Detection of Hepatitis B Virus–Host Junction Sequences in Urine of Infected Patients

Selena Y. Lin,1 Yih-Ping Su,2 Evan R. Trauger,1 Benjamin P. Song,1 Emilie G.C. Thompson,2 Malcolm C. Hoffman,1 Ting-Tsung Chang,3 Yih-Jyh Lin,4 Yu-Lan Kao,2 Yixiao Cui,2 Hie-Won Hann,3 Grace Park,5 Fwu-Shan Shieh,1 Wei Song,1 and Ying-Hsiu Su2

Integrated hepatitis B virus (HBV) DNA, found in more than 85% of HBV-associated hepatocellular carcinomas (HBV-HCCs), can play a significant role in HBV-related liver disease progression. HBV-host junction sequences (HBV-JSs), created through integration events, have been used to determine HBV-HCC clonality. Here, we investigate the feasibility of analyzing HBV integration in a noninvasive urine liquid biopsy. Using an HBV-targeted next-generation sequencing (NGS) assay, we first identified HBV-JSs in eight HBV-HCC tissues and designed short-amplicon junction-specific polymerase chain reaction assays to detect HBV-JSs in matched urine. We detected and validated tissue-derived junctions in five of eight matched urine samples. Next, we screened 32 urine samples collected from 25 patients infected with HBV (5 with hepatitis, 10 with cirrhosis, 4 with HCC, and 6 post-HCC). Encouragingly, all 32 urine samples contained HBV-JSs detectable by HBV-targeted NGS. Of the 712 total HBV-JSs detected in urine, 351 were in gene-coding regions, 11 of which, including TERT (telomerase reverse transcriptase), had previously been reported as recurrent integration sites in HCC tissue and were found only in the urine patients with cirrhosis or HCC. The integration breakpoints of HBV DNA detected in urine were found predominantly (~70%) at a previously identified integration hotspot, HBV DR1-2 (down-regulator of transcription 1-2). Conclusion: HBV viral–host junction DNA can be detected in urine of patients infected with HBV. This study demonstrates the potential for a noninvasive urine liquid biopsy of integrated HBV DNA to monitor patients infected with HBV for HBV-associated liver diseases and the efficacy of antiviral therapy. (Hepatology Communications 2021;5:1649-1659).
transcriptase], MLL4 [mixed-lineage leukemia-like protein], and CCNE1 [cyclin E1]) has been reported repeatedly in patients with HCC.\(^{(9)}\)

One outcome of HBV-DNA integration is the creation of a unique HBV-host junction sequence (HBV-JS), representing a specific molecular signature of the infected cell. HBV-DNA integrations detected in tissue have been used to determine the clonality of recurrent HCC tumors.\(^{(10,11)}\) Recent studies have shown that integrated HBV DNA can be detected in the circulation as well.\(^{(12-15)}\) We have previously demonstrated that urine of patients with HCC contains cell-free DNA (cfDNA) with HCC-associated modifications,\(^{(16-19)}\) whereas urine of patients infected with HBV contains fragmented (but not intact) HBV DNA.\(^{(20)}\) We hypothesize that urine can serve as a noninvasive liquid biopsy for the investigation of HBV integration in patients infected with HBV.

In this proof-of-concept study, we assessed the feasibility of detecting HBV-JSs in urine from patients infected with HBV. We first identified HBV-JSs in HCC tissue, followed by detection in matched urine by sensitive junction-specific polymerase chain reaction (PCR) assays. Next, we detected HBV-JSs in urine from patients infected with HBV with hepatitis, cirrhosis, and HCC, including those undergoing recurrence monitoring (post-HCC), using an HBV-targeted next-generation sequencing (NGS) assay. These findings support the potential of urine as a noninvasive liquid biopsy for integrated HBV DNA, which can be used to monitor the dynamics of HBV integration in the infected liver and to assess HBV treatment efficacy and disease progression.

**Patients and Methods**

**STUDY SUBJECTS AND SPECIMENS**

Two studies, A and B, were performed. Study A specimens included archived DNA isolated from eight pairs of matched HBV-HCC formalin-fixed paraffin-embedded tumor tissue and urine samples. The specimens were obtained from patients in stages I-II undergoing surgery (Table 1) at the National Cheng-Kung University Hospital, Taiwan, as described previously.\(^{(17,19)}\) Study B was performed on urine samples collected from 25 patients infected with HBV, including 5 with hepatitis, 10 with cirrhosis, 4 with HCC, and 6 post-HCC, at the Thomas Jefferson University Hospital (Table 2). Five patients (B7, B15, B17, B19, and B20) provided more than one urine collection over the course of the study, resulting in a total of 32 urine samples. Additionally, urine samples from 8 HBV-negative donors (4 male, 4 female), aged 23-60 years, were collected and served as controls. All samples were collected with written, informed consent and in accordance with the guidelines of the institutional review board.

