Application Note

Genomics

Agilent

Trusted Answers

Targeted Single-Cell Sequencing for the Nanopore Sequencing Platform

Using Agilent SureSelect technology for long-read sequencing

Authors

Gento Fujii, Masahide Seki, Junko Zenko, Ayako Suzuki, and Yutaka Suzuki
University of Tokyo
Kashiwa-shi, Chiba, Japan

Satoshi Namai, Rumiko Saito, and Fumiko Yoshizki
Agilent Technoliges, Inc.
Hachioji-shi, Tokyo, Japan

Abstract

Long-read, single-cell RNA sequencing allows transcript analysis to be performed at single-cell resolution while maintaining more accurate transcript information, such as isoforms and fusion transcripts. This application note introduces a method for target enrichment of a 10x Genomics Chromium single-cell library using Agilent SureSelect technology for Oxford Nanopore sequencing. This method improves the number of valid reads available for single-cell analysis and provides more efficient analysis of the genes of interest. Additionally, we also demonstrate that this method can be applied to Visium spatial analysis.

Introduction

Single-cell RNA-Seq (scRNA-Seq) has become widely used in recent years as a method to analyze gene expression at single-cell resolution. In conventional bulk analysis, transcripts from individual cells are pooled to obtain the overall average data, whereas single-cell analysis obtains and analyzes independent data for each cell. Therefore, scRNA-Seq unveils cell heterogeneity and can provide deeper insights for understanding a range of biological events. Single-cell data from various types of human cells are being collected under the Human Cell Atlas (HCA) consortium to understand human health and disease. In cancer research, single-cell technology is used to study tumor heterogeneity, tumor microenvironment, and clonality.

The 10x Genomics Chromium is a widely used platform for generating single-cell libraries. Using this platform, one cell and one bead with a cell barcode are enclosed in a single droplet. This allows molecules derived from each cell to be uniquely barcoded.

Although conventional scRNA-Seq is done using short-read sequencing, long-read sequencing technology can improve 3'-/5'- bias as well as detect splicing isoforms and fusion transcripts more precisely. Long-read sequencing, however, comes with higher costs and lower throughput. Therefore, when using long-read scRNA-Seq, efficient gene analysis is key. Agilent SureSelect reagent kits are hybridization capture-based target enrichment products that can be used for the enrichment of long-read sequencing (LRS) libraries.[1] Furthermore, these kits can also be applied to long-read scRNA-Seq.

In this application note, we demonstrate the capture of long-read scRNA-Seq libraries using the Chromium system with Agilent SureSelect reagent kits, followed by sequencing on the Oxford Nanopore platform. Additionally, using the same strategy, we introduce an example of target enrichment for a spatial library prepared with the 10x Genomics Visium platform.

Experimental

The workflow is summarized in Figure 1, and the materials used in this application note are listed in Table 1. Table 2 provides the primers and blockers required for this study.

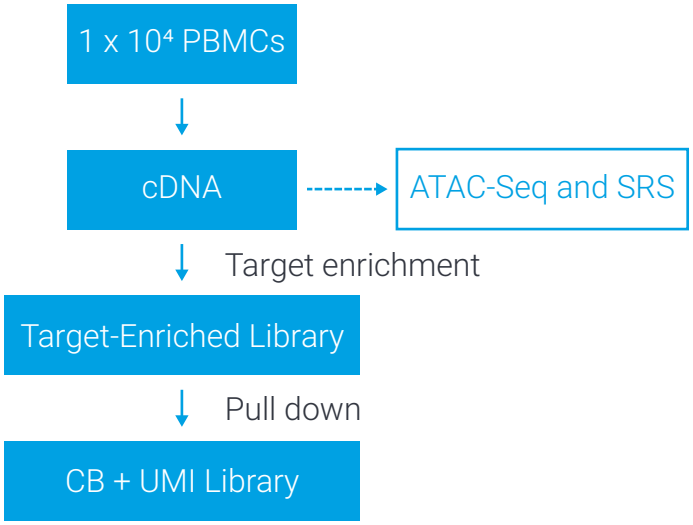

Figure 1. Overview of workflow. The steps explained in this application note are highlighted in blue, including the generation of cDNA from peripheral blood mononuclear cells (PBMCs), target enrichment, and the generation of libraries with cell barcodes (CBs) and unique molecular identifiers (UMIs). Assays for transposase-accessible chromatin sequencing (ATAC-Seq) and short-read sequencing (SRS) are not included. ATAC-Seq is optional, whereas SRS is required to acquire data for performing single-cell long-read (SiCeLoRe) data processing.[2]

Table 1. Materials used in this application note.

