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).

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

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

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

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

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

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

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

122 We applied EPINUC to generate high-dimensional data, comprising of the three layers of
123 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

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

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

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

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

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

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

204 As proof-of-concept, we applied EPINUC followed by sequencing (EPINUC-seq) to two plasma
205 samples of late stage CRC probed for H3K4me3 and H3K27me3 (Fig. 4c,d,e, Supplementary Fig.
206 6c,d, Methods). Single-molecule mapped reads, corresponding to modified nucleosomes, were
207 intersected with unique antibody peak signatures generated from ENCODE ChIP-seq data for
208 various tissues and primary cell lines, followed by bootstrapping simulations to calculate
209 significance. Reinforcing our hypothesis, we found that both plasma samples showed significant
210 overlap with colon-specific H3K4me3 and H3K27me3 peaks, indicating colon as the main tissue-
211 of-origin (Fig. 4d and Supplementary Fig. 6c,d). Validating our approach, a similar analysis of a
212 plasma sample from lung cancer patient revealed lung tissue as the main origin (Supplementary
213 Fig. 6e). Moreover, comparing our data to a recent ChIP-seq study of H3K4me3 in plasma showed
214 significant overlap with profiles obtained from CRC patients[13], but not with healthy plasma (Fig.
215 4e). H3K27me3 mapped reads showed a broader pattern, overlapping with peaks corresponding
216 to hematopoietic lineage as well as colon (Fig. 4d and Supplementary Fig. 6d). Interestingly, we

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

219 Finally, we combined a complementary single-molecule approach to identify the tumor tissue-of-
220 origin, by single-molecule profiling of the DNA modification 5-Hydroxymethylcytosine (5hmC).
221 5hmC is known to play important roles in gene regulation and cancer[38,39]. We captured 5hmC-
222 enriched DNA from plasma as previously described[40], followed by single-molecule DNA
223 sequencing (Supplementary Fig. 7a, Methods). In agreement with previous reports[41], we found
224 5hmC in cfDNA to be enriched at gene bodies and promoter proximal regions (Supplementary
225 Fig. 7b), which are also known to be marked with H3K36me3. Thus, we generated unique 5hmC
226 read signatures for healthy and CRC samples, and examined their overlap with H3K36me3 peak
227 signatures from various tissues and primary cell lines (Methods). Similar to sequencing of the
228 histone marks (Fig. 4d), analysis of 5hmC in the plasma of the same patient showed highly
229 significant overlap with the colon-specific profile, validating this strategy for identification of the
230 tumor tissue-of-origin (Fig. 4f and Supplementary Fig. 7c). Finally, we showed correct
231 identification of colon origin also for early stage CRC patients (Fig. 4g and Supplementary Fig.
232 7d).

233 Our work establishes EPINUC as a novel liquid biopsy approach that analyzes multiple histone
234 and DNA modifications, as well as protein biomarkers, at single-molecule precision. EPINUC
235 distinguishes between CRC patients to healthy individuals at high specificity and sensitivity. We
236 showed that this multi-parametric approach is suitable also for detection of early stage patients,
237 although expanding the analysis to a larger cohort is needed. The main challenges with analyzing
238 plasma nucleosomes are (1) their minute amount- in 1 ml of plasma there are ~1000 genome
239 copies[13,42]; (2) The plasma is highly dense with additional proteins, rendering enzymatic or binding
240 approaches to capture nucleosomes difficult; (3) There is high variability between different
241 individuals, stressing the need for quantitative methodologies to allow comparison between
242 samples; and (4) Multi-parametric data is needed to achieve high specificity and confidence in
243 detection. Our EPINUC approach addresses these challenges by enabling single-molecule
244 combinatorial detection of epigenetic marks, DNA sequencing and protein biomarkers from
245 limited input material. In addition to the unique epigenetic analysis, the single-molecule system
246 outperforms the classical ELISA assay for measuring protein biomarkers. ELISA is of relatively
247 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.

310 **Methods**

311

312 Patients

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

315 Plasma collection

316 Blood samples were collected in VACUETTE K3 EDTA tubes and transferred immediately to ice.
317 Next, blood was centrifuged (10 minutes, 1,500g, 4°C) and the supernatant was transferred to a
318 fresh 50-ml tubes and centrifuged again (10 minutes, 3,000g, 4°C). The supernatant was collected
319 and used as plasma for all experiments. Plasma was analyzed fresh or flash-frozen and stored at
320 −80°C for future analysis.

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

322 Tagging and tailing of cfNucleosomes was carried out as following: 20 µl of plasma or 5X (DDW
323 diluted) concentrated apoptotic medium was incubated at 37°C for 1 hour with the following
324 reaction mixture: 10 µl 10X Green Buffer (Enzymatics, B0120), 416 µM CoCl2 (Enzymatics,
325 B0220), 1:60 PI (SIGMA, P8340), 83.3 nM fluorescently labeled dATP (Jena Bioscience, NU-
326 1611-Cy3/Cy5), 83.3 µM dATP (Thermo Fisher Scientific, R0181), 5 µl of Klenow Fragment
327 (3'→5' exo-, NEB, M0212S) 3µl of T4 Polynucleotide Kinase (NEB, M0201L) and 4 µl of
328 Terminal deoxynucleotidyl transferase (TdT, Enzymatics, P7070L). Following incubation,
329 samples were inactivated by immediate transfer to ice. For nucleosome sequencing, 1.67 µl of
330 ddATP was added (SIGMA, GE27-2051-01).

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

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

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

344 Cell culture and apoptosis

345 Cell lines were maintained at 37˚C with 5% CO2. HEK-293 cells were cultured in 150 cm plates
346 (10×10^6 cells in 20 ml of media) in DMEM supplemented with 10% FBS and 1% P/S, and
347 passaged every week. For induction of apoptosis, 6 µM of Staurosporine (STS, Holland-Moran,
348 62996-74-1.25) was added to medium of confluent cells. 72 hours later, medium was collected and
349 immediately processed. To verify fragment sizes along with nucleosome labeling, 10 ul of the
350 nucleosomes and 10ul of AMPure extracted DNA (either directly from concentrated medium or
351 after the tagging and tailing reaction), were loaded on High Sensitivity D1000 ScreenTapes
352 (Agilent, 5067-5584) and 6% TBE gel (ThermoFisher Scientific, EC62655BOX), and imaged with
353 4200 TapeStation (Agilent) or Typhoon imager (Amersham Biosciences), respectively. Apoptotic
354 medium cfNucleosomes were concentrated and recovered using Centricon Plus-70 centrifugation
355 filteres (Merck, UFC710008) according to the manufacture protocol. PI was supplemented 1:100
356 following concentration.

357 Surface preparation for single-molecule imaging

358 PEGylated-Biotin and PEGylated-poly T coated microscope slides were prepared based on the
359 protocol described by Chandradoss et al[43]. Ibidi glass coverslips (25 mm x 75 mm, IBIDI, IBD-
360 10812) were cleaned with (1) MilliQ H2O (3X washes, 5 minutes sonication, 3X washes); (2) 2%
361 Alconox (SIGMA, 242985) (20 min sonication followed by 5X washes with MilliQ H2O); and (3)
362 100% Acetone (20 min sonication followed by 3X washes with MilliQ H2O). Slides were further
363 cleaned and functionalized (Hydroxylated) by incubation in 1 M KOH (SIGMA, 484016) solution
364 for 20 minutes while sonicated, followed by 3X washes with MilliQ H2O. Slides were sonicated
365 twice for 10 minutes in 100% HPLC ethanol (J.T baker 8462-25) prior to applying amino-
366 silanization chemistry. Next, slides were incubated for 24 minutes in a mixture of 3% 3-
367 Aminopropyltriethoxysilane (ACROS Organics, 430941000) and 5% acetic acid in HPLC EtOH,
368 with 1 minute sonication in the middle. Slides were then washed with HPLC EtOH (3X) and
369 MilliQ H2O (3X) and dried with N2. Surface functionalization along with first passivation step
370 was performed by applying mPEG: PEGylated-Biotin/PEG-Azide solution [20 mg PEGylated-
371 Biotin (Laysan, Biotin-PEG-SVA-5000), 180 mg mPEG (Laysan, MPEG-SVA-5000) or 20 mg
372 PEG-Azide (JenKem, A5088), 180 mg mPEG (Laysan, MPEG-SVA-5000)] dissolved in 1560 ul
373 0.1 M Sodium Bicarbonate (SIGMA, S6297) and degassed (centrifugation at 1 minute at 16,000*g*).

374 Next, 140 µl of solution was applied on one surface, followed by immediate assembly of another
375 surface on top. Each pair of assembled surfaces were incubated overnight in a dark humid
376 environment.
377 For PEGylated-Biotin surfaces: At the next day, surfaces were washed with MilliQ H2O and dried
378 with N2 followed by a second passivation step. MS(PEG)4 (ThermoFisher Scientific, TS-22341)
379 was diluted in 0.1 M of sodium bicarbonate to a final concentration of 11.7 mg/ml and applied on
380 one surface, followed by the assembly of another surface on top. Each pair of assembled surfaces
381 were incubated overnight in dark humid environment. The next day, surfaces were disassembled,
382 washed with MilliQ H2O and dried with nitrogen. After nitrogen flush, surfaces were stored in -
383 20°C.
384 For PEGylated-poly T surfaces, following PEG-Azide coating, surfaces were washed with MilliQ
385 H2O and dried with N2. To enable anchoring of dT50 to surface via click chemistry, 10 µM of
386 5'heyxynyl-dT50 (IDT) were mixed with 2 mM of CuSO4 (SIGMA, C1297) and DDW. Next, 100
387 µl of the mixture was applied on one surface, followed by immediate assembly of another surface
388 on top. Each pair of assembled surfaces was incubated overnight in a dark humid environment. In
389 the next day, a second passivation step [MS(PEG)4] was carried out, similarly to PEGylated-Biotin
390 preparation. Surfaces were stored in -20°C post nitrogen flush in a similar fashion.
391 Antibody labeling
392 Capture and detection antibodies were labeled using Biotin conjugation kit (Abcam, ab201796)
393 and Alexa flour antibody labeling kits (Thermo Fisher Scientific, A20181/ A10237/A20186)
394 according to the manufacture protocol. Labeled antibodies were purified by size exclusion
395 chromatography using Bio-Spin 6 columns (Bio-Rad, 7326200) followed by measurement of
396 protein concertation using *Nanodrop* 2000 at 260 nm.
397 TIMP-1 imaging and siRNA transfections
398 siRNA transfection was performed using INTERFERin (Polyplus, 409-10) according to the
399 manufacturer's protocol. Briefly, cells were plated in 6-well plates (1.5×10^5 in 2.5 ml per well)
400 overnight, and the 200 µl of transfection complex was added directly to medium, at final
401 concentration of 25 nM of siRNA. RNA and protein samples were isolated from cells 72 hours
402 after transfection. The following siRNA was used: SMARTpool: ON-TARGETplus Human
403 TIMP1 siRNA (L-011792-00-0005, Dharmacon). For single-molecule imaging, medium was
404 collected from plates, followed by centrifugation at max speed in 4°C and collection of supernatant
405 to separate proteins from cell debris. Protein concentration was determined by Pierce™ BCA

