Multiplexed Single-Molecule Epigenetic Analysis of Plasma-Isolated Nucleosomes for Cancer Diagnostics

Vadim Fedyuk[1+], Nir Erez[1+], Noa Furth[1], Olga Beresh[1], Ekaterina Andreishcheva[2], Abhijeet Shinde[2], Daniel Jones[2], Barak Bar Zakai[3], Yael Mavor[3], Tamar Peretz[4], Ayala Hubert[4], Jonathan E Cohen[4], Azzam Salah[4], Mark Temper[4], Albert Grinshpun[4], Myriam Maoz[4], Aviad Zick[4], Guy Ron[5], Efrat Shema[1*].

[1]Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel

[2]SeqLL Inc., Woburn, MA 01801, USA

[3]Department of Surgery A, Kaplan Medical Center, Rehovot, Israel

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

[5]Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel.

* Corresponding author:

Efrat.shema@weizmann.ac.il

[+] Equal contribution

The analysis of cell-free DNA (cfDNA) in plasma represents a rapidly advancing field in medicine, providing information on pathological processes in the body. Blood cfDNA is in the form of nucleosomes, which maintain their tissue- and cancer-specific epigenetic state. We developed EPINUC, a single-molecule multi-parametric assay to comprehensively profile the Epigenetics of Plasma Isolated Nucleosomes, DNA methylation and cancer-specific protein biomarkers. Our system allows high-resolution detection of six active and repressive histone modifications, their ratios and combinatorial patterns, on millions of individual nucleosomes by single-molecule imaging. In addition, it provides sensitive and quantitative data on plasma proteins, including detection of non-secreted tumor-specific proteins such as mutant p53. Applying this analysis to a cohort of plasma samples detected colorectal cancer at high accuracy and sensitivity, even at early stages. Finally, combining EPINUC with direct single-molecule DNA sequencing revealed the tissue-of-origin of the tumor. EPINUC provides multi-layered clinical-relevant information from limited liquid biopsy material, establishing a novel approach for cancer diagnostics.

Non-invasive liquid biopsy methods, based on the analysis of cfDNA, potentiate a new generation of diagnostic approaches. The cfDNA that circulates in the plasma and serum of healthy individuals originates predominantly from death of normal blood cells[1]. In cancer patients, however, a fraction of cfDNA is tumor-derived, termed circulating tumor DNA (ctDNA). ctDNA-based sequence analysis has been shown to reveal tumor-specific genetic alterations and provide the means for non-invasive molecular profiling of tumors[2,3]. Despite encouraging data, these approaches are limited, as they require genetic differences (i.e. mutations) in order to distinguish between the normal and tumor DNA. Liquid biopsy approaches based on analysis of non-genetic features have emerged recently, most prominently methodologies that utilize tissue- and cancer-specific DNA methylation, as well as differential fragmentation patterns of cfDNA[4,5,6,7,8].

cfDNA in the plasma appears predominantly in the form of nucleosomes (cfNucleosomes), the basic unit of chromatin that consists of ~150 base pairs of DNA wrapped around the octamer of core histone proteins. Histones are extensively modified by covalent attachment of various chemical groups, forming combinatorial epigenetic patterns that are unique to each tissue, and provide information on gene expression and regulatory elements within cells[9,10,11,12]. There is evidence that cfNucleosomes retain at least some of their epigenetic modifications, and a recent study applied Chromatin Immunoprecipitation and sequencing (ChIP-seq) to identify certain marks[13,14,15]. Moreover, deep sequencing of cfDNA revealed nucleosome occupancy patterns correlating with the tissue of origin[16,17,18]. While these approaches provide the first glimpse into the rich epigenetic information present in plasma that has so far remained mostly inaccessible, they have major limitations. Mainly, they require large amounts of input material, have a limited dynamic range (ChIP-seq), or are costly and require deep sequencing. Most importantly, these methodologies have limited output and sensitivity, as they usually measure a single layer of information (i.e. DNA methylation OR a single histone modification OR nucleosome occupancy, etc.). Thus, high-resolution approaches that integrate information from multiple parameters spanning different types of analytes are required.

Colorectal Cancer (CRC) is the third most common cancer worldwide, causing approximately 700,000 deaths every year[19]. Early metastatic seeding has been recently demonstrated in CRC[20], underlining the necessity to develop better diagnostic tools to improve patient outcome. In this study, we developed a single-molecule-based liquid biopsy approach, to analyze multiple parameters from less than 1 ml of plasma sample and demonstrated its value for CRC diagnosis. We coined the technology “EPINUC” for Epigenetics of Plasma Isolated Nucleosomes (Fig. 1a).

The technology builds on our recent development of a single-molecule system to image combinatorial histone modifications by Total Internal Reflection (TIRF) microscopy[21]. To capture nucleosomes from plasma, we developed high-efficiency enzymatic reactions to fluorescently tag and polyadenylate nucleosomes (Supplementary Fig.1a,b,c,d,e, Methods). Tailed, intact nucleosomes were then immobilized on a PEGylated surface via hybridization, and the status of their post-translational modifications (PTMs) was recorded by TIRF imaging with fluorescently tagged antibodies (Fig. 1b and Supplementary Fig. 1f). Binding and dissociation of antibodies to target PTMs was imaged over 90 minutes, leveraging the TIRF narrow excitation range (~100 nm). Integration of binding events assured maximal detection of modified histones (Fig. 1c and Supplementary Fig. 2a).

EPINUC relies on direct counting of single-molecules in a population, yielding data amenable to absolute quantification and comparisons between samples. Each antibody was verified for specificity and linearity of binding with a panel of recombinant modified nucleosomes, yielding six antibodies that passed the quality control criteria (Supplementary Fig. 2b,c,d). These antibodies target the tri-methylations on histone H3 lysine 9 (H3K9me3) and lysine 27 (H3K27me3), associated with gene silencing and heterochromatin, as well as antibodies targeting marks associated with active transcription: tri-methylation of histone H3 on lysine 4 (H3K4me3) and lysine 36 (H3K36me3), and acetylation on lysine 9 (H3K9ac). In addition, our panel includes an antibody targeting mono-methylation of histone H3 on lysine 4 (H3K4me1), a mark associated with enhancers[22,23].

Nucleosomes from each plasma sample were tagged with Cy3 (green), and imaged with three pairwise combinations of antibodies labeled with AF488 (cyan) or AF647 (red). Thus, we obtained multi-parametric data for six histone PTMs, comprising of the percentage of modified nucleosomes in each sample, the ratio between various histone modifications, and the percentage of nucleosomes that harbor a combinatorial pattern of two modifications (Fig. 1d,e,f). To the best of our knowledge, EPINUC is the only technology that enables counting of multiple histone PTMs, as well as combinatorially-modified nucleosomes, at a single-molecule precision, from low volume plasma sample (<1ml).

To extend the number of analytes measured beyond histone PTMs, we exploited the single-molecule system for quantification of protein biomarkers. We modulated surface chemistry to contain PEG-streptavidin, allowing anchoring of biotin-conjugated antibodies that target plasma

proteins. Following incubation with plasma, bound proteins are imaged by fluorescent detection antibodies. Multiplexed simultaneous detection of three biomarkers is achieved through the use of distinct fluorophores (Fig. 2a). We imaged two proteins known to increase in plasma of CRC patients: Carcinoembryonic antigen (CEA), a canonical biomarker measured routinely by clinicians[24], and Tissue inhibitor of metalloproteinase-1 (TIMP-1), a glycoprotein reported to have diagnostic value in screening for CRC[25]. In addition, we measured the mammalian sterile 20-like kinase 1 (MST1), an inhibitor of cell proliferation that decreases in CRC patients[26] (Fig. 2b). We verified linear detection and specificity using cell-culture systems and knockdown experiments (Fig. 2c,d and Supplementary Fig. 3).

Counting of single molecules confers high sensitivity[27,28], thus we explored whether we could also quantify non-secreted tumor-specific plasma proteins that are undetectable by conventional technologies. We focused on the tumor suppressor p53, which is frequently mutated in CRC; p53 mutations lead to its stabilization and accumulation in tumor cells[29]. We captured p53 from plasma and applied simultaneous detection with two distinct antibodies; an antibody targeting both the wild type and mutant forms of p53, or another antibody specifically targeting the mutant conformation (Fig. 2e). Time-lapse imaging enabled the accumulation of p53 signal, overcoming the transient binding dynamics of the detection antibodies (Fig. 2f and Supplementary Fig. 3d). Indeed, we observed higher levels of total and mutant p53 in the plasma of CRC patients with confirmed p53 mutations (Fig. 2g), establishing our system's capabilities in specific detection of mutant proteins that originate directly from tumor cells.

DNA methylation is often deregulated in cancer, and specifically in colorectal cancer[30,31]. We therefore aimed to combine our analysis with quantitative single-molecule detection of DNA methylation levels in plasma. We incubated Methyl-CpG-binding domain protein 2 (MBD2-biotin), which specifically binds to methylated DNA[32], with fluorescently labeled plasma cfDNA. Bound complexes were anchored to the surface and imaged (Fig. 2h). Specificity and sensitivity were validated using synthetic methylated/unmethylated DNA, as well as DNA from cells treated with the DNA methyl transferase (DNMT) inhibitor 5-Aza-2′-deoxycytidine (Fig. 2i and Supplementary Fig. 4a,b). Finally, we verified detection of cfDNA methylation levels from plasma of CRC and healthy subjects (Fig. 2j).

We applied EPINUC to generate high-dimensional data, comprising of the three layers of information; histone PTMs, DNA methylation and protein biomarkers, from 33 plasma samples of

healthy subjects and 29 samples taken from 23 late stage CRC patients (stages III-IV; six patients were sampled twice at different times during cancer progression and treatment). CRC samples were obtained from patients prior to surgery or from patients that underwent surgical resection procedure and chemotherapy. In accordance with its use in clinical diagnostics[24], single-molecule counting of CEA showed higher levels in CRC patients (Fig. 3a and Supplementary Fig. 5a,b), and a reduction in patients after resection. Interestingly, high CEA levels were also observed in a few healthy individuals, generating a 'false positive' signal (for example, sample 19, marked by * in Supplementary Fig. 5a). Simultaneous probing of MST1, an anti-proliferative factor, allowed us to derive the CEA/MST1 ratio, resulting in better classification of samples and highlighting an advantage of combinatorial biomarker detection (Fig. 3b,c). Of note, plasma from CRC patients following resection exhibited altered CEA/MST1 ratio compared to non-resected patients, showing higher similarity to healthy individuals (Fig. 3a,b). This demonstrates the potential applicability of our technology to monitor treatment, while underlying the need to collect additional information from each sample to allow correct sample classification.

EPINUC also provides quantitative measurements of the total number of cfNucleosomes, six histone PTMs, their pairwise combinations and ratios per plasma sample (Fig. 1). In agreement with the literature, CRC patients had higher cfNucleosomes in their plasma compared to healthy controls[33] (Supplementary Fig. 5c). While most epigenetic parameters did not change, several showed significant differences: CRC patients had higher levels of H3K27me3-, H3K9me3- , H3K9ac- and H3K4me1-modified nucleosomes, and higher ratio of H3K9ac to H3K4me1 (Fig. 3d and Supplementary Fig. 5c,d). Interestingly, the combinatorial pattern of H3K9me3+H3K36me3-modified nucleosomes decreased in CRC, concomitant with an increase in 'bivalent' nucleosomes marked by H3K4me3+H3K27me3 (Fig. 3d). As bivalent chromatin is strongly implicated in many types of cancers[21,34], this result further confirms the diagnostic value of single-molecule quantification of combinatorial histone marks. DNA methylation was reduced in CRC samples, in agreement with previous studies[35,36] (Fig. 3e).

The identification of epigenetic and biomarkers alterations in late stage CRC motivated us to apply EPINUC to eight plasma samples from individuals diagnosed with early stage CRC (stage II). As in the later stage, the levels of DNA methylation, CEA and CEA/MST1 ratio significantly differed in early stage cancer patients versus healthy (Fig. 3f and Supplementary Fig. 5e). Interestingly, TIMP1, whose levels did not alter between our cohort of healthy and late stage CRC, was elevated

at the early stage (Supplementary Fig. 5e,f). This may reflect the downregulation of TIMP1 following chemotherapy[37], rendering it a significant biomarker only for early stage. Of note, plasma from stage II CRC patients also showed elevated levels of H3K27me3- and H3K9me3-modified nucleosomes, as seen in the late stage (Fig. 3f). Interestingly, we did not observe increased levels of cfNucleosomes in early stage CRC, likely due to the low tumor burden (Supplementary Fig. 5e). While the levels of H3K9ac- and H3K4me1-modified nucleosomes did not differ significantly from the healthy group, the combinatorial pattern of H3K4me1 and H3K9ac was lower in early CRC patients compared to healthy (Fig. 3f). Finally, we calculated the predictive score of each parameter alone to discriminate between the healthy and the distinct groups of CRC patients (Fig. 3g, Methods). These results highlight EPINUC's capabilities in providing multiplexed single-molecule measurements of protein biomarkers, epigenetic modifications and their combinations for CRC diagnostics.

To visualize the distribution of samples across the most significant and predictive parameters, we performed Principal Component Analysis (PCA). The PCA showed spatial separation between the groups, with the early stage CRC samples positioned in between the healthy and the late-stage CRC, potentially reflecting a transition stage (Fig. 3h). While samples from healthy individuals formed a tight cluster, the cancer samples showed greater variability, likely due to inherent heterogeneity between tumors. CRC patients who underwent resection also exhibited a high heterogeneity; interestingly, patients who received both primary tumor resection and metastectomy were positioned closer to the healthy group (samples 4118, 4211 and 4050).

A few CRC patients in our cohort were sampled twice along the course of the study, allowing us to examine the projection of samples taken from the same individual in the PCA plot (Fig. 3h, marked in *, $^{\#}$, and $^{+}$). Sample 4090 was collected two months following sample 4075 from a CRC patient who underwent tumor resection and extensive treatments. Unfortunately, her condition did not improve and she passed away a month later; indeed the later sample projects further from healthy on both principle components. A similar trend can be seen for samples 3488 and 4059, taken 6.5 months apart. These results highlight the potential of EPINUC to monitor patients' positive or negative response to treatment, and the power of collecting multiple layers of information from each sample.

Finally, in order to integrate all measurements and fine-tune the discrimination between healthy and CRC samples, we employed machine-learning classification (Fig. 4a, Methods). The best

predictive model displayed high diagnostic potential by generating a 0.96 AUC [95% confidence interval (CI) 0.935 - 0.981], and sensitivity of 88% [95% CI 82.9 – 93.3] at 90% specificity [95% CI 84 – 94.8] and 91% precision [95% CI 87.1 – 95.3], outperforming predictive models relying solely on protein biomarkers or DNA methylation coupled with biomarkers (Fig. 4a and Supplementary Fig. 6a). Intriguingly, this high discrimination power is achieved without including DNA sequencing. This is mainly due to the combination of multiple parameters spanning various cellular pathways into a single assay, and the high accuracy of the single-molecule technology that allows for digital counting of molecules.

We hypothesized that introducing a sequencing feature for samples that were classified as cancerous by the machine-learning algorithm would provide yet another layer of specificity and sensitivity. As different tissues vary in their epigenetic modifications, it may allow detection of the tissue-of-origin of the circulating nucleosomes, thus revealing the origin of the cancer. To that end, we coupled the epigenetic analysis with single-molecule DNA sequencing[21] (Fig. 4b and Supplementary Fig. 6b). Briefly, following detection of histone PTMs on cfNucleosomes, the histone proteins are evicted, and the DNA is subjected to repeated cycles of sequencing-by-synthesis using an automated fluidics system. Each cycle consists of incorporation of A, C, T or G by DNA polymerase and imaging; following 120 cycles, the data is integrated to build a strand that can be aligned to the genome, corresponding to the position of the modified nucleosome.

