10x GENOMICS®

TECHNICAL NOTE

Chromium™ Single Cell 3' v2 Libraries – Sequencing Performance on Illumina® NextSeq® 500 Flow Cells

INTRODUCTION

The Chromium™ Single Cell 3' v2 Solution (CG00052) produces Illumina®-compatible sequencing libraries. A Single Cell 3' Library comprises Illumina paired-end constructs with P5 and P7 sequences at either end. These libraries have been validated on the following Illumina sequencing platforms:

- MiSeq®
- NextSeq® 500/550
- HiSeq® 2500 Rapid Run (RR) and High Output (HO)
- HiSeq® 3000/4000
- NovaSeq®

While sequencing metrics differ between sequencing platforms (e.g. MiSeq versus HiSeq 4000), they are generally consistent across runs on the same platform (see *Technical Note Chromium™ Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina® Sequencers* - CG000089). We have found, however, that Single Cell 3' v2 libraries sequenced on NextSeq 500/550 can show significant variability in sequencing metrics across different flow cell lots. This Technical Note presents a comparison of sequencing metrics for Single Cell 3' v2 libraries run on NextSeq 500 Illumina flow cells.

METHODS

We prepared a Chromium Single Cell 3' v2 library covering ~8,400 peripheral blood mononuclear cells (PBMCs) from a healthy donor by following the protocol in the *Single Cell 3' Reagent Kits v2 User Guide* (CG00052). This library was initially sequenced using paired-end sequencing (26bp Read 1 and 98bp Read 2) with a single sample index (8bp) on an Illumina HiSeq 4000. Results of the Cell Ranger™ analysis of this sequencing run are available on our Support website as the 'pbmc8k' dataset (https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/pbmc8k).

We next used the same library to assess the performance of individual NextSeq 500/550 flow cell lots. To minimize user-introduced variability, three 20 pM aliquots of the Single Cell 3' v2 library were denatured following the Illumina denaturation protocol. These aliquots were then further subdivided into single use 100 µl aliquots that were stored at -20°C.

The study design included four sections:

i. *Inter-lot* variability – We sequenced the same library on the same Illumina® NextSeq® 500 instrument at the same loading concentration with one run on each of nine unique flow cell lots.
ii. *Intra-lot* variability – We sequenced the same library on the same Illumina NextSeq 500 instrument at the same loading concentration with multiple runs on each of four different flow cell lots.
iii. Sequencing performance for additional flow cell lots – We compared our internal data with external customer data.
iv. Application performance for data sequenced on three flow cell lots – We compared Cell Ranger™ data outputs across these runs.

We report the following sequencing metrics for the sequencing runs:

- Cluster densities (K/mm^2)
- Yield (Gb) per Lane for Read 1 (R1) and Read 2 (R2)
- Phred quality scores (%bases >= Q30) for R1, i7 index and R2
- Mapping rate (%) of Read 2 transcript read to reference (GRCh38)

Table 1 provides the comparison of sequencing metrics for a single library that was sequenced on nine unique flow cell lots.

Library ID	Loading Conc. (pM)	PhiX Spike in (%)	Instrument	Flow Cell Lot Number	Cluster Density	Yield per Lane (Gb)		% Bases >= Q30			Mapping Rate (%)
						R1	R2	R1	i7	R2	
1	1.3	1	NextSeq 500	20071445	194 K/mm^2	2.8	10.9	95.66	88.35	61.66	53.5
1	1.3	1	NextSeq 500	20103772	117 K/mm^2	1.8	7.0	97.90	92.27	74.56	57.5
1	1.3	1	NextSeq 500	20103773	139 K/mm^2	2.2	8.3	98.10	95.50	83.96	60.0
1	1.3	1	NextSeq 500	20108160	170 K/mm^2	2.6	10.1	97.25	93.16	78.96	57.5
1	1.3	1	NextSeq 500	20109647	128 K/mm^2	1.7	6.6	93.02	70.11	32.35	37.3
1	1.3	1	NextSeq 500	20109650	155 K/mm^2	1.8	7.1	90.13	66.95	25.12	25.2
1	1.3	1	NextSeq 500	20109654	139 K/mm^2	2.0	7.1	95.34	82.58	47.89	48.9
1	1.3	1	NextSeq 500	20111907	120 K/mm^2	1.8	7.0	96.87	88.85	60.13	53.4
1	1.3	1	NextSeq 500	20112725	122 K/mm^2	1.9	7.3	97.84	91.84	74.10	57.8

Table 1. Metrics for one library sequenced across 9 unique NextSeq 500 flow cells.

