Urine as a non-invasive alternative to blood for germline and somatic mutation detection in hepatocellular carcinoma

Amy K. Kim[1], Selena Y. Lin[2], Surbhi Jain[2], Yixiao Cui[3], Terence Gade[4], Fwu-Shan Shieh[2], Max Chao[2], John Shieh[2], Jonathan Cheng[2], James P. Hamilton[1], Hie-Won Hann[5], Dmitry Goryunov[2], Zhili Wang[2], Ying-Hsiu Su[3,*]

[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

*Corresponding author: Ying-Hsiu Su, PhD, 3805 Old Easton Rd, Doylestown, PA 18902, Telephone: 215-489-4907.

Email: Ying-Hsiu.Su@bblumberg.org

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

Competing Interest Statement: S.J., S.L., Z.W., and D.G. are shareholders of JBS Science Inc.

Classification: Biological Sciences – Medical Sciences.

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

This PDF file includes:

Main Text

Figures 1-5 and Tables 1-2

Supplemental Data (Figure S1 and Tables S1-S10)

Supplemental Information (files SI1 and SI2)

Data deposition: The sequence data for the subjects studied in this work who consented to data archiving have been deposited in the European Genome–Phenome Archive (EGA), www.ebi.ac.uk/ega, which is hosted by the European Bioinformatics Institute (EBI) www.ebi.ac.uk (accession no. pending).

Abstract

Cell-free DNA (cfDNA) from blood has become a promising analyte for cancer genetic liquid biopsy. Urinary cfDNA has been shown to contain mutations associated with non-genitourologic cancers including hepatocellular carcinoma (HCC). In this study, we evaluate urine as a noninvasive alternative to blood-based liquid biopsy in both germline and circulating tumor DNA (ctDNA) genotyping in HCC. Using quantitative PCR (qPCR), whole-genome sequencing (WGS), and targeted NGS, DNA isolated from blood or urine of patients with HCC was analyzed for overall genome coverage, HCC hotspot coverage, and germline or somatic mutation concordance. Targeted NGS of plasma and urine cfDNA was also performed for detection of somatic variants. We found urine cfDNA, similar to plasma cfDNA, showed a major mononucleosomal species of 150-180 bp in both healthy individuals and patients with HCC. By WGS, overall genome coverage breadth was similar between urine and plasma cfDNA, with higher fraction of covered cancer-associated mutation hotspots in urine cfDNA. qPCR analyses of HCC-associated mutations (*TP53, CTNNB1,* and *TERT*) in 101 patients with HCC revealed 78% overall concordance between plasma and urine. Targeted NGS of HCC-associated gene regions in additional 15 HCC patients showed a 97% overall position-level concordance between plasma and urine cfDNA. Collectively, urine DNA can potentially be used as a completely noninvasive liquid biopsy for HCC.

Significance Statement

Hepatocellular carcinoma (HCC) is the most common liver cancer worldwide and the fastest growing gastrointestinal cancer in the U.S. Cell-free DNA (cfDNA) which originates from various cells undergoing apoptosis or necrosis including tumor cells, is present in all body fluids levels including urine. Urinary cfDNA isolated from patients with HCC showed a similar fragment size distribution, overall genome coverage, and comparable sensitivity for detecting HCC-associated variants compared to plasma cfDNA. Urine was also determined to be a reliable source of germline genotype information, similar to peripheral blood mononuclear cells in blood-based liquid biopsies. Urine cfDNA can be used as a completely non-invasive liquid biopsy in HCC.

Introduction

Cell-free DNA (cfDNA) originates from various cells undergoing apoptosis or necrosis and is present in all body fluids including urine (1, 2). CfDNA is highly fragmented via internucleosomal nuclease cleavage, resulting in a major mononucleosomal species of 165-170 bp with lower levels of di- and trinucleosomal fragments (3, 4). Plasma isolated from blood has been the most common source of cfDNA, reflecting the mutational profiles of tumors (5, 6), albeit to varying extents (7-14) and has recently been implemented in the clinic for advanced non-small-cell lung cancer (NSCLC) tumor mutational profiling (14). We and others have shown that urinary cfDNA reflects mutations associated with non-genitourologic cancers (2, 15, 16). Urine cfDNA is a complex mixture of short DNA fragments of prerenal, renal, and postrenal origins (17, 18) of 150-250 bp containing circulation-derived transrenal low-molecular-weight (LMW; 150-250 bp) cfDNA as well as high-molecular-weight (HMW; >1,000 bp) DNA originating from exfoliating urothelial cells (15).

Hepatocellular carcinoma (HCC), despite implemented with surveillance program in a well-defined high risk population, early detection and effective treatment remain challenges resulting in poor prognosis with a 5-year survival of 20% (19). In a previous proof-of-concept study utilizing mutation-specific or methylation-sensitive qPCR, we demonstrated detectability of single-nucleotide variants (SNVs) or methylation events in four major HCC-associated genes (tumor suppressor protein 53 [*TP53*], β-catenin [*CTNNB1*], telomerase reverse transcriptase [*TERT*], and Ras association domain family 1 isoform A [*RASSF1A*]) in urinary cfDNA from patients with HCC (20). These findings suggest that urine can serve as a completely non-invasive source of genetic and epigenetic information for HCC tumors and as an alternative to plasma in liquid biopsy assays for HCC.

Limited studies have investigated urine as an alternative source of cfDNA for cancer genetic profiling (19, 21-25) and its relationship to plasma tumor-derived cfDNA (ctDNA) profiles in cancer patients. In this study, we aimed to further investigate urine as a noninvasive alternative for blood-based liquid biopsy. We used HCC as a model to study the relationship between plasma and urine cfDNA in cancer patients. We characterized fragmentation patterns of urine cfDNA using both capillary gel electrophoresis and NGS. We then employed different sequencing techniques to compare overall genome coverage, HCC hotspot coverage, and germline or somatic mutation concordance in DNA isolated from blood or urine of patients with HCC. We also performed targeted NGS of plasma and urine cfDNA for detection of somatic variants. Our results demonstrate that urine DNA can be used for both germline genotyping and somatic mutation detection in patients with HCC. A noninvasive liquid biopsy using urine as a source for HCC genetic profiling may offer insight to drug development and early detection through more frequent sampling and monitoring.