**URINE COLLECTION AND DNA ISOLATION**

Urine samples were collected and total urine DNA isolated, as previously described.\(^{(21)}\) For study A, cfDNA was obtained by isolating the <1-kb fraction from total urine DNA using carboxylated magnetic beads, a method previously developed by our laboratory.\(^{(22)}\) For study B, the JBS urine cfDNA isolation kit (JBS Science Inc., Doylestown, PA) was used according to the manufacturer’s specifications.
HBV-JS DETECTION BY HBV-TARGETED NGS ASSAY AND VALIDATION BY PCR-SANGER SEQUENCING

Tissue DNA was fragmented by sonication and subjected to Illumina (San Diego, CA) NGS library preparation, as previously described. (23) Ten cycles of library amplification with Herculase II Fusion polymerase were performed (Agilent Technologies, Santa Clara, CA). Urine cfDNA samples underwent NGS library preparation using the ThruPLEX Tag-Seq kit (Cat# R400585; Takara Bio, Mountain View, CA), which contains unique molecular tags (UMTs) and uses double-stranded DNA (dsDNA) as a substrate, according to manufacturer specifications. Of note, the TruPLEX Tag-Seq kit converted dsDNA at least 300 times more efficiently than single-stranded DNA (ssDNA) when control templates of 131-nucleotide (nt) ssDNA and 140–base pair (bp) dsDNA were used to validate the performance of the kit. Library DNA was subjected to an HBV-targeted NGS assay following the manufacturer’s instructions (JBS Science). The captured library DNA was pooled for NGS analysis on the Illumina MiSeq platform (Penn State Hershey Genomics Sciences Facility at Penn State College of Medicine, Hershey, PA) or on the NovaSeq platform (Psomagen Inc., Rockville, MD).

The ChimericSeq analysis software we developed previously (24) was used to detect HBV-JSs in NGS data sets. For NGS libraries containing UMTs, UMT family consolidation and consensus read generation were performed using the software package Connor. (25) Briefly, UMT-containing reads were aligned to human and HBV reference genomes, and paired reads were grouped into families based on the mapping coordinates and combined UMT sequence. For each UMT family, a consensus read pair was generated. The consensus reads were then processed by ChimericSeq for HBV-JS identification.

To validate junctions identified by NGS, junction-specific primers were designed for the host and viral sequences for tissue and urine (Supporting Table S1). Primers used for generating PCR products larger than 100 bp for Sanger sequencing were only used for tissue junction validation, whereas urine primers were designed to target the same junction sequence generating amplicons of 65 bp or less. The junction-specific PCR was performed with Hotstart Plus Taq Polymerase (Qiagen, Valencia, CA). PCR products were visualized by agarose gel electrophoresis on a 2.2% FlashGel DNA Cassette (Lonza Group, Basel, Switzerland) and subsequently subjected to Sanger sequencing, a nested PCR reaction, or restriction endonuclease (RE) digestion using enzymes obtained from New England Biolabs (Ipswich, MA), according to the manufacturer’s specifications.

To control for artifacts generated by the HBV-targeted NGS assay, which may be unavoidable with short PCR templates, an artifact library was generated using as input a mix of 30 ng of sonicated (predominantly 170–200-bp fragments) normal human genomic DNA and 9,000 copies of HBV plasmid DNA. The mix contained 10–100 times more copies of HBV DNA than the >500 urine cfDNA samples examined (including previously described samples (20)), in which HBV-DNA content ranged from undetectable to hundreds of copies per milliliter of urine (as

| Patient ID | Age (Years) | Sex (M/F) | Cirrhosis | Tumor stage* | Tumor Size (cm) |
|------------|-------------|-----------|-----------|--------------|-----------------|
| A1         | 57          | M         | -         | 1            | 4.0             |
| A2         | 29          | M         | +         | 2            | 7.0             |
| A3         | 42          | M         | +         | 1            | 2.0             |
| A4         | 49          | M         | +         | 2            | 3.4             |
| A5         | 61          | M         | -         | 2            | 2.3             |
| A6         | 75          | F         | -         | 1            | 3.0             |
| A7         | 47          | M         | -         | 1            | 4.5             |
| A8         | 39          | F         | -         | 2            | 10.0            |

