**Featured hypomethylation of regions surrounding HBV integration sites revealed by low pass whole genome bisulfite sequencing (WGBS) of plasma cell free DNA from patients with liver diseases**

**Abstract**

DNA methylation changes have been proved a promising direction for noninvasive detection to facilitate earlier cancer detection. Sequencing volume optimization may help to broaden the clinical application of whole genome bisulfite sequencing (WGBS). Here we analyzed the effects of reducing of sequencing read pairs on methylation detection for patients with liver diseases using cell free DNA samples, and showed 5M mappable read pairs were efficient for this kind of analysis. Patients with hepatitis and cirrhosis had significant fewer hypomethatlated genomic regions comparing to HCC patients; nevertheless, they shared certain differentially methylated CpGs (DMCs) with HCC patients. DNA sequences characterized by low pass WGBS mainly came from genomic repeat regions, which also are hotspots for HBV integration sites. Genomic regions around HBV integration sites are also suitable surrogate for methylation level estimation in plasma cfDNA analysis of liver diseases.

**Introduction**

Circulating cell-free DNA (cfDNA) are small double-stranded DNA fragments(1) in plasma, unrine, and other body fluids(2), deriving from cell apoptosis and necrosis(3). Analysis of cfDNA could be regard as a way to perform “liquid biopsy”, which have demonstrated the potential not only for cancer diagnosis and prognosis(4, 5), but also for finding patients with premalignant states, inflammatory or traumas(1). Apoptotic and necrotic tumor cells can release cfDNA into the blood circulation, which can reflect the tumor-related genetic features, including mutations, copy number aberrations and epigenetic changes(4), representing important biomarkers of tumor development.

Hepatocellular carcinoma (HCC) is the third leading cause of cancer death, accounting for 782,000 cases diagnosed and 746,000 deaths in 2012(6). Hepatitis B or C virus infection, cigarette smoking, heavy alcohol consumption, toxicant exposure and other risk factors are involved in the etiology of hepatocellular carcinoma(7, 8). Hepatocarcinogenesis is a complicated, multistep process that involves sustained inflammatory damage, including hepatocytes necrosis and regeneration, associated with fibrotic deposition(6). Accumulation of genetic or epigenetic alterations might transform normal hepatocytes into malignant clones, which promote hepatocarcinogenesis(9, 10). Chronic hepatitis and cirrhosis are recognized as conditions predisposing of developing hepatocellular carcinoma(11, 12), initiating transformation and the development of hepatocellular carcinoma(9). Therefore, biomarkers for tumor early detection have attracted a lot of interests in the research area of chronic conditions predisposing to cancer.

DNA methylation alternation plays a crucial role on tumorigenesis and progression, including the global DNA hypomethylation and the promoter hypermethylation of tumor suppressor genes(13), and is an ideal target for cancer diagnosis(14). A number of researches have focused on cancer diagnosis of cfDNA methylation, such as the potential specific biomarkers(15), pervasive hypomethylation(16) and the tissue origin(17-19). Previous attempts using WGBS for cancer diagnosis require deep sequencing(17, 19), which may limit their wide application. Reducing the sequencing volume will influence the cost consideration in clinical decision making. Chan et al.(16) came up with a genome-wide pervasive hypomethylation percentage method for hepatocellular carcinoma detection and could be detected at low sequencing depth (10 million reads), providing us a clue on the low-pass WGBS cancer diagnosis. Further attempts for sequencing volume optimization and more patients from multiple stages predisposing to HCC involved will contribute to improve its feasibility in clinical application.

In this study, we estimated the optimal sequencing depth of low-pass whole genome bisulfite sequencing (WGBS) in cfDNA methylation profiling and enrolled patients at different stages of liver diseases. We found low-pass WGBS at 5M sequencing read pairs were able to detect the hypomethylation level of plasma cfDNA from patients with liver diseases. We also analyzed the genomic features of sequences represented in low-pass WGBS and examined if some regions are more suitable for methylation evaluation in cfDNA analysis.