Results

Urinary cfDNA size distribution

Electrophoretic profiling

To determine the size distribution of urine cfDNA, we performed capillary electrophoresis of LMW DNA isolated from urine samples of eight healthy individuals and eight patients with HCC. As anticipated, mononucleosomal DNA (150-180 bp) was the predominant cfDNA species in all urine samples (Fig. 1A). Most samples also contained a minor dinucleosomal peak or “shoulder”. In subsequent NGS analyses, three sets of urinary

and plasma cfDNA isolates derived from two HCC patients showed similar profiles (Fig. 1B and 1C).

WGS insert size analysis

Next, we performed shallow WGS (sWGS) of LMW DNA isolated from three sets of matched urine and plasma of these two patients with HCC (Table S1). Samples 649 and 697 were collected from the same patient eight months apart. Samples 649 from both urine and plasma were also re-sequenced at a higher depth. The WGS data are summarized in Table S2.

All plasma samples showed a major peak of 170-180 bp, corresponding to the mononucleosomal fragment size (Fig. 2A-D). A minor dinucleosomal peak was also present. In urine, a similar mononucleosomal peak was also observed in all libraries, but no dinucleosomal peaks were discernable. In addition, both plasma and urine insert size distributions showed a subnucleosomal “sawtooth” pattern with 10 bp periodicity. The pattern was more pronounced in the urine samples. Consistent with this observation, the insert size distribution in the urine samples was wider compared to that in the plasma libraries (Fig. 2E). When sequenced to a higher depth, fragment size distributions in both urine and plasma had no notable differences compared to sWGS. (Fig. 2A and 2D).

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

Overall genome coverage

After unique molecular identifier (UMI)-based consensus building, the total fractions of genomic positions covered at least once in urine and plasma cfDNA were similar within each pair (ID608: 60.6% vs. 53.5%; ID649: 64.6% vs. 66.0%; ID697: 64.6% vs. 64.0%;

ID649 deep: 80.8% vs. 77.6%, respectively; Table S2). 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 3). 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, whereas the fraction of urine-only coverage was 7-13% across all pairs. Moreover, despite the cfDNA inputs of the plasma libraries exceeding those of the urine libraries 2- to 8-fold, the comparable overall coverage and high coverage overlap between plasma and urine were achieved.

Cancer hotspot coverage

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 (190 target regions in 50 genes) (26). The hotspot coverage in urine was comparable to, or somewhat higher than, the corresponding plasma coverage in each pair: 83.0% vs. 74.6% in sample pair 608, 84.3% vs. 85.9% in 649, and 85.6% vs 83.5% in ID697 urine vs. plasma, respectively (Fig 4A). 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 (21, 23, 24).

We then visualized the hotspot coverage in all three sample pairs using Circos plots (Fig. 4B). 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 next 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 (27, 28). 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) (21). These assays can detect their respective target mutations at allele frequencies as low as 0.1% (*TP53*), 0.3% (*CTNNB1*), and 0.15% (*TERT*). Urine and plasma cfDNA samples isolated from a total of 101 patients with HCC were screened. The results are detailed in Table S3 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 S4).

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 urine cfDNA for somatic mutation detection in 15 patients with HCC (Table S5). A 24-gene panel with a total target size of approximately 64 kb was used to capture HCC-associated gene regions from matched urine or plasma (Supplementary file 1). 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 S6). 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 S7). 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]; Fig. 5A). The correlation was even stronger for position-level mean VAFs across

the entire cohort (Pearson r=0.995 [95% CI 0.995-0.995]; Fig. 5B). As expected, putative somatic VAFs (<0.4) correlated less strongly (Pearson r=0.890 [95% CI 0.884-0.895]; Fig. 5C).

Among the 24 targets, *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 S8). Putative somatic mutations with 100% sample-level (n=30) frequencies were found in *ARID1, TERT* promoter, *PCLO, KMT2D, AXIN1, TSC2, TP53*, and *KEAP1* (Supplemental Information file 2). 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 peripheral blood mononuclear cells (PBMCs) to determine the patient's germline genotype. In addition to cfDNA, urine contains exfoliated urothelial cells as well as some leukocytes (18), 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 S9). 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

Numerous studies have shown that plasma cfDNA may reflect the mutational burden in many cancers, including HCC (7, 29-36). In the present study, we demonstrated that urinary cfDNA isolated with our previously developed method (15) from both healthy individuals and patients with HCC has a fragment size distribution similar to that observed in plasma cfDNA. In both plasma and urine, the predominant electrophoretic species of 160-180 bp corresponded to the average size of fragments protected from degradation by the histone core of the nucleosome. Furthermore, urine cfDNA produced overall genome coverage similar to that obtained with plasma cfDNA in patients with HCC, with mostly overlapped positions. HCC-associated mutations were also detectable at similar rates overall between urine and plasma cfDNA by qPCR and targeted NGS. Targeted NGS also revealed high concordance between sequence variants detected in the two body fluids, confirming the feasibility of using urine for somatic mutation detection in HCC. These findings strongly suggest that urine cfDNA can be used as an alternative or complement to plasma for tumor mutation analyses of HCC patients and it also may be a richer source of HCC ctDNA compared to plasma, similar to findings in urothelial cancers (37, 38) and NSCLC (39).

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 concordance was obtained between urine DNA and the matched PBMC when genome coverage was compared. Thus, similar to PBMC-derived DNA in blood-based liquid biopsy, total urinary DNA can potentially be used to extract germline information.

Concordance between plasma and urine ctDNA mutational profiles is often used to evaluate the feasibility of using urine as a source of genetic information for the diagnosis and monitoring of a non-genitourinary cancer. Using mutation-specific qPCR assays, we found a high overall concordance of 78% in 101 patients screened for three HCC hotspots (*TERT* c.-124, *TP53* codon 249, and *CTNNB1* codons 32-37), 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 (40). This is in contrast with other plasma-urine *EGFR* T790M positive agreement reports in other cancers such as NSCLC (23, 36) and bladder (41). The apparent discrepancies in urine/plasma positive agreement between these studies and our results may be attributable to the different sensitivities of the methods, different urinary cfDNA isolation method, or different genes of interest. 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 increased the limit of detection in our study.

