Agilent SureSelect^{XT} Methyl-Seq Applications with Low-Input DNA and Smaller Capture Libraries

Application Note

Authors

Alicia Alonso, Mame Penda Fall, Thadeous James Kacmarczyk

Department of Medicine
Division of Hematology/Oncology
Epigenomics Core Facility
Weill Cornell Medicine
New York, NY

Josh Zhiyong Wang
Carlos Pabon
Mary Napier

Agilent Technologies
Santa Clara, CA

Abstract

The Agilent SureSelect^{XT} Human Methyl-Seq kit can be used to generate methyl-seq libraries with DNA input amounts as low as 250 ng. The sequencing data has comparable read depth coverage to libraries obtained with 1 µg, although at higher duplicate read rates. Lower input amounts can be used by slightly increasing the number of PCR cycles used to amplify the bisulfite-treated libraries (10-11 versus 8 in the standard protocol).

An alternative method for use with low-input DNA amounts that combines Agilent SureSelect^{XT} capture with the Illumina TruSeq Methylation protocol was also evaluated, but initial results generated less useful sequencing data than the SureSelect^{XT} Human Methyl-Seq workflow alone. Guidelines are also provided for working with custom SureSelect^{XT} Methyl-Seq capture libraries with smaller target regions (< 3 Mb).

Introduction

DNA methylation is widely known as a versatile mechanism for gene regulation across organisms, while abnormal DNA methylation has been observed in various human diseases such as cancer and mental disorders, underscoring the utility of methylation sites as diagnostic markers and/or drug targets. Detection of DNA methylation at single-nucleotide resolution is typically achieved by sequencing, including next-generation sequencing (NGS).

Agilent Technologies

While whole-genome bisulfite sequencing (WGBS) is a viable option, a targeted approach such as Agilent SureSelect^{XT} Methyl-Seq is more cost-effective. Agilent SureSelect^{XT} Methyl-Seq enables target enrichment designed specifically for methylomic regions with subsequent bisulfite treatment to detect meaningful methylation events at single base pair resolution. This targeted approach also facilitates data storage and analysis.

This Application Note describes the evaluation of the Agilent SureSelect^{XT} Human Methyl-Seq workflow using a DNA input amount of 250 to 500 ng with protocol modifications, which is much less than the 1 µg of input DNA called for by the standard protocol. It also provides guidance on working with custom methyl-seq capture libraries targeted against smaller regions (< 3 Mb). A method using a combination of the Agilent SureSelect^{XT} Human Methyl-Seq capture library and the TruSeq DNA Methylation kit from Illumina was evaluated for its suitability to detect DNA methylation status at the single nucleotide level with a low-input DNA amount (250 – 500 ng). This combined method did not improve the ability to detect DNA methylation status at the single nucleotide level with these low-input DNA amounts.

Materials and Methods

Samples and Reagents

The samples used for the methyl-seq protocols were either human blood (C77 and C196) or a human lung fibroblast cell line sample (IMR90) used as a reference standard. The Agilent SureSelect^{XT} Methyl-Seq Reagent kit and SureSelect^{XT} Human Methyl-Seq kit were used, respectively, for NGS library preparation and hybrid capture of putative methylated regions of the genome. The Illumina TruSeq DNA Methylation Kit was part of a hybrid protocol for NGS library preparation that also utilized the SureSelect^{XT} Human Methyl-Seq kit for capture. Beckman Coulter Agencourt AMPure XP was used to clean up sheared DNA. The EZ-DNA Methylation-Gold kit from Zymo Research provided bisulfite conversion of captured libraries.

Instruments

DNA shearing was accomplished using a Covaris S220 Focused-ultrasonicator, and the Agilent 2100 Bioanalyzer was used to assess DNA fragment sizes after shearing and library ligation. Final post-capture library yields were measured using a Qubit 3.0 Fluorometer (ThermoFisher Scientific). The final libraries were pooled, clustered at 10 pM on a pair-end read flow cell and sequenced on an Illumina HiSeq 2500 System for 2x 100 cycles at a depth of ~100 M reads per sample. Illumina CASAVA 1.8.2 software was used to perform base calling and demultiplexing.

