Application Note

Genomics

Agilent
Trusted Answers

Target Capture Long-Read Sequencing Using the Agilent SureSelect XT HS2 Target Enrichment System

Authors

Satoshi Namai, Yoshiyuki Ishii, and Fumiko Yoshizaki, Agilent Technologies, Inc.

Masahide Seki and Yutaka Suzuki, University of Tokyo

Abstract

Target enrichment in long-read sequencing is critical for cost-effective and high-throughput analysis of regions of interest. This application note introduces a method for target capture sequencing using the Agilent SureSelect target enrichment system with some protocol modifications. Target-enriched 5 kb libraries for Nanopore sequencing were obtained using this method and sequencing results demonstrated high on-target percentage and excellent coverage.

Introduction

Long-read sequencing (LRS), selected as a technology of the year 2022 by Nature Methods[1], is a powerful technology for analyzing genome sequences that are difficult to study with short-read sequencing (SRS), such as repetitive sequences, pharmacogenomics-related genes, and complex structural variations. Challenges remain in terms of cost-effectiveness and throughput. Additionally, there is a demand for a targeted approach when specific genes or regions are of interest. For example, applications such as cancer can leverage target capture sequencing to efficiently enrich the regions of interest and achieve efficient sequencing.

The Agilent SureSelect XT HS2 target enrichment system is a hybridization-capture-based technology that uses SureSelect probes to enrich regions of interest and is optimized for SRS. For long read sequencing, we previously used methods published by PacBio and Oxford Nanopore that included an insert length of approximately 2 kb and required several micrograms of DNA.[2] In this application note, we demonstrate an improved target enrichment method for LRS, achieving 5.3 kb N50 length using 500 ng of input DNA with Agilent SureSelect Cancer CGP probes.

Materials and Method

The workflow was modified[3] to generate long-read sequencing libraries, summarized in Figure 1. Throughout the experiment, pipetting was used for sample mixing to avoid physical fragmentation of long DNA by vortexing, especially at the hybridization and capture steps.

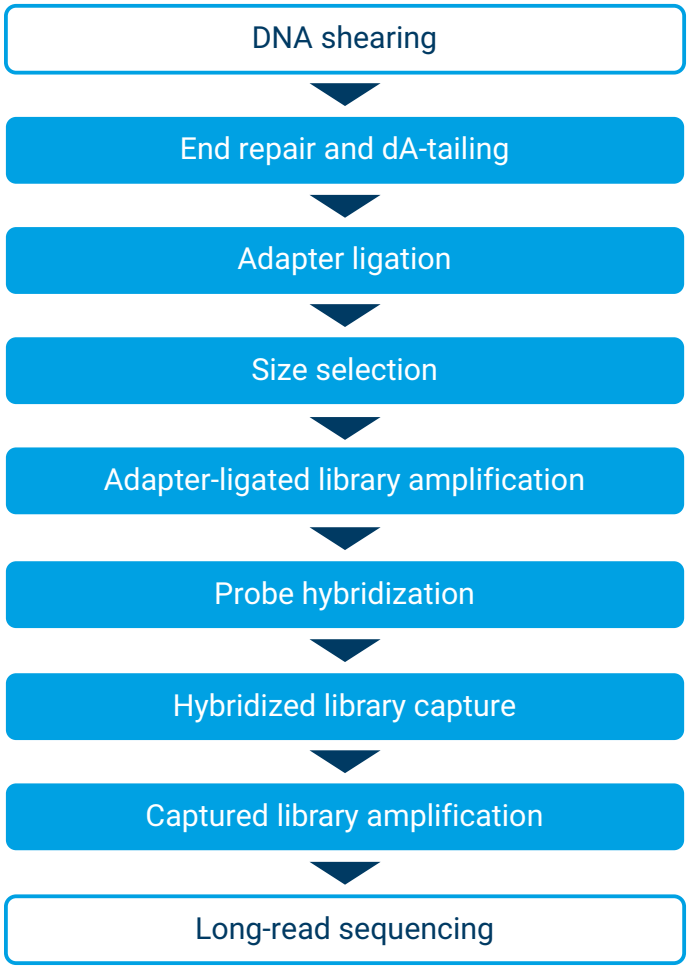


Figure 1. Target capture long-read sequencing library preparation workflow. Modifications to the SureSelect XT HS2 DNA workflow are described in the following sections: DNA shearing size, size selection, pre-capture amplification, denature condition before hybridization, capture condition, elution instead of on-beads during PCR, and post-capture amplification.

Table 1. Required materials.