Our study has several limitations. First, tumor tissue mutational profiles were unavailable for most patients, preventing any analyses of concordance between tissue and plasma/urine DNA. Second, the amounts of plasma and urine cfDNA used for WGS (but not targeted NGS) library preparation were different, precluding rigorous

comparison of some WGS metrics and underlying biological phenomena. Third, the number of patients with HCC analyzed by NGS was relatively small. Fourth, 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 (42). 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 using targeted NGS sequencing as well as different DNA isolation techniques.

In conclusion, urinary cfDNA produces similar overall genome coverage compared to plasma cfDNA in patients with HCC, as well as similar rates of HCC-associated mutations using both qPCR and targeted NGS. It is also a reliable source of germline genotype information. Taken together, our results provide further evidence of non-redundant utility of urinary cfDNA in HCC genetic liquid biopsy.

Materials and Methods

Patients and body fluid collection

Detailed patient information is summarized in Tables S1 and S5. Patient specimen collections were approved by each study center's institutional review board. Urine was mixed with EDTA (10 mM final concentration) immediately upon collection and stored at -80 °C until ready for isolation. Blood was collected in K_2EDTA BD Vacutainer tubes (BD, Franklin Lakes, NJ) and processed to obtain plasma and PBMC.

DNA isolation, quantitation, and size distribution analysis

DNA from urine samples was extracted as previously described (15). The LMW fraction (cfDNA; < 1 kb) was obtained by size selection 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 JPurX plasma cell-free DNA isolation kit (JBS Science, Inc, Doylestown, PA). DNA concentrations were determined with a Qubit fluorometer using High Sensitivity dsDNA reagents (Thermo Fisher Scientific, Waltham, MA). Equal volumes of purified cfDNA samples were loaded on a D5000 High Sensitivity screentape and electrophoretically resolved on a 2200 TapeStation system (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

First, the total copy number of each target gene was determined for each DNA sample using qPCR assays (JBS Science) according to manufacturer's instructions. Only urine DNA samples that contained more than 100 copies/mL were used for analysis. Mutation detection in each of the three markers of interest was then performed in duplicate with the corresponding JBS mutation assay (JBS Science). The maximum DNA input recommended by the manufacturer was used in each assay with plasma DNA. Because the available amounts of urine DNA were limited, in some pairs the urine DNA input was less than that in the matched plasma sample. However, such pairs were included in the analysis only if the urine sample generated a positive result, thus minimizing the confounding effect of varying inputs.

NGS library preparation and sequencing

For WGS, fragmented DNA (untreated cfDNA or sonicated PBMC/total urine DNA) was end-repaired, A-tailed, and ligated to adapters overnight at 16 °C. Double-stranded adapters containing a 6 bp i7 sample index or an 8 bp i5 index followed by a 9 bp random UMI were used. Adapter oligos (Table S10) were annealed by slowly cooling (0.1°C/s) equimolar oligo mixtures heat-treated at 95 °C for 5 min. MagBio beads with isopropanol were used for cleanup after end repair and 3'A-tailing, whereas AMPure beads (Agencourt Bioscience; Beverly, MA) with ethanol were used for post-ligation cleanup. Ligated libraries were PCR-amplified for 15 cycles with Herculase polymerase (Agilent), purified, and quantified by qPCR and the Qubit fluorometer. Final libraries were pooled into a total volume of 20 □L at a combined final concentration of 20 mM, denatured with sodium hydroxide, and sequenced on a NextSeq550 or HiSeq4000 instrument (2×150 bp; Genomics Core Facility at Drexel College of Medicine, Philadelphia, PA). The PBMC/urine DNA inputs ranged from 30 to 50 ng.

For targeted sequencing, 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). Purified libraries were hybridized with a 24-gene hybridization panel (Supplementary file 1) in the presence of xGen Universal Blockers using the xGen Hybridization and Wash Kit (both IDT) according to the manufacturer's protocol. Captured libraries were amplified with KAPA HiFi polymerase (Roche, Wilmington, MA), purified with AMPure beads, and sequenced on a NovaSeq S1 flow cell (Psomagen, Rockville, MD).

NGS data analysis

Raw data were demultiplexed and each paired-end read 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 normalized (downsampled) within each plasma/urine pair. Most of the UMI processing, including coordinate/UMI family consensus building and error correction, was performed with various tools in the fgbio, connor, and picard suites (See Figure S2 for workflow details). A minimum UMI family size of 3 was used in most analyses. Consensus reads were realigned, quality-filtered, and used to generate VCF files where applicable. The VCF files were filtered for common germline variants (dbSNP138) and used to evaluate mutant variant frequencies.

Acknowledgments

This study was supported by NIH 1K08CA237624-01A1, R44HG008700, and R01CA202769.

References

1. Schwarzenbach H, Hoon DS, & Pantel K (2011) Cell-free nucleic acids as biomarkers in cancer patients. *Nature Reviews Cancer* 11(6):426-437.
2. Jain S, Lin SY, Song W, & Su YH (2019) Urine-Based Liquid Biopsy for Nonurological Cancers. *Genet Test Mol Biomarkers* 23(4):277-283.
3. Rumore PM & Steinman CR (1990) Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone. *J Clin Invest* 86(1):69-74.
4. Stroun M, Anker P, Lyautey J, Lederrey C, & Maurice PA (1987) Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol* 23(6):707-712.
5. Jiang P*, et al.* (2015) Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. 112(11):E1317-E1325.
6. Jiang P*, et al.* (2018) Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. 115(46):E10925-E10933.
7. Li Z*, et al.* (2019) Early recurrence detected in hepatocellular carcinoma patients after transcatheter arterial chemoembolization treatment with plasma cell-free DNA. *Eur J Gastroenterol Hepatol* 31(7):885-892.
8. Stover DG*, et al.* (2018) Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *J Clin Oncol* 36(6):543-553.
9. Adam Z, Tai-Jung L, Surbhi J, & Ying-Hsiu S (2018) Urine as an Alternative to Blood for Cancer Liquid Biopsy and Precision Medicine. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*:2820-2525.
10. Dhayat SA & Yang Z (2020) Impact of circulating tumor DNA in hepatocellular and pancreatic carcinomas. *J Cancer Res Clin Oncol* 146(7):1625-1645.
11. Yang JD, Liu MC, & Kisiel JB (2019) Circulating Tumor DNA and Hepatocellular Carcinoma. *Semin Liver Dis* 39(4):452-462.
12. Qu C*, et al.* (2019) Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy. *Proc Natl Acad Sci U S A* 116(13):6308-6312.
13. Wu X*, et al.* (2020) Circulating tumor DNA as an emerging liquid biopsy biomarker for early diagnosis and therapeutic monitoring in hepatocellular carcinoma. *Int J Biol Sci* 16(9):1551-1562.
14. Zhang S*, et al.* (2018) Epidermal growth factor receptor (EGFR) T790M mutation identified in plasma indicates failure sites and predicts clinical prognosis in non-small cell lung cancer progression during first-generation tyrosine kinase inhibitor therapy: a prospective observational study. *Cancer Commun (Lond)* 38(1):28.
15. Su YH*, et al.* (2004) 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. *J Mol Diagn* 6(2):101-107.
16. Su YH, Wang M, Aiamkitsumrit B, Brenner DE, & Block TM (2005) Detection of a K-ras mutation in urine of patients with colorectal cancer. *Cancer Biomark* 1(2-3):177-182.