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

408 Synthetic DNA preparation for DNA methylation assay

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

415 Sequence:

416 CATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAA
417 TCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACA
418 TACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
419 ACA

420 5-Aza-2-Deoxycytidine Treatment

421 HEK-293 cells were plated in 150 cm plates (10×10^6 cells in 20 ml of media) overnight, then
422 treated with 1 uM of 5-Aza-2-deoxycytidine (5 –Aza, SIGMA, A3656) or PBS for 4 days. Next,
423 5×10^6 cells were collected and washed with PBS supplemented with PI (1:100), followed by
424 centrifugation at 3000 rpm for 3 minutes. Cell pellet was resuspended with 1 ml of 0.05% Igepal
425 (SIGMA, I8896) diluted in PBS (supplemented with PI as mentioned above) and centrifuged again
426 at 3000 rpm for 3 minutes. Next, the pellet was resuspended in Lysis buffer [100 mM Tris-HCl
427 pH 7.5 (Gibco, 115567-027), 300 mM NaCl (J.T Baker, 7647145), 2% Triton® X-100 (SIGMA,
428 9002 93-1), 0.2% sodium deoxycholate (SIGMA, D6750), 10 mM CaCl2 (SIGMA, 21115)]
429 supplemented with PI and Micrococcal Nuclease (ThermoFisher Scientific, 88216). The reaction
430 mixture was incubated at 37°C for 10 minutes and then inactivated by addition of EGTA at a final
431 concentration of 20mM. Then, lysate was centrifuged for 10 minutes at max speed and supernatant
432 was transferred to a new tube. DNA extraction, fluorescent labeling and quantification was
433 performed as described for plasma cfDNA.

434 Single-molecule imaging

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

437 imaging of six different samples or various combinations of antibodies on a single surface. For
438 PEGylated-Biotin flowcells, Streptavidin (SIGMA, S4762) was added to a final concentration of
439 0.2 mg/ml followed by 10 minutes incubation and washing with imaging buffer [IMB: 12 mM
440 HEPES pH 8 (Thermo Fisher Scientific, 15630056), 40 mM TRIS pH 7.5 (Gibco, 115567-027)
441 60 mM KCL (SIGMA, 60142), 0.32 mM EDTA (Invitrogen, 15575-038), 3 mM MgCl2 (SIGMA,
442 63069), 10% glycerol (Bio-Lab, 56815) , 0.1 mg/ml BSA (SIGMA, A7906) and 0.02% Igepal
443 (SIGMA, I8896)]. For time-lapse imaging experiments (Histone PTMs, p53), prior to sample
444 application, TetraSpeck beads (ThermoFisher Scientific, T7279) diluted in PBS were added and
445 incubated on surface for at least 10 minutes to allow correction for stage drift in image analysis.
446 Imaging was performed on a total internal reflection (TIRF) microscope, Nikon (Ti2 LU-N4
447 TIRF).

448 **Histone PTMs analysis**

449 PEGylated-poly T coated coverslips were assembled as described and further passivated with 5%
450 BSA (Merck, A7906) for 30 minutes followed by wash with IMB. Next, plasma sample containing
451 tailed and fluorescently labeled cfNucleosomes was incubated with antibodies (diluted 1:60) for
452 30 minutes at room temperature (RT), to allow formation of antibody-cfNucleosomes complexes.
453 Next, samples were loaded on the surface and incubated for 15 minutes to allow hybridization.
454 Flowcell was washed (X3) with IMB, followed by time lapse imaging every 15 minutes, with the
455 three laser channels, across all positions (50 Fields of View (FOVs, 148μm^2) per experiment).

456 **Protein biomarkers analysis**

457 PEGylated-Biotin coated coverslips were assembled and coated with streptavidin. Biotinylated
458 antibodies were incubated on surface in IMB2 [10 mM MES pH 6.5 (Boston Bioproducts Inc,
459 NC9904354), 60 mM KCL, 0.32 mM EDTA, 3 mM MgCl2, 10% glycerol, 0.1 mg/ml BSA and
460 0.02% Igepal] for 30 minutes, followed by wash with IMB2. Next, plasma sample was added to
461 flowcell and incubated on surface for 30 minutes, followed by washes (3X) with IMB2, to allow
462 binding of target proteins. Fluorescently labeled antibodies (detection antibodies) were introduced
463 to the surface for 60 minutes, washed with IMB2, and imaged.

464 **Global DNA methylation analysis**

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

469 **DNA Hydroxymethylation analysis**

470 cfDNA was incubated in 25 µl reaction mixture containing 50 mM HEPES buffer (pH 8), 25 mM
471 $MgCl_2$, 60 µM UDP-6-N_3-Glc (Jena Bioscience, CLK-076Motif) and 12.5 U T4 beta-
472 glucosyltransferase (Thermo Fisher Scientific, EO0831) for 2 hours at 37°C. Next, 5 µl DBCO-S-
473 S-biotin (Click Chemistry Tools, 10 mM stock in DMSO) was directly added to the reaction
474 mixture and incubated overnight at 37°C. DNA was cleaned using Oligo Clean & Concentrator
475 (Zymo, D4060), and immobilized on a PEGylated-Biotin streptavidin coated surface, followed by
476 imaging.

477 Single-molecule DNA sequencing

478 For single-molecule DNA sequencing of cfNucleosomes, PEGylated-poly T surface was blocked
479 with BSA as described above. Poly-A tailed FluoSpheres (described below) along with TetraSpeck
480 beads and cfNucleosomes were applied to surface. PTMs of plasma cfNucleosomes were imaged
481 over time for 169 FOV, as described above. Then flowcell was washed with Wash A buffer [150
482 mM HEPES (KOH, pH 7.0), 1×SSC, 0.1% SDS] and Wash B buffer [150 mM HEPES (KOH, pH
483 7.0), 150 mM NaCl] to evict histones and antibodies, and temperature was increased to 37°C.
484 Sequencing was performed as described previously, using Helicos True Single Molecule
485 Sequencing (http://seqll.com/)[21,45]. FluoSpheres preparation: FluoSpheres (Carboxylate-Modified
486 Microspheres, Thermo Fisher Scientific, F8789) were conjugated to dA50-amine (IDT), tailed as
487 previously described, and hybridized to the surface to serve as reference points for stage drift
488 correction during alignment of sequencing images.

489 Single-Molecule Hydroxymethylation sequencing

490 2.5 ng of plasma cfDNA was added to a 25 µl solution containing 50 mM HEPES buffer (pH 8),
491 25 mM $MgCl_2$, 60 µM UDP-6-N_3-Glc (Jena Bioscience, CLK-076) and 12.5 U T4 beta-
492 glucosyltransferase (Thermo Fisher Scientific, EO0831), and incubated for 2 hours at 37°C. Next,
493 5 µl DBCO-S-S-biotin (Click Chemistry Tools, 10 mM stock in DMSO) was directly added to the
494 reaction mixture and incubated overnight at 37°C. For 5hmC DNA pulldown, samples were
495 incubated at RT with 15 µl of Streptavidin beads (Thermo Fisher Scientific, Dy-11205D) for 1
496 hour, followed by 3 washes with 1X wash buffer, and elution in 20 µl of 125 mM TCEP (Thermo
497 Fisher Scientific, TS-77720). 5hmC eluted DNA was poly-adenylated and sequenced on a
498 PEGylated-poly T surface as described above. Metagene profile was generated using ngs.plot.

499 Image analysis

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

506 Predictive Power Score (PPS)

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

515 Machine learning model for sample classification

516 For sample binary classification, various machine-learning algorithms were trained on the features
517 that showed significant differences between healthy and CRC (Fig. 3c,d,e, Supplementary Fig.
518 5b,c), and evaluated for their performance using a four-fold cross-validation across all samples.
519 The best predictive performance was achieved by a Logistic Regression classifier. To improve
520 classifier performance, we conducted additional feature selection by training the classifier on all
521 possible feature combinations out of the significant features aforementioned. Evaluating the
522 resulting Area Under the Curve (AUC) values of repeated (500 iterations) four-fold cross-
523 validation for each combination revealed an optimal cumulative performance of a five feature
524 combination: H3K27me3/Nuc, H3K9me3/Nuc, CEA/MST1, H3K27me3+Nuc (% of
525 cfNucleosomes harboring H3K27me3) and global DNA methylation. To evaluate the classifier
526 overall performance using the selected features, we performed repeated (10K) 4-fold cross-
527 validation across all samples. For each iteration the sensitivity, specificity, accuracy, precision and
528 the AUC value were calculated and averaged over all iterations. R caret and Rweka packages were
529 used for machine-learning modeling.

