Video Article

Selective Capture of 5-hydroxymethylcytosine from Genomic DNA

Yujing Li[1], Chun-Xiao Song[2], Chuan He[2], Peng Jin[1]

[1]Department of Human Genetics, Emory University School of Medicine

[2]Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago

Correspondence to: Peng Jin at peng.jin@emory.edu

Abstract

5-methylcytosine (5-mC) constitutes ~2-8% of the total cytosines in human genomic DNA and impacts a broad range of biological functions, including gene expression, maintenance of genome integrity, parental imprinting, X-chromosome inactivation, regulation of development, aging, and cancer[1]. Recently, the presence of an oxidized 5-mC, 5-hydroxymethylcytosine (5-hmC), was discovered in mammalian cells, in particular in embryonic stem (ES) cells and neuronal cells[2-4]. 5-hmC is generated by oxidation of 5-mC catalyzed by TET family iron (II)/α-ketoglutarate-dependent dioxygenases[2, 3]. 5-hmC is proposed to be involved in the maintenance of embryonic stem (mES) cell, normal hematopoiesis and malignancies, and zygote development[2, 5-10]. To better understand the function of 5-hmC, a reliable and straightforward sequencing system is essential. Traditional bisulfite sequencing cannot distinguish 5-hmC from 5-mC[11]. To unravel the biology of 5-hmC, we have developed a highly efficient and selective chemical approach to label and capture 5-hmC, taking advantage of a bacteriophage enzyme that adds a glucose moiety to 5-hmC specifically[12].

Here we describe a straightforward two-step procedure for selective chemical labeling of 5-hmC. In the first labeling step, 5-hmC in genomic DNA is labeled with a 6-azide-glucose catalyzed by β-GT, a glucosyltransferase from T4 bacteriophage, in a way that transfers the 6-azide-glucose to 5-hmC from the modified cofactor, UDP-6-N3-Glc (6-N3UDPG). In the second step, biotinylation, a disulfide biotin linker is attached to the azide group by click chemistry. Both steps are highly specific and efficient, leading to complete labeling regardless of the abundance of 5-hmC in genomic regions and giving extremely low background. Following biotinylation of 5-hmC, the 5-hmC-containing DNA fragments are then selectively captured using streptavidin beads in a density-independent manner. The resulting 5-hmC-enriched DNA fragments could be used for downstream analyses, including next-generation sequencing.

Our selective labeling and capture protocol confers high sensitivity, applicable to any source of genomic DNA with variable/diverse 5-hmC abundances. Although the main purpose of this protocol is its downstream application (*i.e.*, next-generation sequencing to map out the 5-hmC distribution in genome), it is compatible with single-molecule, real-time SMRT (DNA) sequencing, which is capable of delivering single-base resolution sequencing of 5-hmC.

Video Link

The video component of this article can be found at http://www.jove.com/video/4441/

Protocol

1. Genomic DNA Fragmentation

Fragment genomic DNA using sonication to a desired size range suited for the genome-wide sequencing platform. (We usually sonicate to ~300 bp.) Verify the size distribution of the fragmented genomic DNA on 1% agarose gel (**Figure 1**).

2. DNA Preparation

Determine the starting DNA amounts based on the abundance of 5-hmC in genomic DNA. Since 5-hmC levels vary significantly in different tissue types, starting DNA amounts depend on the 5-hmC levels of the samples. Please refer to **Table 1** for examples.

3. β-GT Catalyzed Reaction (Glucose Transfer Reaction)

1. Mix by pipetting the mixture as detailed in **Table 2** and incubate in a 37 °C water bath for 1 hr.

2. After incubation, clean up the reaction with QIAquick Nucleotide Removal Kit, using 10 µg of DNA per column. Elute with 30 µl water per column and combine.

4. Biotinylation Reaction (Click Chemistry)

1. Add DBCO-S-S-PEG3-Biotin conjugate working solution (1 mM) in the eluted DNA solution (from step 3) to a final concentration of 150 µM (*i.e.*, 5 µl of working solution per 30 µl DNA solution).
2. Mix by pipetting and incubate in a 37 °C water bath for 2 hr.
3. Clean up the reaction with QIAquick Nucleotide Removal Kit. The ideal total elution volume is 100 µl.
4. Quantify the recovered DNA amount using microliter scale spectrophotometer (*e.g.*, NanoDrop).

