Application Note

Genomics

Agilent

Trusted Answers

Use of a Modified Agilent SureSelect XT HS2 Target Enrichment Procedure for Analysis of Methylation Status in a Model Organism

A review of method and analysis

Author

Dr. Matthieu Egloff MD,
Université de Poitiers,
INSERM 1084, LNEC, F-86000
Poitiers, France

Dr. Quentin Riche-Piotaix, PhD,
CHU de Poitiers,
Service de Génétique, F-86000
Poitiers, France

François Lozach
Agilent Technologies, Inc.

Abstract

Detection and analysis of methylation by next-generation sequencing (NGS) can be challenging, with wet lab and analysis procedures often being long and complex. The aim of this application note is to provide a detailed review of the method, helping researchers using the Agilent SureSelect XT HS2 hybrid capture kit for this specific NGS application. Additionally, to provide some guidelines for methylation analysis. The full procedure for capturing the regions of interest using DNA extracted from the blood and brain of mice is described within this document. In this use-case practical example, bisulfite converted libraries revealed the power of targeted capture and the precision of the data obtained.

Introduction

Methylation in vertebrate DNA occurs on cytosine (C) nucleotides by attachment of a methyl group (CH_3) and is largely found in the GC sequence context in vertebrates. These methylation sites are referred as CpG (p for Phosphate linkage) to distinguish from the CG base pair.

Methylation patterns can be propagated through replication, maintaining the methylation state in daughter cells. Methylation status acts in conjunction with other epigenetic modifiers of gene expression, such as histone modifications, to regulate transcription. For example, CpG methylation at promoter sites can repress transcription by preventing protein binding for transcription initiation or can be a binding site for specific modifying enzymes to produce a complete gene expression control mechanism.[1]

Therefore, CpG are generally located in and around promoters and their methylation impacts gene expression. However, in certain cases, studying the differentially methylated region (DMR) is also important outside the CpG group.[2]

In this application note, a modified Agilent SureSelect XT HS2 target enrichment protocol was used to study the methylation status in mice brain and blood cells. This animal model offers an identical genetic background between individuals, and a controlled environment, making it possible to identify marks specifically linked to the agent under study, exposed to valproic acid (VPA). VPA is an anti-epileptic drug commonly used to generate an environmental animal model of ASD (autism spectrum disorders) by in utero treatment. In this study, we assessed the methylation differences between wild-type mice and mice exposed in utero to VPA. This represents a good example of the utility of the Agilent SureSelect Methyl-Seq target enrichment system and the quality of results that can be obtained.

This application note provides a short description of the SureSelect custom panel used in this study followed by a detailed description of the protocol, with the final section dedicated to analysis procedures.

Custom Panel

The targeted regions of the SureSelect 3 Mb panel correspond to 1407 genomic regions located in known or predicted promoters of 1029 genes, which are known to be involved in ASD. To establish this list of targets, two sources were cross-referenced: the list of autism-related genes from Simons Foundation Autism Research Initiative (SFARI) (https://gene.sfari.org/) and the Eukaryotic Promoter Database (https://epd.expasy.org/epd/). The design strategy placed probes on or around CpG sites utilizing in silico probe performance criteria. Beyond that the 3 Mb panel focused specifically on targeted regions of interest in ASD, the main advantage of its use over the SureSelect full methylome (80 Mb) panel is the ability to perform deep sequencing of each target.

Step-By-Step Protocol

In this section, the Quick Start protocol, SureSelect XT HS2 DNA with Post-capture Pooling, has been used as reference. Please refer to the published protocol where applicable. Some of the protocol parts/sections are modified from the published procedure and all differences are described in detail below. Modified protocol parts for the quick start protocol are highlighted in red below in Figure 8. When not specifically stated, the steps are identical to the original protocol. Running this protocol requires the EZ DNA-Methylation Lightening Kit (Zymo Research).

Sample collection: 500 µL of blood was collected by intracardiac sampling at the time of sacrifice for immediate DNA extraction. After intracardiac perfusion with 50 mL of 0.9% NaCl, brains were harvested and stored at –80 °C. DNA extractions from blood and brain were performed with the DNEasy Blood & Tissue Kit (Qiagen).

Custom panel: See the description above for details.

