Sequencing Metrics & Base Composition of Chromium Fixed RNA Profiling Libraries

Introduction

The Chromium Single Cell Gene Expression Flex workflow produces sequencing-ready libraries for measuring RNA. This Technical Note presents a comparison of sequencing metrics for pooled Chromium Fixed RNA Profiling Gene Expression libraries across 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 on Illumina platforms. Individual results may vary depending on the specific sequencing instrument and/or particular sample and loading characteristics.

Chromium Fixed RNA Profiling offers comprehensive, scalable solutions to measure gene expression in formaldehyde fixed samples. The dual index Gene Expression libraries are generated using Fixed RNA Profiling reagents and protocols (see References).

As shown in Figure 1, the Fixed RNA Profiling Gene Expression libraries comprise standard Illumina paired-end constructs, which begin with P5 and end with P7. These libraries include 16 bp 10x GEM Barcodes (10x BC) encoded at the start of TruSeq Read 1 (Read 1T). Sample index sequences are incorporated as the i5 and i7 index reads. TruSeq Read 1 (Read 1T) and Small RNA Read 2 (Read 2S) are used in paired-end sequencing of Fixed RNA Gene Expression libraries. Read 2 contains the ligated probe insert sequence, constant sequence, and the 8 bp Probe Barcode that identifies the probe set used to hybridize the sample. 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 737k 10x GEM Barcodes).

Tables 1-3 show representative plots and sequencing data derived from pooled single and multiplexed (4-plex, 16-plex) sample libraries. The sequencing configuration and run parameters are provided for each dataset.

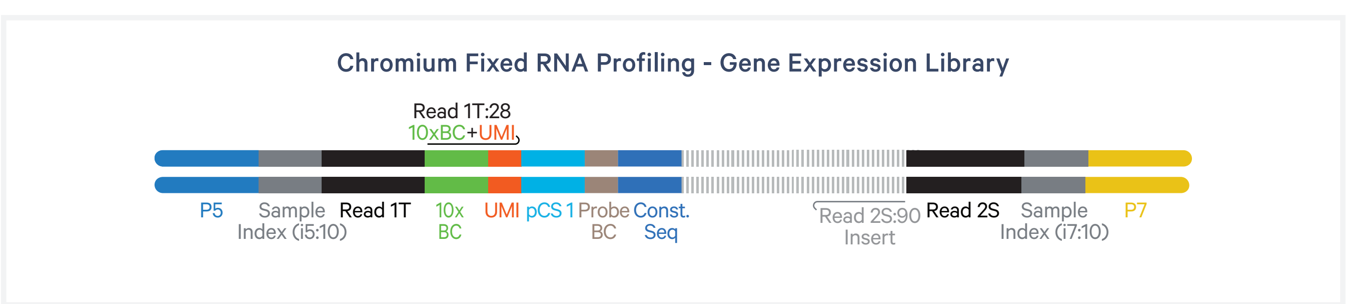


Figure 1. Schematic showing the Chromium Fixed RNA Profiling Gene Expression dual index library components.

Methods Overview

Three Fixed RNA Profiling Gene Expression library pools were generated using either the Chromium Fixed RNA Profiling Reagent Kits for Singleplexed Samples (Optional: with Feature Barcode technology for Cell Surface Protein) User Guide (CG000477) or Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide (CG000527).

- Pool A = four singleplex libraries (BC001)
- Pool B = four singleplex + one 4-plex library (BC001-BC004)
- Pool C = four singleplex + one 4-plex + one 16-plex library (BC001-BC016)

For each pool, the libraries were sequenced with the run parameters described in Tables 1-3. The library plots and metrics illustrate sequencer compatibility. All 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. Refer to 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

The 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 1-3. The "% ≥Q30" for R2 refers to the full 90 bp of read 2, and for multiplexed samples the "Multiplexing Probe BC Q30" refers to the 8 bp Probe Barcode within read 2. The "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, which use patterned flow cells, instead of cluster density.

These representative data demonstrate compatibility of Chromium Fixed RNA Profiling Gene Expression libraries with multiple Illumina sequencers.

Due to the probe design, there are some regions of the sequenced library where Q30 scores decline 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. However, this does not affect assay performance as Cell Ranger’s probe alignment algorithm uses the probe insert region (1-50 bp), as well as the Probe Barcode region in the case of multiplex libraries (69-76 bp). The aligner does not use the constant sequence-NN (51-68 bp) or pCS1-UMI (77-90 bp) regions.