Material	Vendor	Part Number	Notes
g-TUBE	Covaris	52009	
10N NaOH	Sigma Aldrich	72068	Or equivalent
1M Tris HCl pH 8.0	Sigma Aldrich	T3038	Or equivalent
KOD FX Neo	Toyobo	KFX201	KOD One; KOD Extreme for alternative (not tested)
Dynabeads M-270	Thermo Fisher	DB65305	
Native Barcoding Kit 24 V14	Oxford Nanopore	SQK-NBD114.24	
HS Large Fragment 50 kb kit	Agilent	DNF-464-0500	
SureSelect XT HS2 DNA reagent kit with index primer paris 1-16, 16 rxn	Agilent	G9981A	Includes hybridization module
OR			
SureSelect XT HS2 library preparation kit with MBC-free adaptors and index primer, 96 rxn		G9956A-D	Hyb reagents need to be ordered separately (see below)
SureSelect XT HS2 target enrichment kit, overnight hybridization, 96 hybridizations	Agilent	G9957B	
OR			
SureSelect XT HS2 target enrichment kit, 96 hybridizations	Agilent	G9987B	
Target enrichment probes	Agilent		Please contact your local sales representative

1) DNA shearing

OneSeq Human Reference DNA, Female (Agilent, p/n 5190-8850) was sheared with a g-TUBE (Covaris, p/n 52009) according to the manufacturer's protocol for 10 kb shearing. Briefly, 15 µg of DNA was diluted to 150 µL in nuclease-free water and transferred into the g-TUBE. The tube was centrifuged at 7,200 rpm for 1 minute with a 5415R centrifuge (Eppendorf). Then the tube was inverted and recentrifuged using the same parameters. To obtain 10 kb fragments with other input amounts, the 5415R centrifuge can be used with 8 µg DNA in 150 µL of nuclease-free water at 6000 rpm and less than 4 µg DNA in 150 µL at 6000 rpm. According to Covaris, the g-TUBE has been tested with DNA input as low as 100 ng and may require concentration before library preparation. After collecting data for this application note, successful feedback indicated starting with 2 µg DNA in 150 µL of nuclease-free water at 6000 rpm, followed by 50 µL for subsequent library preparation.

Sheared DNA was evaluated using the Agilent 5200 Fragment Analyzer system (p/n: M5310AA) with the Agilent FA 12-Capillary Array Short, 33 cm (p/n: A2300-1250-3355) and Agilent HS Large Fragment 50kb kit (p/n: DNF-464-0500).

2) Library preparation

End repair, dA-tailing, and adapter ligation were performed using the Agilent SureSelect XT HS2 reagent kit according to the manufacturer's protocol.[3] 500 ng of sheared DNA was used for the reaction. Adapter-ligated samples were purified using 0.8X volume of AMPure XP according to the SureSelect XT HS2 protocol.[3]

3) Size selection

Size selection was conducted based on the PacBio size selection protocol.[5] Purified samples were mixed with 129.5 µL of 35% v/v AMPure XP beads (prepared by dilution with 10 mM Tris HCl (pH 8.0) and incubated at room temperature (RT) for 15 minutes to allow for binding. The bead pellet was washed with 80% ethanol twice, then eluted with 19 µL of 10 mM Tris HCl (pH 8.0) with a 2-minute incubation. 18 µL of supernatant was transferred to a new tube.

4) Adapter-ligated library amplification

KOD FX Neo (TOYOBO), an enzyme solution that can amplify long DNA fragments, was used for PCR amplification. The composition and thermal cycling program are described in Table 2 and Table 3, respectively.

Table 2. Composition of pre-capture PCR solution.

Reagent	Volume/1 Rxn
2X PCR buffer for KOD FX Neo	50 µL
2 mM dNTPs	20 µL
SureSelect XT HS2 index primer	10 µL
KOD FX Neo (1.0 U/µL)	2 µL
Adapter-ligated DNA samples	18 µL

Table 3. Thermal cycling program for pre-capture PCR.

Step	Number of Cycles	Temperature	Time
1	1	94 °C	2 minutes
2	10*	94 °C	15 seconds
		60 °C	30 seconds
		68 °C	15 minutes
3	1	68 °C	10 minutes
4	1	4 °C	Hold

** May need to be adjusted depending on DNA input volume.*

The amplified libraries were mixed with 50 µL (0.5X volume of sample) AMPure XP beads and incubated for 15 minutes at RT. The bead pellet was washed with 80% ethanol twice on a magnet separator, then eluted with 22 µL of 10 mM Tris HCl (pH 8.0) with a 2-minute incubation. 20 µL of supernatant was transferred to a new tube after bead separation.

The pre-capture library was then assessed using the 5200 Fragment Analyzer system with the HS Large Fragment DNA kit for the size distribution and concentration.