DNA fragmentation: Performed as described in the Quick Start protocol, SureSelect XT HS2 DNA with Post-capture Pooling, Option 2: Covaris shearing; page 2.

Library preparation: Performed as described the Quick Start protocol, SureSelect XT HS2 DNA with Post-capture Pooling. *Important:* You must follow steps 1, 2, 3, and 4 (page 3) but replace the SureSelect XT HS2 Adaptor Oligo mix (white cap) with 5 µL of the SureSelect Methyl-Seq Methylated Adapter vial (green cap) from the SureSelect Methyl-Seq XT kit. Note: Depending on the initial amount of DNA, optimizing the amount of adaptor dilution may be required.

Assess library quality with the Agilent TapeStation High Sensitivity DNA kit (HSD1000). The expected peak distribution is in the range of 200 to 300 bp. Measure the concentration of the library by integrating the area under the peak.

Hybridization/capture:

1. Hybridization setup

Overnight hybridization using the SureSelect Methly-Seq hybridization kit, boxes one and two.

- ☐ The ligated and purified libraries must be dried down to 4 µL by SpeedVac.
- ☐ Prepare the hybridization buffer mixture (Table 1).

Table 1. Preparation of hybridization buffer.

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)	17 Rxn with Excess (µL)
SureSelect Hyb 1 (orange cap or bottle)	6.63	60	112,7
SureSelect Hyb 2 (red cap)	0.27	2.5	4,59
SureSelect Hyb 3 (yellow cap or bottle)	2.65	24	45
SureSelect Hyb 4 (black cap or bottle)	3.45	31	58,65
	13 µL	117.5 µL	220,94 µL

- ☐ Prepare the SureSelect Block mix (Table 2).

Table 2. Preparation of SureSelect Block mix.

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)
SureSelect Indexing Block 1 (green cap)	2.5	22.5
SureSelect Block 2 (blue cap)	2.5	22.5
SureSelect Methyl-Seq Block 3 (brown cap)	0.6	5.4
	5.6 µL	50.4 µL

- ☐ To each unamplified library, add 5.6 µL of the SureSelect Block mix created above. Vortex for 5 seconds, then spin down.
- ☐ Cap the wells, then transfer the sealed plate or strip tube to the thermal cycler and run the following program shown in Table 3.

Table 3. Thermal cycler program for DNA + Block mix prior to hybridization.

Step	Temperature	Time
Step 1	95 °C	5 minutes
Step 2	65 °C	Hold (at least 5 minutes)

- ☐ Prepare the appropriate dilution of SureSelect RNase Block mix, based on the size of your capture library (Table 4).

Table 4. Dilution of SureSelect RNase Block mix based on capture library size.

Capture Library Size	RNase Block Dilution (x Parts RNase Block: x Parts Water)	Volume of Dilute RNase Block Required per Hybridization Reaction
> 3 Mb	25% (1:3)	2 µL
< 3 Mb	10% (1:9)	5 µL

☐ Prepare the capture library hybridization mix based on the size of your capture library (Tables 5 and 6). Keep the mixture at room temperature briefly until use.

Table 5. Preparation of capture library hybridization mix for > 3 Mb capture libraries.

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)
Hybridization Buffer Mixture	13	117
25% RNase Block Solution	2	18
SureSelect Capture Library	5	45
	20 µL	180 µL

Table 6. Preparation of capture library hybridization mix for > 3 Mb capture libraries.

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)
Hybridization Buffer mixture	13	117
10% RNase Block solution	5	45
SureSelect capture library	2	18
	20 µL	180 µL

☐ Maintain the library + RNase Block mix plate or strip tube at 65 °C while you add 20 µL of the capture library hybridization mix to each sample well. Mix well by pipetting up and down 8 to 10 times.

☐ Incubate the hybridization mixture for 16 hours at 65 °C with a heated lid at 105 °C.
(Cap well to avoid evaporation.)

2. Prepare streptavidin beads

Please refer to the Quick Start protocol, SureSelect XT HS2 DNA with Post-capture Pooling, hybridization/capture step 2; page 5.

3. Capture hybridized library

☐ Check the sample volume after hybridization (> 25 µL).

