Genome-wide human brain DNA 5-hmC profiling using a novel sequence- and strand-specific method

Xueguang Sun, Adam Petterson, Tzu Hung Chung, Marc E. Van Eden, and Xi Yu Jia

Zymo Research Corporation, Irvine, CA

ZYMO RESEARCH

The Beauty of Science is to Make Things Simple

Introduction

5-Hydroxymethylcytosine (5-hmC) is an epigenetic hallmark rapidly gaining much interest within mapping and sequencing disciplines. While the precise role of 5-hmC is not fully understood, it is implicated in regulation of gene expression via active DNA demethylation pathways. Previous studies demonstrate that it plays a role in cell differentiation and carcinogenesis: Cells that are more stem- and progenitor-like have greatly reduced levels of 5-hmC compared with more differentiated cells. Similarly, tumor cells display less 5-hmC than their normal counterparts independent of either grade or stage, suggesting that global loss of 5-hmC may be an early event in carcinogenesis. Several methods have been described to profile 5-hmC at the genomic level: Most are enrichment-based via immunoprecipitation or other bioorthogonal labeling schemes, and several conversion methods have also been described that exploit selective oxidation. Here we employ a new method which combines modification-sensitive restriction enzymes with next-generation sequencing approaches to allow genome-wide 5-hmC mapping at single-site resolution in several families of carcinomas. This new method should provide a unique tool in enhancing our understanding of the interplay of genetic and epigenetic regulations in carcinogenesis.

Methodology

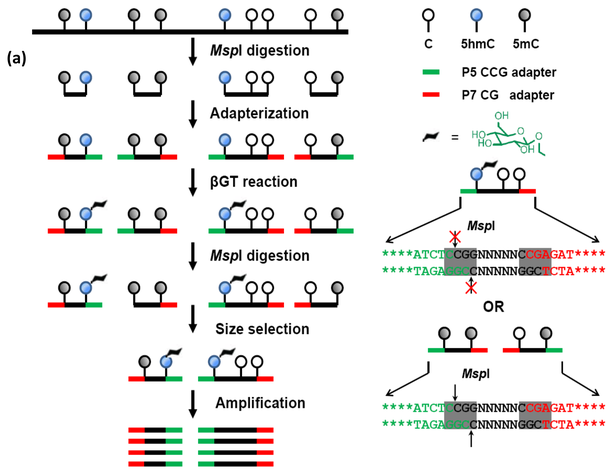

Figure 1. Overview of the Reduced Representation Hydroxymethylation Profiling (RRHP) system. (a) The assay exploits β-glucosyltransferase (β-GT) to label 5-hmC positions at adapter junctions, thus preventing digestion of the adapter from the fragment. Fragments lacking 5-hmC at the junction will not be labeled and the adapter can be digested away. Only fragments with intact adapters on both sides will be amplified for hybridization and sequencing. (b) Agarose gel visualization of amplification products from several RRHP library schemes an input amounts. (c) and (d) are Bioanalyzer traces from, respectively, positive and negative libraries which demonstrate robustly suppressed background signal.

Library amplification qualities and characteristics

In general, the output of the system is extensively dependent upon the 5-hmC content of the sample: That is, higher hydroxymethylation results in more protected junctions for adapterization, thus yielding more moles of intact library for amplification. Similarly, this yields higher final library concentrations following limited amplification.