530 Tissue and plasma signatures

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

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

554 Statistical analysis

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

560 Data availability

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

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

566

567

568

569

570

571

572

573

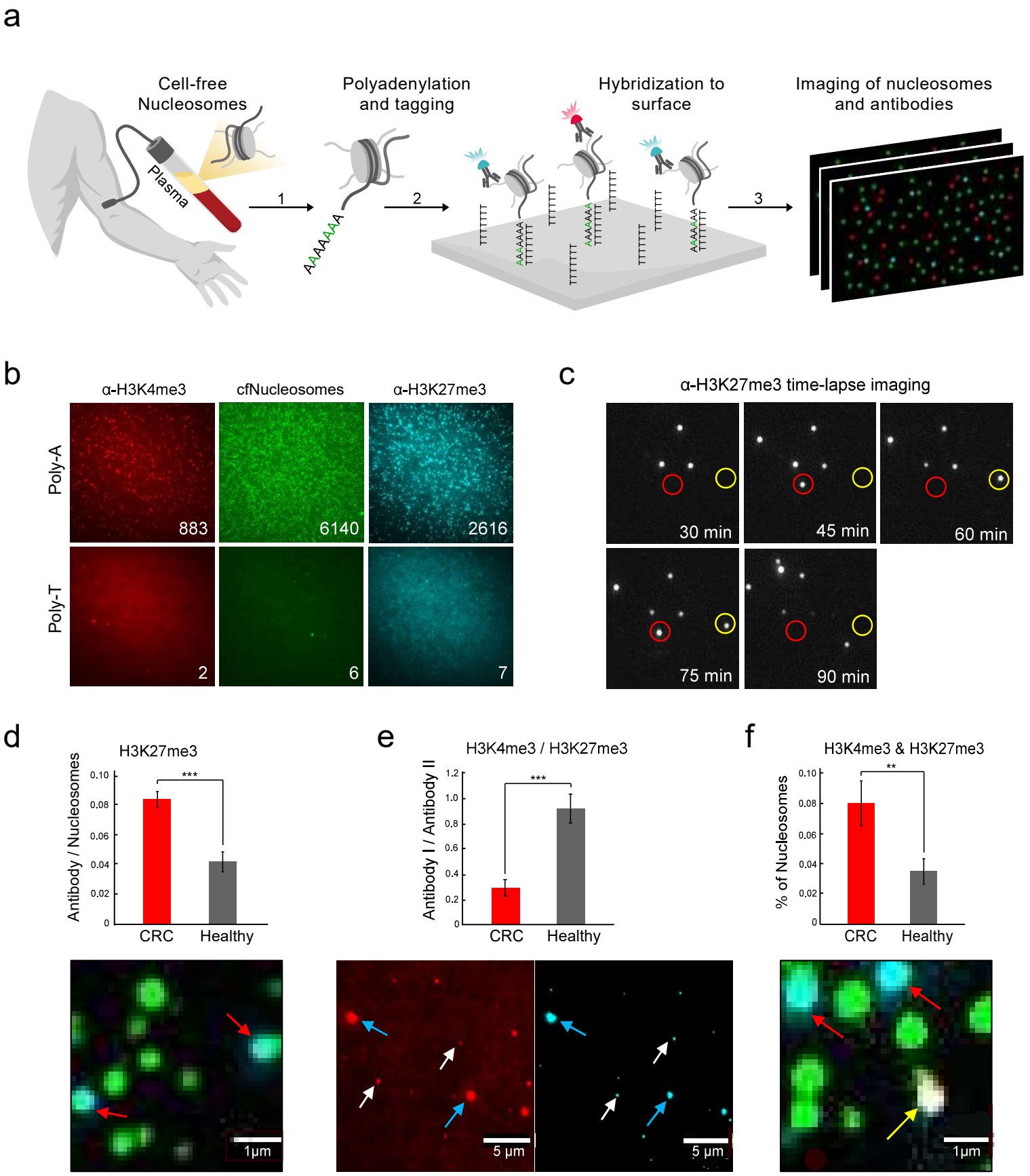
574

575

576

577

578


a
Cell-free Nucleosomes
Plasma
1
Polyadenylation and tagging
AAAAAAA
2
Hybridization to surface
3
Imaging of nucleosomes and antibodies
b
α-H3K4me3
cfNucleosomes
α-H3K27me3
Poly-A
Poly-T
883
6140
2616
2
6
7
c
α-H3K27me3 time-lapse imaging
30 min
45 min
60 min
75 min
90 min
d
H3K27me3
Antibody / Nucleosomes
***
CRC
Healthy
1μm
e
H3K4me3 / H3K27me3
Antibody I / Antibody II
***
CRC
Healthy
5 μm
5 μm
f
H3K4me3 & H3K27me3
% of Nucleosomes
**
CRC
Healthy
1μm

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

580 **a,** Experimental scheme: (1) Sample preparation procedure of cfNucleosomes is carried out in one-
581 step and consists of two enzymatic processes: repair of DNA ends by Klenow polymerase, and
582 addition of a poly A tail by Terminal Transferase (TdT). The reaction contains a mixture of natural
583 dATPs and fluorescently labeled dATPs (Cy3-dATP) to label nucleosomes. (2) cfNucleosmes are
584 captured on a PEGylated-poly T surface via dA:dT hybridization. Immobilized nucleosomes are
585 incubated with fluorescently labeled antibodies targeting different histone modifications. (3) TIRF
586 microscopy is applied to record nucleosome positions and generate time-lapse imaging of
587 antibodies' binding events. **b,** Representative images of plasma-derived cfNucleosomes and the
588 corresponding H3K4me3 and H3K27me3 signal (numbers within images represent counted spots).
589 Each spot corresponds to a single nucleosome. Nucleosomes anchor to the surface specifically via
590 hybridization, as evident from the lack of signal when tailed with dTTP (Poly-dT) rather than
591 dATP (Poly-dA). **c,** Representative images of antibodies' binding and dissociation events over
592 time from individual target molecules (marked by red/yellow circles). **d,e,f** Example for
593 quantification and representative images of the various parameters measured by EPINUC in
594 plasma samples from a healthy subject and a CRC patient. Zoomed-in image segments of entire
595 field of view (148μm^2). **d,** The percentage of cfNucleosomes (green spots, Cy3) that are modified
596 by H3K27me3 (cyan spots, AF488). Red arrows indicate co-localization events. Scale bar = 1µm
597 **e,** Ratio between H3K4me3 (red, AF647) and H3K27me3 (cyan) antibodies. White arrows indicate
598 antibody spots, blue arrows indicate TetraSpeck beads that are used for alignment. Scale bar =
599 5µm. **f,** cfNucleosomes marked by the combinatorial pattern of both H3K27me3 and H3K4me3.
600 Red arrows indicate co-localization events of H3K27me3 only, yellow arrow indicates a
601 combinatorially modified nucleosome. Scale bar = 1µm.

