Sequencing Metrics & Base Composition of Chromium Flex Libraries

Introduction

The Flex Gene Expression workflow with Feature Barcode technology produces sequencing-ready libraries for measuring RNA & protein expression. This Technical Note compares sequencing metrics for pooled Flex Gene Expression & Protein Expression libraries across different Illumina platforms. The expected base percentage profiles and Phred quality scores (Q scores) based on a set of control libraries are described to provide general guidance on the expected range of sequencing metrics. Individual results may vary depending on the specific sequencing instrument and sample and loading characteristics.

The Flex Gene Expression workflow with Feature Barcode technology offers comprehensive, scalable solutions to measure gene and protein expression in formaldehyde-fixed samples. The dual index Flex Gene and Protein Expression libraries are generated using Flex reagents and protocols (see References).

Flex Gene Expression library components are shown in Figure 1A. When libraries are sequenced with the Read 2 sequencing scheme (see Figure 3), the sequencing primer binds to the TruSeq library adapter. Read 1 begins with a 16 bp 10x GEM Barcode followed by a 12 bp UMI sequence, while Read 2 contains the ligated probe insert sequence, a constant sequence, and an 8 bp Probe Barcode, which identifies the probe set used for sample hybridization. A single library contains reads derived from up to 16 samples, with cell barcodes composed of a Probe Barcode and a 10x GEM Barcode (up to 16 Probe Barcodes and 737,000 10x GEM Barcodes).

Flex Protein Expression (direct capture) library components are shown in Figure 1B. In the Read 2 sequencing scheme, Illumina TruSeq Read 2 (Read 2T) sequences the Feature Barcode. Multiplexing is not supported for this library type.

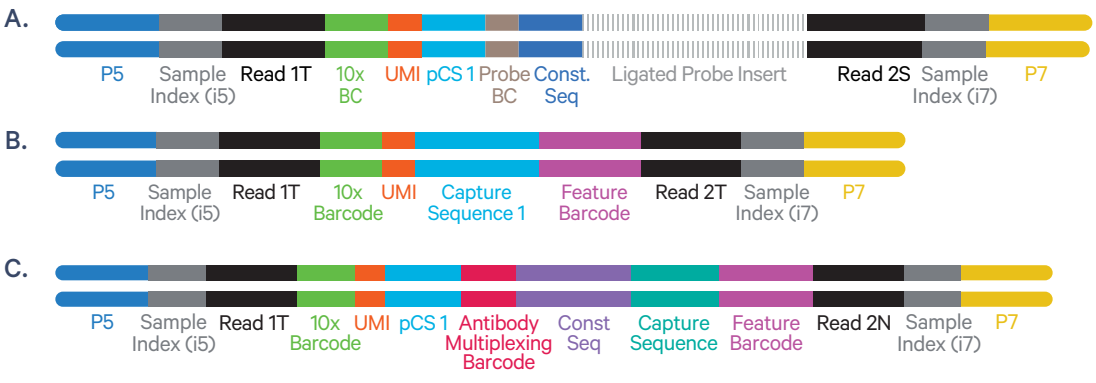


Figure 1. Schematic showing **(A)** Chromium Flex Gene Expression library components. **(B)** Chromium Flex Protein Expression library components. **(C)** Chromium Flex Protein Expression (Barcode Oligo Capture) library components.

Flex Protein Expression (Barcode Oligo Capture) library components are shown in Figure 1C. In the Read 2 sequencing scheme, Illumina Nextera Read 2 (Read 2N) sequences the Feature Barcode, constant sequence, and Antibody Multiplexing Barcode that identifies the sample.

Tables 1-10 show representative plots and sequencing data derived from pooled single and multiplexed (4-plex, 16-plex) sample libraries. A comparison of Read 1 and Read 2 sequencing schemes (Table 4) shows that both perform similarly. In most cases, either Read 1 (Read 1: 48 cycles, i7 index: 10 cycles, i5 index: 10 cycles, Read 2: 50 cycles) or Read 2 (Read 1: 28 cycles, i7 index: 10 cycles, i5 index: 10 cycles, Read 2: 90 cycles)

Methods Overview

Five Flex Gene Expression library pools were generated using the relevant Chromium Flex User Guide (see reference section for details).

- **Pool A** = four singleplex Flex Gene Expression libraries (BC001)
- **Pool B** = four singleplex Flex Gene Expression + one 4-plex Flex Gene Expression library (BC001-BC004)
- **Pool C** = four singleplex Flex Gene Expression libraries + one 4-plex + one 16-plex Flex Gene Expression library (BC001-BC016)
- **Pool D** = two singleplex Flex Gene Expression libraries (80%) and two singleplex Protein Expression (20%) libraries
- **Pool E** = two 4-plex Flex Gene Expression libraries and two 4-plex Protein Expression (Barcode Oligo Capture) libraries

Libraries were quantified with the KAPA DNA Quantification Kit and sequenced with either 1% (singleplex), or 5% or 10% (multiplex) PhiX depending on the workflow and sequencer used. Consult the Library Loading table in the respective user guide for sequencer-specific PhiX recommendations. The additional singleplex and 4-plex libraries in Pools B and C were included to improve i5 and i7 diversity.

Results

Representative **Data by Cycle** plots (% Base by cycle and % ≥Q30 quality scores by cycle) from the Illumina SAV software, along with additional metrics for singleplex, 4-plex, and 16-plex libraries are shown in Tables 1-10. **% ≥Q30** may refer to i5, i7, R1, or R2. For multiplexed samples, the **Multiplexing Probe BC Q30** refers to the 8 bp Probe Barcode. **Valid barcodes** metric refers to 10x Barcode for singleplexed and 10x Barcode + Probe Barcode for multiplexed libraries. Unless otherwise specified, **Mapped reads** are reported from Cell Ranger's Reads Mapped Confidently to the Filtered Probe Set metric. Occupancy (% nanowells containing reads) and percent pass filter (% PF) are reported for the NextSeq 2000 and the NovaSeq 6000, which use patterned flow cells, instead of cluster density.