*HCC tumors were staged using the tumor-node-metastasis staging system. Abbreviations: F, female; M, male.
| Diagnosis        | Patient ID | Age (Years) | Sex (M/F) | ALT (IU/L) | Serum AFP (ng/mL) | HBsAg+ | HBV Serum Viral Load (IU/mL) | HBeAg | HBeAb | Tumor Stage | Tumor Size (cm) |
|------------------|------------|-------------|-----------|------------|-------------------|--------|-----------------------------|--------|--------|-------------|----------------|
| Hepatitis (n = 5)| B1         | 59          | M         | 8          | 1.2               | +      | <20                         | -      | -      | N/A         |                |
|                  | B2         | 56          | M         | 58         | 3.3               | +      | 105                         | -      | +      |             |                |
|                  | B3         | 58          | M         | 24         | 1.5               | +      | <20                         | -      | +      |             |                |
|                  | B4         | 56          | M         | 27         | 1.0               | +      | <20                         | -      | +      |             |                |
|                  | B5         | 53          | M         | 77         | 3.9               | +      | <20                         | -      | +      |             |                |
| Cirrhosis (n = 10)| B6        | 60          | M         | 40         | 2.8               | +      | <20                         | -      | +      | N/A         |                |
|                  | B7*        | 60          | M         | 22         | 2.8               | +      | <20                         | -      | -      |             |                |
|                  | B8         | 74          | M         | 30         | 1.4               | +      | <20                         | -      | -      |             |                |
|                  | B9         | 55          | M         | 31         | 1.8               | +      | <20                         | -      | +      |             |                |
|                  | B10        | 61          | M         | 18         | 1.2               | +      | <20                         | -      | -      |             |                |
|                  | B11        | 47          | M         | 22         | 1.8               | +      | <20                         | -      | +      |             |                |
|                  | B12        | 61          | M         | 16         | 3.1               | +      | <20                         | -      | +      |             |                |
|                  | B13        | 61          | M         | 16         | 1.1               | +      | <20                         | -      | -      |             |                |
|                  | B14        | 57          | M         | 50         | 256.7             | +      | 90                          | +      | -      |             |                |
|                  | B15*       | 57          | M         | 83         | 1.8               | +      | <20                         | -      | +      |             |                |
| HCC (n = 4)      | B16        | 73          | M         | 29         | 88.9              | +      | <20                         | -      | +      | I           | 1.4            |
|                  | B17*       | 78          | F         | NA         | 1.2               | +      | <20                         | -      | +      | I           | 0.7            |
|                  | B18*       | 57          | M         | 38         | 3045.0            | +      | 50                          | +      | -      | I           | 4.8            |
|                  | B19*       | 70          | M         | 23         | 756.4             | +      | <20                         | -      | +      | I           | 3.0            |
| Post-HCC (n = 6)| B20*       | 51          | M         | 28         | 3.5               | +      | <20                         | +      | -      | N/A         |                |
|                  | B21        | 66          | M         | 46         | 1.2               | +      | <20                         | -      | -      |             |                |
|                  | B22        | 58          | M         | 28         | 2.5               | N/A    | <20                         | -      | -      |             |                |
|                  | B23        | 74          | M         | 28         | 35.3              | +      | N/A                         | -      | +      |             |                |
|                  | B24        | 56          | M         | NA         | 47.1              | +      | <20                         | -      | +      |             |                |
|                  | B25        | 56          | M         | 50         | 5.5               | +      | <20                         | -      | +      |             |                |

*More than one urine collection taken in the course of the study.
†HCC tumors were staged using the tumor-node-metastasis staging system.
Recurrence detected.

Abbreviations: +, HBsAg was taken at the time of HBV diagnosis; AFP, alpha-fetoprotein; ALT, alanine aminotransferase; F, female; HBeAb, hepatitis B e antibody; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; M, male; N/A, not available.
determined by quantitative PCR). The artifact library was used to remove HBV-JSs from each test sample (Supporting Table S2).

DATA AVAILABILITY STATEMENT

The sequence data for the subjects studied in this work who consented to data archiving have been deposited in the NCBI Sequencing Read Archive (accession no. PRJNA720423).

STATISTICAL ANALYSIS

The chi-square test was performed for categorical variables. For all statistical tests, the level of significance was defined as $P < 0.05$ to determine whether differences were significant.

Results

DETECTION OF HBV INTEGRATION IN URINE DNA BY JUNCTION-SPECIFIC PCR ASSAYS

To investigate whether HBV-JSs can be detected in urine, we identified eight pairs of archived DNA samples from matched tissue and urine from patients with HBV-HCC (study A; Table 1), with each urine DNA isolate extracted from at least 1 mL of urine. We first detected HBV-JSs in tissue DNA using an HBV-targeted NGS assay, as described previously. HBV integrations were detected in all eight HCC tumor DNA samples (Table 3). The affected regions included the TERT gene, the most frequently reported integrant-containing host gene (patient A5). (9) The junctions were validated by Sanger sequencing (Supporting Fig. S1) using primers listed in Supporting Table S1.

To detect HCC tissue-derived HBV junction DNA in urine, we designed a junction-specific PCR assay for each junction (Supporting Table S1). The amplicons were designed to be short, as urine cfDNA is mostly fragmented to less than 250 bp. (21) Matched tissue and urine DNA samples were subjected to junction-specific PCR amplification. PCR products of expected size were generated from five of the eight matched urine samples tested. To determine whether the PCR products generated from the urine and tissue in each of the five tissue-urine pairs were identical, we used either two-step nested PCR followed by RE digestion of PCR products (patients A5, A6, and A8; Fig. 1A) or Sanger sequencing when no suitable RE site was available in PCR products (patients A1 and A2; Fig. 1B-D). Four (patients A2, A5, A6, and A8) of the five PCR product pairs contained sequences indistinguishable by these methods, indicating that the HCC tissue-derived HBV-JS was present in the matched urine.

