads starting with a "T" in the first cycle.

Impact of T-overhang on read quality

To explore the impact of the T-overhang on read quality and alignment, a 9-plex pool of Illumina Stranded mRNA Prep libraries were run on the NextSeq 550 and NovaSeq 6000 systems. Analysis of the Q-scores for each cycle showed low-quality calls for the first cycle of Read 1 and Read 2 for the NextSeq 550 and NovaSeq 6000 Systems, as expected (Figure 2). However, analysis of performance metrics across the entire run showed minimal impact on data quality, as measured by the percent passing filter (PF), % ≥ Q30, and yield (Table 1).

Table 1: Impact of T-overhang on performance metrics

System	Read length	% PhiX[a]	% Q30	Yield (Gb)
NextSeq 550 System (v2 chemistry)	2 × 75 bp	0%	91.25%	77.82
NovaSeq 6000 System (S1 flow cell)[b]	2 × 75 bp	0%	91.79%	278.81
NovaSeq 6000 System (S4 flow cell)	2 × 75 bp	0%	94.50%	1890

a. No PhiX was loaded in these sequencing runs to present the flow cell with only the T-overhang in the first read cycle.
b. NovaSeq 6000 v1.0 reagents were used for this sequencing run.