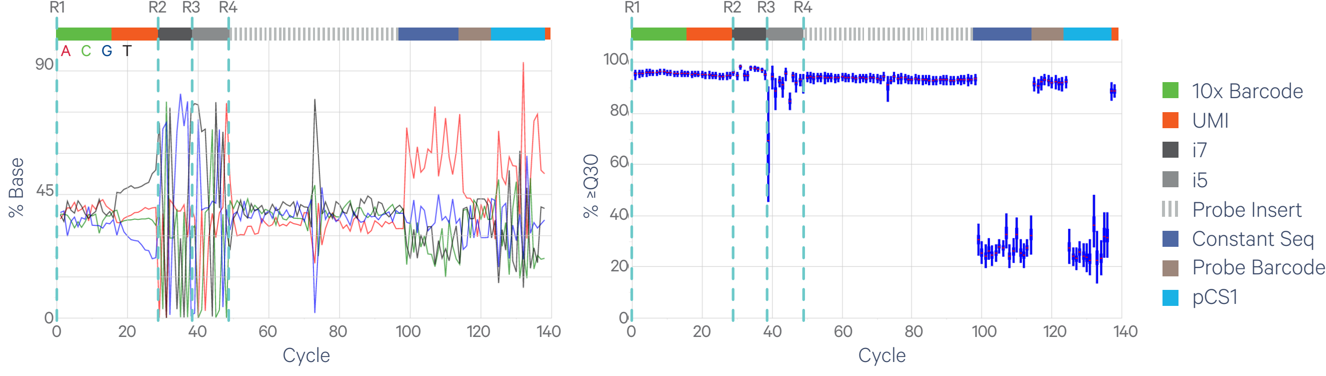
Pool A: Fixed RNA Profiling Gene Expression singleplex library

This pool consists of four singleplex libraries (BC001). The plots below show the components of the library molecule being sequenced per cycle.

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



Sequencer	Loading conc. (pM) Cluster density (K/mm^2) / Occupancy (%) % PF PhiX (%)	% ≥Q30	Yield per Lane (Gb)	Total read pairs (M)	Mapped Reads (%)	Valid barcodes (%)
MiSeq	Loading conc.: 10 Cluster density: 932 PF: 92.4% PhiX: 1%	R1: 98.0 i7: 97.4 i5: 97.8 R2: 96.1	R1: 0.57 R2: 1.87	20.7	96.6	98.4
NextSeq 550	Loading conc.: 2.5 Cluster density: 217.5 PF: 91.2% PhiX: 1%	R1: 96.7 i7: 97.1 i5: 97.3 R2: 69.3	R1: 13.96 R2: 45.97	512.6	96.7	98.5
NextSeq 2000	Loading conc.: 650 Occupancy: 96.5%* PF: 85.6% PhiX: 1%	R1: 96.3 i7: 97.6 i5: 96.2 R2: 75.2	R1: 15.38 R2: 50.70	570.1	97.4	98.5
NovaSeq (S4 flow cell, 1 lane)	Loading conc.: 150 Occupancy: 97.8%* PF: 85.1% PhiX: 1%	R1: 95.5 i7: 96.6 i5: 88.8 R2: 72.4	R1: 88.61 R2: 292.10	3259.3	97.1	98.5

* Occupancy is reported for the NextSeq 2000 and the NovaSeq, 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 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (left of line) and Cell Ranger metrics are shown for a singleplex library (right).

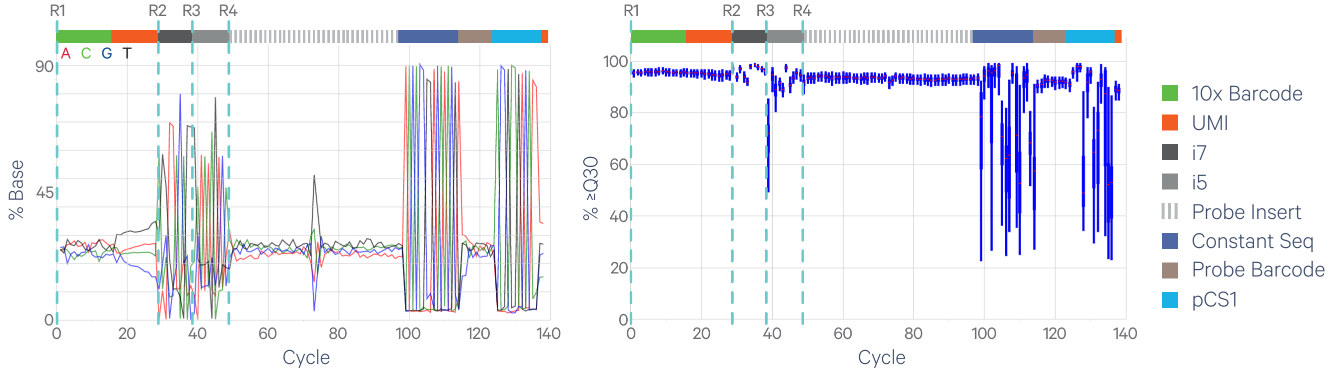
Pool B: Fixed RNA Profiling Gene Expression multiplexed 4-plex library

