ChIP-seq of Urinary Cell-Free Chromatin Infers Tissue Origin and Detects Tumor-Derived Cell-Free DNA in Bladder Cancer

Author list: Matan Lotem[1,2], Israa Sharkia[1,2], Batia Azria[3], Tal Michaeli[3,4], Nir Friedman[1,2*]

1.The Lautenberg Center for Immunology and Cancer Research; Jerusalem, Israel

2.The Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University; Jerusalem, Israel

3.Department of Developmental Biology and Cancer Research Institute for Medical Research Israel-Canada Hebrew University Medical School; Jerusalem, Israel

4.Sharett Institute of Oncology Hadassah-Hebrew University Medical Center; Jerusalem, Israel

*Corresponding author. Email: nir.friedman@mail.huji.ac.il

Abstract

Urinary Cell-free DNA (cfDNA) holds valuable molecular insights into the processes occurring within the urinary system. The analysis of epigenetic markers such as histone post-translational modifications in both cells and in blood circulation offers detailed information regarding cell identity and underlying state. However it remains unclear whether urine contains cell-free chromatin with intact epigenetic marks. Here, we extend the use of cell-free chromatin immunoprecipitation followed by sequencing (cfChIP-seq) to human urine. We demonstrate that cell-free nucleosomes captured in urine preserve multiple histone post-translational

modifications indicative of activation and repression. By analyzing the H3K4me3 promoter mark, we identified the primary tissues contributing to cfDNA in healthy individuals. Notably, we observe distinct populations of circulating nucleosomes in urine and plasma samples with a contribution from the kidney in healthy donors that are not detected in matched urine exfoliated cells or matching plasma samples. Additionally, we show that urine cf-nucleosomes can be used to detect pathologically driven changes in the urine of bladder cancer patients, reflecting tumor-associated transcriptional programs and immune responses. The cfChIP-seq signal from cancer and immune origins can be used to accurately identify cancer patient samples, achieving an area under the curve (AUC) of 0.97 in a validation set. Our findings highlight the potential of urine cell-free nucleosomes as accessible, noninvasive biomarkers for both basic research in renal physiology and monitoring urinary pathologies.

Key Findings:

- Urine cell-free nucleosomes retain multiple histone marks informative of gene promoters and enhancers and can be successfully immunoprecipitated.
- Urine cfChIP-seq identifies bladder, kidney, and immune cells as the major contributing organs to the pool of urine cell-free nucleosomes.
- The populations of cell-free nucleosomes in urine and blood are distinct and mostly disjoint. Suggesting that there is little if any transport of nucleosomes across the blood-urine interface.
- We can detect pathologically driven changes in cell-free nucleosomes from the tumors and immune cells responding to the tumor.
- Urine cell-free nucleosomes as potential noninvasive biomarkers for detections of non-muscle invasive bladder cancer

Introduction

Dying cells release fragments of DNA (cell-free DNA or cfDNA) into the urinary tract (*1*). Urinary cfDNA is rapidly degraded by active DNaseI in the urinary tract and thus it originates from recent, close to urination time, cell death events (*2*, *3*).

The two major sources of urinary cfDNA include dying cells along the urogenital tract and transrenal cfDNA which passes from blood circulation into urine. Urinary cfDNA has proven to be a useful analyte currently used as a source of biomarkers for various pathologies, including urogenital cancers such as prostate and bladder cancer, but also non-urogenital cancers such as glioma (*4*) and colorectal cancer (*5*). Urinary cfDNA has also shown potential applications in monitoring urinary tract infection and organ transplant rejections (*6*, *7*) and in prenatal diagnosis (*8*).

In plasma most cfDNA is packaged in the form of nucleosomes (*9*). Nucleosomes are DNA-protein complexes that help compacted the DNA within the cell's nucleus. These complexes, composed of ~150 bp of DNA and histone proteins, are the basic repeating unit of chromatin (*10*). Histones undergo post-translational modifications reflecting DNA-templated processes including the activity of RNA polymerase at promoters, gene bodies, and enhancers (*11*–*13*). Histone marks offer clear and well-defined relationships to cellular processes, making chromatin analysis valuable for understanding cell identity and state.

Histone proteins have been described in urine (*14*) , however it is not known if they represent intact nucleosomes that carry additional epigenetic information in the form of histone modifications. Necrotic and apoptotic cells release chromatin into their immediate environment (*15*) thus we expect to find in urine cell-free nucleosomes from dying urogenital cells. Conversely, the process of urine formation by glomerular filtration of blood entails that large proteins with molecular weight >45 kilodalton rarely pass filtration (*16–18*), and thus it is expected that plasma circulating nucleosomes (~179 kilodalton) (*19*) are mostly excluded from urine. The unidirectionality of the filtration process suggests absence of urine cf-nucleosomes in plasma. This implies that while transrenal cfDNA might hold similar information in urine and plasma, nucleosome-bound cfDNA captured in cfChIP-seq from urine and plasma may provide mutually exclusive insights.

To investigate the existence of urine cell-free nucleosomes and the information they possess, we performed cell-free chromatin immunoprecipitation and sequencing (cfChIP-Seq (*20*)) in human urine. We targeted both active (H3K4me1/2/3, H3K27ac) and repressive (H3K27me3) histone marks and showed that urine contains cell-free nucleosomes which retain multiple histone marks. Using cfChIP-seq against H3K4me3 (an active/poised promoter histone mark) in a cohort of healthy samples, we characterize the cell of origin of urine cell-free nucleosomes. Moreover, cfChIP-seq profiles of a cohort of samples from non-muscle invasive bladder cancer patients (NMIBC) highlight pathology driven changes. We show these discoveries are otherwise undetectable in plasma cfChIP-seq, reinforcing the notion that urine and plasma are non-overlapping sources of epigenetic information. Taken together these results highlight the

information urine cell-free nucleosomes can provide on molecular states of cells in the renal and urinary system.

Results

Chromatin immunoprecipitation in human urine yields information from several histone post-translational modifications

We performed urine cfChIP-seq on urine samples from healthy individuals with antibodies targeting marks of accessible/active promoters (H3K4me3, H3K27Ac), enhancers (H3K4me2, H3K4me1, H3K27Ac) and polycomb repressed genes (H3K27me3) (Fig. 1A-B, S1A). Assay performed on mere 2 ml of urine yielded sequencing libraries with an average of >13 million estimated unique reads and high alignment rates (Supplementary table 1). Indeed, all tested marks showed the expected enrichment for active (H3K4me1/2/3, H3K27ac), or repressive (H3K27me3) promoters/enhancers (Fig. 1B-C, S1A-B). The urine cell-free ChIP-profile of each histone modification matches the profiles observed in published tissue ChIP-seq (Fig. 1B-C, S1B). Specifically, we see that H3K4me3 is located at promoters of active and poised genes, H3K4me2/1 at regulatory regions (promoters and enhancers), H3K27ac at active regulatory regions, and H3K27me3 surrounding repressed regulatory regions.

Urine cfChIP-seq captures a distinct subpopulation of fragmented nucleosomal cfDNA

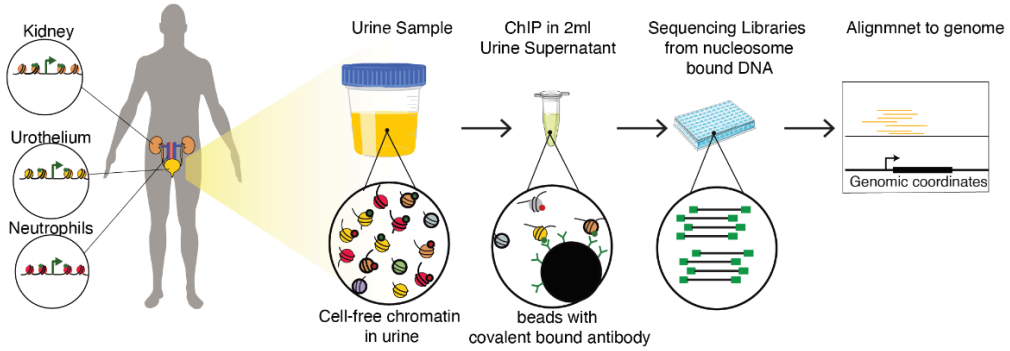
Examining the fragment lengths from multiple urine cfChIP-seq assays that target different histone modifications, we observe a stereotypical pattern (Fig. 1D, S1C-D). ChIP enriches for histone-bound DNA, and thus we will refer to the products of ChIP cell-free nucleosomal DNA (cfnDNA).

The lengths of cfnDNA fragments captured as a continuum in the range of mono-nucleosome (~170 bp) to tri-nucleosome (~400 bp) fragments (Fig. 1D-E, S1C-D). In contrast, urine cfDNA fragments lengths span a wider range short (< 100 bp) to long (>1000 bp) (*21*, *22*). However, the cfDNA pool includes a large proportion of short fragments (<100bp) that do not appear in the cfnDNA pool. To ensure that this is not due to the library preparation and sequencing protocol, we sequenced urine cfDNA using the same protocols used for cfChIP-seq (Fig 1E) and reproduced the same discrepancy in the short fragments distribution. These results suggest that cfChIP-seq captures nucleosomal DNA fragments which by nature are longer than 100bp, while cfDNA sequencing captures an additional pool of non-nucleosomal fragments which include many fragments shorter than 100bp. In contrast, results in plasma cfDNA suggest that most, if not all, the cfDNA fragments are nucleosomal (*9*, *23*, *24*).

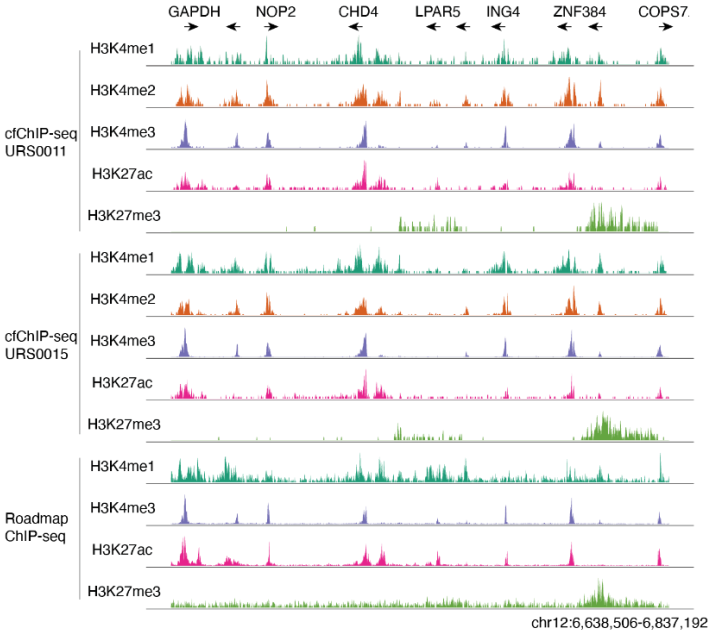
Both urine cfDNA and cfnDNA have a 10 bp length periodicity, consistent with DNaseI digestion pattern (Fig 1E) (*4*, *25*, *26*). Together with the fragment length distribution, this suggests that the fragments recovered by ChIP-seq undergo multi-phase digestion where nucleosomal ladder fragments (presumably digested at cell death) are subsequently digested in the urine by DNaseI.