5) Hybridization

500 ng of amplified libraries were used for hybridization. The SureSelect Cancer CGP assay probes (2.671 Mbp, Design ID: A3416642; 16 rxns: p/n 5280-0035, 96 rxns: p/n 5280-0036), designed for SRS, was used as the capture probe.

For SureSelect XT HS2 target enrichment, fast and overnight hybridization options are available. Both options were compared in this application note. For each method, the thermal cycling parameter of the first denaturation step (Table 4 and Table 5) was changed from the original parameter of 5 min at 95 °C to avoid DNA fragmentation. Other procedures were conducted according to the SureSelect XT HS2 protocols.[3, 4]

Table 4. Thermal cycling program for fast hybridization.

Step	Number of Cycles	Temperature	Time
1	1	95 °C	1 min. / 30 sec.
2	1	65 °C	10 minutes
3*	1	65 °C	1 minute
4	1	65 °C	1 minute
		37 °C	3 seconds
5	1	21 °C	Hold (up to 16h)

** Pause to add probe-hybridization mix.*

Table 5. Hybridization reaction for overnight hybridization.

Step	Number of Cycles	Temperature	Time
1	1	95 °C	1 min. / 30 sec.
2	1	65 °C	Hold (for 16 to 19h)

6) Hybridization library capture

Steptavidin bead preparation

Dynabeads M-270 Streptavidin beads (Thermo Fisher Scientific) were used to capture the hybridized libraries. 50 µL of beads were washed 3 times with 200 µL of Binding Buffer and finally resuspended with 200 µL of Binding Buffer. The prepared bead suspension was pre-heated at 68 °C for 10 minutes prior to the next capturing step.

Wash Buffer 2 was pre-heated at 70 °C before the wash step.

Capture

After hybridization, samples were mixed with the prepared beads at 68 °C by pipetting, then incubated for 5 minutes. The supernatant was removed after bead precipitation on a magnet separator.

Wash 1

DNA-bound beads were resuspended with 200 µL of Wash Buffer 1 by pipetting. The supernatant was removed after bead precipitation on a magnet separator.

Wash 2

The beads were resuspended with 500 µL of 70 °C preheated Wash Buffer 2 and incubated for 5 minutes at 70 °C. The beads were precipitated with a magnet separator and the supernatant was removed. This Wash Buffer 2 wash step was repeated 3 times.

During the third wash, 400 µL of supernatant was removed and the remaining was resuspended and transferred to a new PCR tube. Supernatant was removed after bead precipitation on a magnet separator.

Elution

DNA-captured beads were resuspended with 18 µL of freshly prepared 0.1M NaOH and incubated at RT for 10 minutes. Then, 6 µL of 600 mM Tris HCl (pH 8.0) was added and the beads were separated on a magnet separator. 24 µL of supernatant was transferred to a new PCR tube.

7) Captured library amplification

KOD FX Neo (TOYOBO) was used for amplification of the captured DNA. The composition and thermal cycling program are described in Tables 6 and 7, respectively.

Table 6. Composition of post-capture PCR.

Reagent	Volume/1 Rxn
2X PCR buffer for KOD FX Neo	50 µL
2 mM dNTPs	20 µL
SureSelect Post-Capture primer mix	4 µL
KOD FX Neo (1.0 U/µL)	2 µL
Captured libraries	24 µL

Table 7. Thermal cycling program for post-capture PCR.

Step	Number of Cycles	Temperature	Time
1	1	94 °C	2 minutes
2	16*	94 °C	15 seconds
		60 °C	30 seconds
		68 °C	15 minutes
3	1	68 °C	10 minutes
4	1	4 °C	Hold

** May need to be adjusted depending on design size of capture probe library.*

After amplification, the samples were mixed with 50 µL (0.5X volume of sample) AMPure XP beads and incubated for 15 minutes at RT. The bead pellet was then washed with 80% ethanol twice on a magnet separator and eluted with 42 µL of 10 mM Tris HCl (pH 8.0) with a 2-minute incubation. 40 µL of supernatant was transferred to a new tube.

The post-capture library was assessed using the 5200 Fragment Analyzer system with the HS Large Fragment DNA kit. Quantification was conducted using the Qubit dsDNA BR Assay kit (Thermo Fisher Scientific) before preparation for Nanopore sequencing.

8) Long-read sequencing and data analysis

Libraries for LRS were prepared with the Native Barcoding Kit 24 V14 (Oxford Nanopore Technologies, p/n: SQK-NBD114.24) for multiplex sequencing according to the manufacturer's protocol. Eight libraries were multiplexed and loaded onto the PromethION flow cell (FLO-PRO114M). The sequencing data was basecalled using the Dorado 7.2.13 with basecall model dna_r10.4.1_e8.2_400bps_sup@v4.2.0. Nanoplot 1.42.0 was used to draw the weighted histogram.