As proof-of-concept, we applied EPINUC followed by sequencing (EPINUC-seq) to two plasma samples of late stage CRC probed for H3K4me3 and H3K27me3 (Fig. 4c,d,e, Supplementary Fig. 6c,d, Methods). Single-molecule mapped reads, corresponding to modified nucleosomes, were intersected with unique antibody peak signatures generated from ENCODE ChIP-seq data for various tissues and primary cell lines, followed by bootstrapping simulations to calculate significance. Reinforcing our hypothesis, we found that both plasma samples showed significant overlap with colon-specific H3K4me3 and H3K27me3 peaks, indicating colon as the main tissue-of-origin (Fig. 4d and Supplementary Fig. 6c,d). Validating our approach, a similar analysis of a plasma sample from lung cancer patient revealed lung tissue as the main origin (Supplementary Fig. 6e). Moreover, comparing our data to a recent ChIP-seq study of H3K4me3 in plasma showed significant overlap with profiles obtained from CRC patients[13], but not with healthy plasma (Fig. 4e). H3K27me3 mapped reads showed a broader pattern, overlapping with peaks corresponding to hematopoietic lineage as well as colon (Fig. 4d and Supplementary Fig. 6d). Interestingly, we

also observed a significant overlap with liver-specific H3K27me3 peaks, in agreement with clinical analysis indicating this CRC patient (patient 4044) had liver metastasis.

Finally, we combined a complementary single-molecule approach to identify the tumor tissue-of-origin, by single-molecule profiling of the DNA modification 5-Hydroxymethylcytosine (5hmC). 5hmC is known to play important roles in gene regulation and cancer[38,39]. We captured 5hmC-enriched DNA from plasma as previously described[40], followed by single-molecule DNA sequencing (Supplementary Fig. 7a, Methods). In agreement with previous reports[41], we found 5hmC in cfDNA to be enriched at gene bodies and promoter proximal regions (Supplementary Fig. 7b), which are also known to be marked with H3K36me3. Thus, we generated unique 5hmC read signatures for healthy and CRC samples, and examined their overlap with H3K36me3 peak signatures from various tissues and primary cell lines (Methods). Similar to sequencing of the histone marks (Fig. 4d), analysis of 5hmC in the plasma of the same patient showed highly significant overlap with the colon-specific profile, validating this strategy for identification of the tumor tissue-of-origin (Fig. 4f and Supplementary Fig. 7c). Finally, we showed correct identification of colon origin also for early stage CRC patients (Fig. 4g and Supplementary Fig. 7d).

Our work establishes EPINUC as a novel liquid biopsy approach that analyzes multiple histone and DNA modifications, as well as protein biomarkers, at single-molecule precision. EPINUC distinguishes between CRC patients to healthy individuals at high specificity and sensitivity. We showed that this multi-parametric approach is suitable also for detection of early stage patients, although expanding the analysis to a larger cohort is needed. The main challenges with analyzing plasma nucleosomes are (1) their minute amount- in 1 ml of plasma there are ~1000 genome copies[13,42]; (2) The plasma is highly dense with additional proteins, rendering enzymatic or binding approaches to capture nucleosomes difficult; (3) There is high variability between different individuals, stressing the need for quantitative methodologies to allow comparison between samples; and (4) Multi-parametric data is needed to achieve high specificity and confidence in detection. Our EPINUC approach addresses these challenges by enabling single-molecule combinatorial detection of epigenetic marks, DNA sequencing and protein biomarkers from limited input material. In addition to the unique epigenetic analysis, the single-molecule system outperforms the classical ELISA assay for measuring protein biomarkers. ELISA is of relatively low sensitivity and is therefore limited to proteins that are present at high levels, has lower dynamic

range in quantifying proteins, and is not amenable to multiplexed detection of several proteins[27,28]. We showed that the single-molecule system is capable of detecting the mutant form of p53, which is a non-secreted protein that originates directly from the tumor cells. Importantly, the system is straightforward to adapt for detection of additional proteins, thus increasing sensitivity and enabling disease-specific biomarkers analysis.

EPINUC is built on the idea that integration of multiple parameters would provide specific diagnosis, which is independent on DNA sequencing. This would potentially render the EPINUC approach fast and inexpensive, paving the way to a mass screening method for a variety of cancers. Nevertheless, further studies are needed to test whether this technology can differentiate between various cancer types based solely on the epigenetic and biomarker profiles. Such studies may also provide insights on whether the epigenetic differences detected by EPINUC originate solely from the tumor cells, or may represent, at least in part, epigenetic alterations occurring in non-cancer cells. Importantly, we showed that the combination of EPINUC with single-molecule DNA sequencing provides unequivocal evidence of the tumor tissue-of-origin, thus further elucidating the type of cancer of each individual. This technology greatly expands the already burgeoning field of liquid biopsies and has the potential to be applicable for early detection of cancer and monitoring.

Acknowledgments: We thank Dr. R. Rozenzweing, Dr. O. Fasust, Mr. M. Maurer, and Ms. R. Irwin for their contribution in establishing a protocol for Methyl-CpG-binding domain protein 2 labeling. We are grateful for the important comments made by I. Ulitsky while reading the manuscript.

Funding: E.S. is an incumbent of the Lisa and Jeffrey Aronin Family Career Development chair. This research was supported by grants from the European Research Council (ERC801655, ERC_PoC_963863), The Israeli Science Foundation (1881/19), The German-Israeli Foundation for Scientific Research and Development and Minerva.

Author contributions: V.F, N.E, and E.S designed the study and wrote the manuscript. V.F and N.E performed the experiments and analyzed the data. N.F and O.B assisted in the experiments, G.R assisted with data analysis. B.B.Z, Y.M, T.P, A.H, J.E.C, A.S, M.T, A.G, M.M and A.Z collected the plasma samples of early and late stage CRC patients. D.J, A.S, K.A contributed to

the development of single-nucleosomes imaging technology and sequencing experiments described in this study.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data is available in the main text or the supplementary materials.

Methods

Patients

All clinical studies were approved by the local ethics committees (Helsinki applications 091-2020 and 0198-14-HMO). Informed consent was obtained from all individuals before blood sampling.

Plasma collection

Blood samples were collected in VACUETTE K3 EDTA tubes and transferred immediately to ice. Next, blood was centrifuged (10 minutes, 1,500g, 4°C) and the supernatant was transferred to a fresh 50-ml tubes and centrifuged again (10 minutes, 3,000g, 4°C). The supernatant was collected and used as plasma for all experiments. Plasma was analyzed fresh or flash-frozen and stored at −80°C for future analysis.

Cell-free nucleosomes (cfNucleosomes) preparation for single-molecule imaging

Tagging and tailing of cfNucleosomes was carried out as following: 20 µl of plasma or 5X (DDW diluted) concentrated apoptotic medium was incubated at 37°C for 1 hour with the following reaction mixture: 10 µl 10X Green Buffer (Enzymatics, B0120), 416 µM CoCl2 (Enzymatics, B0220), 1:60 PI (SIGMA, P8340), 83.3 nM fluorescently labeled dATP (Jena Bioscience, NU-1611-Cy3/Cy5), 83.3 µM dATP (Thermo Fisher Scientific, R0181), 5 µl of Klenow Fragment (3'→5' exo-, NEB, M0212S) 3µl of T4 Polynucleotide Kinase (NEB, M0201L) and 4 µl of Terminal deoxynucleotidyl transferase (TdT, Enzymatics, P7070L). Following incubation, samples were inactivated by immediate transfer to ice. For nucleosome sequencing, 1.67 µl of ddATP was added (SIGMA, GE27-2051-01).

Plasma cell-free DNA (cfDNA) isolation and fluorescent labeling

cfDNA was extracted from 4 ml of healthy human blood plasma, or from 0.5 ml of plasma from CRC patients, using the Mag-Bind cfDNA Kit (Omega Bio-Tek, M3298-01). For optimized yield, protocol was modified by increasing elution time to 20 minutes on a thermomixer, at 1,600 rpm, in 15 µl elution buffer at room temperature. Sample concentration was measured using Qubit Fluorometer (Thermo Fisher Scientific). For fluorescent labeling of plasma isolated DNA, 10 µl of cfDNA was incubated at 37°C for 1 hour with the following reaction mixture: NEBuffer™ 2 (NEB, B7202), 0.25 mM MnCl2 (SIGMA, M1787), 33 µM fluorescently labeled dATP (Jena Bioscience, NU-1611-Cy3), 1.5 µl of Klenow Fragment (3'→5' exo-, NEB, M0212S) and 1.5 µl of T4 Polynucleotide Kinase (NEB, M0201L). Following incubation, samples were inactivated by addition of EDTA (Invitrogen, 15575-038) at a final concentration of 20mM. Next, DNA was

purified by AMPure SPRI beads (Beckman Coulter, A63881), and quantified by Qubit (Thermo Fisher Scientific).

Cell culture and apoptosis

Cell lines were maintained at 37˚C with 5% CO2. HEK-293 cells were cultured in 150 cm plates (10×10^6 cells in 20 ml of media) in DMEM supplemented with 10% FBS and 1% P/S, and passaged every week. For induction of apoptosis, 6 µM of Staurosporine (STS, Holland-Moran, 62996-74-1.25) was added to medium of confluent cells. 72 hours later, medium was collected and immediately processed. To verify fragment sizes along with nucleosome labeling, 10 ul of the nucleosomes and 10ul of AMPure extracted DNA (either directly from concentrated medium or after the tagging and tailing reaction), were loaded on High Sensitivity D1000 ScreenTapes (Agilent, 5067-5584) and 6% TBE gel (ThermoFisher Scientific, EC62655BOX), and imaged with 4200 TapeStation (Agilent) or Typhoon imager (Amersham Biosciences), respectively. Apoptotic medium cfNucleosomes were concentrated and recovered using Centricon Plus-70 centrifugation filteres (Merck, UFC710008) according to the manufacture protocol. PI was supplemented 1:100 following concentration.

Surface preparation for single-molecule imaging

PEGylated-Biotin and PEGylated-poly T coated microscope slides were prepared based on the protocol described by Chandradoss et al[43]. Ibidi glass coverslips (25 mm x 75 mm, IBIDI, IBD-10812) were cleaned with (1) MilliQ H2O (3X washes, 5 minutes sonication, 3X washes); (2) 2% Alconox (SIGMA, 242985) (20 min sonication followed by 5X washes with MilliQ H2O); and (3) 100% Acetone (20 min sonication followed by 3X washes with MilliQ H2O). Slides were further cleaned and functionalized (Hydroxylated) by incubation in 1 M KOH (SIGMA, 484016) solution for 20 minutes while sonicated, followed by 3X washes with MilliQ H2O. Slides were sonicated twice for 10 minutes in 100% HPLC ethanol (J.T baker 8462-25) prior to applying amino-silanization chemistry. Next, slides were incubated for 24 minutes in a mixture of 3% 3-Aminopropyltriethoxysilane (ACROS Organics, 430941000) and 5% acetic acid in HPLC EtOH, with 1 minute sonication in the middle. Slides were then washed with HPLC EtOH (3X) and MilliQ H2O (3X) and dried with N2. Surface functionalization along with first passivation step was performed by applying mPEG: PEGylated-Biotin/PEG-Azide solution [20 mg PEGylated-Biotin (Laysan, Biotin-PEG-SVA-5000), 180 mg mPEG (Laysan, MPEG-SVA-5000) or 20 mg PEG-Azide (JenKem, A5088), 180 mg mPEG (Laysan, MPEG-SVA-5000)] dissolved in 1560 ul 0.1 M Sodium Bicarbonate (SIGMA, S6297) and degassed (centrifugation at 1 minute at 16,000*g*).

Next, 140 µl of solution was applied on one surface, followed by immediate assembly of another surface on top. Each pair of assembled surfaces were incubated overnight in a dark humid environment.

For PEGylated-Biotin surfaces: At the next day, surfaces were washed with MilliQ H2O and dried with N2 followed by a second passivation step. MS(PEG)4 (ThermoFisher Scientific, TS-22341) was diluted in 0.1 M of sodium bicarbonate to a final concentration of 11.7 mg/ml and applied on one surface, followed by the assembly of another surface on top. Each pair of assembled surfaces were incubated overnight in dark humid environment. The next day, surfaces were disassembled, washed with MilliQ H2O and dried with nitrogen. After nitrogen flush, surfaces were stored in -20°C.

For PEGylated-poly T surfaces, following PEG-Azide coating, surfaces were washed with MilliQ H2O and dried with N2. To enable anchoring of dT50 to surface via click chemistry, 10 µM of 5'heyxynyl-dT50 (IDT) were mixed with 2 mM of CuSO4 (SIGMA, C1297) and DDW. Next, 100 µl of the mixture was applied on one surface, followed by immediate assembly of another surface on top. Each pair of assembled surfaces was incubated overnight in a dark humid environment. In the next day, a second passivation step [MS(PEG)4] was carried out, similarly to PEGylated-Biotin preparation. Surfaces were stored in -20°C post nitrogen flush in a similar fashion.

Antibody labeling

Capture and detection antibodies were labeled using Biotin conjugation kit (Abcam, ab201796) and Alexa flour antibody labeling kits (Thermo Fisher Scientific, A20181/ A10237/A20186) according to the manufacture protocol. Labeled antibodies were purified by size exclusion chromatography using Bio-Spin 6 columns (Bio-Rad, 7326200) followed by measurement of protein concertation using *Nanodrop* 2000 at 260 nm.

TIMP-1 imaging and siRNA transfections

siRNA transfection was performed using INTERFERin (Polyplus, 409-10) according to the manufacturer's protocol. Briefly, cells were plated in 6-well plates (1.5×10^5 in 2.5 ml per well) overnight, and the 200 µl of transfection complex was added directly to medium, at final concentration of 25 nM of siRNA. RNA and protein samples were isolated from cells 72 hours after transfection. The following siRNA was used: SMARTpool: ON-TARGETplus Human TIMP1 siRNA (L-011792-00-0005, Dharmacon). For single-molecule imaging, medium was collected from plates, followed by centrifugation at max speed in 4°C and collection of supernatant to separate proteins from cell debris. Protein concentration was determined by Pierce™ BCA

Protein Assay (Thermo Fisher Scientific, 23225), followed by addition (1:100) of protease inhibitor cocktail (PI, SIGMA, P8340).

Synthetic DNA preparation for DNA methylation assay

DNA fragments were generated by conventional PCR (primer sites underlined) supplementing the reaction with either methylated (NEB, N0356S) or un-methylated cytosine (Thermo Fisher Scientific, R0181), followed by purification with AMPure SPRI beads. The size (~200 bp) was chosen to mimic the size of mono-nucleosomal DNA fragments previously identified in blood plasma[44]. Fragment labeling, purification and quantification was performed as described for plasma cfDNA.

Sequence:

CATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAA TCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACA TACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC ACA

5-Aza-2-Deoxycytidine Treatment

HEK-293 cells were plated in 150 cm plates (10×10^6 cells in 20 ml of media) overnight, then treated with 1 uM of 5-Aza-2-deoxycytidine (5 –Aza, SIGMA, A3656) or PBS for 4 days. Next, 5×10^6 cells were collected and washed with PBS supplemented with PI (1:100), followed by centrifugation at 3000 rpm for 3 minutes. Cell pellet was resuspended with 1 ml of 0.05% Igepal (SIGMA, I8896) diluted in PBS (supplemented with PI as mentioned above) and centrifuged again at 3000 rpm for 3 minutes. Next, the pellet was resuspended in Lysis buffer [100 mM Tris-HCl pH 7.5 (Gibco, 115567-027), 300 mM NaCl (J.T Baker, 7647145), 2% Triton® X-100 (SIGMA, 9002 93-1), 0.2% sodium deoxycholate (SIGMA, D6750), 10 mM CaCl2 (SIGMA, 21115)] supplemented with PI and Micrococcal Nuclease (ThermoFisher Scientific, 88216). The reaction mixture was incubated at 37°C for 10 minutes and then inactivated by addition of EGTA at a final concentration of 20mM. Then, lysate was centrifuged for 10 minutes at max speed and supernatant was transferred to a new tube. DNA extraction, fluorescent labeling and quantification was performed as described for plasma cfDNA.

Single-molecule imaging

PEGylated-Biotin and PEGylated-poly T coated coverslips were assembled into an Ibidi flowcell (Sticky Slide VI hydrophobic, IBIDI, IBD-80608) generating a six lane flowcell, which enables

imaging of six different samples or various combinations of antibodies on a single surface. For PEGylated-Biotin flowcells, Streptavidin (SIGMA, S4762) was added to a final concentration of 0.2 mg/ml followed by 10 minutes incubation and washing with imaging buffer [IMB: 12 mM HEPES pH 8 (Thermo Fisher Scientific, 15630056), 40 mM TRIS pH 7.5 (Gibco, 115567-027) 60 mM KCL (SIGMA, 60142), 0.32 mM EDTA (Invitrogen, 15575-038), 3 mM MgCl2 (SIGMA, 63069), 10% glycerol (Bio-Lab, 56815) , 0.1 mg/ml BSA (SIGMA, A7906) and 0.02% Igepal (SIGMA, I8896)]. For time-lapse imaging experiments (Histone PTMs, p53), prior to sample application, TetraSpeck beads (ThermoFisher Scientific, T7279) diluted in PBS were added and incubated on surface for at least 10 minutes to allow correction for stage drift in image analysis. Imaging was performed on a total internal reflection (TIRF) microscope, Nikon (Ti2 LU-N4 TIRF).