Interestingly, the sizes of the junction PCR products from patient A1 were different between tissue and urine (Fig. 1C). Sanger sequencing analysis identified a 23-nt insert, indicating a re-arrangement between chromosomes 5 and 10 in the urine DNA sample.

To determine whether the urine-identified integration site could be found in tumor-tissue DNA, we

| TABLE 3. HBV INTEGRATIONS IDENTIFIED IN ARCHIVED TUMOR-TISSUE DNA OF STUDY A PATIENTS |
|---------------------------------|----------------|----------------|----------------|
| Patient ID | HBV–Host Junction Sequences | HBV Junction Breakpoint Nucleotide Position* | Gene |
| A1 | ccttgaggcgttccccaaacccagagcccct | Chr.10: 31192695 | LOC101929352 |
| A2 | agagagtgtgtgtgtgttcaaaatttgcctga | Chr.1: 165248017 | None |
| A3 | gtgggggggggggttcgcttcatttctaggagg | Chr.11: 920486 | LOC101929352 |
| A4 | gccccagccccacctttggtacgccggaagttt | Chr.16: 30207023 | None |
| A5 | gtcgtcggtgtagcttcattgcgggtctgctgctggtctg | Chr.5: 1292170 | TERT |
| A6 | ctgggggtgtgtgcctgcggccctgccgatgcttcata | Chr.14: 67004392 | GPHN |
| A7 | gatctgtgtgtagcttcattgcgggtctgctgctggtctg | Chr.5: 165559760 | None |
| A8 | gatctgtgtgtagcttcattgcgggtctgctgctggtctg | Chr.14: 103176895 | None |

Note: Lowercase sequences represent HBV DNA. Underlined and capitalized sequences represent human DNA. Underlined, lowercase, bold bases represent overlapping human/HBV sequences. “None” indicates that no known gene was present within 100 kb of the integration site.

*HBV accession #NC_003977.2; human genome GRCh38.
designed a primer across the chimeric chromosome 5/10 sequence (Fig. 1D) to specifically amplify the 23-nt insert–containing HBV-JS species. Remarkably, the tripartite HBV-JS was successfully amplified and validated in the matched tissue DNA sample (Fig. 1D).

In total, we were able to detect and validate five of the eight HCC-tissue-derived HBV-JSs in matched urine samples.

**DETECTION AND CHARACTERIZATION OF HBV-JSs IN THE URINE OF HBV-INFECTED PATIENTS BY AN HBV-TARGETED NGS ASSAY**

After detecting liver-derived HBV-JSs in the urine of patients with HBV-HCC by junction-specific PCR assays, we next assessed whether it was feasible to analyze HBV integrations in urine without prior knowledge of junction sequences from tumor tissue and before disease progression to HCC. A total of 32 urine specimens were collected from 25 HBV-infected individuals, including 5 patients with hepatitis, 10 with cirrhosis, 4 with HCC, and 6 post-HCC (study B; summarized in Table 2). Five patients provided more than one urine specimen in the course of the study. For HBV-negative controls, we included urine samples from 4 males and 4 females who had no known history of HBV infection. Urine DNA was isolated, enriched for HBV sequences, and subjected to NGS followed by HBV-JS identification as described previously. We used a threshold of two unique (UMT-consolidated) supporting reads (SRs) for junction detection for two reasons. First, this cutoff
is consistent with other studies of HBV integration in HCC tissue.\(^{(8,26–28)}\) Second, the PCR-Sanger sequencing validation rate we obtained in a previous study of 16 HBV-HCC tumor tissue samples was 88%, which is in line with validation rates reported in the literature.\(^{(23,26,28)}\) No HBV-JSs with two or more SRs were observed in the NGS data obtained from the 8 HBV-negative controls. In contrast, 31 of 32 HBV-infected urine samples contained HBV-JSs at this cutoff.

Of the total of 712 HBV-JSs identified, 84, 369, 41, and 218 junctions were detected in 4 hepatitis, 12 cirrhosis, 6 HCC, and 9 post-HCC urine samples, respectively. Table 4 lists the major HBV-JSs of each individual urine sample, defined as the two most abundant HBV-JSs supported by at least 3% of total junction reads. As expected, we detected integrations in both gene-coding and noncoding regions in all disease categories. The latter included repeat regions, such as LINEs, SINEs, and simple repeats,\(^{(25)}\) which have been previously reported as HBV integration sites.

