Urinary cell-free DNA as a noninvasive liquid biopsy for hepatocellular carcinoma: a novel tool for diagnostics and monitoring

Amy K. Kim[1], Selena Y. Lin[2], Hsin-Ni Liu[3], Surbhi Jain[2], Terence P. Gade[4], Fwu-Shan Shieh[2], Max Chao[2], James Hamilton[1], Hie-Won Hann[5], Ting-Tsung Chang[6], Dmitry Goryunov[2], Zhili Wang[2], Ying-Hsiu Su[3,*]

Author's affiliations:

[1]Division of Gastroenterology and Hepatology, Johns Hopkins School of Medicine. Baltimore, MD 21205

[2]JBS Science Inc., Doylestown, PA 18902

[3]Baruch S. Blumberg Institute, Doylestown, PA 18902

[4]Department of Radiology, University of Pennsylvania Hospital, Philadelphia, PA 19104

[5]Division of Gastroenterology and Hepatology, Thomas Jefferson University Hospital, Philadelphia, PA 19107

[6] Department of Internal Medicine, National Cheng Kung University Medical College, Center of infectious Disease and Signaling Research, National Cheng Kung University, Tainan, Taiwan, Republic of China

Corresponding Author:

*Ying-Hsiu Su, The Baruch S. Blumberg Research Institute, 3805 Old Easton Rd, Doylestown, PA, 18902; Tel: 215-489-4949; Fax: 215-489-4920; E-mail: Ying-hsiu.su@bblumberg.org.

Author Contributions: A.K. obtained patient specimens and co-wrote the manuscript; S.L. devised the experimental plans, supervised execution, and analyzed the data; S.J. conceived and executed the experiments; F.S.S., H.L., and M.C. performed data analysis and produced figures; Z.W. developed cfDNA isolation for the study. D.G. performed the data analysis and co-wrote the manuscript; T.G., J.H., H.W.H., and T-T.C. obtained patient specimens; Y.H.S. conceived the study, coordinated the experimental plans, and co-wrote the manuscript.

Financial Support: This work was supported by the National Institute of Health grant NCI W81XWH-20-1-0605 (PI Kim), 1K08CA237624-01A1 (PI Kim), R44HG008700 (PI Wang), R01CA202769 (PI Su), U01CA275648 (PI Su), R44CA165312-07 (PI Lin).

Conflict of Interest: S.J., S.L., Z.W., and Y.S are shareholders of JBS Science Inc. All other authors declare no competing interests.

Keywords: cell-free DNA, hepatocellular carcinoma, urine, somatic mutation, liquid biopsy

List of Abbreviations: cell-free DNA (cfDNA), cell-free tumor DNA (ctDNA), genitourinary (GU), hepatitis B virus (HBV), hepatocellular carcinoma (HCC), high-molecular-weight (HMW), low molecular weight (LMW), next generation sequencing (NGS), peripheral blood mononuclear cells (PBMC), quantitative PCR (qPCR), shallow whole genome sequencing (sWGS), transrenal DNA (trDNA), unique molecular identifier (UMI), variant allele frequencies (VAFs), whole genome sequencing (WGS).

Abstract

Background & Aims: Cell-free DNA (cfDNA) has advanced cancer genetic profiling through liquid biopsy. While plasma is traditionally the primary source, emerging evidence highlights urinary cfDNA as a novel and noninvasive alternative. This study aimed to comprehensively assess transrenal DNA (trDNA) as a novel noninvasive biomarker source in HCC patients, compared to blood-based liquid biopsy.

Approach & Results: HBV DNA was used as a biomarker for trDNA. HBV-targeted and HCC-focused next generation sequencing (NGS) and whole genome sequencing (WGS) were used to compare fragment insert-sizes, the genome coverage, and germline genotyping accuracy. Urinary cfDNA overall exhibited a predominantly mononucelosomal pattern similar to plasma cfDNA, but with shorter fragments, broader size distribution and a more pronounced 10-bp periodicity. In contrast, trDNA were shorter and more variable among all patients. In HCC patients, trDNA was even shorter, with distinct 4-mer end motifs, compared to non-HCC trDNA. Higher concentrations of HCC-distinctive 4-mer end motif and TP53 mutations were found in urine compared to plasma. The overall genome coverage breadth by WGS was similar between urine and plasma cfDNA, with a higher fraction of covered cancer-associated mutation hotspots in urine cfDNA. In 101 HCC patients, there was a 78% overall concordance of HCC-associated mutations (*TP53, CTNNB1,* and *hTERT*) and in select 15 patients, 97% overall position-level concordance by targeted NGS between plasma and urine cfDNA.

Conclusion: Urine cfDNA has comparable features with distinct characteristics to plasma cfDNA and is a promising tool for liver cancer studies.

Introduction

Cell-free tumor DNA (ctDNA) in biofluids has created new opportunities for cancer diagnostics and management. While most applications have focused on ctDNA in plasma, ctDNA from the bloodstream can filter through the kidneys into urine. Urine has been shown to contain ctDNA from both urological and non-urological cancers, including hepatocellular carcinoma (HCC) (1-18). This suggests that urine can serve as a noninvasive source of genetic and epigenetic information for HCC and as a potential alternative or complement to plasma for HCC liquid biopsy.

Urine provides a noninvasive approach that addresses the challenges of blood collection, such as limitations in specimen volume and frequency. Additionally, the convenience of at-home urine collection has the potential to enhance access to cancer screening, particularly in underserved populations (19).

Despite the promise of transrenal ctDNA, reports on its detection efficiency compared to blood ctDNA have been mixed, in both genitourinary (GU) (7, 20, 21), and non-GU cancers (14, 22-25). Urinary DNA is a complex mixture originating from prerenal, renal, and postrenal sources, it includes high-molecular-weight (HMW) DNA (>1,000 bp) from exfoliating urothelial cells and cell-free DNA (cfDNA) originating transrenally or from urinary tract cells (6, 18, 26). Unlike plasma, urine is not under stringent homeostatic regulation, and ctDNA detection is significantly affected by preanalytical factors and the assays used (27, 28). For instance, there is considerable variability in the reported sizes of transrenal DNA fragments, ranging from ~50 bp to 150-300 bp, likely due to differences in isolation and assay platforms (29, 30). This variability complicates the assessment of the urine liquid biopsy utility.

This study evaluated urine liquid biopsy against plasma-based liquid biopsy in HCC. Using hepatitis B (HBV) DNA in urine samples from patients with chronic HBV infection (hepatitis), cirrhosis, and HCC, we analyzed transrenal DNA fragment size. Different sequencing techniques were employed to compare overall genome coverage, 4mer-end motifs, HCC hotspot coverage, and germline or somatic mutations concordance. Targeted NGS of matched plasma and urine further confirmed the detection of somatic variants. This approach offers a promising noninvasive alternative for genetic profiling, facilitating frequent sampling for early detection and disease monitoring of urine liquid biopsy in clinical practice.

Materials and Methods

Patients and body fluid collection

Detailed patient information is summarized in **Tables S1, S2,** and **S3**. Specimen collections were approved by each study center's institutional review board. This study was conducted in accordance with both the Declarations of Helsinki and Istanbul. Urine was mixed with EDTA (10 mM final concentration) immediately

upon collection and stored at -80°C until isolation. Blood was collected in K_2EDTA BD Vacutainer tubes (BD, Franklin Lakes, NJ) and processed to obtain plasma and peripheral blood mononuclear cells (PBMC).

DNA isolation, quantitation, and size distribution analysis

DNA was extracted from urine samples as previously described (31). The low molecular weight (LMW) fraction (cfDNA; < 1 kb) was size selected with HighPrep PCR magnetic beads (MagBio Genomics, Gaithersburg, MD) at 0.55X bead/sample ratio. Plasma cfDNA was extracted using the Quick-cfDNA Serum & Plasma kit (Zymo Research, Irvine, CA) or JBS plasma cell-free DNA isolation kit (Cat# 08874, JBS Science, Inc, Doylestown, PA). DNA concentrations were determined using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Waltham, MA). DNA Size distribution was analyzed on Agilent's 2200 TapeStation system using High Sensitivity D5000 screentapes (Agilent, Santa Clara, CA). PBMC DNA was isolated using the Genomic DNA Isolation Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions.