17. Su YH, *et al.* (2008) 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* 1137:82-91.
18. Lin SY, Linehan JA, Wilson TG, & Hoon DS (2017) Emerging Utility of Urinary Cell-free Nucleic Acid Biomarkers for Prostate, Bladder, and Renal Cancers. *European urology focus*.
19. Chen VL, Xu D, Wicha MS, Lok AS, & Parikh ND (2020) Utility of Liquid Biopsy Analysis in Detection of Hepatocellular Carcinoma, Determination of Prognosis, and Disease Monitoring: A Systematic Review. *Clin Gastroenterol Hepatol*.
20. Jain S, *et al.* (2015) Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. *Hepatol Res* 45(11):1110-1123.
21. Lin SY, *et al.* (2011) A locked nucleic acid clamp-mediated PCR assay for detection of a p53 codon 249 hotspot mutation in urine. *J Mol Diagn* 13(5):474-484.
22. Lin SY, Linehan JA, Wilson TG, & Hoon DSB (2017) Emerging Utility of Urinary Cell-free Nucleic Acid Biomarkers for Prostate, Bladder, and Renal Cancers. *Eur Urol Focus* 3(2-3):265-272.
23. Chen S, Zhao J, Cui L, & Liu Y (2017) Urinary circulating DNA detection for dynamic tracking of EGFR mutations for NSCLC patients treated with EGFR-TKIs. *Clin Transl Oncol* 19(3):332-340.
24. Song BP, *et al.* (2012) Detection of hypermethylated vimentin in urine of patients with colorectal cancer. *J Mol Diagn* 14(2):112-119.
25. Cheng THT, *et al.* (2019) Noninvasive Detection of Bladder Cancer by Shallow-Depth Genome-Wide Bisulfite Sequencing of Urinary Cell-Free DNA for Methylation and Copy Number Profiling. *Clin Chem* 65(7):927-936.
26. Leong KW, *et al.* (2019) A nuclease-polymerase chain reaction enables amplification of probes used for capture-based DNA target enrichment. *Nucleic Acids Res* 47(22):e147.
27. Huang A, *et al.* (2016) Detecting Circulating Tumor DNA in Hepatocellular Carcinoma Patients Using Droplet Digital PCR Is Feasible and Reflects Intratumoral Heterogeneity. *J Cancer* 7(13):1907-1914.
28. Jiao J, *et al.* (2018) Telomerase reverse transcriptase mutations in plasma DNA in patients with hepatocellular carcinoma or cirrhosis: Prevalence and risk factors. *Hepatol Commun* 2(6):718-731.
29. Dudley JC, *et al.* (2019) Detection and Surveillance of Bladder Cancer Using Urine Tumor DNA. *Cancer Discov* 9(4):500-509.
30. An Y, *et al.* (2019) The diagnostic and prognostic usage of circulating tumor DNA in operable hepatocellular carcinoma. *Am J Transl Res* 11(10):6462-6474.
31. Christensen E, *et al.* (2018) Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. *Sci Rep* 8(1):1917.
32. Howell J, *et al.* (2019) Identification of mutations in circulating cell-free tumour DNA as a biomarker in hepatocellular carcinoma. *Eur J Cancer* 116:56-66.
33. Kaseb AO, *et al.* (2019) Molecular Profiling of Hepatocellular Carcinoma Using Circulating Cell-Free DNA. *Clin Cancer Res* 25(20):6107-6118.
34. Labgaa I, *et al.* (2018) A pilot study of ultra-deep targeted sequencing of plasma DNA identifies driver mutations in hepatocellular carcinoma. *Oncogene* 37(27):3740-3752.

35. Liao W, *et al.* (2016) Noninvasive detection of tumor-associated mutations from circulating cell-free DNA in hepatocellular carcinoma patients by targeted deep sequencing. *Oncotarget* 7(26):40481-40490.
36. Reckamp KL, *et al.* (2016) A Highly Sensitive and Quantitative Test Platform for Detection of NSCLC EGFR Mutations in Urine and Plasma. *J Thorac Oncol* 11(10):1690-1700.
37. Avogbe PH, *et al.* (2019) Urinary TERT promoter mutations as non-invasive biomarkers for the comprehensive detection of urothelial cancer. *EBioMedicine* 44:431-438.
38. Ou Z, *et al.* (2020) Detection of bladder cancer using urinary cell-free DNA and cellular DNA. *Clin Transl Med* 9(1):4.
39. Yu H, Liu M, Qiu H, & Yang K (2019) Urinary and Plasma Cell-Free DNA Comparison for Lung Cancer Patients Treated With Epidermal Growth Factor Receptor-Thyroxine Kinase Inhibitors. *Am J Med Sci* 357(1):29-36.
40. Wu Z, *et al.* (2019) Differences in the genomic profiles of cell-free DNA between plasma, sputum, urine, and tumor tissue in advanced NSCLC. *Cancer Med* 8(3):910-919.
41. Christensen E, *et al.* (2017) Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur Urol* 71(6):961-969.
42. Melkonyan HS, *et al.* (2008) Transrenal nucleic acids: from proof of principle to clinical tests. *Ann N Y Acad Sci* 1137:73-81.