☐ Transfer the hybridization mixture to the bead solution at room temperature (RT). Mix up and down to fully resuspend the beads.

☐ Incubate at RT for 30 minutes at 1400 rpm.

☐ Place on a magnetic separator. Discard the supernatant.

☐ Resuspend the beads with 200 µL SureSelect Wash 1, then incubate for 15 minutes at RT.

☐ Place on a magnetic separator. Discard the supernatant.

☐ Wash the beads with 200 µL (prewarmed, 65 °C) SureSelect Wash 2, incubate each was at 65 °C for 10 minutes on a thermal cycler (heated lid). Repeat 3 times.

☐ Prepare fresh 0.1 M NaOH for elution of the captured libraries.

☐ Add 20 µL of the freshly-prepared 0.1 M NaOH solution to the bead-bound samples and mix on a vortex mixer for 5 seconds to resuspend the beads.

☐ Incubate the samples for 20 minutes at RT.

☐ Collect the beads from the elution mixture on a magnetic separator and transfer the supernatant from each well to wells of a fresh plate or strip tube.

Bisulfite conversion:

1. Hybridization setup

☐ Add 130 µL of **Lightning Conversion Reagent** to 20 µL of a DNA sample in a PCR tube. Mix, then centrifuge briefly to ensure there are no droplets in the cap or on the sides of the tube.

Note: If the volume of DNA is less than 20 µL, bring to volume with nuclease-free water.

Note: Samples > 20 µL must be processed using multiple conversion reactions. Replicate reactions can be cleaned using the same column for each by repeating steps 3 to 5.

☐ Place the PCR tube in a thermal cycler and perform the following steps:

1. 98 °C for 8 minutes
2. 54 °C for 60 minutes
3. 4 °C storage for up to 20 hours

☐ Add 600 µL of **M-Binding Buffer** to a Zymo-SpinTM IC Column and place the column into a provided Collection Tube.

Note: The capacity of the collection tube with the column inserted is 800 µL.

Empty the collection tube whenever necessary to prevent

contamination of the column contents by the flow-through.

- ☐ Load the sample (from step 2) into the Zymo-SpinTM IC Column containing the M-Binding Buffer. Close the cap and mix by inverting the column several times.
- ☐ Centrifuge at full speed (> 10,000 x g) for 30 seconds. Discard the flow-through.
- ☐ Add 100 µL of **M-Wash Buffer** to the column. Centrifuge at full speed for 30 seconds.
- ☐ Add 200 µL of **L-Desulphonation Buffer** to the column and let stand at room temperature (20 to 30 °C) for 17 minutes. After the incubation, centrifuge at full speed for 30 seconds.
- ☐ Add 200 µL of **M-Wash Buffer** to the column. Centrifuge at full speed for 30 seconds. Repeat this wash step.
- ☐ Place the column into a 1.5 mL microcentrifuge tube and add 20 µL of H_2O directly to the column matrix. Centrifuge for 30 seconds at full speed to elute the DNA.

2. PCR amplification

First PCR using the SureSelect Methyl-Seq library prep kit.

- ☐ Prepare the PCR #1 reaction mix (Table 7).

Table 7. Preparation of PCR #1 reaction mix.

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)
Nuclease-Free Water	30	270
SureSelect Methyl-Seq PCR Master Mix	50	450
Methyl-Seq PCR1 Primer F	1	9
Methyl-Seq PCR1 Primer R	1	9
	82 µL	738 µL

- ☐ Add 82 µL of PCR #1 mix to each 20 µL DNA sample, mix well.
- ☐ Incubate in the thermal cycler on PCR #1 cycling profile (Table 8).

Table 8. PCR #1 cycling program for bisulfite-converted libraries.

PCR #1 Cycling Profile (100 µL rxn vol)			
Segment	**Number of Cycles**	**Temperature**	**Time**
1	1	95 °C	2 min
2	To be determined by design size	95 °C	30 sec
		60 °C	30 sec
		72 °C	30 sec
3	1	72 °C	7 min
4	1	4 °C	Hold

3. Ampure cleanup

- ☐ Vortex the AMPure XP beads.
- ☐ Incubate the DNA sample with 100 µL AMPure XP beads at RT for 5 minutes.
- ☐ Wash with fresh 70% ethanol (200 µL), 2 times.
- ☐ Elute with 25 µL nuclease-free water (mix by vortexing, incubate 5 minutes).