602

603

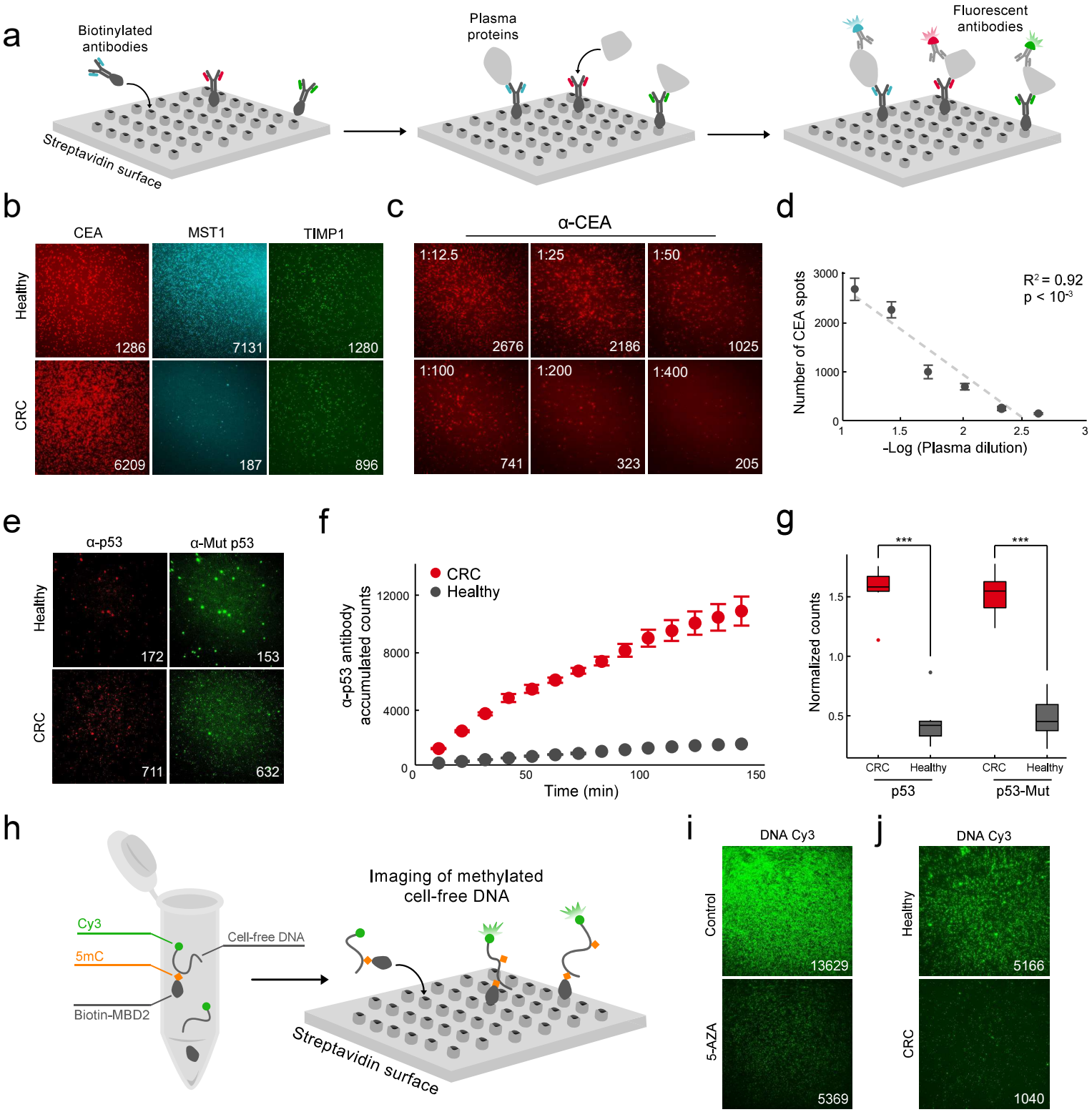
604

605

606

607

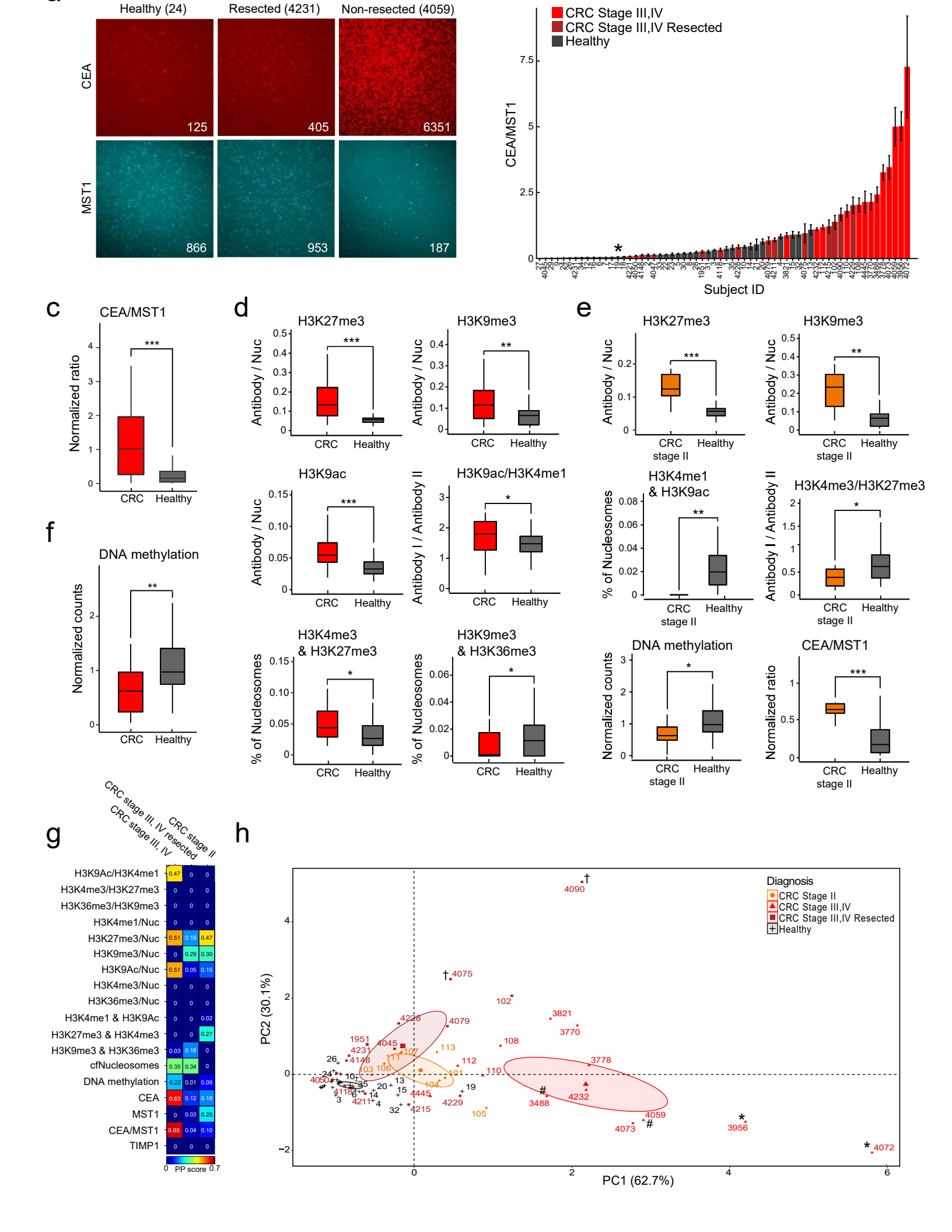
608


a
Biotinylated antibodies
Streptavidin surface
Plasma proteins
Fluorescent antibodies
b
CEA
MST1
TIMP1
Healthy
CRC
1286
7131
1280
6209
187
896
c
α-CEA
1:12.5
1:25
1:50
1:100
1:200
1:400
2676
2186
1025
741
323
205
d
Number of CEA spots
-Log (Plasma dilution)
R² = 0.92
p < 10⁻³
e
α-p53
α-Mut p53
Healthy
CRC
172
153
711
632
f
CRC
Healthy
α-p53 antibody accumulated counts
Time (min)
g
Normalized counts
CRC
Healthy
p53
p53-Mut
h
Cy3
5mC
Biotin-MBD2
Cell-free DNA
Imaging of methylated cell-free DNA
Streptavidin surface
i
DNA Cy3
Control
5-AZA
13629
5369
j
DNA Cy3
Healthy
CRC
5166
1040

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

611 **a,** Experimental scheme: biotinylated capture antibodies targeting distinct proteins are anchored
612 to a PEG-streptavidin surface. Plasma proteins are captured on surface, followed by detection with
613 fluorescently labeled antibodies and TIRF imaging. Multiplexed detection of up to three proteins
614 is achieved by labeling antibodies with different fluorophores. Each spot represents a single protein
615 bound on the surface. **b,** Representative TIRF images of selected CRC biomarkers measured
616 simultaneously for each plasma sample: CEA (red), MST1 (cyan) and TIMP-1(green). Images
617 reveal distinct biomarkers profiles for healthy and CRC. **c,** Representative TIRF images depicting
618 α-CEA antibody binding events (spots) over serial plasma dilutions. **d,** Regression analysis of the
619 number of spots as a function of plasma concentration highlights the linearity of detection. Data is
620 presented as the mean +/- s.d. of 50 FOVs for each concentration. **e,f,g** Single-molecule detection
621 of p53 in the plasma of healthy and CRC patients with known p53 mutations. **e,** Representative
622 TIRF images. Detection is carried out simultaneously with antibodies targeting all p53 (red) and
623 with antibodies that are specific to the mutant p53 conformation (green). Large diameter spots
624 correspond to TetraSpeck beads used for alignment. **f,** p53 signal accumulation over time in late
625 stage CRC and healthy plasma. Data is presented as the mean +/- s.d. of 50 FOVs for each time
626 point. **g,** Total and mutant p53 levels in plasma show significantly higher levels in CRC patients
627 versus healthy individuals (n=6 for each group). Box plots limits: 25–75% quantiles, middle:
628 median, upper (lower) whisker to the largest (smallest) value no further than 1.5× interquartile
629 range from the hinge. P values were calculated by Wilcoxon rank sum exact test. *** P value <
630 0.001. **h,** Experimental scheme for single-molecule imaging of global DNA methylation: MBD2-
631 biotin is incubated with Cy3-labeled (green) DNA, and binds specifically to methylated DNA
632 molecules. Next, biotin-MBD-meDNA complexes are immobilized on a PEG-streptavidin surface,
633 followed by TIRF imaging. Each spot represents a single bound complex, number of spots
634 correspond to the level of DNA methylation in plasma. **i,** Representative TIRF images of DNA
635 methylation in HEK293 cells treated with 5-Aza-2′-deoxycytidine, demonstrating significant
636 reduction in methylation compared to control cells. **j,** Representative TIRF images of global
637 cfDNA methylation levels in the plasma of CRC and healthy subjects, showing lower DNA
638 methylation levels in CRC. For all images, numbers within images represent counted spots.

639


Healthy (24)
Resected (4231)
Non-resected (4059)
CEA
MST1
125
405
6351
866
953
187
CRC Stage III,IV
CRC Stage III,IV Resected
Healthy
CEA/MST1
Subject ID
c
CEA/MST1
Normalized ratio
CRC
Healthy
d
H3K27me3
H3K9me3
H3K9ac
H3K9ac/H3K4me1
H3K4me3 & H3K27me3
H3K9me3 & H3K36me3
Antibody / Nuc
Antibody I / Antibody II
% of Nucleosomes
e
H3K4me1 & H3K9ac
H3K4me3/H3K27me3
DNA methylation
CRC stage II
f
DNA methylation
Normalized counts
g
CRC stage III, IV
CRC stage III, IV resected
CRC stage II
H3K9Ac/H3K4me1
H3K4me3/H3K27me3
H3K36me3/H3K9me3
H3K4me1/Nuc
H3K27me3/Nuc
H3K9me3/Nuc
H3K9Ac/Nuc
H3K4me3/Nuc
H3K36me3/Nuc
H3K4me1 & H3K9Ac
H3K27me3 & H3K4me3
H3K9me3 & H3K36me3
cfNucleosomes
DNA methylation
CEA
MST1
CEA/MST1
TIMP1
PP score
h
Diagnosis
CRC Stage II
CRC Stage III,IV
CRC Stage III,IV Resected
Healthy
PC2 (30.1%)
PC1 (62.7%)