This is consistent with the high abundance of DNaseI in urine, which is also more active in urine than in plasma (*27*, *28*), and partial unwrapping of nucleosomes mainly due to the high salt and urea concentrations in urine (*29*).

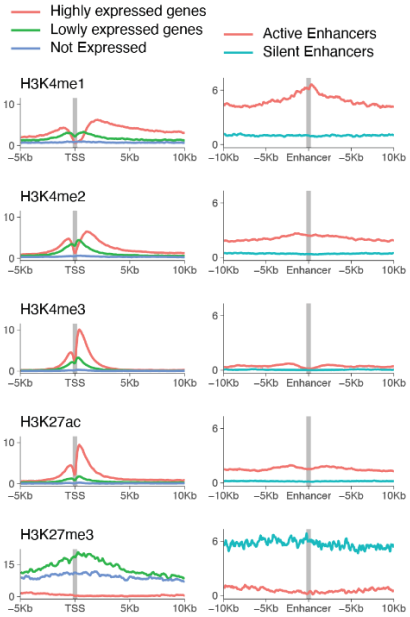
A Capturing urine cell-free nucleosomes using cfChIP-seq



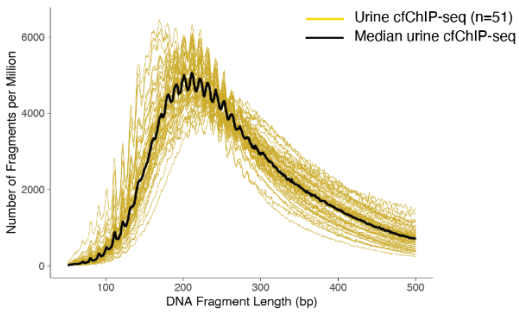
B Urine cfChIP-seq signal in multiple histone marks



C Average cfChIP-seq signal over regulatory regions



D Urine cfnDNA fargment pattern



E Fragmentation of urine cfDNA and cfnDNA

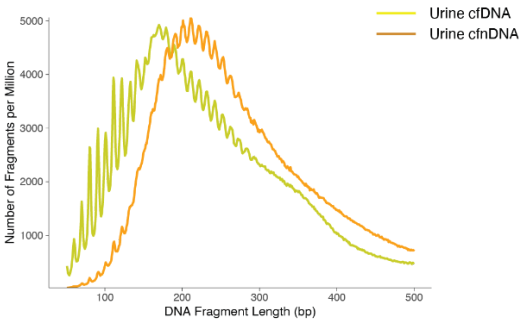


Fig. 1 cfChIP-seq of multiple histone modifications in human urine

A. Urine cfChIP-seq outline. Different cells release chromatin fragments into the urine. These fragments are immunoprecipitated using beads with covalently bound antibodies. DNA fragments are released and sequencing libraries are generated.

B. Genome-browser view of histone modifications signal. Three sets of tracks are shown: Two from cfChIP-seq of healthy donors and one from published epithelial-tissue ChIP-seq data (*30*) .

C. Metagene of normalized cfChIP-seq signal (Methods) over promoters and enhancers. Line colors indicate activity levels of the genes based on >2000 genes and enhancers from published epithelial-tissue ChIP-seq data(*30*) . Graphs scale are coverage of fragments per million.

D. cfChIP-seq DNA fragment size distribution. Thin yellow lines show DNA fragments from 51 urine cfChIP-seq samples of various histone modifications (H3K27ac n=6,H3K27me3 n=5,H3K4me1 n=13, H3K4me2 n= 13, H3K4me3 n= 14), thick black line shows the median of all samples.

E. Fragmentation patterns of urine ChIP-seq cfnDNA and urine cfDNA (brown and orange lines).

H3K4me3 promoter-marks identifies the tissue of origin within the urine cf-nucleosome pool

To further analyze the tissue origin of urine cf-nucleosomes, we focused on the H3K4me3 histone modification which marks active or poised promoters and was shown to correlate with gene transcription and be informative of cell identity and cell state (*20*). We used published ChIP-seq data to define tissue-specific signatures as promoters that have high signal only in one tissue-type (Methods, Fig. 2A-B).

In nearly all healthy urine donors, we observed a strong signal of urinary epithelium (urothelium), kidney and neutrophils, and in a few samples macrophages, monocytes, B-cells and T-cells signatures. We do not see contributions from other tissue types tested (Fig. 2B, S2A). These are the expected cell-types according to observations from cfDNA methylation in urine (*3*) and single-cell RNA-seq of urine exfoliated-cells(*31*, *32*)

Men and women differ genomically and anatomically and these differences are reflected in assay results. Urine cfChIP-seq can clearly differentiate between the sexes based on Y chromosome read coverage (Fig. 2C). Additionally in comparison of urine tissue signatures, we saw several sex based differences (Fig. 2D, S2B). We observed significantly more urothelium in females compared to males. This is concordant with the phenomenon of the female urethra using cell exfoliation as a defense mechanism against pathogens (*33*). We observed higher macrophage levels in females but did not observe significant sex based differences in the levels of kidney and neutrophils. We did not observe sex differences in assay quality (Figure S2C).

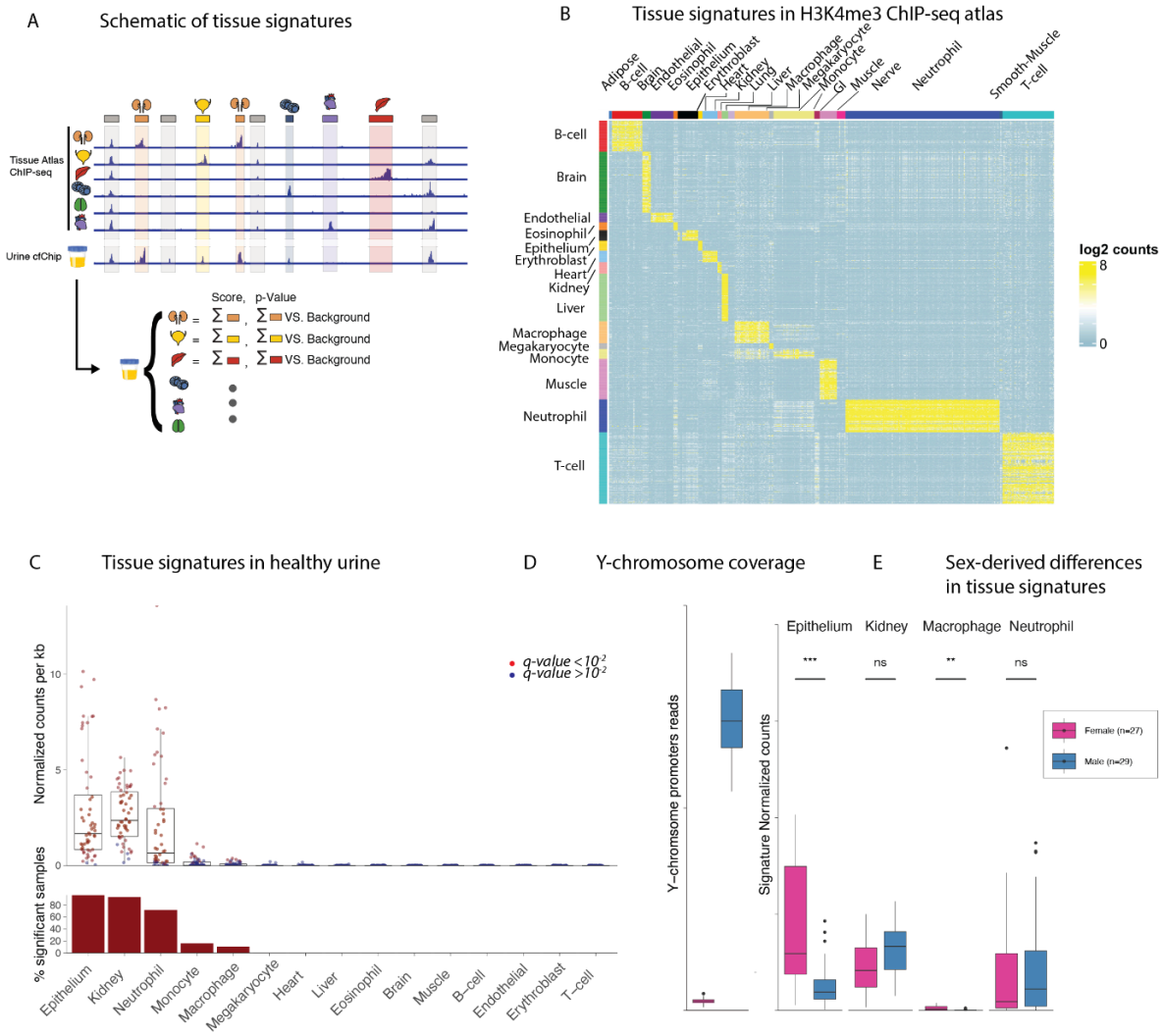

A
Schematic of tissue signatures
Tissue Atlas ChIP-seq
Urine cfChip
Score, p-Value
VS. Background
B
Tissue signatures in H3K4me3 ChIP-seq atlas
Adipose
B-cell
Brain
Endothelial
Eosinophil
Epithelium
Erythroblast
Heart
Kidney
Lung
Liver
Macrophage
Megakaryocyte
Monocyte
GI
Muscle
Nerve
Neutrophil
Smooth-Muscle
T-cell
log2 counts
8
0
C
Tissue signatures in healthy urine
Normalized counts per kb
% significant samples
q-value <10-2
q-value >10-2
Epithelium
Kidney
Neutrophil
Monocyte
Macrophage
Megakaryocyte
Heart
Liver
Eosinophil
Brain
Muscle
B-cell
Endothelial
Erythroblast
T-cell
D
Y-chromosome coverage
Y-chromosome promoters reads
E
Sex-derived differences in tissue signatures
Epithelium
Kidney
Macrophage
Neutrophil
***
ns
**
ns
Signature Normalized counts
Female (n=27)
Male (n=29)

Fig. 2 cfChIP-seq detects cfDNA cell of origin

A. Schematic of genomic regions used as tissue-specific signatures based on reference ChIP-seq tissue atlas. For each tissue (rows in genome browser view) we identify genomic locations that have high signal only in the target cell type (column with colored matching icons. Colored squares represent tissue specific windows. Gray windows are non-specific or background windows). Given a urine cfChIP-seq sample (bottom row and right-side of illustration), we sum the signal at signature locations and test against the null hypothesis of non-specific background coverage (Methods).