**Results**

**The efficacy of low pass whole genome bisulfite sequencing (WGBS) strategy illustrated by re-sampling reads from medium WGBS data**

In pilot medium WGBS for cfDNA, we analyzed 5 samples from 1 healthy individual, 1 hepatitis patient, 1 cirrhosis patient and 2 HCC patients (1 before surgery and 1 after surgery), with a final read count at the mean of 58 million (M) read pairs (Supplementary Table 1). We divided the genome into 1-Mb regions and calculated the average methylation level for each region (Methyl1MB). The Methyl1MB for all 2734 autosomal regions were used for global methylation level calculation (Method). To determine the effective sequencing depth in low pass WGBS of cfDNA, we randomly sampled 1M to 10M mappable read pairs from each sequencing dataset of about 58M read pairs respectively and calculated Methyl1MB. During each extraction, we calculated Methyl1MB forall 1-Mb regions, and adopted correlation coefficient to show their consistency with those based on total sequencing reads. For each sequencing depth, we repeated the random extraction for 100 times to examine the variation of correlation coefficient, and the difference (coefficient of variation, CV) among 100 values of correlation coefficient demonstrated sampling bias. The final results showed this CV was only around 4% (Fig 1). With sequencing reads increasing, the Methyl1MB was closer to the value calculated using total sequencing reads (Fig 1).When sequencing depth increased from 1M to 10M read pairs, the correlation coefficient stopped obvious increasing when using 5M or more. High correlation of Methyl1MB was observed between 5M reads and all sequencing reads (R > 0.84, P < 2.2e-16, Pearson’s correlation test, Figure S1), and methylation level remained consistent during 100 resampling processes (coefficient of variation is 3.7%, 0.6%, 0.6%, 0.8%, 0.6% for H1, P1, P2, P3 and P4, respectively, Fig 1).Therefore, we required 5M mappable read pairs without redundancy in low pass WGBS, which we believed were reliable to evaluate methylation level of cfDNA samples.

**The methylation level of plasma cfDNA from hepatitis and cirrhosis patients resemble healthy individuals**

Next, we applied low pass WGBS for plasma cfDNA samples from 54 individuals, including 17 HCC (4 early stage HCC, 4 advanced HCC and 9 HCC after surgery), 17 with hepatitis, 17 with cirrhosis and 3 healthy volunteers (Supplementary Table 2). On average, each sample obtained 10.2M mappable read pairs (Supplementary Table 3). For methylation level evaluation, we adopted the strategy developed by Chan et al (16) to define the hyper- or hypo-methylated 1-Mb regions (Method), with Methyl1MB in healthy individuals as the baseline level. The percentage of hyper- or hypo-methylated 1-Mb regions was shown for each patient (Fig 2; Supplementary Table 3).

Hepatitis and cirrhosis patients had cfDNA methylation levels similar to healthy individuals (Fig 2). Moreover, they had fewer either hyper- or hypo-methylated 1-Mb regions. Specifically, hyper ones only accounted for less than 3% (Fig 2A); meanwhile, hypo ones accounted for 0 to 21.8%, with only three patients with the percentage over 10% (Fig 2B; Supplementary Table 3). In early stage HCC patients, hyper-methylated 1-Mb regions were even fewer, less than 0.1%, but hypo-methylated ones accounted for 2.3% to 58%. In advanced HCC patients, no hyper-methylated 1-Mb regions were identified, and hypo-methylated regions accounted for more than 60%. (Fig 2; Supplementary Table 3). As expected, most HCC patients (8/9) after surgery demonstrated similar cfDNA methylation level to healthy individuals and patients with hepatitis or cirrhosis. Nevertheless, one out of nine HCC patients after surgery has higher proportion of hypo-methylated regions (63.5%, Fig 3B; Supplementary Table 3), indicating tumor existence.

**Differentially methylation CpGs (DMCs) and genes**

On average, each dataset for corresponding cfDNA sample had 61,018 CpGs with sequencing depth over 5 (Method, Supplementary Table 3). Totally, advanced HCC patients had 1841 DMCsidentified (Supplementary Table 4), of which all the DMCs were hypo-methylated comparing to healthy individuals. Among those, 23 DMCs located in gene body of six genes: *HFM1, PMF1, PMF1-BGLAP, SENP5, SLCO5A1, REXO1L1P*. In the four early stage HCC patients, we identified 193 DMCs (Supplementary Table 5), of which 155 were in common with those observed in advanced HCC patients and 8 located within *PMF1* and *PMF1-BGLAP* (Fig 3A; Table 1).