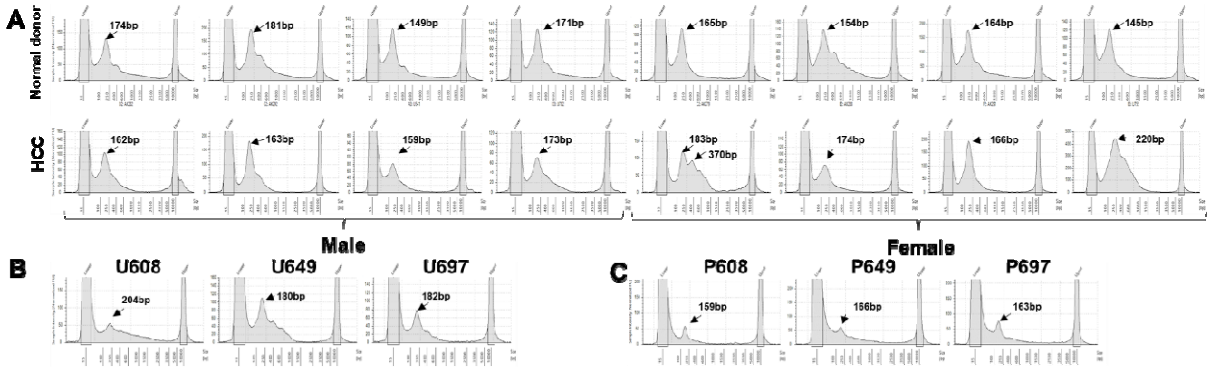
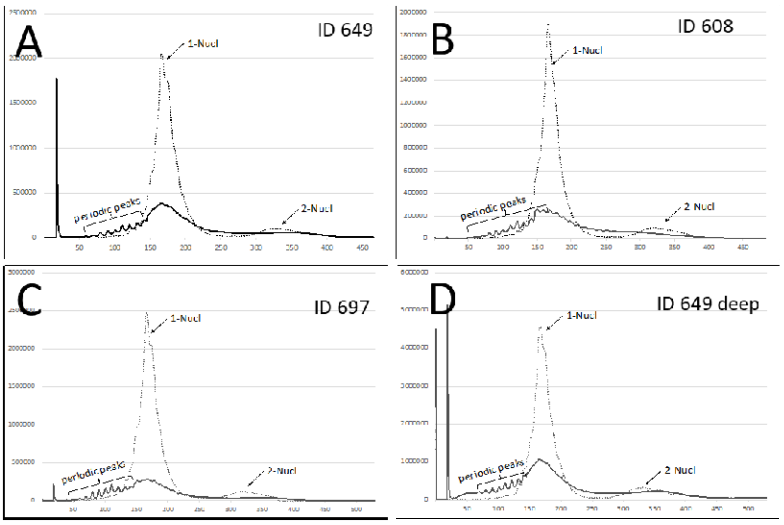


Figure 1. Electrophoretic size profiling of LMW DNA isolated from urine and plasma. (A) Eight healthy and eight HCC urine LMW DNA samples were resolved on a TapeStation as described in Materials in Methods. Healthy (top) and HCC (bottom) sample panels are indicated. Equivalents of approximately 0.1 mL of healthy-donor urine and 0.057 mL of HCC-patient urine were loaded. Matched urinary (B) or plasma (C) cfDNA samples from two patients with HCC were separated (equivalents of 0.033 mL and 0.0175 mL of urine or plasma, respectively). Plasma DNA was size-selected (<500 bp) prior to analysis. Samples 649 and 697 were collected from the same patient 8 months apart (Table S1).



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 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 25th/75th percentile insert sizes for each sample. In ID649, the bottom urine/plasma rows (*) are for the deeper WGS run. See Table S2 for details.

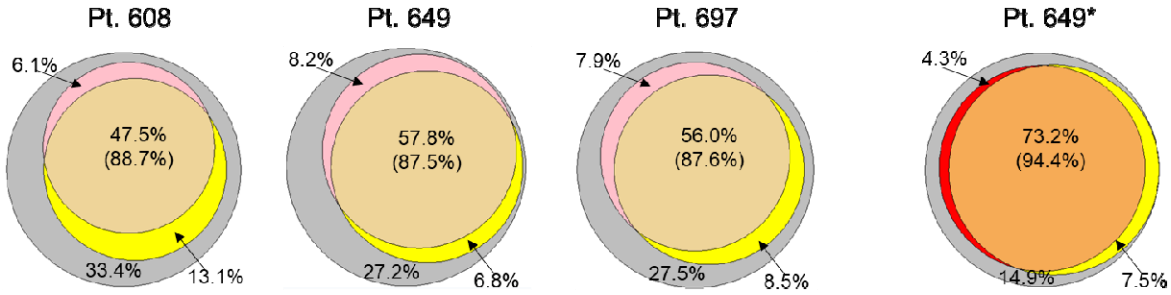


Figure 3. Genomic coverage of LMW cfDNA isolated from matched urine (yellow) and plasma (pink/red). Euler diagram of genomic positions in urine and plasma covered in each pair. The urine/plasma overlap ranged from 48% to 58% of the genome in the three shallow-sequenced sample pairs and increased from 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.

A

	Pt608		Pt649		Pt697	
	Plasma	Urine	Plasma	Urine	Plasma	Urine
Avg. Depth (X)	5.6	3.5	7.4/23.4	5.3/15.9	5.3	3.8
Coverage (%)	74.6	83.0	86.9/ 94.6	84.3/ 94.5	83.5	85.6