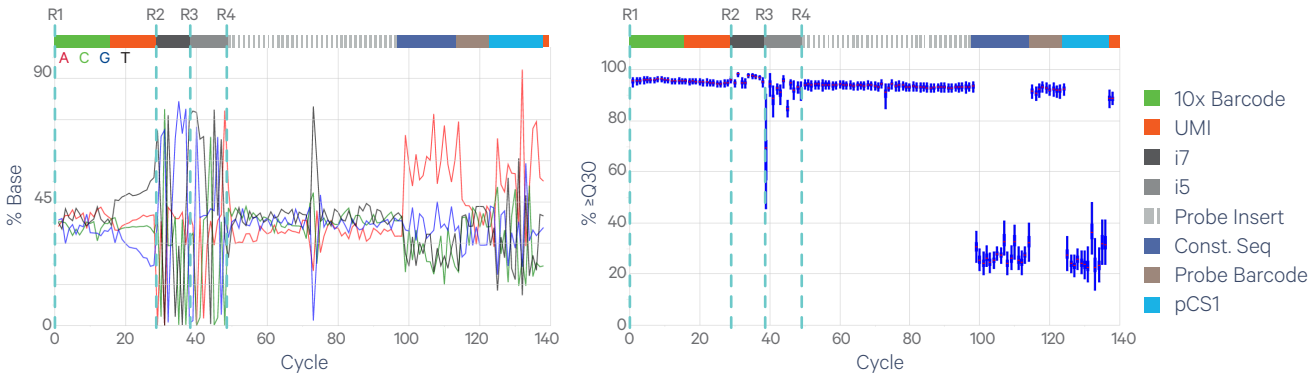
These data demonstrate compatibility of Chromium Flex Gene Expression libraries with multiple Illumina sequencers. Due to probe design, some regions have low Q30 scores as a result of reduced base diversity. As the data by cycle plot annotations indicate, these declines occur in the constant sequence and pCS1 regions. This does not affect assay performance as Cell Ranger's probe alignment algorithm uses the probe insert region as well as the Probe Barcode region in the case of multiplex libraries. The aligner does not use the constant sequence-NN or pCS1-UMI regions.

Pool A: Flex Gene Expression Singleplex Library

Pool: Four singleplex libraries (BC001)

Sequencing configuration & run parameters: Minimum sequencing depth: Gene Expression 10,000 read pairs/cell.

Paired-end, dual indexing:
Read 1: 28 cycles
i7 Index: 10 cycles
i5 Index: 10 cycles
Read 2: 90 cycles



Plots show the components of the library molecule being sequenced per cycle.

	MiSeq	NextSeq 500/550	NextSeq 2000	NovaSeq 6000*
Loading concentration	10 pM	2.5 pM	650 pM	150 pM
Cluster density / Occupancy	932 K/mm^2	217.5 K/mm^2	96.5%**	97.8%**
% PF	92.4	91.2	85.6	85.1
% PhiX	-	-	1	1
% ≥Q30	R1: 98.0, i7: 97.4, i5: 97.8, R2: 96.1	R1: 96.7, i7: 97.1, i5: 97.3. R2: 69.3	R1: 96.3, i7: 97.6, i5: 96.2. R2: 75.2	R1: 95.5, i7: 96.6, i5: 88.8. R2: 72.4
Yield per lane	R1: 0.57 Gb R2: 1.87 Gb	R1: 13.96 Gb R2: 45.97 Gb	R1: 15.38 Gb R2: 50.70 Gb	R1: 88.61 Gb R2: 292.10 Gb
Total read pairs	20.7 M	512.6 M	570.1 M	3259.3 M
Mapped reads %	96.6	96.7	97.4	97.1
Valid barcodes %	98.4	98.5	98.5	98.5

*S4 flow cell, 1 lane, **Occupancy is reported for the NextSeq 2000 and the NovaSeq 6000, which use patterned flow cells, instead of cluster density.

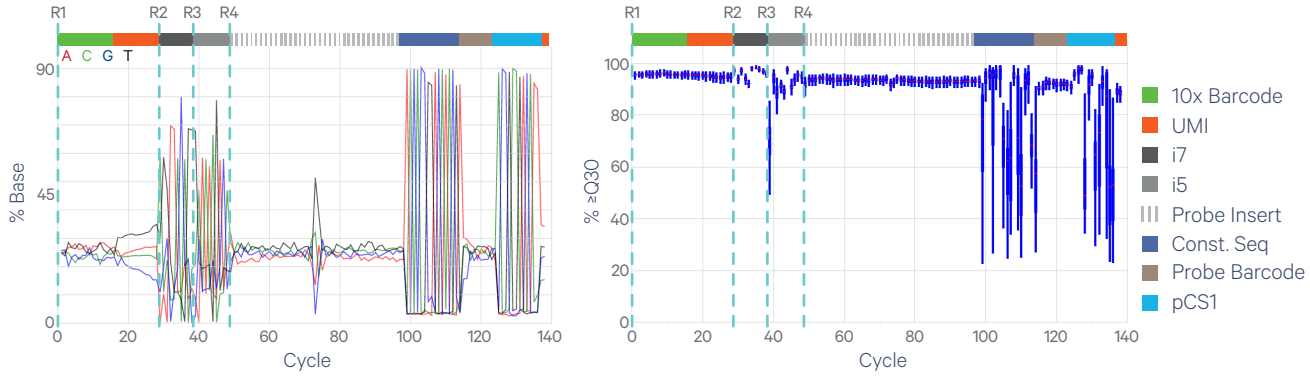
Table 1. The data by cycle plots shown are from the library pool sequenced on one lane of a NovaSeq 6000 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for a singleplex library (below black line).

Pool B: Flex Gene Expression Multiplexed 4-plex Library

Pool: Four singleplex libraries and a 4-plex library (BC001-BC004)

Sequencing configuration & run parameters: Minimum sequencing depth: Gene Expression 10,000 read pairs/cell.

Paired-end, dual indexing:
Read 1: 28 cycles
i7 Index: 10 cycles
i5 Index: 10 cycles
Read 2: 90 cycles



Plots show the components of the library molecule being sequenced per cycle.

	MiSeq	NextSeq 500/550	NextSeq 2000	NovaSeq 6000*
Loading concentration	12 pM	2.5 pM	650 pM	150 pM
Cluster density / Occupancy	892 K/mm^2	160.5 K/mm^2	92.3%**	97.2%**
% PF	94.7	94.9	82.8	86.8
% PhiX	5	5	5	10
% ≥Q30	R1: 98.7, i7: 95.7, i5: 96.9, R2: 96.0	R1: 97.8, i7: 97.9, i5: 97.6, R2: 75.2	R1: 96.6, i7: 96.3, i5: 96.1, R2: 74.2	R1: 95.3, i7: 96.1, i5: 90.0, R2: 87.9
Yield per lane	R1: 0.56 Gb R2: 1.84 Gb	R1: 10.71 Gb R2: 35.26 Gb	R1: 14.88 Gb R2: 48.91 Gb	R1: 89.00 Gb R2: 293.38 Gb
Total read pairs	20.3 M	394.5 M	551.4 M	3324.4 M
Mapped reads %	98.4	98.7	98.8	98.6
Multiplexing Probe BC Q30 %	96.2	88.8	91.0	92.9
Valid barcodes %	96.6	94.2	94.5	96.8

*S4 flow cell, 1 lane, **Occupancy is reported for the NextSeq 2000 and the NovaSeq 6000, which use patterned flow cells, instead of cluster density.