Data Analysis

Flexbar 2.4 software (https://github.com/seqan/flexbar) was used to remove adaptor sequences from sequencing reads. Mapping of sequence reads to the reference genome was performed using Bismark v0.14.4 (http://www.bioinformatics.babraham.ac.uk/projects/bismark/Bismark_User_Guide_v0.14.4.pdf) and Bowtie2 v2.2.5 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) per the following command lines (sample IMR90_1 used here):

```
flexbar -r IMR90_1_R1_PF.fastq -p
IMR90_1_R2_PF.fastq -t IMR90_1 -a
Truseq_adapters.fa -f i1.8 -n 10 -ao 6
-m 21 -at 2 -ae RIGHT -u 2 -j -am 3 -ai
-3 -ag -20
```

```
/Bismark-0.14.4/bismark -q -un
--ambiguous -L 32 -X 500 /genomes/
indices/Homo_sapiens/Bismark_
bt2_hg19 -1 IMR90_1_R1.fastq -2
IMR90_1_R2.fastq
```

Data analysis of sequencing reads percent on-target and read depth was performed using a pipeline integrated with the bedtools utilities (http://bedtools.readthedocs.io/en/latest/).

Results and Discussion

Evaluation of SureSelect^{XT} Methyl-Seq with less than 1-3 µg of input DNA

The catalog SureSelect^{XT} Human Methyl-Seq kit from Agilent Technologies provides targeted coverage of 84 Mb of the human genome. The protocol recommends an input DNA amount of 3 µg or 1 µg with a workflow that consists of constructing a genomic DNA library, followed by capture hybridization, post-capture bisulfite treatment and then PCR. However, the requirement of 1-3 µg genomic DNA is high compared to array-based methylation assays, which typically require 250-500 ng of DNA.

In order to determine if the SureSelect^{XT} Methyl-Seq Reagent kit could be used successfully with less than 1 µg of input DNA, two human blood DNA samples (C77 and C196) were processed at 500 ng

and 250 ng input for a total of 4 data points using the catalog kit. The post-capture PCR cycles were increased to 10 and 12, respectively, for the 500 ng and 250 ng input, as a first attempt.

The final post-capture library yields were determined using Qubit fluorometric quantitation and are shown in Table 1. With 500 ng input DNA and 10 cycles of PCR for the bisulfite-treated libraries amplification step, we obtained 7.18 ng/µL for the C77 sample and 9.68 ng/µL for the C196 sample. These values were equivalent to 35.9 nM and 48.4 nM respectively. As the goal for a final library concentration was >10-20 nM, the results indicated it was feasible to start with 500 ng input DNA and 9-10 cycles of PCR. Similarly, for 250 ng input DNA, 10-11 cycles instead of 12 cycles of PCR may be appropriate for the bisulfite-treated libraries amplification step.

All these libraries were sequenced with pair-end 2x100 bp using an Illumina HiSeq 2500 instrument with 15-25 Gb raw sequencing output, along with a control captured library made with 3 µg of input IMR90 DNA. Human hg19 genome build was used as input for Bismark software, a tool to map bisulfite converted sequence reads and determine cytosine methylation states, for reference generation[1]. Passing-filter reads had their adaptor sequences removed using Flexbar version 2.4 software and subsequently mapped to the reference genome with Bismark v0.14.4 and Bowtie2 v2.2.5, using the command lines provided in the Data Analysis section.

Approximately 75% mapping efficiency was achieved, and Table 2 shows the SureSelect^{XT} Methyl-Seq target enrichment performance after normalizing bam files to around 100 M reads. The percentage of reads in targeted regions remained at a very high level (81.9% to 87.6%) at input levels that range from 250 ng to 3 µg. The same was true for the percentage reads in targeted regions +/- 200 bps, which ranged from 94.2 – 96.3%. These results demonstrate that a lower DNA input of 250 to 500 ng does not impact the performance against this QC metric.