The reported Q30 scores for R1, i7 and R2 are plotted in Figure 1. R2 (which sequences the transcript) showed the highest degree of variability, with Q30 scores ranging between approximately 25% and 84% (correlating with mapping rates ranging between 25% and 60%). Of all the flow cells tested, lot number 20109650 showed the lowest Q30 scores across all three reads. Consequently, mapping rates were below 30% for this flow cell lot. Note that although cluster densities varied substantially (between 117K/mm^2 and 194K/mm^2), they did not correlate with Q30 scores.

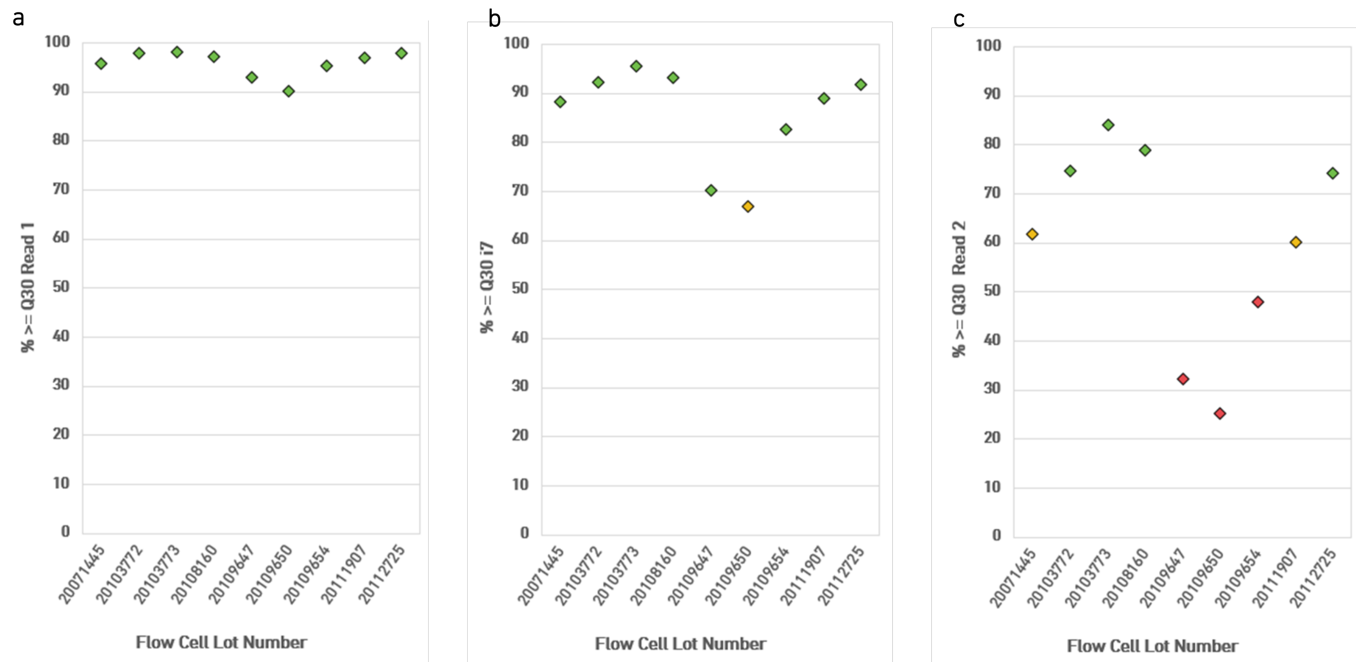

Figure 1. Read quality (%Q30 scores) for a) Read 1, b) i7 Read and c) Read 2 are plotted for one library sequenced across 9 unique NextSeq® 500 flow cells. Data points in green indicate Q30 quality scores >70%, data points in yellow indicate Q30 quality scores between 50 and 70%, and data points in red indicate Q30 quality scores <50%.