Mutation detection by qPCR

Archived DNA from 101 HCC patients (15) was used for mutation testing. First, the total copy number of each target gene was determined for each DNA sample using quantitative PCR (qPCR) assays (JBS Science) according to manufacturer's instructions. Only urine DNA samples that contained >100 copies/mL were used for analysis. Mutation detection in three markers of interest was then performed in duplicate using JBS mutation assay (JBS Science).

NGS library preparation and sequencing and data analysis for HBV insert size and variants

For whole genome sequencing (WGS), fragmented DNA (LMW cfDNA or sonicated PBMC/total urine DNA) with DNA inputs ranged from 30 to 50 ng was subjected to Illumina NGS library preparation and sequenced on a NextSeq550 or HiSeq4000 instrument (2×150 bp; Genomics Core Facility at Drexel College of Medicine, Philadelphia, PA).

For targeted sequencing by JBS HBV or JBS HCC panel, libraries were prepared from 10 ng of urine or plasma cfDNA using the NEBNext Ultra II DNA Kit (NEB, Ipswich, MA) and xGen UDI-UMI adapters (IDT, Cornwall, IA) and subjected to enrichment according to the manufacturer's protocol, and sequenced on a NovaSeq S1 flow cell (Psomagen, Rockville, MD) for paired-end sequencing. The HCC panel targets 24 genes with a panel size of 64 kb (**Table S4**).

Raw data were demultiplexed and each paired-end read was tagged with the UMI. Unmapped BAM (uBAM) files were then converted to fastq format for initial alignment using BWA MEM and hg19 as a reference. The resulting BAM files were downsampled within each plasma/urine pair. UMI processing was

performed using fgbio, connor, and picard suites, as detailed in **Figure S1**. A minimum UMI family size of three was used in most analyses. Consensus reads were realigned, quality-filtered, and used to generate VCF files that were filtered for common germline variants (dbSNP138) and used to evaluate mutant variant frequencies.

Results

Urine cfDNA size distribution similar to plasma cfDNA

We performed capillary electrophoresis of LMW DNA isolated from urine samples of eight healthy individuals and eight HCC patients. As anticipated, mononucleosomal DNA (150-180 bp) was the predominant cfDNA species in all samples (**Figure 1**). Most samples also contained a minor dinucleosomal peak or "shoulder".

Next, shallow whole genome sequencing (sWGS) of LMW DNA isolated from three sets of matched urine and plasma samples from two patients with HCC (**Table S1,** LMW DNA profiles shown in **Figure S2**) was performed. Sample ID649 and ID697 were collected from the same patient eight months apart. Samples 649 from both urine and plasma were also resequenced at a higher depth. The WGS data are summarized in **Table S5**.

All plasma samples showed a major peak at 170-180 bp, corresponding to the mononucleosomal fragment size (**Figure 2A-D**) with a minor dinucleosomal peak. Urine cfDNA showed a similar mononucleosomal peak, but no dinucleosomal peaks were discernable. Both plasma and urine insert size distributions exhibited a subnucleosomal "sawtooth" pattern with a 10 bp periodicity, more pronounced urine. Consistent with this observation, the insert size distribution in the urine samples was wider compared to that in the plasma libraries (**Figure 2E**). Sequencing at a higher depth revealed no notable differences in fragment size distributions in both urine and plasma compared to sWGS.

Urinary HBV DNA is significantly shorter than human cfDNA in urine

Urinary HBV DNA which is liver-derived was used as a biomarker to examine the size of transrenal DNA (trDNA). Human cfDNA in urine (blue, **Figure 3A**) exerted a mono-nucleosomal peak with pronounced a subnucleosomal "sawtooth" pattern with 10 bp periodicity, similar to the earlier size distribution pattern (**Figures 1 & 2**).

We compared HBV DNA insert sizes (red, **Figure 3A**) which represent fewer reads than human cfDNA, in urine samples from two patients: Pt 1 with HBV and Pt 2, a patient with HBV-associated HCC. The HBV DNA is significantly shorter than human cfDNA in urine ($p<0.001$, by student t-test) for both examples.

Next, we compared the HBV DNA size distribution between urine and plasma from individuals collected on the same day to control for temporal variations and technical variables (**Figure 3B**). To avoid possible

contamination from viral particles in the blood, we selected matched urine and plasma from three patients with undetectable serum viral load, Hepatitis (I) and HCC (I and II). A hepatitis B patient with high viral load, 6.9 log IU/mL, Hepatitis (II) was included as a control. As expected, plasma HBV DNA showed typical nucleosomal sizes, whereas HBV trDNA showed significantly shorter fragment sizes, peaking in a wider range of 80-150 bp which suggests greater individual variation (**Figure 3B**). By examining urine and plasma from a high viral load patient Hepatitis (II), trDNA size distribution was similar to that of patients with undetectable viral load, Hepatitis (I) and HCC (I & II). The size of circulating HBV DNA in plasma of the patient with high viral load (Hepatitis II) was less focused in nucleosomal sizes and peaked at 228 bp in this individual, possibly derived from circulating viral particles.

Comparison of HBV DNA insert size among three disease groups, hepatitis, cirrhosis, and HCC

It has been demonstrated that ctDNA is shorter than cfDNA in the blood (32). Since trDNA in urine is derived from blood, we hypothesized that shorter urinary HBV DNA would be observed in patients with liver cancer. We thus compared HBV DNA size distribution in urine from patients with HCC (n=17) to those with hepatitis (n=14) and cirrhosis (n=13) (**Figure 4A-C)**. As expected, urine HBV DNA differed significant among the three disease groups ($p<0.001$, one way ANOVA) as summarized in **Figure 4D**. HBV DNA in HCC patients was the shortest (cirrhosis-HBV: $p<0.001$, cirrhosis-HCC: $p=0.048$, HBV-HCC: $p<0.001$, pair-wise comparison).

Comparison for 4-mer end motifs of HBV DNA between patients with or without HCC in urine and plasma

cfDNA fragmentation has been shown to be nonrandom with various DNA nucleases generating sequence-specific cleavage or unique 5' end motifs. These end motifs have been found to distinguish between cancer and non-cancer patient plasma samples including those with HCC (33-36). Hence, we analyzed the 4-mer end motifs of HBV DNA between HCC and non-HCC HBV patients to determine whether distinct signatures can also be observed in trDNA and whether these signatures are similar between urine and plasma.

We compared the motifs frequency of the 10 most frequent HBV DNA 4-mer end motifs in plasma (**Figure 5A**, top panel) and in urine (**Figure 5B**, top panel) from HBV patients with HCC (HCC, n=17, blue) and without HCC (CHB, n=14, red) (**Figure 5**, bottom panels). Overall, there is only one, CGTT, overlapped between the 10 most frequent motifs identified in plasma and urine. CGTT is the 10th most frequent observed motif in plasma and the 3rd most frequent motif in urine, but for both plasma and urine, frequency was much lower in HCC than non-HCC. Although the first 9 most frequent motifs in plasma did not exert detectable difference between HCC and non-HCC, the first two, CCCA and CCAA, in urine were found to be much more abundant in HCC than in non-HCC. Specifically, in plasma, CCAA, had no appreciable difference in frequency

between patients with and without HCC, but found to be three times more abundant in urine from HCC patients compared to non-HCC. Similarly, the second most frequent 4-mer end motifs in urine, GAGA, was 6-fold more in HCC urine than in non-HCC urine, but no detectable difference found in plasma. Note, GAGA was not part of the 10 most frequent 4-mer end motifs identified in plasma. Collectively, the comparison of these 4-mer motifs suggests the difference between HCC and non-HCC is more pronounced in trDNA compared to cfDNA in plasma.