4. Indexing PCR

SureSelect XT HS2 target enrichment kit ILM Hyb module, box 2 and SureSelect XT HS2 index primer pairs for ILM (pre-PCR).

- ☐ Prepare the PCR #2 reaction mix (Table 9).

Table 9. Preparation of PCR #2 reaction mix..

Reagent	Per Rxn (µL)	8 Rxn with Excess (µL)
Nuclease-Free Water	9	81
5X Herculase II Rxn Buffer with dNTPs (clear)	10	90
Herculase II Fusion DNA Polymerase (red)	1	9
	20 µL	180 µL

- ☐ Add 5 µL of the appropriate SureSelect XT HS2 indexing primer pair to each sample in the PCR reaction.
- ☐ Add 20 µL of PCR #2 mix to each sample in the PCR reaction. Mix well.
- ☐ Incubate in the thermal cycler on PCR #2 cycling profile (Table 10).

Table 10. PCR #2 cycling program for indexing amplification.

PCR #2 Cycling Profile (50 µL rxn vol)			
Segment	**Number of Cycles**	**Temperature**	**Time**
1	1	98 °C	2 min
2	8	98 °C	30 sec
		60 °C	30 sec
		72 °C	1 min
3	1	72 °C	5 min
4	1	4 °C	Hold

5. Final Ampure cleanup

- ☐ Vortex the AMPure XP beads.
- ☐ Incubate the amplified library supernatant with 50 µL AMPure XP beads at RT for 5 minutes.
- ☐ Wash with fresh 70% ethanol (200 µL), 2 times.
- ☐ Dry the samples on a heat block 37°C for ~2 minutes or at RT ~ 10 minutes.
- ☐ Elute with 25 µL nuclease-free water (mix by vortexing, incubate 2 minutes).

Analysis Procedures

In the experimental study, four groups of eight mice were classified according to sex (male/female) and treatment (saline/VPA). A total of 64 samples (32 blood and 32 brain) were processed according to the protocol described above. Sequencing was performed on a mid-output flow cell on NextSeq550 (Illumina) by multiplexing 16 samples per run with the following parameters: paired end, read length 151, double index, generating 10 million reads per sample.

The analysis has been performed in two ways:

- Command line scripts
- Commercial software

The command line pipeline created for the analysis mainly utilized Bismark (https://github.com/FelixKrueger/Bismark) for the cleaning, alignment (using the Bowtie2 (https://github.com/BenLangmead/bowtie2) aligner), and call phases.

The BiSeq R package (https://www.bioconductor.org/packages/release/bioc/html/BiSeq.html) was then used to calculate the coverages, to check the quality of the sequencing and alignment, and to carry out the analysis to determine the differentially methylated regions (DMRs).

Alignment Stats

Sequence pairs analysed in total	9943476
Paired-end alignments with a unique best hit	6456110
Pairs without alignments under any condition	3370155
Pairs that did not map uniquely	117211
Genomic sequence context not extractable (edges of chromosomes)	21

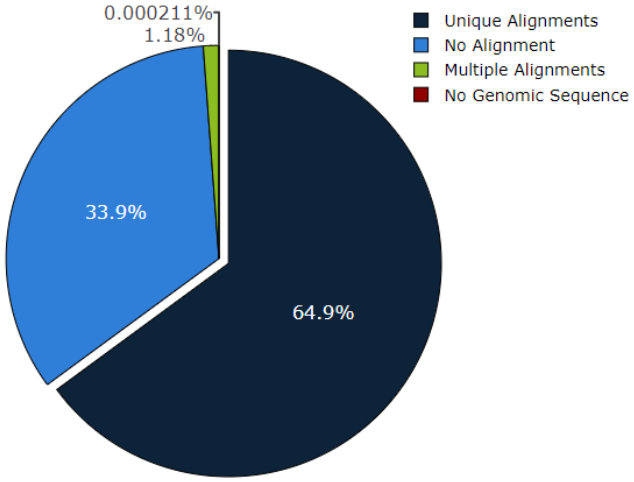

Figure 1. Bismark report after bisulfite-treated sample read alignment. It groups the basic alignment metrics against the genomic build used.