Four of the nine flow cell lots tested in this study (lot # 20071445, 20108160, 20109650, 20111907) showed large differences in sequencing metrics and were selected for additional assessment of intra-lot variability. Table 2 provides a comparison of sequencing metrics for the same library, sequenced multiple times in independent runs, on flow cells from these four lots. Flow cell lot number 20109650, which had the lowest performance metrics in the inter-lot comparison (Table 1), also showed low Q30 scores in the intra-lot comparison (average of R1: 92.2%, i7: 75.1%, R2: 33.23%), suggesting that poor sequencing performance can be linked to specific flow cell lots, rather than variability in instrument run performance or cluster density.

Library ID	Loading Conc. (pM)	PhiX Spike in (%)	Instrument	Flow Cell Lot Number	Cluster Density	Yield per Lane (Gb)		% Bases >= Q30			Mapping Rate (%)
						R1	R2	R1	i7	R2	
1	1.3	1	NextSeq 500	20071445	194 K/mm^2	2.8	10.9	95.66	88.35	61.66	53.5
1	1.3	1	NextSeq 500		138 K/mm^2	2.1	8.2	97.62	92.40	75.06	58.0
1	1.3	1	NextSeq 500		127 K/mm^2	1.9	7.6	97.52	89.80	69.03	56.3
1	1.3	1	NextSeq 500	20108160	170 K/mm^2	2.6	10.1	97.25	93.16	78.96	57.5
1	1.3	1	NextSeq 500		108 K/mm^2	1.7	6.5	97.93	92.23	69.70	56.7
1	1.3	1	NextSeq 500	20109650	155 K/mm^2	1.8	7.1	90.13	66.95	25.12	25.2
1	1.3	1	NextSeq 500		142 K/mm^2	1.9	7.5	93.1	75.66	35.46	42
1	1.3	30	NextSeq 500		164 K/mm^2	2.2	8.7	93.53	82.68	39.11	45.9
1	1.3	1	NextSeq 500	20111907	120 K/mm^2	1.8	7.0	96.87	88.85	60.13	53.4
1	1.3	1	NextSeq 500		147 K/mm^2	2.2	8.4	95.96	87.62	55.48	52.3

Table 2. Metrics for one library sequenced across 6 unique NextSeq® 500 flow cells from four different lots.

We next plotted the fraction of reads mapped to the transcriptome (GRCh38) against the corresponding Q30 quality scores for R2 for all of the previously described sequencing runs (9 from the inter-lot study + 6 from the intra-lot study; Figure 2). As expected, mapping rates were positively correlated with Q30 scores for R2.

For comparison, we also sequenced the same library on multiple flow cells of other supported Illumina® platforms and consistently obtained R2 Q30 quality scores of ~80% and transcriptome mapping rates of >60% (compared to an average of ~58% and ~51%, respectively, across the 15 NextSeq®; see also Technical Note *Chromium™ Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina Sequencers* (CG000089)).