Comparison of genome representations in urinary and plasma cfDNA from patients with HCC

The total fractions of overall genomic positions covered at least once in urine and plasma cfDNA were determined and demonstrated similarity within each pair as detailed in **Table S5**. Overall, 87.5% to 94.4% of the urine cfDNA coverage overlapped with the genomic positions covered in the matched plasma cfDNA. To examine the difference in sWGS genome coverage between matched plasma and urine cfDNA in more detail, we visualized the overlap of genomic positions covered in each pair using Euler diagrams (**Figure 6**). The urine/plasma overlap ranged from 48% to 58% of the genome in the three shallow-sequenced sample pairs and increased to 60% to 73% with a higher depth sequencing of samples (pt.649). Positions covered in plasma, but not in urine comprised 6-8% of the genome, whereas the fraction of urine-only coverage was 7-13% across all pairs. Moreover, despite the cfDNA inputs of the plasma libraries being 2- to 8-fold higher than those of the urine, comparable overall coverage and high coverage overlap between plasma and urine were observed.

Next, we compared cancer hotspot coverage obtained in the two matched sample types using the downsampled sWGS data. The hotspot regions analyzed corresponded to the NEBNext™ Direct hotspot cancer panel which is comprised of 190 target regions in 50 genes (37). The hotspot coverage in urine was comparable to, or slightly higher than, the corresponding plasma coverage in each pair as detailed **Figure S3A**. In deep-sequenced pair of ID649 samples, both plasma and urine cfDNA covered approximately 94.5% of the hotspot target positions. These findings are consistent with previous reports of cancer-associated mutations being enriched in urine DNA in other solid tumors and our earlier work (5, 16, 38).

We then visualized the hotspot coverage in all three sample pairs using Circos plots (**Figure S3B and C**). Although the coverage profiles were divergent between plasma and urine, they were similar within each sample type. This observation suggests that while the genome may be represented differentially in plasma and urine cfDNA, the biases are qualitatively reproducible within sample types. This finding was further supported by hotspot coverage profiles in deep-sequenced sample pair ID649 being similar to those seen in the shallow-sequenced counterparts (data not shown).

TP53, CTNNB1, and TERT mutation prevalence in matched urine and plasma cfDNA from patients with HCC by qPCR

Because the sWGS results showed differential genomic coverage in urine and plasma cfDNA, we compared the prevalence of HCC-associated mutations between the two sources of ctDNA. HCC is most frequently linked to mutations in the *TERT*, *TP53*, and *CTNNB1* genes (39, 40). We have previously developed short-amplicon qPCR assays for the detection of mutations in the *TERT* promoter (c.-124), *TP53* (codon 249), and *CTNNB1* (codons 32-37) (4, 16) with sensitivities at 0.15%, 0.1%, and 0.3%, respectively. Urine and plasma cfDNA samples isolated from a total of 101 patients with HCC were screened. The results are detailed in **Table S6** and summarized in **Table 1**.

The difference in overall non-matched detection rate was not statistically significant between plasma and urine (18.4% vs. 16.9%; p=0.797, Fisher's exact test). The overall concordance (i.e. across all patients, target genes, and urine-plasma pairs, including those with double negative results) was 78%. However, the number dropped to 24% when only urine-plasma pairs with at least one positive result were counted (positive agreement) among all target genes. In this group, *TERT* showed the highest concordance (31%), followed by *CTNNB1* (20%) and *TP53* (0%). Notably, *TP53* mutations were detected more in urine than in plasma (p=0.023; Fisher's exact test). No significant differences in *CTNNB1* or *TERT* mutation prevalence were observed between plasma and urine within pairs with at least one positive result (**Table S7**).

Targeted NGS analyses of urine and plasma cfDNA from HCC patients

Next, we performed hybridization-based target enrichment followed by NGS to evaluate the potential of utilizing matched urine and plasma cfDNA for somatic mutation detection in 15 patients with HCC (**Table S3**). All data were downsampled to approximately 2.2×10^7 paired-end reads for comparative analysis and consolidated by UMI with a minimum family size of three, resulting in mean per-base on-target coverage of approximately 1,500× and 1,280× for plasma and urine, respectively (**Table S8**). The global (i.e. across all patients/target positions) variant call rates were significantly higher in urine compared to plasma (0.0266 vs. 0.0186; McNemar χ^2=1,942.6; p<0.0001), potentially indicating either a higher background noise in the urine data, urine being enriched for ctDNA relative to plasma, or both. The global position-level concordance was 96.8%, while the both-variant positive agreement was 17.7%. The concordances for plasma- and urine- variant positive positions were 36.5% and 25.5%, respectively (**Table S9**). The global mean variant allele frequencies (VAFs) were similar between the two body fluids, albeit the difference was significant (0.046% for plasma vs. 0.048% for urine; Wilcoxon signed-rank test p<0.0001). Consistent with this result, plasma and urine VAFs for double-positive positions showed strong global correlation (Pearson r=0.966 [95% CI 0.966-0.966]; **Figure S4A**). The correlation was even stronger for position-level mean VAFs across the entire cohort (Pearson

r=0.995 [95% CI 0.995-0.995]; **Figure S4B**). As expected, putative somatic VAFs (<0.4) correlated less strongly (Pearson r=0.890 [95% CI 0.884-0.895]; **Figure S4C**).

Among the 24 target genes, *TERT* promoter showed the highest somatic variant call rate in both plasma and urine with 20.2% and 21.9%, respectively, followed by *KMT2D* (3.7 and 4.2%), *TP53* (2.8 and 3.5%), and *ARID1A* (2.4 and 3.5%). Mutation rates were overall similar between urine and plasma, but in six targets, urine had more than a two-fold higher variant rate than in plasma (**Table S10**). Putative somatic mutations with 100% sample-level (n=30) frequencies were found in *ARID1, hTERT* promoter, *PCLO, KMT2D, AXIN1, TSC2, TP53*, and *KEAP1* (**Supplemental File 1**). Taken together, these results suggest that the mutational profiles of HCC tumors are likely to be reflected similarly in urine and plasma cfDNA.

Comparison of blood- and urine-derived DNA as a source of germline genotype information

Liquid biopsy assays using blood as the source of genetic material often rely on sequencing DNA extracted from PBMCs to determine the patient's germline genotype. In addition to cfDNA, urine contains exfoliated urothelial cells as well as some leukocytes (41), which normally contain the individual's germline genetic information. For a urine-based liquid biopsy assay to be self-sufficient, both somatic and germline genotype information needs to be recoverable from urine DNA. To demonstrate the feasibility of germline genotyping using total urine DNA, we first performed WGS of PBMC/total urine DNA pairs isolated from six healthy subjects (**Table S11**). In all pairs, the genome fraction covered in urine DNA was ≥ 99.3% relative to the matched PBMC coverage. Next, we evaluated genotype concordance across all positions sequenced at a minimum depth of 10 in both sample types. Perfect concordance was observed in all six sample pairs (**Table 2**). These results demonstrate that total urinary DNA can be used for germline genotyping similarly to DNA isolated from PBMCs.

Discussion

This study demonstrated that urinary cfDNA provides a viable, noninvasive source for genetic profiling, capable of detecting both germline and somatic mutations. We analyzed genomic features including DNA fragment size distribution, 4-mer end motifs, HCC-associated mutation detection, and genome coverage of trDNA in comparison to plasma cfDNA to evaluate the potential of urine-based liquid biopsy as a noninvasive alternative or complement to blood-based liquid biopsy for HCC. Notably, based on HCC characteristic 4-mer end motifs, *TP53* mutations, and the HCC-focused NGS (**Table S8**) VAF analysis, urine may contain a higher concentration of some HCC ctDNA. By WGS, cfDNA from urine and plasma have similar overall genome coverage with mostly overlapped positions. HCC-associated mutations were also detectable at similar rates

overall between urine and plasma cfDNA by qPCR and targeted NGS which revealed high concordance between sequence variants detectability in the two body fluids.