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

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

660

661

662

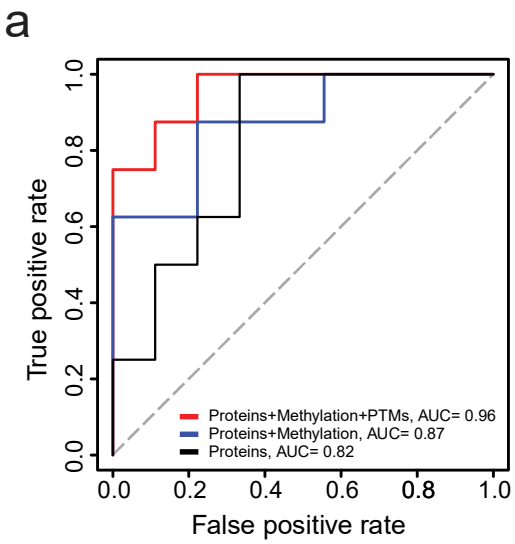
663

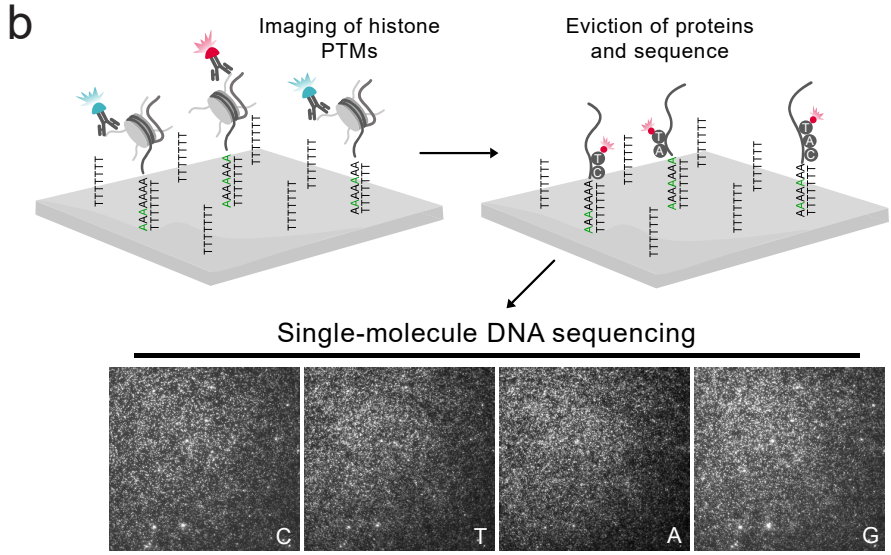
664

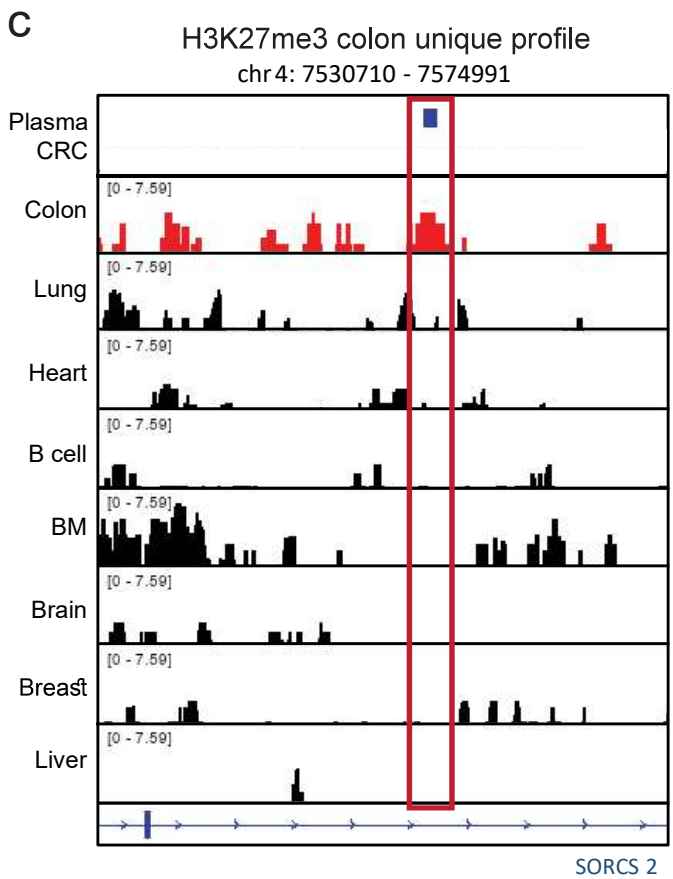
665

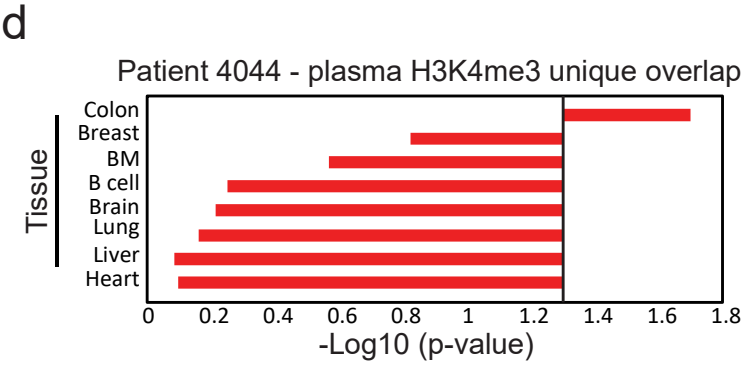
666

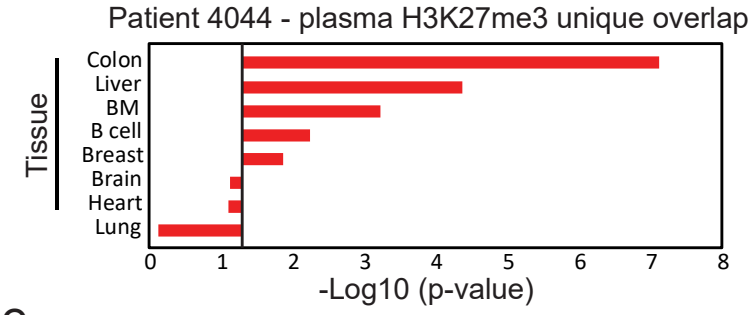
667

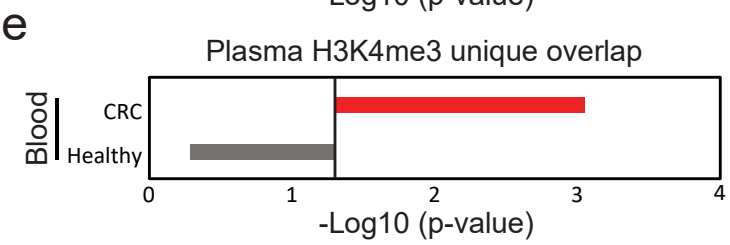

a
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Proteins+Methylation+PTMs, AUC= 0.96
Proteins+Methylation, AUC= 0.87
Proteins, AUC= 0.82

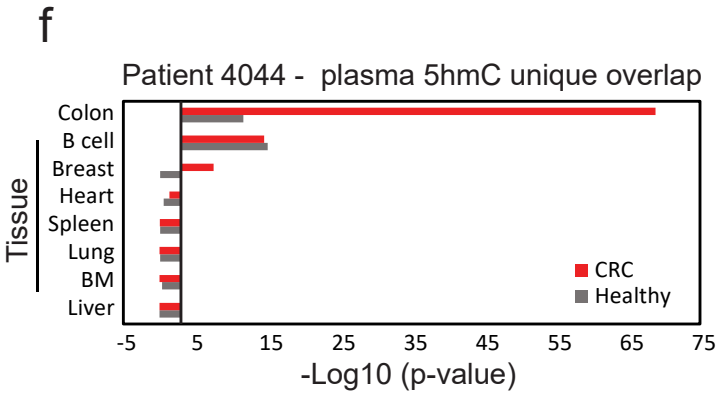

b
Imaging of histone PTMs
Eviction of proteins and sequence
Single-molecule DNA sequencing
C
T
A
G

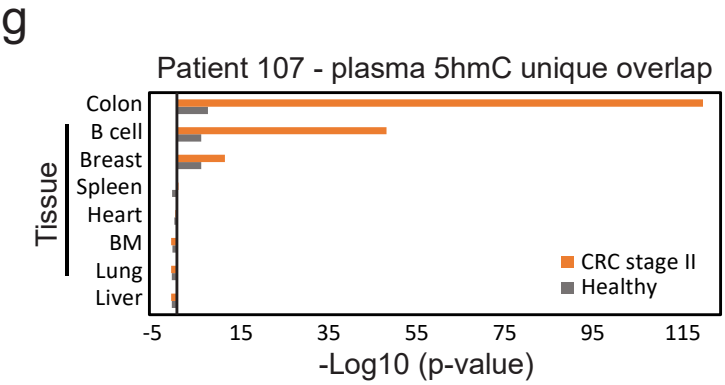

c
H3K27me3 colon unique profile
chr 4: 7530710 - 7574991
Plasma CRC
Colon
Lung
Heart
B cell
BM
Brain
Breast
Liver
[0 - 7.59]
SORCS 2


d
Patient 4044 - plasma H3K4me3 unique overlap
Tissue
Colon
Breast
BM
B cell
Brain
Lung
Liver
Heart
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
-Log10 (p-value)


Patient 4044 - plasma H3K27me3 unique overlap
Tissue
Colon
Liver
BM
B cell
Breast
Brain
Heart
Lung
0
1
2
3
4
5
6
7
8
-Log10 (p-value)


e
Plasma H3K4me3 unique overlap
Blood
CRC
Healthy
0
1
2
3
4
-Log10 (p-value)


f
Patient 4044 - plasma 5hmC unique overlap
Tissue
Colon
B cell
Breast
Heart
Spleen
Lung
BM
Liver
CRC
Healthy
-5
5
15
25
35
45
55
65
75
-Log10 (p-value)


g
Patient 107 - plasma 5hmC unique overlap
Tissue
Colon
B cell
Breast
Spleen
Heart
BM
Lung
Liver
CRC stage II
Healthy
-5
15
35
55
75
95
115
-Log10 (p-value)

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

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

688

689

690

691

692

693

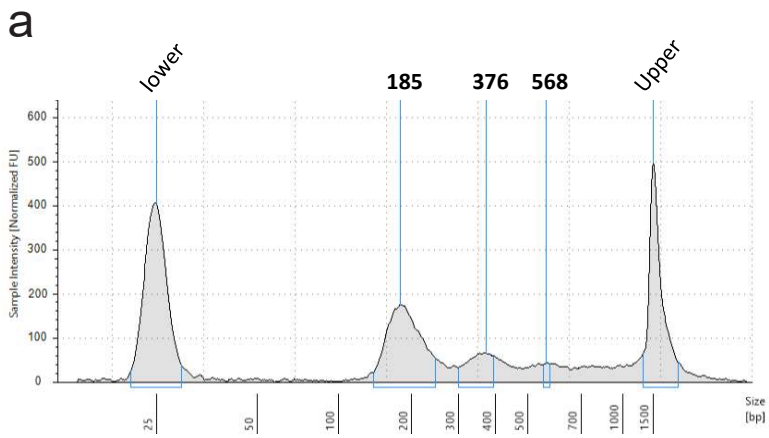
694

695

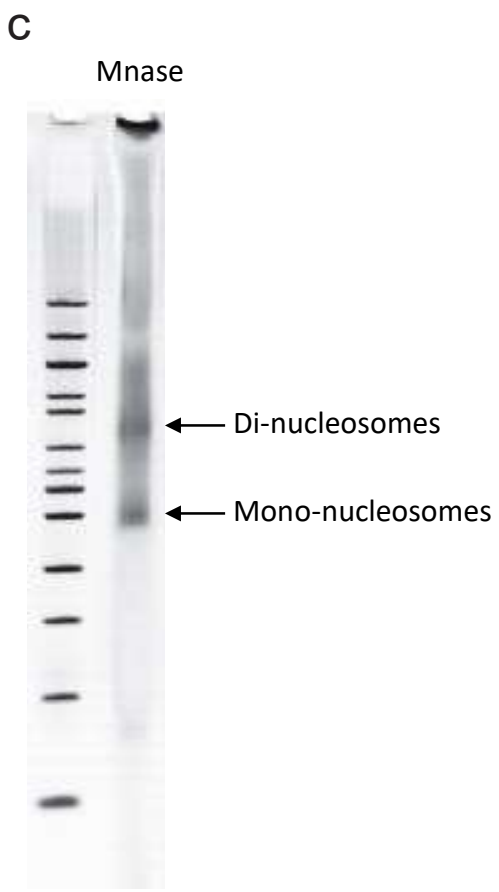
696

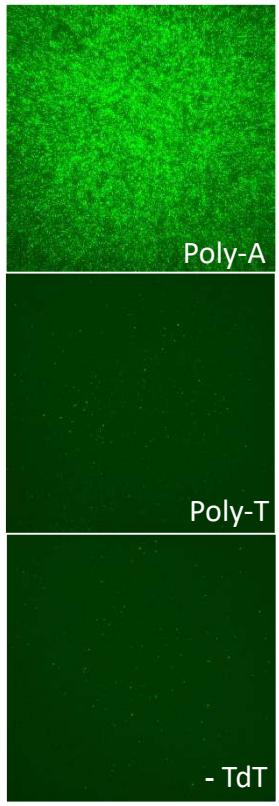
697

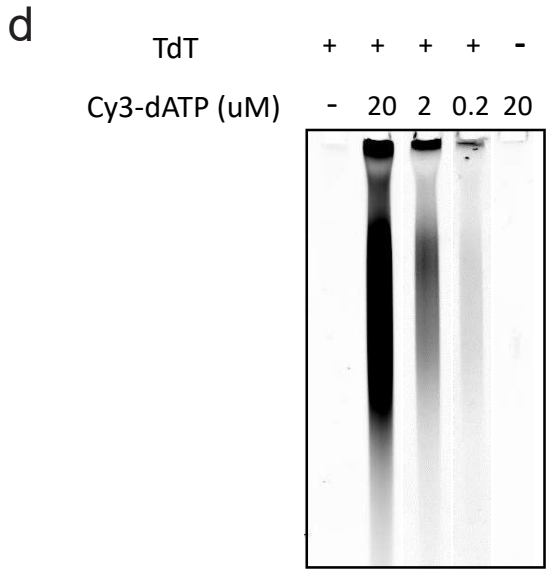
698

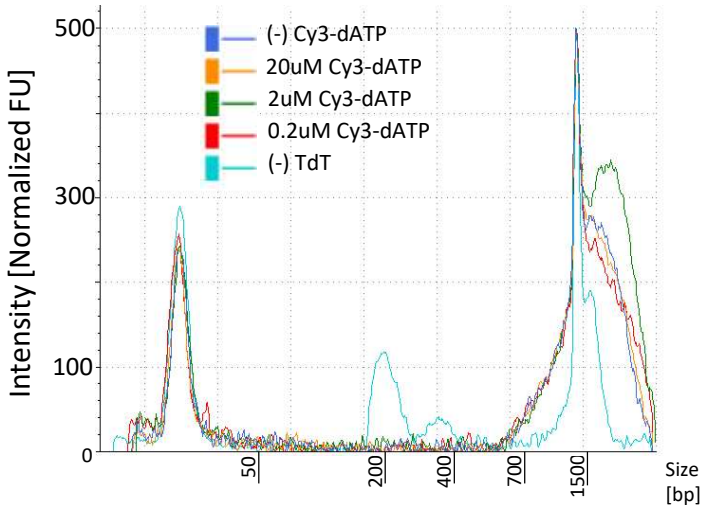

a
lower
185
376
568
Upper
Sample Intensity [Normalized FU]
600
500
400
300
200
100
0
25
50
100
200
300
400
500
700
1000
1500
Size [bp]

b
Medium cfNucleosomes
STS
-
+


c
Mnase
Di-nucleosomes
Mono-nucleosomes


f
Medium cfNucleosomes
Poly-A
Poly-T
- TdT


d
TdT
+ + + + -
Cy3-dATP (uM)
- 20 2 0.2 20


e
(-) Cy3-dATP
20uM Cy3-dATP
2uM Cy3-dATP
0.2uM Cy3-dATP
(-) TdT
Intensity [Normalized FU]
500
300
100
0
50
200
400
700
1500
Size [bp]