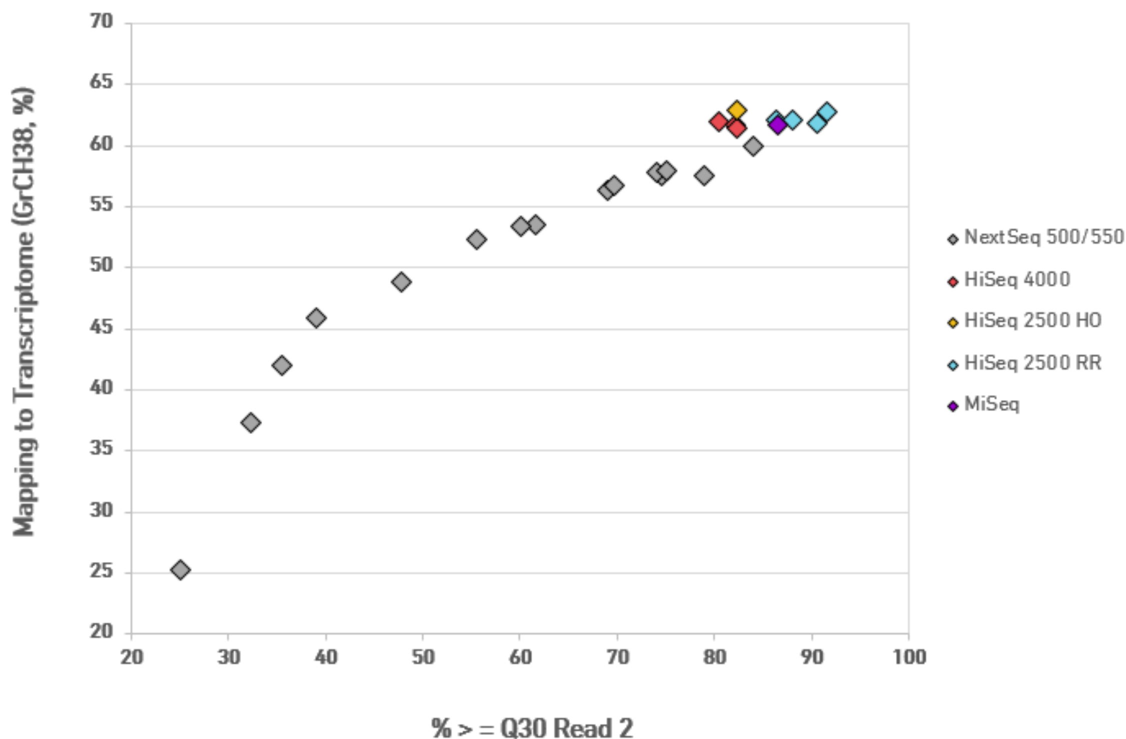

Figure 2. Fraction of reads mapped to the transcriptome (%) and corresponding Q30 scores for R2 for the same library sequenced across 15 different NextSeq 500 flow cells, as well as one MiSeq, one HiSeq 2500 High Output, four HiSeq 2500 Rapid Run and three HiSeq 4000 flow cells.

To complement and expand upon our internal study, we also collected metrics from libraries sequenced by external laboratories. Figure 3 shows data for a diverse set of Single Cell 3' v2 libraries sequenced across a total of 33 flow cell lots on the NextSeq 500/550 platform. We ranked flow cell lot numbers based on average Q30 quality score for R2 (Fig. 3a). We then grouped flow cells into three Q30 score bins (Fig. 3b). Mean values for all three bins were calculated at 41%, 60% and 77%.