We previously demonstrated that HBV DNA detected in urine originates from chronically infected hepatocytes unless there is blood contamination. Using HBV DNA as a trDNA marker to compare trDNA with total cfDNA in urine, and samples between HCC and non-HCC patients, several notable characteristics were found. **First**, HBV trDNA from non-cancer patients is shorter and has a broader range in its size distribution than plasma cfDNA, with more variance in size peaking at ~50 bp, 80 bp, 100 bp or even longer. Our analysis was from all DNA isolated without pre-processing centrifugation to avoid losing cell debris associated trDNA (31, 42). This may explain discrepancies for the trDNA size reported in the literature (43) ranging from mostly 50 bp or less, with median sizes of 80 or 100 bp (29, 30). **Second**, trDNA is shorter in HCC than non-HCC, which is consistent with previous studies in blood. **Third**, the most frequent 4-mer end motif, CCCA, which was previously identified in plasma cfDNA, is also the most frequent motif identified in HBV DNA in plasma, although the abundance of HBV CCCA motif shows no detectable difference in plasma between HCC and non-HCC, as shown in a previous study (33). Although CCAA is not among the 10 most frequent motifs observed in urine, it was significantly more abundant in HCC patients compared to non-HCC patients. This finding contrasts with previous plasma-based studies, where CCAA was observed to be less abundant in HCC patients compared to non-HCC patients. Other studies do not show a significant difference in CCAA between cancers and controls as analyzed by cfOmics (44, 45). While there may be some inconsistency in 4-mer motifs comparison between studies, our data using HBV DNA motifs suggests the end motifs frequencies in urine may show better distinction between HCC and non-HCC, compared to plasma. This is consistent even in the detection of TP53 mutation, where significantly more TP53 mutations were detected in urine than in plasma. This is similar to findings in urothelial cancers (46, 47) and NSCLC (25).

Using mutation-specific qPCR assays, we found a high overall concordance of 78% in 101 patients screened for three HCC hotspots mutations and the overall non-matched detection rates were not significantly different between plasma and urine. Positive agreement between plasma and urine was only 24%, suggesting that ctDNA representation may differ between urine and plasma even in patients with non-genitourinary malignancies (48). This is in contrast with other plasma-urine *EGFR* T790M positive agreement reports in other cancers such as NSCLC (11, 38) and bladder (49). The apparent discrepancies in urine/plasma positive agreement between these studies and our results may be attributable to the pre-analytic variables such as detection assays and DNA isolation method and fragmentation extent of different genes. Similar reasons may underlie the relatively low plasma-urine positive agreement (18%) in the targeted NGS data from additional 15 patients with HCC. It is also important to note that most patients in our cohort had early-stage disease and our library inputs were only 10 ng, both of which may have substantially reduced the limit of detection sensitivity in

our study. Tumor tissue mutational profiles were unavailable for most patients, limiting our ability to analyze any concordance between tissue and plasma/urine DNA. However, it is well-known that mutation profiles often differ between tissue samples and plasma. This discrepancy arises because tissue analysis typically relies on fine needle biopsies which is only a very small fraction of the highly heterogenous tumor nodule(s) whereas plasma cfDNA represents a collection of cfDNA from the dying cells from the entire tumor (50).

Lastly, urine was also determined to be a reliable source of germline genotype information, similar to PBMCs in blood-based liquid biopsies with high genome fraction coverage. Perfect genome coverage concordance was obtained between urine DNA and the matched PBMC. Thus, similar to PBMC-derived DNA in blood-based liquid biopsy, total urinary DNA can be used to extract germline information.

Our study has several limitations. First, the disparate quantities of plasma and urine cfDNA used for WGS (but not targeted NGS) hindered a rigorous comparison of some WGS metrics and underlying biological phenomena. Second, due to the limited DNA available, the number of HCC patients analyzed by NGS was relatively small. Third, the DNA isolation method we used to obtain urinary cfDNA may not allow for efficient recovery of fragments <100 bp, which may be enriched for tumor-derived somatic variants (51). Lastly, the validity of comparative analyses of many liquid biopsy studies is greatly limited by the lack of a standardized approach to both DNA isolation and data acquisition technologies. Even within the same methodology, seemingly minor differences (e.g., sequencing depth or input amounts in the case of NGS) can affect study outcomes and conclusions drastically. It is therefore critical to reproduce our findings in larger-scale studies.

In conclusion, our results highlight urinary cfDNA's potential as a noninvasive tool for liver cancer diagnostics, monitoring, and personalized treatment strategies. It offers a promising complement—or even alternative—to plasma-based approaches. The ability to reliably detect key genomic features and mutations underscores its clinical utility, while the ease of urine collection and potential for improved patient compliance make it particularly attractive for broader application in liquid biopsy. By addressing challenges associated with blood-based methods, urinary cfDNA could play a transformative role in early detection, disease monitoring, and access to personalized care in hepatocellular carcinoma. Future studies will be essential to validate these findings and further optimize this approach for widespread clinical adoption.

References

1. Su Y-H, Song J, Wang Z, Wang XH, Wang M, Brenner DE, Block TM. Removal of high-molecular-weight DNA by carboxylated magnetic beads enhances the detection of mutated K-ras DNA in urine. Ann N Y Acad Sci 2008;1137:82-91.

2. Hann HW, Jain S, Park G, Steffen JD, Song W, Su Y-H. Detection of urine DNA markers for monitoring recurrent hepatocellular carcinoma. Hepatoma Research 2017;3:105-111.

3. Jain S, Xie L, Boldbaatar B, Lin SY, Hamilton JP, Meltzer SJ, Chen S-H, et al. Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. Hepatology Research 2014;45:1110-1123.

4. Lin SY, Chang T-T, Steffen JD, Chen S, Jain S, Song W, Lin Y-J, et al. Detection of CTNNB1 Hotspot Mutations in Cell-Free DNA from the Urine of Hepatocellular Carcinoma Patients. Diagnostics 2021;11:1475.

5. Song BP, Jain S, Lin SY, Chen Q, Block TM, Song W, Brenner DE, et al. Detection of hypermethylated vimentin in urine of patients with colorectal cancer. Journal of Molecular Diagnostics 2012;14:112-119.

6. Su Y-H, Wang M, Aiamkitsumrit B, Brenner DE, Block TM. Detection of K-ras mutation in urine of patients with colorectal cancer. Cancer Biomarkers 2005;1:177-182.

7. Nuzzo PV, Berchuck JE, Korthauer K, Spisak S, Nassar AH, Abou Alaiwi S, Chakravarthy A, et al. Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes. Nature medicine 2020;26:1041-1043.

8. Dudley JC, Schroers-Martin J, Lazzareschi DV, Shi WY, Chen SB, Esfahani MS, Trivedi D, et al. Detection and surveillance of bladder cancer using urine tumor DNA. Cancer discovery 2019;9:500-509.

9. Lee DH, Yoon H, Park S, Kim JS, Ahn Y-H, Kwon K, Lee D, et al. Urinary exosomal and cell-free DNA detects somatic mutation and copy number alteration in urothelial carcinoma of bladder. Scientific reports 2018;8:1-7.

10. Bach S, Wever BM, van de Wiel MA, Veltman JD, Hashemi SM, Kazemier G, Bahce I, et al. Dynamics of methylated cell-free DNA in the urine of non-small cell lung cancer patients. Epigenetics 2021;17:1057-1069.

11. Reckamp KL, Melnikova VO, Karlovich C, Sequist LV, Camidge DR, Wakelee H, Perol M, et al. A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. Journal of Thoracic Oncology 2016;11:1690-1700.

12. Wever B, Bach S, Tibbesma M, Ter Braak T, Wajon D, Dickhoff C, Lissenberg-Witte B, et al. Detection of non-metastatic non-small-cell lung cancer in urine by methylation-specific PCR analysis: A feasibility study. Lung Cancer 2022;170:156-164.