5. Capture of 5-hmC-containing DNA

1. Wash 50 µl of Dynabeads MyOne Streptavidin C1 3 times with 1 ml of 1X B&W buffer according to the manufacturer's instructions. Separate the beads with a magnetic stand.
2. Add equal volume of 2X B&W buffer to the recovered biotinylated DNA (100 ul) to the washed beads.
3. Incubate for 15 min at room temperature with gentle rotation on a rotator.
4. Separate the beads with a magnetic stand and wash the beads 3 times with 1 ml of 1X B&W buffer.
5. Elute the DNA by incubating the beads in 100 µl of freshly prepared 50 mM DTT for 2 hr at room temperature with gentle rotation on a rotator.
6. Separate the beads with a magnetic stand. Aspirate the eluent and load onto a Micro Bio-Spin 6 Column according to the manufacture instruction to remove the DTT. The target DNA is in the solution now.
7. Purify the eluted DNA from the previous step by Qiagen MinElute PCR Purification Kit and elute DNA in 10 µl of EB buffer. Quantify DNA using Qubit Fluorometer, or NanoDrop if concentration is higher than 20 ng/ul. The DNA is ready for downstream genome-wide sequencing library preparation.

6. Representative Results

If the quality of genomic DNA is high, typical recovery yields after the β-GT and biotinylation reactions are ~60-70%. However, the capture efficiency vary significantly with different tissue types depending on the 5-hmC levels of the samples. Typically, the capture efficiency for brain genomic DNA is ~4-9%, and in some extreme cases, the efficiency may reach up to 12%. For ES cells, the average capture efficiency is ~2-4%, in contrast to ~0.5% for neural stem cells. The lowest efficiency seen so far was for genomic DNA from cancer cells. All enriched DNA is ready for standard next-generation library preparation protocols. In addition, the captured DNA can also be used as template for real-time PCR to detect the enrichment of some fragments compared to the input DNA, if the related primers are available.
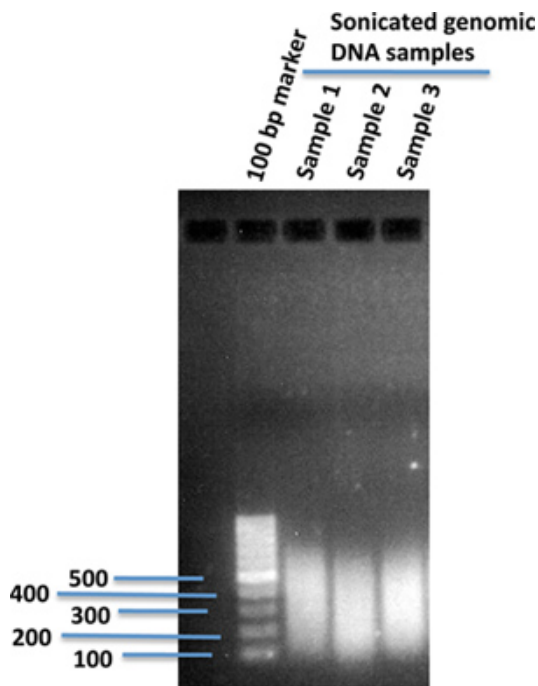
Figure 1. Sonicated human genomic DNA fragments in 1% agarose gel. 10 µg of genomic DNA isolated from human iPS cells in 120 µl of 1X TE buffer was sonicated using a sonication device (Covaris). After sonication, 2 µl of the sonicated DNA was loaded onto 1% agarose gel using 100 bp of DNA marker to compare the sizes of the sonicated DNA fragments.

Component	Volume	Final Concentration
Water	_ µl	
10 X β-GT Reaction Buffer	2 µl	1 X
Up to 10 µg genomic DNA	_ µl	Up to 500 ng/µl
UDP-6-N$_3$-Glc (3 mM)	0.67 µl	100 µM
β-GT (40 µM)	1 µl	2 µM
Total volume	20 µl	

i) For tissue genomic DNA (high 5-hmC content > 0.1%)

Component	Volume	Final Concentration
Water	_ µl	
10 X β-GT Reaction Buffer	10 µl	1 X
Up to 20 µg genomic DNA	_ µl	Up to 500 ng/µl
UDP-6-N3-Glc (3 mM)	1.33 µl	100 µM
β-GT (40 µM)	2 µl	2 µM
Total volume	40 µl	

ii) For stem cell genomic DNA (median 5-hmC content ~0.05%)

Component	Volume	Final Concentration
Water	_ µl	
10 X β-GT Reaction Buffer	10 µl	1 X
Up to 50 µg genomic DNA	_ µl	Up to 500 ng/µl
UDP-6-N3-Glc (3 mM)	3.33 µl	100 µM
β-GT (40 µM)	5 µl	2 µM
Total volume	100 µl	

iii) For cancer cell genomic DNA (low 5-hmC content ~0.01%)

Table 1. Examples of amounts of input DNA and labeling reactions using the samples with various 5-hmC levels by the selective chemical labeling method.

Sample	5-hmC level	Starting DNA (µg)	Recovery after labeling (input to beads) (µg)	Recovery yield	Pull-down DNA (ng)	Pull-down yield
Adult mouse cerebellum	0.4%	10	7.5	75%	236	3.1%
Postnatal day 7 mouse cerebellum	0.1%	11	9	82%	140	1.6%
Mouse ES cell E14	0.05%	60	42	70%	350	0.8%

Table 2. Representative results from mouse brain tissues and ES cells.