699 **Supplementary Figure 1**

700 **In-vitro system for sample preparation of cfNucleosomes.**

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

715

716

717

718

719

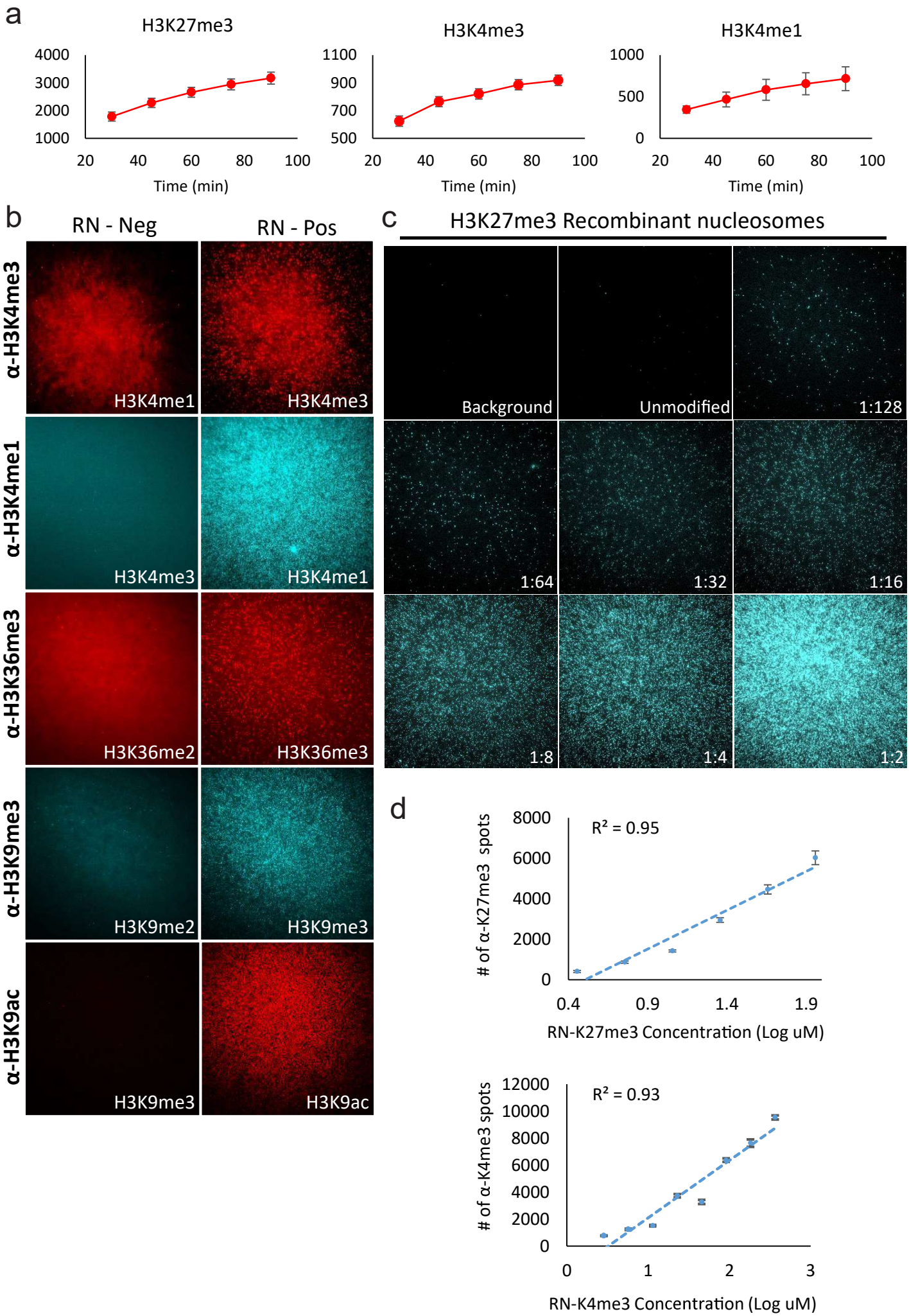
720

721

722

723

724


a
H3K27me3
4000
3000
2000
1000
20
40
60
80
100
Time (min)
H3K4me3
1100
900
700
500
20
40
60
80
100
Time (min)
H3K4me1
1000
500
0
20
40
60
80
100
Time (min)
b
RN - Neg
RN - Pos
α-H3K4me3
H3K4me1
H3K4me3
α-H3K4me1
H3K4me3
H3K4me1
α-H3K36me3
H3K36me2
H3K36me3
α-H3K9me3
H3K9me2
H3K9me3
α-H3K9ac
H3K9me3
H3K9ac
c
H3K27me3 Recombinant nucleosomes
Background
Unmodified
1:128
1:64
1:32
1:16
1:8
1:4
1:2
d
R² = 0.95
# of α-K27me3 spots
8000
6000
4000
2000
0
0.4
0.9
1.4
1.9
RN-K27me3 Concentration (Log uM)
R² = 0.93
# of α-K4me3 spots
12000
10000
8000
6000
4000
2000
0
0
1
2
3
RN-K4me3 Concentration (Log uM)

725 **Supplementary Figure 2**

726 **Single-molecule measurements of antibodies' specificity and dynamics.**

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

740

741

742

743

744

745

746

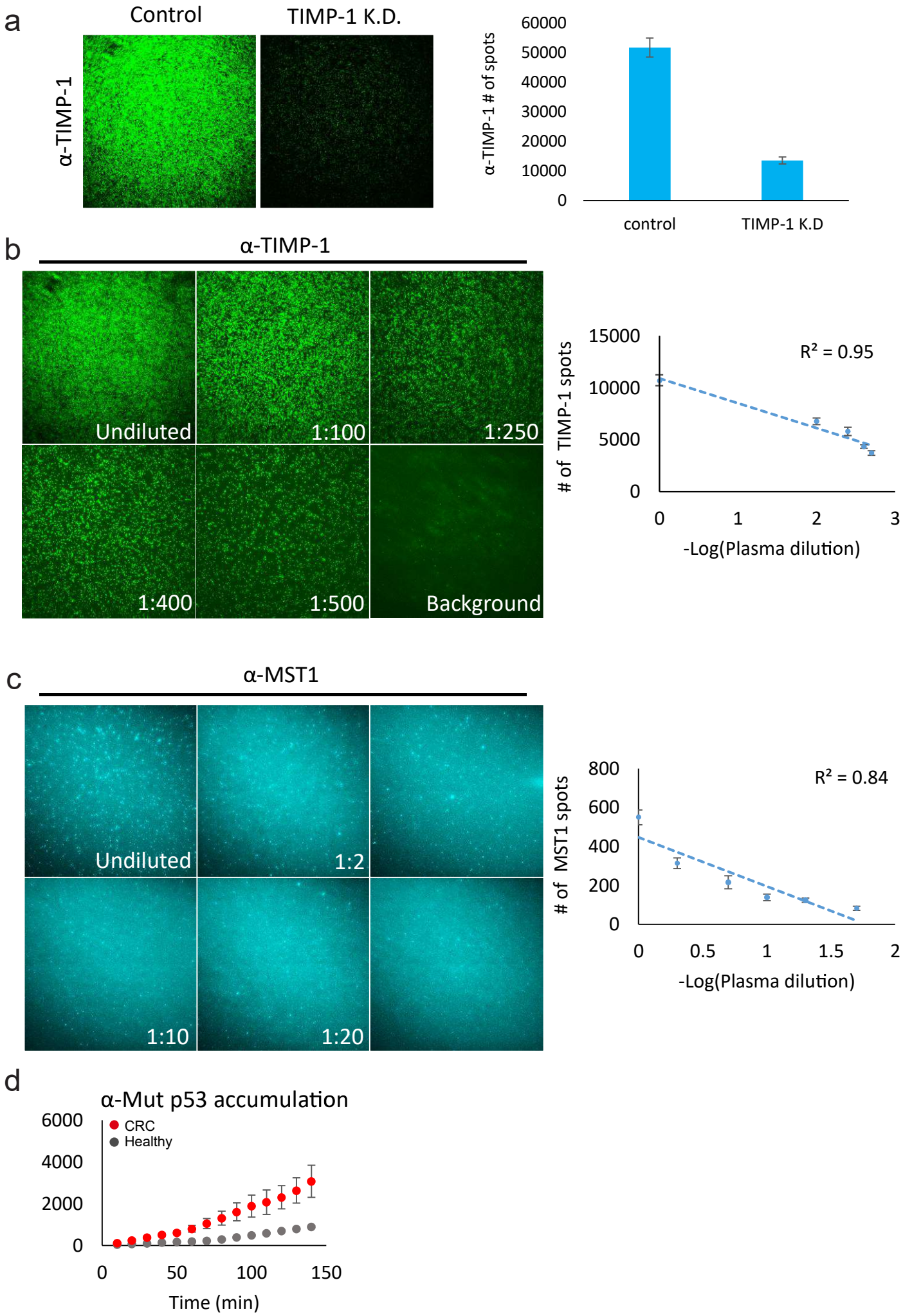
747

748

749

750

751


a
Control
TIMP-1 K.D.
α-TIMP-1
α-TIMP-1 # of spots
60000
50000
40000
30000
20000
10000
0
control
TIMP-1 K.D
b
α-TIMP-1
Undiluted
1:100
1:250
1:400
1:500
Background
# of TIMP-1 spots
15000
10000
5000
0
R² = 0.95
0
1
2
3
-Log(Plasma dilution)
c
α-MST1
Undiluted
1:2
1:10
1:20
# of MST1 spots
800
600
400
200
0
R² = 0.84
0
0.5
1
1.5
2
-Log(Plasma dilution)
d
α-Mut p53 accumulation
CRC
Healthy
6000
4000
2000
0
0
50
100
150
Time (min)