13. Liu B, Ricarte Filho J, Mallisetty A, Villani C, Kottorou A, Rodgers K, Chen C, et al. Detection of Promoter DNA Methylation in Urine and Plasma Aids the Detection of Non–Small Cell Lung CancerUrine Lung Cancer Epigenetic. Clinical Cancer Research 2020;26:4339-4348.
14. Mouliere F, Smith CG, Heider K, Su J, van der Pol Y, Thompson M, Morris J, et al. Fragmentation patterns and personalized sequencing of cell-free DNA in urine and plasma of glioma patients. EMBO Molecular Medicine 2021;13:e12881.
15. Kim AK, Hamilton JP, Lin SY, Chang T-T, Hann H-W, Hu C-T, Lou Y, et al. Urine DNA biomarkers for hepatocellular carcinoma screening. British Journal of Cancer 2022;126:1432-1438.
16. Lin SY, Dhillon V, Jain S, Chang T-T, Hu C-T, Lin Y-J, Chen S-H, et al. A locked nucleic acid clamp-mediated PCR assay for detection of a p53 codon 249 hotspot mutation in urine. Journal of Molecular Diagnostics 2011;13:474-484.
17. Lin SY, Halegoua-DeMarzio D, Block P, Kao Y-L, Civan J, Shieh F-s, Song W, et al. Persistently elevated HBV viral-host junction DNA in urine as a biomarker for hepatocellular carcinoma minimum residual disease and recurrence: a pilot study. Diagnostics 2023;13:1537.
18. Su Y-H, Wang M, Brenner DE, Ng A, Melkonyan H, Umansky S, Syngal S, et al. Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. Journal of Molecular Diagnostics 2004;6:101-107.
19. Chen VL, Xu D, Wicha MS, Lok AS, Parikh ND. Utility of liquid biopsy analysis in detection of hepatocellular carcinoma, determination of prognosis, and disease monitoring: a systematic review. Clinical Gastroenterology & Hepatology 2020;18:2879-2902.
20. Smith CG, Moser T, Mouliere F, Field-Rayner J, Eldridge M, Riediger AL, Chandrananda D, et al. Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. Genome Med 2020;12:23.
21. Christensen E, Nordentoft I, Birkenkamp-Demtröder K, Elbæk SK, Lindskrog SV, Taber A, Andreasen TG, et al. Cell-Free Urine and Plasma DNA Mutational Analysis Predicts Neoadjuvant Chemotherapy Response and Outcome in Patients with Muscle-Invasive Bladder Cancer. Clin Cancer Res 2023;29:1582-1591.
22. Yu H, Han L, Yuan J, Sun Y. Circulating tumor cell free DNA from plasma and urine in the clinical management of colorectal cancer. Cancer Biomarkers 2020;27:29-37.
23. Wu Z, Yang Z, Li CS, Zhao W, Liang ZX, Dai Y, Zhu Q, et al. Differences in the genomic profiles of cell-free DNA between plasma, sputum, urine, and tumor tissue in advanced NSCLC. Cancer medicine 2019;8:910-919.

24. Zuo Z, Tang J, Cai X, Ke F, Shi Z. Probing of breast cancer using a combination of plasma and urinary circulating cell-free DNA. Bioscience Reports 2020;40.
25. Yu H, Liu M, Qiu H, Yang K. Urinary and Plasma Cell-Free DNA Comparison for Lung Cancer Patients Treated With Epidermal Growth Factor Receptor-Thyroxine Kinase Inhibitors. Am J Med Sci 2019;357:29-36.
26. Jain S, Lin SY, Song W, Su Y-H. Urine-based liquid biopsy for nonurological cancers. Genetic Testing Molecular Biomarkers 2019;23:277-283.
27. Zhang A, Lee T-J, Jain S, Su Y-H. Urine as an Alternative to Blood for Cancer Liquid Biopsy and Precision Medicine. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2018: IEEE; 2018. p. 2820-2525.
28. Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. Quantitative analysis of the transrenal excretion of circulating EBV DNA in nasopharyngeal carcinoma patients. Clinical Cancer Research 2008;14:4809-4813.
29. Markus H, Zhao J, Contente-Cuomo T, Stephens MD, Raupach E, Odenheimer-Bergman A, Connor S, et al. Analysis of recurrently protected genomic regions in cell-free DNA found in urine. Science Translational Medicine 2021;13.
30. van Der Pol Y, Tantyo NA, Evander N, Hentschel AE, Wever BM, Ramaker J, Bootsma S, et al. Real‐time analysis of the cancer genome and fragmentome from plasma and urine cell‐free DNA using nanopore sequencing. EMBO molecular medicine 2023;15:e17282.
31. Lin SY, Luo Y, Marshall MM, Johnson BJ, Park SR, Wang Z, Su Y-H. A New Method for Improving Extraction Efficiency and Purity of Urine and Plasma Cell-Free DNA. Diagnostics 2021;11:650-659.
32. Jiang P, Chan CW, Chan KA, Cheng SH, Wong J, Wong VW-S, Wong GL, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proceedings of the National Academy of Sciences 2015;112:E1317-E1325.
33. Jiang P, Sun K, Peng W, Cheng SH, Ni M, Yeung PC, Heung MMS, et al. Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. Cancer Discov 2020;10:664-673.
34. Guo W, Chen X, Liu R, Liang N, Ma Q, Bao H, Xu X, et al. Sensitive detection of stage I lung adenocarcinoma using plasma cell-free DNA breakpoint motif profiling. eBioMedicine 2022;81.
35. Zhou Z, Ma ML, Chan RWY, Lam WKJ, Peng W, Gai W, Hu X, et al. Fragmentation landscape of cell-free DNA revealed by deconvolutional analysis of end motifs. Proc Natl Acad Sci U S A 2023;120:e2220982120.

36. Jin C, Liu X, Zheng W, Su L, Liu Y, Guo X, Gu X, et al. Characterization of fragment sizes, copy number aberrations and 4-mer end motifs in cell-free DNA of hepatocellular carcinoma for enhanced liquid biopsy-based cancer detection. Mol Oncol 2021;15:2377-2389.
37. Leong KW, Yu F, Adalsteinsson VA, Reed S, Gydush G, Ladas I, Li J, et al. A nuclease-polymerase chain reaction enables amplification of probes used for capture-based DNA target enrichment. Nucleic Acids Res 2019;47:e147.
38. Chen S, Zhao J, Cui L, Liu Y. Urinary circulating DNA detection for dynamic tracking of EGFR mutations for NSCLC patients treated with EGFR-TKIs. Clinical and Translational Oncology 2017;19:332-340.
39. Huang A, Zhang X, Zhou SL, Cao Y, Huang XW, Fan J, Yang XR, et al. Detecting Circulating Tumor DNA in Hepatocellular Carcinoma Patients Using Droplet Digital PCR Is Feasible and Reflects Intratumoral Heterogeneity. J Cancer 2016;7:1907-1914.
40. Jiao J, Watt GP, Stevenson HL, Calderone TL, Fisher-Hoch SP, Ye Y, Wu X, et al. Telomerase reverse transcriptase mutations in plasma DNA in patients with hepatocellular carcinoma or cirrhosis: Prevalence and risk factors. Hepatol Commun 2018;2:718-731.
41. Lin SY, Linehan JA, Wilson TG, Hoon DS. Emerging Utility of Urinary Cell-free Nucleic Acid Biomarkers for Prostate, Bladder, and Renal Cancers. European urology focus 2017;3:265-272.
42. Kim AK, Lin SY, Wang Z, Luu H, Hamilton JP, Song W, Su YH. Impact of Cell-Debris and Room-Temperature Storage on Urine Circulating Tumor DNA from Hepatocellular Carcinoma. Journal of Molecular Diagnostics 2023;25:913-920.
43. Tsui NB, Jiang P, Chow KC, Su X, Leung TY, Sun H, Chan KA, et al. High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. PloS one 2012;7:e48319.
44. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, Liu B, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. Cell Res 2017;27:1231-1242.
45. Li M, Zhou T, Han M, Wang H, Bao P, Tao Y, Chen X, et al. cfOmics: a cell-free multi-Omics database for diseases. Nucleic Acids Research 2024;52:D607-D621.
46. Avogbe PH, Manel A, Vian E, Durand G, Forey N, Voegele C, Zvereva M, et al. Urinary TERT promoter mutations as non-invasive biomarkers for the comprehensive detection of urothelial cancer. EBioMedicine 2019;44:431-438.
47. Ou Z, Li K, Yang T, Dai Y, Chandra M, Ning J, Wang Y, et al. Detection of bladder cancer using urinary cell-free DNA and cellular DNA. Clin Transl Med 2020;9:4.

48. Wu Z, Yang Z, Li CS, Zhao W, Liang ZX, Dai Y, Zhu Q, et al. Differences in the genomic profiles of cell-free DNA between plasma, sputum, urine, and tumor tissue in advanced NSCLC. Cancer Med 2019;8:910-919.