Among the targets of the 351 unique integrations in gene-coding regions detected in urine, 11 genes had been previously reported in liver tissue from patients with severe liver diseases, such as cirrhosis and HCC. These genes are \(AC007880.1\), \(ADAM12\) [disintegrin and metalloproteinase domain-containing protein 12], \(ARSA\) [aryl sulfatase A], \(ATF6\) [activating transcription factor 6], \(CLEC2L\) [C-type lectin domain family 2 member L], \(FOCAD\) [focadhesin], \(PPP2R2C\) [protein phosphatase 2 regulatory subunit Bgamma], \(TNFRSF9\) [TNF receptor superfamily member 9], \(TXNDC16\) [thioredoxin domain containing 16], \(VEGFC\) [vascular endothelial growth factor C], and \(TERT\). Most interestingly, all of the genes, except for \(AC007880.1\) and \(ARSA\), have been associated with the development of HCC \((ADAM12,\text{ATF6, }^{(30)}\text{CLEC2L, }^{(32)}\text{TERT, and }\text{VEGFC})^{(33)}\) or other cancers \((FOCAD,\text{PPP2R2C, }^{(35)}\text{TNFRSF9,}^{(36)}\text{and }\text{TXNDC16})^{(37)}\).

**DISTRIBUTION OF INTEGRATED HBV BREAKPOINTS DETECTED IN URINE**

We next examined the HBV genome breakpoints of the 712 junction sequences detected in urine. The HBV sequences of these junctions were plotted based on their locations in the HBV genome (Fig. 2A), and the SR statistics and HBV breakpoint regions are summarized in Fig. 2B. The number of SRs for a given junction varied widely within each disease type. Of the 712 junctions detected, 70% had two SRs. Approximately 70% of HBV-DNA breakpoints detected in urine were clustered in the HBV DR1-2 region, a known integration hotspot\(^{(8)}\) (Fig. 2B). The second-most frequent of the previously reported HBV breakpoint regions, PreS,\(^{(8)}\) was also identified in urine, with higher frequencies found in hepatitis (7.1%), cirrhosis (10.5%), and post-HCC (7.3%) as compared with HCC (4.9%). However, the association of integrated HBV breakpoint location (DR1-2, PreS1/2, or other) with disease type was not statistically significant \((\chi^2 [6, 712] = 6.3); \ P = 0.39) (Fig. 2B).

**PROFILING OF DETECTABLE HBV INTEGRATION SITES IN SERIALLY COLLECTED URINE SAMPLES**

Urine cfDNA containing HBS-JSs may reflect the HBV integration profile of the infected liver. For a preliminary assessment of the possibility of using HBS-JSs for disease monitoring, we collected serial urine samples from 5 patients (B7, B15, B17, B19, and B20) and compared the HBV-JSs detected in different collections (Table 4). Patients B15 and B19, both with cirrhosis, had two collections 1 week apart and showed consistent integration in SINE and simple repeat regions, respectively, although the integration breakpoints were different. This finding indicates that the detection of these junctions does not reflect clonal expansion, but rather a high incidence of integration in the repeat regions. Similarly, patient B20 (post-HCC) had four collections at 3-month intervals, all showing integration in noncoding regions (e.g., simple repeats), although the specific sites varied. Patient B17 (HCC) had two urine collections 2 months apart, with different junction sequences detected. Patient B7, with cirrhosis, had two urine samples collected 7 months apart, with integrations initially detected in a noncoding region followed by integrations in gene-coding regions. Interestingly, patient B7 developed HCC 4 years after the second collection. This initial comparison suggests that different HBV integrations can be captured in urine through serial monitoring. Collectively, our data demonstrate the feasibility of analyzing HBV integration using urine liquid biopsy.
| Patient ID | Diagnosis          | No. of HBV-JS Species Identified | Major Junctions (% of Total Junction Reads)* |
|------------|--------------------|----------------------------------|---------------------------------------------|
| B1         | Hepatitis (n = 5)  | 48                               | Chr4 (5.0%), SR (4.3%)                      |
| B2†        |                    | 0                                | N/A                                         |
| B3         |                    | 2                                | Chr1-a (50.0%), Chr1-b (50.0%)              |
| B4         |                    | 20                               | CHCHD2P9 (20.7%), HSPE1P8 (17.2%)           |
| B5         |                    | 10                               | SINE (19.0%), SEMA6D (14.3%)                |
| B6         | Cirrhosis (n = 12) | 21                               | MLLT4 (10.2%), FOCAD (8.2%), PAK3 (8.2%)    |
| B7a†       |                    | 0                                | N/A                                         |
| B7b        |                    | 5                                | GCSPH2 (20.0%), PPP2R2C (20.0%)             |
| B8         |                    | 30                               | SR (6.3%), LINE (4.8%)                      |
| B9         |                    | 10                               | SR (28.0%)                                  |
| B10        |                    | 40                               | XRCC3 (9.6%), MSL1 (8.0%)                   |
| B11        |                    | 18                               | TERT (48.8%), Chr16 (7.1%)                  |
| B12        |                    | 53                               | TERT (30.5%), GCSPH2 (6.9%)                 |
| B13        |                    | 54                               | SR (5.1%), LINC0148 (4.4%)                  |
| B14        |                    | 13                               | ALR54742.1 (13.8%), Chr5 (10.3%)            |
| B15a‡      |                    | 46                               | SINE (13.2%), LINE (3.9%)                   |
| B15b       |                    | 58                               | SINE (8.0%), EXD3 (6.8%)                    |
| B16        | HCC (n = 6)        | 4                                | KIF14 (25.0%), TET (25.0%)                  |
| B17a‡      |                    | 2                                | VPS8 (50.0%), Chr9 (50.0%)                  |
| B17b       |                    | 5                                | SR (20.0%), AC011193.1 (20.0%)              |
| B18        |                    | 11                               | SR (36.4%), TERT (9.1%)                     |
| B19a‡      |                    | 17                               | CTD-3229J4.1 (50.8%), SR (9.2%)             |
| B19b       |                    | 2                                | SR-a (50.0%), SR-b (50.0%)                  |
| B20a‡      | Post-HCC (n = 9)   | 16                               | SR (25.0%), LTR (12.5%)                     |
| B20b       |                    | 63                               | LMF1-AS1 (5.6%), hsd17b6 (5.1%)             |
| B20c       |                    | 14                               | GCSPH2 (60.8%), SR (10.1%)                  |
| B20d       |                    | 13                               | SR (21.4%), LACTB2-AS1 (7.1%)               |
| B21        |                    | 4                                | SR (75.0%)                                  |
| B22        |                    | 13                               | TRPM3 (37.0%), ANKRD9 (13.0%)               |
| B23        |                    | 2                                | PCDHG5 (50.0%), HBB (50.0%)                 |
| B24        |                    | 11                               | Chr12 (20.7%), GALNT6 (10.3%)               |
| B25        |                    | 82                               | LINE (12.4%), LTR (5.8%)                    |