Materials	Vendor/Part Number	Comment
Single Cell Multiome ATAC + Gene Expression	10x Genomics, 100285 (4 reactions)	Single Cell 3' Kit can be also used
SureSelect XT HS Reagent Kit	Agilent, G9702A (16 reactions)	SureSelect XTHS2 DNA (p/n G9982A for 16 reactions) can also be used; Contact an Agilent representative
SureSelect Cancer CGP Capture Probe	Agilent, 5280-6990	Any HS probes of interest can be used
KOD FX Neo DNA Polymerase	Toyobo, KFX-201 (200 U)	Appropriate DNA polymerase for amplifying long fragments should be chosen
AMPure XP Beads	Beckman Coulter Genomics, A63880 (5 mL)	
SPRIselect Beads	Beckman Coulter Genomics, B23317 (5 mL)	
Dynabeads M-270 Streptavidin	Thermo Fisher Scientific, 65305 (2 mL)	Used for hybridization capture and wash
Dynabeads M-280 Streptavidin	Thermo Fisher Scientific, 60210 (2 mL)	Used for pull-down step
Ligation Sequencing Kit	Oxford Nanopore Technologies, SQK-LSK114	Any other ligation-based sequencing kit that is suitable for your flow cell can be used
10 N NaOH	Any qualified vendor	
1 M Tris-HCl (pH 7.5)	Any qualified vendor	
1 M Tris-HCl (pH 8.0)	Any qualified vendor	
5 M NaCl Solution	Any qualified vendor	
0.5 M EDTA (pH 8.0)	Any qualified vendor	
TapeStation High Sensitivity D5000 ScreenTape	Agilent	The appropriate automated electrophoresis system and instrument should be chosen (for example, Agilent HS Large Fragment kit for Agilent Fragment Analyzer systems, Agilent High Sensitivity DNA kit for Agilent Bioanalyzer system)
High Sensitivity D5000 Reagents	Agilent	

Table 2. Primers and blockers required.

Name	Sequence (5' to 3')
Illumina Read 1 Primer	CTACACGACGCTCTTCCGATCT
Template Switch Oligo (TSO) Primer	AAGCAGTGGTATCAACGCAGAGTACAT
Biotin-Illumina Read 1 Primer	Biotin-CTACACGACGCTCTTCCGATCT
TSO Blocker	CCCATGTACTCTGCGTTGATACCACTGCTT
10x Blocker	IIIIIIIIIIIIIIIIIIIIIIIIIIIIITTTTTTTTTTTTTTTTTTTTTTTTTTTTTT*

T: inverted dT*

Full-length cDNA synthesis

Peripheral blood mononuclear cells (PBMCs) were obtained from a healthy donor. Nuclei from 1 × 10^4 PBMCs were isolated according to the manufacturer's demonstrated protocol.[3]

After transposition, the generation of Gel Beads-in-emulsion (GEMs), barcoding, and solid-phase reversible immobilization (SPRI) purification were performed using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle according to manufacturer's instructions.[4] cDNA was purified using 60 µL of SPRIselect reagent (Beckman Coulter Genomics). The sample was incubated for 5 minutes at room temperature. The beads were washed twice with 200 µL of 80% ethanol. After residual ethanol was removed, the sample was dried for 2 minutes at room temperature and eluted in 40 µL of Buffer EB (Qiagen). The sample was then divided into two parts: one for short-read sequencing and one for long-read sequencing.

cDNA amplification

Barcoded cDNA from PBMCs was amplified by PCR. PCR reaction mixtures (50 µL split into two 25 µL reactions) contained 13 µL of template cDNA (10 ng), 25 µL of 2x PCR buffer for KOD FX Neo, 10 µL of 2 mM dNTPs, 1 U of KOD FX Neo (all from TOYOBO), and 0.5 µL each of 100 µM Illumina read 1 primer and TSO primer. The cycling conditions were as follows: 94 °C for 2 minutes, followed by 8 cycles of 94 °C for 15 seconds, 63 °C for 30 seconds, and 68 °C for 15 minutes. This was followed by a final extension at 68 °C for 15 minutes.

Each cDNA product was purified using 20 µL of AMPure XP (Beckman Coulter Genomics), and samples were incubated for 5 minutes at room temperature. The beads were then washed twice with 200 µL of 70% ethanol. After residual ethanol was removed, the samples were dried at 37 °C for 5 minutes and eluted in 17 µL of nuclease-free water. The two samples were combined and concentrated at room temperature for 45 minutes by the Eppendorf Concentrator plus. Next, the combined sample was resuspended in 10.5 µL of nuclease-free water.

The quality of purified and amplified cDNA was assessed using the Agilent High Sensitivity D5000 ScreenTape and reagent with the Agilent 4200 TapeStation system. Part of the precapture (unenriched) library was used for long-read sequencing.

Target enrichment

Amplified cDNA (500 ng) was hybridized using the Agilent SureSelect Cancer CGP DNA assay capture library (target size: 2.671 Mb, 679 genes) and the Agilent SureSelect XT HS reagent kit.