Over the captured regions, the read depth often exceeded 100X (Figure 2). However, the total number of reads per sample was limited (only 10 million reads per sample) as expected from the custom design strategy (see custom design section for details). With an on-target percentage > 75% for all samples, it shows the highly efficient capture performance of the Agilent SureSelect 3 Mb Methyl-Seq custom design.

Based on the detailed results from Bismark, it was possible to represent the data according to the different parameters studied and then highlight the cytosines, genes, or regions of interest. (Data not shown.)

For all the analyses relating to single cytosines or whole genes, specific Perl scripts were developed.

A Github section has been created to share all the script used for this project. Please refer to: GitHub - qRp/Methylation_scripts.

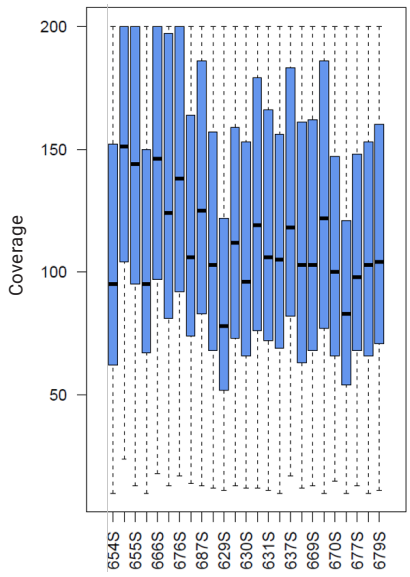

Figure 2. Representation of the average coverage per sample on the regions captured. More than 100X depth were obtained on average.

The results were also analyzed using a commercial software: StrandNGS, (https://www.strand-ngs.com/). For analysis in StrandNGS, the BAM files generated during the previous steps were used. StrandNGS collects the "tags" generated on sequence reads by Bismark and integrates them into a built-in analysis pipeline. The analysis can be restricted to the region of interest by incorporating the custom design target BED file into the system. The rest of the analysis followed the StrandNGS wizard-driven workflow for Methyl-Seq analyses. For the study presented here, default settings were used.

The advantage of this approach is its ease of use for non-bioinformatic experts. The successive steps of the analysis include the detection of methylation in each sample followed by the evaluation of methylation differences at the cytosines scale (DMCs), and the analysis of these DMCs into differentially methylated regions (DMRs). The results were then annotated (by genes involved, GO analysis, or pathway analysis). StrandNGS software also includes various graphical data representations and database integrations that can be used easily by non-statistical experts.

The results of this study were visualized using the StrandNGS integrated genome browser (Figure 3), histogram plots (Figure 4), and lollipop plots (Figure 5). Intergroup comparisons allowed the generation of lists of DMCs and DMRs, compiled easily into Venn diagrams using the StrandNGS software.