Histone PTMs analysis

PEGylated-poly T coated coverslips were assembled as described and further passivated with 5% BSA (Merck, A7906) for 30 minutes followed by wash with IMB. Next, plasma sample containing tailed and fluorescently labeled cfNucleosomes was incubated with antibodies (diluted 1:60) for 30 minutes at room temperature (RT), to allow formation of antibody-cfNucleosomes complexes. Next, samples were loaded on the surface and incubated for 15 minutes to allow hybridization. Flowcell was washed (X3) with IMB, followed by time lapse imaging every 15 minutes, with the three laser channels, across all positions (50 Fields of View (FOVs, 148μm^2) per experiment).

Protein biomarkers analysis

PEGylated-Biotin coated coverslips were assembled and coated with streptavidin. Biotinylated antibodies were incubated on surface in IMB2 [10 mM MES pH 6.5 (Boston Bioproducts Inc, NC9904354), 60 mM KCL, 0.32 mM EDTA, 3 mM MgCl2, 10% glycerol, 0.1 mg/ml BSA and 0.02% Igepal] for 30 minutes, followed by wash with IMB2. Next, plasma sample was added to flowcell and incubated on surface for 30 minutes, followed by washes (3X) with IMB2, to allow binding of target proteins. Fluorescently labeled antibodies (detection antibodies) were introduced to the surface for 60 minutes, washed with IMB2, and imaged.

Global DNA methylation analysis

PEGylated-Biotin coated coverslips were assembled and coated with streptavidin. 2 µl of MBD2-Biotin (Thermo Fisher Scientific, A11148) was incubated with 8 µl of Cy3 labeled cfDNA fragments for 30 minutes, to allow MBD2-Biotin binding to methylated DNA. Next, the reaction mixture was immobilized on the surface and incubated for 10 minutes, followed by TIRF imaging.

DNA Hydroxymethylation analysis

cfDNA was incubated in 25 µl reaction mixture containing 50 mM HEPES buffer (pH 8), 25 mM $MgCl_2$, 60 µM UDP-6-N_3-Glc (Jena Bioscience, CLK-076Motif) and 12.5 U T4 beta-glucosyltransferase (Thermo Fisher Scientific, EO0831) for 2 hours at 37°C. Next, 5 µl DBCO-S-S-biotin (Click Chemistry Tools, 10 mM stock in DMSO) was directly added to the reaction mixture and incubated overnight at 37°C. DNA was cleaned using Oligo Clean & Concentrator (Zymo, D4060), and immobilized on a PEGylated-Biotin streptavidin coated surface, followed by imaging.

Single-molecule DNA sequencing

For single-molecule DNA sequencing of cfNucleosomes, PEGylated-poly T surface was blocked with BSA as described above. Poly-A tailed FluoSpheres (described below) along with TetraSpeck beads and cfNucleosomes were applied to surface. PTMs of plasma cfNucleosomes were imaged over time for 169 FOV, as described above. Then flowcell was washed with Wash A buffer [150 mM HEPES (KOH, pH 7.0), 1×SSC, 0.1% SDS] and Wash B buffer [150 mM HEPES (KOH, pH 7.0), 150 mM NaCl] to evict histones and antibodies, and temperature was increased to 37°C. Sequencing was performed as described previously, using Helicos True Single Molecule Sequencing (http://seqll.com/)[21,45]. FluoSpheres preparation: FluoSpheres (Carboxylate-Modified Microspheres, Thermo Fisher Scientific, F8789) were conjugated to dA50-amine (IDT), tailed as previously described, and hybridized to the surface to serve as reference points for stage drift correction during alignment of sequencing images.

Single-Molecule Hydroxymethylation sequencing

2.5 ng of plasma cfDNA was added to a 25 µl solution containing 50 mM HEPES buffer (pH 8), 25 mM $MgCl_2$, 60 µM UDP-6-N_3-Glc (Jena Bioscience, CLK-076) and 12.5 U T4 beta-glucosyltransferase (Thermo Fisher Scientific, EO0831), and incubated for 2 hours at 37°C. Next, 5 µl DBCO-S-S-biotin (Click Chemistry Tools, 10 mM stock in DMSO) was directly added to the reaction mixture and incubated overnight at 37°C. For 5hmC DNA pulldown, samples were incubated at RT with 15 µl of Streptavidin beads (Thermo Fisher Scientific, Dy-11205D) for 1 hour, followed by 3 washes with 1X wash buffer, and elution in 20 µl of 125 mM TCEP (Thermo Fisher Scientific, TS-77720). 5hmC eluted DNA was poly-adenylated and sequenced on a PEGylated-poly T surface as described above. Metagene profile was generated using ngs.plot.

Image analysis

Image analysis was performed with the open-source software Cell Profiler (http://www.cellprofiler.org/). Image analysis pipelines are available upon request. Briefly, time-lapse images of antibody binding events and TetraSpeck beads are aligned, stacked and summed to one image. Antibody spots can be differentiated from TetraSpeck beads spots based on spot size and intensity. Summed antibodies images are aligned with cfNucleosomes images to count colocalization events.

Predictive Power Score (PPS)

PPS analysis on the data was conducted using a previously published algorithm (https://github.com/8080labs/ppscore). Briefly, by calculating a cross-validated decisions tree for the target variable (e.g., diagnosis) using only one of the markers, it is possible to determine which of the markers in the datasets contributes most to the target variable. The PPS is normalized to the most common assignment in order to provide a baseline for comparison. Using the PPS rather than a simple correlation measure allows us to account for non-linear effects and provides an alternative formulation for correlation which also treats categorical variables (e.g., diagnosis, or disease state – see **Supplementary Table 3**).

Machine learning model for sample classification

For sample binary classification, various machine-learning algorithms were trained on the features that showed significant differences between healthy and CRC (Fig. 3c,d,e, Supplementary Fig. 5b,c), and evaluated for their performance using a four-fold cross-validation across all samples. The best predictive performance was achieved by a Logistic Regression classifier. To improve classifier performance, we conducted additional feature selection by training the classifier on all possible feature combinations out of the significant features aforementioned. Evaluating the resulting Area Under the Curve (AUC) values of repeated (500 iterations) four-fold cross-validation for each combination revealed an optimal cumulative performance of a five feature combination: H3K27me3/Nuc, H3K9me3/Nuc, CEA/MST1, H3K27me3+Nuc (% of cfNucleosomes harboring H3K27me3) and global DNA methylation. To evaluate the classifier overall performance using the selected features, we performed repeated (10K) 4-fold cross-validation across all samples. For each iteration the sensitivity, specificity, accuracy, precision and the AUC value were calculated and averaged over all iterations. R caret and Rweka packages were used for machine-learning modeling.

Tissue and plasma signatures

We downloaded and combined two independent *Homo sapiens* based ChIP-seq tracks for each tissue from the Encyclopedia of DNA Elements (ENCODE, **Supplementary Table 4**). To generate a unique antibody peak profile for a given tissue, we discarded reads found to overlap with at least one of the other seven tissues tested, retaining only tissue specific peaks. Of note, brain H3K36me3 peaks were available only for embryonic tissue, therefore were replaced with spleen tissue H3K36me3 peaks. To generate unique plasma H3K4me3 ChIP-seq peaks, we obtained data from healthy (H) and CRC (C) plasma (n=3 for each) produced by Sadeh et. al.[13] (**Supplementary Table 4**). For each group, reads were intersected and only shared reads across all samples were kept for further analysis. Non-overlapping reads between the overlapping healthy and overlapping CRC reads were defined as the unique plasma signature.

Bootstrapping simulation to analyze single-molecule reads overlap with various tissues Overlap significance was assessed as following: First, single-molecule sequenced plasma antibody aligned reads were extended by 100bp from each side to resemble nucleosome length. Then, for each chromosome, we randomly selected a number of 230bp-long DNA segments that is equivalent to the number of antibody positive plasma reads for this chromosome. Random reads were intersected with each unique tissue/plasma signature and overlapping events were recorded. These bootstrapping simulations were iterated 10K times for each tissue for a given antibody to generate a distribution of overlap by chance. Finally, single-molecule sequenced plasma antibody reads were intersected with all tissue signatures, and contrasted against the corresponding distribution of random overlap for that tissue to evaluate overlap significance (two tailed z-test or Wilcoxon rank sum test). Signature and overlap analysis was performed using an in-house R script (EPINUC-overlap), where minimal overlap was defined as 1bp overlap. The accession numbers for the ENCODE chip-seq datasets are summarized in **Supplementary Table 4**.

Statistical analysis

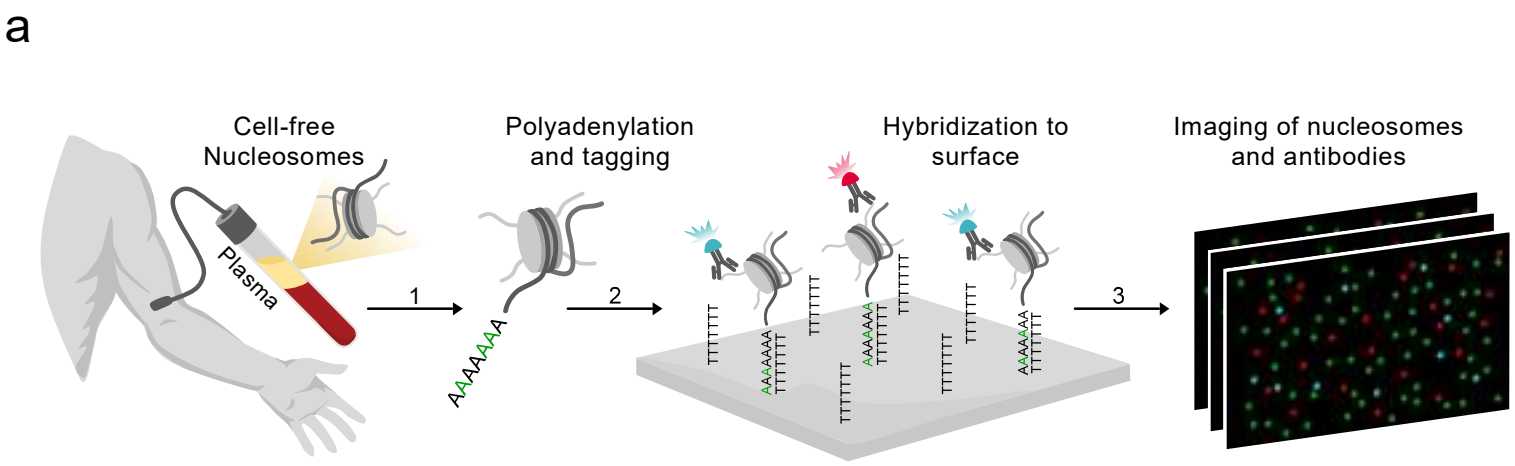
All the statistical analysis was conducted using the statistical programming language R. Multiple comparison (Supplementary Fig. 5d) was calculated for cfNucleosomes, DNA methylation and all other significant parameters via One-way ANOVA, Pairwise comparisons using Wilcoxon rank sum test with continuity correction and Asymptotic K-Sample Brown-Mood Median Test, respectively.

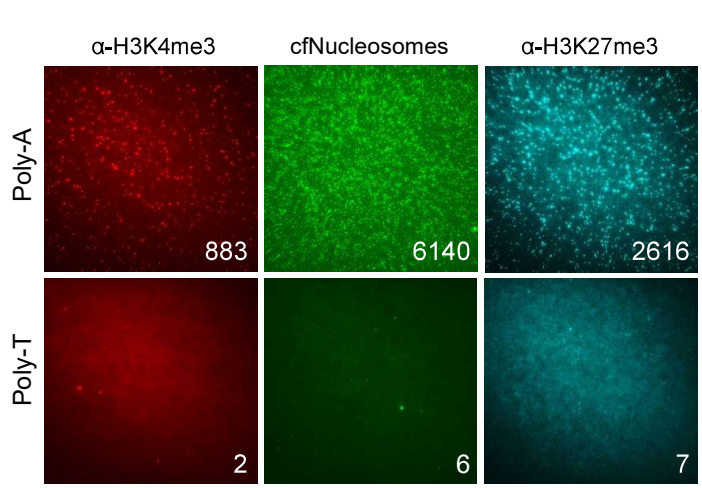
Data availability

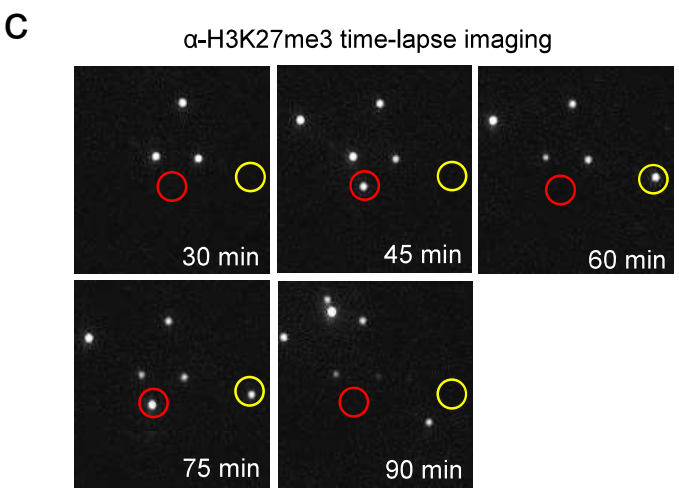
Datasets generated and analyzed during this study are summarized in **Supplementary Table 4**, BED files of plasma sequenced reads are available upon request.

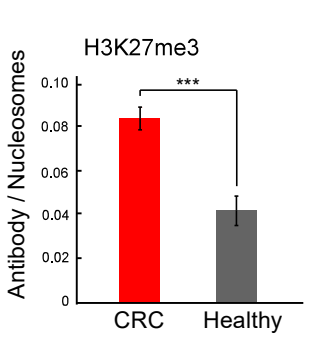
Code availability

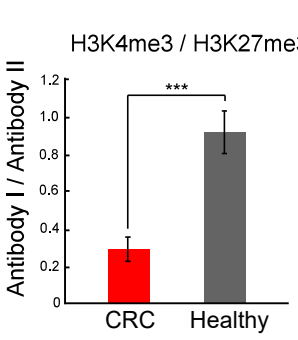
R code for performing overlap analysis is available at https://github.com/Vadim-Fed/EPINUC-overlap.

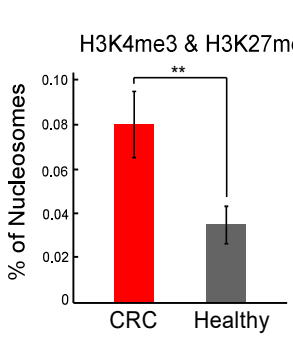


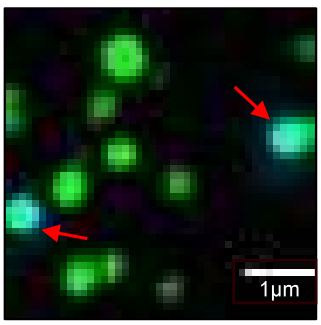


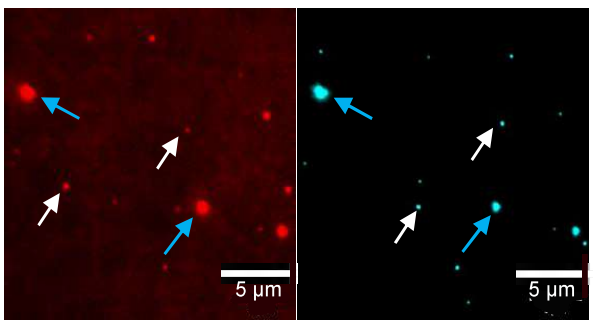












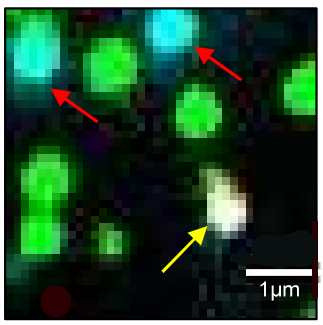


Fig.1: EPINUC decodes the combinatorial epigenetic states of plasma cell-free nucleosomes.

a, Experimental scheme: (1) Sample preparation procedure of cfNucleosomes is carried out in one-step and consists of two enzymatic processes: repair of DNA ends by Klenow polymerase, and addition of a poly A tail by Terminal Transferase (TdT). The reaction contains a mixture of natural dATPs and fluorescently labeled dATPs (Cy3-dATP) to label nucleosomes. (2) cfNucleosmes are captured on a PEGylated-poly T surface via dA:dT hybridization. Immobilized nucleosomes are incubated with fluorescently labeled antibodies targeting different histone modifications. (3) TIRF microscopy is applied to record nucleosome positions and generate time-lapse imaging of antibodies' binding events. **b,** Representative images of plasma-derived cfNucleosomes and the corresponding H3K4me3 and H3K27me3 signal (numbers within images represent counted spots). Each spot corresponds to a single nucleosome. Nucleosomes anchor to the surface specifically via hybridization, as evident from the lack of signal when tailed with dTTP (Poly-dT) rather than dATP (Poly-dA). **c,** Representative images of antibodies' binding and dissociation events over time from individual target molecules (marked by red/yellow circles). **d,e,f** Example for quantification and representative images of the various parameters measured by EPINUC in plasma samples from a healthy subject and a CRC patient. Zoomed-in image segments of entire field of view (148µm^2). **d,** The percentage of cfNucleosomes (green spots, Cy3) that are modified by H3K27me3 (cyan spots, AF488). Red arrows indicate co-localization events. Scale bar = 1µm **e,** Ratio between H3K4me3 (red, AF647) and H3K27me3 (cyan) antibodies. White arrows indicate antibody spots, blue arrows indicate TetraSpeck beads that are used for alignment. Scale bar = 5µm. **f,** cfNucleosomes marked by the combinatorial pattern of both H3K27me3 and H3K4me3. Red arrows indicate co-localization events of H3K27me3 only, yellow arrow indicates a combinatorially modified nucleosome. Scale bar = 1µm.