49. Christensen E, Birkenkamp-Demtröder K, Nordentoft I, Høyer S, van der Keur K, van Kessel K, Zwarthoff E, et al. Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. Eur Urol 2017;71:961-969.

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

51. Melkonyan HS, Feaver WJ, Meyer E, Scheinker V, Shekhtman EM, Xin Z, Umansky SR. Transrenal nucleic acids: from proof of principle to clinical tests. Ann N Y Acad Sci 2008;1137:73-81.

Figures legend

Figure 1. Electrophoretic size profiling of LMW DNA isolated from urine. Eight healthy and eight HCC urine LMW DNA samples were resolved on a TapeStation as described in Materials in Methods. Healthy (top) and HCC (bottom) sample panels are indicated. Equivalents of approximately 0.1 mL of healthy-donor urine and 0.057 mL of HCC-patient urine were loaded.

Figure 2. Analysis of WGS insert size distributions in matched urine and plasma cfDNA samples from patients with HCC. (A-C) Insert size distributions in shallow-sequenced samples. (D) Samples 649 resequenced at higher depths. Plasma and urine samples are dashed and solid lines, respectively. Mono- (1-Nucl) and di- (2-Nucl) nucleosomal peaks are indicated. The 10 bp periodic peaks are indicated with brackets. (E) Median and $25^{th}/75^{th}$ percentile insert sizes for each sample. In ID649, the bottom urine/plasma rows (*) are for the deeper WGS run. See **Table S5** for details.

Figure 3. trDNA size distribution as compared to human cfDNA in urine (A) and liver-derived cfDNA in plasma (B). (A) Size distribution between trDNA (HBV DNA, red) and urine cfDNA (human, blue), as indicated from one patient with HBV hepatitis, Pt 1 and one HBV-associated HCC patient, Pt 2. (B) HBV DNA size distribution between matched urine and plasma from two hepatitis B (I & II) and two HCC (I & II) patients. Note, HBV hepatitis (II) patient had a 6.9 log IU/mL serum viral load at the time of specimen collection.

Figure 4. Comparison of HBV DNA fragment size in urine from 14 hepatitis, 13 cirrhosis or 17 HCC patients. Insert sizes of the HBV DNA collectively from three disease groups, hepatitis (A), cirrhosis (B), and HCC (C) are plotted by NGS read counts per size. Panel D compares the median with P25 and P75 from each disease group.

Figure 5. Comparison of the 10 most frequent 4-mer end motifs abundance in HBV DNA fragments in plasma (A) and in urine (B, top panel) between HCC (blue) and non-HCC HBV hepatitis (red) patients. The 10 most frequent 4-mer end motifs identified in plasma (A) and in urine (B) listed by the motif frequency (%)(top panel) and corresponding matched plasma or urine (bottom panel) from HCC (blue) and non-HCC (HBV hepatitis, red) patients are compared.

Figure 6. Genomic coverage of LMW cfDNA isolated from matched urine (yellow) and plasma (pink/red). Euler diagram of genomic positions in urine and plasma covered in each pair. The urine/plasma overlap ranged from 48% to 58% of the genome in the three shallow-sequenced sample pairs and increased from 60% to 73% when the ID649 samples were sequenced to higher depths. Positions covered in plasma, but not in urine comprised 6-8% of the genome (pink area), whereas the fraction of urine-only coverage was 7-13% across all pairs (yellow). * re-sequenced to higher depth. Grey indicates the genomic sequence not covered by urine or plasma.

Table 1. Detection of ctDNA markers by qPCR in matched urine and plasma samples from patients with HCC (n=101).

	TP53	***CTNNB1***	***TERT***	**TOTAL**	**Cohen's κ**
Overall concordance (all patients/pairs) %, [concordant/total]	87.8 [72/82]	83.3 [40/48]	64.9 [50/77]	78.3 [162/207]	0.565
95% CI	78.7-94.0	69.8-92.5	53.2-75.5	72.0-83.7	
Positive concordance with at least one positive test (positive agreement) %, [concordant/total]	0.0 [0/10]	20.0 [2/10]	30.8 [12/39]	23.7 [14/59]	-0.525
95% CI	0.0-30.8	2.5-55.6	17.0-47.6	13.6-36.6	

Table 2. Germline genotype concordance between PBMC and total urine DNA.

ID	Number of bases covered at least 10X in both DNA samples	Number of concordant bases	Concordance
#1	2,455,224,425	2,455,224,425	100%
#2	2,348,155,979	2,348,155,979	100%
#3	2,443,568,212	2,443,568,212	100%
#4	2,417,859,203	2,417,859,203	100%
#5	2,120,546,181	2,120,546,181	100%
#6	2,365,129,319	2,365,129,319	100%

Figure 1

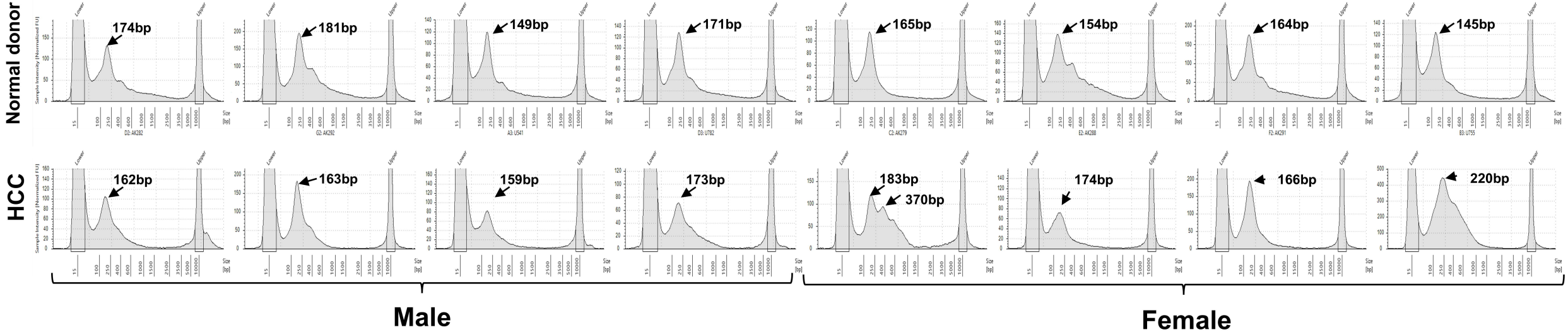
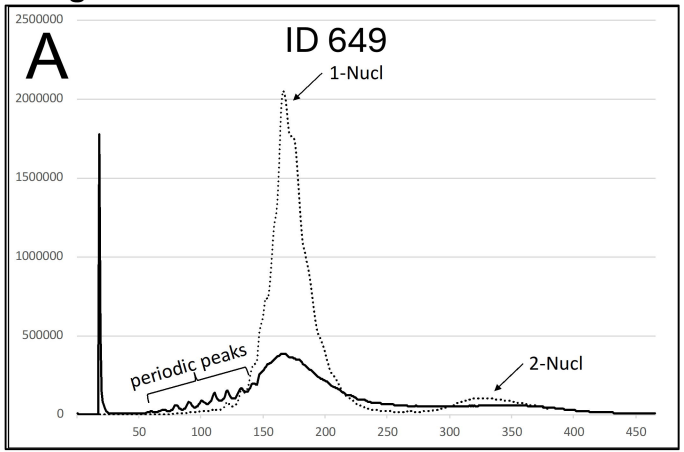
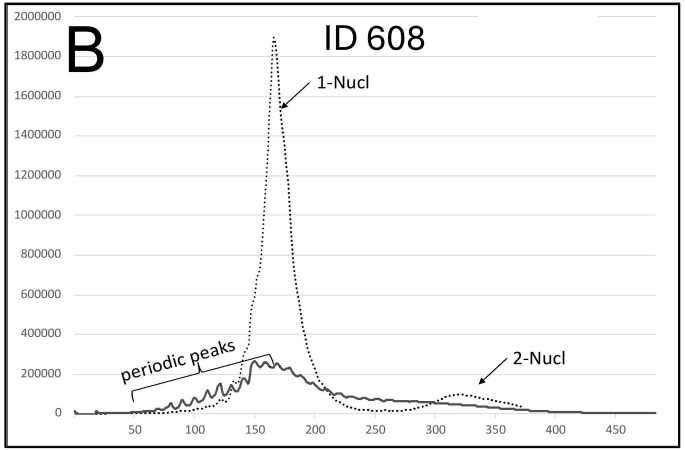
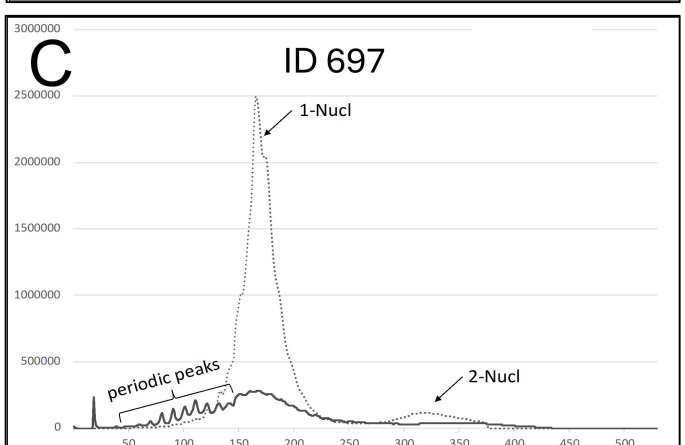
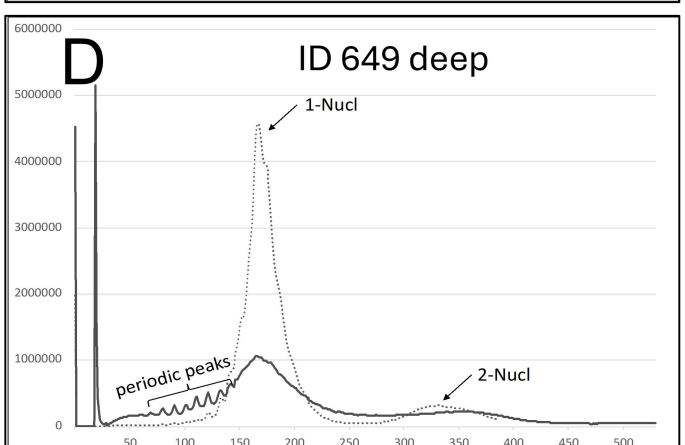