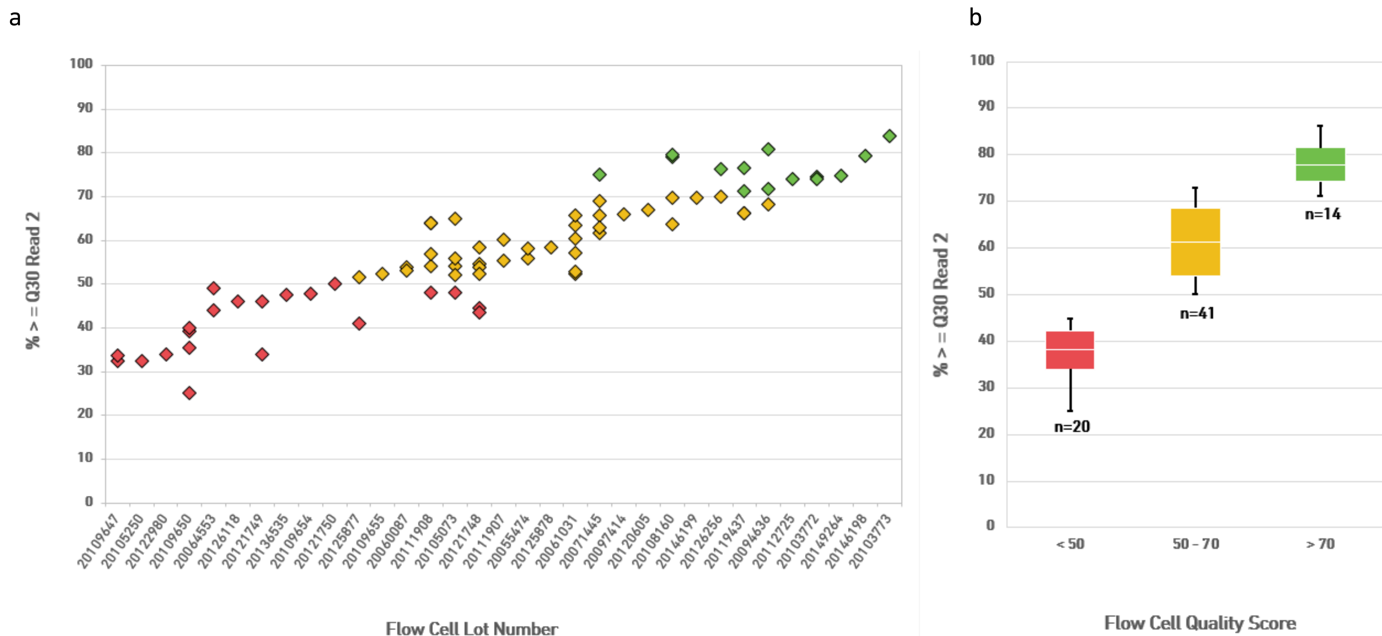

Figure 3. NextSeq® 500/550 flow cell lots ranked by increasing R2 Q30 quality scores for 75 sequenced SC3' v2 libraries. Data points in green indicate Q30 quality scores >70%, data points in yellow indicate Q30 quality scores between 50 and 70%, and data points in red indicate Q30 quality scores <50%. a) – Individual flow cell lots listed. b) – Flow cell lots binned into Q30 quality scores that are < 50%, between 50 and 70%, and > 70%.

Finally, we compared application performance and data quality from libraries that were sequenced on three different flow cell lots. We used one flow cell lot from each quality score bin (see Fig. 3), and processed the resulting sequencing data with Cell Ranger™ 2.0. Table 3 summarizes key metrics that are reported in the Cell Ranger *web_summary.html* file. As expected, application performance metrics that depend on the sequencing quality of R2 are lowest in the library that was sequenced on flow cell lot 20109650, which produced poor sequencing data (see Table 1), including an R2 Q30 score of ~25%. Metrics that were impacted include: "Number of Reads", "Q30 Bases in RNA Read", "Reads Mapped Confidently to the Reference" and "Mean Reads per Cell" (highlighted in yellow). Other key metrics including "Valid Barcodes", "Estimated Number of Cells", "Fraction Reads in Cells", and "Total Genes Detected" remained relatively stable across all three flow cells tested (highlighted in green). These metrics are not dependent on Read 2 Q30 quality scores but instead largely rely on Q30 quality scores for Read 1 (Valid Barcodes), number of cells loaded (Estimated Number of Cells), number of transcripts associated with a cell-containing partition (Fraction Reads in Cells) and complexity/number of transcripts associated with a given sample (Total Genes Detected).

We then down-sampled the data to 20,000 Mean Reads per Cell for a more direct comparison and for downstream analysis in Loupe Cell Browser™. Specifically, we used the *AMLBloodCell.csv* gene list to identify each of the major cell subtypes found in human PBMCs:

- Lymphocytes: T cells, B cells
- Myeloid cells: Monocytes

The gene list can be downloaded on the Support website at https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/tutorial-celltypes).

Sequencing Quality			
Flow cell quality score	< 50%	50-70%	> 70%
Flow cell lot number	20109650	20111907	20103773
Number of Reads	174,859,616	310,978,971	326,083,311
Valid Barcodes	97.40%	98.00%	98.20%
Reads Mapped Confidently to Transcriptome	25.20%	52.30%	58.70%
Reads Mapped Confidently to Exonic Regions	26.90%	55.60%	62.40%
Reads Mapped Confidently to Intronic Regions	10.50%	21.50%	23.00%
Reads Mapped Confidently to Intergenic Regions	2.00%	3.20%	3.00%
Sequencing Saturation	51.60%	78.90%	80.40%
Q30 Bases in Barcode	92.10%	96.50%	98.40%
Q30 Bases in RNA Read	25.40%	57.40%	84.30%
Q30 Bases in Sample Index	75.40%	91.20%	96.60%
Q30 Bases in UMI	90.00%	95.90%	97.80%
Cells			
Estimated Number of Cells	8,236	8,313	8,341
Fraction Reads in Cells	93.10%	93.40%	93.30%
Mean Reads per Cell	21,231	37,408	39,094
Median Genes per Cell	813	1,090	1,139
Total Genes Detected	19,079	20,397	20,503
Median UMI Counts per Cell	2,050	3,240	3,501
Post Normalization (20,000 raw reads per cell)			
Estimated Number of Cells	8,236	8,313	8,341
Mean Reads per Cell	20,000	20,000	20,000
Median Genes per Cell	797	984	1,028
Median UMI Counts per Cell	1,995	2,792	3,025

Table 3. Sequencing and application performance metrics for three libraries sequenced on three different NextSeq® 500 flow cell lots. Performance metrics reported by Cell Ranger 2.0. Key metrics that are substantially different across all three sequencing runs are highlighted in yellow. Key metrics that remained stable are highlighted in green.