This pool consists of four singleplex libraries and a 4-plex library (BC001-BC004). The plots below show the components of the library molecule being sequenced per cycle.

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



Sequencer	Loading conc. (pM) Cluster density (K/mm²) / Occupancy (%) % PF PhiX (%)	% ≥Q30	Yield per Lane (Gb)	Total read pairs (M)	Mapped Reads (%)	Multiplexing Probe BC Q30	Multiplexing BC correction %	Valid barcodes (%)
MiSeq	Loading conc.: 12 Cluster density: 892 PF: 94.7% PhiX: 5%	R1: 98.7 i7: 95.7 i5: 96.9 R2: 96.0	R1: 0.56 R2: 1.84	20.3	98.4	96.2	2.6	96.6
NextSeq 550	Loading conc.: 2.5 Cluster density: 160.5 PF: 94.9% PhiX: 5%	R1: 97.8 i7: 97.9 i5: 97.6 R2: 75.2	R1: 10.71 R2: 35.26	394.5	98.7	88.8	4.7	94.2
NextSeq 2000	Loading conc.: 650 Occupancy: 92.3%* PF: 82.8% PhiX: 5%	R1: 96.6 i7: 96.3 i5: 96.1 R2: 74.2	R1: 14.88 R2: 48.91	551.4	98.8	91.0	5.9	94.5
NovaSeq (S4 flow cell, 1 lane)	Loading conc.: 150 Occupancy: 97.2%* PF: 86.8% PhiX: 10%	R1: 95.3 i7: 96.1 i5: 90.0 R2: 87.9	R1: 89.00 R2: 293.38	3324.4	98.6	92.9	3.1	96.8

* Occupancy is reported for the NextSeq 2000 and the NovaSeq, 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 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (left of line) and Cell Ranger metrics are shown for the 4-plex library (right).

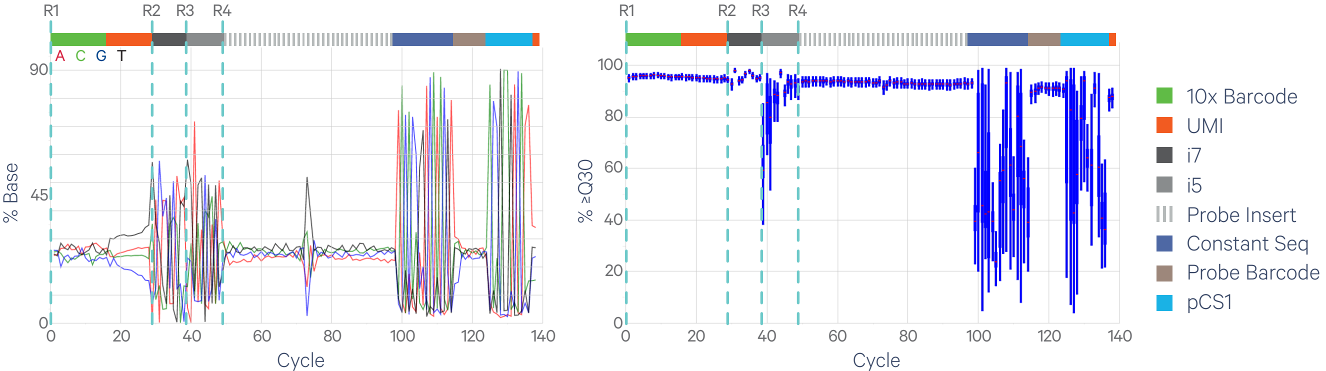
Pool C: Fixed RNA Profiling Gene Expression multiplexed 16-plex library

This pool consisted of four singleplex libraries, a 4-plex library, and a 16-plex library (BC001-BC016). The plots below show the components of the library molecule being sequenced per cycle.