Briefly, 500 ng of amplified cDNA was adjusted to 9 µL, followed by the addition of 1.5 µL of 100 µM 10x blocker, 1.5 µL of 100 µM TSO blocker, and 5 µL of Agilent SureSelect XT HS blocker mix. The cDNA was then denatured at 95 °C for 30 seconds and incubated at 65 °C for 10 minutes.

While maintaining the samples at 65 °C, the cDNA with each blocker was mixed with 2 µL of 25% Agilent SureSelect RNase block mix, 2 µL of the SureSelect Cancer CGP DNA assay capture library, 6 µL of Agilent SureSelect Fast Hybridization buffer, and 3 µL of nuclease-free water. The cDNA was hybridized for 60 cycles of 65 °C for 1 minute and 37 °C for 3 seconds. This was followed by a final incubation at 65 °C for 10 minutes and then at 21 °C for up to 16 hours.

The capture process was performed according to the long-read sequencing library preparation method using the Agilent SureSelect XT HS2 target enrichment system.[1] Brefly, the capture process was performed using Dynabeads M-270 Streptavidin (Thermo Fisher Scientific). The beads were washed three times and resuspended with 200 µL of Agilent SureSelect binding buffer. The suspended beads were then heated to 68 °C for 10 minutes and mixed with the samples. For sample washing, DNA-bound beads were resuspended in 200 µL of SureSelect Wash Buffer 1, followed by the addition of 500 µL of SureSelect Wash Buffer 2 (prewarmed to 70 °C) and incubation at 70 °C for 5 minutes. A total of three washes were performed. Samples were eluted with 18 µL of freshly prepared 0.1 N NaOH and incubated for 10 minutes at room temperature. Then 18 µL of 200 mM Tris-HCl (pH 8.0) was added.

Pull-down (template switch oligo artifact removal)

Pull-down steps were based on the protocol published by Oxford Nanopore Technologies (ONT) to eliminate PCR artifacts and improve sequencing efficiency.[5]

The captured cDNA library was amplified and biotin-tagged by pre-pull-down PCR. For amplification, PCR reaction mixtures (200 µL split into two 100 µL reactions) contained 36 µL of template cDNA, 100 µL of 2x PCR buffer for KOD FX Neo, 40 µL of 2 mM dNTPs, 4 U of KOD FX Neo, 2 µL each of 100 µM biotin-Illumina read 1 primer and TSO primer, and 16 µL of nuclease-free water. The cycling conditions were as follows: 94 °C for 2 minutes, followed by 12 cycles at 94 °C for 15 seconds, 63 °C for 30 seconds, and 68 °C for 15 minutes. This was followed by a final extension at 68 °C for 15 minutes.

Amplified cDNA products were then combined and purified using 160 µL of AMPure XP (Beckman Coulter Genomics). The samples were incubated for 5 minutes at room temperature. The beads were washed twice with 200 µL of 70% ethanol. After residual ethanol was removed, the samples were dried at 37 °C for 5 minutes and eluted in 42 µL of nuclease-free water. The quality of the purified and amplified library was assessed using the High Sensitivity D5000 ScreenTape and reagent with the 4200 TapeStation system.

For pull-down, pull-down wash buffer (PWB) and capture beads were prepared. The recipe for 2x PWB is described in Table 3. A total of 7 mL 1x PWB was prepared by diluting 2x PWB twice with nuclease-free water. Pull-down beads were prepared by washing 5 µL of M-280 Streptavidin beads (10 µg/ µL) (Thermo Fisher Scientific) three times with 1,000 µL of 1x PWB, and then finally diluting them with 10 µL of 2x PWB.

Table 3. Composition of 2x pull-down wash buffer (PWB).

Reagent	Volume
1 M Tris-HCl (pH 7.5)	40 µL
5 M NaCl	1,600 µL
0.5 M EDTA (pH 8.0)	8 µL
Nuclease-Free Water	2,352 µL
Total	**4,000 µL**

Next, 10 µL of sample containing 100 ng amplified cDNA was mixed with 10 µL of prepared pull-down beads. After mixing with bead suspensions, samples were incubated at room temperature for 20 minutes using a rotator. For washing, 1,000 µL of 1x PWB was added and vortexed for 5 seconds. A total of three washes were performed, followed by the addition of 200 µL of 10 mM Tris-HCl (pH 7.5) and vortexing for 5 seconds. After, 13 µL of nuclease-free water was added to the bead-bound samples.

For post-pull-down amplification, PCR reaction mixtures (50 µL) contained 13 µL of bead-bound pull-down sampels, 25 µL of 2x PCR buffer for KOD FX Neo, 10 µL of 2 mM dNTPs, 1 U of KOD FX Neo, and 0.5 µL each of 100 µM Illumina read 1 primer and TSO primer. The cycling conditions were as follows: 94 °C for 2 minutes, followed by 4 cycles of 94 °C for 15 seconds, 63 °C for 30 seconds, and 68 °C for 15 minutes. This was followed by a final extension at 68 °C for 15 minutes.