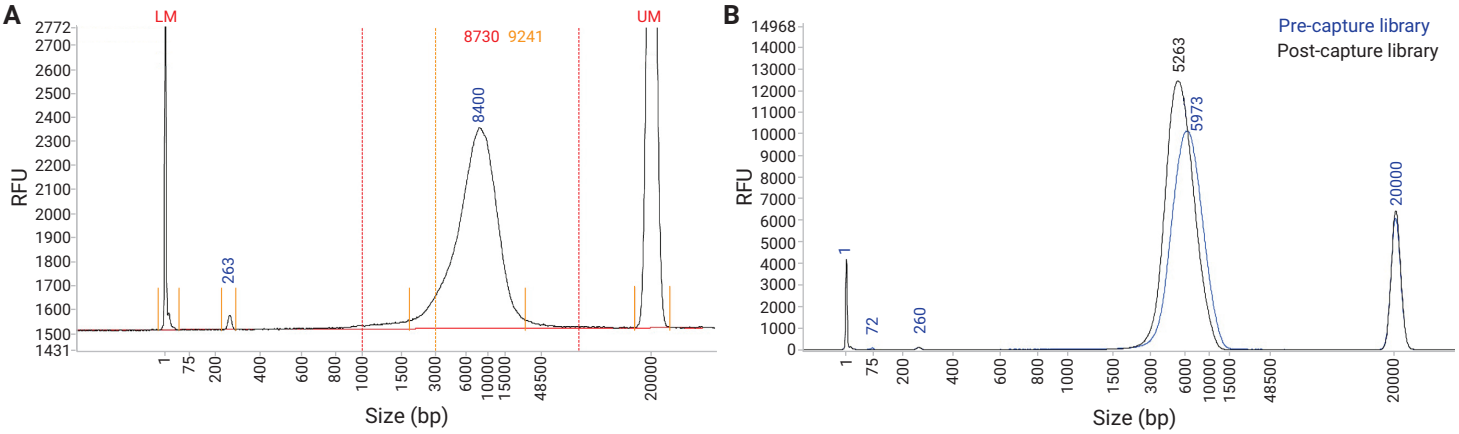
Sequencing data was downsampled to 2.9 million reads per sample and used for downstream analysis. Nanopore and Illumina adapter sequences were removed using Porechop 0.2.4 and cutadapt 2.6. Trimmed reads ware aligned to the hg38 reference genome using Minimap2-2.26. Percent on-target, mean-coverage, fold enrichment, percent coverage, and AT/GC dropout were analyzed using Picard 3.1.1 HsMetrics (Broad Institute).

Results

Size distribution and yield of fragmented DNA and pre-post-capture library

Size distribution of DNA sheared by the g-TUBE was assessed using the 5200 Fragment Analyzer system with the HS Large Fragment kit (Figure 2A). Due to less than 3000 bp fragments being removed at the size selection step, it is recommended to evaluate the size distribution of the sheared DNA at this step. Optimization of input volume/PCR cycles/size selection strategy would be required for DNA fragments less than 3000 bp.

Size distribution of the adapter-ligated and amplified libraries were 1 to 1.5 kb smaller than the sheared DNA (Figure 2B and Table 7). When skipping the size selection step, overall size distribution got smaller due to preferential amplification of small fragments (data not shown). Pre-capture library #1 was used for target enrichment. Size distribution of the target-enriched library was about 1 kb smaller than the pre-capture library (Table 8).



ID	Range	ng/µL	% Total	nmole/L	Avg. Size	%CV
C5: Fragmented DN	1000 bp to 100000 b	0.6228	98.5	0.1174	8730	72.53
	3000 bp to 100000 b	0.5777	91.4	0.1029	9241	68.06

Figure 2. Sizing distribution examples of (A) sheared DNA and (B) pre- and post-capture libraries.

Table 8. Summary of pre- and post-capture library quality control results.

Pre-Capture Library				
Sample Number			Average Size (bp)	Yield (ng)
1			6,499	4,815.2
2			6,68	6,502.4
3			7,712	5,856.4
4			6,241	7,130.6
Post-Capture Library				
Sample Number	Hybridization Option	Denaturation Duration	Average Size (bp)	Yield (ng)
1	Overnight	1 minute	5,455	2,425.5
2	Overnight	30 seconds	5,388	2,779.8
3	Fast	1 minute	5,483	1,939.2
4	Fast	30 seconds	5,623	2,637.0

Sequencing results

The captured and amplified libraries were ligated with the Nanopore adapter and sequenced on the PromethION. The sequencing metrics are summarized in Table 9. Sequence length N50 is about 4.9 to 5.4 kb, consistent with the electrophoresis result (Figure 3). When comparing the hybridization methods, fast hybridization tends to preserve larger fragments. Additionally, when comparing the denaturation condition, 30 seconds also favors the preservation of larger fragments. However, the sample size for each condition was one and no statistical information was available.