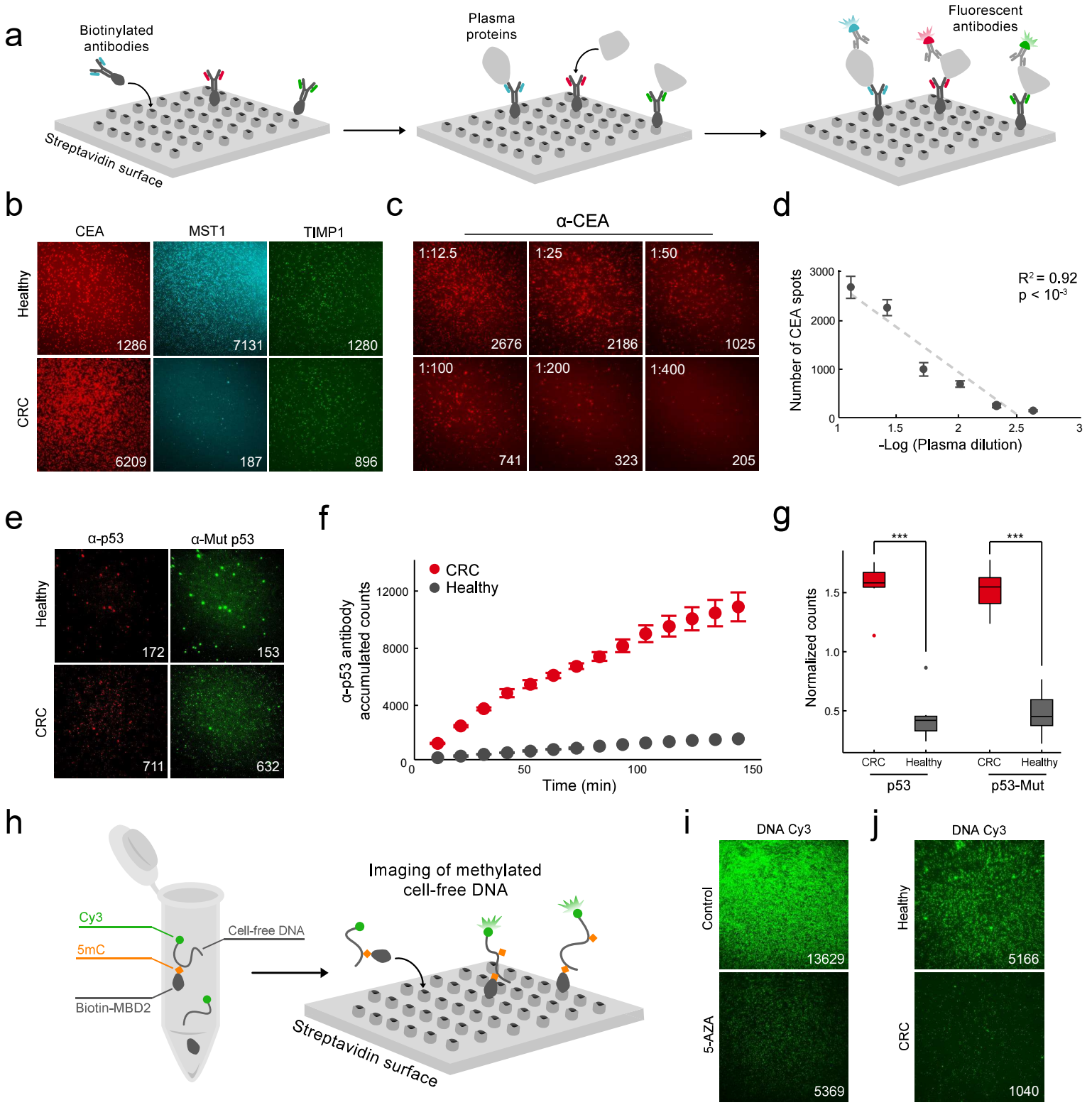
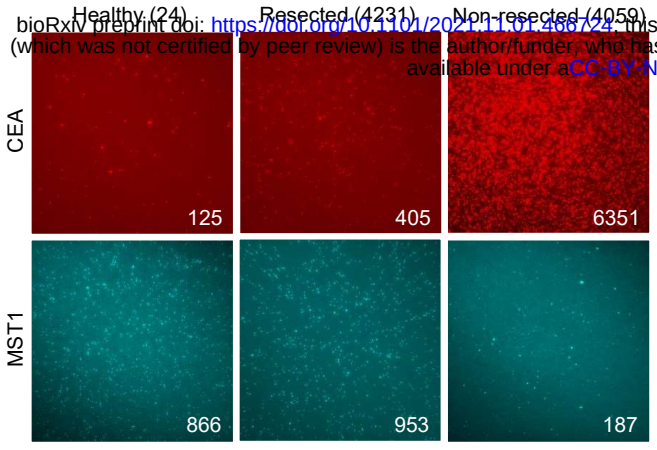
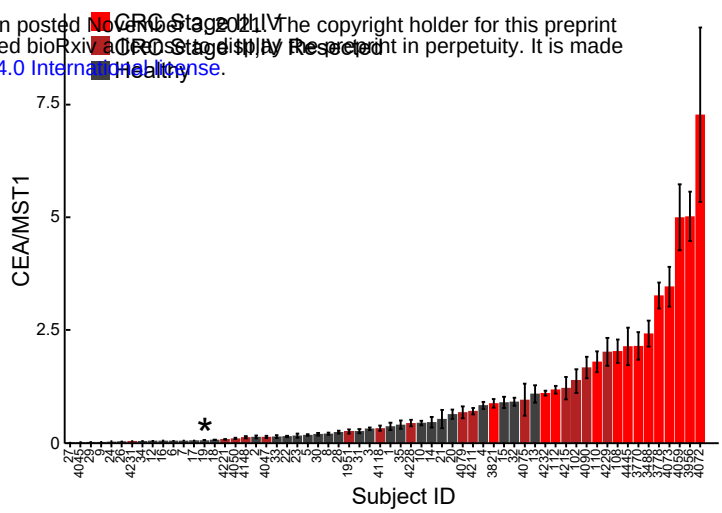
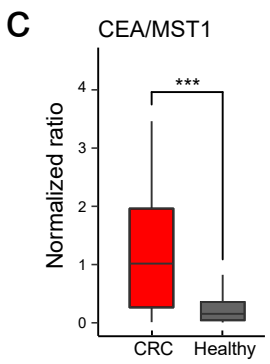


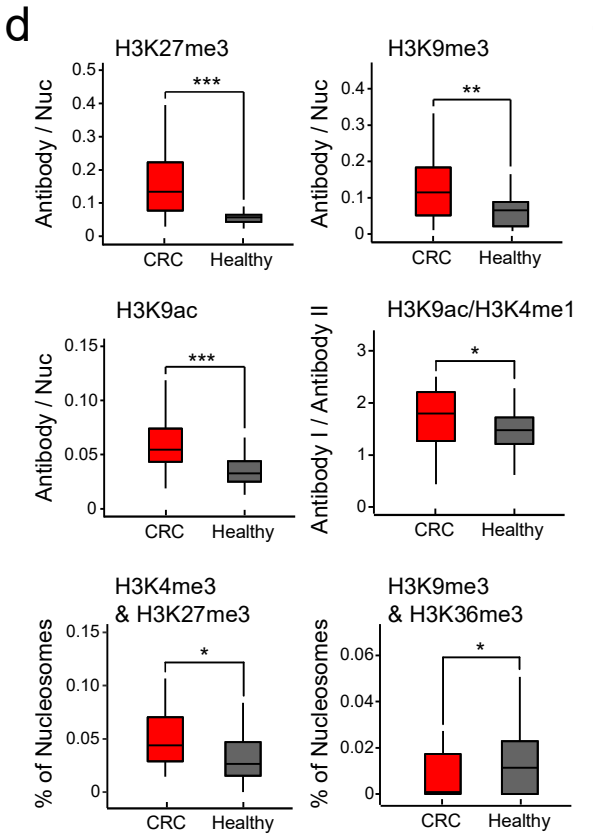
Fig.2: Multiplexed single-molecule detection of cancer-associated protein biomarkers, mutant p53 and DNA methylation.

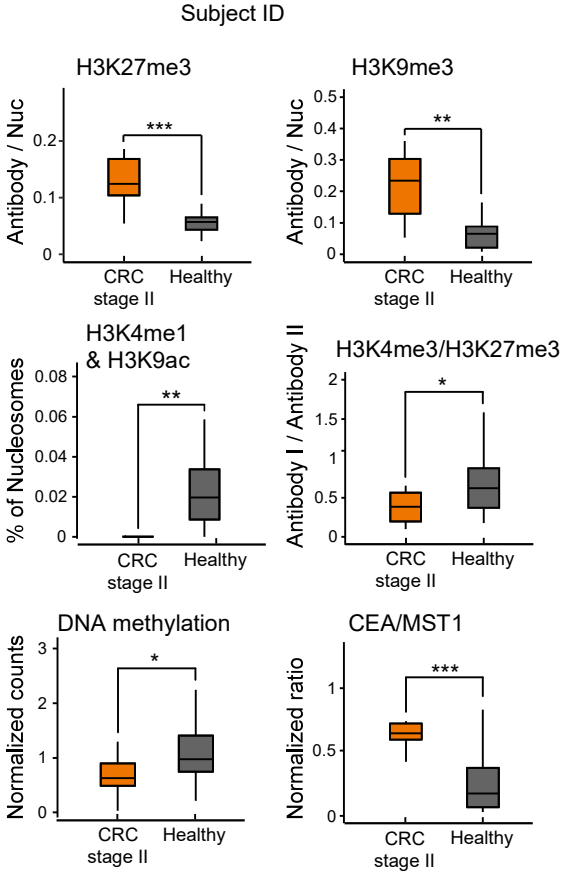
a, Experimental scheme: biotinylated capture antibodies targeting distinct proteins are anchored to a PEG-streptavidin surface. Plasma proteins are captured on surface, followed by detection with fluorescently labeled antibodies and TIRF imaging. Multiplexed detection of up to three proteins is achieved by labeling antibodies with different fluorophores. Each spot represents a single protein bound on the surface. **b,** Representative TIRF images of selected CRC biomarkers measured simultaneously for each plasma sample: CEA (red), MST1 (cyan) and TIMP-1(green). Images reveal distinct biomarkers profiles for healthy and CRC. **c,** Representative TIRF images depicting α-CEA antibody binding events (spots) over serial plasma dilutions. **d,** Regression analysis of the number of spots as a function of plasma concentration highlights the linearity of detection. Data is presented as the mean +/- s.d. of 50 FOVs for each concentration. **e,f,g** Single-molecule detection of p53 in the plasma of healthy and CRC patients with known p53 mutations. **e,** Representative TIRF images. Detection is carried out simultaneously with antibodies targeting all p53 (red) and with antibodies that are specific to the mutant p53 conformation (green). Large diameter spots correspond to TetraSpeck beads used for alignment. **f,** p53 signal accumulation over time in late stage CRC and healthy plasma. Data is presented as the mean +/- s.d. of 50 FOVs for each time point. **g,** Total and mutant p53 levels in plasma show significantly higher levels in CRC patients versus healthy individuals (n=6 for each group). Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than 1.5× interquartile range from the hinge. P values were calculated by Wilcoxon rank sum exact test. *** P value < 0.001. **h,** Experimental scheme for single-molecule imaging of global DNA methylation: MBD2-biotin is incubated with Cy3-labeled (green) DNA, and binds specifically to methylated DNA molecules. Next, biotin-MBD-meDNA complexes are immobilized on a PEG-streptavidin surface, followed by TIRF imaging. Each spot represents a single bound complex, number of spots correspond to the level of DNA methylation in plasma. **i,** Representative TIRF images of DNA methylation in HEK293 cells treated with 5-Aza-2′-deoxycytidine, demonstrating significant reduction in methylation compared to control cells. **j,** Representative TIRF images of global cfDNA methylation levels in the plasma of CRC and healthy subjects, showing lower DNA methylation levels in CRC. For all images, numbers within images represent counted spots.