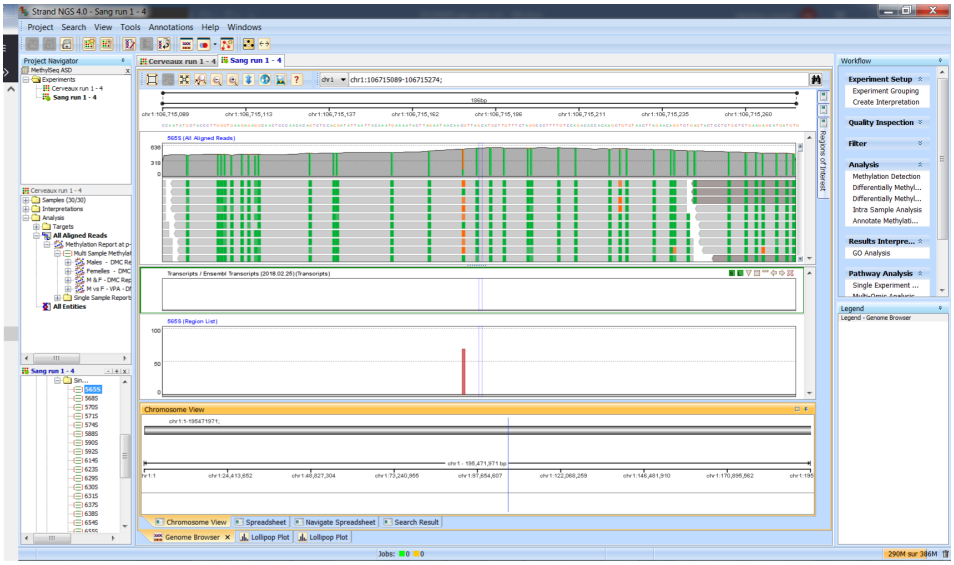

Figure 3. Visualization in the StrandNGS integrated genome browser of a region captured on chromosome 1, showing a cytosine with a methylation rate of approximately 75%.

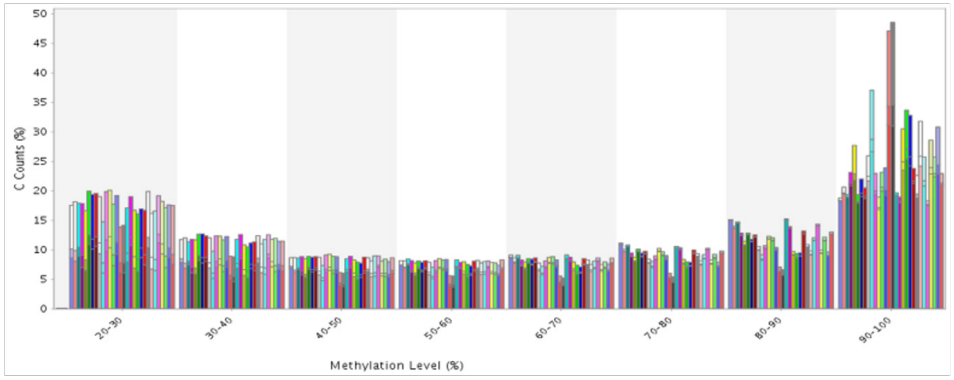

Figure 4. Visualization in the form of a histogram of the global distribution of methylation in the samples analyzed (n=30). The majority of cytosines analyzed have a methylation rate > 90% or < 30%.

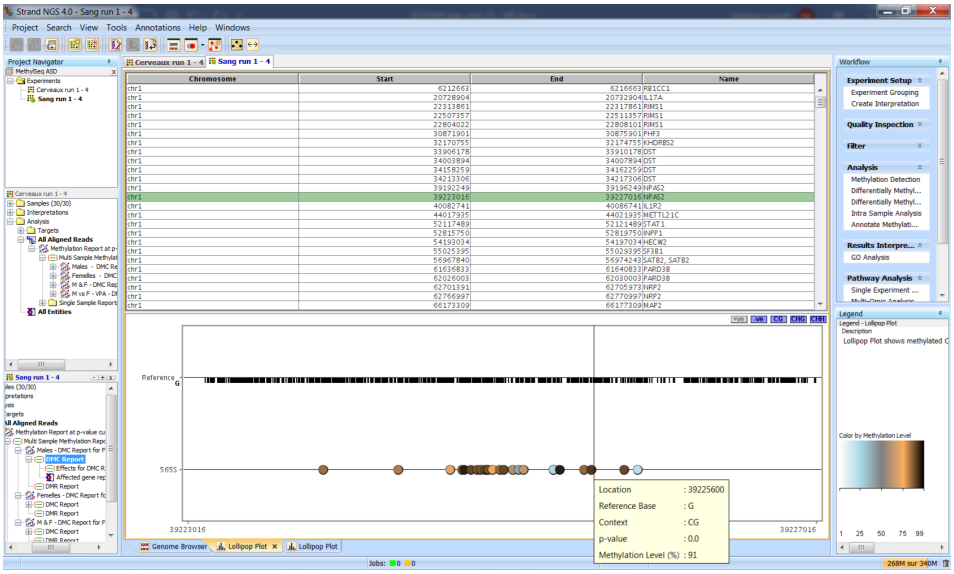

Figure 5. Lollipop plot showing the distribution of DMCs in a region of chromosome 1 corresponding to the NPAS2 gene in one blood sample.