Relative high proportions of hypo-methylated 1-Mb regions (>10%) were observed in one hepatitis and two cirrhosis patients (Fig 2B), indicating their high HCC risk. Totally, all four groups had 134 DMCs in common (Fig 3A), which suggested that methylation changes may occur in early stage of liver disease progression prior to HCC. Table 1 displayed the promoter or the gene body with DMCs in the four comparisons. Particularly, *SENP5* gene had 7 significantly hypo-methylated DMCs with consistently high sequencing coverage across all individuals (149 reads, on average, Supplementary Fig 2, and Fig 3B). Intriguingly, all 7 DMCs that we found in intron 2 of *SENP5* were located near previously reported HBV integration sites in HCC (Fig 3C) (20).

**Over representing of DMCs in repeat regions and surrounding HBV integration sites**

Genome feature distribution of CpGs illustrated that they tended to locate at intergenic region and repeat regions (Methods, Supplementary Fig 3A), and CpGs in repeat regions had much higher sequencing depth in this low pass sequencing strategy than those in other regions (P < 2.2e-16, Wilcoxon rank sum test; Supplementary Fig 3B). On average, 64% of all these CpGs were in the repeat regions (Supplementary Fig 3C), and this percentage varied from 49% to 87% among individuals. Differential methylation analysis required the CpG sites having sequencing reads over 5 in all individual (Method), and qualified CpG were over represented in repeat regions. Finally, 91% of DMCs of advanced HCC patients located within repeat regions (Fig 4A).

Considering repeat region is a known feature of HBV integration sites (21, 22), we subsequently analyzed the location of DMCs relative to reported HBV integration sites (20, 23-28). Among 1841 DMCs in advanced HCC patients, twenty completely overlapped with the HBV integration sites, including two in *SENP5* (Supplementary Table 8). Meanwhile, 36% of DMCs located within 100bp region either upstream or downstream of integration sites, and 95.1% of DMCs within 5K regions (Fig 4A). Overall, these DMCs were more significantly enriched in HBV integration sites compared with promoter and gene body regions (Fig 4B).

Then, we wondered if methylation levels of CpGs near HBV integration sites could mirror the hypo-methylation statutes of cfDNA from HCC patients. We first analyzed CpGs with depth over 5 in all the 54 samples that within the 100 bp upstream or downstream of HBV integration sites and calculated the percentage of hypo-CpGs. These CpGs were significantly hypo-methylated in advanced HCC patients, with 53% to 59% of CpGs were hypo (Fig 4C; Supplementary Table 3). Then, we incorporated all the CpGs in each samples and calculated the average methylation level of the CpGs within the 100bp of the reported HBV integration sites (Supplementary Table 2), and the advanced HCC patients still showed significantly hypo-methylation level (<48.4%; Fig4D; Supplementary Table 3). However, for early stage HCC patients, this value was relatively higher, from 66.5% to 71%. We adopted the median of early stage HCC patients (68.83%) as a cutoff for the detection. Moreover, four patients with hepatitis or cirrhosis (P2, P14, P18, P19) showed slightly hypo-methylated with these sites. One hepatitis patient, P14 had the average methylation level at 67.4%, with abnormal AFP level (141.9 ng/ml; Table 2) and this patient was diagnosed with HCC later. Patient P19 was diagnosed as alcoholic cirrhosis, with the transaminase indicator and jaundice mildly abnormal at each of the follow-up time point. For patient P2 with chronic hepatitis and P18 with nash-related cirrhosis, both of them had no abnormal measurements observed and haven’t had tumor detected yet.

**Discussion**

Although WGBS of cell-free DNA has been proved effective for cancer detection (16, 19, 29), the cost of cfDNA WGBS in cancer patients is one of challenges for wide application. We explored the cfDNA methylome of hepatitis, cirrhosis and HCC patients and examined the feasibility of HCC detection using low-pass WGBS. Although we required 5M qualified read pairs for low pass WGBS for 54 samples, there were 2 samples only having 3.6M read pairs (Supplementary Table 3). The average correlation coefficient of 100 re-sampling were larger than 0.8 using 3M read pairs (Fig 1), and theoretically they were sufficient to evaluate the methylation level.

Previous studies have been shown that the fragmentation process of cell free DNA is not random (30, 31).Our results showed low-pass WGBS for cfDNA tended to capture fragments from repeat regions and HBV integration sites. More than 49% of CpGs located in the repeat regions and had a higher sequencing depth. When decreasing the sequencing volume, over representing of genomic repeat regions was observed in our data. This suggested that the repeat regions could remain adequate sequencing depth in low pass WGBS. Since HBV integrations tend to locate at repeat regions, DMCs of advanced HCC patients were also enriched in previously reported HBV integration sites. Notably, CpGs near the HBV integration sites were likely to have methylation levels reflecting hypo-methyation status of tumor genome in HCC patients, which can discern the HCC patients from patients in other stages of liver diseases.