B. Heatmap of tissue signatures (y-axis) in H3K4me3 chip-seq signal of the tissue atlas (x-axis). Color indicates the log2 normalized read-count of each genomic-window in each tissue.

C. Top, Tissue signatures in cfChIP-seq in 56 healthy samples from 21 donors. Each dot is a sample. Box limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than 1.5× interquartile range from the hinge. Dots marked in red indicate values significantly higher than background levels (Methods). Bottom, percent of samples with significant signal for each signature.

D. Normalized promoter coverage in the Y-chromosome in urine samples from female (27 samples from 9 donors) and male (29 samples from 12 donors) donors.

E. Tissue signatures in female and male donors as in C. Box plots description as in B. Wilcox-test p-values for comparison between the sexes are presented for each tissue signature. . */**/***/**** correspond to $P < 0.05/0.01/0.001/0.0001$.

Tissue contributions differ in matching urine cfnDNA, plasma cfnDNA, and exfoliated cells in urine

Next we asked if urine cf-nucleosomes can provide insights absent in plasma cf-nucleosomes and in exfoliated cells found in urine. Therefore, we compared urine cf-nucleosomes to chromatin of urine-derived exfoliated cells and plasma cf-nucleosomes from the same individual (Fig. 3A-B, S3A) and systematically across multiple samples (Fig. 3C).

In both urine cfChIP-seq and exfoliated cells ChIP-seq, we observe a significant signal associated with the epithelial signature. This aligns with previous findings that the majority of exfoliated cells in urine are of epithelial origin (*32*, *34*). Notably, we detected evidence of kidney-derived cells exclusively in the cell-free fraction of urine (Fig. 3C, Fig. S3A). This is consistent with earlier reports indicating that these cells are rare in healthy urine samples (*32*). Our findings suggest that in healthy donors, most kidney-derived cfDNA arises from *in-situ* cell death rather than exfoliation and subsequent cell death within the bladder.

In plasma samples, we observe megakaryocyte-derived and monocyte cfnDNA, most likely due to megakaryocytes cell-death in the bone marrow and monocyte turnover in the circulation (*35*). Despite a large contribution in all plasma samples, we do not observe megakaryocyte signatures in both urine fractions, suggesting that in healthy donors, there is effective filtration of these nucleosomes in the kidney. We do observe low levels of monocyte signature in urine cfnDNA, potentially due to monocyte passage through the urogenital tract.

In contrast to the above pattern of monocyte-derived DNA, we observe macrophages, which are the tissue-resident form of monocytes, appear in exfoliated cells compared to the minimal signal in urine or plasma, and suggests these are tissue-resident macrophages along the urogenital tract.

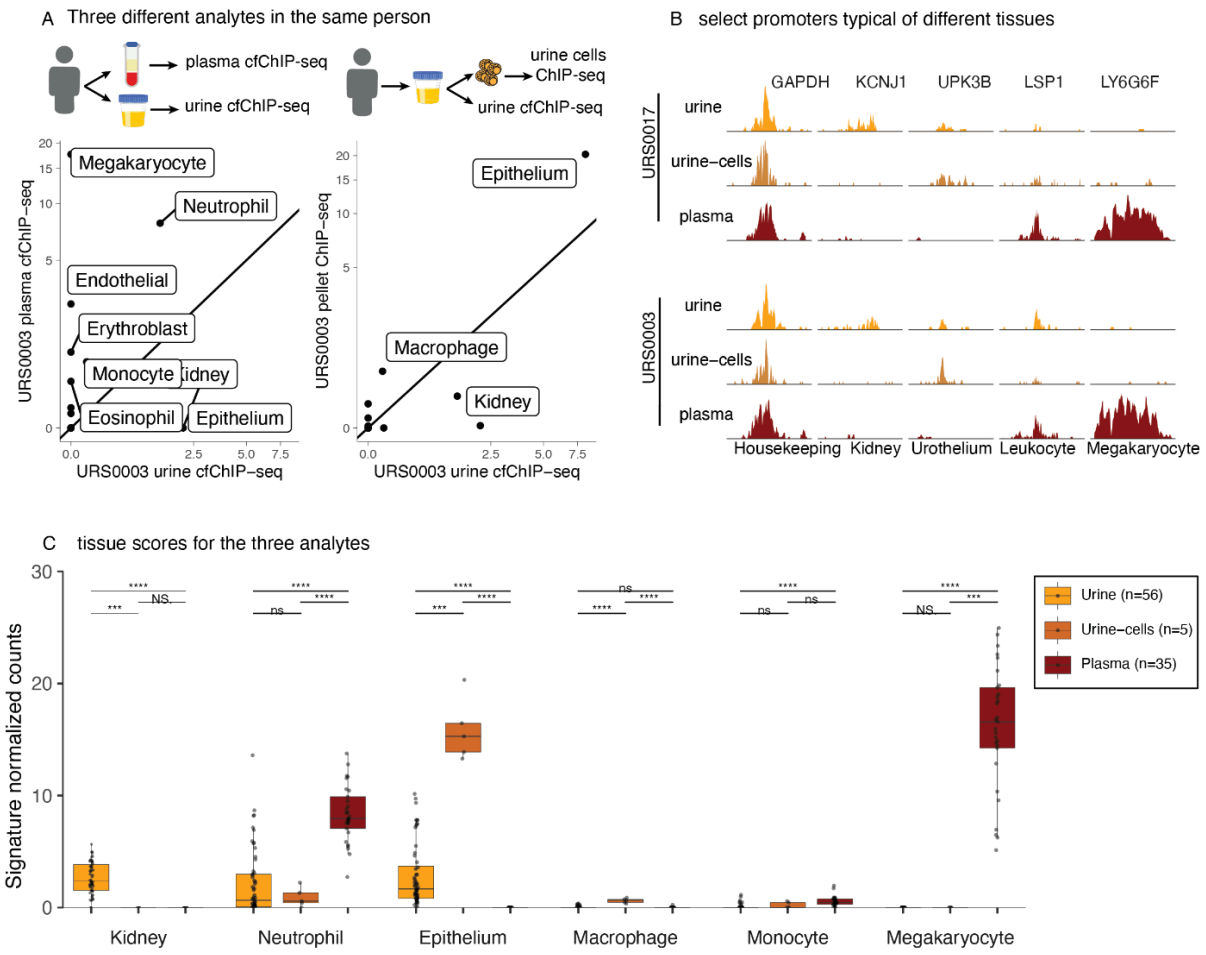


Figure 3 - Differences in promoter signal of three analytes

A. Comparison of urine cfChIP-seq (x-axis) vs. donor matching plasma cfChIP-seq (y-axis left) and ChIP-seq of urine-cells (y-axis right). Dots represent signature- scores of labeled tissues.

B. Genome browser comparing the three analytes (urine cfChIP-seq, plasma cfChIP-seq and urine-cells ChIP-seq) in two individuals. Each column features a promoter representative of a tissue-type.

C. Tissue signature distribution of multiple samples in the three analytes: urine cfChIP-seq (56 samples from 21 donors), urine-cellsChIP-seq (5 samples from 4 donors) and plasma cfChIP-seq (35 samples from 35 donors). Boxplot description and statistical test as in figure 2D.

Bladder-cancer driven changes are detected in urine in cell-free chromatin

To test the utility of urine cf-nucleosomes in differentiating between transcription patterns within the source tissue, we performed the cfChIp-seq on urine samples from patients diagnosed with non-muscle invasive bladder cancer (NMIBC, table S2). Bladder cancer is one of the most common cancers in humans with more than half a million new cases worldwide and hundreds of thousands of subsequent deaths (*36*). Bladder cancer affects ~4 times more men than women and disease incidence increases with age. Over 70% of patients diagnosed with bladder cancer are diagnosed with NMIBC (*37*). This cancer originates from the urothelium, which lines the inner surface of the bladder. NMIBC has a high probability of recurrence and progression (*38*). Currently, the gold-standard in NMIBC monitoring is frequent cystoscopies (*39*) which carry adverse effects on the patients' quality of life (*40*). Non-invasive approaches to monitor and study this cancer are an unmet need.

H3K4me3 urine cfChIP-seq performed on Individual samples from NMIBC patients shows hundreds of genes with increased signal compared to a healthy reference (Fig. 4A). To systematically evaluate these differences, a cohort of 25 urine samples from men diagnosed with NMIBC were compared to a cohort of 29 urine samples from self-proclaimed healthy males (Supplementary table S2). At the tissue level, the NMIBC-cohort does not show increased levels of healthy tissues (Fig. S4A), however, testing for differentially marked genes between the two cohorts displays 177 genes with significantly elevated coverage and 61 genes with significantly decreased coverage (Fig 4B; Methods).

Despite major differences in urine cfChIP-seq, no differences were found between the plasma cfChIP-seq results of healthy individuals and those diagnosed with NMIBC (Fig. S4B-C), reiterating the value of urine as a valuable tool for study and diagnosis of urinary system pathologies. To ensure these differences in signal are not the results of the age differences between cohorts we sampled three men ages over 70 without bladder cancer and saw very little difference in gene signal compared to our healthy cohort, and does not overlap with the signal we see in cancer patients (Fig S4D).

The differentially up-marked gene set includes several previously annotated expression biomarkers for bladder cancer (e.g KRT20 (*41*), GRHL3 (*42*) FXYD3 (*43*), CA9 (*44*), UPK1B (*45*)). Comparison against a reference H3K4me3 ChIP-seq cell type atlas (Methods, Fig. 4D, S4D) identified many of these up-marked genes are enriched for T cells (22/177 up marked genes, EnrichR (*46*) ARCHS4 Tissues (*47*) $q<10^{-12}$) and phagocytes (18/177, EnrichR ARCHS4 Tissues $q<10^{-16}$). These genes have diminished signal in healthy urine cfChIP-seq and are indicative of a

differential immune-cell state as response to the tumor. Additional up-marked genes were categorized as having signal in multiple solid tissues (109/177, solid tissue) . Several up-marked genes had absent levels of signal in the reference atlas (15/177, absent), suggesting aberrant activation of pathways that are normally repressed in healthy adults (e.g UPK1B (*45*) and MAGEA10 (*48*)). Down-marked genes are enriched for the kidney (EnrichR ARCHS4 Tissues $q<10^{-8}$) in agreement with lower levels of kidney observed in cancer samples (Fig. S4G-H). This is most likely a consequence of increased signal of other contributing tissues such as T-cells, phagocytes and the tumor in NMIBC samples. Utilizing genes that are present in cancer urines but unmarked in healthy urines we can define a cancer signature (43 genes, termed cancer-signature, table S3). This signature performs well in identifying the urines of cancer patients when tested against a validation cohort of 23 healthy urine samples and 12 cancer urine samples withheld from analysis (Figure 4E-F, Receiver Operating Characteristic (ROC) curve AUC of 0.97, Methods).