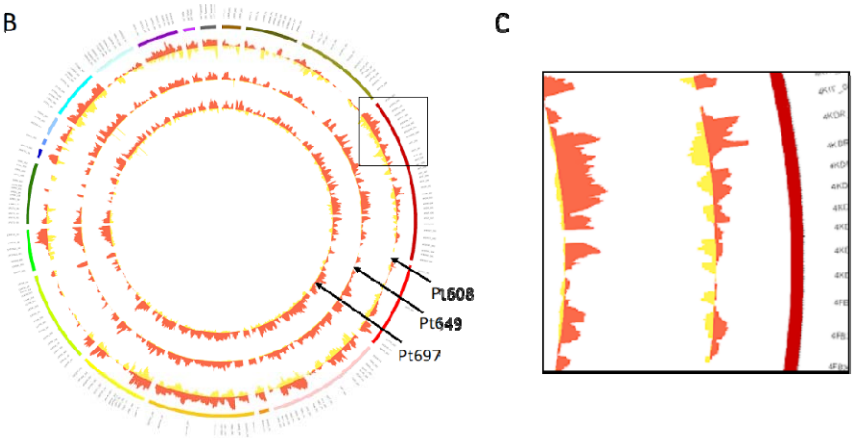


Figure 4. Comparison of cancer hotspot coverage between matched urine and plasma cfDNA from patients with HCC. (A) Summary of sWGS hotspot coverage in three matched urine/plasma pairs. In samples 649, the second number represents data from the deep-sequencing run. (B) Circos plot depicting the comparison of shallow WGS hotspot coverage profiles of three matched plasma (red) and urine (yellow) cfDNA sample pairs. The insert in the center is a zoom-in of the area in the rectangle.

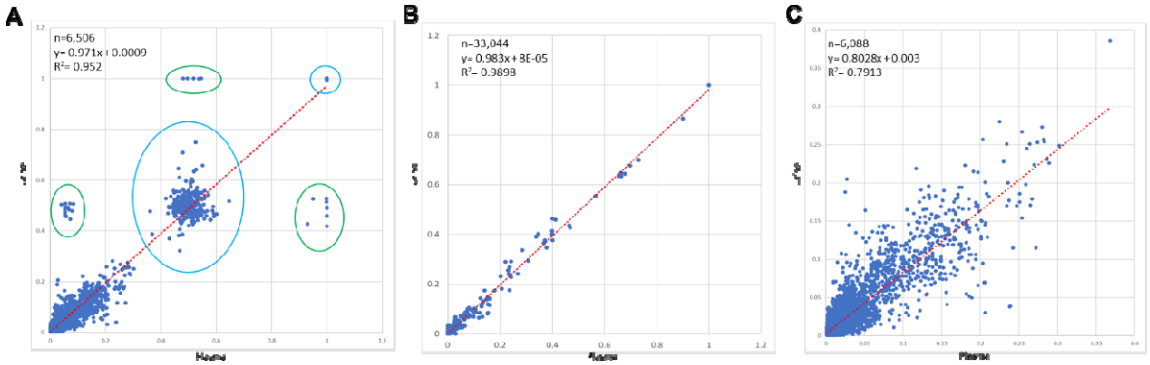


Figure 5. Correlation of all (A), cohort mean (B), and likely somatic (C) variant allele frequencies detected by targeted NGS in plasma or urine cfDNA of patients with HCC. Green circles indicate variants likely attributable to early embryonic somatic mutations and originating from circulating or uroepithelial cell gDNA. Blue circles indicate likely germline variants. Only variants supported by ≥3 unique reads (UMI families) were included in the analyses. n, number of double-positive positions.

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

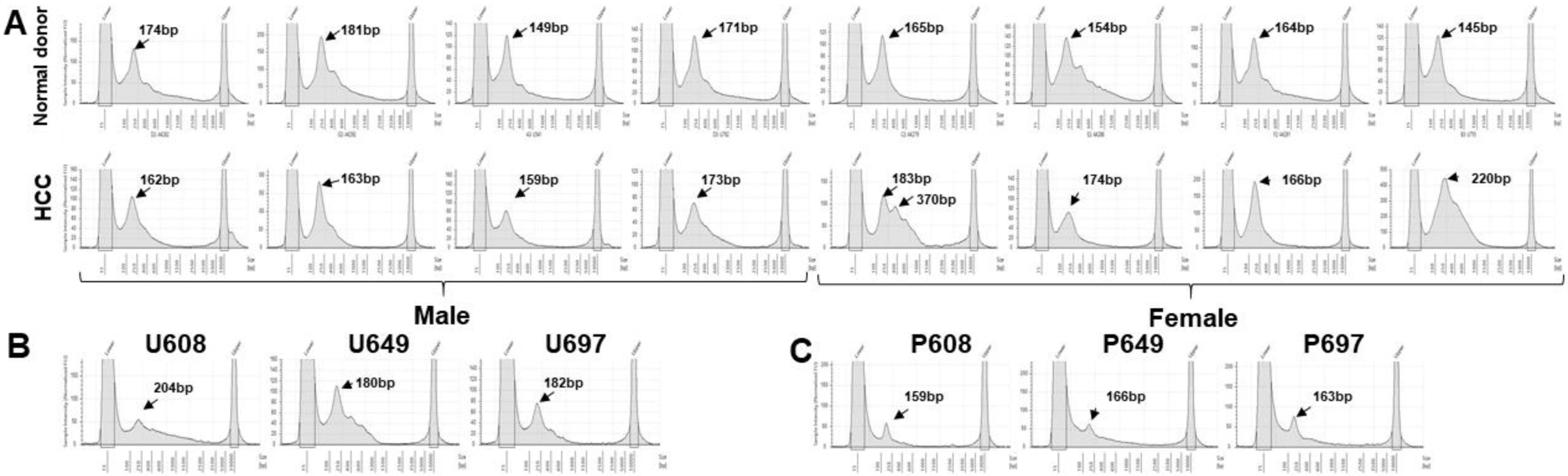
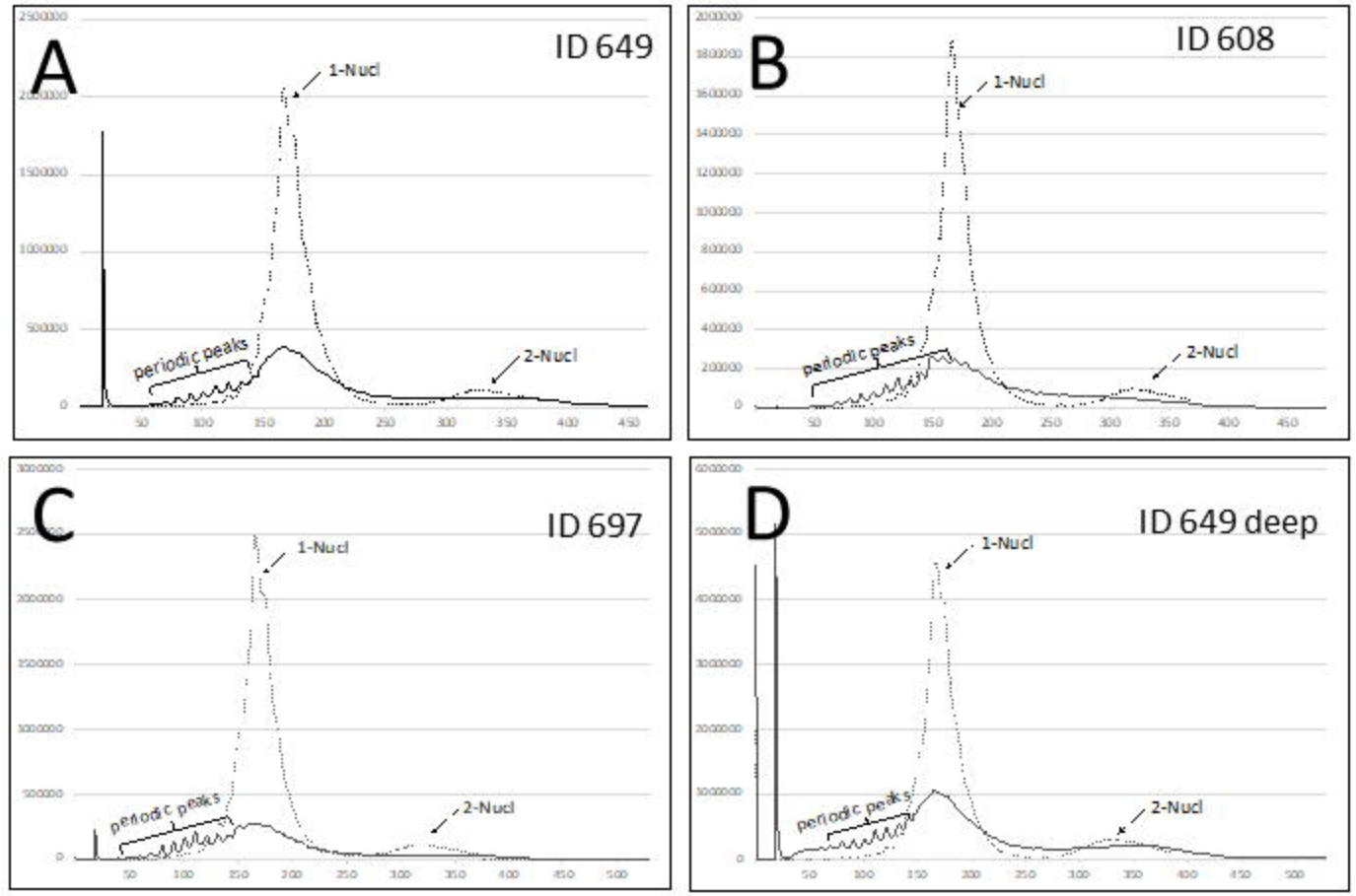