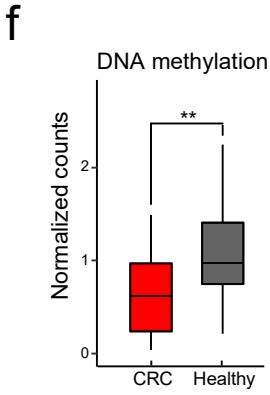


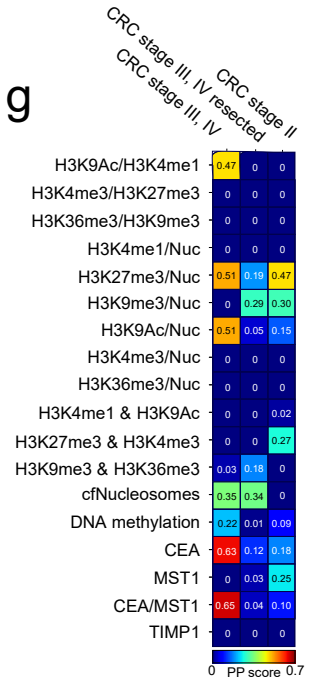












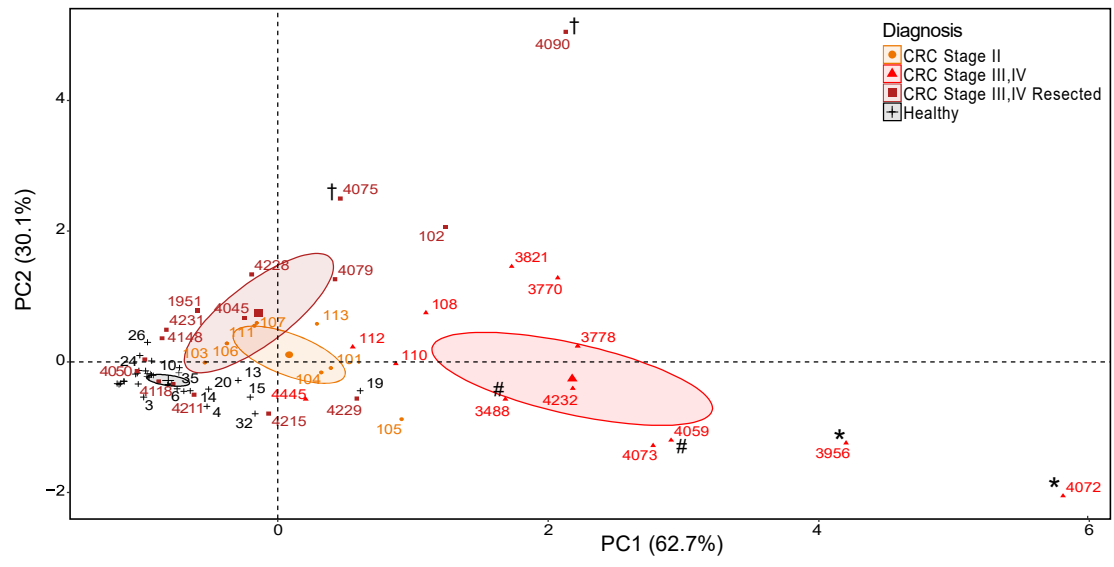
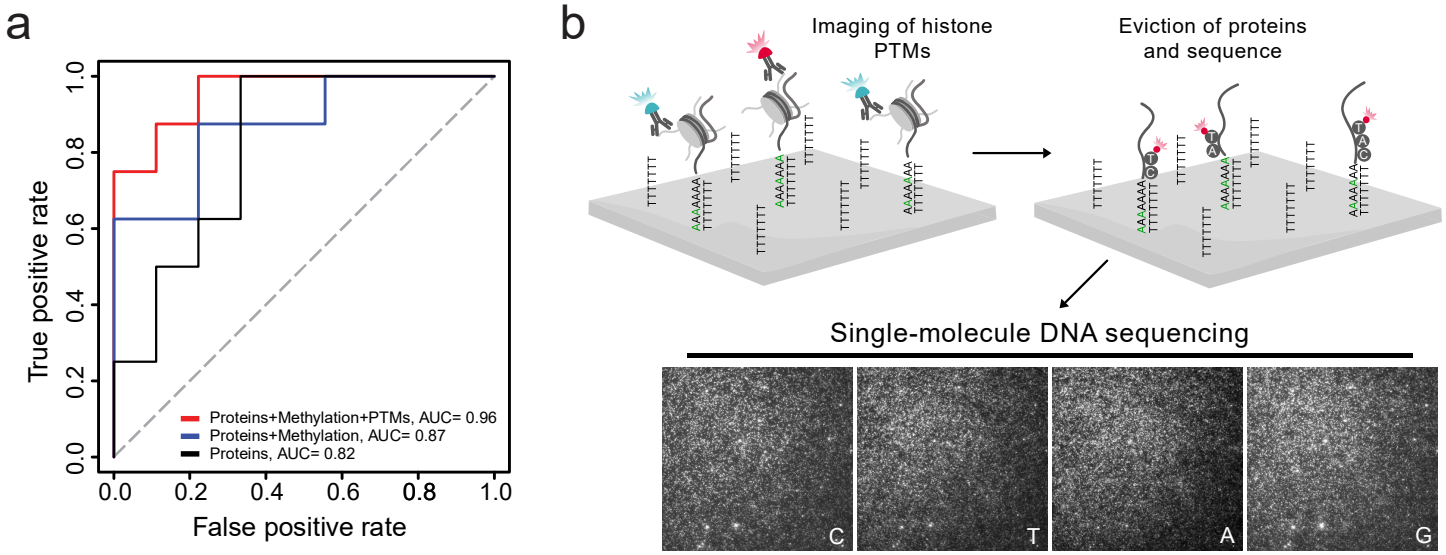
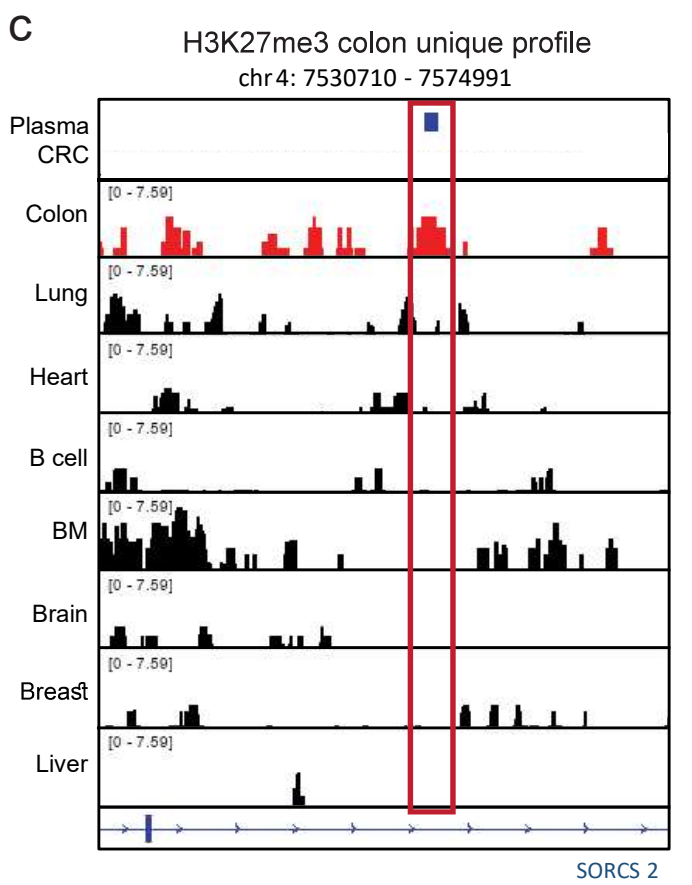
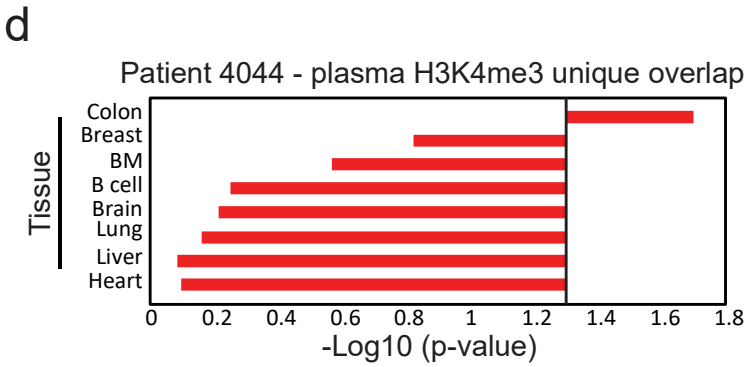


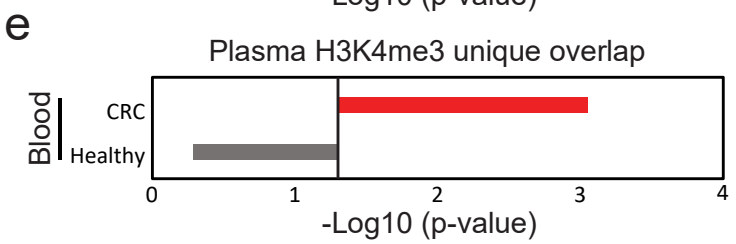
Fig.3: EPINUC reveals significant epigenetic and biomarkers alterations in the plasma of CRC patients.

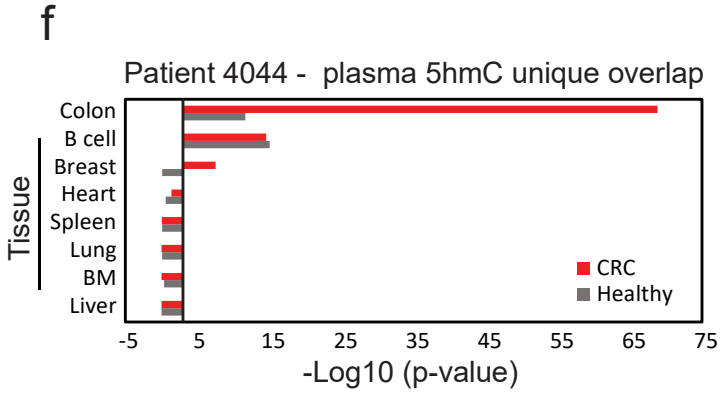
a, Representative TIRF images depict changes in protein biomarker levels in the plasma of healthy, CRC patient, and CRC following tumor resection. Numbers within images represent counted spots. **b,** CEA/MST1 normalized levels in the plasma of CRC patients and healthy individuals. Each bar represents a subject, data is presented as the mean +/- s.d. of 50 fields of view per sample. **c,** Box plot representation of the data in B (healthy = 33, CRC = 29). Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than 1.5× interquartile range from the hinge. P values were calculated by Welch's t-test. *** P value < 0.001. **d,** Histone PTMs, ratios and combinations (as indicated on the graphs) that significantly differ between healthy and CRC late stage samples (healthy = 33, CRC = 29). P values were calculated by Welch's t-test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **e,** Global DNA methylation levels, measured as in Fig. 2h, in the same cohort as (d). **f,** Histone PTMs, ratios and combinations (as indicated on the graphs) that significantly differ between healthy and early stage CRC patients (healthy = 33, early CRC = 8). P values were calculated by Wilcoxon rank sum exact test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **g,** Individual parameters predictive power score (PPS) analysis for the various subgroups (see Methods). Color scale represents PPS value. **h,** Principal Component Analysis (PCA) with input parameters of H3K27me3/Nuc, CEA/MST1 and CEA. Sample groups are color-coded as indicated, each dot represents a plasma sample.











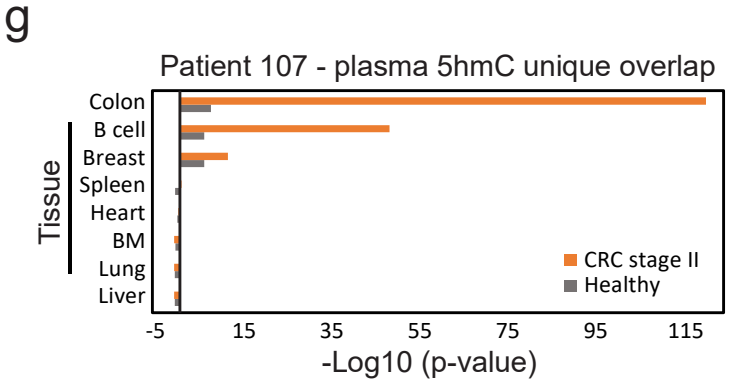
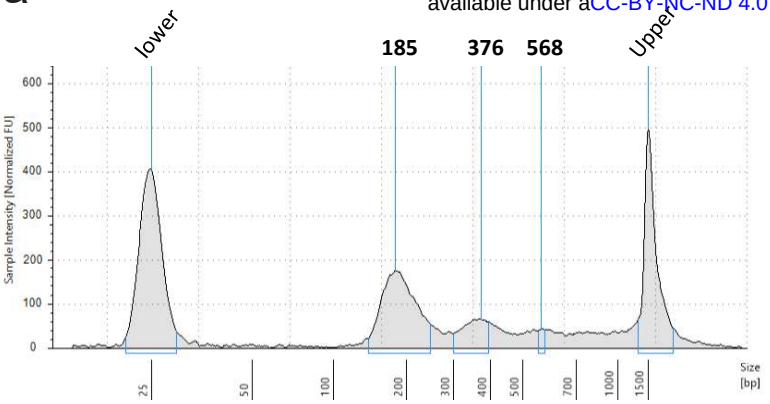


Fig.4: EPINUC differentiates healthy versus CRC patients and informs the tumor tissue-of-origin.

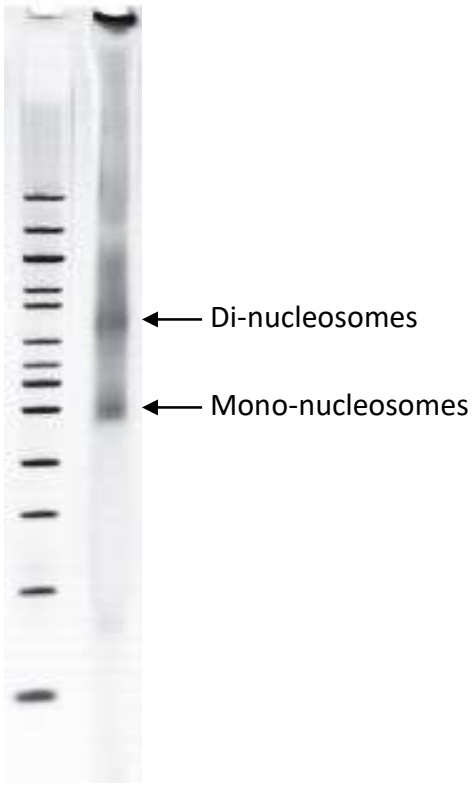
a, ROC curves discriminate between healthy (n=33) and all CRC samples (n=37) using a logistic regression model. The area under the curve (AUC) is shown for protein biomarkers only (Proteins; CEA + CEA/MST1, black line), protein biomarkers in combination with global DNA methylation levels (Proteins + Methylation, blue line) and the complete dataset generated by EPINUC (Proteins + Methylation + PTMs, Red line). Gray diagonal line indicates expected curve for random classification. **b,** Experimental scheme for EPINUC-seq. Histone PTMs are decoded as shown in Fig. 1A. Next, histones are evicted by increasing salt concentration, retaining DNA strands at identical positions. Single-molecule DNA sequencing-by-synthesis is performed by cycles of incorporation of fluorescently labeled nucleotides and TIRF imaging[21]. Images represent four sequencing cycles, showing incorporation of cytosine (C), thymine (T), adenosine (A) and Guanine (G). For each x,y coordinate on the surface, sequence data is analyzed and integrated with the initial images that registered histone PTMs, revealing the modification state of the corresponding nucleosome. **c,** Single-molecule reads of H3K27me3 (blue) from CRC patient overlap with ChIP–seq profile of H3K27me3 in the colon, but not the other indicated tissues. BM=bone marrow. **d,** Tissues and primary cell lines ranked by overlap significance with single-molecule plasma H3K4me3 (top) or H3K27me3 (bottom) positive reads. Black line corresponds to P value of 0.05. P values were determined by Z-test. **e,** Overlap significance of H3K4me3 single-molecule reads with ChIP-seq data[13] performed in the plasma of healthy and CRC patients. **f,g,** Overlap significance of tissues and primary cell lines unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus late stage CRC (f) or early stage CRC (g), calculated as in (e). Black line corresponds to P value of 0.05. P values were determined by Z-test.