Discussion

5-hydroxymethylcytosine (5-hmC) is a recently identified epigenetic modification present in substantial amounts in certain mammalian cell types. The method presented here is for determining the genome-wide distribution of 5-hmC. We use T4 bacteriophage β-glucosyltransferase to transfer an engineered glucose moiety containing an azide group onto the hydroxyl group of 5-hmC. The azide group can be chemically modified with biotin for detection, affinity enrichment, and sequencing of 5-hmC-containing DNA fragments in mammalian genomes. This protocol has advantages over 5-hmC antibody-based hydroxymethylated DNA immunoprecipitation (hMeDIP-Seq)[13-15], although the hMeDIP is simple, it has a strong bias towards high-density 5-hmC regions and usually gives inconsistent results from independent pull-downs. Our method would not introduce such bias.

There are, however, two possible concerns in terms of our labeling and capture method. The first concern could be the specificity of the labeling and capture. Since a recent report indicated that 5-hydroxymethyluracil (5-hmU) can also serve as a substrate of β-GT, false-positive signals might be introduced, leading to enriched fragments that are nonspecific to 5-hmC-relatedness[16]. This concern might be unfounded, though, since highly active 5-hydroxymethyluracil-DNA glycosylases are constantly removing this DNA damage, resulting in essentially undetectable 5-

hmU in mammalian genome[17,18]. The second concern is the efficiency of the labeling and the capture. Since 5-hmC levels vary significantly in different tissues and cells, ranging from 0.5% in brain tissues to 0.05% in mES cells and 0.01% in cancer cells, it is essential that our methods be applicable to all these tissues in terms of the downstream applications of the labeled and captured genomic DNA. Our experience shows that the methods described here are indeed sensitive and specific enough to analyze all these tissues for the purpose of either quantification-based comparison of 5-hmC levels among the tissues or for downstream applications, such as deep sequencing to map the distribution of the 5-hmC in the genome[19-21].

The key to our method is the use of an appropriate amount of starting genomic DNA. If the genomic DNA concentration is too low, the labelling efficiency and specificity will decrease accordingly. Based on our experience, even though the abundance of 5-hmC is high enough in DNA from brain tissues, if the concentration is lower than 25 ng/µl in 20 µl of reaction, the labelling and capture efficiency, as well as the specificity, will be significantly reduced. For the low-abundance DNA samples, in addition to the concentration requirement, the total amount of DNA is essential to get high and specific capture efficiency. Thus, although one needs only 25 ng of 5-hmC-enriched DNA fragments for the downstream standard library generation protocols employed by next-generation sequencing platforms, the starting amount of genomic DNA from brain tissues should not be less than 2 µg (and the ideal starting amount is 5-10 µg). Through trial and error, we have optimized the starting amount of genomic DNA from different tissues with variable 5-hmC abundances to get high labeling efficiency and specificity, as detailed in **Table 2**.

In summary, our method is qualified to precisely label genomic 5-hmC and specifically capture the 5-hmC-containing fragments, making it a successful protocol in terms of sensitivity and specificity for downstream next-generation sequencing assays.

Disclosures

No conflicts of interest declared.

Acknowledgements

References

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 Suppl.**, 245-254 (2003).
2. Ito, S., *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature.* **466**, 1129-1133 (2010).
3. Tahiliani, M., *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* **324**, 930-935 (2009).
4. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science.* **324**, 929-930 (2009).
5. Ko, M., *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature.* **468**, 839-843 (2010).
6. Koh, K.P., *et al.* Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell.* **8**, 200-213 (2011).
7. Iqbal, K., Jin, S.G., Pfeifer, G.P., & Szabo, P.E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America.* **108**, 3642-3647 (2011).
8. Wossidlo, M., *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241 (2011).
9. Gu, T.P., *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature.* **477**, 606-610 (2011).
10. Dawlaty, M.M., *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell.* **9**, 166-175 (2011).
11. Huang, Y., *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One.* **5**, e8888 (2010).
12. Song, C.X., *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68-72 (2011).
13. Pastor, W.A., *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* **473**, 394-397 (2011).
14. Matarese, F., Carrillo-de Santa Pau, E., & Stunnenberg, H.G. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? *Mol. Syst. Biol.* **7**, 562 (2011).
15. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
16. Terragni, J., Bitinaite, J., Zheng, Y., & Pradhan, S. Biochemical characterization of recombinant β-glucosyltransferase and analysis of global 5-hydroxymethylcytosine in unique genomes. *Biochemistry.* DOI: 10.1021/bi2014739 (2012).
17. Rusmintratip, V. & Sowers, L.C. An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14183-14187 (2000).
18. Globisch, D., *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One.* **5**, e15367 (2010).
19. Yildirim, O., *et al.* Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell.* **147**, 1498-1510 (2011).
20. Szulwach, K.E., *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
21. Szulwach, K.E., *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607-1616 (2011).