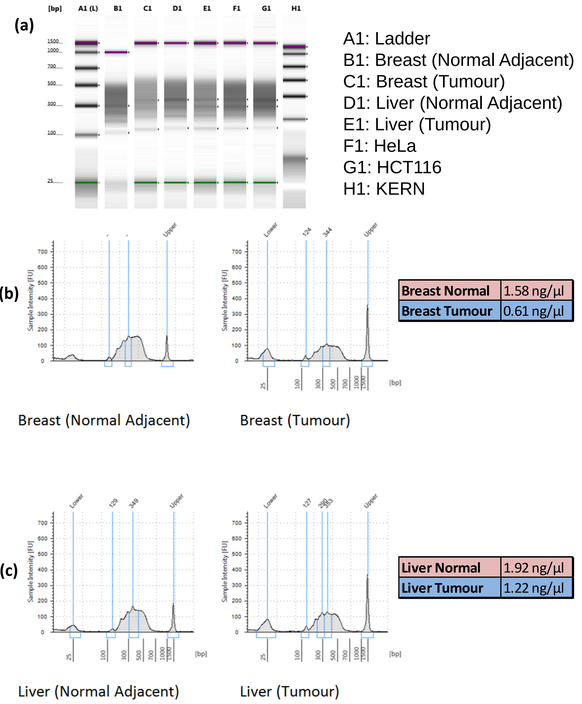

Figure 2. Genomic libraries prepared from various normal and cancerous tissues demonstrate extensive dependence upon 5-hmC content. (a) RRHP libraries were prepared from several cancerous and noncancerous tissues and subjected to limited amplification. Region traces (b) and (c) indicate the intensity, size distribution, and concentration of final libraries. Quantitative analysis demonstrates that tumour-derived libraries have lower final concentrations than libraries from normal adjacent biopsies; this indicates a depletion of 5-hmC in the cancerous tissues when compared to normal tissues. This result correlates strongly with previous findings indicating profound depletion of 5-hmC in various carcinomas.

Sequencing performance

	Simulation 40-430 bp	RRHP-HpaII (0.5 ug)	RRHP-MspI-1 (0.5 ug)	RRHP-MspI-2 (0.5 ug)	RRHP-MspI-3 (0.1 ug)	RRHP (- βgt control)	RRBS
Total reads	1,845,334	23,702,341	23,383,403	9,505,230	19,539,657	5,518	41,066,513
Mapped reads	1,845,025	20,605,538	22,271,499	9,017,016	18,482,119	4,373	15,668,523
Mappability	99.98%	86.93%	95.24%	94.86%	94.59%	79.25%	38.15%
# of tagged reads	1,845,025	17,763,501	21,081,749	8,531,903	17,328,911	3,230	NA
tagged reads %	100%	86%	95%	95%	94%	74%	NA
# of 5hmc sites	1,845,014	1,878,394	1,737,993	1,550,791	1,674,080	3,171	5,330,488

Table 1. Brief overview of several indicators of sequencing output and quality. All samples were sequenced on the Illumina HiSeq 2000 platform. Mappability for all sample preparations remains high, allowing for the identification of more unique sites even with deliberately reduced sequencing depth (RRHP-MspI-2). Samples prepared from low input (RRHP-MspI-3) demonstrate comparable mappability and unique sites to the increased inputs (RRHP-MspI-1, 2). The negative control library without glucosylation demonstrates extremely low reads, low tag rate, and the lowest number of uniquely mapped sites. Standard RRBS was prepared from the same sample for comparison.

Sequencing performance, continued

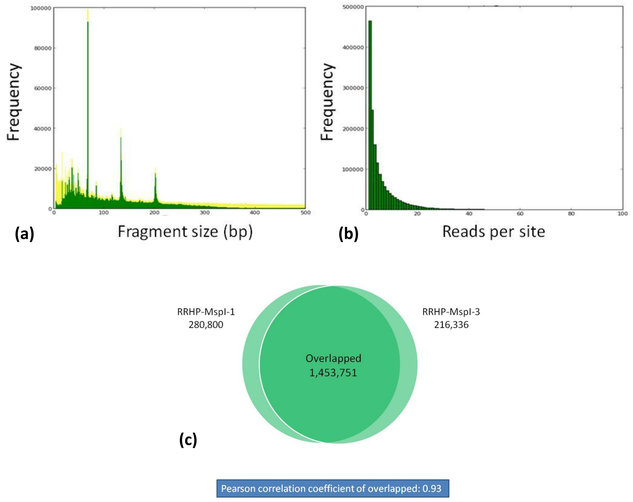

Figure 3. RRHP libraries demonstrate close correlation to *in silico* predictions as well as between one another. (**a**) The observed fragment size distributions from sequencing (green) are laid over the size distributions predicted from *in silico* digestion (yellow). The sizes mirror those observed in Bioanalyzer traces from Figure 1 (shifted up for addition of adapters). (**b**) Reads per site indicate detection of sites with few intact fragments, suggesting proportionality between 5-hmC density and generated reads. (**c**) Correlation study between high input (RRHP-MspI-1) and low input (RRHP-MspI-3) libraries indicates the ability of the assay to discover and map the same unique sites from reduced inputs.

