



## Fast and convenient 5-hydroxymethylcytosine enrichment workflow for next-generation sequencing

The Scientific™ EpiJET™ 5-hmC Enrichment Kit is highly specific for different DNA samples containing 5-hydroxymethylcytosine (5-hmC), an extensively studied DNA epigenetic modification. This tool, when combined with next-generation sequencing, offers new ways of analyzing 5-hmC at the genomic level. The data show that this tool can be used for both locus-specific and whole genome-specific analysis, yielding 5-hmC distribution patterns across different genomes.

### Introduction

The Thermo Scientific EpiJET 5-hmC Enrichment Kit is a new tool for 5-hmC DNA enrichment. This kit uses highly specific enzymatic-based labeling of 5-hmC followed by chemical biotin labeling and enrichment via streptavidin-coated magnetic beads. Compared to other methods, the kit takes less time to enrich 5-hmC-specific DNA. In addition, the kit requires low amounts of starting DNA, has a low background and does not exhibit any bias toward unspecific sequences. The kit is fully compatible with next-generation sequencing (NGS) library-preparation tools and sequencing platforms. This versatility allows researchers to quickly analyze any genome for epigenetic modifications. Until a few years ago, only one DNA modification was well known in mammalian cells—5-methylcytosine (5-mC). This modification has been extensively studied, and a number of its important epigenetic functions (e.g., gene regulation, X chromosome imprinting) are known. In 2009, 5-hmC, a forgotten DNA modification, was rediscovered, and this resulted in a new age of epigenetics<sup>1</sup>. 5-hmC immediately became an intensively studied modification, and subsequent studies revealed not only the mechanism by which this base is produced *in vivo* via TET1-mediated oxidation<sup>2</sup>, but also the mechanism for generating two more DNA modifications, 5-formylcytosine and 5-carboxylcytosine<sup>3</sup>. The enrichment methods took the analysis of these modifications to a new level, and the unique capabilities of the Thermo Scientific EpiJET 5-hmC Enrichment Kit make this analysis simple and fast and suit any research needs.

### Method overview

The total DNA analysis with NGS is done in three simple steps and takes less than 4 h (Fig. 1), and the Thermo Scientific EpiJET 5-hmC Enrichment Kit specifically enriches only the 5-hmC-containing DNA fraction for further analysis. The first step in the 5-hmC analysis workflow with NGS is library preparation by fragmentation of 5-hmC-containing DNA. This can be achieved using a variety of physical methods (e.g., sonication or hydrodynamics) with Thermo Scientific ClaSeek™ library kits. Alternatively, one can use enzymatic (e.g., transposon-based) fragmentation methods with MuSeek™ library-preparation kits, which takes less than 15 min.

The most important step is the second, where DNA is tagged using a specifically formulated 5-hmC-modifying enzyme included in the EpiJET 5-hmC Enrichment Kit; this step is completed in just 1 h. The tagged DNA is covalently conjugated with biotin in only 5 min with biotin-conjugation solution, and the biotin-labeled DNA is enriched using proprietary streptavidin-coated magnetic beads. After enrichment, the DNA is conveniently eluted in water and can be used directly for quantitative PCR (qPCR) (if specific loci are analyzed), microarrays or sequencing (if a whole genome is analyzed). The whole EpiJET 5-hmC Enrichment Kit procedure is straightforward and can be completed in less than 3 h.

The last step is prepared-library analysis by NGS. After 5-hmC enrichment and PCR amplification, libraries can be analyzed by NGS using either Ion Torrent™ or Illumina® platforms.

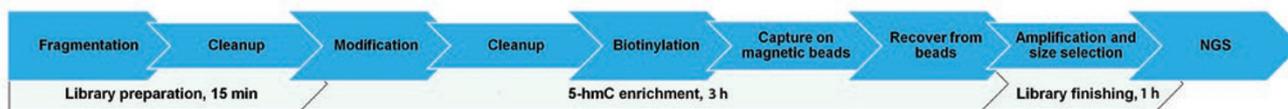
### Model system: high specificity for 5-hmC

To analyze the specificity of the method for 5-hmC enrichment, we developed a bacterial genome-based control system. *Staphylococcus aureus* genomic DNA was modified *in vitro* in such a way that GCGC sequences were converted to GhmCGC with nearly 100% efficiency. To control the specificity of our enrichment method, we used *Escherichia coli* genomic DNA with all CG sites methylated.