Normal donor
HCC
174bp
181bp
149bp
171bp
165bp
154bp
164bp
145bp
162bp
163bp
159bp
173bp
183bp
370bp
174bp
166bp
220bp
Male
Female

Figure 2









E

Sample Type		Insert Size bp (P25/Median/p75)	
ID608	Urine	149/ 181/ 242	
	Plasma	161/ 171/ 186	
ID649	Urine	148/ 179/ 233	
		126/ 174/ 255	*
	Plasma	163/ 174/ 190	
		162/ 174/ 193	*
ID697	Urine	137/ 172/ 216	
	Plasma	161/ 172/ 188	

Figure 3

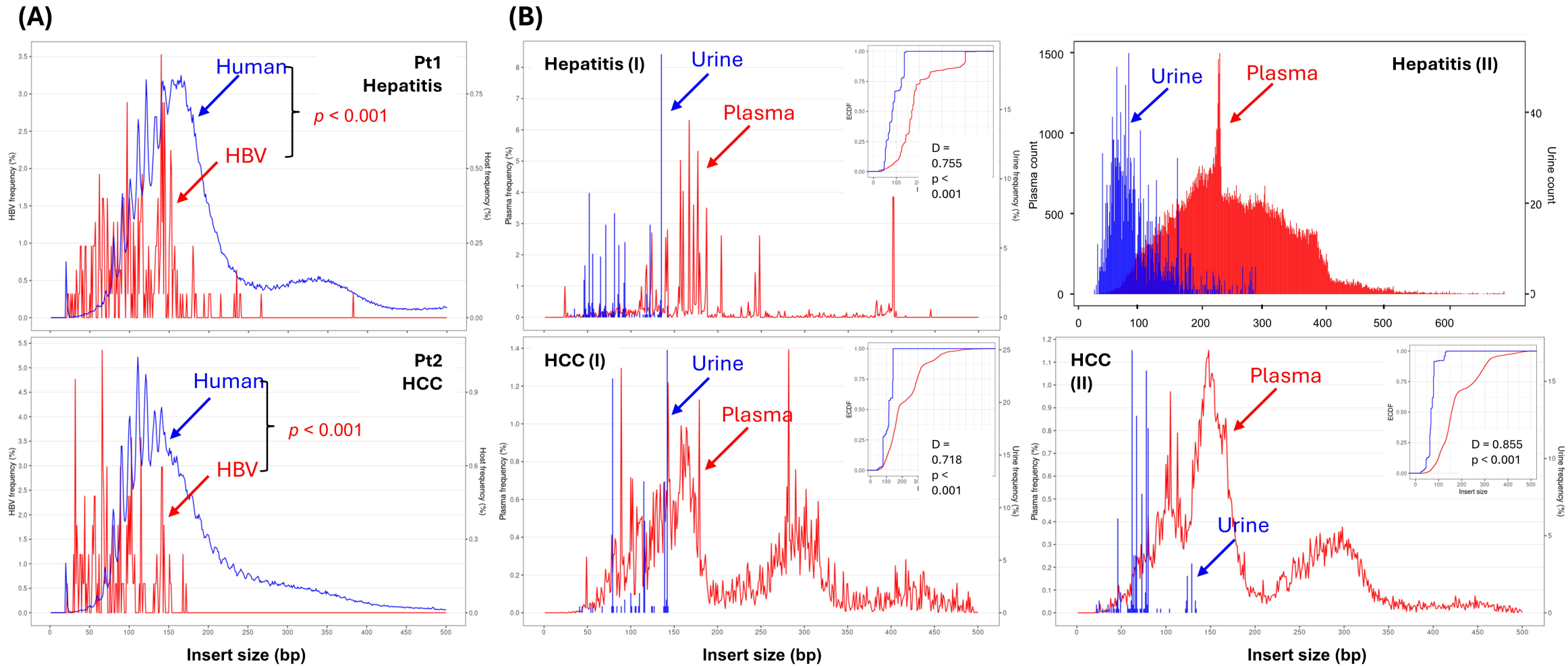
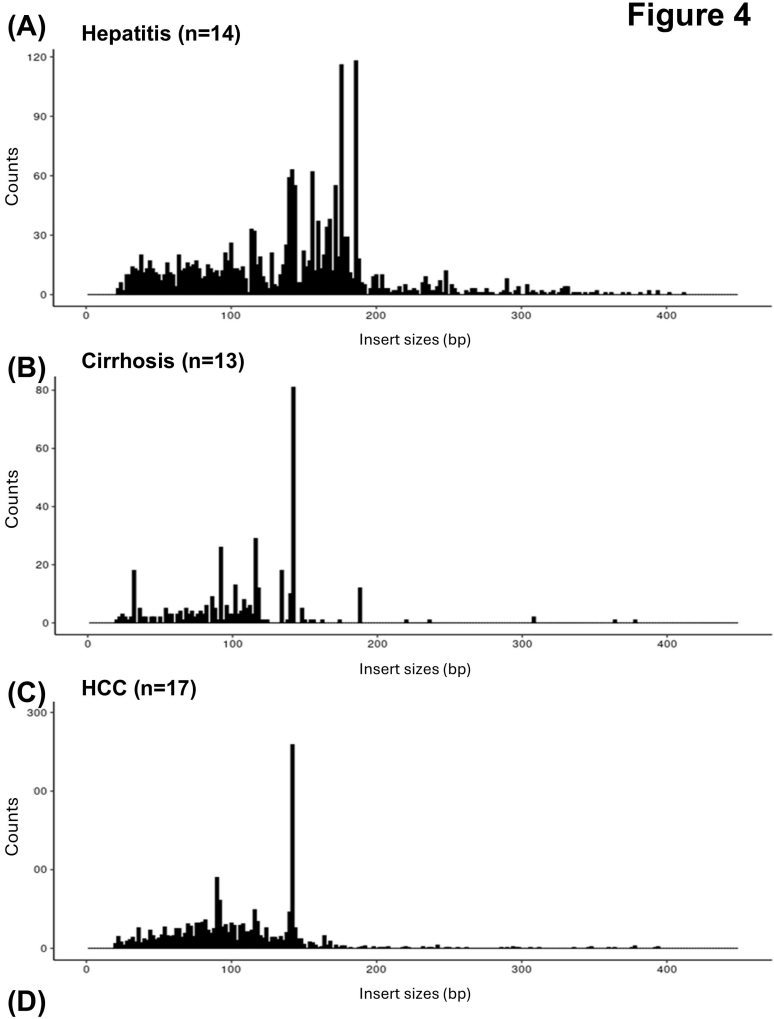