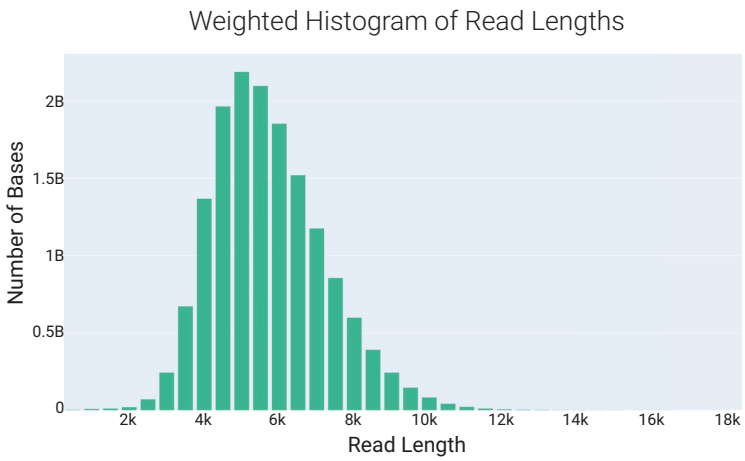


Figure 3. Weighted histogram of read length distribution. Histogram of 1D pass reads of the library with the fast hybridization option and 30-second denaturation condition, shown as a representation.

Table 9. Sequencing result summary.

		Overnight Hybridization		Fast Hybridization	
		1 minute	30 seconds	1 minute	30 seconds
Total Reads	Number of reads	3,176,795	3,850,233	3,340,509	3,351,175
	Total bp	15,027,621,729	18,619,540,009	16,650,843,463	17,133,017,839
	N50	4,947	5,064	5,229	5,356
	Max length (bp)	408,325	442,635	439,222	574,014
	Average length (bp)	4,730	4,836	4,985	5,113
	Median length (bp)	4,575	4,676	4,827	4,941
	Average q-score	15.2	15.1	14.8	15.0
≥ 10 Mean_Q-Score Reads	Number of reads	2,909,038	3,523,069	2,984,778	3,060,786
	Pass max length	408,325	401,776	270,356	331,457
	Pass average length	4,733	4,840	4,995	5,119
	Pass average q-score	15.9	15.8	15.7	15.7
Target Enrichment Performance	% on-target	73.5%	76.6%	84.5%	85.2%
	Mean coverage	496.22	524.01	680.86	700.89
	Fold enrichment	119.42	122.91	154.91	155.31
	Fold 80 base penalty	3.42	3.34	3.96	4.51
	10X % taget bases	98.8%	98.7%	99.1%	99.0%
	20X % target bases	97.7%	97.8%	98.4%	97.9%
	50X % target bases	93.8%	94.4%	94.7%	93.2%
	100X % target bases	84.7%	86.2%	87.1%	85.1%
	AT dropout	0.56	0.51	5.45	5.16
	GC dropout	17.69	17.30	1.89	2.04

Target capture efficiency

Each sequence dataset was downsampled to 2.9 million reads and evaluated using sequencing metrics with Picard HsMetrics. On-target percentage exceeded 70% for all experimental conditions, with fast hybridization showing a higher percentage (Figure 4A). For mean coverage, fold enrichment, and percent coverage with 10X, 20X, 50X, and 100X, fast hybridization showed higher values.

Regarding AT/GC dropout, overnight hybridization demonstrated better values for AT dropout, while fast hybridization demonstrated even better values for GC dropout. As a result, it can be concluded that fast hybridization achieves a balanced performance for both AT and GC dropouts. This may be due to the probe design of the SureSelect Cancer CGP assay being optimized for fast hybridization, and there may be room for performance improvements.

As to the denaturing step of hybridization (30 seconds versus 1 minute), there was no obvious difference in the hybridization performance. There was initial concern about the shorter denature time (30 seconds) potentially affecting GC-rich region performance. However, there was no significant difference observed in GC dropout, therefore the 30-second denaturation time appears sufficient. Average read lengths were slightly longer with a 30-second denaturation, also concluding that it is ideal for this application.