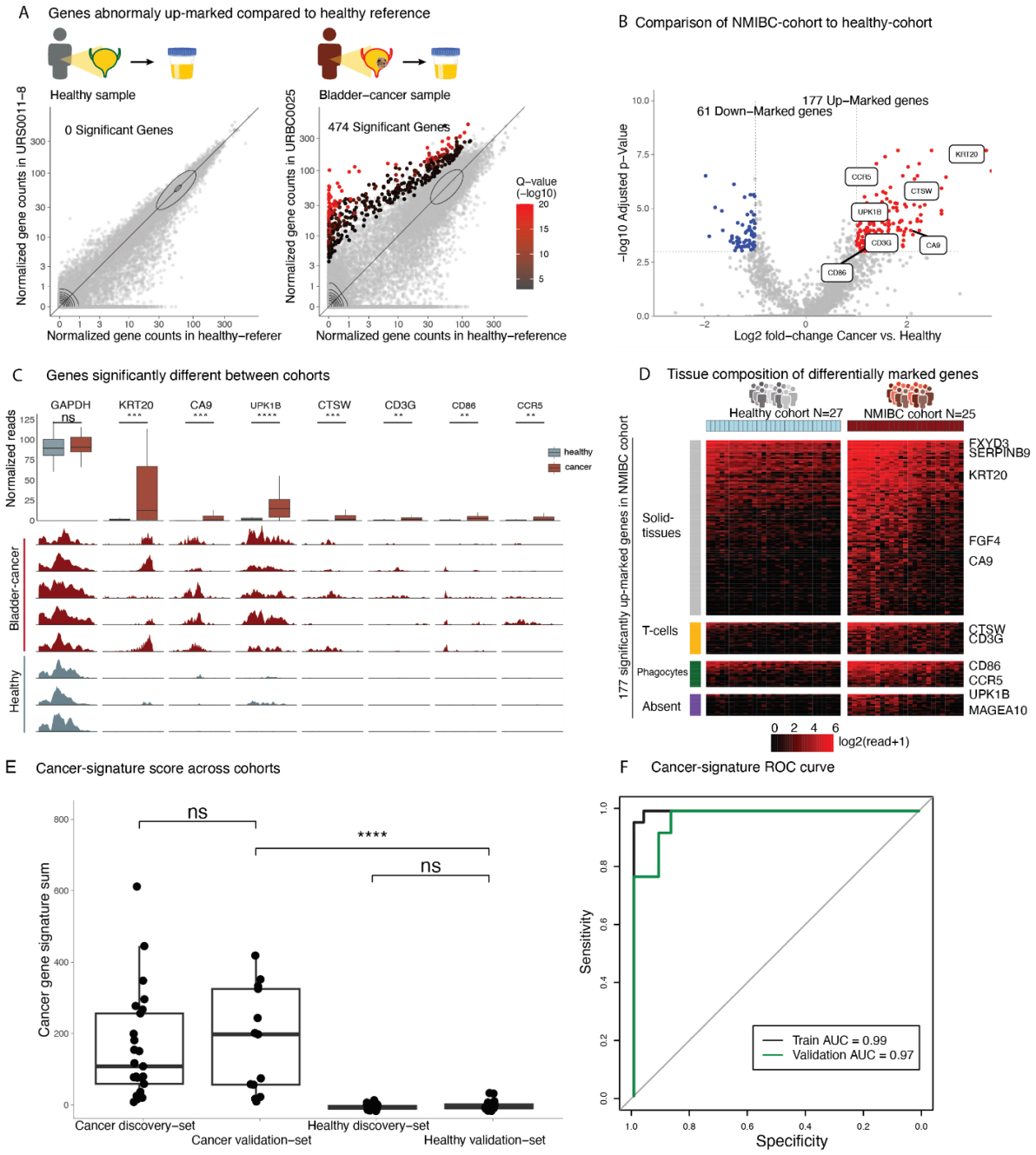

A
Genes abnormaly up-marked compared to healthy reference
Healthy sample
0 Significant Genes
Normalized gene counts in URS0011-8
Normalized gene counts in healthy-referer
Bladder-cancer sample
474 Significant Genes
Normalized gene counts in URBC0025
Normalized gene counts in healthy-reference
Q-value (-log10)
B
Comparison of NMIBC-cohort to healthy-cohort
177 Up-Marked genes
61 Down-Marked genes
-log10 Adjusted p-Value
Log2 fold-change Cancer vs. Healthy
KRT20
CCR5
CTSW
UPK1B
CD3G
CA9
CD86
C
Genes significantly different between cohorts
GAPDH
KRT20
CA9
UPK1B
CTSW
CD3G
CD86
CCR5
Normalized reads
healthy
cancer
Bladder-cancer
Healthy
D
Tissue composition of differentially marked genes
Healthy cohort N=27
NMIBC cohort N=25
177 significantly up-marked genes in NMIBC cohort
Solid-tissues
T-cells
Phagocytes
Absent
FXYD3
SERPINB9
KRT20
FGF4
CA9
CTSW
CD3G
CD86
CCR5
UPK1B
MAGEA10
log2(read+1)
E
Cancer-signature score across cohorts
Cancer gene signature sum
ns
****
ns
Cancer discovery-set
Cancer validation-set
Healthy discovery-set
Healthy validation-set
F
Cancer-signature ROC curve
Sensitivity
Specificity
Train AUC = 0.99
Validation AUC = 0.97

Fig. 4 urine detects pathology-driven changes in NMIBC cohort

A. Detection of genes with significant high coverage in a sample from a healthy urine sample (left) and the urine of a patient diagnosed with NMIBC (right). For each gene, we compared mean normalized coverage in a reference healthy cohort (methods, x axis) against the normalized coverage in the cfChIP-seq sample (y axis). Significance tests whether the observed number of reads is significantly higher than expected based on the distribution of values in healthy samples (Methods).

B. Cohort comparison across genes. Statistical significance (y-axis) and fold change of NMIBC group over healthy (x-axis) of genes in both groups. Significance values represent FDR-corrected q value of likelihood ratio test of two groups (methods).

C. Top, normalized count distribution of representative genes significantly different between the two cohorts (red boxes are NMIBC cohort samples, gray boxes are healthy samples) . Bottom, gene browser view of representative genes significantly different between the two cohorts (red tracks are NMIBC cohort samples, gray tracks are healthy samples).

D. Heat map representation of 177 genes that are significantly higher in the NMIBC cohort compared to the healthy cohort. Rows (genes) are clustered by cell type category. Heat map values (colors) are log transformed normalized values (log2 (1+normalized gene reads)). Colored row annotations are the tissues in which the row (gene) is highly expressed in. Abbreviations: Uro - urothelium, Phag- phagocytes, Solid - solid tissues, Absent - genes with negligible signal in reference tissue atlas.

E. Comparison of cancer signature levels across the healthy and NMIBC discovery cohorts and the validation cohorts. Boxplot description and statistical test as in figure 2D.

F. The Receiver Operating Characteristic (ROC) curve illustrates the performance of a classifier using the cancer signature by plotting the true positive rate (sensitivity) on the y-axis against the false positive rate (1-specificity) on the x-axis.

Discussion

Here we modify cfChIP-seq for use in urine as a new method for the inquiry of chromatin discharged by dying cells. This method captures several histone marks (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K27me3) and provides an agnostic genome-wide view of the epigenomic landscape of cells contributing cell-free nucleosomes to the urine. As such, it is applicable for the study of varied epigenomic functions, including gene activation and repression.

Our results show that cfnDNA fragmentation pattern in urine displays characteristics of digestion patterns observed in plasma and of cfDNA in urine, suggesting a two-phase digestion mechanism—initial digestion during cell death (similar to plasma cfDNA) followed by in-urine digestion (similar to urine cfDNA).

Focusing on the promoter mark H3K4me3, We applied the assay on >100 urine samples from healthy donors and NIMBC patients. We delineated the tissues contributing to the cf-nucleosome pool in human urine. We observed sex based differences in assay results which most likely reflect physiological differences.

The observed differences in cell-free nucleosome populations strongly suggest that in healthy donors plasma cell-free nucleosomes rarely reach the urine. Megakaryocytes contribute much of the observed contribution to cfDNA in plasma (*20*, *35*) and are not observed in urine cf-nucleosomes while kidney cf nucleosomes are absent in plasma cell-free nucleosomes. It is conceivable that this barrier is removed in various kidney pathologies.

Urine cfChIP-seq's ability to capture kidney-derived cfnDNA presents a novel avenue for studying epigenetics in kidney physiology and potential utility in transplantation monitoring. This information is not retrievable in the plasma of healthy individuals and is scarce in urine exfoliated-cells.

Beyond cell-type variability, we used urine cf-nucleosomes to detect differences in cell-states. In the urine of NMIBC-patients we observed increased signal of genes known as bladder-cancer expression markers as well as genes indicative of an immune response.

Taken together, these findings demonstrate non-invasive urine cf-nucleosomes range of uses spanning basic biological research as well as potential clinical utility opening new possibilities in the field of urine liquid biopsy.

Materials and Methods

Donors

All clinical studies were approved by the relevant local ethics committees. The study was approved by the Ethics Committees of the Hadassah-Hebrew University Medical Center of Jerusalem. Helsinki: 0525-21-HMO. Informed consent was obtained from all individuals before blood and urine sampling.

Sample collection

For urine: ~20 ml Urine samples were collected in standard plastic collection cups supplemented with EDTA to a final concentration for 10mM to prevent DNase digestion. Samples were transferred to cone 50ml tubes and centrifuged immediately at room-temperature at speed of 3000xg for 10 minutes. Urine supernatant was transferred to a second 50 ml cone tube whilst being careful not to disturb any cell debris. The urine supernatant was centrifuged and transferred a second time using the same settings as before to ensure no cellular debris. The doubly centrifuged urine was used for the cfChIP-seq experiments. Urine cfChIP-seq was performed immediately or after storage at 4C for at most 2 weeks per our preanalytic experiment results (see below).

For plasma: Plasma samples were collected as in Sadeh et. al. 2021. Briefly, Blood samples were collected in VACUETTE K3 EDTA tubes and transferred to ice. The blood was centrifuged (10 min, 1,500g, 4 °C) and the supernatant was transferred to a fresh 15-ml tube and centrifuged again (10 min, 3,000g, 4 °C). The supernatant plasma was used for ChIP experiments. The plasma was used fresh or flash frozen and stored at −80 °C for long-term storage.

