TECHNICAL NOTE

Sequencing Metrics & Base Composition of Chromium Single Cell ATAC Libraries

Introduction

The Chromium Single Cell ATAC Reagent Kits workflow produces Single Cell ATAC libraries for detecting accessible chromatin regions for epigenetic profiling. This Technical Note presents a comparison of sequencing metrics for a Single Cell ATAC library across Illumina® platforms. The expected base percentage profiles and Phred quality scores based on a control library is described to provide general guidance on the expected range of sequencing metrics on Illumina® platforms. Individual results may vary depending on the specific sequencing instrument and/or particular sample and loading characteristics.

Method

A Single Cell ATAC library was generated from one sample (cell line) following the Single Cell ATAC Reagents Kits User guide (Document GC000168). Approximately 500 nuclei were targeted using the Chromium Chip E Single Cell to generate a library.

The library fragments contain an accessible chromatin genomic insert, a 10x Barcode, a Chromium sample index N, standard Illumina® paired-end P5 and P7 sequences, and Read 1 (Read 1N) and Read 2 (Read 2N) sequences (Figure 1). The library was quantified using a KAPA DNA Quantification Kit and sequenced with 1% PhiX on various Illumina® sequencers to evaluate performance. The following platforms were tested using the run layout specified in Table 1:

- MiSeq™ - one flowcell
- NextSeq™ 500/550 - one flowcell
- HiSeq™ 2500 (RR) - one flowcell
- HiSeq™ 4000 - one flowcell
- NovaSeq™ - one S2 flowcell

Figure 1. Schematic of final library construct of a Chromium Single Cell ATAC library. The 10x Barcode (16 bp) is sequenced in the i5 index read and the Sample Index N is sequenced in the i7 index read. The accessible chromatin insert is sequenced in Read 1N and Read 2N.

Table 1. Sequencing configuration.

Sequencing Read	Number of Cycles
Read 1N	50 cycles
i7 Index	8 cycles
i5 Index	16 cycles
Read 2N	50 cycles

Data were analyzed using the Cell Ranger ATAC pipeline (1.0), which processes the sequencing data to align reads, detect cells, and identify and annotate open chromatin regions in clusters of cells.

Results

To assess performance across Illumina® platforms, metrics for the same Single Cell ATAC library run on different sequencers are shown in Table 2. The loading concentration and cluster density for each sequencer is specified. The sequencing yield for each run is in line with the expected yield for each sequencing platform.

Q30 quality scores: In general, the Q30 quality scores for read 1 (Read 1N) were higher than those for the sample index read (i7), followed by the quality of the 10x barcode read (i5) and read 2 (Read 2N) (Table 2). Read 2N Q30 scores were lower for the HiSeq™ 2500 RR and HiSeq™ 4000 compared to the other three platforms, but this did not affect the downstream assay performance based on the %≥mapQ30 scores.

Data by Cycle: Two features available from the 'data by cycle' panel in the Illumina® Sequencing Analysis Viewer (SAV) software were assessed between the sequencing platforms. All sequencers tested had similar profiles. Representative examples for the % Bases by cycle (Figure 2A) and quality score of ≥30 (%≥Q30; Figure 2B) distribution are shown for the Illumina® MiSeq™ run data.

The fluctuation in the % Bases across the cycles arises from the different sequence features contained within the Single Cell ATAC libraries, including the 10x Barcode and sequence bias of the Tn5 transposase. Table 3 provides an interpretation of the expected profile for the Single Cell ATAC libraries.

The % of Phred quality scores ≥Q30 across the cycles for the Illumina® MiSeq™ platform are shown in Figure 2B. The Phred quality score reflects the base calling accuracy and is used to determine how much of the data from a given sequencing run can be used. Sequencing runs with lower quality scores can result in a significant portion of reads being unmappable. The index i5 read for the 10x Barcode was lower for the NovaSeq™ run compared to the other platforms, but this did not affect the overall assay performance based on the % ≥mapQ30 scores (Table 2).

In summary, % Bases by cycle and % ≥Q30 quality score distribution showed highly consistent profiles for all 5 sequence platforms.