Purification of 50 µL of post-pull-down products was performed using 30 µL of AMPure XP. The samples were incubated for 5 minutes at room temperature. The beads were washed twice with 200 µL of 70% ethanol. After residual ethanol was removed, the samples were dried at 37 °C for 5 minutes and eluted in 17 µL of nuclease-free water. The quality of the purified and amplified library was assessed using the High Sensitivity D5000 ScreenTape and reagent with the 4200 TapeStation system.

Sequencing and analysis

Using the 40 ng (53 fmol) prepared sample, the library for the PromethION (Oxford Nanopore Technologies) was prepared with the Ligation Sequencing Kit (Oxford Nanopore Technologies) according to the manufacturer's instructions. The libraries were loaded onto the PromethION with Flow Cell (part number FLO-PRO114M). The sequencing data was base called using Guppy 6.5.7.

FeatureCounts metrics were analyzed using Minimap2 (v2.24-r1122), featureCounts (v2.0.3), and Picard RNASeqMetrics (v3.1.1).

Results and discussion

Protocol modifications for 10x Genomics cDNA library capture

Procedures for capturing cDNA libraries were based primarily on the protocol described in the application note for long-read sequencing capture.[1] Because cDNA libraries include shorter fragments, the size-selection procedure was skipped.

The 10x Genomics Chromium single-cell (and Visium spatial) library contains TSO artifacts, which include TSO sequencing at both ends and lack barcodes and UMIs. These artifacts reduce the number of reads available for analysis (Figure 2). The pull-down step was added after hybrdization and capture. In this step, fragments containing a CB and UMI were tagged by PCR using the biotinylated Illumina read 1 primer. The target fragment was then pulled down using streptavidin-beads, and TSO artifacts were washed away. The library length distributions before capture (unenriched library) and after capture and pull-down (targeted library) show no significant differences (Figure 3).

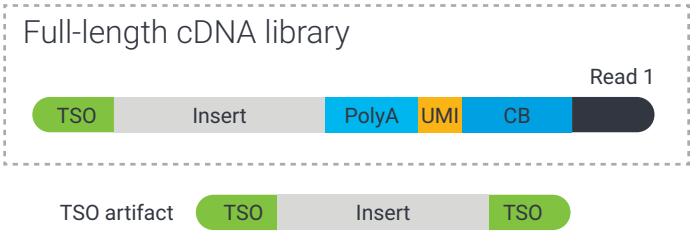

Figure 2. Structure of the 10x Genomics single-cell library. Inserts (gray) are surrounded by sequencing and analysis elements, including the template switching oligo (TSO, green), polyA (light blue), unique molecular identifier (UMI, orange), cell barcode (CB, blue), and Read 1 (black). The 10x Genomics spatial library obtained from Visium has the same structure and is processed using the same strategy.

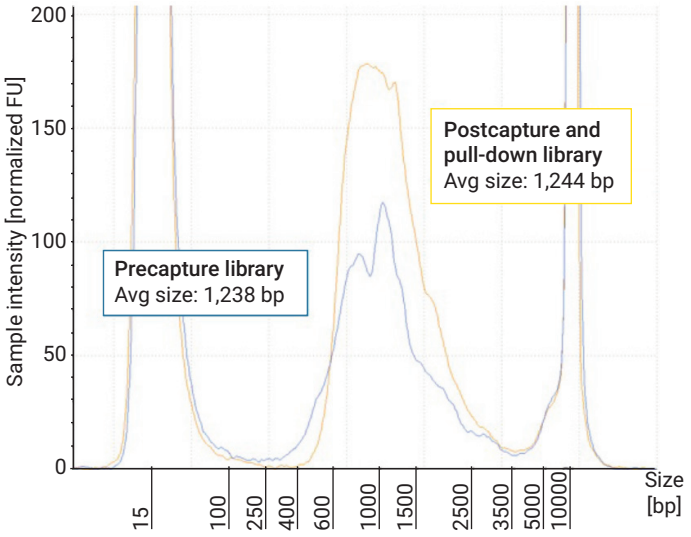

Figure 3. Electropherogram examples of libraries. Electropherograms of the precapture cDNA library (blue) and postcapture, pull-down library (orange) acquired using the Agilent TapeStation High Sensitivity D5000 ScreenTape assay and Agilent 4200 TapeStation system.

Sequencing results

Libraries were prepared for ONT sequencing using the Ligation Sequencing Kit (Oxford Nanopore Technologies) and then sequenced. Results were processed using the SiCeLoRe suite of tools, which facilitates error correction of ONT sequencing by using short-read sequencing data from the unenriched scRNA-Seq library.[2] Sequencing and analysis results are summarized in Table 4.