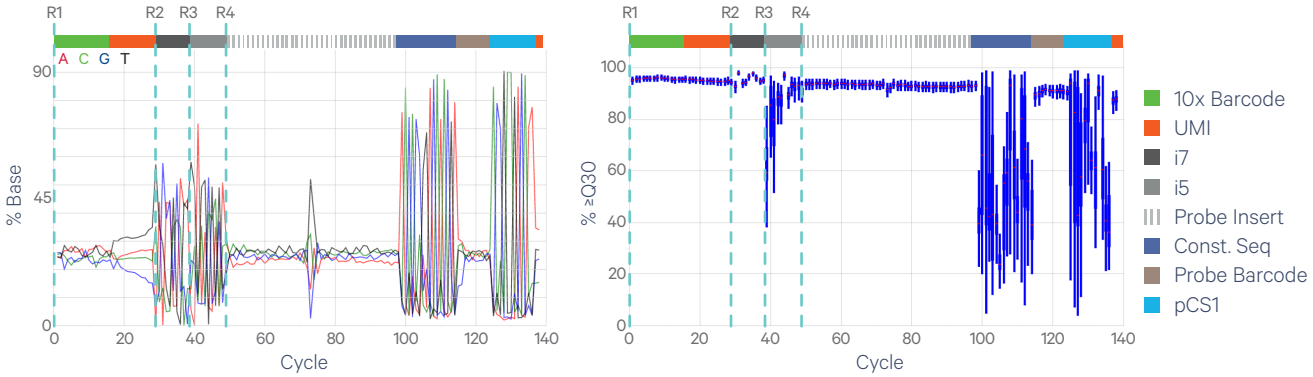
Table 2. The data by cycle plots shown are from the library pool sequenced on one lane of a NovaSeq 6000 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for the 4-plex library (below black line).

Pool C: Flex Gene Expression Multiplexed 16-plex Library

Pool: Four singleplex libraries, a 4-plex library, and a 16-plex library (BC001-BC016)

Sequencing configuration & run parameters: Minimum sequencing depth: Gene Expression 10,000 read pairs/cell.

Paired-end, dual indexing:
Read 1: 28 cycles
i7 Index: 10 cycles
i5 Index: 10 cycles
Read 2: 90 cycles



Plots show the components of the library molecule being sequenced per cycle.

	MiSeq	NextSeq 500/550	NextSeq 2000	NovaSeq 6000*
Loading concentration	12 pM	2.5 pM	650 pM	150 pM
Cluster density / Occupancy	888 K/mm^2	174.3 K/mm^2	92.9%**	97.5%**
% PF	93.6	93.8	83.6	86.4
% PhiX	5	5	5	10
% ≥Q30	R1: 98.3, i7: 96.6, i5: 97.6, R2: 95.1	R1: 97.5, i7: 97.3, i5: 97.4. R2: 70.7	R1: 96.5, i7: 96.5, i5: 96.7. R2: 74.7	R1: 95.5, i7: 95.5, i5: 85.9. R2: 81.8
Yield per lane	R1: 0.55 Gb R2: 1.82 Gb	R1: 11.49 Gb R2: 49.31 Gb	R1: 15.07 Gb R2: 49.68 Gb	R1: 85.72 Gb R2: 282.57 Gb
Total read pairs	20.0 M	423.7 M	556.8 M	3309.1 M
Mapped reads %	98.2	98.4	98.6	98.3
Multiplexing Probe BC Q30 %	93.9	90.6	88.8	89.2
Valid barcodes %	96.0	95.6	93.2	94.9

*S4 flow cell, 1 lane, **Occupancy is reported for the NextSeq 2000 and the NovaSeq 6000, which use patterned flow cells, instead of cluster density.

Table 3. The data by cycle plots shown are from the library pool sequenced on one lane of a NovaSeq 6000 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for the 4-plex library (below black line).

Figure 2 shows an example of a poor-quality sequencing run for multiplexed libraries, sequenced with 3% PhiX on a NovaSeq 6000. In this example, the constant sequence drops the Q30 scores for the multiplexing Probe Barcode region because there is low or no sequence diversity and sequencers are occasionally unable to recover Q30s in the Probe Barcode region. Increasing PhiX helps to mitigate this issue, and thus the Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide (CG000527) recommends sequencing with 10% PhiX for multiplexed samples on a NovaSeq 6000.

In the case of multiplexed sample libraries, poor sequencing quality in the Probe Barcode region can interfere with correct assignment of reads to respective samples in Cell Ranger, and thus affect downstream analysis. However, barcode correction should help to recover some of the data. For singleplex libraries, the Probe Barcode region is not required and poor sequencing quality in this region is not expected to impact downstream analysis.

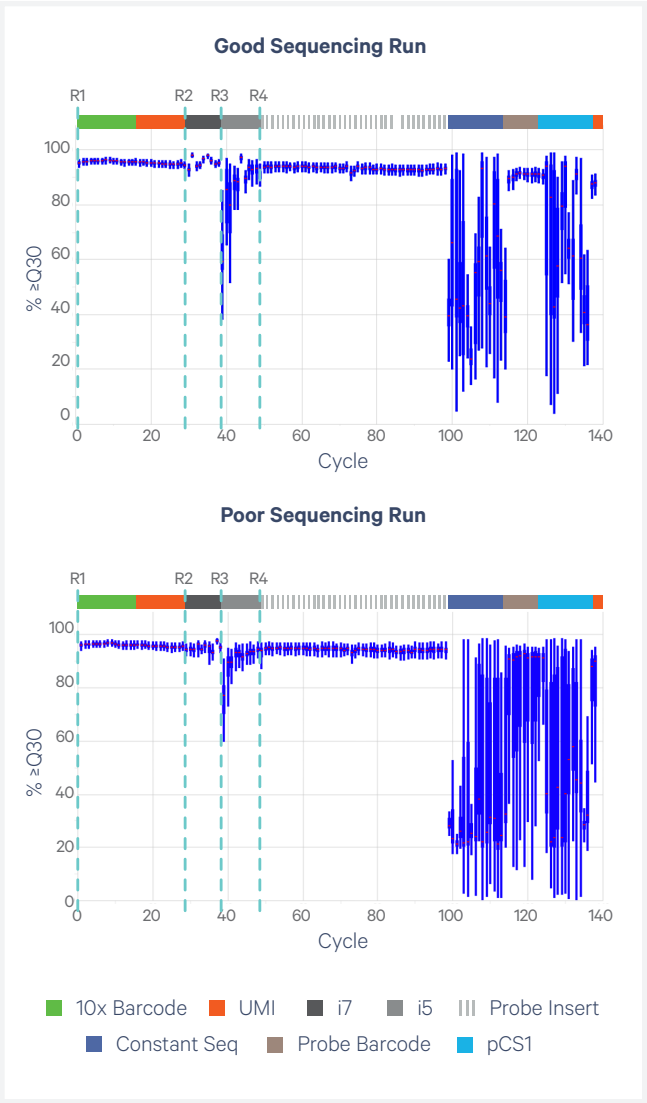


Figure 2. Example of a good and a poor % ≥Q30 quality scores by cycle plot for multiplexed libraries sequenced on a NovaSeq 6000. The good sequencing run was sequenced with 10% PhiX. The poor sequencing run was sequenced with 3% PhiX, and the Q30 scores are degraded in the Probe Barcode region.