On the other hand, as expected, the percentage of duplicate reads increased dramatically from 8.3% for 3 µg input DNA to 40-50% for 500 ng input DNA and 50-65% for 250 ng input DNA (Table 2). Consequently, the average read depth dropped from 84x for 3 µg input DNA to 47x-55x for 500 ng input DNA and 32x-45x for 250 ng DNA. Taken together, these results show that lower-input DNA at 250-500 ng did not affect the percentage of on-target reads, but did result in higher duplicates.

Interestingly, when looking at percentage of bases with at least 10x read coverage, even with 250 ng input DNA, 81-87% of the targeted 84 M bases have achieved this metric, suggesting a high level of utility of the result (Table 2). Indeed, the percentage of bases with at least 10x read coverage can increase further with additional sequencing output. For example, sample C77 provided 81.4% of the bases at 10x reads coverage with 100 M reads of sequencing output (Table 2). However, using the same libraries made with 250 ng starting DNA, 167 M reads sequencing output yielded 91.6% of the bases at 10x reads coverage (data not shown). Similarly, sample C196 yielded 87.7% of bases at 10x reads coverage with 100 M reads sequencing output (Table 2), versus 95.0% of bases at 10x reads coverage with 195 M reads sequencing output (data not shown).

Overall, the results obtained with 250 ng and 500 ng of input material demonstrate that when low-input DNA amounts are used, appropriate protocol modifications must be made to achieve the best possible results.

Samples	DNA Input (ng)	Concentration (ng/µL)	Concentration (nM)
C77	500	7.18	35.9
C196	500	9.68	48.4
C77	250	10.6	53
C196	250	24.8	124

Table 1. Library yield comparison using SureSelect^{XT} Human Methyl-Seq with 250 ng and 500 ng of DNA input.

Workflow and sequencing data comparison of SureSelect^{XT} Methyl-Seq library prep + SureSelect^{XT} capture versus TruSeq DNA Methylation kit + SureSelect^{XT} capture

In an effort to explore the application of methyl-seq using low-input DNA, an alternative method called PBAT (post-bisulfite adaptor tagging) was described for use with input DNA as low as 10 ng and 100 ng[2]. This method shears genomic DNA for capture hybridization first, followed by bisulfite treatment and adaptor tagging/random priming steps, etc. to construct the library after bisulfite treatment. However, this method requires ordering a large number of components from multiple vendors for buffer and reagent preparations and lacks reagent QC or related support.

There are striking similarities between the adaptor tagging steps in PBAT and the workflow steps in a commercial kit, EpiGnome/TruSeq DNA Methylation kit (Illumina) which is typically used for whole genome bisulfite sequencing. A description of the TruSeq DNA methylation kit is available by contacting Illumina. Combining the TruSeq kit with the Human SureSelect^{XT} capture library may provide a more convenient alternative than PBAT when working with the SureSelect^{XT} Methyl-Seq kit, for investigating methylation information in targeted regions. Table 3 shows the workflow comparison between these two strategies, with both workflows undergoing the same capture hybridization/wash/bisulfite treatment steps and taking about two and a half days, but differing regarding DNA shearing size, DNA treatment pre-capture and DNA treatment post-bisulfite treatment.