* Defined as up to two most abundant junctions with at least two supporting reads.
† No one junction sequence meets the major-junction criteria.
‡ Multiple collections were obtained from the patient.

Abbreviations: AC011193.1, ENSG0000025582.1; ALR54742.1, microRNA ENSG00000251912; ANKRD9, ankyrin repeat domain 9; BAGE2, BAGE family member 2; CHCHD2P, coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene; Chr, chromosome location; CTD-3229J4.1, Long Intergenic Non-Protein Coding RNA ENSG00000261555; EXD3, exonuclease 3'-5' domain containing 3; FOCA, focadhesin; GALNT6, polypeptide N-acetylgalactosaminyltransferase 6; GCSPH2, glycine cleavage system protein H pseudogene 2; HBB, hemoglobin subunit beta; HSD17B6, hydroxysteroid 17-beta dehydrogenase 6; HSPE1P8, heat shock protein family E (Hsp10) member 1 pseudogene 8; KIF14, kinesin family member 14; LACTB2-AS1, LACTB2 antisense RNA 1; LINC0148, long Intergenic Non-Protein Coding RNA 1468; LINE, long interspersed nuclear elements; LMF1-AS1, LMF1 antisense RNA 1; LTR, Long terminal repeat; MLLT4, Myeloid/Lymphoid or Mixed-Lineage Leukemia; MSL1, macrophage stimulating 1 like; N/A, not available; PAK3, p21 (RAC1) activated kinase 3; PCDHG5, protocadherin gamma subfamily C, 5; PPP2R2C, protein phosphatase 2 regulatory subunit Bgamma; SEMA6D, semaphorin 6; SINE, short interspersed nuclear elements; SR, simple repeat; TERT, telomerase reverse transcriptase; Translocated To, 4; TRPM3, transient receptor potential cation channel subfamily M member 3; VPS8, VPS8 subunit of CORVET complex; XRCC3, X-ray repair cross complementing 3.
Discussion

In this proof-of-concept study, we provide evidence for noninvasive detection of integrated HBV DNA in urine of patients infected with HBV with liver disease ranging from hepatitis to HCC. The importance of this study is threefold. First, we report the detection of integrated HBV DNA in urine. Second, we includes a comparative analysis of HBV integration in hepatitis, cirrhosis, HCC, and post-HCC using solely ectopic samples. Third, we provide evidence that urine contains liver-derived DNA, by taking advantage of the liver tropism of HBV and unique HBV-JSs as molecular signatures of infected hepatocytes. Taken together, our results demonstrate the potential of urine liquid biopsy for integrated HBV burden and HBV-related liver disease monitoring.

Our previous study demonstrated that only fragmented HBV DNA was detected in the urine of patients with HBV infection even when the blood viral load was as high as $10^8$ copies/mL. Urine cfDNA is mostly monocleosomal and dinucleosomal in size and largely derived from apoptotic cells. Therefore, an HBV sequence containing urine cfDNA likely originates from degradation of host-integrated HBV genomes protected by nucleosomes in apoptotic HBV-infected hepatocytes. Hepatocytes from the HBV-infected liver have a higher turnover rate than hepatocytes from uninfected normal liver. Thus, urine provides a unique biopsy source to noninvasively detect integrated HBV DNA and monitor infected-cell burden, neither of which is achievable with current clinical HBV tests.