Comparing Read 1 and Read 2 Schemes

When sequencing Flex Gene Expression libraries on a NovaSeq 6000 instrument (Standard workflow) with a loading concentration of 150 pM, a comparison of Read 1 (Read 1: 48 cycles, i7 index: 10 cycles, i5 index: 10 cycles, Read 2: 50 cycles) and Read 2 (Read 1: 28 cycles, i7 index: 10 cycles, i5 index: 10 cycles, Read 2: 90 cycles) schemes indicates that their performance is similar (Table 4). In most cases, the two schemes can be used interchangeably. However, for effective data analysis, it may be important to understand the specific conditions under which Cell Ranger extracts information from Read 1 versus Read 2, as well as the rationale behind these selections.

For Cell Ranger v7.2 and earlier, the multiplexing Probe Barcode must be in the same position for all library types. Libraries created with multiplexed custom probes without a constant sequence and Proteintech Genomics (PTG)-derived antibody libraries do not have the Probe Barcode in the same position on Read 2 as libraries prepared using the 10x WTA Probe Panel (Figure 3). However, regardless of library type, the Probe Barcode is always located at the same base position on Read 1, allowing Cell Ranger to extract this information consistently from Read 1. If libraries are sequenced with the Read 1 scheme, Cell Ranger can autodetect and extract multiplexing Probe Barcode information from Read 1.

Cell Ranger v8.0 introduces the ability to specify per-library chemistry in the libraries section of the multi config CSV. Different library types within a multilibrary analysis can now have the multiplexing Probe Barcode in different positions, provided the chemistries are specified in the multi config CSV. This allows Cell Ranger to extract barcode information from different positions on Read 2. However, a single library type (e.g., Antibody Capture specified in the multi config CSV) can only have one chemistry definition. For example, if the Antibody Capture library is a combination of PTG and BioLegend antibodies, Cell Ranger cannot extract multiplexing Probe Barcode information from Read 2 because the barcode is not aligned for PTG versus BioLegend-derived reads. In such cases, Cell Ranger must use Read 1 for barcode extraction.

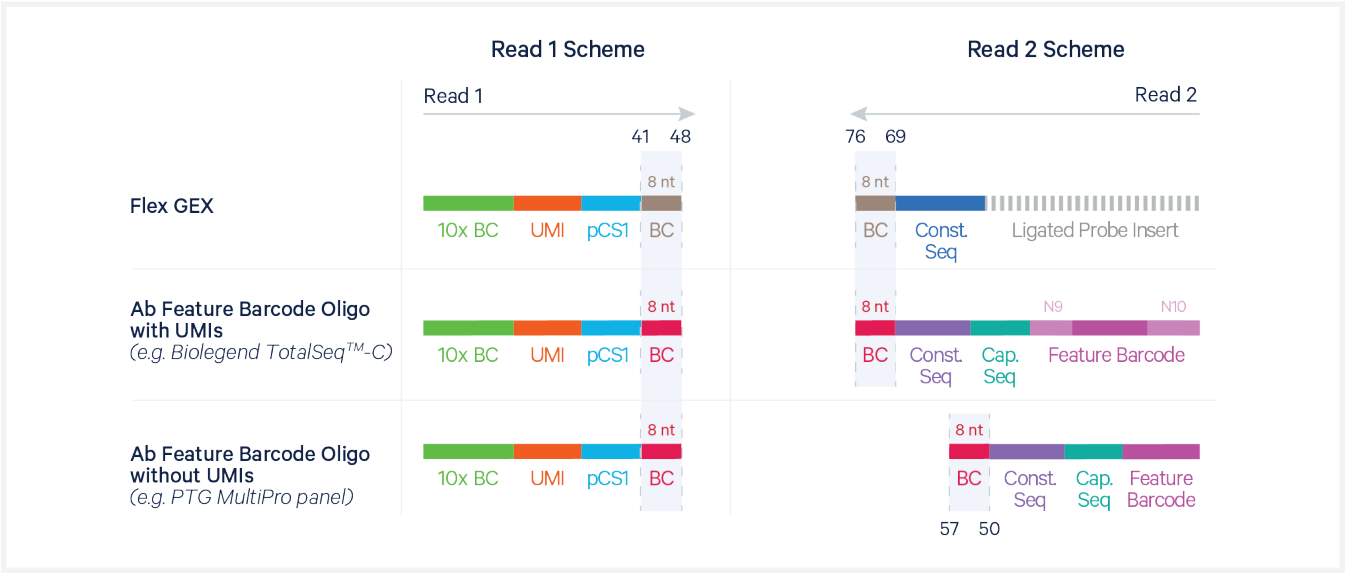
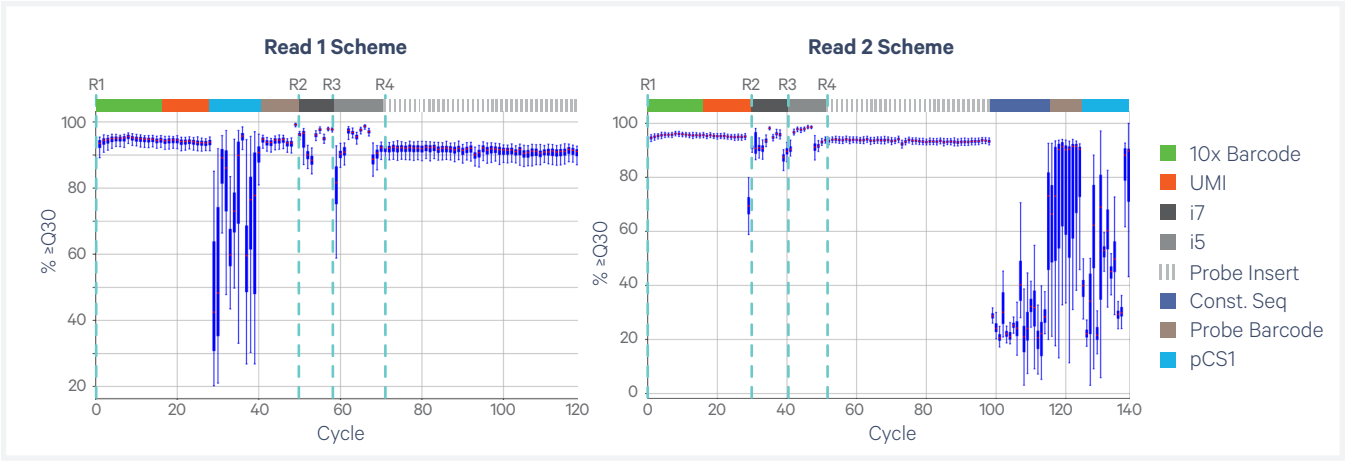


Figure 3. The location of the multiplexing Probe Barcode sequence in Read 1 versus Read 2 of Flex Gene Expression and two Flex Protein Expression libraries (BioLegend barcode oligo capture and PTG).