a

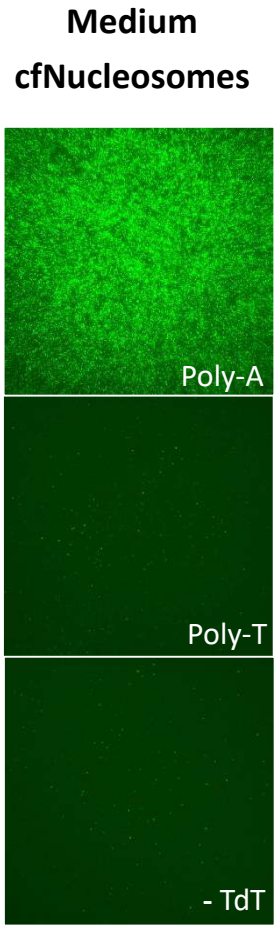


b

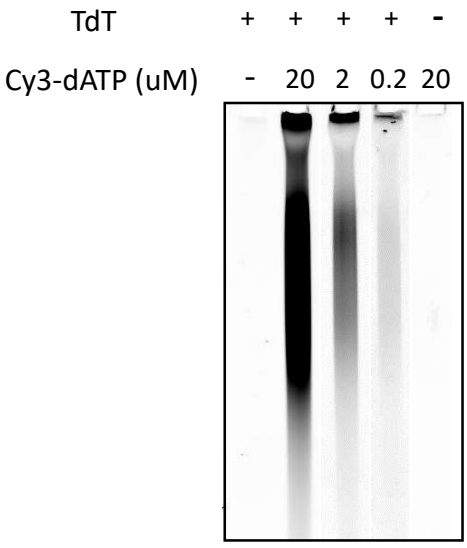
c



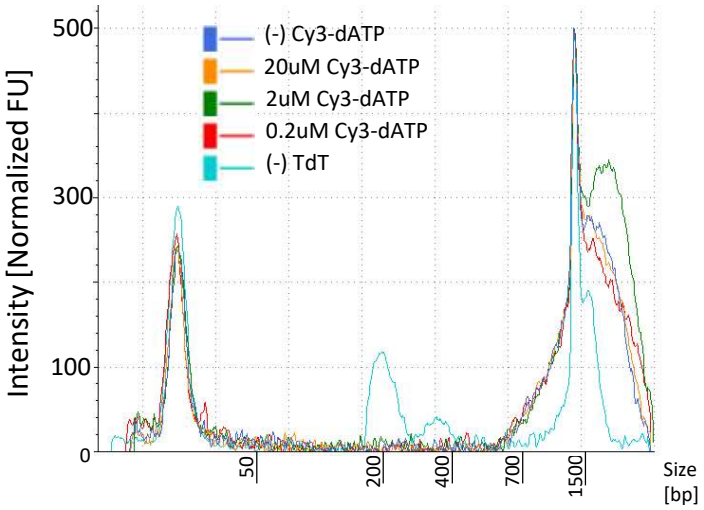
f



d



e

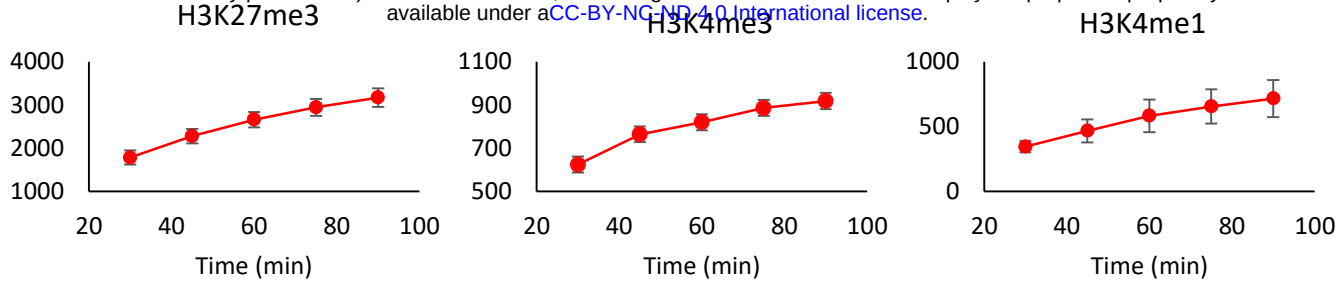


Supplementary Figure 1

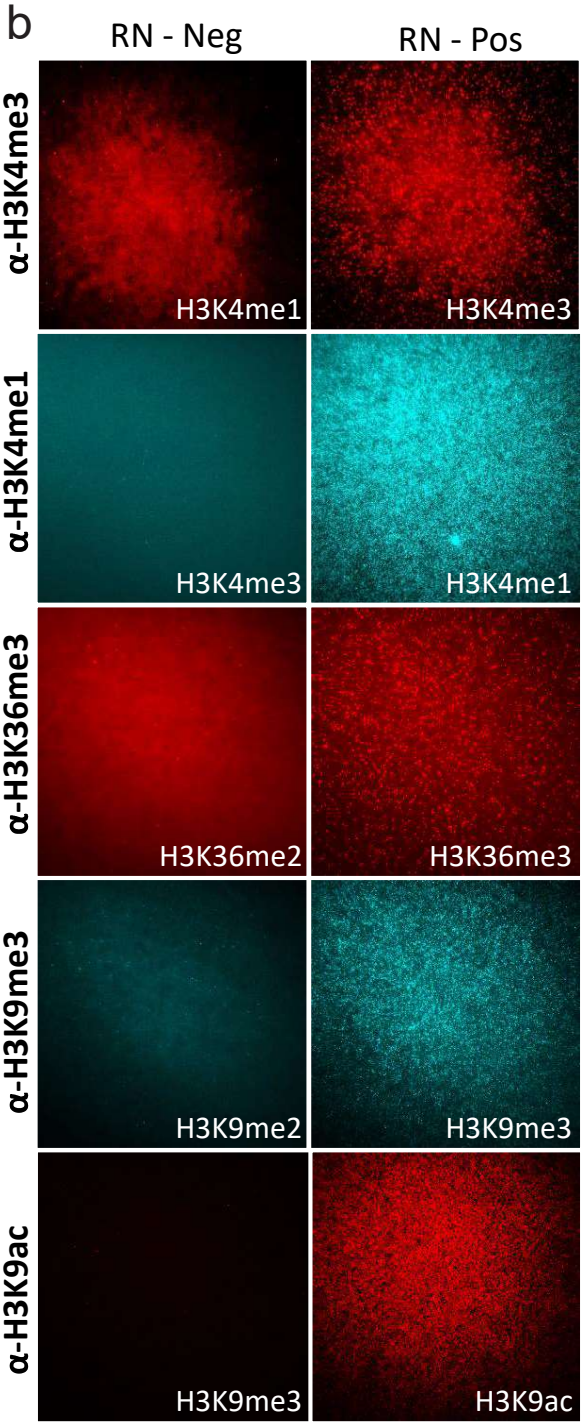
In-vitro system for sample preparation of cfNucleosomes.

(a) TapeStation fragment size analysis of DNA isolated from medium of HEK293 cells treated with the apoptosis-inducing factor Staurosporine (STS). DNA exhibits canonical apoptotic DNA fragmentation pattern. **(b)** Nucleosomes from medium of HEK293 cells (described in a) treated with STS (+) or with PBS (-) were resolved on a 6% TBE gel and visualized by Typhoon laser scanner. **(c)** Nucleosomes were extracted from HEK293 cells by digestion with MNase[21], resolved on a 6% TBE gel and visualized by Typhoon laser scanner. Bands correspond to mono-nucleosomes and di-nucleosomes. **(d)** Nucleosomes from medium of HEK293 cells (described in a) were subjected to tagging and tailing reaction (see Methods), resolved on a 6% TBE gel and visualized by Typhoon laser scanner. Image depicts fluorescent labeling via incorporation of Cy3-dATP during polyadenylation reaction. Incorporation is dependent on the presence of TdT in the reaction, and level of florescent signal correlates with the concentration of Cy3-dATP (Left). **(e)** TapeStation fragment size analysis of the samples in (d). **(f)** Poly-dA tailed and Cy3-dATP labeled cfNucleosomes show specific anchoring to PEGylated poly-dT surfaces through A:T hybridization.

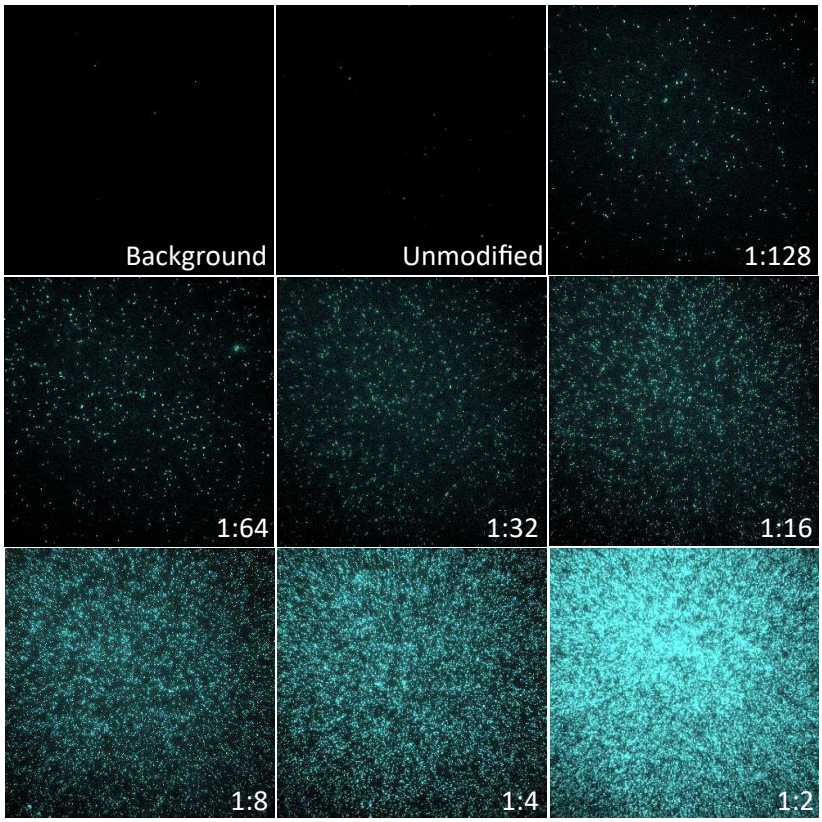
a



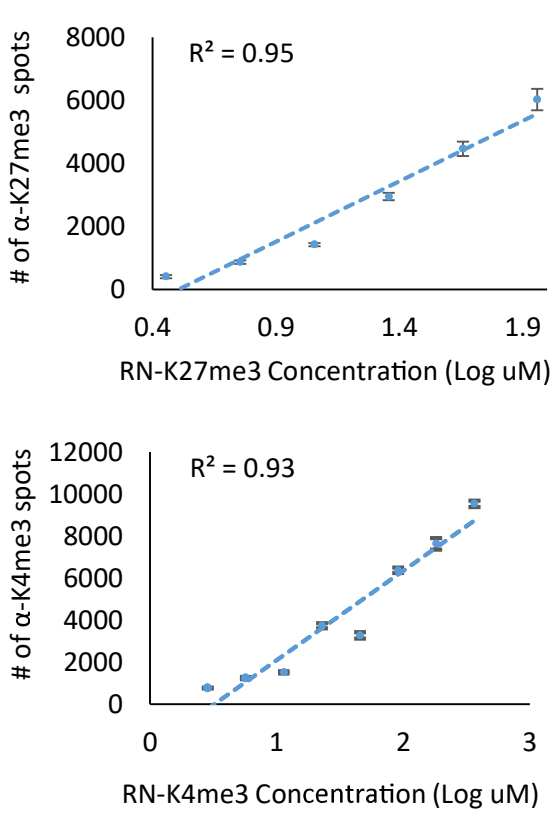
b



c



d

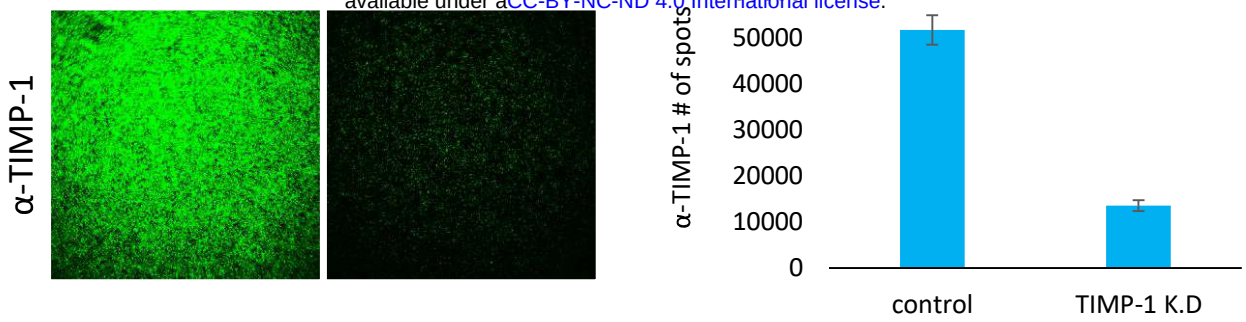


Supplementary Figure 2

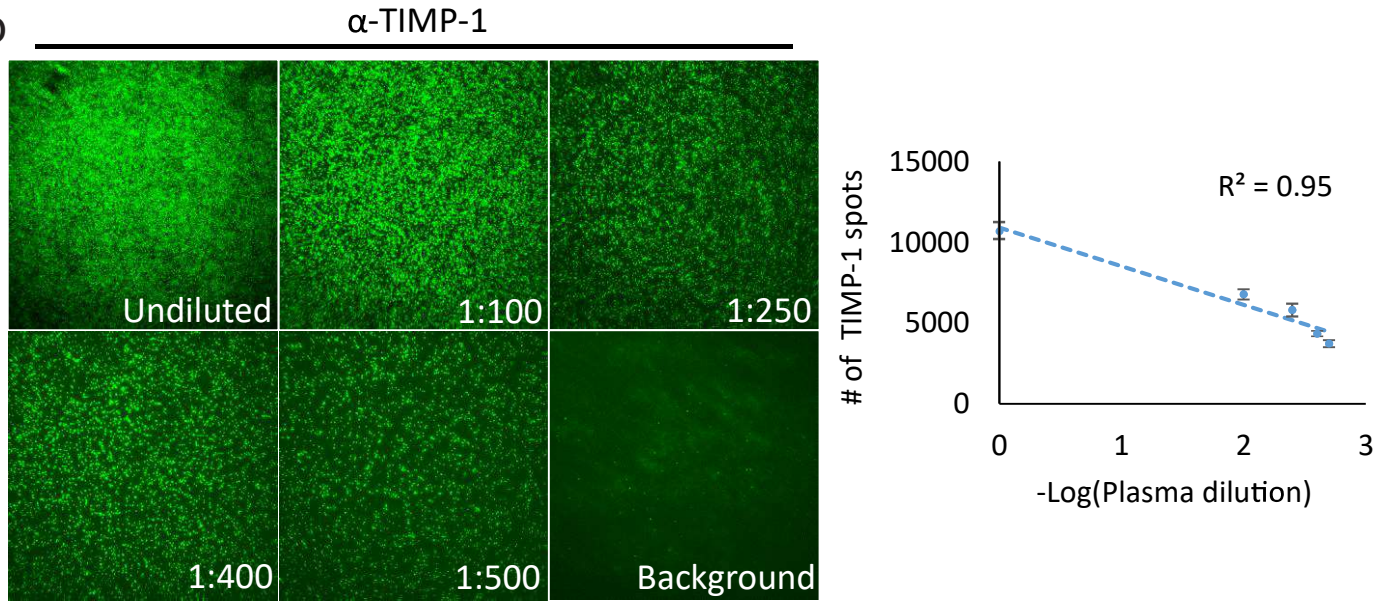
Single-molecule measurements of antibodies' specificity and dynamics.

(a) Accumulation of unique antibody binding events over time. Data is presented as the mean +/- standard deviation (s.d.) of 50 FOVs for each time point. **(b)** Representative TIRF images demonstrate antibodies binding to various recombinant nucleosomes (RN) carrying different histone PTMs. Left images: negative control, showing very low un-specific binding of fluorescent antibodies to recombinant nucleosomes that do not carry the target PTM. Right: binding of antibodies to recombinant nucleosomes that carry the target modification. The type of modification is indicated at the bottom of the images. **(c)** Representative TIRF images depicting binding of the antibody targeting H3K27me3 to an empty surface (Background), unmodified recombinant nucleosomes (Unmodified), and H3K27me3-modified recombinant nucleosomes diluted as indicated. **(d)** Regression analysis of number of spots per FOV as a function of target recombinant nucleosomes' concentration demonstrates low variability along the regression line, supporting linearity of binding. 1:2 dilution was excluded from analysis due to high density. Data is presented as the mean +/- s.d. of 50 FOVs for each concentration.

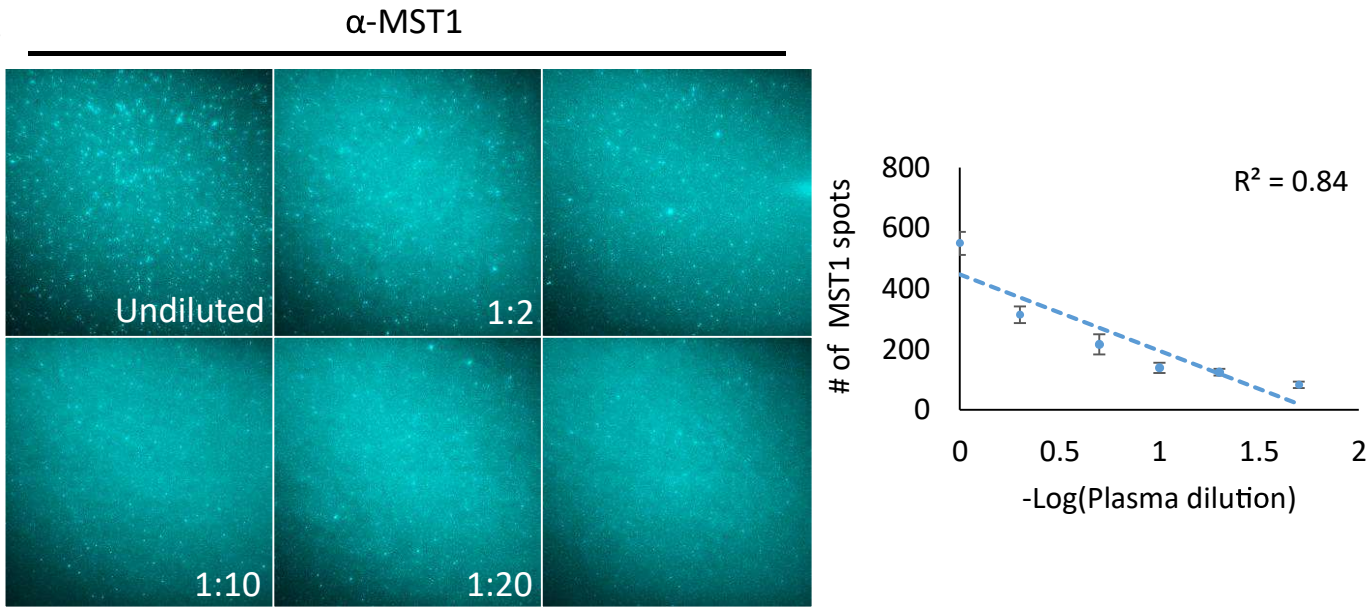
a



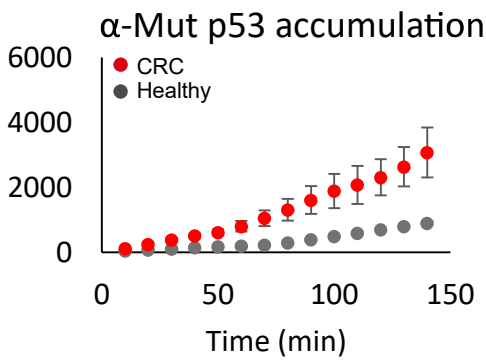
b



c



d



752 **Supplementary Figure 3**

753 **Single-molecule imaging of MST1, TIMP-1 and mutant p53.**

754 **(a)** Representative TIRF images (Left) and quantification (Right) of TIMP-1 protein levels in
755 SW480 medium, following TIMP-1 knockdown versus control cells. Data is presented as the mean
756 +/- s.d. of 50 FOVs for each treatment. **(b, c)** Representative TIRF images and standard curves of
757 antibodies targeting MST1 and TIMP-1 on serial plasma dilutions, depicting linear detection of
758 molecules within this concentration range. Data is presented as the mean +/- s.d. of 50 FOVs for
759 each concentration. **(d)** Signal accumulation of mutant p53 over time for late stage CRC and
760 healthy plasma samples. Each time point is presented as the mean +/- s.d. of 50 FOVs.