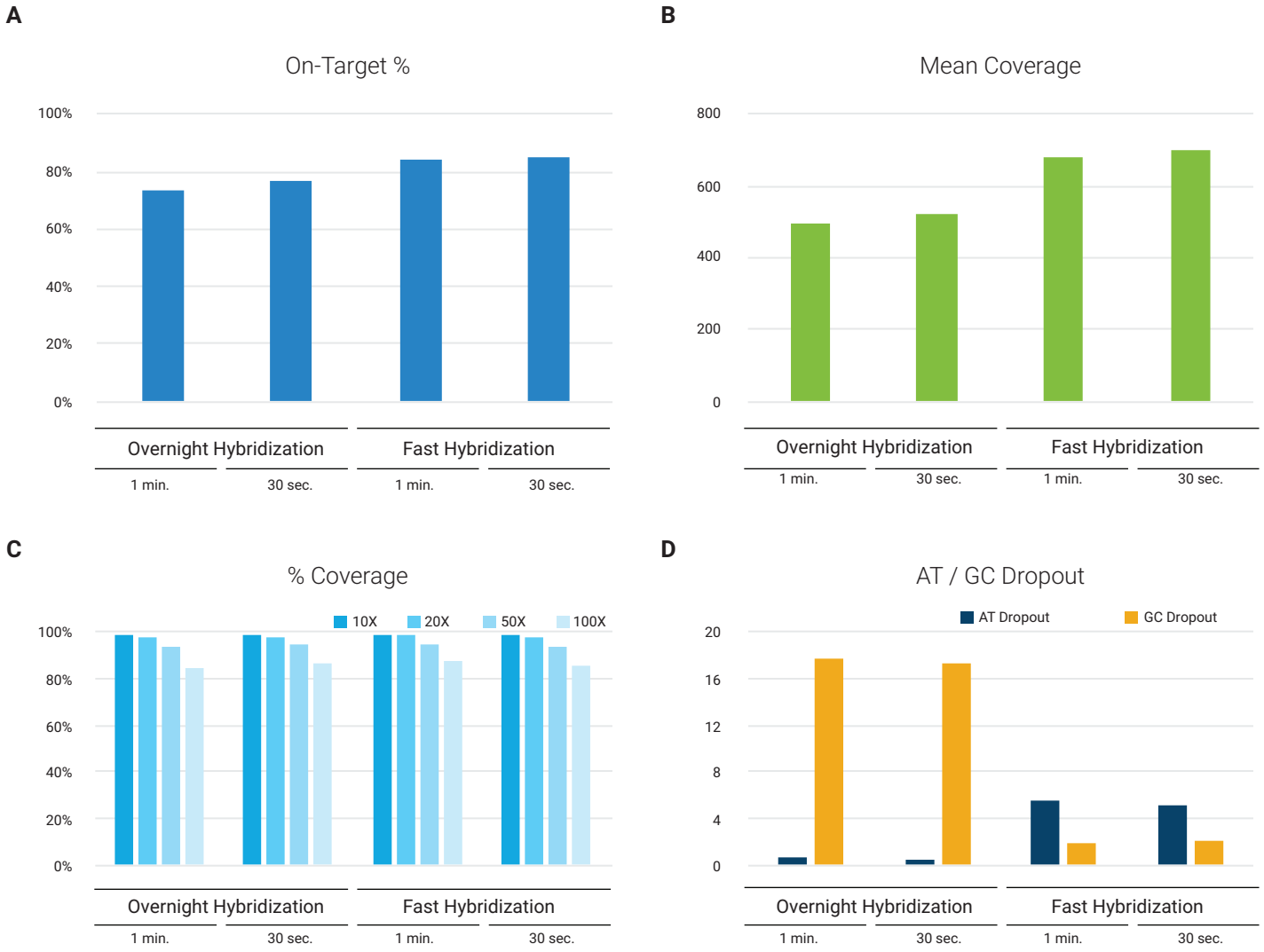


Figure 4. Summary of target enrichment performance metrics.(A) On-target percentage, (B) mean coverage, (C) percent coverage (D) AT/GC dropout.

Target region examples

Examples of target coverages are shown in Figure 5. As expected with the sequencing metrics for on-target percentage being higher than 70%, each region shows piled-up reads with, and around, the targeted regions. These findings indicate that target enrichment was conducted effectively.

There are regions in the human genome that are difficult to sequence with SRS, such as GC-rich regions, genes with pseudo genes, and repetitive sequences. Among them, the TERT promoter region has known oncogenic variants, such as positions C228T and C250T for certain cancer types, known to be GC-rich and difficult to sequence with SRS. In the data featuring target-enriched long reads, those regions exhibit more uniform coverage with minimal or no dropouts (Figures 5 and 6). Another example is RB1 exon 15, another notoriously difficult-to-read region surrounded by poly-A and poly-T. The target-enriched long reads displayed consistent coverage (Figures 5 and 6).