Table 4. Sequence and analysis results summary.

Object	Description	Unenriched Library	Targeted Library
Total Reads	No. of reads	97,607,035	107,604,305
	Total bp	80,368,537,295	98,718,514,096
	N50	882	938
	1D max length (bp)	1,540,991	1,366,182
	1D average length (bp)	823	917
	1D median length (bp)	750	821
	Average Q score	14.2	14.2
1D Pass Reads (mean_Q score ≥ 10)	No. of reads	88,548,773	97,605,413
	Max length	466.281	127,922
	Average length	828	918
	Average Q score	15.0	15.0
1D Pass and CB + UMI Reads (Processed Using SiCeLoRe)	No. of reads	**21,932,370 (24.7% of 1D pass reads)**	**63,603,994 (65.1% of 1D pass reads)**
	Assigned (to transcript)	14,908,626	56,727,993
	Unassigned (no features)	4,809,450	676,466
	Unassigned (ambiguity)	2,214,294	6,199,535
	Unassigned (unmapped)	0	0
On-Target Reads	No. of reads (panel-related)	**322,093 (0.36% of 1D pass reads)**	**34,931,913 (35.70% of 1D pass reads)**
	No. of unique UMIs	245,779	480,981
	Average no. of reads/UMI	**1.31**	**72.6**
	Read length (average)	838.0 bp	918.9 bp
	Read length (min)	200 bp	200 bp
	Read length (max)	99,719 bp	99,898 bp
	No. of > 5,000 bp reads	**24**	**1,328**

In the unenriched library, the proportion of reads with a CB and UMI that are valid for single-cell analysis is low at 24.7%. In contrast, in the captured (targeted) library, the proportion of valid reads is 2.6-fold higher at 65.1% (Figure 4). The proportion of assigned reads is 3.3-fold higher in the targeted library, indicating that pull-down significantly improves the number of valid reads and enables more efficient data acquisition.

In the unenriched library, only 0.36% of total reads are mapped to the target genes, whereas 35.7% of reads are mapped in the targeted library, corresponding to approximately a 100-fold increase (Figure 4). Furthermore, the majority of target genes exhibit positive fold changes, whereas off-target genes exhibit strong negative fold changes, indicating that target genes are efficiently enriched (Figure 5A). Some genes show no change, which is thought to be due to low expression levels initially or to the effect of cDNA length.

The correlation of the number of genes detected per CB is also high between the unenriched and targeted libraries (R = 0.94), as shown in Figure 5B. This suggests that no bias was caused by the capture and pull-down process. The number of UMIs differs by only two-fold when comparing the unenriched and targeted libraries (Table 4). However, the higher number of reads per UMI in the targeted library is thought to contribute to improved sequence accuracy.

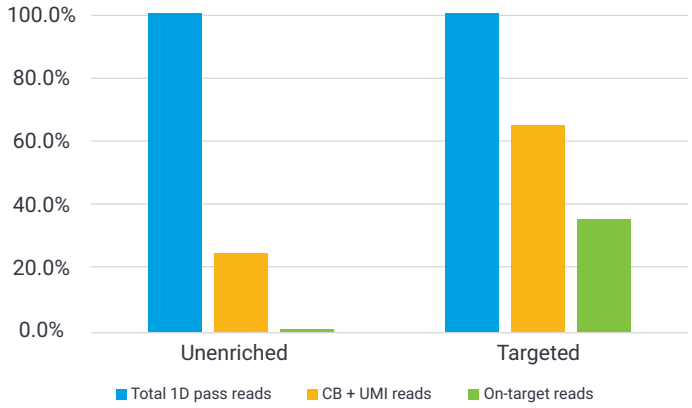

Figure 4. Ratio of reads available for single-cell and target sequencing analysis (summarized from Table 4).