	Read 1 Scheme	Read 2 Scheme
Sequencing scheme	Read 1: 28 cycles i7 index: 10 cycles i5 index: 10 cycles Read 2: 90 cycles	Read 1: 48 cycles i7 index: 10 cycles i5 index: 10 cycles Read 2: 50 cycles
Loading concentration	150 pM	150 pM
% Occupancy	97.51	93
% PF	78.24	85
% PhiX	10	10
% ≥Q30	R1: 88.7, i7: 95.0, i5: 93.6. R2: 91.3	R1: 96.8, i7: 96.9, i5: 96.5, R2: 84.6
Yield per lane	R1: 46.98 Gb R2: 48.95 Gb	R1: 29.07 Gb R2: 95.82 Gb
Total read pairs	872.0 M	942.6 M
Mapped reads %	95	96
Multiplexing Probe BC Q30 %	95	92
Valid barcodes %	95.5	96.6

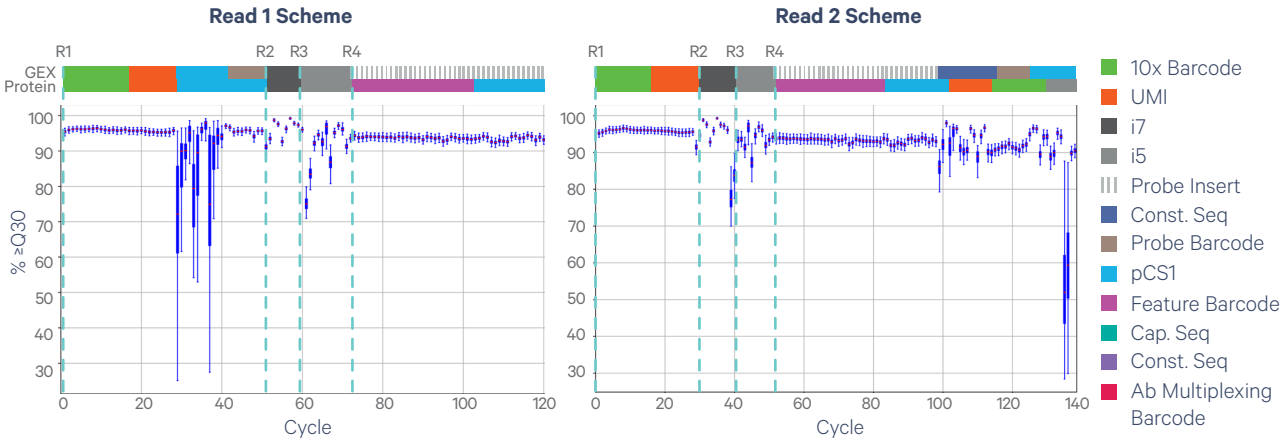
Table 4. The data by cycle plots shown are from the 4-plex library sequenced on one lane of a NovaSeq 6000 SP 100 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for the 4-plex library (below black line).

Pool D: Flex Gene Expression (80%) with Protein Expression (20%)

Pool: Two singleplex Flex Gene Expression libraries and two singleplex Protein Expression libraries. Gene and Protein Expression libraries were combined at a ratio of 4:1.

Sequencing configuration & run parameters: Minimum sequencing depth: Gene Expression 10,000 read pairs/cell. Protein library 5,000 reads pairs/cell.

Paired-end, dual indexing:
Read 1: 28 cycles
i7 Index: 10 cycles
i5 Index: 10 cycles
Read 2: 90 cycles
Plots show the components of the library molecule being sequenced per cycle.



	MiSeq	NextSeq 500/550	NextSeq 2000	NovaSeq 6000*	NovaSeq 6000 Xp*
Loading concentration	12 pM	2.5 pM	650 pM	150 pM	150 pM
Cluster density / Occupancy	820 K/mm^2	130 K/mm^2	90.9%**	96.7%**	82%**
% PF	95.7	95.9	83.0	86.8	77.2
% PhiX	5	5	5	10	10
% ≥Q30	R1: 99.0, i7: 98.4, i5: 98.3, R2: 96.7	R1: 98.2, i7: 97.7, i5: 98.0, R2: 96.6	R1: 96.4, i7: 96.8, i5: 97.6, R2: 92.2	R1: 95.9, i7: 96.1, i5: 90.4. R2: 92.0	R1: 96.2, i7: 96.6, i5: 93.1. R2: 92.9
Yield per lane	R1: 0.51 Gb R2: 1.67 Gb	R1: 8.76 Gb R2: 28.82 Gb	R1: 14.86 Gb R2: 49.00 Gb	R1: 29.55 Gb R2: 97.37 Gb	R1: 60.57 Gb R2: 199.64 Gb
Total read pairs	16.4 M	274.9 M	527.4 M	900.8 M	1925.3 M
Mapped reads %	94	94	95	94	95
Valid barcodes %	99.0	99.0	99.0	99.0	99.0

* S4 flow cell, 1 lane, ** Occupancy is reported, which use patterned flow cells, instead of cluster density.

Table 5. The data by cycle plots shown are from the library pool sequenced on one lane of a NovaSeq 6000 SP 100 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for the 4-plex library (below black line). Values for the NovaSeq 6000 in the table are from the Read 1 sequencing scheme. The multiplexing barcode Q30 and valid barcode metrics are only for Flex Gene Expression libraries.

Pool E: Flex Gene Expression (80%) with Protein Expression (Barcode Oligo Capture, 20%)

Pool: Two 4-plex Flex Gene Expression libraries (BC1-4) and two 4-plex Protein Expression libraries (Barcode Oligo Capture). Gene and Protein Expression libraries were combined at a ratio of 4:1.

Sequencing configuration & run parameters: Minimum sequencing depth: Gene Expression 10,000 read pairs/cell. Protein library 5,000 reads pairs/cell.

Paired-end, dual indexing:
Read 1: 28 cycles
i7 Index: 10 cycles
i5 Index: 10 cycles
Read 2: 90 cycles

Plots show the components of the library molecule being sequenced per cycle.