761

762

763

764

765

766

767

768

769

770

771

772

773

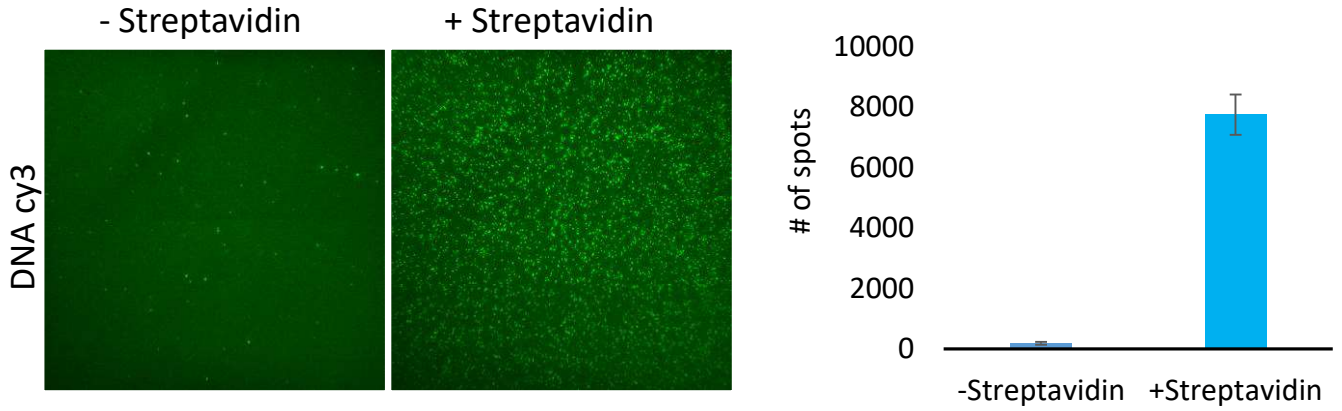
774

775

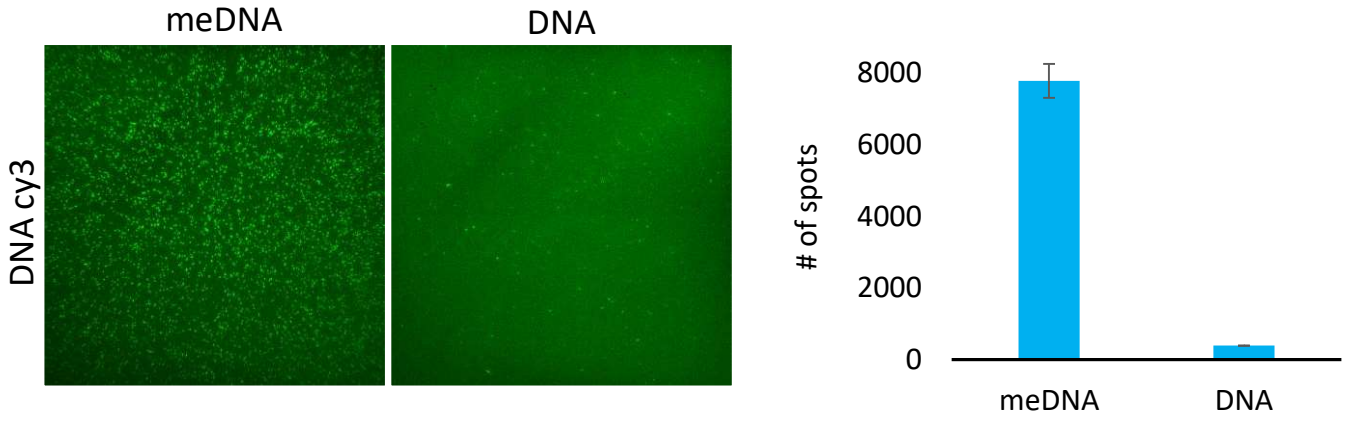
776

777

a



b

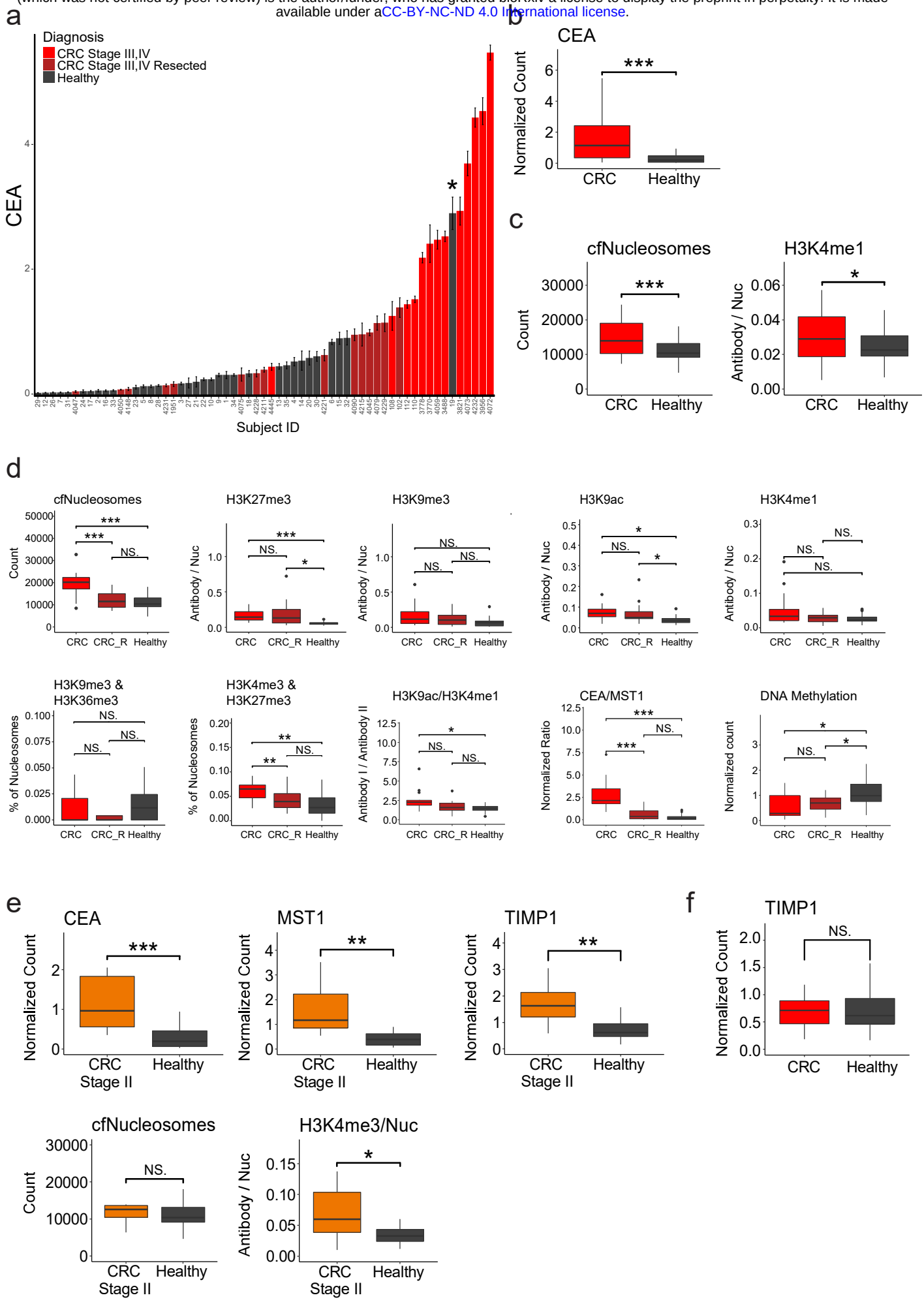


778 **Supplementary Figure 4**

779 **Single-molecule detection of DNA methylation levels.**

780 **(a)** Representative TIRF images (Left) and quantification (Right) indicating specific anchoring of
781 Biotin-MBD-meDNA complexes to streptavidin-coated surfaces. Data is presented as the mean
782 +/- s.d. of 50 FOV for each treatment. **(b)** Biotin-MBD specifically binds methylated DNA.
783 Representative TIRF images (Left) and quantification (Right) of global DNA methylation levels
784 following incubation of MBD2-biotin with Cy3-labeled (green) methylated (meDNA) or un-
785 methylated synthetic DNA fragments. Data is presented as the mean +/- s.d. of 50 FOV for each
786 treatment.

787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807

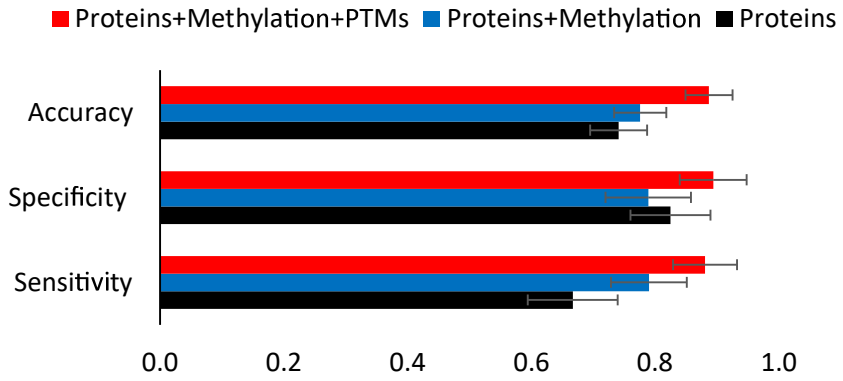


Supplementary Figure 5

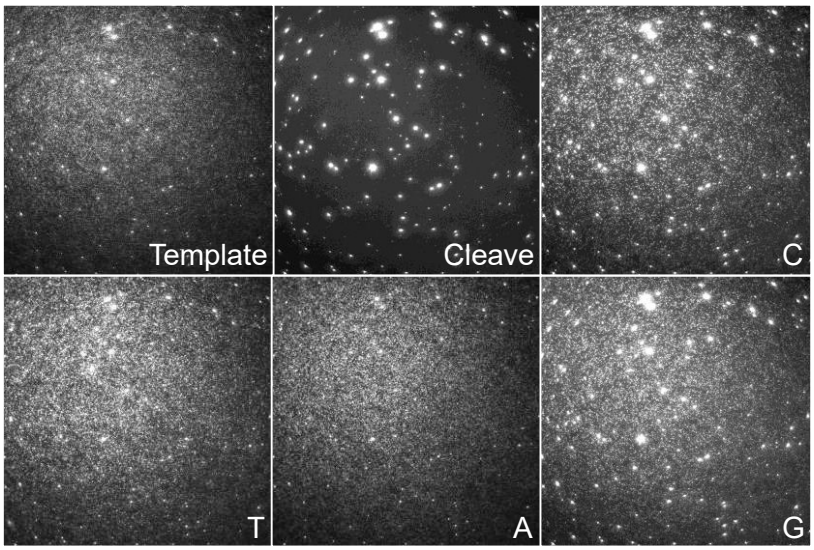
Analysis of histone PTMs, protein biomarkers and DNA methylation in the cohort of plasma samples from healthy and CRC subjects.

(a) CEA normalized levels in the plasma of CRC patients and healthy individuals. Each bar represents a subject, data is presented as the mean +/- s.d. of 50 FOVs per sample. Sample 19 (Healthy) denoted by *. **(b)** Box plot representation of the data in A (healthy = 33, CRC = 29). Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than 1.5× interquartile range from the hinge. P values were calculated by Welch's t-test. *** P value < 0.001. **(c)** Global level of cfNucleosomes, and levels of H3K4me1-modified nucleosomes, significantly differ between healthy and CRC late stage samples (healthy = 33, CRC = 29). P values were calculated by Welch's t-test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(d)** Multiple comparison of all significant parameters between CRC, CRC resected (CRC_R) and healthy samples (CRC = 13, CRC_R=16, healthy = 33), corresponding to the data presented in main Fig. 3c,d,e. Of note, while all of these parameters differ between healthy versus the combined cohort of all CRC patients, this figure shows the differences between CRC patients with/without tumor resection. In some parameters, resected patients show higher similarity to healthy, and in other parameters, they are similar to CRC patients prior to tumor resection. See methods for P value calculation. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(e)** Levels of CEA, MST1 and TIMP1, as well as H3K4me3-modified nucleosomes, significantly differ between healthy and early stage CRC patients (healthy = 33, early CRC = 8). Total levels of cfNucleosomes do not show significant difference between the groups, likely due to low tumor burden at early stage patients. P values were calculated by Wilcoxon rank sum exact test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(f)** TIMP1 levels do not significantly differ in the cohort of healthy versus late-stage CRC patients.