Milda Kaniušaitė, Eimantas Astromskas, Gediminas Alzbutas, Renata Bružaitė & Arūnas Lagunavičius

Thermo Fisher Scientific, Vilnius, Lithuania. Correspondence should be addressed to A.L. (arunas.lagunavicius@thermofisher.com) or E.A. (eimantas.astromskas@thermofisher.com).

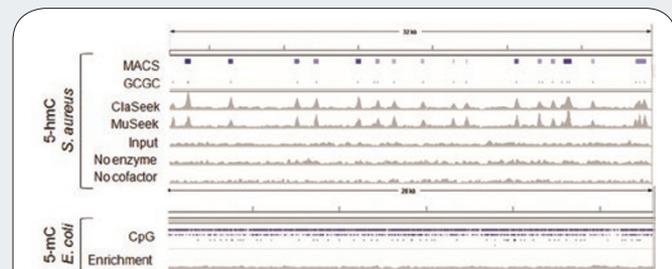
## APPLICATION NOTE



**Figure 1** | Overview of EpiJET 5-hmC Enrichment Kit workflow for NGS. Library preparation using ClaSeek or MuSeek protocols is followed by 5-hmC enrichment, which takes less than 3 h for six samples. Libraries can be amplified and size selected depending on the sequencing platform's requirements.

The 5-mC-modified *E. coli* DNA showed no enrichment, which demonstrated the ability of the enrichment method to discriminate between 5-hmC and 5-mC.

To demonstrate the kit's compatibility with different NGS library-preparation methods, we prepared NGS libraries using the ClaSeek Library Preparation Kit, Ion Torrent compatible and the MuSeek Library Preparation Kit for Ion Torrent. Following PCR amplification and size selection, we sequenced these libraries before and after 5-hmC enrichment using the Ion PGM™ System, which resulted in deep enough coverage for analysis of enrichment specificity (**Fig. 2**). After sequencing, bacterial genomic DNA peaks were called with MACS software. Only *S. aureus* (modified with 5-hmC) contained clear and reliable peaks. More than 90% of GCGC sequences were detected as positives, which demonstrates the high specificity of the EpiJET 5-hmC Enrichment Kit for this DNA modification (**Fig. 2**).

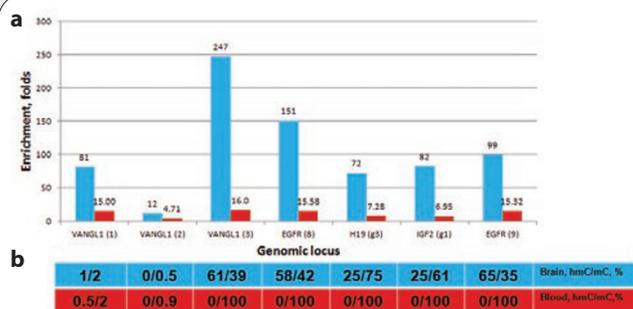


**Figure 2** | High specificity of 5-hmC enrichment with different library-preparation methods. No enrichment was observed with 5-mC-containing sequences or without the indicated reaction components. For library preparation, 1 µg of bacterial DNA was used with the ClaSeek Library Preparation Kit, Ion Torrent compatible (lane: ClaSeek), and 100 ng was used for the MuSeek Library Preparation Kit for Ion Torrent (lane: MuSeek). After being sequenced on the Ion PGM System, the libraries were visualized with a genome viewer, and a snapshot was taken of a 32 kb 5-hmC-modified *S. aureus* region and a 28-kb 5-mC-modified *E. coli* region. MACS, peaks called by MACS software. GCGC, positions of 5-hmC-modified GmCGC sequences in the *S. aureus* genome. Input, library with no enrichment. No enzyme, enriched without 5-hmC-modifying enzyme. No cofactor, enriched without a cofactor for the 5-hmC-modifying enzyme. CpG, positions of 5-mC-modified CG sequences in the *E. coli* genome. Enrichment, libraries enriched using EpiJET™ kit protocol with 5-mC-modified *E. coli* DNA at all CpG sites.