As expected, HBV genome integration breakpoints detected in urine were found to be mostly...
Directed T-cell immunotherapies, as targeting HBV may be highly applicable in the development of HBV-related HCC. Characterization of HBV-JSs in urine can be feasible for patients at risk of developing primary or recurrent HCC, monitoring dynamics of HBV-JS quantity and quality can provide critical molecular profiling to aid in delineating liver pathogenesis in patients with HBV. Because integrations into known HCC drivers were only found in urine of patients with cirrhosis or HCC in this study cohort, characterization of HBV-JSs may hold potential for application in assessing HBV-related liver disease in infected patients. In the present proof-of-concept study, we demonstrated that detection of HBV-JSs in urine is feasible. Further study with a much larger sample size and quantitative measurement in addition to qualitative characterization is needed to evaluate its potential clinical applications.

REFERENCES

1) Tu T, Budzinska MA, Vondran FW, Shackel NA, Urban S. Hepatitis B virus DNA integration occurs early in the viral life cycle in an in vitro infection model via NTCP-dependent uptake of enveloped virus particles. J Virol 2018;92.e02007-17.
2) Chauhan R, Churchill ND, Mulrooney-Cousins PM, Michalak TJ. Initial sites of hepatadnavirus integration into host genome in human hepatocytes and in the woodchuck model of hepatitis B-associated hepatocellular carcinoma. Oncogenesis 2017;6:e317.
3) Chauhan R, Shimizu Y, Watashi K, Wakita T, Fukasawa M, Michalak TJ. Retrotransposon elements among initial sites of hepatitis B virus integration into human genome in the HepG2-NTCP cell infection model. Cancer Genet 2019;235:39-56.
4) Zhao K, Liu A, Xia Y. Insights into hepatitis B virus DNA integration-55 years after virus discovery. Innovation 2020;1:100034.
5) Shafritz DA, Shouval D, Sherman HI, Hadziyiannis SJ, Kew MC. Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. N Engl J Med 1981;305:1067-1073.
6) Matsubara K, Tokino T. Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. Mol Biology Med 1990;7:243-260.
7) Bonilla GR, Roberts LR. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. J Hepatol 2005;42:760.
8) Yan H, Yang Y, Zhang L, Tang G, Wang YuZhao, Xue G, 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.
9) Li X, Zhang J, Yang Z, Kang J, Jiang S, Zhang T, et al. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. J Hepatol 2014;60:975-984.
10) Esumi M, Aritaka T, Arii M, Suzuki K, Tanikawa K, Mizuo H, et al. Clonal origin of human hepatoma determined by integration of hepatitis B virus DNA. Can Res 1986;46:5767-5771.
11) Esumi M, Tanaka Y, Totsuka S, Shikata T. Clonal state of human hepatocellular carcinoma and non-tumorous hepatocytes. Cancer Chemother Pharmacol 1989;23:S1-S3.
12) Li C-L, Ho M-C, Lin Y-Y, Tseng S-T, Chen Y-J, Pai H-H, et al. Cell-free virus-host chimera DNA from Hepatitis B virus integration sites as a circulating biomarker of hepatocellular cancer. Hepatology 2020;72:2063-2076.
13) Chen W, Zhang KE, Dong P, Fanning G, Tao C, Zhang H, et al. Noninvasive chimeric DNA profiling identifies tumor-originated
HBV integrants contributing to viral antigen expression in liver cancer. Hepatol Commun 2020;4:1181-1191.
14) Qu C, Wang Y, Wang P, Chen K, Wang M, Zeng H, et al. Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy. Proc Natl Acad Sci 2019;116:6308-6312.
15) Li W, Cui X, Huo Q, Qi Y, Sun Y, Tan M, et al. Profile of HBV integration in the plasma DNA of hepatocellular carcinoma patients. Curr Genomics 2019;20:61-68.
16) Hann H-W, Jain S, Park G, Steffen JD, Song W, Su Y-H. Detection of urine DNA markers for monitoring recurrent hepatocellular carcinoma. Hepatoma Res 2017;3:105-111.
17) Jain S, Xie L, Boldbaatar B, Lin SY, Hamilton JP, Meltzer SJ, et al. Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. Hepatology Res 2015;45:1110-1123.
18) Lin SY, Jain S, Song W, Hu C-T, Su Y-H. Strategic assay developments for detection of hlv 1762t/1764a double mutation in urine of patients with HBV-associated hepatocellular carcinoma. In: Hepatocellular Carcinoma—Clinical Research. Chinese University of Hong Kong, China; 2012.
19) Lin SY, Dhillon V, Jain S, Chang T-T, Hu C-T, Lin Y-J, et al. A locked nucleic acid clamp-mediated PCR assay for detection of a p53 codon 249 hotspot mutation in urine. J Mol Diagn 2011;13:474-484.