A
Normal donor
174bp
181bp
149bp
171bp
165bp
154bp
164bp
145bp
HCC
162bp
163bp
159bp
173bp
183bp
370bp
174bp
166bp
220bp
Male
Female
B
U608
204bp
U649
180bp
U697
182bp
C
P608
159bp
P649
166bp
P697
163bp

Figure 2



F

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

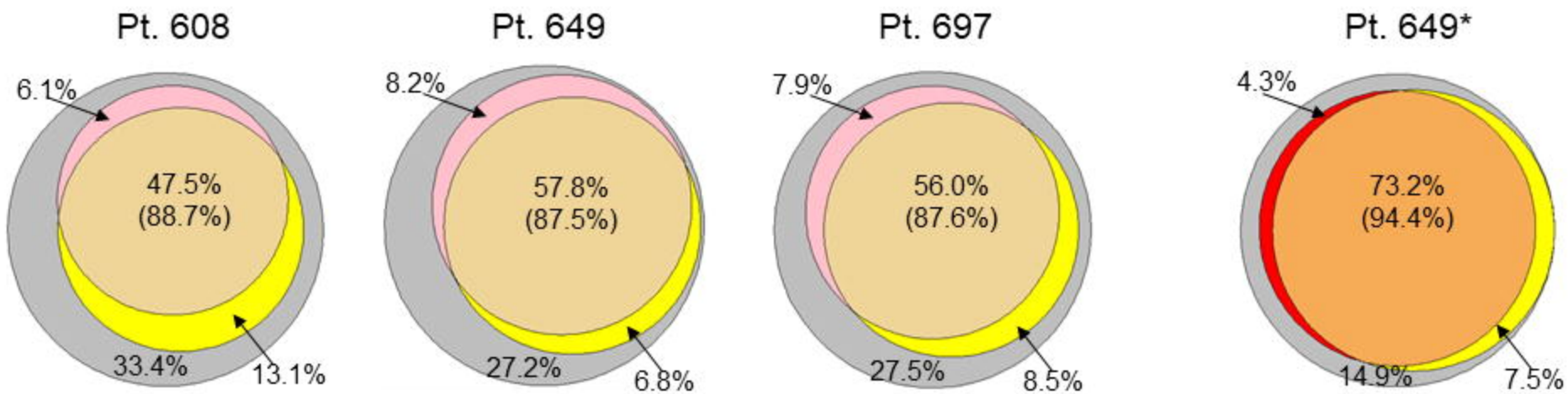
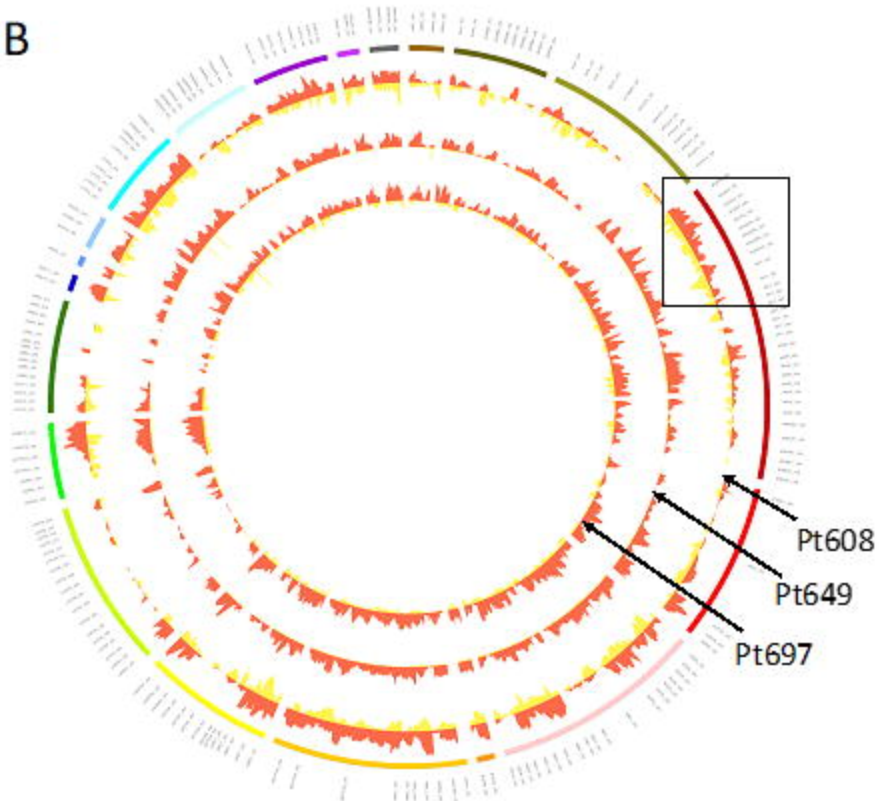