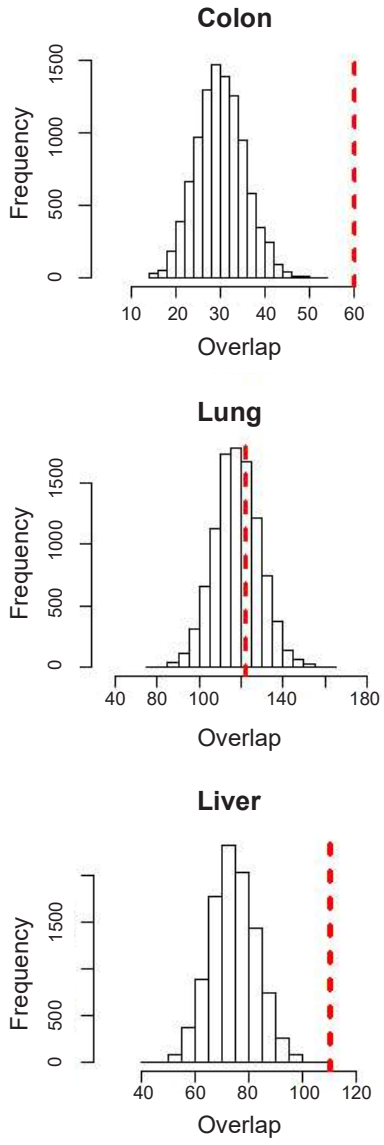
a



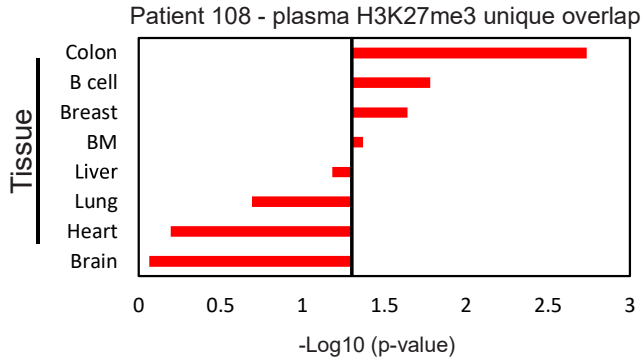
b



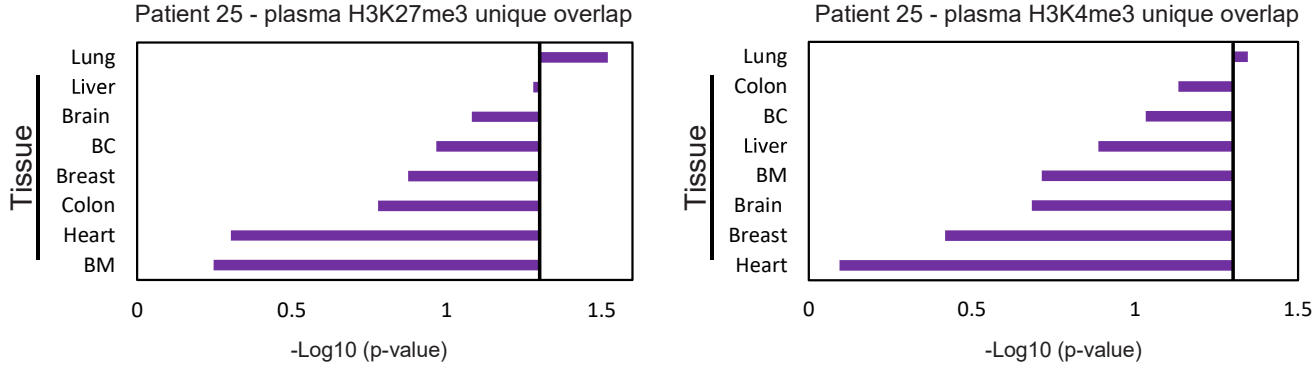
c



d



e

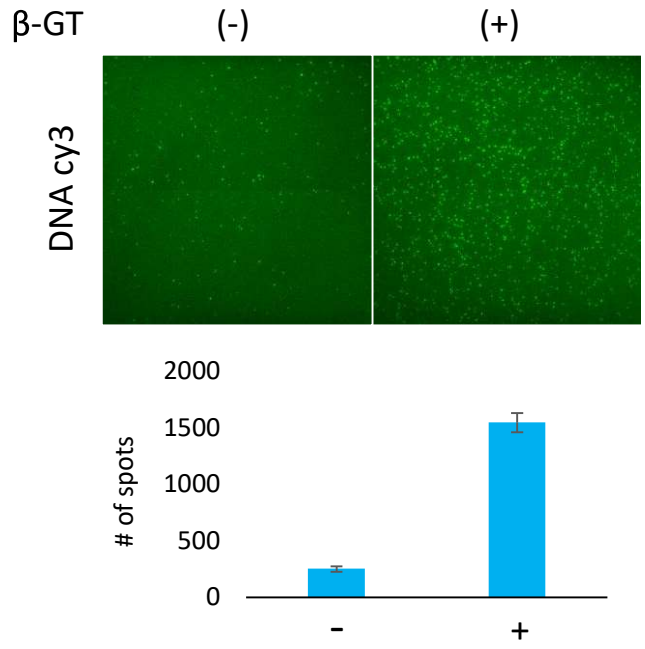


Supplementary Figure 6

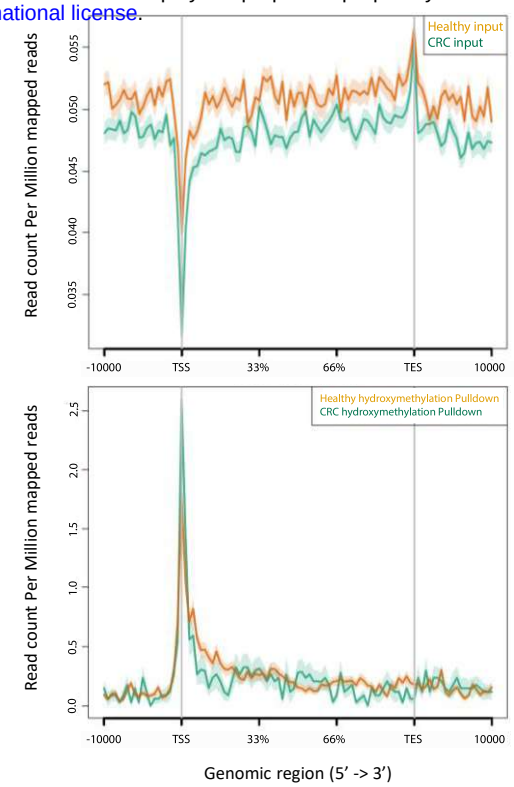
Prediction model performance and Tissue-of-origin analysis based on EPINUC-seq.

(a) Logistic model performance increases with integration of additional layers of information. Corresponds to Fig. 4a. **(b)** Representative TIRF images of antibodies' binding (Template) followed by cleavage of fluorophore (Cleave) and sequencing cycles (C,T,A,G). High diameter spots correspond to FluoSpheres used for image alignment. **(c)** Histograms portray colon, lung and liver frequency distributions of random reads (equivalent to the number of H3K27me3 plasma CRC positive reads) that overlap with tissue H3K27me3 unique peaks (bootstrapping simulation, 10K iterations, Methods). Red line corresponds to the number of H3K27me3 unique tissue peaks that overlap with CRC (patient 4044) H3K27me3 positive plasma reads. (**d**) EPINUC-seq analysis of an additional CRC patient, similar to the analysis shown in Fig. 4d. (**e**) EPINUC-seq analysis of plasma from stage IV lung cancer patient (patient 25). Tissues and primary cell lines ranked by overlap significance with single-molecule H3K27me3 (Left) or H3K4me3 (Right) positive reads, indicating lung as the tissue-of-origin. Black line corresponds to P value of 0.05. P values were determined by two tailed Z-test or Wilcoxon rank sum test.

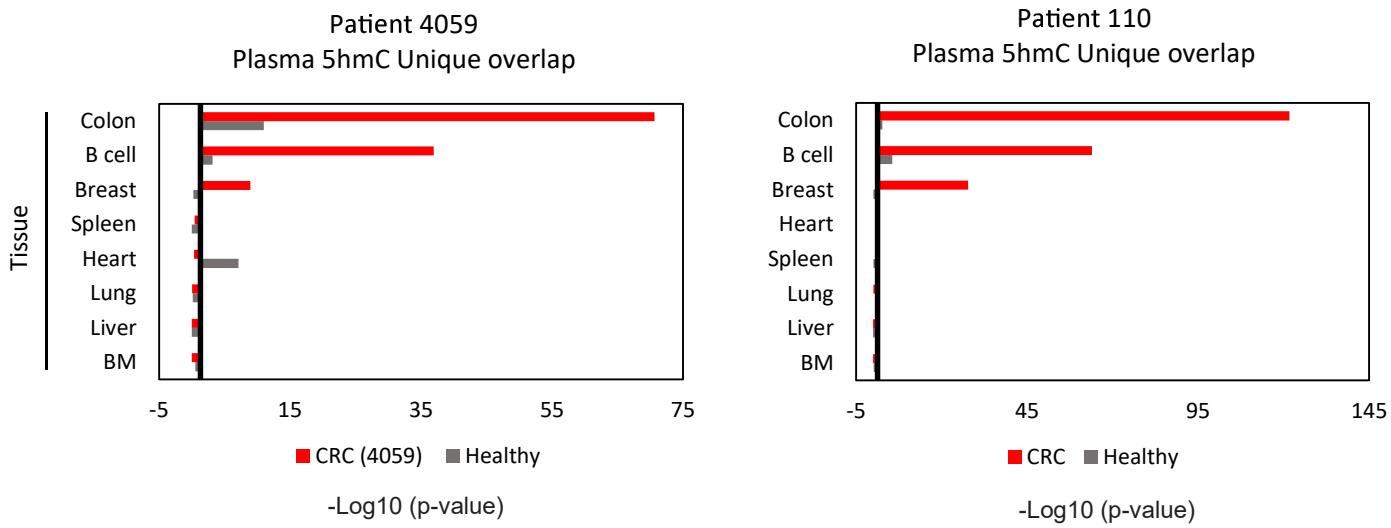
a



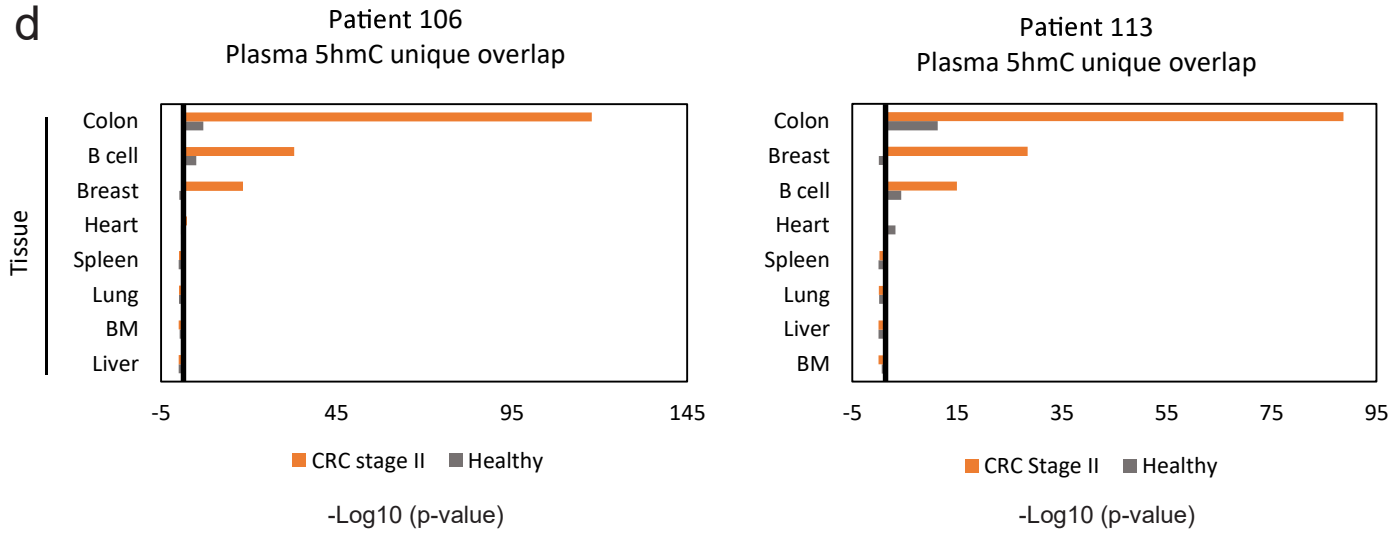
b



c



d



Supplementary Figure 7

Tissue-of-origin analysis based on single-molecule 5hmC sequencing.

(a) Representative TIRF images (Top) and quantification (Bottom) of fluorescently labeled nucleosomal DNA (Cy3, green) enriched for 5hmC, with (+) or without (-) the biotin conjugating enzyme beta-glucosyltransferase (β-GT). Data is presented as the mean +/- s.d. of 50 FOV for each treatment. **(n)** Metagene profiles of input (Top) and 5hmC enriched (Bottom) cfDNA sequenced from healthy (n=3, green) and CRC (n=3, Orange) samples, exhibiting 5hmC enrichment at gene bodies. **(c-d)** Overlap significance of tissues and primary cell lines unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus late stage CRC **(c)** or early stage CRC **(d)**, similar to the analysis shown in Fig. 4f,g. Each panel represents a different patient. Black line corresponds to P value of 0.05. P values were determined by Z-test.

References

1. Wan, J. C. M. *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17,** 223–238 (2017).
2. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **17,** 100087 (2019).
3. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20,** 71–88 (2019).
4. Lo, Y. M. D., Han, D. S. C., Jiang, P. & Chiu, R. W. K. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science (80-.).* **372,** (2021).
5. Xu, R. H. *et al.* Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16,** 1155–1162 (2017).
6. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9,** 5068 (2018).
7. Kang, S. *et al.* CancerLocator: Non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* **18,** 1–12 (2017).
8. Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563,** 579–583 (2018).
9. Reinberg, D. & Vales, L. D. Chromatin domains rich in inheritance only certain histone posttranslational modifications qualify as being epigenetic. *Science (80-.).* **361,** 33–34 (2018).
10. Shema, E., Bernstein, B. E. & Buenrostro, J. D. Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.* **51,** 19–25 (2019).
11. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17,** 487–500 (2016).
12. Mancarella, D. & Plass, C. Epigenetic signatures in cancer: proper controls, current challenges and the potential for clinical translation. *Genome Med. 2021 131* **13,** 1–12 (2021).
13. Sadeh, R. *et al.* ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nat. Biotechnol.* **39,** 586–598 (2021).
14. Gezer, U. *et al.* Histone methylation marks on circulating nucleosomes as novel blood-based biomarker in colorectal cancer. *Int. J. Mol. Sci.* **16,** 29654–29662 (2015).
15. Van den Ackerveken, P. *et al.* A novel proteomics approach to epigenetic profiling of circulating nucleosomes. *Sci. Rep.* **11,** 1–12 (2021).
16. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an in Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164,** 57–68 (2016).
17. Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48,** 1273–1278 (2016).

18. Sun, K. *et al.* Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Res.* **29,** 418–427 (2019).

19. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136,** E359–E386 (2015).

20. Hu, Z. *et al.* Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51,** 1113–1122 (2019).

21. Shema, E. *et al.* Single-molecule decoding of combinatorially modified nucleosomes. *Science (80-.).* **352,** 717–721 (2016).

22. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).

23. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129,** 823–837 (2007).

24. Tiernan, J. P. *et al.* Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting. *Br. J. Cancer* **108,** 662–667 (2013).

25. Meng, C. *et al.* TIMP-1 is a novel serum biomarker for the diagnosis of colorectal cancer: A meta-analysis. *PLoS One* **13,** e0207039 (2018).

26. Yu, J. *et al.* Identification of MST1 as a potential early detection biomarker for colorectal cancer through a proteomic approach. *Sci. Rep.* **7,** (2017).

27. Mandal, S. *et al.* Direct Kinetic Fingerprinting for High-Accuracy Single-Molecule Counting of Diverse Disease Biomarkers. *Acc. Chem. Res.* (2020). doi:10.1021/acs.accounts.0c00621

28. Furth, N. *et al.* Unified platform for genetic and serological detection of COVID-19 with single-molecule technology. *PLoS One* **16,** e0255096 (2021).

29. Nakayama, M. & Oshima, M. Mutant p53 in colon cancer. *J. Mol. Cell Biol.* **11,** 267–276 (2019).

30. Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F. & Goel, A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat. Rev. Gastroenterol. Hepatol.* **17,** 111–130 (2020).

31. Dawson, M. A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science (80-.).* **355,** 1147–1152 (2017).

32. Wood, K. H. & Zhou, Z. Emerging molecular and biological functions of MBD2, a reader of DNA methylation. *Front. Genet.* **7,** (2016).

33. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6,** 224ra24-224ra24 (2014).

34. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* **14,** 747–753 (2014).

35. Kerachian, M. A. *et al.* Crosstalk between DNA methylation and gene expression in colorectal cancer, a potential plasma biomarker for tracing this tumor. *Sci. Rep.* **10,** 1–13 (2020).

36. King, W. D. *et al.* A cross-sectional study of global DNA methylation and risk of colorectal adenoma. *BMC Cancer* **14,** 1–9 (2014).

37. Frederiksen, C. *et al.* Plasma TIMP-1 levels and treatment outcome in patients treated with XELOX for metastatic colorectal cancer. *Ann. Oncol.* **22,** 369–375 (2011).

38. Li, W. *et al.* 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* **27,** 1243–1257 (2017).

39. Lio, C. W. J., Yuita, H. & Rao, A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood* **134,** 1487–1497 (2019).

40. Zhang, L. *et al.* Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. *Nat. Commun.* **4,** (2013).

41. Song, C. X. *et al.* 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* **27,** 1231–1242 (2017).

42. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34,** 547–555 (2016).

43. Chandradoss, S. D. *et al.* Surface passivation for single-molecule protein studies. *J. Vis. Exp.* (2014). doi:10.3791/50549

44. Fleischhacker, M. & Schmidt, B. Circulating nucleic acids (CNAs) and cancer-A survey. *Biochim. Biophys. Acta - Rev. Cancer* **1775,** 181–232 (2007).

45. Harris, T. D. *et al.* Single-Molecule DNA Sequencing of a Viral Genome. *Science (80-.).* **320,** 106–109 (2008).