Here we adopted 100 bp upstream and downstream HBV integration sites as surrogate regions for plasma hypomethylation analysis in HCC patients. Although we chose HBV integration sites as the indicator, it did not mean the analysis was only suitable for patients with HBV infection. In our patients we also included 3 patients without HBV infection (P1, P18 and P19; Supplementary Table 2). This kind of regions may have some biological features suitable for HBV integrations, and here we also demonstrated their methylation changes may be common in HCC development independent of HBV infection. Chan et al. has proposed the proportion of hypo methylation 1-Mb regions as the HCC indicator in cfDNA methylation analysis (16). According to their method, the median of first stage HCC patients is 17%. Indeed, the percentage of hypomethylated regions for one hepatitis patient in our study, P14, was 6.4%, far below the median cutoff of HCC indication (Table 2; Supplementary Table 3). However, according to our indicator, average methylation level of regions around known HBV integration sites, the value was 67.4% for P14, only slightly lower than the up limit median cutoff for HCC. We followed up this patient and found him diagnosed as HCC within half a year. Another hepatitis patient, P2, the proportion of hypo methylation 1-Mb regions was 19%, and the average methylation level around HBV integration sites was 67.7%; while in the re-visit after half a year, the former value became only 3% and the latter one was 69%, seemingly free of HCC risk. For a HCC indicator, the most challenging part is to determine cutoffs for candidate diseases, which need a large sample size in further study. Nevertheless, our study successfully illustrated it is necessary to monitor the patients with suspicious methylation changes in cfDNA according to multiple indicators to combine their powers together.

One limit of cfDNA detection is the application of early stage cancer, our attempt showed that the sensitivity of early stage HCC detection was much lower than advanced HCC. For early stage HCC patients, P35 and P36, both the proportion of hypomethylation (2.38% and 6.37%) and the average methylation level around HBV integration sites (70.48% and 71.48%) were similar to the healthy individuals and hepatitis patients. Both of these two patients had small tumor size (P35, 1.5cm; P36, less than 2.5cm, multiple; Supplementary Table 2).Another application of cfDNA methylation analysis is to evaluate the residual tumor or tumor recurrence after surgery. After tumor removal, the methylation level of cfDNA should resemble with those from healthy individuals or patients with other chronic liver diseases. We found one HCC patient after interventional therapy, P45, showed significantly hypomethylation of the CpGs near the HBV integration sites (60.87%), indicating the tumor existence and he died two months later in follow-up due to multiple and recurrent lesion.

Although we have found some stable pattern at low-pass WGBS, we still need to enlarge sample size to validate the sensitivity and specificity of this pattern to obtain more precise information for HCC diagnosis and surveillance. The most challenging part in future attempt is to set the cutoff for disease stage definition, and especially for those with small tumors. Furthermore, blood samples from HCC patients at multiple time points were also needed to trace the disease progression.

**Materials and Methods**

**Sample collection**

All the blood samples of patients were collected from Beijing You’an Hospital. Healthy individuals enrolled by Beijing Institute of Genomics were collected as controls. The diagnosis was made according to the guidelines for the prevention and treatment of chronic hepatitis B: a 2015 update (32). The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Ethics Committee of Beijing You’an Hospital and Beijing Institute of Genomics. An informed consent was obtained from all patients and volunteers.

**Cell free DNA extraction and whole genome bisulfite sequencing**

Ten microliters (ml) of whole blood was collected from each patient in Streck Cell-Free DNA BCT® tubes (Streck, Omaha, NE) and immediately transmitted to Beijing Institute of Genomics. Upon arrival, the blood collected in Streck BCT tubes were centrifuged at 3,000 × g for 15 minutes at 4°C within two hours. Subsequently, the plasma was transferred into a fresh microcentrifuge tube, followed by a 2nd centrifugation at 16,000 × g for 10 minutes at room temperature. Five ml of resultant plasma was used for cfDNA extraction using a QIAamp Circulating Nucleic Acid Kit (Qiagen, Valencia, CA). After extraction, total DNA was quantified using a Qubit dsDNAHS Assay kit (Life technologies, Grand Island, NY, USA). All DNA samples were stored at -80°C before sequencing library construction.