Urine sample preanalytics

Urine samples collected as above were kept in three storage conditions and then cfChIP-seq was performed (below). Conditions tested were three time points (3 days, 7 days and 14 days) in room temperature or 4C and two time points (7 days, 14 days and 30 days) in -80C. We found urine cfChIP-seq quality rapidly deteriorates in room-temperature but is maintained in 4C and -80C over-time relative to freshly processed samples (Fig. S5) .

Immunoprecipitation, NGS library preparation, and sequencing

Immunoprecipitation, library preparation and sequencing were performed by Senseera LTD as previously reported(*20*) with slight modifications. Briefly, ChIP antibodies were covalently immobilized to paramagnetic beads and incubated with urine at room temperature overnight or plasma at 4C overnight. Barcoded sequencing adaptors were ligated to chromatin fragments and DNA was isolated and next-generation sequenced.

Sequencing Analysis

Reads were aligned to the human genome (hg19) using bowtie2 (2.4.2) with 'no-mixed' and 'no-discordant' flags. Fragments with low alignment scores (-q 2) and duplicates were discarded. See Supplementary Table 1 for total number of reads, alignment statistics and number of unique fragments for each sample. BAM files were processed by samtools (1.7) to BED files and by R (4.0.2) scripts (see 'Code availability').

Preprocessing of sequencing data was performed as previously described. Briefly, the human genome was segmented into windows representing TSS according to a and background (rest of the windows) according to a curated genomic catalog. The sequenced reads covering each of these regions were quantified and used for further analysis, with non-specific reads estimated per sample and subtracted resulting in the specific signal in every window. Signal -counts were normalized and scaled to 1 million reads in healthy reference accounting for sequencing depth differences. Detailed information regarding these steps can be found at supplementary note in(*20*)

Epigenomics atlas.

We downloaded 414 samples of H3K4me3 ChIP-seq aligned read data from the ENCODE Epigenomics consortium (https://www.encodeproject.org/) and BLUEPRINT database (https://www.blueprint-epigenome.eu/). List of all cell-types used as reference data is in table S5. The BED files were then processed with the same scripts as cfChIP-seq samples.

H3K4me3 ChIP-seq of urine exfoliated cells

Urine samples were collected as above with the exception of urine exfoliated-cells collection after initial centrifugation. After moving urine to a fresh tube, pelleted cells were washed using ice-cold 1X PBS supplemented with protease inhibitors (Roche cat. 5056489001). Cells were counted [Bio-Rad TC20 Cell Counter] and resuspended using supplemented NP buffer [10 mM Tris pH 7.4, 50 mM NaCl, 5 mM MgCl2, 1 mM CaCl2, and 0.075% NP-40, supplement with fresh 1 mM β-mercaptoethanol, 500 µM spermidine]. Chromatin was digested with Micrococcal nuclease [Worthington Biochemical LS004798] at a concentration of 5 units per 1 million cells. Chromatin was collected by centrifugation (14,000xg, 10 minutes at 4C) and quantified (Qubit 4 Fluorometer, Agilent TapeStation). Native H3K4me3 ChIP-seq was performed in 500ul SDSless RIPA [10 mM Tris pH 8.0, 140 mM NaCl, 1 mM EDTA, , 1% Triton X-100] or Bovine Serum Heat inactivated (rockland D500-05-0500) using the same protocol as for urine cfChIP-seq. Data processing was also done as above.

Urine and plasma cfDNA sequencing

cfDNA from urine and plasma samples (collected as above) was extracted using Zymo cfDNA extraction kit [Zymo Research ZR-R1072] according to manufacturer's instructions. Sequencing libraries were prepared as above. Data processing was also done as above.

Tissue signatures.

Using the ENCODE and Roadmap Epigenomics metadata table we defined sets of samples that belong to a tissue or group of tissues (see Supplementary Table 12). We then defined for each

group of samples the set of specific windows as meeting the following criteria: 1. The window is not predefined as a background region in the genome 2. The window is not located on the Y-chromosome 3. At least two samples of the window satisfy the condition that they have >10 normalized reads, and <5 in all other samples outside the group, or >30 normalized reads and <10 in all other samples - the last condition allows minimal signal from contaminating populations such as immune-populations in solid tissues. Groups for which we found less than 20 specific windows were considered to be without signature. The signature-scores are defined as the sum of normalized reads across signature windows divided by window lengths (normalized reads / kb).

Cancer-signature (Figure 4E-F)

Significantly differentially up-marked genes in the NMIBC-cohort were filtered for those with a maximum of 3 normalized reads in healthy samples. This set of 43 genes was termed cancer-signature (Table S3).

Statistical analysis.

All custom statistical analyses are thoroughly described in(*20*). Computed p-values were corrected for multiple hypotheses using FDR and estimated as q values (R function p.adjust()).

Comparison to background (Figure 2B) - We test whether the total sum of reads over a collection of windows (a signature, gene promoter windows, etc.) is larger than we would expect from the background signal. The null hypothesis is that the number of reads in the windows of interest is Poisson distributed according to the estimated background rate at these windows.

Comparison to healthy reference (Figure 4A, S4B) - We test whether the total sum of reads over a collection of windows is higher than we would expect according to mean and variance in healthy donor reference (either in plasma or in urine references).

Comparison of two groups of samples (Figures 4B, S4C) - For each gene in each group, mean and variance were estimated under a model distribution. We use a likelihood ratio test (LRT) where, under the null assumption, the distribution of the statistic is approximately a Chi-squared distribution with 2 degrees of freedom, which allows us to compute the p-value.

Acknowledgements

We thank the members of the Friedman lab for discussions and comments on this manuscript. This work was supported by the European Research Council's AdG Grant cfChIP 101019560 (to N.F.) and Israel Science Foundation IPMP Grant 3751/21 (to N.F).

Author contributions

Conceptualization: ML,IS,NF

Methodology: ML,IS

Investigation: ML

Visualization: ML

Funding acquisition: NF

Project administration: ML,IS,TFM,NF

Supervision: NF, TFM

Writing – original draft: ML

Writing – review & editing: ML,IS,BA,TFM,NF

Competing interests

Nir Friedman and Israa Sharkia are co-founders and shareholders of Senseera LTD. all other authors have no disclosures.

Data Transparency Statement:

Study material will be made available on the EGA European Genome-Phenome Archive upon publication.

Data availability

All datasets used in this study are in the process of being deposited to public repositories.

Code availability

All script files used in the analysis in this manuscript will be available upon submission.

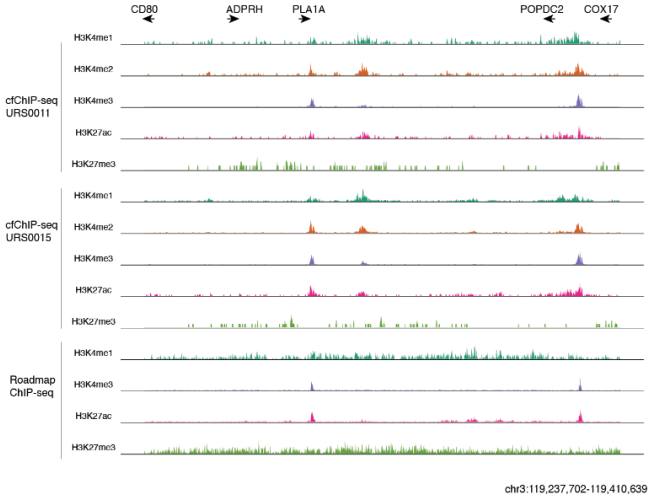
Supplementary information

- Table S1 - Sequencing statistics for samples sequenced in this study

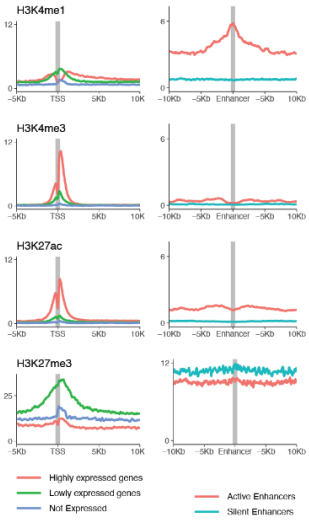
- Table S2 - Bladder cancer patients clinical information
- Table S3 - Bladder cancer signature genes
- Table S4 - List of all cell-types used as reference data from published ChIP-seq datasets
- Fig. S1-S4

Supplementary figures

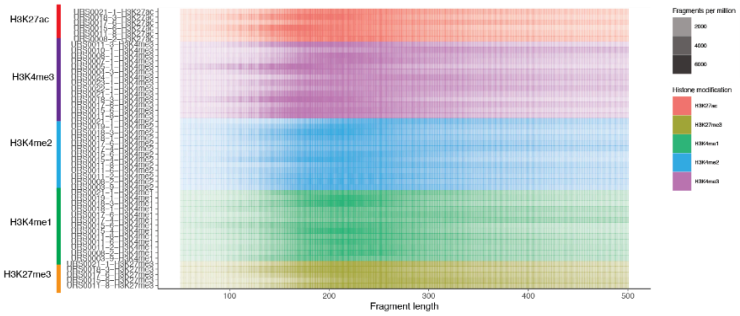
A Urine cfChIP-seq signal in another genomic region



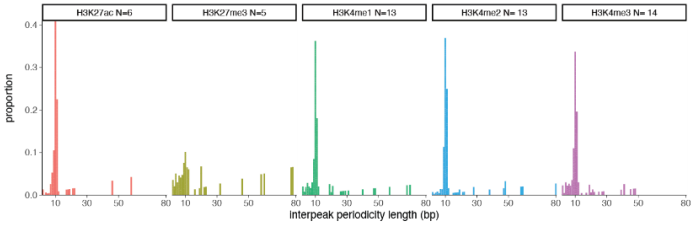
B ChIP-seq signal in reference atlas



C cfnDNA fragment profiles in multiple histone marks



D cfnDNA fragment-length peaks periodicity in multiple histone marks



E Fragmentation patterns of cfDNA and cfnDNA in urine and plasma

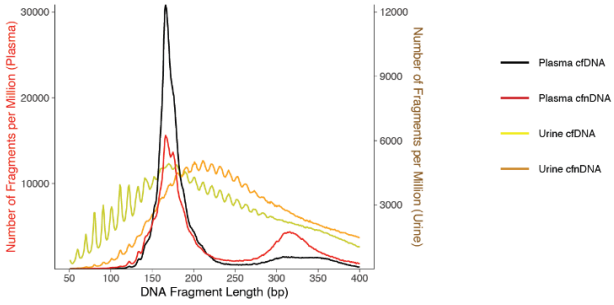


Figure S1: Urine cfChIP-seq fragment size distribution and periodicity

A. Genome browser view as in Fig. 1B

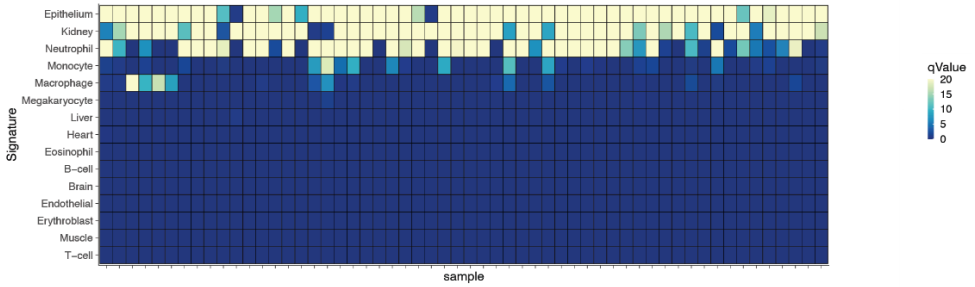
B. DNA fragment size distribution across 5 histone PTMs: H3K27ac (N=6), H3K27me3, (N=5), H3K4me1 (N=13), H3K4me2 (N = 13),H3K4me3 (N=14). Each row depicts the number of fragments (opacity) for each fragment length (x axis). Colors indicate different modifications.