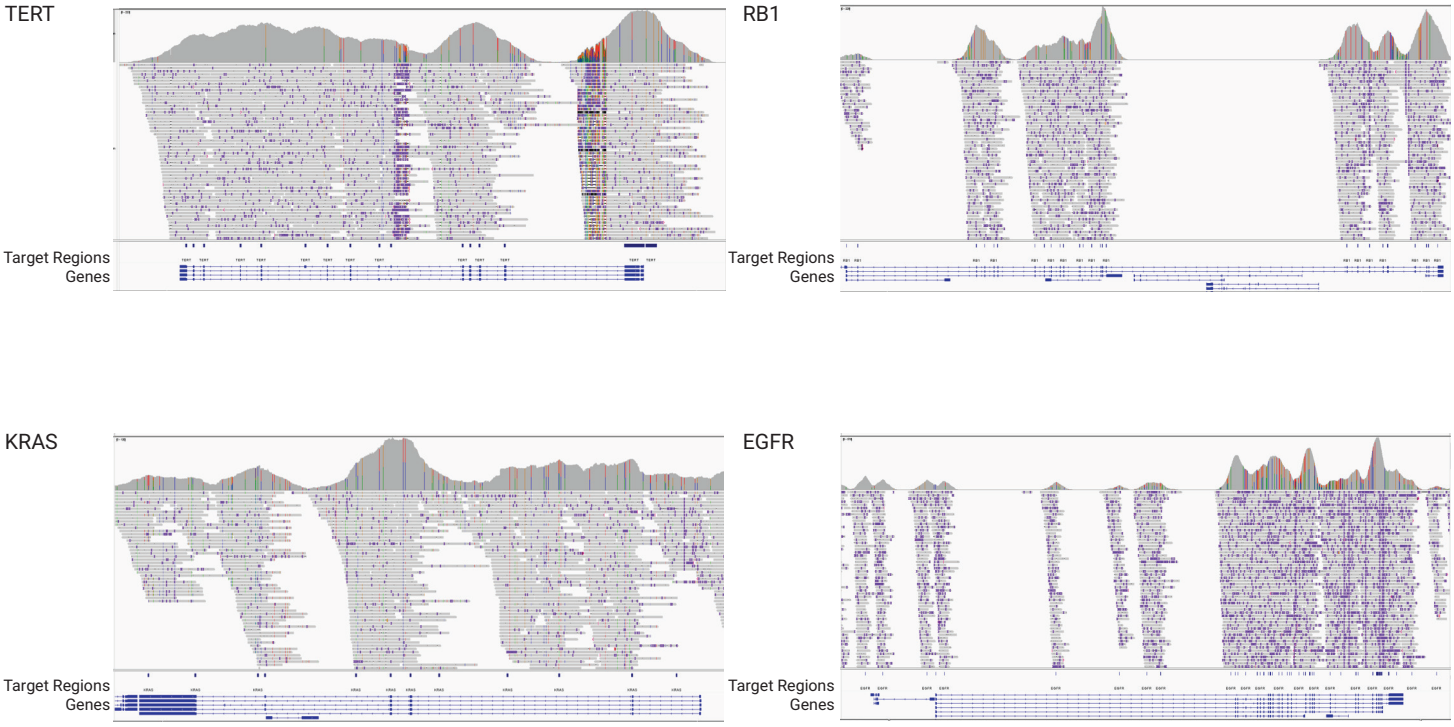


Figure 5. Examples of target regions.