	MiSeq	NextSeq 500/550	NextSeq 2000	NovaSeq*	NovaSeq 6000 Xp*
Loading conc.	12 pM	2.5 pM	650 pM	150 pM	150 pM
Cluster density / Occupancy	762 K/mm^2	129 K/mm^2	87.0%**	97.3%**	93.3%**
% PF	96.0	95.9	78.7	85.3	85.7
% PhiX	5	5	5	10	10
% ≥Q30	R1: 99.1, i7: 97.8, i5: 98.6, R2: 96.7	R1: 98.2, i7: 97.7, i5: 98.7, R2: 94.9	R1: 96.4, i7: 96.4, i5: 98.2, R2: 83.1	R1: 95.6, i7: 96.3, i5: 92.5. R2: 89.9	R1: 96.0, i7: 96.9, i5: 93.9. R2: 91.1
Yield per lane	R1: 0.49 Gb R2: 1.60 Gb	R1: 8.69 Gb R2: 28.53 Gb	R1: 14.09 Gb R2: 46.42 Gb	R1: 29.33 Gb R2: 96.68 Gb	R1: 66.57 Gb R2: 219.42 Gb
Total read pairs	12.4 M	219.1 M	380.2 M	2938.5 M	1592.5 M
Mapped reads %	96	96	96	96	96
Multiplexing Probe BC Q30 %	97	97	96	93	93
Valid barcodes %	96.4	96.4	97.0	96.8	96.9

* S4 flow cell, 1 lane, ** Occupancy is reported, which use patterned flow cells, instead of cluster density.

Table 6. The data by cycle plots shown are from the library pool sequenced on one lane of a NovaSeq 6000 SP 100 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown for the 16-plex library (below black line). The multiplexing Probe BC Q30 % and valid barcode % metrics are only for Flex Gene Expression libraries.

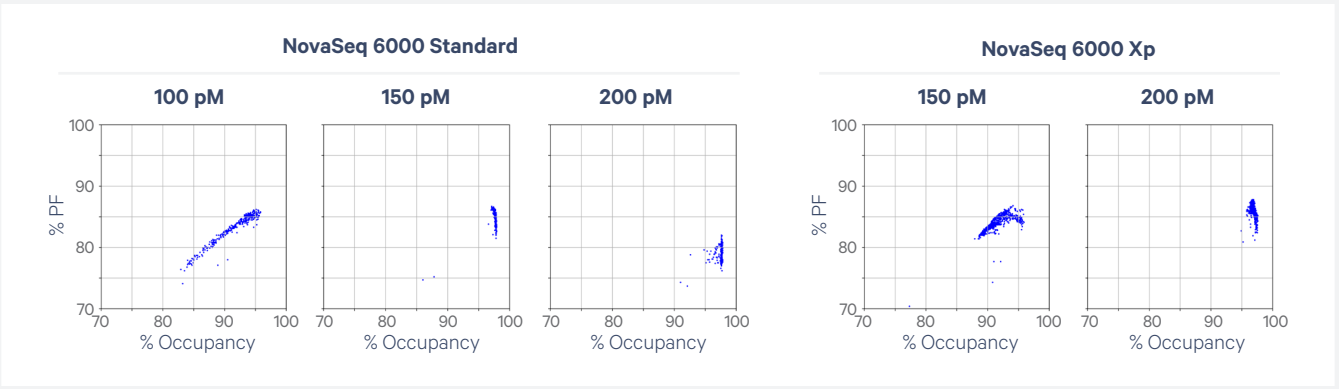
Loading Concentration for NovaSeq 6000 Instruments

Optimal loading concentration is a critical factor in achieving high-quality sequencing results.

NovaSeq 6000 Standard vs. NovaSeq 6000 Xp

Optimal loading concentrations for NovaSeq 6000 Standard and NovaSeq 6000 Xp workflows were evaluated using a comparative study with two 4-plex Flex Gene Expression libraries. For NovaSeq 6000 Standard, the flow cell exhibits slight underclustering at 100 pM and some overclustering at 150 pM. Therefore, the recommended concentration for optimal loading for the NovaSeq 6000 Standard workflow is between 100 pM and 150 pM.

For NovaSeq 6000 Xp, a loading concentration of 150 pM exhibits slight underclustering and slight overclustering at 200 pM. The recommended loading concentration range for the NovaSeq 6000 Xp workflow is between 150 pM and 200 pM.



	NovaSeq 6000 Standard	NovaSeq 6000 Xp
Loading concentration	150 pM	150 pM
Occupancy	93%	91%
% PF	85	83.8
% PhiX	10	10
% ≥Q30	R1: 96.8, i7: 96.9, i5: 96.5, R2: 84.6	R1: 96.1, i7: 97.0, i5: 96.4. R2: 80.2
Yield per lane	R1: 29.07 Gb R2: 95.82 Gb	R1: 65.25 Gb R2: 215.09 Gb
Total read pairs	3770.4 M	2147.6 M
Mapped reads %	96	96
Multiplexing Probe BC Q30 %	92	89
Valid barcodes %	96.6	95.9

Table 7. The secondary analysis metrics for one of the 4-plex libraries loaded at 150 pM are shown for NovaSeq 6000 Standard SP 100 cycles v1.5 and Xp S2 100 cycles v1.5 workflows. Sequencing metrics are shown for the flow cell (above black line) and Cell Ranger metrics are shown below the black line.

Loading Concentration Variability Between Flow Cells for the Same Instrument

There can be minor differences in clustering across different instruments, flow cells, and sequencing reagent lots (Figure 4). To account for variations in the percent occupied and percent passing filter metrics, it is advisable to adjust the loading concentrations for each new lot of flow cells or instrument.

Users of the NovaSeq 6000 instrument should titrate loading concentrations between 100–150 pM for the standard workflow, or between 150–200 pM for the Xp workflow.

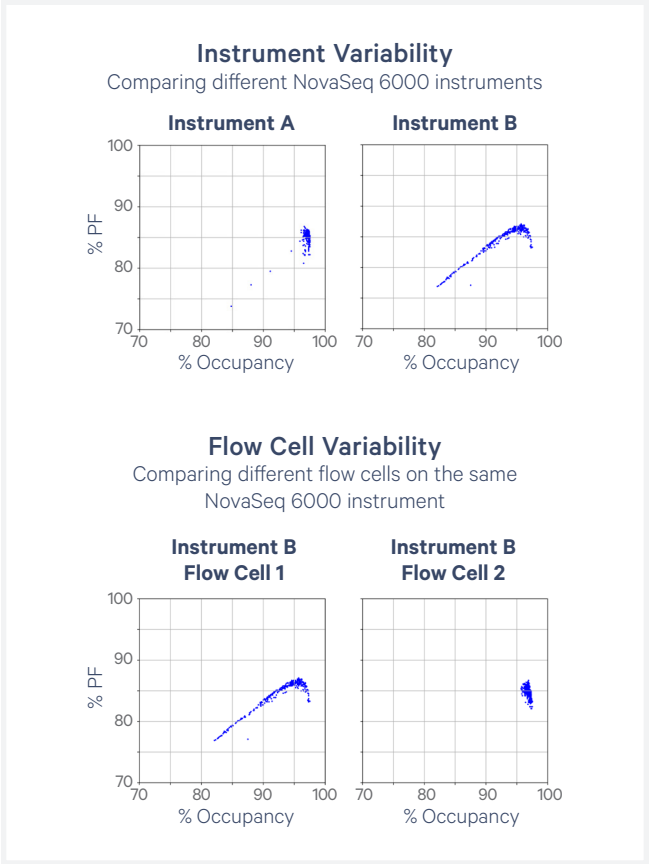


Figure 4. Clustering across different NovaSeq 6000 instruments (A and B) and across flow cell lots (1 and 2) run on the same instrument (B). Loading concentration 150 pM.