### Human DNA analysis: high 5-hmC levels in brain DNA

To further demonstrate the capabilities of our kit, we analyzed human brain DNA samples. Mammalian brain DNA is known to contain high levels of 5-hmC and was used to discover this DNA modification<sup>1</sup>. First, we analyzed the enriched DNA using well-established methods such as qPCR and restriction enzyme digestion coupled with T4 β-glucosyltransferase-based glucosylation (**Fig. 3**). The results showed that the analyzed regions exhibited high 5-hmC levels at a number

of CCGG sites, as described previously<sup>4</sup>. Following the EpiJET 5-hmC enrichment protocol and qPCR analysis, we confirmed that these regions were highly 5-hmC enriched (**Fig. 3a**).



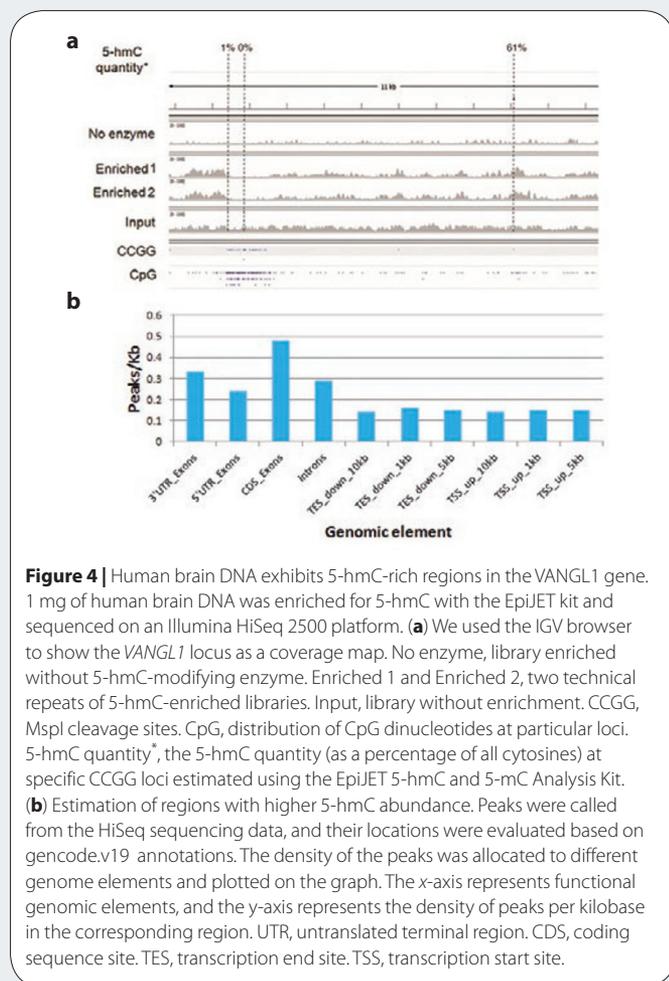
**Figure 3** | Human brain DNA exhibits high 5-hmC levels. (**a,b**) 1 µg of human brain and blood DNA was enriched for 5-hmC with the EpiJET kit, and the enrichment of several loci was evaluated by qPCR. (**a**) Enrichment was calculated as the ratio of brain or blood DNA loci to nonspecific spike-in control DNA yield. (**b**) Specific CCGG sites at different genomic locations were analyzed for 5-mC and 5-hmC levels (as a percentage of all cytosines) using the EpiJET 5-hmC and 5-mC Analysis Kit.