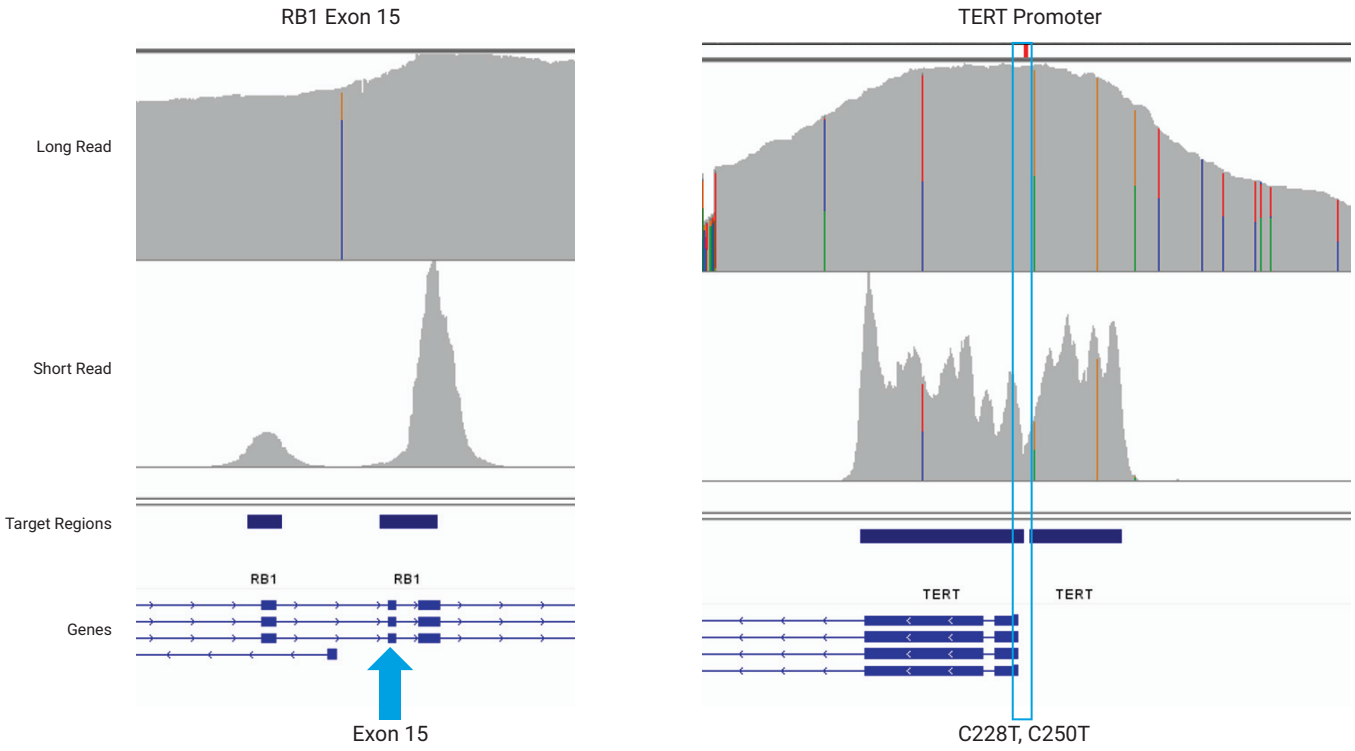


Figure 5. Examples of sequence coverage for challenging regions. Short-read libraries were prepared from Agilent OneSeq Female Reference DNA and enriched using the Agilent SureSelect Cancer CGP assay probe. Libraries were sequenced using the Illumina HiSeq4000 with 150 bp pair end.

Conclusion

In this application note, the results demonstrated that modifying the Agilent SureSelect XT HS2 protocol together with KOD FX Neo can achieve target enrichment for longer fragments (5 to 6 kb) with high on-target percentage and good uniformity. Moreover, the capture probe design for short reads can also be used for target enrichment of long fragments.

Future considerations involve optimizing the probe design for long-read target enrichment to improve both performance and cost-effectiveness of the overall design strategy. This method is expected to be applicable for target enrichment of dark genes, haplotyping (including HLA regions), pharmacogenomics, repeat expansion detection, full-length cDNA, and pathogen genomes. Additionally, compatibility with PacBio sequencing is also anticipated through modifications to the adapter ligation protocols. In conclusion, this technique can enable cost-effective analysis of long-read sequencing, whether through multiplexing or small flow cells.

References

1. Method of the Year 2022: Long-Read Sequencing. *Nature Methods* **2023**, 20(1), 1–1. https://doi.org/10.1038/s41592-022-01759-x.
2. Use of Agilent SureSelect to perform targeted long-read nanopore sequencing. *Agilent Technologies Application Note*, **2017**, publication number 5991-8056EN.
3. SureSelect XT HS2 DNA Kits with Library Preparation (+/-MBC) / Fast-Hyb Target Enrichment /Post-capture Pooling Workflow, *Agilent Technologies User Manual*, **2023**, publication number G9983-90000.
4. SureSelect XT HS2 DNA Kits with Library Preparation (+/-MBC) / Overnight-Hyb Target Enrichment /Post-capture Pooling Workflow, *Agilent Technologies User Manual*, **2023**, publication number G9956-90000.
5. Procedure & Checklist – Using AMPure® PB Beads for Size-Selection. *Pacific Biosciences*, **2020**, Part Number 101-854-900 Version 02 https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Using-AMPure-PB-Beads-for-Size-Selection.pdf.

www.agilent.com

PR7001-2884

Published in the USA, August 1, 2024
5994-7612EN

Agilent
Trusted Answers