**Whole genome bisulfite sequencing**

Using the TruSeq DNA Methylation Kit (Illumina Inc.) according to the manufacturers’ protocol. XX ng cfDNA was used for sequencing library construction. Bisulfite conversion of cfDNA was performed using the EZ DNA Methylation-Gold Kit (Zymo Research) according to the instruction manual. During conversion, 0.5% methylated lambda DNA was included as a spike-in DNA control to estimate the conversion efficiency of unmodified cytosine. The sequencing libraries were then performed paired end sequencing (2 × 100 bp) on an Illumina HiSeq 4000 (Illumina Inc., San Diego, CA, USA). All the sequencing data was deposited in the BIG Data Center (http://bigd.big.ac.cn/bioproject/) under the BioProject accession code XXXXXXXX.

**WGBS data processing**

After base calling, all paired-end fastq files were trimmed using cutadapt (v 1.8.3)(33) to removed adapter sequences and low quality bases with parameters ‘-q 15 --minimum-length 36’. HG19 reference genome was downloaded from ENSEMBL. Lambda genome was also included in the reference sequence for calculating bisulfite conversion rate. Filtered paired-end bisulfite sequencing data were mapped with Bismark (v0.14.5)(34) using with default parameters. After alignment, read duplicates were removed using the deduplicate\_bismark application included in the bismark software. Then the BAM files produced by Bismark were sorted using samtools (v 0.1.19) and overlapping paired-end reads were clipped using ClipOverlap function of bamUtil (<https://github.com/statgen/bamUtil>) to prevent counting twice from the same observation. For each CpG, the methylation level was combined from both DNA strands and estimated as mi/(mi + ui), where mi was defined as the number of methylated cytosines and ui was defined as the number of unmethylated cytosines. The number of methylated and unmethylated cytosines of 1-Mb regions were generated using R package methylKit. The average methylation level of each 1-Mb region (Methyl1MB) was calculated as the total number of cytosines divided by the number of methylated cytosines.

**Randomly re-sampling lower reads from medium WGBS data**

A randomly sampling method was used to obtain low depth WGBS for 5 medium WGBS of cell-free DNA. (a) 1 M to 10 M read pairs (increasing by 1 M step) was randomly extracted from each medium WGBS data set. (b) For each re-sampling, the average methylation level for each 1-Mb region (Methyl1MB) from autosome were calculated and pearson correlation coefficient was used to show the correlation of all the autosomal Methyl1MB between this re-sampling reads and total WGBS reads. This process was repeated for 100 times. (c) For each re-sampling, coefficient of variation (CV) for correlation coefficient was calculated across 100 randomly re-sampling to examine the variability of 100 extraction.

**Identification of hyper- or hypo-Methyl1MB**

We adopted the method of Chan et al.(16) to define the hyper- or hypo-Methyl1MB compared to the healthy ref group. Only autosomes were included in this analysis. A 1-Mb region of a sample was defined as hyper- or hypo-methylated if its average methylation level was at least 3 SDs above or below the mean of the corresponding region of the healthy individuals. Then the number and percentage of hyper- or hypo-Methyl1MB within the genome was calculated.

**Identification and annotation of the differentially methylated CpGs (DMCs)**

The identification of DMCs was generated using R package methylKit. The significance of the DMCs between case (hepatitis, cirrhosis, early stage HCC, advanced HCC) group and healthy group was performed by logistic regression test with at least 5-fold coverage. P-value was adjusted for multiple testing with the method of Hochberg and Benjamini. The CpG sites were considered differentially, if the Benjamini Hochberg corrected P value ≤ 0.05 and the methylation level difference was ≥ 0.2. Each DMCs was annotated for each RefSeq transcript obtained from ENSEMBL GRCh37. Promoters are defined as regions 2kb upstream from TSS for each RefSeq transcript. RepeatMasker annotations were obtained from UCSC.

**The enrichment score in each genomic region**

The enrichment score for CpGs or DMCs was calculated by the following formula:

The enrichment scorein the genomic element = log2 (# DMCsin the genomic element/# expected). # expected was computed as: # DMCsin the genome × # CpG sitesin the genomic element/# total CpG sitesin the genome. # means the number of sites.