C. cfnDNA fragment distributions per histone mark. X axis presents DNA fragment length, y axis are urine cfChIP-seq samples and colors represent several histone modifications with shade representing the number of fragments per million.

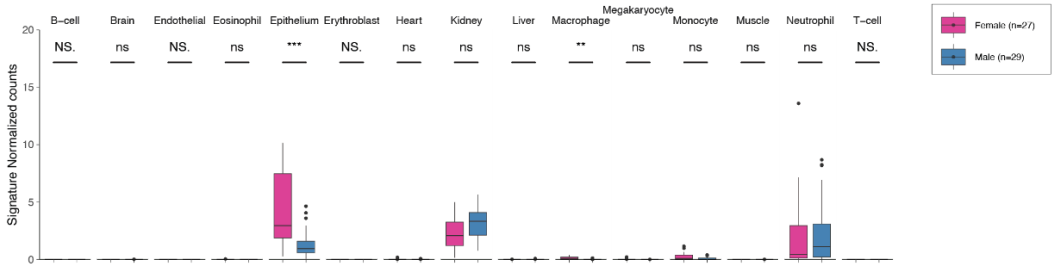
D. Proportion of interpeak periodicity length in each of the histone modifications in this study. In each histone PTM, the length between local maxima in DNA fragments is tallied for each sample, the proportions are then plotted.

E. Fragmentation pattern from urine ChIP-seq cfnDNA, urine cfDNA (brown and orange lines) and from plasma cfChIP-seq cfnDNA and plasma cfDNA (red and dark-red lines).

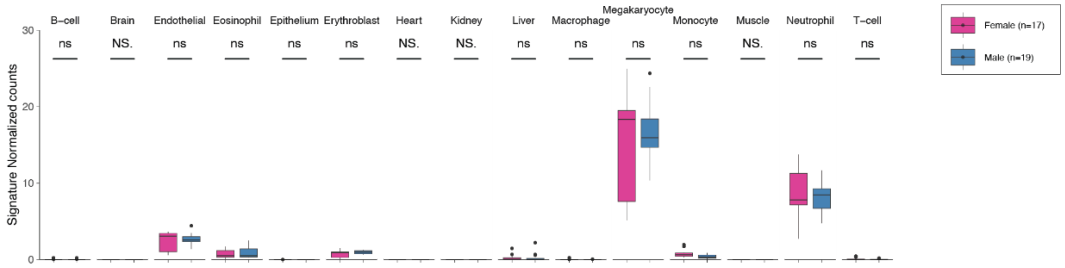
A Tissue signatures acorss healthy samples



B Tissue signatures acorss sexes in urine cfChIP-seq



C Tissue signatures acorss sexes in plasma cfChIP-seq



D QC parameters across sexes in urine cfChIP-seq

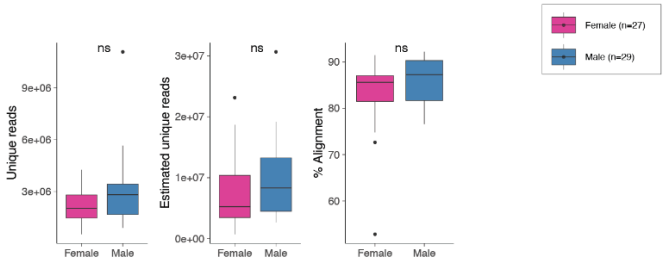


Figure S2: Tissue signatures

A. Tissue signatures (rows) of all healthy samples (columns, N=56) color in heatmap represents the adjusted p-value (q-value) of the null hypothesis that the signature's signal is equivalent to noise.
B. Tissue signature comparison between male and female samples of urine cfChIP-seq as in Fig. 2D.
C. Tissue signature comparison between male and female samples of plasma cfChIP-seq as in Fig. 2D.
D. Quality control (QC) parameters in both sexes. Statistical comparison done via Mann-Whittney U-test, "ns" - "not significant"

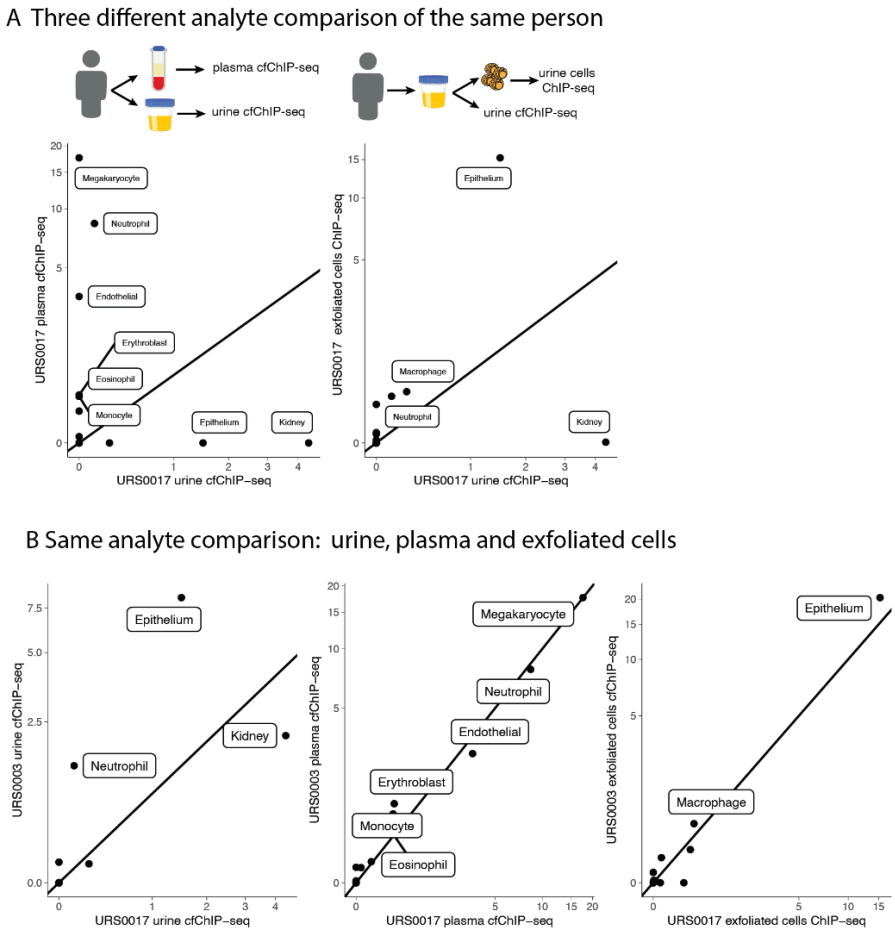


Figure S3: Comparison of Urine, exfoliated urine-cells and plasma

A. Same as Figure 3A - signature comparison of three analytes from the same individual.

B. signature comparison of the same analyte from two different individuals. Left, urine cfChIP-seq. Middle, plasma cfChIP-seq. Right, urine exfoliated cells ChIP-seq.