20) Jain S, Su Y-H, Su Y-P, McCcloud S, Lee T-J, et al. Characterization of the hepatitis B virus DNA detected in urine of chronic hepatitis B patients. BMC Gastroenterol 2018;18:40.
21) Su Y-H, Song J, Wang Z, Wang X, Wang M, Brenner DE, et al. Removal of high molecular weight DNA by carboxylated magnetic beads enhances the detection of mutated K-ras DNA in urine. Ann NY Acad Sci 2008;1137:82-91.
22) Ding D, Lou X, Hua D, Yu W, Li L, Wang J, 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.
23) Shieh F-S, Jorgensen P, Steffen JD, Lin S, Jain S, Song W, et al. ChimericSeq: an open-source, user-friendly interface for analyzing NGS data to identify and characterize viral-host chimeric sequences. PLoS One 2017;12:e0182843.
24) University of Michigan. UM BRCP Bioinformatics Core. https://github.com/umich-brcf-bionformics/Connor. Accessed June, 2020.
25) Sung W-K, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765-769.
26) Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Knoennmer MI, Guan Y, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res 2012;22:593-601.
27) Li W, Zeng XI, Lee NP, Liu X, Chen S, Guo B, et al. HIVID: an efficient method to detect HBV integration using low coverage sequencing. Genomics 2013;102:338-344.
28) Lee VW, Bachtar M, Choo CC, Lee CG. Comprehensive review of hepatitis B virus-associated hepatocellular carcinoma research through text mining and big data analytics. Birol Rev Camb Philos Soc 2019;94:353-367.
29) Le Pabic H, Bonnier D, Wewer UM, Coutand A, Musso O, Baffet G, et al. ADAM12 in human liver cancers: TGF-β-regulated expression in stellate cells is associated with matrix remodeling. Hepatology 2003;37:1056-1066.
30) Shuda M, Kondoh N, Imazeki N, Tanaka K, Okada T, Mori K, et al. Activation of the ATF6, XB1, and grp78 genes in human hepatocellular carcinoma: a possible involvement of the ER stress pathway in hepatocarcinogenesis. J Hepatol 2003;38:605-614.
31) Zhao J, Ye J, Lin Y, Bu K, Mai R, Liu Z, Gao X, et al. Identification and validation of a ten-gene set variation score as a diagnostic and prognostic stratification tool in hepatocellular carcinoma. Am J Transl Res 2020;12:5683-5695.
32) Zhuang P-Y, Shen J, Zhu X-D, Lu L, Wang L, Tang Z-Y, Sun H-C. Prognostic roles of cross-talk between peritumoral hepatocytes and stromal cells in hepatocellular carcinoma involving peritumoral VEGF-C, VEGFR-1 and VEGFR-3. PLoS One 2013;8:e64598.
33) Weren RD, Venkataraman R, Cazier JB, Farin HF, Kets CM, De Voer RM, et al. Germline deletions in the tumour suppressor gene FOCAD are associated with polyposis and colorectal cancer development. J Pathol 2015;236:155-164.
34) Bluehm EG, Spencer ES, Mechem B, Gordon RR, Coleman I, Lewinshtein D, et al. PPP2R2C loss promotes castration-resistance and is associated with increased prostate cancer-specific mortality. Mol Cancer Res 2013;11:568-578.
35) El Kramani N, Elsherbiny NM, El-Gayar AM, Ebrahim MA, Al-Gayar MMJC. Clinical significance of the TNF-α receptors, TNFRSF2 and TNFRSF9, on cell migration molecules Fascin-1 and Versican in acute leukemia. Cytokine 2018;111:523-529.
36) Han F, Zhu Y, Han C, Xu Q, Wu J, Dai BO, et al. Identification and validation of an eight-gene expression signature for predicting high Fuhrman grade renal cell carcinoma. Int J Cancer 2017;140:1199-1208.
37) Summers J, Gilburt AR, Yang W, Aldrich CE, Saputelli J, Litwin S, et al. Hepatocyte turnover during resolution of a transient hepaviral infection. Proc Nat Acad Sci 2003;100:11652-11659.
38) Kim AK, Hamilton JP, Lin SY, Chang T-T, Hann H-W, Hu C-T, et al. Urine biomarker: novel approach to hepatocellular carcinoma screening. medRxiv Oncol 2020.
39) Summers J, Gilburt AR, Yang W, Aldrich CE, Saputelli J, Litwin S, et al. Hepatocyte turnover during resolution of a transient hepaviral infection. Proc Nat Acad Sci 2003;100:11652-11659.
40) Kim AK, Hamilton JP, Lin SY, Chang T-T, Hann H-W, Hu C-T, et al. Urine biomarker: novel approach to hepatocellular carcinoma screening. medRxiv Oncol 2020.
41) Tan AT, Yang N, Lee Krishnamoorthy T, Ooi V, Chua A, Zhao X, et al. Use of expression profiles of HBV-DNA integrated into genomes of hepatocellular carcinoma cells to select T cells for immunotherapy. Gastroenterology 2019;156:1862-1876.e1869.
42) Ferguson JR, Wallace Z, Connolly MM, Woon AP, Suckling RJ, Hine DW, Barber C, et al. Immune-mobilizing monoclonal T cell receptors mediate specific and rapid elimination of Hepatitis B-infected cells. Hepatology 2020;72:1528-1540.

Supporting Information
Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1783/suppinfo.