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

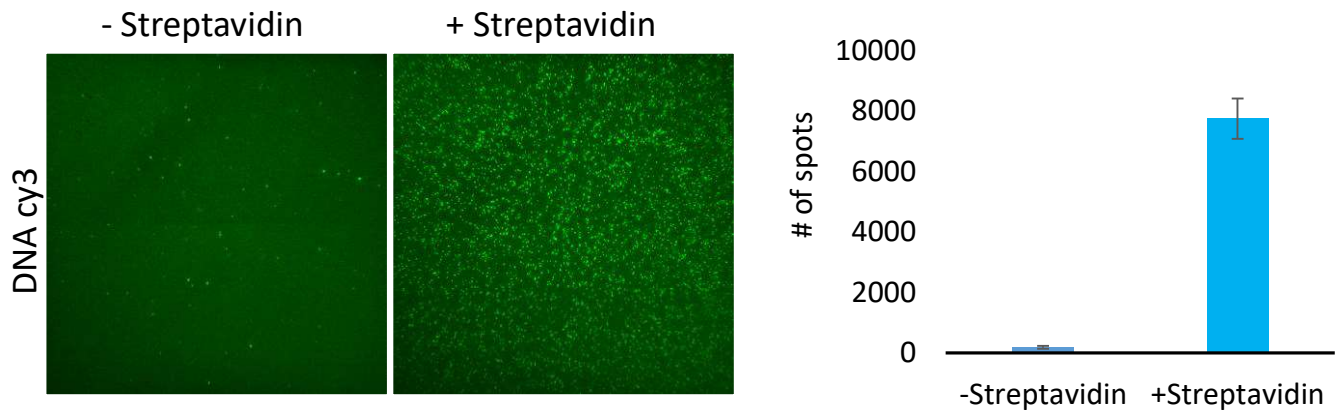
774

775

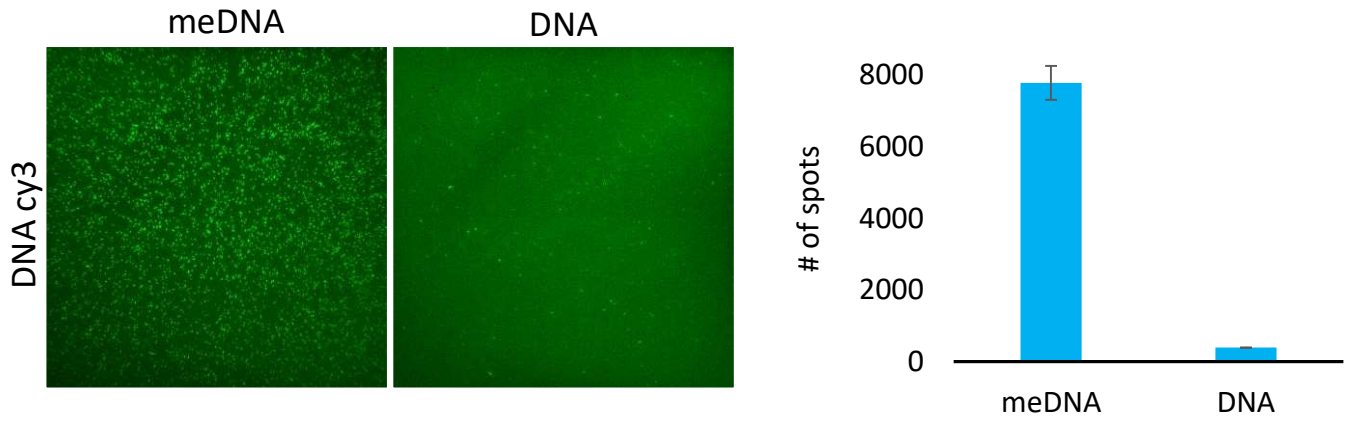
776

777

a


- Streptavidin
+ Streptavidin
DNA cy3
10000
8000
6000
4000
2000
0
# of spots
-Streptavidin
+Streptavidin

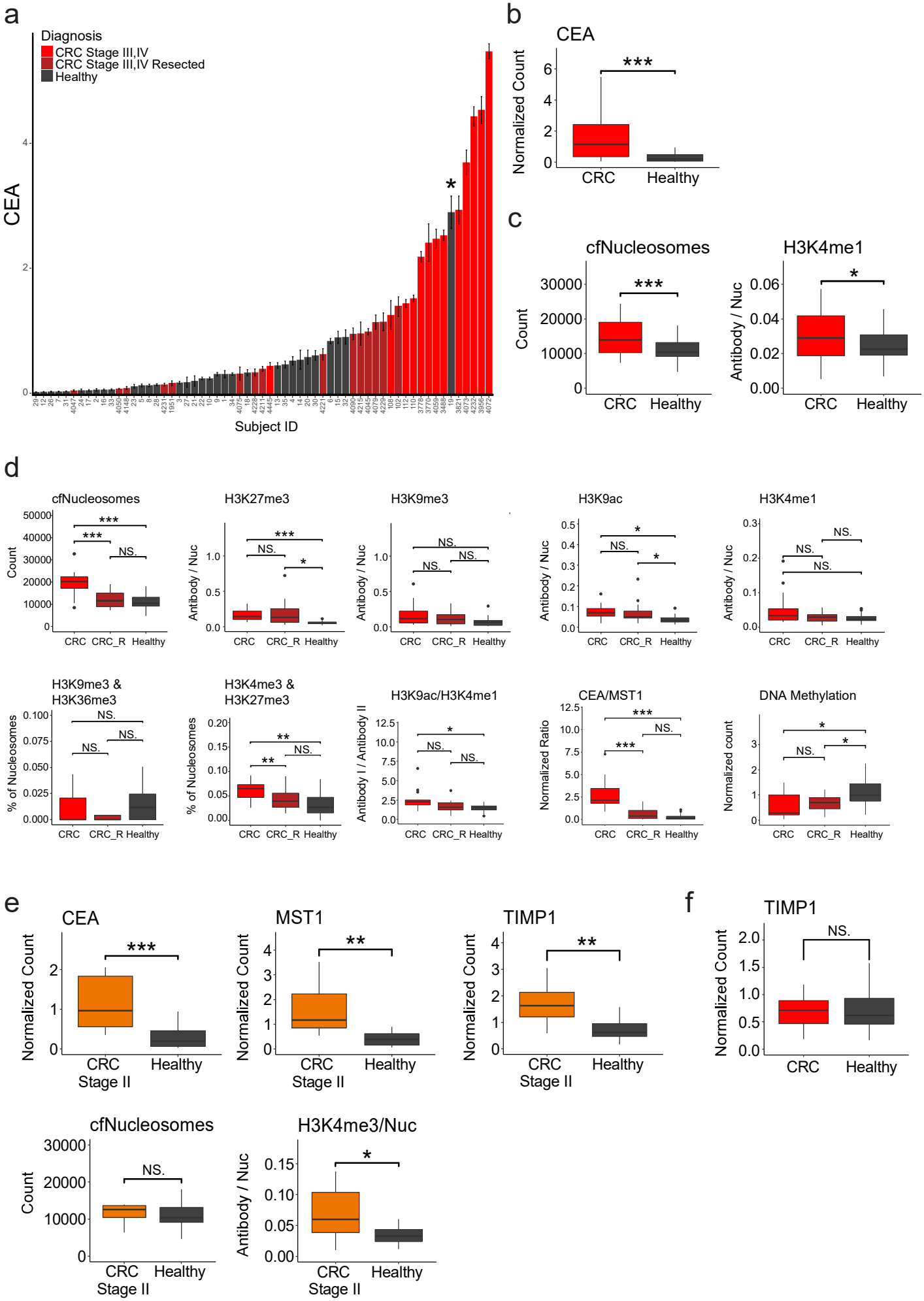
b


meDNA
DNA
DNA cy3
8000
6000
4000
2000
0
# of spots
meDNA
DNA

Supplementary Figure 4

Single-molecule detection of DNA methylation levels.

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

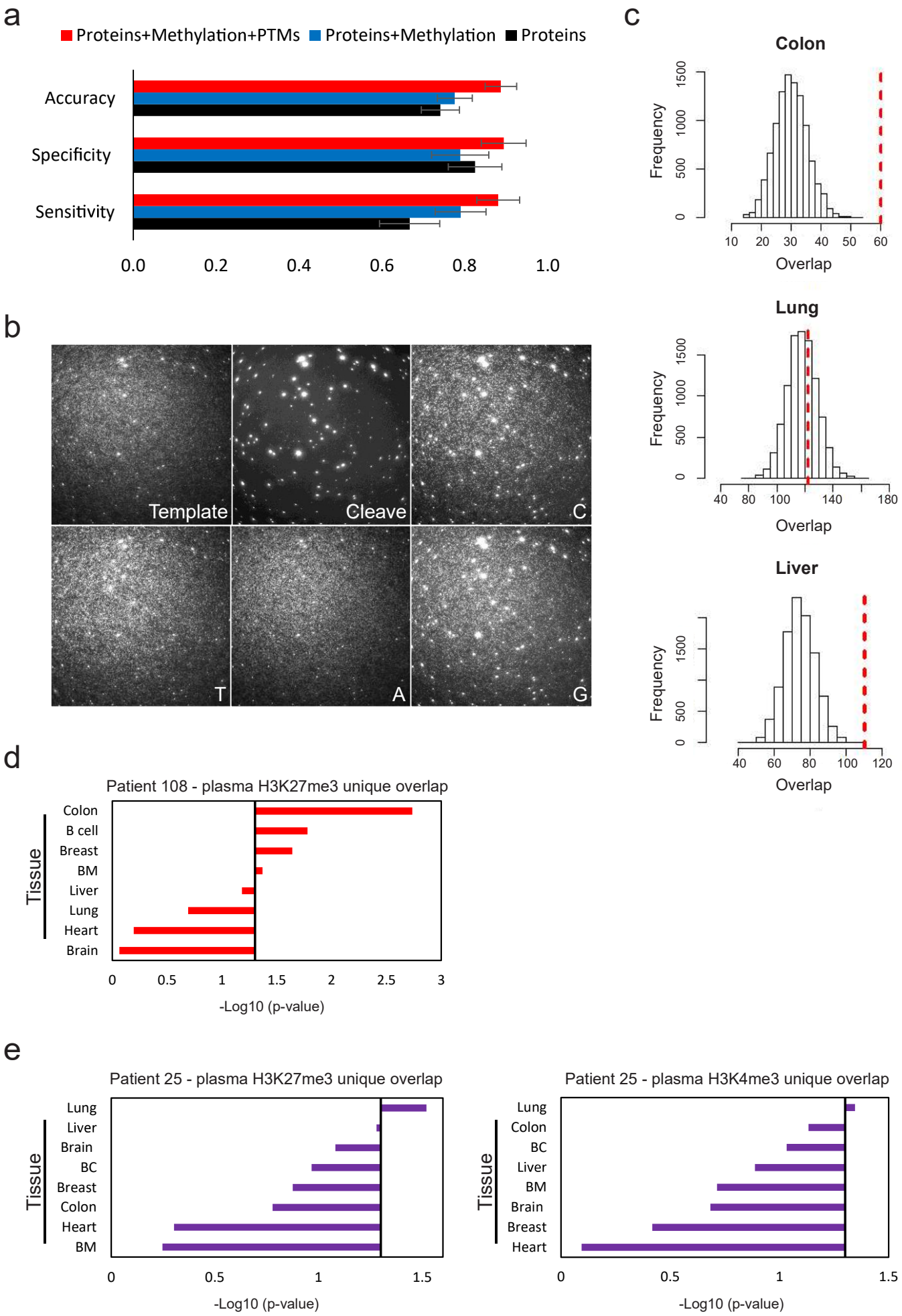

a
Diagnosis
CRC Stage III,IV
CRC Stage III,IV Resected
Healthy
CEA
Subject ID
b
CEA
Normalized Count
CRC
Healthy
c
cfNucleosomes
Count
H3K4me1
Antibody / Nuc
d
cfNucleosomes
H3K27me3
H3K9me3
H3K9ac
H3K4me1
H3K9me3 & H3K36me3
H3K4m3 & H3K27me3
H3K9ac/H3K4me1
CEA/MST1
DNA Methylation
% of Nucleosomes
Antibody I / Antibody II
Normalized Ratio
Normalized count
CRC_R
e
CEA
MST1
TIMP1
Stage II
cfNucleosomes
H3K4me3/Nuc
f
TIMP1