Library Quantification Accuracy

Accurate quantification of library concentration is required for precise loading. The loading concentrations in this Technical Note were determined using the KAPA qPCR DNA Quantification Kit, with average library size determined by the Agilent Bioanalyzer, Revvity LabChip, or Agilent TapeStation QC. Using alternative methods for final library quantification, such as Qubit instead of KAPA qPCR, may lead to underquantification and, consequently, overloading (see Figure 5).

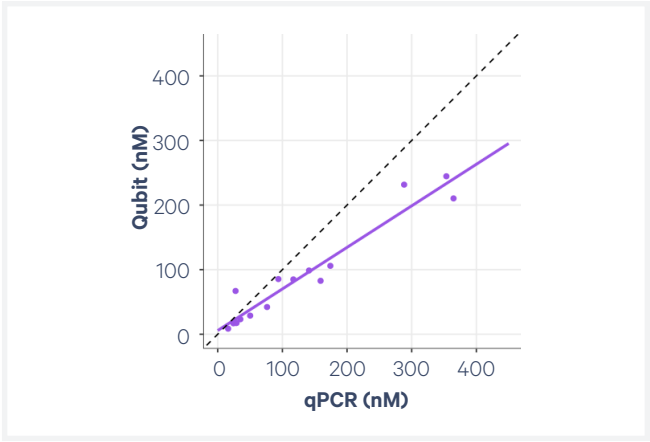


Figure 5. Comparison of final library concentration as measured by KAPA qPCR and Qubit. Gene Expression libraries were quantified using Qubit™ 1X dsDNA High Sensitivity Assay (ThermoFisher, Q33231) or by qPCR using the KAPA Library Quantification Kit for Illumina Platforms (Roche, KK4824). The average size of the libraries was calculated using LabChip using the DNA High Sensitivity Regent Kit (Revvity CLS760672) over 150–300 bp. Qubit systematically results in lower reported concentrations than KAPA qPCR, and can result in overloading flow cells.

Pooling for Different Library Types

Flex Gene Expression and Visium CytAssist libraries

For pooled Flex and CytAssist libraries, 10x recommends the following loading concentrations: 150 pM for NovaSeq, 11–12 pM for MiSeq, 650 pM for NextSeq 2000, and 1.8–2.5 pM for NextSeq 500/550.

Although the recommendation to use the same loading concentration for both libraries is based on data from NovaSeq 6000, it should be applicable to other sequencers as well.

10x does not advise loading Flex and CytAssist pools on the NovaSeq X, as the recommended concentrations differ significantly: 150–200 pM for Flex versus 300–400 pM for CytAssist.

Singleplex Flex Gene Expression libraries can be combined with CytAssist libraries in any ratio when 1% PhiX is included in the pool.

Multiplex Flex Gene Expression libraries can also be combined with CytAssist libraries in any ratio when 10% PhiX is included in the pool. However, it is advisable for at least 10% of the pool to consist of the Flex library.

Initially, it was recommended to avoid pooling CytAssist libraries with Multiplex Flex libraries due to concerns about potential sequencing interference. This recommendation has been updated based on further understanding; the polyA capture sequence present in CytAssist libraries (positions 51-80) does not interfere with the multiplexing Probe Barcode (positions 69-76), provided that 10% PhiX is included in the sequencing pool.

	Flex 10%, CytAssist 90%	Flex 90%, CytAssist 10%
Loading concentration	150 pM	150 pM
% Occupancy	98	97
% PF	80	86
% PhiX	10	10
% ≥Q30	R1: 94.7, i7: 95.6, i5: 90.7, R2: 76.6	R1: 95.5, i7: 96.5, i5: 95.6. R2: 85.0
Yield per lane	R1: 21.55 Gb R2: 90.83 Gb	R1: 132.82 Gb R2: 437.70 Gb
Total read pairs (Flex metrics)	105.3 M	3928.8 M
Multiplexing Probe BC Q30 % (Flex metrics)	85	93.3
Valid barcodes % (Flex metrics)	96	97

Table 8. The table shows two different library pooling combinations. Library pools comprising Flex Gene Expression (10%) and CytAssist (90%) were sequenced on one lane of a NovaSeq 6000 (Standard) SP 100 cycles v1.5 flow cell. The Flex Gene Expression (90%) and CytAssist (10%) data are from one lane of a NovaSeq 6000 (Standard) S2 100 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell above black line and Cell Ranger metrics are shown below black line. The multiplexing Probe BC Q30 %, total read pairs, and valid barcode metrics are only for Flex Gene Expression libraries.

Flex Gene Expression Combined with Either 3’ Gene Expression or 5’ Gene Expression

For pooled Flex and 3' or 5' Gene Expression libraries, 10x recommends the following loading concentrations: 150 pM for NovaSeq, 12 pM for MiSeq, 650 pM for NextSeq 2000, and 1.6–2.5 pM for NextSeq 500/550.

Although the recommendation to use the same loading concentration for both libraries is based on data from NovaSeq 6000, it should be applicable to other sequencers as well.

Singleplex Flex Gene Expression libraries can be combined with 3’/5’ libraries in any ratio when 1% PhiX is included in the pool.

Multiplex Flex Gene Expression libraries normally require 20% PhiX, but can be mixed with 1% PhiX if the pool comprises at least 20% 3’/5’ Gene Expression libraries.

Increasing PhiX concentration to 10% is not required because the 3’/5’ libraries contribute adequate diversity to the multiplexing Probe Barcode of Multiplex Flex Gene Expression libraries (positions 69-76).

Up to 80% Singleplex or Multiplex Flex Gene Expression libraries can be mixed with 3’/5’ libraries without affecting the sequencing quality of 3' or 5' libraries. Larger percentages of Flex libraries (>80%) may start to adversely affect the 3'/5' alignment and ultimately the 3'/5' data quality.

