10x GENOMICS®

TECHNICAL NOTE

Base Composition of Sequencing Reads of Chromium™ Single Cell 3' v2 Libraries

INTRODUCTION

The Chromium™ Single Cell 3' v2 Protocol (CG00052) produces Single Cell 3' libraries, ready for Illumina® sequencing. Single Cell 3' libraries incorporate standard Illumina paired-end constructs with P5 and P7 sequences at opposite ends. The 16bp 10x™ Barcode and the UMI is encoded at the start of Read 1, while sample index sequence information is incorporated into the i7 index read. Read 1 and Read 2 are standard Illumina sequencing primer sites used in paired-end sequencing (Figure 1). This Technical Note describes base percentage profiles and Phred quality scores (shown as %Q30 in Illumina SAV software) that are characteristic of 10x Single Cell 3' v2 libraries for different sequencing run configurations.

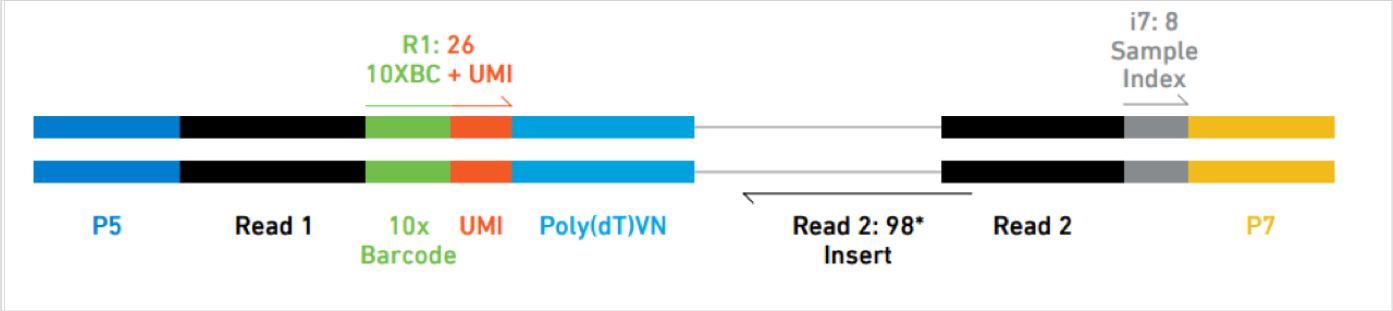

Fig. 1. Schematic of a fragment from a final Chromium Single Cell 3' v2 library. *Can be adjusted.

METHOD

We prepared a Chromium Single Cell 3' v2 library with ~8,400 peripheral blood mononuclear cells (PBMCs) from a healthy donor following the *Chromium Single Cell 3' Reagent Kits v2 User Guide* – CG00052. Raw and processed data is freely available from: https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.0.1/pbmc8k. We sequenced this library on the Illumina HiSeq 4000 in two independent runs with different sequencing run configurations (Table 1).

DISCUSSION

Single Cell 3' v2 libraries are run using paired-end sequencing with a single index read per sample. We recommend the sequencing run parameters listed in Table 1, column 2. Alternative sequencing formats can also be run (Table 1, column 3). However, please note that alternative sequencing formats are only recommended for 10x Single Cell 3' v2 libraries that are sequenced alone or are pooled with other 10x Single Cell 3' v2 libraries for sequencing on the same flow cell lane. We *do not* recommend to sequence 10x Single Cell 3' v2 libraries with a dual-index configuration. We also *do not* recommend to pool 10x Single Cell 3' v2 libraries with 10x Genome/ Exome libraries or other non-10x libraries for sequencing on the same flow cell

lane, as this will compromise sequencing data quality. Specifically, the low diversity past cycle 26 in Read 1 will impact Phred quality scores of non-10x libraries which may impact application performance.

Sequencing Read	Recommended Number of Cycles	Alternative Number of Cycles
Read 1	26 cycles	150 cycles
i7 index	8 cycles	8 cycles
i5 index	0 cycles	0 cycles
Read 2	98 cycles	150 cycles

Table 1. Recommended sequencing run parameter for Chromium Single Cell 3' v2 libraries.

Figure 2 illustrates the distribution of base composition along Read 1(R1 in plot), i7 index read (R2 in plot), and Read 2 (R3 in plot) that we typically observe after a successful sequencing run of a Single Cell 3' v2 library that was prepared according to the Single Cell 3' Reagent Kits v2 User Guide. The profiles are characteristic for Single Cell 3' v2 libraries that are sequenced with the recommended number of cycles and can be explained as follows:

- Cycle 1 to 26 (Read 1 – R1): Base percentages fluctuate due to sequences from the 16bp 10x barcode and the 10bp UMI that is attached to the Gel Bead primers. These non-normally distributed sequences will have a different distribution than the human transcriptome, which causes the apparent shift in percentage of each base.
- Cycle 27 to 34 (i7 index – R2): Base percentages are dissimilar due to sequences from the 8bp sample index.
- Cycle 35 to ~60 (Read 2 – R3): Beginning of the transcript read. Base percentages typically fluctuate due to sequences from the RT Primer at the 5' end of the cDNA from a small subset of library fragments
- Cycle ~60 – 132 (Read 2 – R3): Base percentages are expected to be stable as the actual transcript is sequenced for all library fragments.