The SureSelect^{XT} capture/TruSeq methylation combination protocol was used to process IMR90 DNA at 500 ng and 250 ng input in duplicate. DNA was sheared to ~600 bp peak with the Covaris instrument, purified using Agencourt AMPure and hybridized with the SureSelect^{XT} Human Methyl-Seq capture library. The captured product was finally eluted in 20 µL elution buffer, incubated at room temperature for 20 minutes and subjected to bisulfite treatment at 64 °C for 2.5 hours using the EZ-DNA Methylation-Gold kit (Zymo Research), and finally eluted in 9 µL elution buffer. The eluted DNA then served as starting material for the TruSeq DNA Methylation kit workflow, and qPCR

	IMR90	C77	C196	C77	C196
DNA quantity (ng)	**3000**	**500**	**500**	**250**	**250**
Total HQ uniquely mapped reads:	96.5M	93.4M	94.4M	92.4M	93.4M
Percent duplicate reads:	8.3%	50%	40.5%	65.5%	50.1%
Total HQ reads after removing duplicates:	88.5M	46.7M	56.1M	31.8M	46.6M
Number of reads in targeted regions:	72.5M	40.7M	47.6M	27.9M	38.9M
Percentage reads in targeted regions:	**81.92%**	**87.24%**	**84.83%**	**87.65%**	**83.54%**
Percentage reads in regions +/- 100bp:	93.04%	95.55%	95.29%	95.65%	95.04%
Percentage reads in regions +/- 200bp:	94.23%	96.13%	96.43%	96.19%	96.34%
Average Read Depth:	**84x**	**47x**	**55x**	**32x**	**45x**
Percentage of targeted bases covered by...					
...at least 1 read:	98.89%	98.16%	98.56%	97.67%	98.36%
...at least 5 read:	97.14%	94.04%	95.30%	91.42%	94.33%
...at least 10 read:	**94.48%**	**87.65%**	**89.97%**	**81.44%**	**87.70%**
...at least 20 read:	87.92%	72.98%	77.14%	60.14%	72.13%
...at least 30 read:	80.33%	58.58%	63.99%	42.18%	56.99%

Table 2. SureSelect^{XT} Human Methyl-Seq target enrichment NGS sequencing performance using different DNA input amounts.

was performed to determine the appropriate cycle number for PCR after completion of double-strand DNA synthesis.

In the DNA samples that were evaluated, 12-17 cycles of PCR were determined to be appropriate, and final library yields are shown in Table 4. The 500 ng duplicates both provided good library yield (~4ng/µL), but one 250 ng sample had 10-fold lower yield (0.3 ng/µL) compared to the other one. Since it was difficult to determine the reason for the inconsistent result with 250 ng input, the decision was made to sequence only the duplicate libraries made with 500 ng input.

On a separate note, when overlaying the Bioanalyzer traces for the final libraries using TruSeq methylation + SureSelect^{XT} capture workflow (Table 3) and the final libraries using standard SureSelect^{XT} Methyl-Seq workflow (Table 1), the peaks are similar at 250-300 bp (data not shown).

These libraries were sequenced with pair-end 2x100 bp chemistry on the HiSeq 2500 instrument to 15-25 Gb raw sequencing output. Table 5 shows the target enrichment performance comparison between SureSelect^{XT} capture/TruSeq methylation (Method 2) and SureSelect^{XT} Methyl-Seq alone (Method 1), after normalizing bam files to around 100 M reads.

With the same 500 ng of input DNA, the percentage of reads in targeted regions using the SureSelect^{XT}/TruSeq combination method was at a respectable 66.6-67.2%, but lower than 84.8-87.2% achieved with SureSelect^{XT} Methyl-Seq method alone (Table 5). The more striking observation was that the percentage of duplicate reads for the SureSelect^{XT}/TruSeq combination method was at a very high 80.6-86.7%, compared to 40-50% achieved with the SureSelect^{XT} Methyl-Seq method alone.

Consequently, the average read depth stayed low at 8x - 12x for the SureSelect^{XT}/TruSeq combination method vs 47x - 55x for the SureSelect^{XT} Methyl-Seq method alone.

Altogether, with 500 ng input DNA, the SureSelect^{XT}/TruSeq combination method did not generate sequencing data as useful as that generated using the SureSelect^{XT} Methyl-Seq method alone, even though the workflows seem comparable. This result suggests that the library complexity with the SureSelect^{XT}/TruSeq combination method may be more compromised than that using the SureSelect^{XT} Methyl-Seq method alone. It is also possible that additional protocol optimization is needed when considering the SureSelect^{XT}/TruSeq combination method for targeted methyl-seq applications.