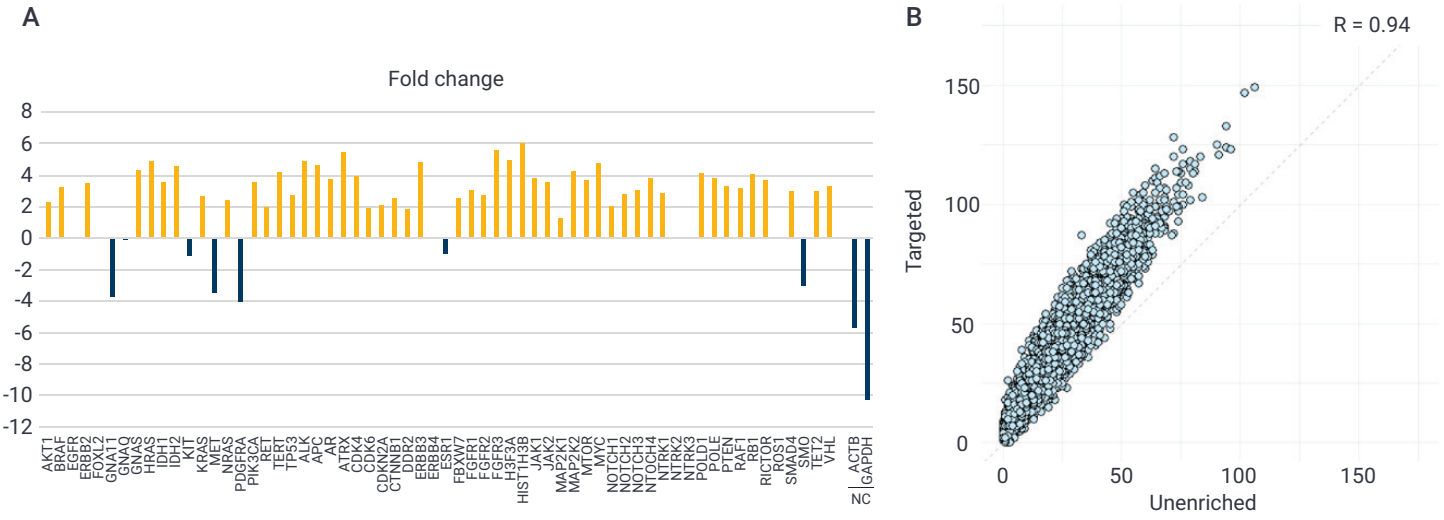

Figure 5. Enrichment efficiency and correlation between targeted and unenriched single-cell long-read sequencing. A) Log2 fold changes of selected targeted genes. ACTB and GAPDH are examples of nontargeted transcripts (orange: positive fold change; blue: negative fold change). (B) Correlation of the number of targeted genes detected per CB between the unenriched and targeted libraries.

Characteristics of long-read scRNA-Seq data

The average read length did not differ significantly, with an average of 838 bp in the unenriched library and 919 bp in the targeted library. However, there is a significant increase in the number of > 5,000 bp reads in the targeted library (Table 4), suggesting that longer cDNAs are analyzed efficiently. Figure 6 shows examples of longer target regions (JAK1 and TET2), demonstrating that the longer transcripts are covered by a single read and that reads are efficiently enriched at these target regions.

The coverage of the unenriched library is biased at the 3' end, which indicates that immature cDNA transcripts are included. The bias is improved in the captured (targeted) library, where uniform cDNA coverage has been obtained (Figure 7).

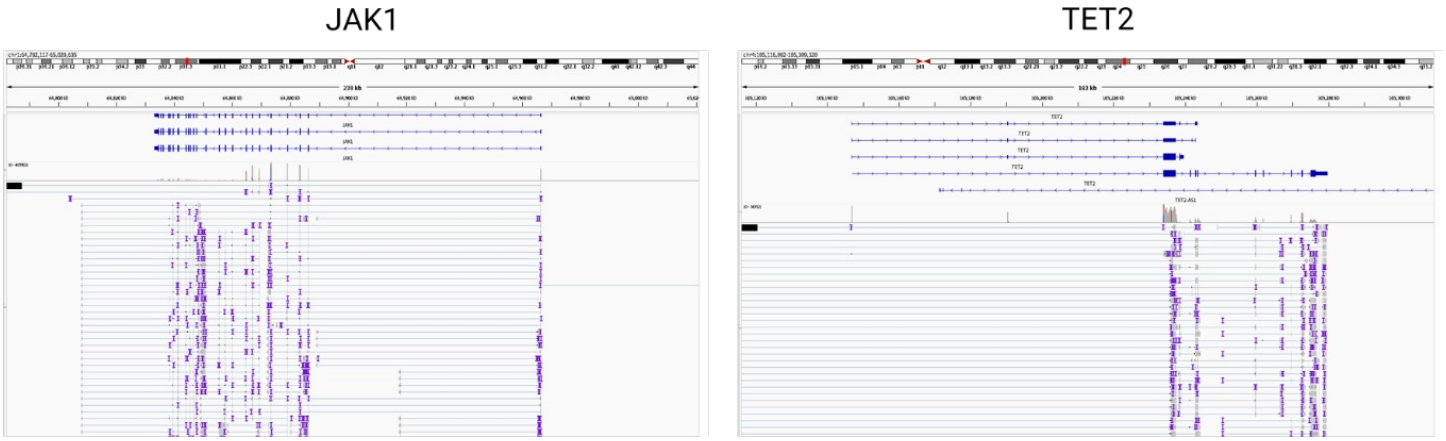

Figure 6. Examples of long target transcripts. Integrative Genomics Viewer (IGV) displays of JAK1 (5092 nt: ENST00000342505.5) and TET2 (9589 nt: ENST00000380013.9) are shown as examples of successfully captured long transcripts.