The combination of these two complementary bioinformatics approaches lead to the identification of 67 common genes between all groups having at least one DMC (Figure 6). This joint analysis confirms that differences in methylation in the brain largely occur on "isolated" cytosines (not located in CpG islands), while in blood no difference in the CpG or CHG/CHH context of methylation distribution was observed. (Figure 7). Furthermore, utilizing target enrichment Methyl-Seq makes visualization of very small methylation variation possible by focusing into specific regions of interest.

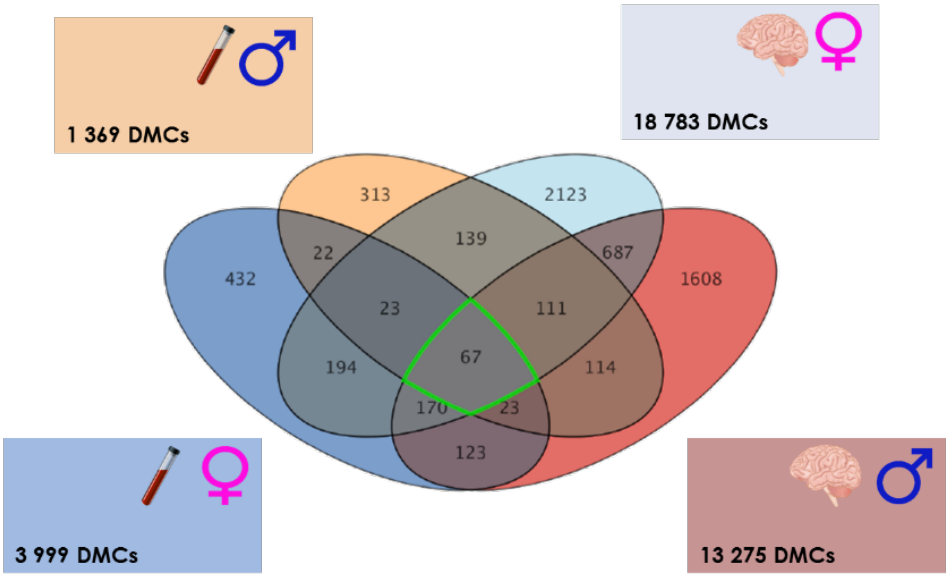

Figure 6. Venn diagram showing gene transcripts affected by DMCs between groups.

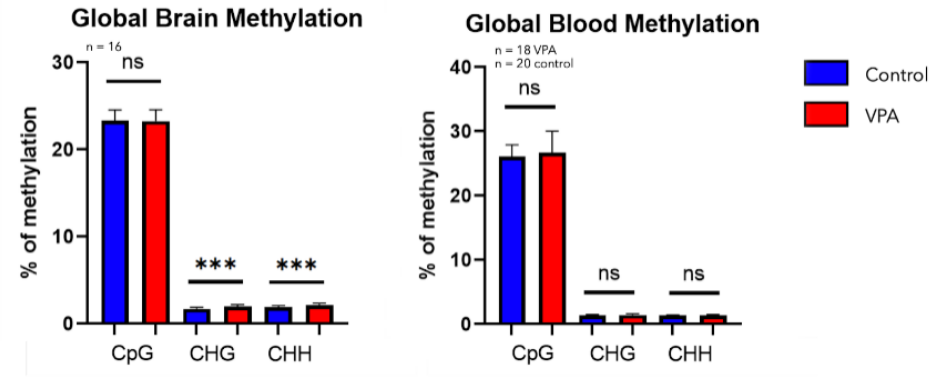

Figure 7. Comparison of global methylation levels in the brain (left) and in the blood (right) of control versus VPA mice, according to the context of cytosine localization (CpG islands or outside CpG: CHG or CHH). Mann-Whitney statistical test.