Techniques based on qPCR allow analysis of only a few loci in the genome, giving no clear picture of whole-genome 5-hmC distribution. To analyze the whole genome, we prepared 5-hmC-enriched human brain DNA libraries and sequenced them on an Illumina® HiSeq® 2500 platform (**Fig. 4a**). Two technical replicates were analyzed, resulting in ~360 M pair-end reads each. As controls, we used 'no enzyme' enrichment and nonenriched 'input' libraries. The reads were mapped on the GRCh37 human reference genome, and peaks were called with MACS v.1.4.2. A closer look at the *VANGL1* locus revealed that the promoter region of the gene contained higher 5-hmC levels than a gene body. Moreover, the CpG island just before the gene was devoid of any 5-hmC, whereas the gene itself contained medium levels of 5-hmC. When we compared NGS data from the EpiJET 5-hmC Analysis Kit and the EpiJET DNA Methylation Analysis Kit, we found a close correlation between the two methods (*VANGL1* locus; **Figs. 3 and 4a**).

### Human DNA analysis: high 5-hmC levels in exon sequences

To further analyze our NGS dataset, we studied the 5-hmC distribution over different genetic elements at the genome level (**Fig. 4b**). This analysis showed that most 5-hmC modifications were located at the coding regions of genes (CDS-exons), followed by the 3'-UTRs of exons. Regions flanking transcription start or stop sites were less

5-hmC abundant. Importantly, these data are well in accordance with an earlier analysis of 5-hmC distribution in the mammalian brain genome<sup>5</sup>, yet again confirming the high specificity of the Thermo Scientific EpiJET 5-hmC Enrichment Kit for 5-hmC-containing DNA sequences.



**Figure 4** | Human brain DNA exhibits 5-hmC-rich regions in the VANGL1 gene. 1 mg of human brain DNA was enriched for 5-hmC with the EpiJET kit and sequenced on an Illumina HiSeq 2500 platform. **(a)** We used the IGV browser to show the VANGL1 locus as a coverage map. No enzyme, library enriched without 5-hmC-modifying enzyme. Enriched 1 and Enriched 2, two technical repeats of 5-hmC-enriched libraries. Input, library without enrichment. CCGG, MspI cleavage sites. CpG, distribution of CpG dinucleotides at particular loci. 5-hmC quantity\*, the 5-hmC quantity (as a percentage of all cytosines) at specific CCGG loci estimated using the EpiJET 5-hmC and 5-mC Analysis Kit. **(b)** Estimation of regions with higher 5-hmC abundance. Peaks were called from the HiSeq sequencing data, and their locations were evaluated based on gencode.v19 annotations. The density of the peaks was allocated to different genome elements and plotted on the graph. The x-axis represents functional genomic elements, and the y-axis represents the density of peaks per kilobase in the corresponding region. UTR, untranslated terminal region. CDS, coding sequence site. TES, transcription end site. TSS, transcription start site.

## Conclusion

The Thermo Scientific EpiJET 5-hmC Enrichment Kit is a new, fast, simple and highly specific method for 5-hmC DNA enrichment. The present study demonstrates the utility of the EpiJET 5-hmC Enrichment Kit as a fast, simple and versatile tool for efficient enrichment of 5-hmC-containing DNA over unmodified and 5-mC-containing DNA. The kit takes advantage of the 5-hmC modifying enzyme, which is formulated for highly specific and efficient modification of 5-hmC present in CpG dinucleotides of DNA, and it does not have any activity on unmodified or methylated cytosines. The 5-hmC DNA enrichment procedure can be completed in just 3 h without compromising yields or efficiency. It requires small amounts of starting material and is compatible with different NGS library-preparation solutions. 5-hmC-enriched libraries can be analyzed with either Ion Torrent or Illumina platforms. The resulting DNA exhibits a clearly enriched profile after NGS, indicating most of the regions containing 5-hmC in the analyzed DNA.

## ACKNOWLEDGMENTS

We are grateful to S. Serva, A. Berezniakovas, L. Zakrys, J. Lubienė, V. Šeputienė, J. Vitkutė, A. Leipus, A. Petronis and A. Lubys for technical support and valuable discussions during the project. The 5-hmC labeling and enrichment method was exclusively licensed from S. Klimašauskas' group (Vilnius University, Institute of Biotechnology, Lithuania).

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
2. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
3. to, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
4. Kinney, S.M. *et al.* Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J. Biol. Chem.* **286**, 24685–24693 (2011).
5. Wen, L. *et al.* Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15**, R49 (2014).

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.