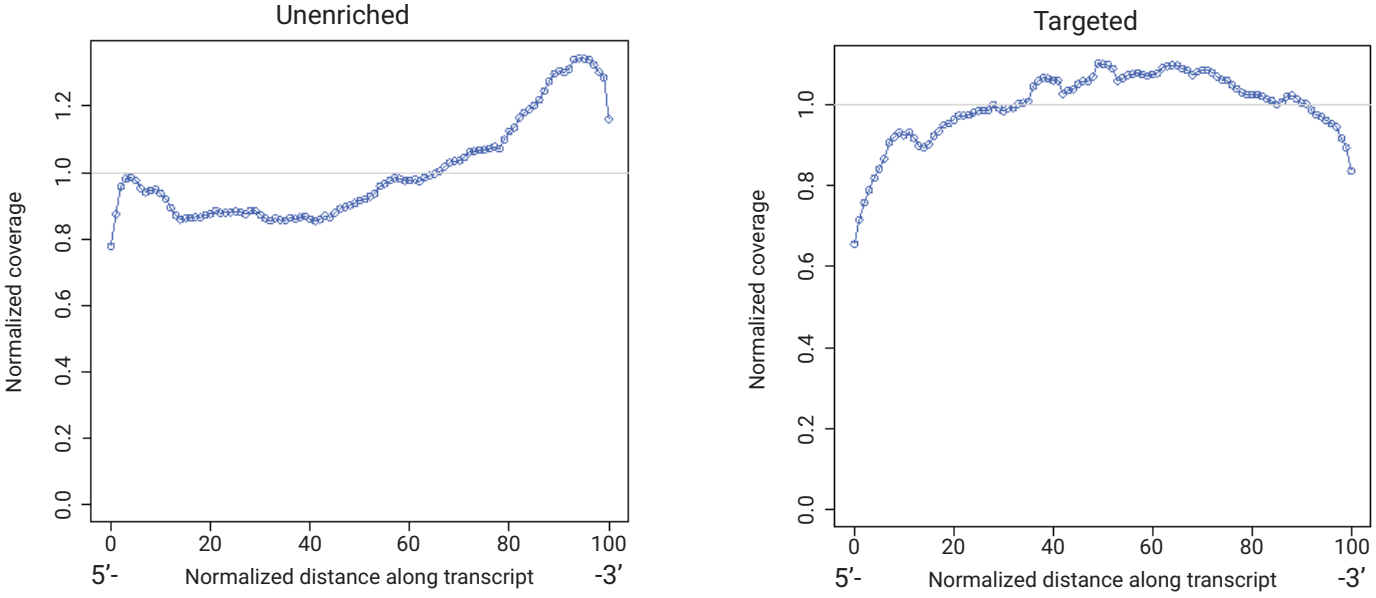

Figure 7. Coverage distributions in unenriched and targeted libraries, normalized by transcript length.

Target enrichment of the Visium spatial library

The 10x Genomics Visium is a platform for spatial transcriptome analysis with a library structure identical to Chromium (Figure 2). Thus, target enrichment of Chromium-spatial libraries is theoretically possible using the method described in this application note. As a proof of concept, target enrichment of a Visium spatial long-read library was performed using the previous version of this protocol.

Because typical Visium analysis is done with short-read sequencing of 3' ends, mutations far from the 3' ends cannot be detected. Figure 8 shows an example of long-read spatial analysis.[6] Using this method, a full-length cDNA sequence was analyzed with a spatial barcode, and a deletion located over 1,000 bp from the 3' end was detected with spatial information.

Despite several differences in experimental conditions for Visium analysis, including skipping the pull-down step, more efficient spatial analysis can be performed using the latest method described in this application note.