Pt. 608
6.1%
47.5%
(88.7%)
33.4%
13.1%
Pt. 649
8.2%
57.8%
(87.5%)
27.2%
6.8%
Pt. 697
7.9%
56.0%
(87.6%)
27.5%
8.5%
Pt. 649*
4.3%
73.2%
(94.4%)
14.9%
7.5%

Figure 4

A

	Pt608		Pt649		Pt697	
	Plasma	Urine	Plasma	Urine	Plasma	Urine
Avg. Depth (X)	5.6	3.5	7.4/23.4	5.3/15.9	5.3	3.8
Coverage (%)	74.6	83.0	85.9/ 94.5	84.3/ 94.6	83.5	85.6

B



C

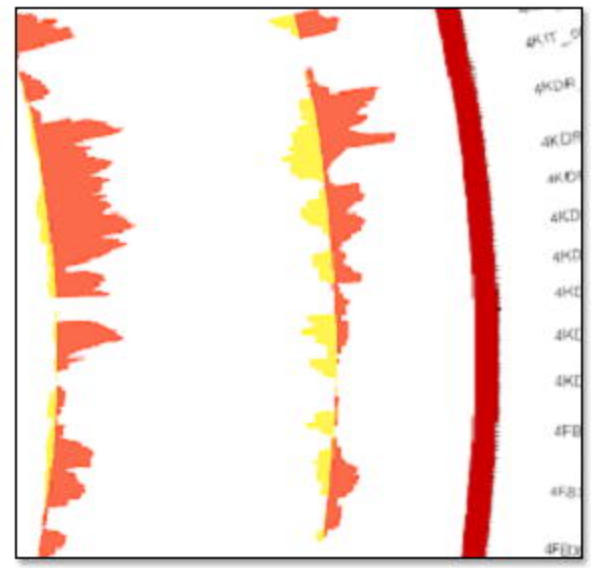
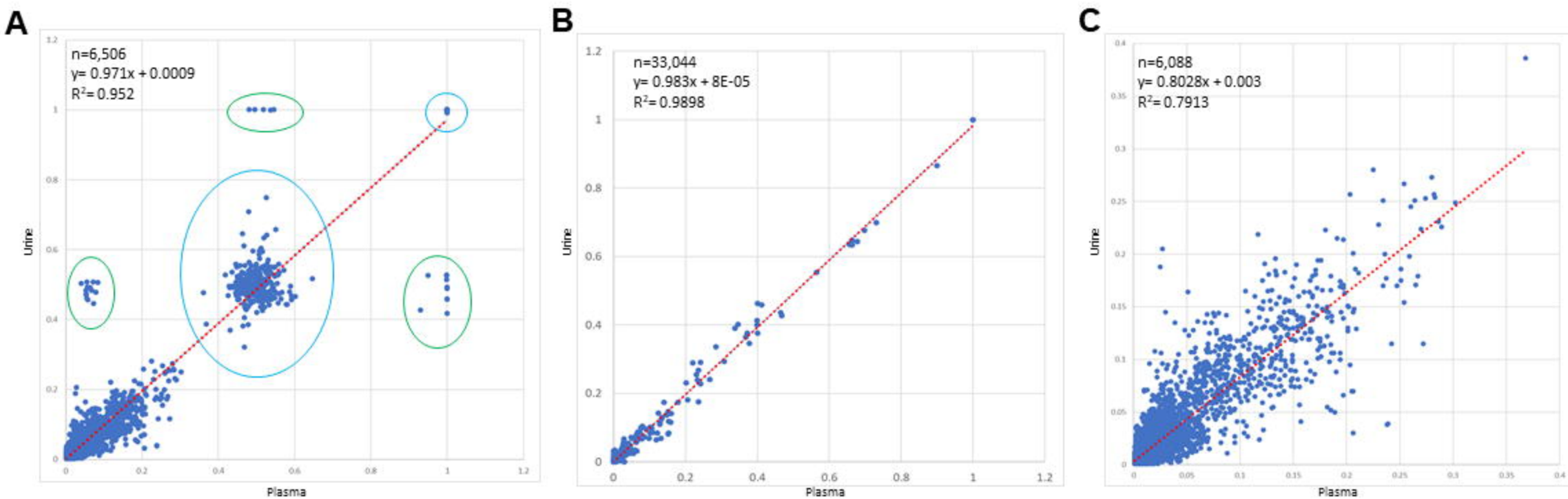


Figure 5


A
n=6,506
y= 0.971x + 0.0009
R²= 0.952
Urine
Plasma
B
n=33,044
y= 0.983x + 8E-05
R²= 0.9898
Urine
Plasma
C
n=6,088
y= 0.8028x + 0.003
R²= 0.7913
Urine
Plasma