Using the *cellranger aggr* command, reads from each library (stored in the molecule_info.h5 file produced by *cellranger count*) were subsampled until all libraries had the same mean reads per cell, to control for sequencing batch effects. The normalized libraries were aggregated into a single gene-barcode matrix, and then secondary analysis (PCA, clustering, differential expression and t-SNE) was performed. We found that each of the three data sets clustered similarly and a total of 17 sub-clusters were identified via Graph based clustering. In addition, despite the differences in sequencing quality, all major cell subtypes were clearly identified in the three data sets based on marker gene expression.

Graph Based Clustering			
Flow cell quality score	< 50%	50-70%	> 70%
Flow cell lot number	20109650	20111907	20103773
Graph based clustering (# of clusters = 17)			
Normalized Expression Levels			
T-cells			
Enrichment of CD3D Color Scale 0.0 4.6			
Enrichment of CD3E Color Scale 0.0 3.5			
Enrichment of CD4 Color Scale 0.0 2.8			
Cytotoxic/CD8 T-cells			
Enrichment of GNLY Color Scale 0.0 7.3			
Enrichment of GZMB Color Scale 0.0 5.6			
B-cells			
Enrichment of CD79A Color Scale 0.0 4.9			

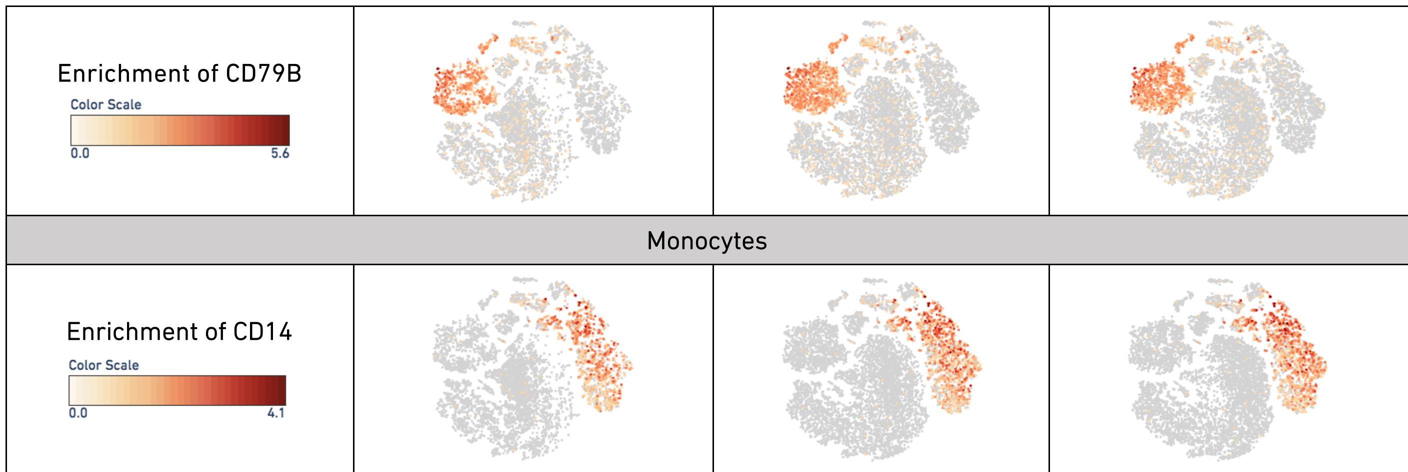

Figure 4. Cluster analysis with Loupe Cell Browser™ from one library sequenced on three NextSeq® 500 flow cell lots with significantly different sequencing performance. Graph-based clustering and normalized expression levels for marker genes of major human PBMC subpopulations are shown.

DISCUSSION

Our studies highlight the variability in sequencing quality (in particular, Q30 score for Read 2) for Single Cell 3' v2 libraries run on the NextSeq® 500 platform. The data suggests that the variability correlates with different lots of flow cells. While one library, sequenced on one flow cell lot, produces Read 2 Q30 scores greater than 70%, the same library can produce Q30 scores between 20-40% on different flow cell lots. The same library consistently produced Q30 scores for Read 2 in the expected range with very minimal variability on other sequencer platforms (see Fig. 2). Note that we did not observe the same degree of variability in sequencing quality across different NextSeq 500 reagent or buffer lots (data not shown).