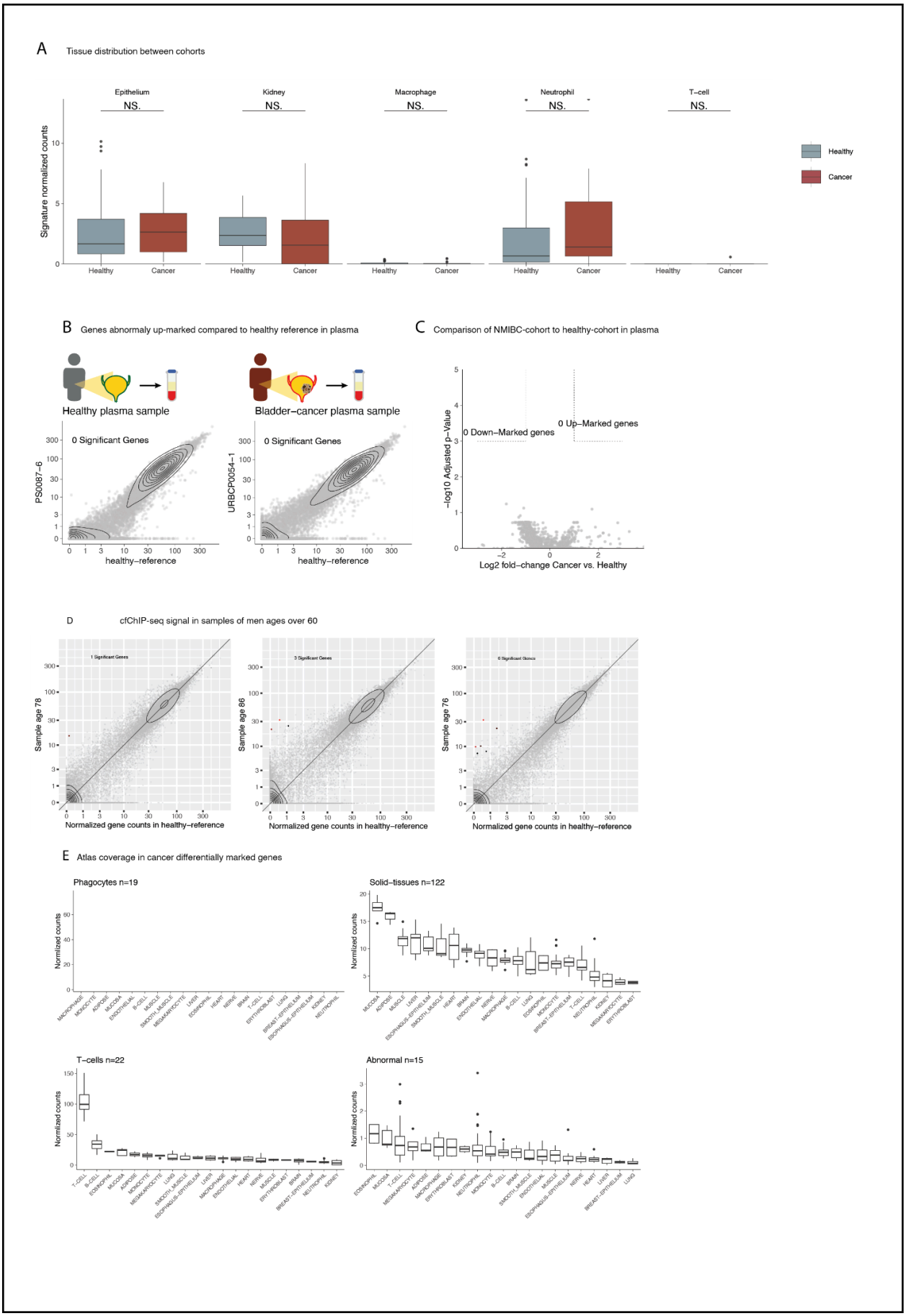

A
Tissue distribution between cohorts
Epithelium
Kidney
Macrophage
Neutrophil
T−cell
NS.
Signature normalized counts
Healthy
Cancer
B
Genes abnormaly up-marked compared to healthy reference in plasma
Healthy plasma sample
Bladder−cancer plasma sample
0 Significant Genes
PS0087−6
URBCP0054−1
healthy−reference
C
Comparison of NMIBC-cohort to healthy-cohort in plasma
0 Down−Marked genes
0 Up−Marked genes
−log10 Adjusted p−Value
Log2 fold−change Cancer vs. Healthy
D
cfChIP-seq signal in samples of men ages over 60
Sample age 78
Sample age 86
Sample age 76
Normalized gene counts in healthy−reference
E
Atlas coverage in cancer differentially marked genes
Phagocytes n=19
Solid−tissues n=122
T−cells n=22
Abnormal n=15
Normlized counts

Figure S4: Urine cfChIP-seq of NMIBC cohort

A. Tissue distribution of the healthy (gray boxes) and NMBC (dark red boxes) cohorts. Boxplot descriptions and statistical tests as in figure 2D.

B. Detection of genes with significant high coverage in a sample from a healthy plasma sample (left) and the plasma of a patient diagnosed with NMIBC (right). For each gene, we compared mean normalized coverage in a reference healthy cohort (x axis) against the normalized coverage in the cfChIP-seq sample (y axis). Significance tests whether the observed number of reads is significantly higher than expected based on the distribution of values in healthy samples (Methods).

C. Cohort comparison across genes. Statistical significance (y-axis) and fold change of NMIBC group over healthy (x-axis) of genes in both groups. Significance values represent FDR-corrected q value of likelihood ratio test of two groups (methods).

D. Comparisons of gene signals from urine samples from male donors (y-axis) ages 76 (left) 86 (middle) and 76 (right) to a healthy cohort of urine samples (x-axis). For each gene, we compared mean normalized coverage in a reference healthy cohort (methods, x axis) against the normalized coverage in the cfChIP-seq sample (y axis). Significance tests whether the observed number of reads is significantly higher than expected based on the distribution of values in healthy samples (Methods).

E. 177 significantly up-marked genes in the NMIBC cohort are categorized by tissues which they are highly marked in (methods). Boxes are different tissues in the ChIP-seq reference atlas (x-axis), box height is the mean normalized gene count for the set of genes in each category.

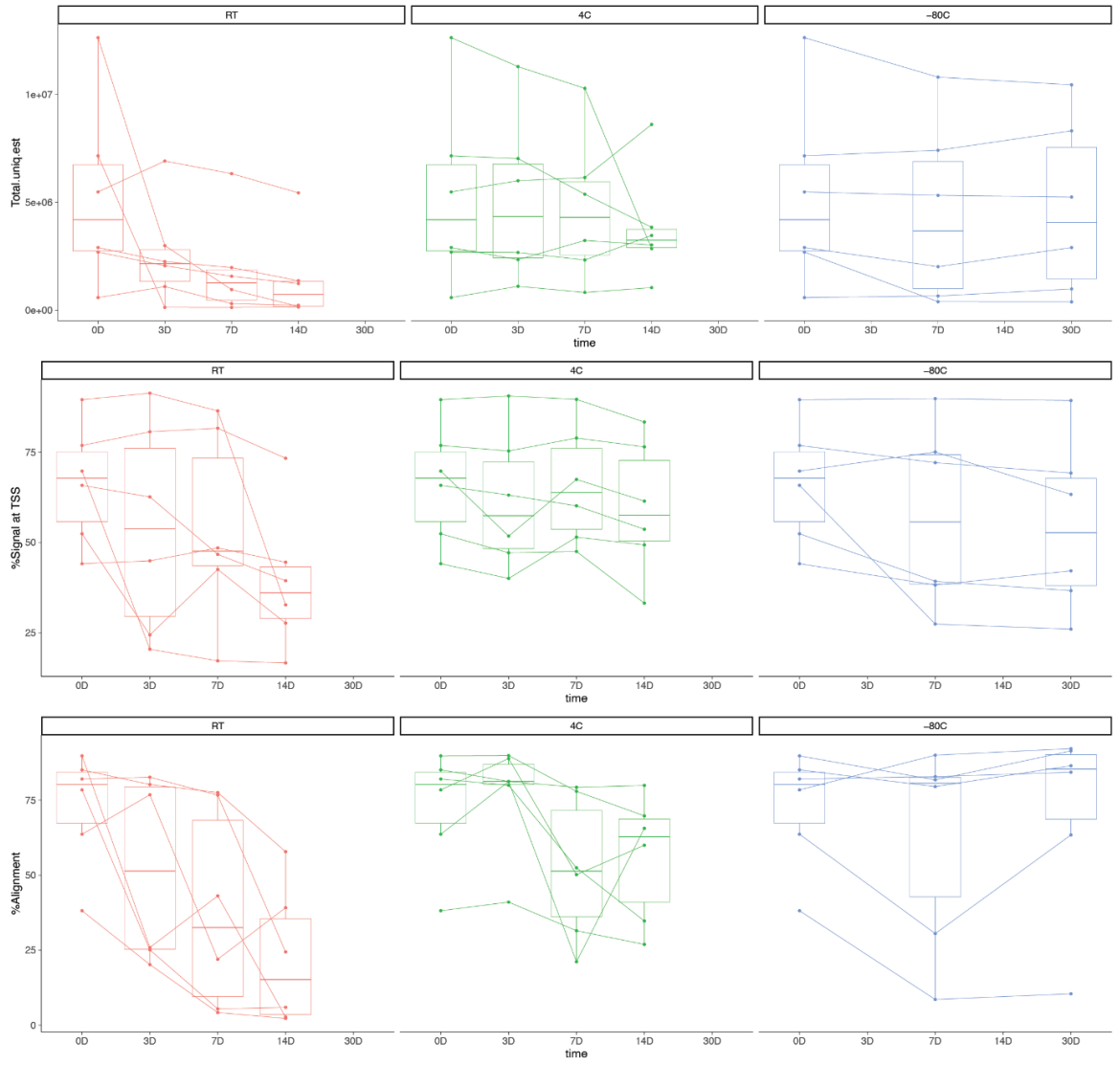


Figure S5 - Urine Sample preanalytics:

Urine cfChIP-seq was applied to 6 samples (3 Male and 3 Female) after storage in three common lab conditions: room-temperature, 4C and -80C. Samples were processed after 3,7 and 14 days in room temperature and 4C, and after a week and after a month in -80C storage. Top row, total unique estimated reads which reflects sequencing library complexity.

Middle, %signal in TSS which reflects the proportion of reads that were mapped to promoter windows in the promoter-mark H3K4me3 in urine cfChIP-seq.

Bottom row, proportion of aligned reads in sequencing.

1. T. Lu, J. Li, Clinical applications of urinary cell-free DNA in cancer: current insights and promising future. *Am. J. Cancer Res.* **7**, 2318–2332 (2017).

2. Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: A qualitative study. *Gene* **590**, 142–148 (2016).

3. T. H. T. Cheng, P. Jiang, J. C. W. Tam, X. Sun, W.-S. Lee, S. C. Y. Yu, J. Y. C. Teoh, P. K. F. Chiu, C.-F. Ng, K.-M. Chow, C.-C. Szeto, K. C. A. Chan, R. W. K. Chiu, Y. M. D. Lo, Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. *Clin. Biochem.* **50**, 496–501 (2017).

4. F. Mouliere, C. G. Smith, K. Heider, J. Su, Y. van der Pol, M. Thompson, J. Morris, J. C. M. Wan, D. Chandrananda, J. Hadfield, M. Grzelak, I. Hudecova, D.-L. Couturier, W. Cooper, H. Zhao, D. Gale, M. Eldridge, C. Watts, K. Brindle, N. Rosenfeld, R. Mair, Fragmentation patterns and personalized sequencing of cell-free DNA in urine and plasma of glioma patients. *EMBO Mol. Med.* **13**, e12881 (2021).

5. Y.-H. Su, M. Wang, D. E. Brenner, P. A. Norton, T. M. Block, Detection of Mutated K-*ras*DNA in Urine, Plasma, and Serum of Patients with Colorectal Carcinoma or Adenomatous Polyps. [Preprint] (2008). https://doi.org/10.1196/annals.1448.027.

6. J. Zhang, K. L. Tong, P. K. Li, A. Y. Chan, C. K. Yeung, C. C. Pang, T. Y. Wong, K. C. Lee, Y. M. Lo, Presence of donor- and recipient-derived DNA in cell-free urine

samples of renal transplantation recipients: urinary DNA chimerism. *Clin. Chem.* **45**, 1741–1746 (1999).

7. P. Burnham, D. Dadhania, M. Heyang, F. Chen, L. F. Westblade, M. Suthanthiran, J. R. Lee, I. De Vlaminck, Urinary cell-free DNA is a versatile analyte for monitoring infections of the urinary tract. *Nat. Commun.* **9**, 2412 (2018).

8. N. B. Y. Tsui, P. Jiang, K. C. K. Chow, X. Su, T. Y. Leung, H. Sun, K. C. A. Chan, R. W. K. Chiu, Y. M. D. Lo, High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. *PLoS One* **7**, e48319 (2012).

9. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57–68 (2016).