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



Sequencer	Loading conc. (pM) Cluster density (K/mm²) / Occupancy (%) % PF PhiX (%)	% ≥Q30	Yield per Lane (Gb)	Total read pairs (M)	Mapped Reads (%)	Multiplexing Probe BC Q30	Multiplexing BC correction %	Valid barcodes (%)
MiSeq	Loading conc.: 12 Cluster density: 888 PF: 93.6% PhiX: 5%	R1: 98.3 i7: 96.6 i5a: 97.6 R2: 95.1	R1: 0.55 R2: 1.82	20.0	98.2	93.9	3.8	96.0
NextSeq 550	Loading conc.: 2.5 Cluster density: 174.3 PF: 93.8% PhiX: 5%	R1: 97.5 i7: 97.3 i5: 97.4 R2: 70.7	R1: 11.49 R2: 49.31	423.7	98.4	90.6	4.1	95.6
NextSeq 2000	Loading conc.: 650 Occupancy: 92.9%* PF: 83.6% PhiX: 5%	R1: 96.5 i7: 96.5 i5: 96.7 R2: 74.7	R1: 15.07 R2: 49.68	556.8	98.6	88.8	6.0	93.2
NovaSeq (S4 flow cell, 1 lane)	Loading conc.: 150 Occupancy: 97.5%* PF: 86.4% PhiX: 10%	R1: 95.5 i7: 95.5 i5: 85.9 R2: 81.8	R1: 85.72 R2: 282.57	3309.1	98.3	89.2	5.6	94.9

* Occupancy is reported for the NextSeq 2000 and the NovaSeq, 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 S4 200 cycles v1.5 flow cell. Sequencing metrics are shown for the flow cell (left of line) and Cell Ranger metrics are shown for the 16-plex library (right).

Figure 2 shows an example of a poor quality sequencing run for multiplexed libraries, sequenced with 3% PhiX on a NovaSeq. 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.

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 singleplexed libraries, the Probe Barcode region is not required and poor sequencing quality in this region is not expected to impact downstream analysis.

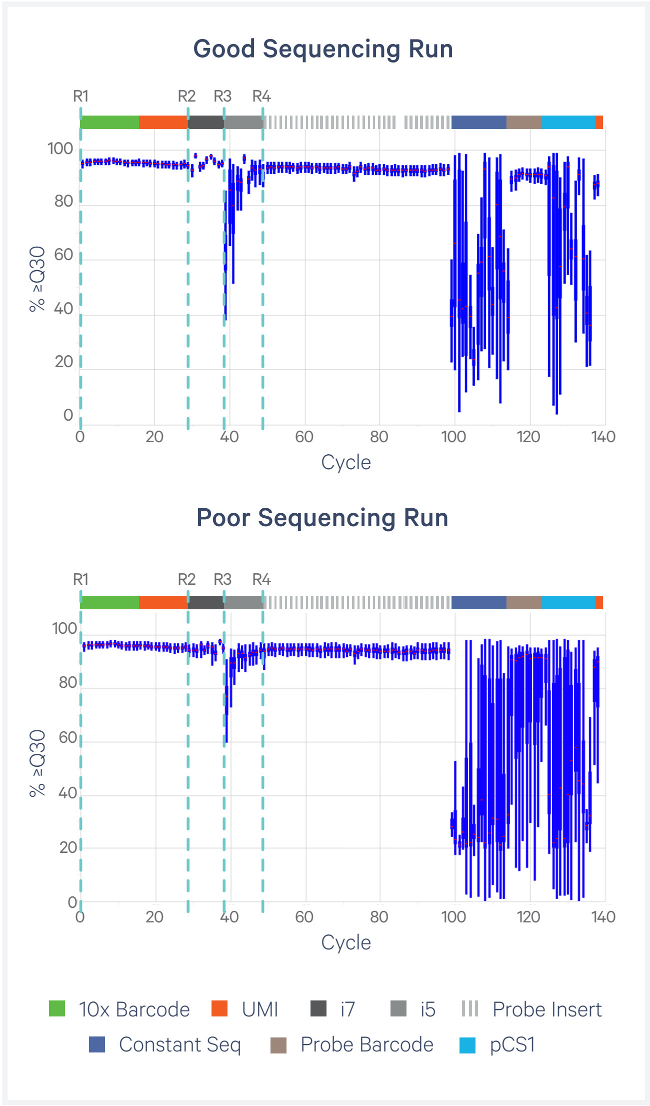


Figure 2. Example of a good and a poor % ≥Q30 quality scores by cycle plot for multiplexed libraries sequenced on a NovaSeq. The good sequencing run was sequenced with 10% PhiX. The poor sequencing run was sequenced with 3% PhiX, and the Q30 scores are degraded in the Probe Barcode region.