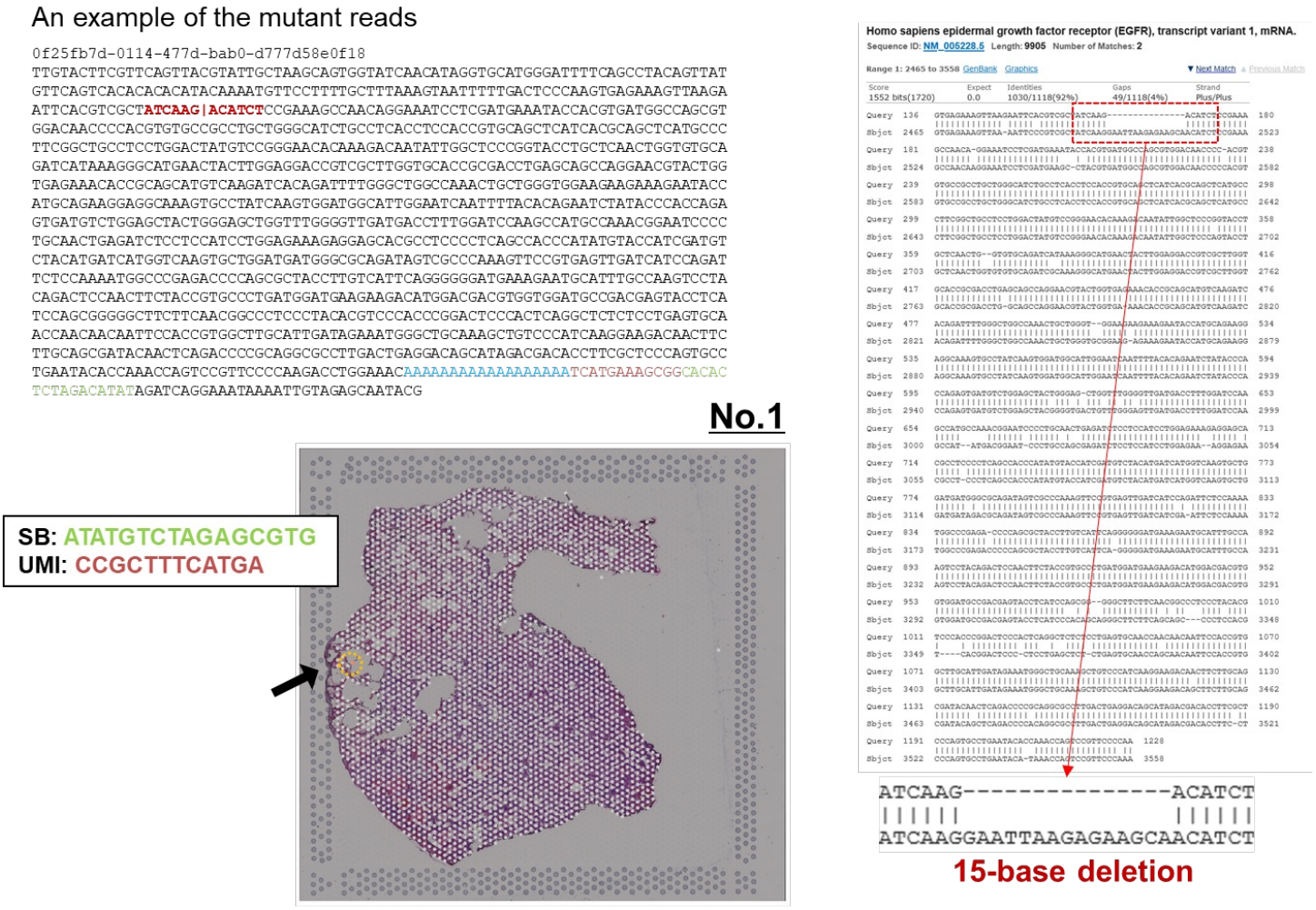

Figure 8. Example of target enrichment for a spatial library. A mutation located far from the 3' end was detected.

Conclusion

In this application note, the Agilent SureSelect target enrichment system was shown to be applicable for single-cell libraries prepared using the 10x Genomics Chromium system and sequenced with Oxford Nanopore Technology long-read sequencing. This method enables efficient single-cell long-read analysis by increasing the ratio of valid reads and the coverage for genes of interest.

Additionally, this method improves sequencing accuracy through high coverage, efficiently obtains full-length cDNA reads, and can also be applied to 10x Genomics Visium spatial libraries with the same library structure.

References

1. Namai, S.; Ishii, Y.; Yoshizaki, F.; Seki, M.; Suzuki, Y. Targeted Capture Long-Read Sequencing Using the Agilent SureSelect XT HS2 Target Enrichment System. *Agilent Technologies application note*, publication number 5994-7612EN, **2024**.
2. Lebrigand, K.; Magnone, V.; Barbry, P.; Waldmann, R. High throughput error corrected Nanopore single cell transcriptome sequencing. *Nat. Commun*. **2020**, *11*, 4025. DOI: 10.1038/s41467-020-17800-6.
3. Demonstrated Protocol Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing; *10x Genomics*, publication number CG000365 Rev C, **2022**.
4. User Guide: Chromium Next GEM Single Cell Multiome ATAC + Gene Expression; *10x Genomics*, publication number CG000338 Rev F, **2022**.
5. Single-cell transcriptomics with cDNA prepared using 10X Genomics; *Oxford Nanopore Technologies*, publication number SST_v9148_v111_revB_12Jan2022, **2022**.
6. Takano, Y.; Suzuki, J.; Nomura, K.; Fujii, G.; Zenkoh, J.; Kawai, H.; Kuze, Y.; Kashima, Y.; Nagasawa, S.; Nakamura, Y.; et al. Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat. Commun*. **2024**, *15*, 10637. DOI: 10.1038/s41467-024-54671-7.

www.agilent.com

PR7001-4615

Published in the USA, July 5, 2025
5994-8490EN

Agilent
Trusted Answers