Time	Steps	SureSelect^{XT} Methyl-Seq	TruSeq Methylation +SureSelect^{XT} Capture
Day 1	**Pre-capture**	DNA sheared to ~150 bp peak End-repair, dA tailing, adaptor ligation to obtain DNA library SureSelect^{XT} hybridization overnight	DNA sheared to ~600 bp peak
	Capture hybridization	SureSelect^{XT} hybridization overnight	SureSelect^{XT} hybridization overnight
Day 2	**Post capture**	SureSelect^{XT} capture/wash, bisulfite treatment	SureSelect^{XT} capture/wash, bisulfite treatment Primer annealing, DNA synthesis and DNA tagging, index
	Post bisulfite	PCR amplify bisulfite-treated library (8-14 cycles)	PCR of bisulfite-treated library (12-17 cycles) overnight
Day 3	**Post bisulfite**	Index modified libraries by PCR (6 cycles), final QC	Final QC
		Kit component used for this step	
Day 1	**NGS library prep**	SureSelect^{XT}	N/A
	Capture library	SureSelect^{XT}	SureSelect^{XT}
Day 2/3	**Binding/wash buffers**	SureSelect^{XT}	SureSelect^{XT}
	EZ-DNA Methylation Gold kit	SureSelect^{XT}	SureSelect^{XT}
Day 2/3	**NGS library prep**	N/A	TruSeq

N/A - not applicable

Table 3. Workflow comparison between the standard SureSelect^{XT} Human Methyl-Seq workflow and the combination of SureSelect^{XT} Capture and TruSeq Methylation.

Guidance for working with a custom methyl-seq capture library with smaller target regions (< 3 Mb)

SureSelect^{XT} target enrichment is an extremely popular platform that enables enrichment of specific genomic regions from <100 kb to over 100 Mb, but the current SureSelect^{XT} Methyl-Seq protocol only supports the catalog human Methyl-Seq design, which is 84 Mb. For researchers interested in analyzing methylation information in genomic regions smaller than 3 Mb, the following modifications should be considered.

1. Regarding capture library volume used during hybridization: SureSelect^{XT} Methyl-Seq protocol version C.0 suggests using 5 µL per hybridization. However, for a custom library, this volume will need to be adjusted based on capture library size. For a library smaller than 3 Mb, 2 µL per hybridization should be used, as indicated in Table 6 for capture hybridization mix preparation.

Samples	DNA Input (ng)	Concentration (ng/µL)	Concentration (nM)
IMR90	500	4.9	24.5
IMR90	500	3.4	17.1
IMR90	250	3.0	15.1
IMR90	250	0.3	1.5

Table 4. Final library yields using the combination of SureSelect^{XT} Capture and TruSeq Methylation.

	SureSelectXT Methyl-Seq alone		SureSelectXT/TruSeq combination	
Sample	C77	C196	IMR90-1	IMR90-2
Total HQ uniquely mapped reads:	93.4 M	94.4 M	86.1 M	85.9 M
Percent duplicate reads:	50%	40.5%	80.6%	86.7%
Total HQ reads after removing duplicates:	46.7 M	56.1 M	16.7 M	11.4 M
Number of reads in targeted regions:	40.7 M	47.6 M	11.1 M	7.6 M
Percentage reads in targeted regions:	**87.24%**	**84.83%**	**66.64%**	**67.27%**
Percentage reads in regions +/- 100bp:	95.55%	95.29%	83.65%	84.36%
Percentage reads in regions +/- 200bp:	96.13%	96.43%	90.10%	90.74%
Average Read Depth:	**47x**	**55x**	**12x**	**8x**
Percentage of targeted bases covered by...				
...at least 1 read:	98.16%	98.56%	90.08%	85.94%
...at least 5 read:	94.04%	95.30%	68.29%	57.92%
...at least 10 read:	**87.65%**	**89.97%**	**47.45%**	**34.12%**
...at least 20 read:	72.98%	77.14%	21.54%	10.82%
...at least 30 read:	58.58%	63.99%	9.32%	3.28%

Table 5. Target enrichment performance comparison between the SureSelect^{XT}/TruSeq combination method vs the SureSelect^{XT} Methyl-Seq method alone, using 500 ng of input DNA.