Library Pooling and Sequencing

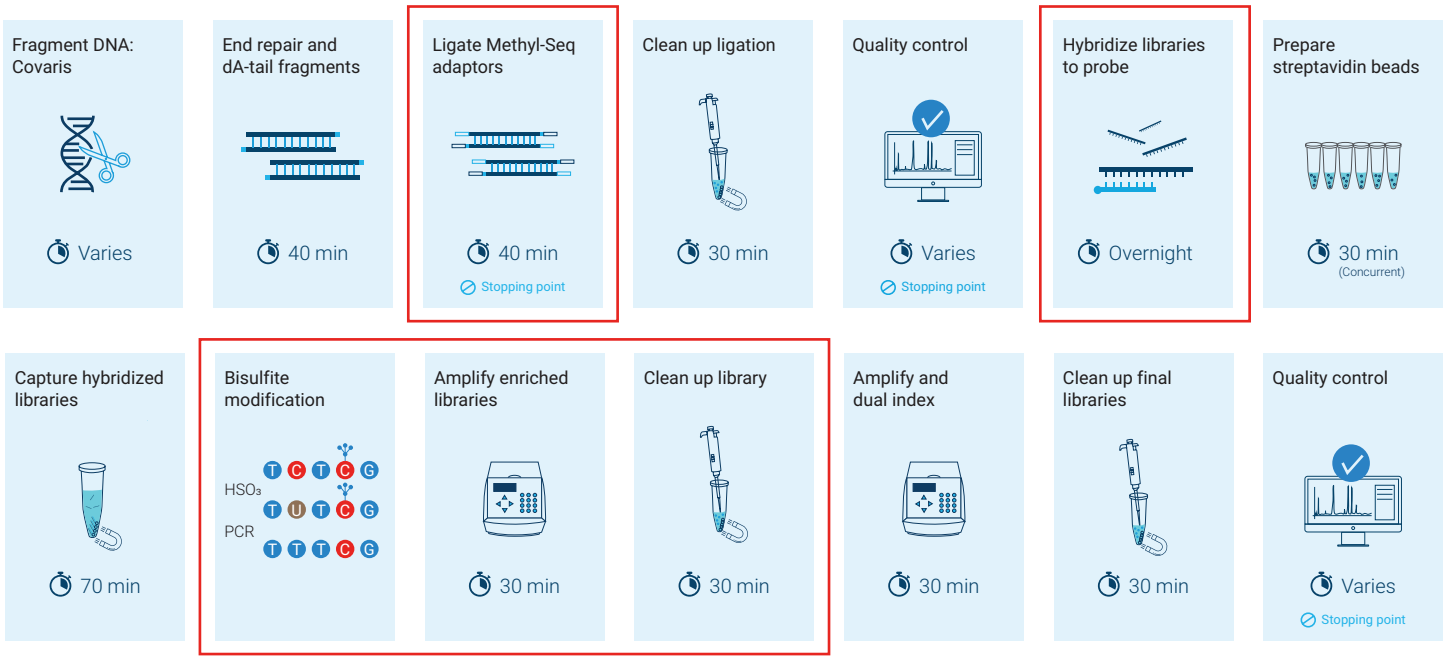

Figure 8. Modified Methyl-Seq SureSelect targeted capture protocol at a glance.

Conclusion

Using a custom capture panel for this project had three major advantages. First, it gives the possibility to perform the analysis on a list of preselected candidate genes and genomic regions, therefore allowing more samples to be multiplexed and ultimately limiting the overall sequencing costs. Second, the capture enables the analysis of cytosines outside the CpG islands, which is particularly interesting in embryonic stem cells or in neurons which present high levels of methylation outside the CpG context.[3] Finally, by using target enrichment, the sequencing depth obtained was increased (> 100X on average and > 400X in certain regions of interest) enabling the reliable detection of methylation difference by a few percentages.

The full results of this study will be discussed at a forthcoming symposium and are subject to publication.

References

1. Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2012**, *38* (1), 23–38. https://doi.org/10.1038/npp.2012.112.
2. Jang, H. S.; Shin, W.; Lee, J. E.; Tae, J. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes* **2017**, *8* (6), 148. https://doi.org/10.3390/genes8060148.
3. Lister, R.; Mukamel, E. A.; Nery, J. R.; Urich, M. A.; Puddifoot, C. A.; Johnson, N.; Lucero, J.; Huang, Y.; Dwork, A. J.; Schultz, M. D.; et al. R. Global Epigenomic Reconfiguration during Mammalian Brain Development. *Science* **2013**, *341* (6146). https://doi.org/10.1126/science.1237905.

www.agilent.com

PR7001-1658

Published in the USA, Decemberr 5, 2023
5994-6841EN

Agilent
Trusted Answers