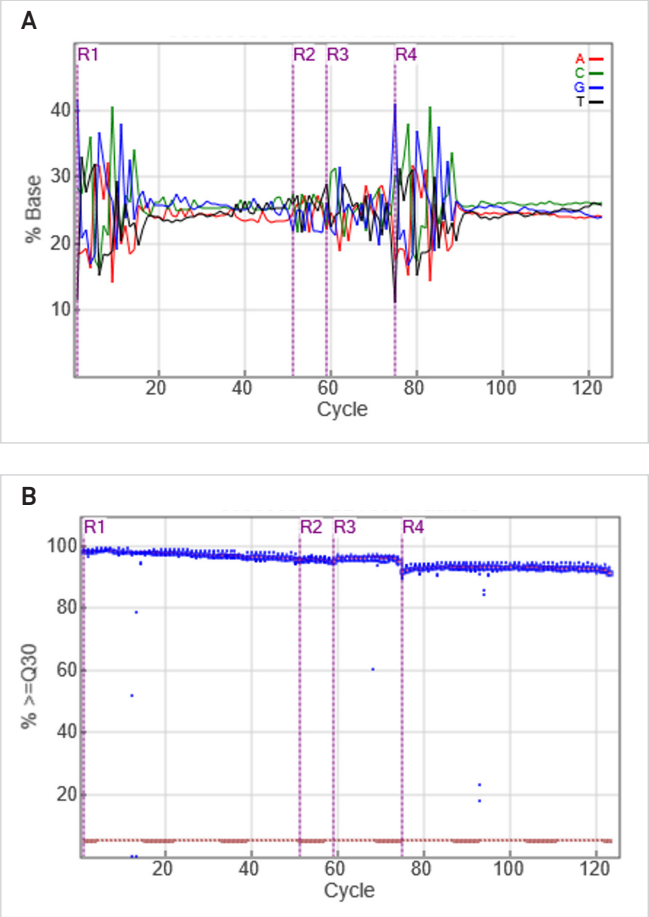

Figure 2. 'Data by Cycle' plot from the Illumina® SAV software displaying the percentage of base calls (A) and Q30 quality scores (B) from a representative Single Cell ATAC library sequenced on Illumina® MiSeq™ using the MiSeq™ Reagent Kit v3.

Conclusion

The sequencing metrics and base composition of sequencing reads generated for Chromium Single Cell ATAC libraries described in this Technical Note serve as guidelines for assessing the quality of Single Cell ATAC library sequencing. Additional factors that may contribute to overall success of a sequencing run and impact downstream application performance metrics include:

- Sample preparation to obtain a high quality single nuclei suspension.
- Final libraries with fragment length of 200-2,000 bp and a significant number of inserts between 200–600 bp for optimal cluster formation on Illumina® flowcells.
- Reliable and accurate library quantification using the KAPA DNA Quantification Kit using the average insert size determined by Agilent Bioanalyzer QC.
- Sequencing platform loading concentration.

References

- Chromium Single Cell ATAC Reagent Kits User Guide (Document CG000168)

Table 2. Sequencing metrics obtained when sequencing the same Single Cell ATAC library across different sequencers.

Instrument	Loading Concentration (pM)	Cluster Density or % PF	Yield per Lane (Gb)		% ≥Q30				% ≥ mapQ30
			Read 1N	Read 2N	Read 1N	i7	i5	Read 2N	
MiSeq™	11	1025 K/mm^2	21.1	1.1	97.2	95.6	95.9	92.9	88.8
NextSeq™ 500	1.7	172 K/mm^2	4.9	4.8	95.4	94.2	91.8	92.0	87.1
HiSeq™ 2500 RR	11	1159 K/mm^2	9.8	9.6	96.0	92.9	91.0	87.1	88.3
HiSeq™ 4000	180	80.60%*	19.1	18.7	98.9	95.0	90.5	88.9	85.7
NovaSeq™	250	80.10%*	113.0	110.7	95.1	89.0	72.2	94.6	87.5

*Percent PF is reported for HiSeq™ 4000 and NovaSeq™ instead of cluster density due to the patterned flowcell.

Table 3. Interpretation of base percentages from Illumina® SAV 'Data by Cycle' plot.

Read Number	Cycle Number	Expected Profile
Read 1N	1-16	Base percentages fluctuate due to the transposase sequence bias.
	17-50	Base percentages reflect the expected base composition of the accessible chromatin in the genome.
i7	51-58	Base percentages fluctuate due to the 8 bp sample index.
i5	59-74	Base percentages fluctuate due to sequences from the 16 bp 10x Barcode. Each base is represented in roughly equal proportions.
Read 2N	75-90	Base percentages fluctuate due to the transposase sequence bias.
	91-124	Base percentages reflect the expected base composition of the accessible chromatin in the genome.

Contact:
support@10xgenomics.com
10x Genomics
7068 Koll Center Parkway
Suite 401
Pleasanton, CA 94566 USA