Functional analysis of mapped 5-hmC sites

	RRHP-MspI-1 (0.5 ug)		RRHP-MspI-3 (0.1 ug)	
Total	1737993		1674080	
CpG island	171982	10%	144413	9%
Promoter	133938	8%	109537	7%
5' UTR	441174	25%	416594	25%
Coding exon	167857	10%	157064	9%
Intron	1147615	66%	1108226	66%
3' UTR	177478	10%	170934	10%
Bivalent	206811	12%	177243	11%
H3K4me3	988758	57%	762340	46%
H3K27me3	429361	25%	390551	23%
HCP	125982	7%	110205	7%
ICP	34957	2%	32966	2%
LCP	0	0%	0	0%
7X Regulatory potential	683973	39%	644236	38%

Table 2. Breakdown of 5-hmC sites profiled by RRHP into specific annotated genomic elements and binding regions. Mapped and aligned fragments from high- and low-input RRHP libraries from the same sample were analyzed per annotated functional regions of the genome. Consistent annotation rates are observed between the two libraries. Despite the high bias of the assay for CG-dense regions, only 10% of the aligned fragments are mapped back to CpG islands (CGIs), indicating appropriateness for interrogating a wide scope of the genome. Nevertheless, >88% of annotated CGIs demonstrate at least one 5-hmC position from RRHP.

5-hmC distribution in normal and tumour tissues

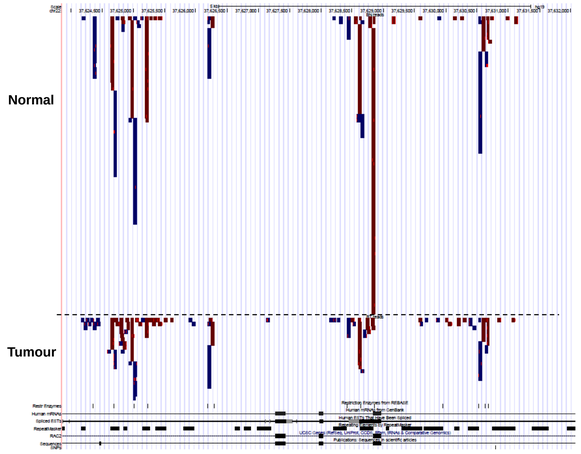

Figure 4. Significant loss of 5-hmC is observed across the *RAC2* locus in breast tumour tissue. Sequencing readout is displayed as piled read tracks proportional to the 5-hmC content at a coordinate. Reads originating from the (+) strand are indicated in blue with reads from the (-) strand in red. RAC2 is a member of the Rho GTPases shown to play a role in the regulation of mitosis.

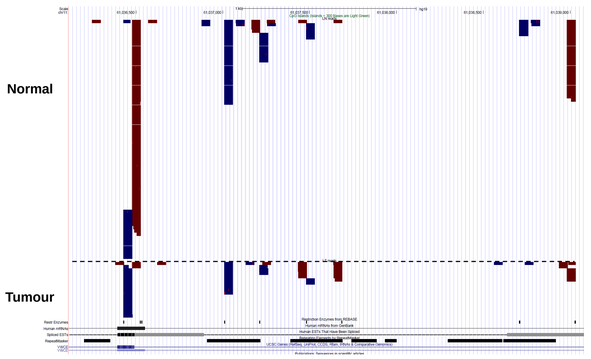

Figure 5. RRHP analysis indicates the distribution of 5-hmC in several sites over *VWCE*. Visualization of RRHP reads across a 3-kb segment of *VWCE* shows pronounced depletion of 5-hmC in introns, exons, and intron-exon junctions. VWCE serves as a regulatory element of beta-catenin signaling and has been indicated as a chemopreventative target in hepatocellular cancer therapeutics.

Conclusions

Our RRHP method offers a strong, positive-display output for the stringent identification and mapping of 5-hmC sites across the genome. The directionally unbiased library construction also allows for strand-specific information about 5-hmC localization. Importantly, the sensitivity of the method renders it exceptionally useful in genomes where the 5-hmC content is known to be severely diminished, such as primary carcinoma tissues. The ability to detect and map the mark with high resolution and high fidelity in such samples makes possible the identification of discrete and robust biomarkers for early detection, disease stratification, and prediction of treatment susceptibility across several cancer pathologies.

References

Jin *et al*. *Cancer Research* 71 (24): 7360-5.

Song *et al*. *Nat. Biotechnol*. 29: 68-72.

Ko *et al*. *Nature* 468: 839-43.

Lian *et al*. *Neoplasia* 5 (3): 229-44.