Reagent	Volume for 1 reaction (µL)
Hybridization buffer mix	13
10% RNase Block*	5
Capture library (< 3 Mb)	2
Total	20

*Prepared at 1:9 dilution from stock RNase Block.

Table 6. Preparation of custom capture library (<3 Mb) hybridization mix.

Input DNA (ng)	Capture Size	SureSelect^{XT} Methyl-Seq	TruSeq Methylation + SureSelect^{XT} capture
1000-3000	1 kb up to 0.5 Mb	13-14	
1000-3000	0.5 – 1.49 Mb	11-12	
1000-3000	> 1.5 Mb	10	
1000-3000	84 Mb catalog	8*	
500	84 Mb catalog	9-10	12-17, qPCR-dependent
250	84 Mb catalog	10-11	12-17, qPCR-dependent

*The cycle number specified in the SureSelect^{XT} Methyl-Seq protocol

Table 7. Suggested PCR cycle number to amplify bisulfite-treated libraries.

2. Post-capture PCR cycle number: The SureSelect^{XT} Methyl-Seq protocol suggests 8 cycles for the first PCR to amplify bisulfite-treated library, then 6 cycles for the index PCR, but for a custom library the cycle number in the first PCR should be adjusted. Table 7 provides suggested cycle numbers for either the SureSelect^{XT} Methyl-Seq protocol or the TruSeq methylation kit plus SureSelect^{XT} capture workflow. The cycle number is only changed for the first PCR that is used to amplify the bisulfite-treated library, while the 6 cycles for the second index PCR remain unchanged. Starting PCR cycle numbers for DNA inputs ranging from 250 ng to 3 µg are also provided.

Conclusions

The Agilent SureSelect^{XT} Methyl-Seq target enrichment platform offers a highly reliable and efficient method with single-base resolution for the study of methylated genomic regions. Current offerings include a catalog Human Methyl-Seq capture design as well as Mouse and Rat Methyl-Seq designs.

The flexibility to target any regions of interest (from < 100 kb to > 100 Mb) can be easily achieved with custom Methyl-Seq designs with some modifications in the published workflow. This study shows that the SureSelect^{XT} Methyl-Seq workflow is capable of generating results with DNA input at both 500 ng and 250 ng that are comparable in read depth coverage to the recommended DNA input amount of 1 to 3 µg.

Details of protocol modification when working with a custom Methyl-Seq library are also provided (Table 7). The utility of the TruSeq Methylation kit together with the SureSelect^{XT} capture library for low- input DNA was also evaluated, but it failed to achieve results comparable to those generated using the SureSelect^{XT} Methyl-Seq method alone, which may have been due to reduced library complexity or insufficient protocol optimization.

References

1. "Agilent SureSelect^{XT} Human Methyl-Seq for the Quantitative Analysis of DNA Methylation with Single-Base Resolution", Agilent publication number 5991-0166EN.

2. Miura F and Ito T. "Highly sensitive targeted methylome sequencing by post-bisulfite adaptor tagging." *DNA Res* **22**, 13-18 (2014).

Trusted Answers. Together.

PR7000-0620

Published in the USA, January 27, 2017
5991-7838EN

LEARN MORE AT:
www.agilent.com/genomics

Find an Agilent customer center in your country:
www.agilent.com/genomics/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Europe
info_agilent@agilent.com

Asia Pacific
info_agilent@agilent.com

For Research Use Only. Not for use in diagnostic procedures.

Agilent Technologies