These pooling experiments were conducted only with 3' libraries using the Read 2 sequencing scheme. The Read 1 scheme is not recommended when pooling with 3'/5' GEX libraries due to the shorter Read 2 sequencing length (50 bp). 5' libraries are structurally similar to 3' and are expected to behave in the same manner.

	Flex 100%	Flex 80%, 3' GEX 20%	Flex 20%, 3' GEX 80%
Loading concentration	150 pM	150 pM	150 pM
Occupancy	97.2%	97.5%	96.5%
% PF	87.4	80.5	80
% PhIX	10	1	1
% ≥Q30	R1: 96.7, i7: 97.1, i5: 92.9, R2: 85.0	R1: 94.4, i7: 96.1, i5: 84.4, R2: 89.7	R1: 94.5, i7: 96.6, i5: 91.9. R2: 91.2
Yield per lane	R1: 30.12 Gb R2: 99.29 Gb	R1: 27.58 Gb R2: 90.87 Gb	R1: 27.41 Gb R2: 90.35 Gb
Total read pairs (Flex metrics)	972.4 M	643.5 M	187.2 M
Multiplexing Probe BC Q30 % (Flex metrics)	92	92.6	93.8
Valid barcodes % (Flex metrics)	96	96.8	97.1

Table 9. The data by cycle plots shown are from libraries sequenced on a NovaSeq 6000 SP 100 cycles v1.5. Flex metrics are highlighted in this table. Three different library pooling combinations (3’ Flex 100%, Flex 80% GEX 20%, Flex 20% 3’ GEX 80%) are shown. Sequencing metrics are shown for the flow cell above black line and Cell Ranger metrics are shown below the black line.

	3' GEX 100%	Flex 80%, 3' GEX 20%	Flex 20%, 3' GEX 80%
Loading concentration	300 pM	150 pM	150 pM
Occupancy	96.6%	97.5%	96.5%
% PF	76.3	80.5	80
% PhiX	1	1	1
% ≥Q30	R1: 94.3, i7: 97.2, i5: 91.9, R2: 90.8	R1: 94.4, i7: 96.1, i5: 84.4, R2: 89.7	R1: 94.5, i7: 96.6, i5: 91.9. R2: 91.2
Yield per lane	R1: 26.64 Gb R2: 87.75 Gb	R1: 27.58 Gb R2: 90.87 Gb	R1: 27.41 Gb R2: 90.35 Gb
Total read pairs (3' GEX metrics)	817.0 M	161.5 M	686.9 M
Q30 bases in RNA read % (3' GEX metrics)	90.8	84.3	90.4
Mapped reads % (3' GEX metrics)	77.7	77.5	78
Median genes per cell (3' GEX metrics)	1713	1662	1702
Median UMI counts per cell (3' GEX metrics)	3727	3595	3719

Table 10. The data by cycle plots shown are from libraries sequenced on a NovaSeq 6000 SP 100 cycles v1.5. 3' GEX metrics are highlighted in this table. Three different library pooling combinations (3' GEX 100%, Flex 80% GEX 20%, Flex 20% 3' GEX 80%) are shown. Sequencing metrics are shown for the flow cell above black line and Cell Ranger metrics are shown below the black line.

Conclusions

In summary, % Base by cycle, % ≥Q30 quality score, and metric distributions show highly consistent profiles for all sequencing platforms tested. The data serve as guidelines for assessing the sequencing quality of Flex Gene Expression and Protein Expression libraries. Additional factors that may contribute to the overall success of a sequencing run and impact downstream application performance metrics include:

- Sample preparation to obtain a high-quality single cell suspension.
- Final libraries with fragment lengths within the expected size range for each library type, for optimal cluster formation on Illumina flow cells.
- Reliable and accurate library quantification using the KAPA qPCR DNA Quantification Kit using the average insert size determined by Agilent Bioanalyzer, Revvity LabChip, or Agilent TapeStation QC. Alternative methods to KAPA qPCR, such as Qubit, for final library quantification may result in under quantification and, consequently, overloading (Figure 5).
- Sequencing platform loading concentration follows recommendations in the Flex User Guides, which are based on KAPA qPCR quantification. Overloading/overclustering may result in poor run performance, decrease sequencing quality, and result in lower total data output as compared to optimally loaded runs. The loading recommendations for an individual sequencer are listed as general guidance and additional optimization may be required.
- Visium V2 CytAssist libraries have a poly(A) stretch that will overlap the multiplexing Probe Barcode bases in Read 2, which could potentially lower nucleotide diversity if pooled with Flex libraries unless 10% PhiX is included in the sequencing pool.

References

1. Fixation of Cells & Nuclei for Chromium Fixed RNA Profiling Demonstrated Protocol (CG000478)
2. Tissue Fixation & Dissociation for Chromium Fixed RNA Profiling (CG000553)
3. Sample Preparation from FFPE Tissue Sections for Chromium Fixed RNA Profiling (CG000632)
4. Chromium Fixed RNA Profiling Reagent Kits for Singleplexed Samples User Guide (CG000691)
5. Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide (CG000527)
6. Chromium Fixed RNA Profiling Reagent Kits for Singleplexed Samples With Feature Barcode technology for Protein User Guide (CG000477)
7. Chromium Fixed RNA Profiling Reagent Kits for Singleplexed Samples with Feature Barcode technology for Protein using Barcode Oligo Capture User Guide (CG000674)
8. Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples with Feature Barcode technology for Protein using Barcode Oligo Capture User Guide (CG000673)
9. Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide (CG000527)
10. GEM-X Flex Gene Expression Reagent Kit for Singleplex samples (CG000786)
11. GEM-X Flex Gene Expression Reagent Kit for Multiplex samples (CG000787)
12. GEM-X Flex Gene Expression Reagent Kit for Singleplex samples with Feature Barcode technology for Protein Expression (CG000788)
13. GEM-X Flex Gene Expression Reagent Kit for Multiplex samples with Feature Barcode technology for Protein Expression (CG000789)
14. Visium CytAssist Spatial Gene Expression Reagent Kits (CG000495)

Document Revision Summary

Document Number	CG000677
Title	Sequencing Metrics & Base Composition of Chromium Flex Libraries
Revision	Rev B to Rev C
Revision Date	November 2024

Description of Changes:

- Updated Figure 1 to include Flex Protein Expression libraries (p. 1)
- Added comparison of Read 1 and Read 2 sequencing schemes (p. 7-8)
- Added data for Flex Protein Expression libraries (p. 9, 10)
- Updated NovaSeq 6000 loading concentration recommendations (p. 11-12)
- Added pooling guidance for Flex libraries with non-Flex libraries (p. 13-15)
- Changed Fixed RNA Profiling to Flex in assay and library names
- Updated for minor consistency of language and terms throughout

Contact:
support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588 USA