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 quality of Chromium Fixed RNA Profiling Gene Expression dual index 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 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, Perkin Elmer 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 3).
- Sequencing platform loading concentration follows recommendations in the Chromium Fixed RNA Profiling 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.
- Please note that Visium CytAssist libraries have a poly(A) stretch that will overlap the multiplexing Probe Barcode bases in Read2, which could potentially lower nucleotide diversity if pooled with Fixed RNA Profiling libraries.

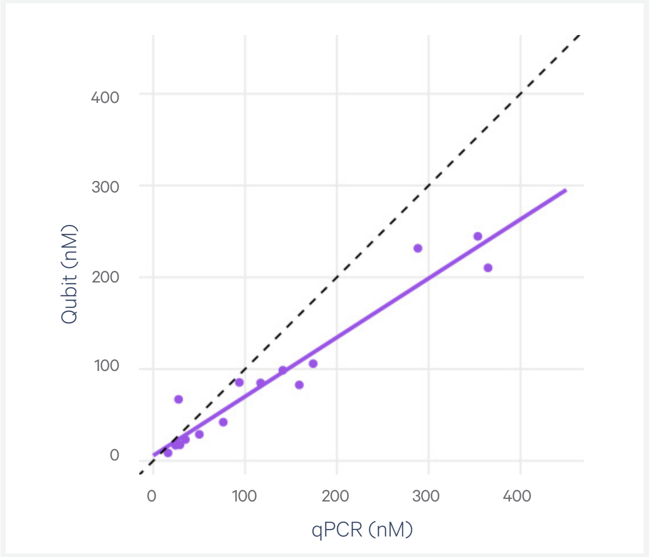


Figure 3. Comparison of final library concentration as measured by KAPA qPCR and Qubit. Gene Expression libraries were quantified using Qubit™ 1X dsDNA High Sensitivity Assay (ThermoFisher, Q33231) or by qPCR using the KAPA Library Quantification Kit for Illumina Platforms (Roche, KK4824). The average size of the libraries was calculated using LabChip using the DNA High Sensitivity Regent Kit (Perkin Elmer, CLS760672) over 150-300 bp. Qubit systematically results in lower reported concentrations than KAPA qPCR, and can result in overloading flow cells.

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. Isolation of Cells from FFPE Tissue Sections for Chromium Fixed RNA Profiling (CG000632)
4. Chromium Fixed RNA Profiling Reagent Kits for Singleplexed Samples (Optional: with Feature Barcode technology for Cell Surface Protein) User Guide (CG000477)
5. Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide (CG000527)

Document Revision Summary

Document Number	CG000677
Title	Sequencing Metrics & Base Composition of Chromium Fixed RNA Profiling Gene Expression Libraries
Revision	Rev A to Rev B
Revision Date	September 2023
General Changes	Update NovaSeq loading concentration recommendations.

© 2023 10x Genomics, Inc. (10x Genomics). All rights reserved. Duplication and/or reproduction of all or any portion of this document without the express written consent of 10x Genomics, 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. A non-exhaustive list of 10x Genomics' marks, many of which are registered in the United States and other countries can be viewed at: www.10xgenomics.com/trademarks. 10x Genomics may refer to the products or services offered by other companies by their brand name or company name solely for clarity, and does not claim any rights in those third-party marks or names. 10x Genomics products may be covered by one or more of the patents as indicated at: www.10xgenomics.com/patents. All products and services described herein are intended FOR RESEARCH USE ONLY and NOT FOR USE IN DIAGNOSTIC PROCEDURES.

The use of 10x Genomics products in practicing the methods set forth herein has not been validated by 10x Genomics, and such non-validated use is NOT COVERED BY 10X GENOMICS STANDARD WARRANTY, AND 10X GENOMICS 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 terms and conditions of sale for the Chromium Controller or the Chromium Single Cell 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 that it currently or will at any time in the future offer or in any way support any application set forth herein.

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