808 **Supplementary Figure 5**

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

811 **(a)** CEA normalized levels in the plasma of CRC patients and healthy individuals. Each bar
812 represents a subject, data is presented as the mean +/- s.d. of 50 FOVs per sample. Sample 19
813 (Healthy) denoted by *. **(b)** Box plot representation of the data in A (healthy = 33, CRC = 29).
814 Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest
815 (smallest) value no further than 1.5× interquartile range from the hinge. P values were calculated
816 by Welch's t-test. *** P value < 0.001. **(c)** Global level of cfNucleosomes, and levels of H3K4me1-
817 modified nucleosomes, significantly differ between healthy and CRC late stage samples (healthy
818 = 33, CRC = 29). P values were calculated by Welch's t-test. * P value < 0.05 ** P value < 0.01.
819 *** P value < 0.001. **(d)** Multiple comparison of all significant parameters between CRC, CRC
820 resected (CRC_R) and healthy samples (CRC = 13, CRC_R=16, healthy = 33), corresponding to
821 the data presented in main Fig. 3c,d,e. Of note, while all of these parameters differ between healthy
822 versus the combined cohort of <u>all CRC</u> patients, this figure shows the differences between CRC
823 patients with/without tumor resection. In some parameters, resected patients show higher similarity
824 to healthy, and in other parameters, they are similar to CRC patients prior to tumor resection. See
825 methods for P value calculation. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(e)**
826 Levels of CEA, MST1 and TIMP1, as well as H3K4me3-modified nucleosomes, significantly
827 differ between healthy and early stage CRC patients (healthy = 33, early CRC = 8). Total levels of
828 cfNucleosomes do not show significant difference between the groups, likely due to low tumor
829 burden at early stage patients. P values were calculated by Wilcoxon rank sum exact test. * P value
830 < 0.05 ** P value < 0.01. *** P value < 0.001. **(f)** TIMP1 levels do not significantly differ in the
831 cohort of healthy versus late-stage CRC patients.

832
833
834
835
836
837
838

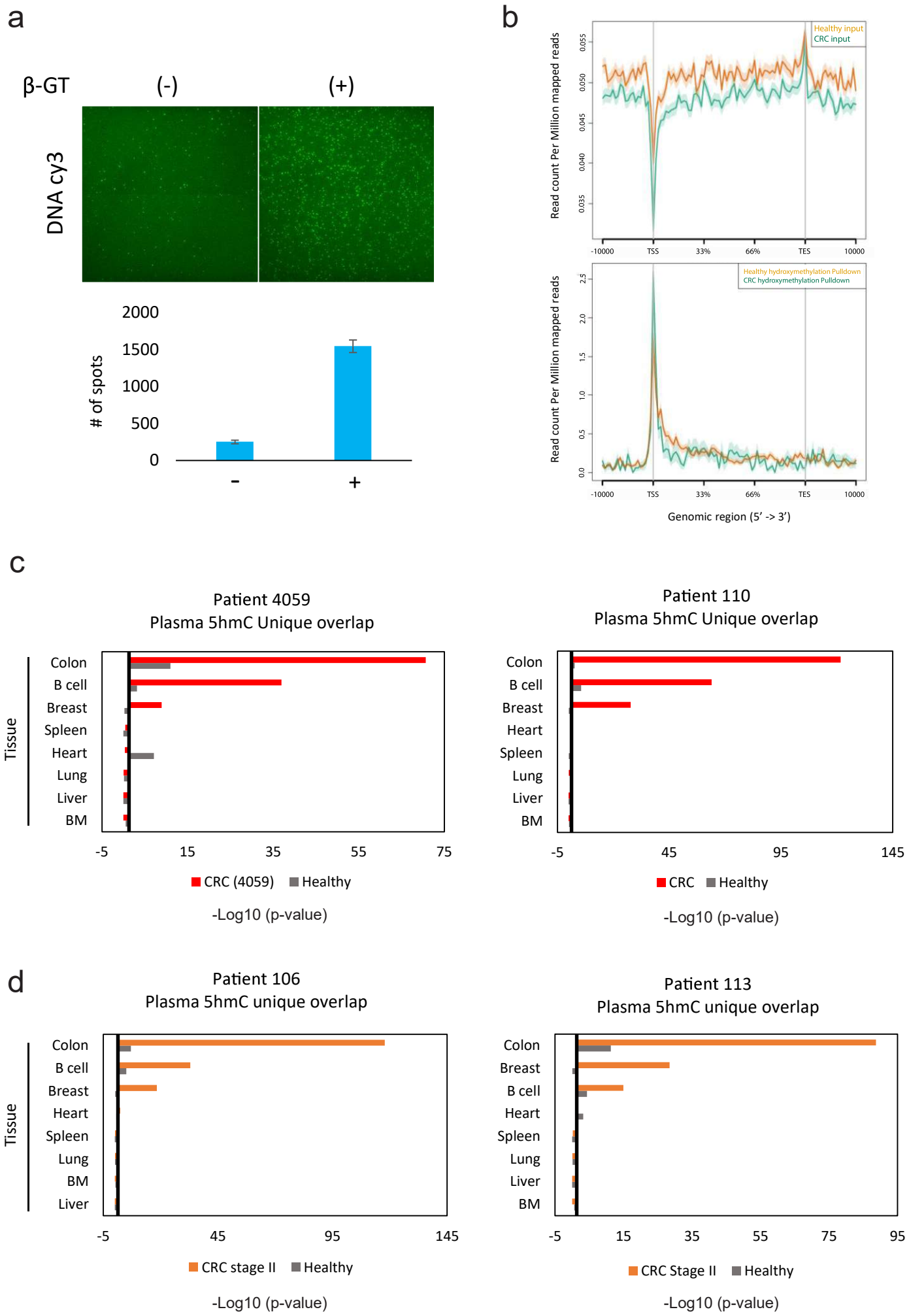

a
Proteins+Methylation+PTMs
Proteins+Methylation
Proteins
Accuracy
Specificity
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
b
Template
Cleave
C
T
A
G
c
Colon
Lung
Liver
Frequency
Overlap
d
Patient 108 - plasma H3K27me3 unique overlap
Tissue
Colon
B cell
Breast
BM
Liver
Lung
Heart
Brain
-Log10 (p-value)
e
Patient 25 - plasma H3K27me3 unique overlap
Patient 25 - plasma H3K4me3 unique overlap
Tissue
Lung
Liver
Brain
BC
Breast
Colon
Heart
BM
-Log10 (p-value)

839 **Supplementary Figure 6**

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

841 **(a)** Logistic model performance increases with integration of additional layers of information.
842 Corresponds to Fig. 4a. **(b)** Representative TIRF images of antibodies' binding (Template)
843 followed by cleavage of fluorophore (Cleave) and sequencing cycles (C,T,A,G). High diameter
844 spots correspond to FluoSpheres used for image alignment. **(c)** Histograms portray colon, lung and
845 liver frequency distributions of random reads (equivalent to the number of H3K27me3 plasma
846 CRC positive reads) that overlap with tissue H3K27me3 unique peaks (bootstrapping simulation,
847 10K iterations, Methods). Red line corresponds to the number of H3K27me3 unique tissue peaks
848 that overlap with CRC (patient 4044) H3K27me3 positive plasma reads. (**d**) EPINUC-seq analysis
849 of an additional CRC patient, similar to the analysis shown in Fig. 4d. (**e**) EPINUC-seq analysis
850 of plasma from stage IV lung cancer patient (patient 25). Tissues and primary cell lines ranked by
851 overlap significance with single-molecule H3K27me3 (Left) or H3K4me3 (Right) positive reads,
852 indicating lung as the tissue-of-origin. Black line corresponds to P value of 0.05. P values were
853 determined by two tailed Z-test or Wilcoxon rank sum test.

854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869


a
β-GT
(-)
(+)
DNA cy3
# of spots
2000
1500
1000
500
0
-
+
b
Healthy input
CRC input
Read count Per Million mapped reads
0.055
0.050
0.045
0.040
0.035
-10000
TSS
33%
66%
TES
10000
Healthy hydroxymethylation Pulldown
CRC hydroxymethylation Pulldown
2.5
2.0
1.5
1.0
0.5
0.0
Genomic region (5' -> 3')
c
Patient 4059
Plasma 5hmC Unique overlap
Tissue
Colon
B cell
Breast
Spleen
Heart
Lung
Liver
BM
-5
15
35
55
75
CRC (4059)
Healthy
-Log10 (p-value)
Patient 110
Plasma 5hmC Unique overlap
-5
45
95
145
CRC
Healthy
-Log10 (p-value)
d
Patient 106
Plasma 5hmC unique overlap
CRC stage II
Healthy
-Log10 (p-value)
Patient 113
Plasma 5hmC unique overlap
CRC Stage II
Healthy
-Log10 (p-value)

870 **Supplementary Figure 7**

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

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

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

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).

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

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

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

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

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

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

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

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

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

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

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

962 29. Nakayama, M. & Oshima, M. Mutant p53 in colon cancer. *J. Mol. Cell Biol.* **11,** 267–276
963 (2019).

964 30. Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F. & Goel, A. Epigenetics of
965 colorectal cancer: biomarker and therapeutic potential. *Nat. Rev. Gastroenterol. Hepatol.*
966 **17,** 111–130 (2020).

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

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

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

973 34. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised
974 epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* **14,** 747–753
975 (2014).

976 35. Kerachian, M. A. *et al.* Crosstalk between DNA methylation and gene expression in
977 colorectal cancer, a potential plasma biomarker for tracing this tumor. *Sci. Rep.* **10,** 1–13
978 (2020).

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

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

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

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

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

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

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

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

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

997 45. Harris, T. D. *et al.* Single-Molecule DNA Sequencing of a Viral Genome. *Science (80-.).*
998 **320,** 106–109 (2008).

999