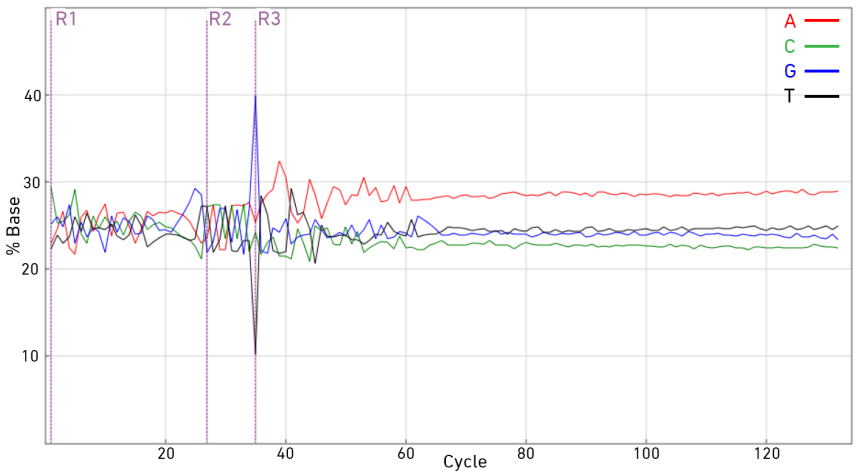

Fig. 2. Representative example of the 'Data by Cycle' plot in the Sequencing Analysis Viewer software (Illumina®). Shown is the percentage of clusters for which the selected base has been called (% base: y axis) along the sequencing length (x axis). Profile is based on sequencing 10x library by itself with no other library type sequenced alongside.

The Phred quality score assesses base calling accuracy and is typically used to determine how much of the data from a given sequencing run can be used. Sequencing data with lower quality scores can result in a significant portion of reads being unusable. Figure 3 and Table 2 outline the Q30 quality metrics that we typically achieve with Single Cell 3' v2 libraries run on the Illumina HiSeq® 4000. Percentages of Q30 are relatively stable across cycles for Read 1 and the i7 Read. The percentages drop at the beginning of Read 2 but recover immediately after a few cycles into the sequencing read.

In some cases, Single Cell 3' v2 libraries may be run with an alternative number of cycles (see Table 1) to accommodate the sequencing requirements of non-10x libraries that are sequenced on the same flow cell. To ensure successful sequencing of both Single Cell 3' v2 and non-10x libraries, we strongly recommend to run these libraries in separate sequencing lanes. Single Cell 3' v2 libraries (as pool or alone) will sequence successfully, however, the distribution of base composition along Read 1 and Read 2 will differ (Figure 4).

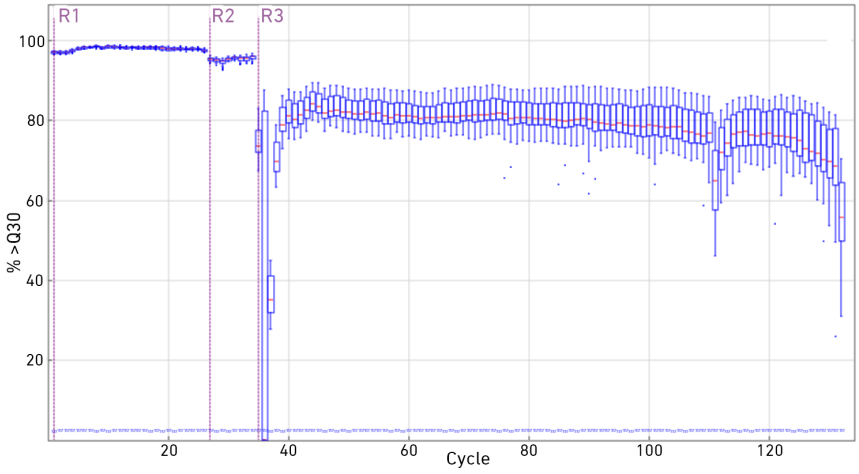

Fig. 3. Representative example of the 'Data by Cycle' plot in the Sequencing Analysis Viewer software (Illumina). Shown is the Q30 percentage along the sequencing length. Profile is based on sequencing 10x library by itself with no other library type sequenced alongside.

The profiles are characteristic for 10x Single Cell 3' v2 libraries that are sequenced with the alternative number of cycles (2x150bp) and can be explained as follows:

- Cycle 1 to 26 (Read 1- R1): Base percentages fluctuate due to sequences from the 16bp 10x barcode and the 10bp UMI that is attached to the Gel Bead primers. Sequences will have a different distribution than the human transcriptome, which causes the shift in percentage of each base.
- Cycle 27 to 150 (Read 1 – R1): Base percentages are dissimilar due to sequencing into the polyT-region of the barcoded Gel Bead primers.
- Cycle 151 to 158 (i7 index – R2): Base percentages are dissimilar due to sequences from the 8bp sample index.
- Cycle 159 to ~183 (Read 2 – R3): Beginning of the transcript read. Base percentages may fluctuate due to sequences from the RT Primer at the 5' end of the cDNA from a small subset of library fragments
- Cycle ~184 – 300 (Read 2 – R3): Base percentages are expected to be stable as the actual transcript is sequenced for all library fragments. An increase in "A" is expected towards the end as the proportion of sequences containing the polyA-tail increases.