Overall, the majority of sequenced libraries in this study (55 out of 75, see Fig. 3b) showed Read 2 Q30 scores greater than 50%, which matches or exceeds our specifications for the NextSeq 500 sequencer (Technical Note *Chromium™ Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina® Sequencers* (CG000089)). Cell Ranger analysis suggests that key metrics such as the number of "Valid Barcodes", "Estimated Number of Cells", "Fraction Reads in Cells", and "Total Genes Detected" are largely unaffected by variation in Read 2 Q30 scores. Likewise, cluster analysis with Loupe Cell Browser™ indicates that individual PBMC subtypes can be identified from even low quality sequencing runs. We therefore encourage users to carefully analyze their data with Cell Ranger and Loupe Cell Browser even if initial sequencing metrics appear to be low. Note that our analysis was performed on human PBMCs and that other cell types may perform differently.

If a Chromium Single Cell 3' v2 library-consistently shows low Q30 quality scores for R2 (<50%), we recommend the following steps:

- Assess the fraction of reads mapped to the transcriptome. Mapping rates are ~55-60% on average for our publicly available datasets (https://support.10xgenomics.com/single-cell-gene-expression/datasets), but will vary depending on the cell type and the quality of the reference used.
- Re-sequence the library on a different lot NextSeq flow cell, preferably one that has been shown to be high-performing (see Figure 3a).
- Re-sequence the library on a HiSeq® flow cell. Note that Chromium Single Cell 3' v2 libraries can be sequenced in a different sequencing run format (e.g. 2x100 bp paired end) to accommodate non-10x libraries on the same flow cell. See Technical Note *Base Composition of Sequencing Reads of Chromium™ Single Cell 3' v2 Libraries* (CG000080) for more details.
- Sequence the library on a MiSeq® for quality control and to rule out library-specific sources of low sequencing quality.

CONCLUSION

We have discussed sequencing parameters for Chromium™ Single Cell 3' v2 libraries that were sequenced on different NextSeq® 500 flow cell lots. Sequencing metrics and flow cell lot number information reported in this Technical Note serve as a guideline to assess sequencing run quality of Chromium Single Cell 3' v2 libraries.

REFERENCES

- *Chromium™ Single Cell 3' Reagent Kits v2 User Guide* (CG00052)
- *Demonstrated Protocol – Single Cell Protocols – Cell Preparation Guide* (CG00053)
- *Base Composition of Sequencing Reads of Chromium™ Single Cell 3' v2 Libraries* (CG000080)
- *Chromium™ Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina Sequencers* (CG000089)

Notices

Document Number

CG000085 Rev A *Technical Note*

Legal Notices

© 2017 10x Genomics, Inc. All rights reserved. Duplication and/or reproduction of all or any portion of this document without the express written consent of 10x Genomics, Inc., is strictly forbidden. Nothing contained herein shall constitute any warranty, express or implied, as to the performance of any products described herein. Any and all warranties applicable to any products are set forth in the applicable terms and conditions of sale accompanying the purchase of such product. 10x Genomics provides no warranty and hereby disclaims any and all warranties as to the use of any third party products or protocols described herein. The use of products described herein is subject to certain restrictions as set forth in the applicable terms and conditions of sale accompanying the purchase of such product. “10x”, “10x Genomics”, “Changing the Definition of Sequencing”, “Chromium”, “GemCode”, “Loupe”, “Long Ranger”, “Cell Ranger” and “Supernova” are trademarks of 10x Genomics, Inc. All other trademarks are the property of their respective owners. All products and services described herein are intended FOR RESEARCH USE ONLY and NOT FOR USE IN DIAGNOSTIC PROCEDURES.

The use of 10x Product(s) in practicing the methods set forth herein has not been validated by 10x, and such non-validated use is NOT COVERED BY 10X STANDARD WARRANTY, AND 10X HEREBY DISCLAIMS ANY AND ALL WARRANTIES FOR SUCH USE.

Nothing in this document should be construed as altering, waiving or amending in any manner 10x Genomics, Inc., terms and conditions of sale for the Chromium™ Controller, consumables or software, including without limitation such terms and conditions relating to certain use restrictions, limited license, warranty and limitation of liability, and nothing in this document shall be deemed to be Documentation, as that term is set forth in such terms and conditions of sale. Nothing in this document shall be construed as any representation by 10x Genomics, Inc that it currently or will at any time in the future offer or in any way support any application set forth herein.

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