10. R. D. Kornberg, Y. Lorch, Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. [Preprint] (1999). https://doi.org/10.1016/s0092-8674(00)81958-3.

11. B. Li, M. Carey, J. L. Workman, The Role of Chromatin during Transcription. [Preprint] (2007). https://doi.org/10.1016/j.cell.2007.01.015.

12. S. Venkatesh, J. L. Workman, Histone exchange, chromatin structure and the regulation of transcription. [Preprint] (2015). https://doi.org/10.1038/nrm3941.

13. S. L. Berger, The complex language of chromatin regulation during transcription. [Preprint] (2007). https://doi.org/10.1038/nature05915.

14. A. Marimuthu, R. N. O'Meally, R. Chaerkady, Y. Subbannayya, V. Nanjappa, P. Kumar, D. S. Kelkar, S. M. Pinto, R. Sharma, S. Renuse, R. Goel, R. Christopher, B. Delanghe, R. N. Cole, H. C. Harsha, A. Pandey, A comprehensive map of the human urinary proteome. *J. Proteome Res.* **10**, 2734–2743 (2011).

15. Z. Hu, H. Chen, Y. Long, P. Li, Y. Gu, The main sources of circulating cell-free DNA: Apoptosis, necrosis and active secretion. *Crit. Rev. Oncol. Hematol.* **157**, 103166 (2021).

16. J. E. Hall, *Guyton and Hall Textbook of Medical Physiology* (Elsevier, 2020).

17. B. Haraldsson, J. Nyström, W. M. Deen, Properties of the glomerular barrier and mechanisms of proteinuria. *Physiol. Rev.* **88**, 451–487 (2008).

18. L. Jia, L. Zhang, C. Shao, E. Song, W. Sun, M. Li, Y. Gao, An Attempt to Understand Kidney's Protein Handling Function by Comparing Plasma and Urine Proteomes. *PLoS One* **4** (2009).

19. K. Mayanagi, K. Saikusa, N. Miyazaki, S. Akashi, K. Iwasaki, Y. Nishimura, K. Morikawa, Y. Tsunaka, Structural visualization of key steps in nucleosome reorganization by human FACT. *Sci. Rep.* **9**, 1–14 (2019).

20. R. Sadeh, I. Sharkia, G. Fialkoff, A. Rahat, J. Gutin, A. Chappleboim, M. Nitzan, I. Fox-Fisher, D. Neiman, G. Meler, Z. Kamari, D. Yaish, T. Peretz, A. Hubert, J. E. Cohen, A. Salah, M. Temper, A. Grinshpun, M. Maoz, S. Abu-Gazala, A. Ben Ya'acov, E. Shteyer, R. Safadi, T. Kaplan, R. Shemer, D. Planer, E. Galun, B. Glaser, A. Zick, Y. Dor, N. Friedman, ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nat. Biotechnol.* **39**, 586–598 (2021).

21. O. E. Bryzgunova, P. P. Laktionov, Extracellular Nucleic Acids in Urine: Sources, Structure, Diagnostic Potential. *Acta Naturae* **7**, 48–54 (2015).

22. Y. van der Pol, N. A. Tantyo, N. Evander, A. E. Hentschel, B. M. Wever, J. Ramaker, S. Bootsma, M. F. Fransen, K. J. Lenos, L. Vermeulen, F. L. Schneiders, I. Bahce, J. A. Nieuwenhuijzen, R. D. Steenbergen, D. M. Pegtel, N. Moldovan, F. Mouliere, Real-time analysis of the cancer genome and fragmentome from plasma and urine cell-free DNA using nanopore sequencing. *EMBO Mol. Med.* **15**, e17282 (2023).

23. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1317–25 (2015).

24. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the

fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).

25. O. E. Bryzgunova, T. E. Skvortsova, E. V. Kolesnikova, A. V. Starikov, E. Y. Rykova, V. V. Vlassov, P. P. Laktionov, Isolation and comparative study of cell-free nucleic acids from human urine. *Ann. N. Y. Acad. Sci.* **1075**, 334–340 (2006).

26. H. Markus, J. Zhao, T. Contente-Cuomo, M. D. Stephens, E. Raupach, A. Odenheimer-Bergman, S. Connor, B. R. McDonald, B. Moore, E. Hutchins, M. McGilvrey, M. C. de la Maza, K. Van Keuren-Jensen, P. Pirrotte, A. Goel, C. Becerra, D. D. Von Hoff, S. A. Celinski, P. Hingorani, M. Murtaza, Analysis of recurrently protected genomic regions in cell-free DNA found in urine. *Sci. Transl. Med.* **13** (2021).

27. Immunochemical assay for deoxyribonuclease activity in body fluids. *J. Immunol. Methods* **325**, 96–103 (2007).

28. D. Nadano, T. Yasuda, K. Kishi, Measurement of deoxyribonuclease I activity in human tissues and body fluids by a single radial enzyme-diffusion method. *Clin. Chem.* **39**, 448–452 (1993).

29. T. Furuyama, C. A. Codomo, S. Henikoff, Reconstitution of hemisomes on budding yeast centromeric DNA. *Nucleic Acids Res.* **41**, 5769–5783 (2013).

30. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A.

Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

31. Y. Wang, Y. Zhao, Z. Zhao, D. Li, H. Nie, Y. Sun, X. Feng, T. Zhang, Y. Ma, J. Nie, G. Cai, X. Chen, W. Zuo, Single-cell RNA-Seq analysis identified kidney progenitor cells from human urine. *Protein Cell* **12**, 305–312 (2021).

32. A. Abedini, Y. O. Zhu, S. Chatterjee, G. Halasz, K. Devalaraja-Narashimha, R. Shrestha, M. S Balzer, J. Park, T. Zhou, Z. Ma, K. M. Sullivan, H. Hu, X. Sheng, H. Liu, Y. Wei, C. M. Boustany-Kari, U. Patel, S. Almaani, M. Palmer, R. Townsend, S. Blady, J. Hogan, L. Morton, K. Susztak, TRIDENT Study Investigators, Urinary Single-Cell Profiling Captures the Cellular Diversity of the Kidney. *J. Am. Soc. Nephrol.* **32**, 614–627 (2021).

33. C. M. Kunin, C. Evans, D. Bartholomew, D. G. Bates, The antimicrobial defense mechanism of the female urethra: a reassessment. *J. Urol.* **168**, 413–419 (2002).

34. H. H. L. Wu, E. M. Goldys, C. A. Pollock, S. Saad, Exfoliated Kidney Cells from Urine for Early Diagnosis and Prognostication of CKD: The Way of the Future? *Int. J. Mol. Sci.* **23** (2022).

35. J. Moss, J. Magenheim, D. Neiman, H. Zemmour, N. Loyfer, A. Korach, Y. Samet, M. Maoz, H. Druid, P. Arner, K.-Y. Fu, E. Kiss, K. L. Spalding, G. Landesberg, A. Zick, A. Grinshpun, A. M. J. Shapiro, M. Grompe, A. D. Wittenberg, B. Glaser, R. Shemer, T. Kaplan, Y. Dor, Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).

36. A. Richters, K. K. H. Aben, L. A. L. M. Kiemeney, The global burden of urinary bladder cancer: an update. *World J. Urol.* **38**, 1895–1904 (2020).

37. M. Babjuk, M. Burger, O. Capoun, D. Cohen, E. M. Compérat, J. L. Dominguez Escrig, P. Gontero, F. Liedberg, A. Masson-Lecomte, A. H. Mostafid, J. Palou, B. W. G. van Rhijn, M. Rouprêt, S. F. Shariat, T. Seisen, V. Soukup, R. J. Sylvester, European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). *Eur. Urol.* **81**, 75–94 (2022).

38. R. J. Sylvester, A. P. M. van der Meijden, W. Oosterlinck, J. A. Witjes, C.

Bouffioux, L. Denis, D. W. W. Newling, K. Kurth, Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur. Urol.* **49**, 466–5; discussion 475–7 (2006).

39. V. Golla, K. Chamie, Oncological Monitoring of NonMuscle Invasive Bladder Cancer (NMIBC). *Bladder Cancer*, 123–138 (2021).

40. K. Stav, D. Leibovici, E. Goren, A. Livshitz, Y. I. Siegel, A. Lindner, A. Zisman, Adverse effects of cystoscopy and its impact on patients' quality of life and sexual performance. *Isr. Med. Assoc. J.* **6**, 474–478 (2004).

41. J. Breyer, R. M. Wirtz, W. Otto, P. Erben, M. C. Kriegmair, R. Stoehr, M. Eckstein, S. Eidt, S. Denzinger, M. Burger, A. Hartmann, BRIDGE Consortium, In stage pT1 non-muscle-invasive bladder cancer (NMIBC), high KRT20 and low KRT5 mRNA expression identify the luminal subtype and predict recurrence and survival. *Virchows Arch.* **470**, 267–274 (2017).

42. F. Wezel, J. Lustig, A. Azoitei, J. Liu, S. Meessen, G. Najjar, V. Zehe, P. Faustmann, F. Zengerling, A. John, T. Martini, C. Bolenz, C. Günes, Grainyhead-Like 3 Influences Migration and Invasion of Urothelial Carcinoma Cells. *Int. J. Mol. Sci.* **22** (2021).

43. Y.-P. Hsu, H.-W. Yang, N.-S. Li, Y.-T. Chen, H.-H. Pang, S.-T. Pang, Instrument-Free Detection of FXYD3 Using Vial-Based Immunosensor for Earlier and Faster Urothelial Carcinoma Diagnosis. *ACS Sens* **5**, 928–935 (2020).

44. K. J. Turner, J. P. Crew, C. C. Wykoff, P. H. Watson, R. Poulsom, J. Pastorek, P. J. Ratcliffe, D. Cranston, A. L. Harris, The hypoxia-inducible genes VEGF and CA9 are differentially regulated in superficial vs invasive bladder cancer. *Br. J. Cancer* **86**, 1276–1282 (2002).

45. F.-H. Wang, X.-J. Ma, D. Xu, J. Luo, UPK1B promotes the invasion and metastasis of bladder cancer via regulating the Wnt/β-catenin pathway. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 5471–5480 (2018).

46. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G.

McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).

47. A. Lachmann, D. Torre, A. B. Keenan, K. M. Jagodnik, H. J. Lee, L. Wang, M. C. Silverstein, A. Ma'ayan, Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* **9**, 1366 (2018).

48. C. Mengus, E. Schultz-Thater, J. Coulot, Z. Kastelan, E. Goluza, M. Coric, G. C. Spagnoli, T. Hudolin, MAGE-A10 cancer/testis antigen is highly expressed in high-grade non-muscle-invasive bladder carcinomas. *Int. J. Cancer* **132**, 2459–2463 (2013).