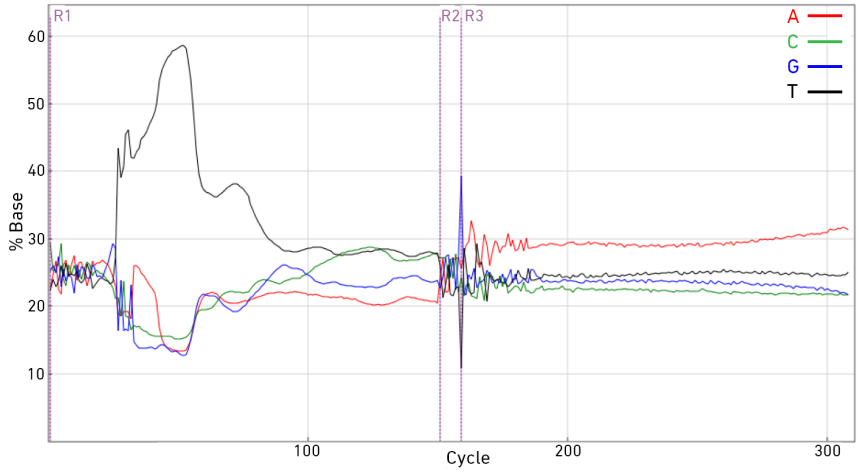

Fig. 4. Representative example of the 'Data by Cycle' plot in the Sequencing Analysis Viewer software (Illumina). Shown is the percentage of clusters for which the selected base has been called along the sequencing length. Profile is based on sequencing 10x library by itself with no other library type sequenced alongside.

The Q30 score during a 2x150bp sequencing run will drop accordingly after cycle 27 and will recover at the beginning of the i7 index read (Figure 5). Please note that the Cell Ranger™ software pipeline will not use any cycles beyond 26 of Read 1. Additional cycles for the transcript read (Read 2) will go into analysis and may improve mapping rates to the reference transcriptome.

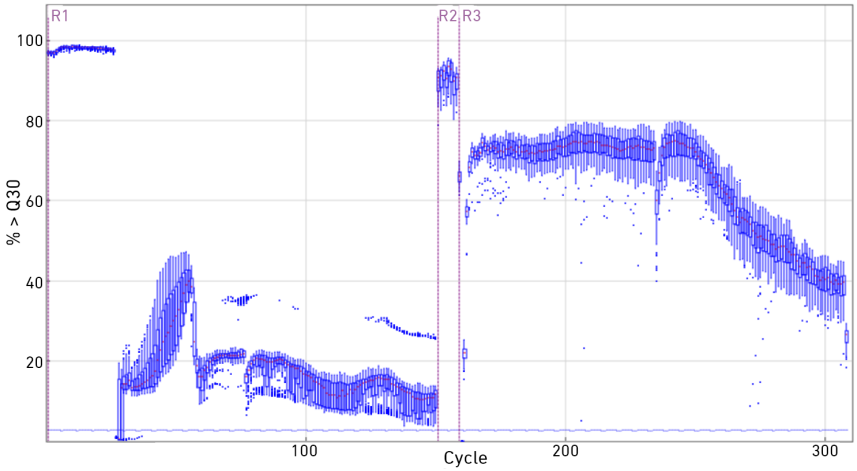

Fig. 5. Representative example of the 'Data by Cycle' plot in the Sequencing Analysis Viewer software (Illumina®). Shown is the Q30 percentage along the sequencing length. Profile is based on sequencing 10x library by itself with no other library type sequenced alongside.

Run Format	Read 1 > Q30	Read 2 > Q30
26 x 98	98.1 %	78.7 %
150 x 150	31.4 %	63.4 %

Table 2. Q30 percentages for Read 1 and Read 2 compared across two different sequencing run formats .

CONCLUSION

We have discussed sequencing parameters for Chromium Single Cell 3' v2 libraries and the resulting apparent base composition of sequencing reads generated from two different sequencing run configurations. Base percentages and quality scores across the library fragments are correlated with the sequences flanking the library insert. The representative example profiles of Chromium Single Cell 3' v2 Libraries demonstrated here thus serve as a reference for what constitutes a successful sequencing run using this library type.

REFERENCES

- *Chromium™ Single Cell 3' Reagent Kits v2 User Guide* (CG00052)

Notices

Document Number

CG00080 Rev B *Technical Note*

Legal Notices

Customer Information and Feedback

For technical information or advice, please contact our Customer Technical Support Division online at any time.

Email: support@10xgenomics.com

10x Genomics 7068 Koll Center Parkway

Suite 401

Pleasanton, CA 94566 USA