**The analysis of CpGs near the HBV integration sites**

*Identification of hypo-CpGs within the 100 bp upstream or downstream of HBV integration sites.* The HBV integration sites were extracted from previous reports(20, 23-28).We extracted CpG within the 100 bp upstream or downstream of HBV integration sites. Only autosomal CpGs and CpGs with depth over 5 in all the 54 samples were included in the hypo-CpGs analysis. Similar to the identification of hypo-Methyl1MB, a CpG of a sample was defined as hypo-methylated if its methylation level was 3 SDs or more below the mean of the corresponding CpGs of the healthy individuals. Then the percentage of hypo-CpGs was calculated. *Average methylation level of the CpGs within the 100bp of the HBV integration sites.* For each sample, the average methylation level of CpGs within the 100 bp upstream or downstream of HBV integration sites was included in all the CpGs. This value was calculated as the number of the total number of cytosines divided by the number of methylated cytosines within the 100bp of the HBV integration sites.

**Reference**

1. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer - A survey. Biochimica Et Biophysica Acta-Reviews on Cancer 2007;1775:181-232.

2. Chan AK, Chiu RW, Lo YM, Clinical Sciences Reviews Committee of the Association of Clinical B. Cell-free nucleic acids in plasma, serum and urine: a new tool in molecular diagnosis. Ann Clin Biochem 2003;40:122-130.

3. Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, Rossier A, et al. The origin and mechanism of circulating DNA. Ann N Y Acad Sci 2000;906:161-168.

4. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nature Reviews Cancer 2011;11:426-437.

5. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017;17:223-238.

6. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. The Lancet 2018;391:1301-1314.

7. Chen CJ, Yu MW, Liaw YF. Epidemiological characteristics and risk factors of hepatocellular carcinoma. J Gastroenterol Hepatol 1997;12:S294-308.

8. Montesano R, Hainaut P, Wild CP. Hepatocellular carcinoma: from gene to public health. J Natl Cancer Inst 1997;89:1844-1851.

9. Stauffer JK, Scarzello AJ, Jiang Q, Wiltrout RH. Chronic inflammation, immune escape, and oncogenesis in the liver: a unique neighborhood for novel intersections. Hepatology 2012;56:1567-1574.

10. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. Nat Rev Cancer 2006;6:674-687.

11. Aihara T, Noguchi S, Sasaki Y, Nakano H, Imaoka S. Clonal analysis of regenerative nodules in hepatitis C virus-induced liver cirrhosis. Gastroenterology 1994;107:1805-1811.

12. Schutte K, Bornschein J, Malfertheiner P. Hepatocellular carcinoma--epidemiological trends and risk factors. Dig Dis 2009;27:80-92.

13. Kulis M, Esteller M. DNA methylation and cancer. Adv Genet 2010;70:27-56.

14. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet 2012;13:679-692.

15. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, Yi S, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater 2017;16:1155-1161.

16. Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, Chan SL, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci U S A 2013;110:18761-18768.

17. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, Grimes B, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 2017;18:53.

18. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, Rubertsson S, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A 2016;113:E1826-1834.

19. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, Chan WK, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A 2015;112:E5503-5512.

20. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765-769.

21. Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. Viruses 2017;9.

22. Yan H, Yang Y, Zhang L, Tang G, Wang Y, Xue G, Zhou W, et al. Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma. Hepatology 2015;61:1821-1831.

23. Jiang S, Yang Z, Li W, Li X, Wang Y, Zhang J, Xu C, et al. Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis. PLoS One 2012;7:e40363.

24. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012;44:760-764.

25. Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, Lee W, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res 2012;22:593-601.

26. Ding D, Lou X, Hua D, Yu W, Li L, Wang J, Gao F, et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. PLoS Genet 2012;8:e1003065.

27. Li W, Zeng X, Lee NP, Liu X, Chen S, Guo B, Yi S, et al. HIVID: an efficient method to detect HBV integration using low coverage sequencing. Genomics 2013;102:338-344.

28. Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, Ronaghi M, et al. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. Carcinogenesis 2013;34:787-798.

29. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 2017;49:635-642.

30. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, Wong J, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc Natl Acad Sci U S A 2018.

31. Chan KC, Jiang P, Sun K, Cheng YK, Tong YK, Cheng SH, Wong AI, et al. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. Proc Natl Acad Sci U S A 2016;113:E8159-E8168.

32. Hou J, Wang G, Wang F, Cheng J, Ren H, Zhuang H, Sun J, et al. Guideline of Prevention and Treatment for Chronic Hepatitis B (2015 Update). J Clin Transl Hepatol 2017;5:297-318.

33. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 2011;17:3.

34. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27:1571-1572.