(A)
(B)
Pt1
Hepatitis
Human
HBV
p < 0.001
HBV frequency (%)
Host frequency (%)
Pt2
HCC
Human
HBV
p < 0.001
Insert size (bp)
Hepatitis (I)
Urine
Plasma
D = 0.755
p < 0.001
Plasma frequency (%)
Urine frequency (%)
ECDF
HCC (I)
Urine
Plasma
D = 0.718
p < 0.001
Hepatitis (II)
Urine
Plasma
Plasma count
Urine count
HCC (II)
Plasma
Urine
D = 0.855
p < 0.001
Insert size

Figure 4



(D)

Insert sizes (bp)	Hepatitis (n=14)	Cirrhosis (n=13)	HCC (n=17)
Min / mode / max	22 / 177 / 412	21 / 142 / 378	20 / 142 / 395
P25 / median / P75	99 / 145 / 177	85 / 113 / 142	76 / 100 / 141

Figure 5

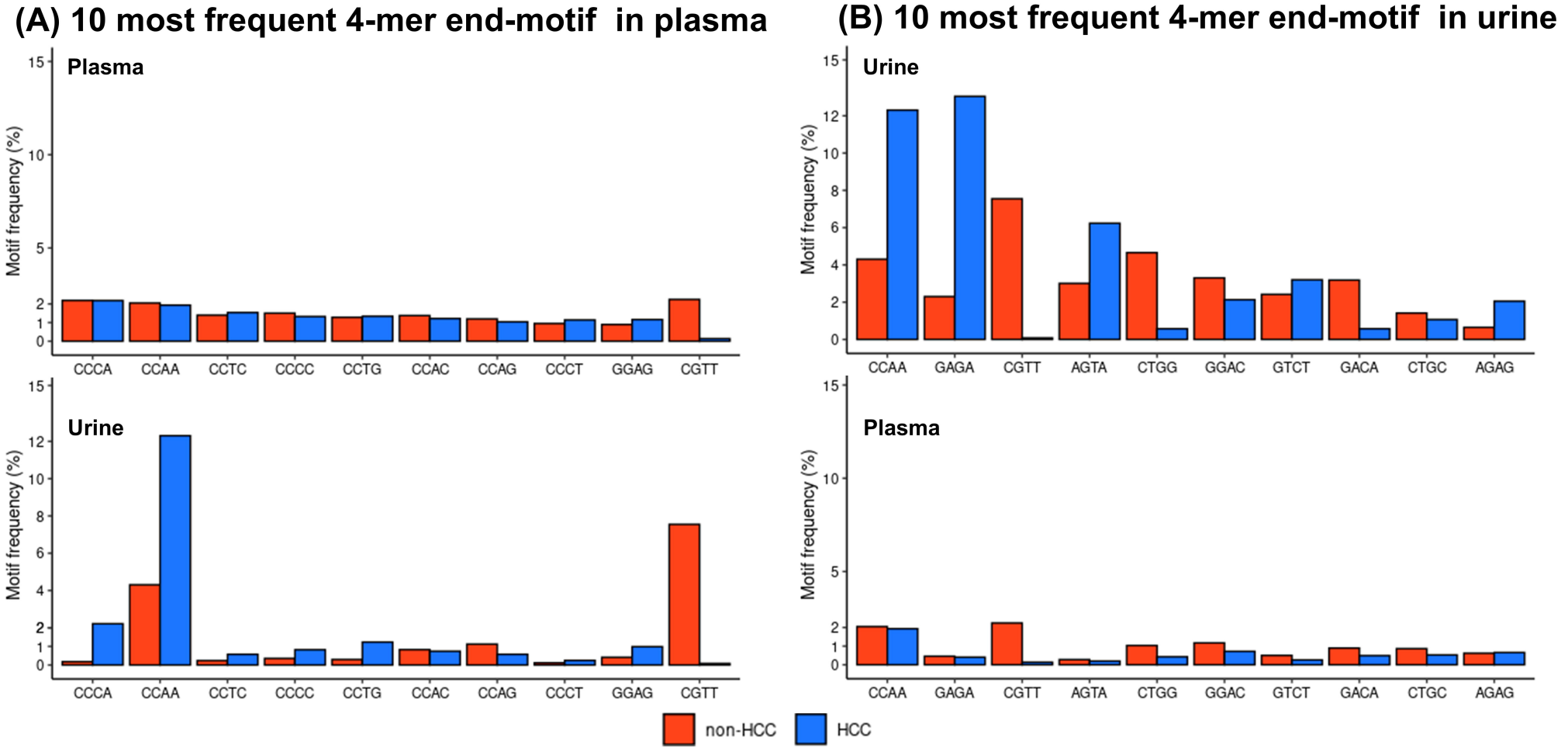
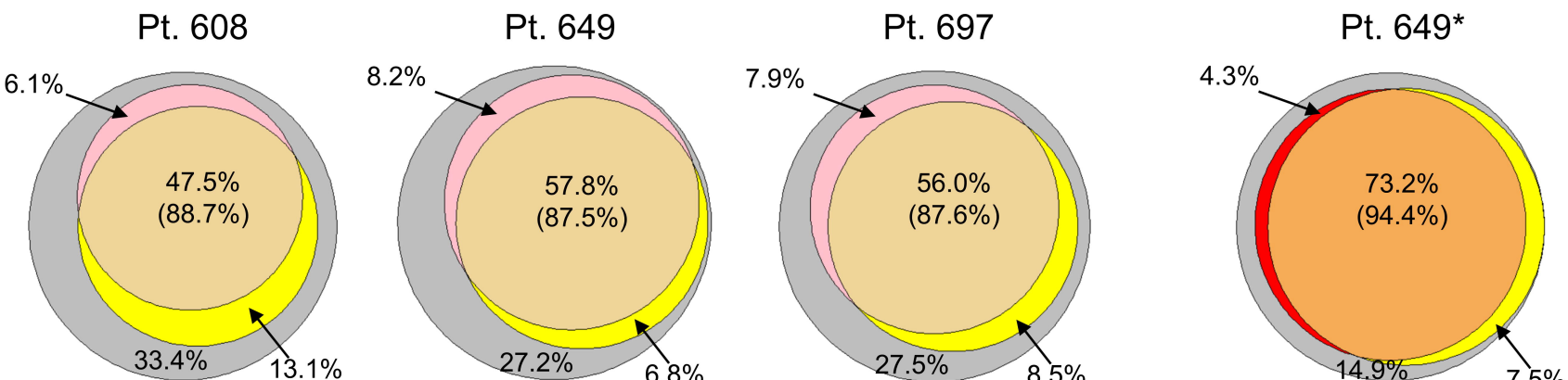

(A) 10 most frequent 4-mer end-motif in plasma
(B) 10 most frequent 4-mer end-motif in urine
Plasma
Urine
Urine
Plasma
Motif frequency (%)
CCCA
CCAA
CCTC
CCCC
CCTG
CCAC
CCAG
CCCT
GGAG
CGTT
CCAA
GAGA
CGTT
AGTA
CTGG
GGAC
GTCT
GACA
CTGC
AGAG
non-HCC
HCC

Figure 6


Pt. 608
6.1%
47.5%
(88.7%)
33.4%
13.1%
Pt. 649
8.2%
57.8%
(87.5%)
27.2%
6.8%
Pt. 697
7.9%
56.0%
(87.6%)
27.5%
8.5%
Pt. 649*
4.3%
73.2%
(94.4%)
14.9%
7.5%