# 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 colorectal, pancreatic, lung and breast tumors. 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]. TIRF provides a powerful means for detecting single fluorescent molecules that are within ~100 nm of a solid surface; a light source creates an evanescent wave that decays exponentially in intensity, thus eliminating background fluorescence from outside the focal plane. To capture nucleosomes from plasma, we developed high-efficiency enzymatic reactions to fluorescently tag and polyadenylate nucleosomes (Supplementary Fig.1a-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, verifying minimal spectral overlap (Fig. 1b and Supplementary Fig. 1f,g). Binding and dissociation of antibodies to target PTMs was imaged over 90 minutes, leveraging the TIRF narrow excitation range. 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-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-f). Repeated measurements of the same plasma samples produced highly similar results, attesting to the quantitative nature of this technology (Supplementary Fig.3). 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).

99 To extend the number of analytes measured beyond histone PTMs, we exploited the single-
100 molecule system for quantification of protein biomarkers. We modulated surface chemistry to
101 contain PEG-streptavidin, allowing anchoring of biotin-conjugated antibodies that target
102 plasma proteins. Following incubation with plasma, bound proteins are imaged by fluorescent
103 detection antibodies. Multiplexed simultaneous detection of three biomarkers is achieved
104 through the use of distinct fluorophores (Fig. 2a). We imaged two proteins known to increase
105 in plasma of CRC patients: Carcinoembryonic antigen (CEA), a canonical biomarker measured
106 routinely by clinicians[24], and Tissue inhibitor of metalloproteinase-1 (TIMP-1), a glycoprotein
107 reported to have diagnostic value in screening for CRC[25]. In addition, we measured the
108 mammalian sterile 20-like kinase 1 (MST1), an inhibitor of cell proliferation that decreases in
109 CRC patients[26] (Fig. 2b). We verified linear detection and specificity using cell-culture systems
110 and knockdown experiments (Fig. 2c,d and Supplementary Fig. 4).

111 Counting of single molecules confers high sensitivity[27,28], thus we explored whether we could
112 also quantify non-secreted tumor-specific plasma proteins that are undetectable by
113 conventional technologies. We focused on the tumor suppressor p53, which is frequently
114 mutated in CRC; p53 mutations lead to its stabilization and accumulation in tumor cells[29]. We
115 captured p53 from plasma and applied simultaneous detection with two distinct antibodies; an
116 antibody targeting both the wild type and mutant forms of p53, or another antibody specifically
117 targeting the mutant conformation (Fig. 2e). Time-lapse imaging enabled the accumulation of
118 p53 signal, overcoming the transient binding dynamics of the detection antibodies (Fig. 2f and
119 Supplementary Fig. 4d). Indeed, we observed higher levels of total and mutant p53 in the
120 plasma of CRC patients with confirmed p53 mutations (Fig. 2g), establishing our system's
121 capabilities in specific detection of mutant proteins that originate directly from tumor cells.

122 DNA methylation is often deregulated in cancer, and specifically in colorectal cancer[30,31]. We
123 therefore aimed to combine our analysis with quantitative single-molecule detection of DNA
124 methylation levels in plasma. We incubated Methyl-CpG-binding domain protein 2 (MBD2-
125 biotin), which specifically binds to methylated DNA[32], with fluorescently labeled plasma
126 cfDNA. Bound complexes were anchored to the surface and imaged (Fig. 2h). Of note, bound
127 DNA molecules may harbor one or more methylation sites. Specificity and sensitivity were
128 validated using synthetic methylated/unmethylated DNA, as well as DNA from cells treated
129 with the DNA methyl transferase (DNMT) inhibitor 5-Aza-2′-deoxycytidine (Fig. 2i and
130 Supplementary Fig. 5a-c). Finally, we verified detection of cfDNA methylation levels from
131 plasma of CRC and healthy subjects (Fig. 2j and Supplementary Fig. 5d).

132 We applied EPINUC to generate high-dimensional data, comprising of the three layers of
133 information; histone PTMs, DNA methylation and protein biomarkers, from 33 plasma
134 samples of healthy subjects and 46 samples taken from 40 late stage CRC patients (stages III-
135 IV; six patients were sampled twice at different times during cancer progression and treatment).
136 CRC samples were obtained from patients prior to surgery or from patients that underwent
137 surgical resection procedure and chemotherapy. In accordance with its use in clinical
138 diagnostics[24], single-molecule counting of CEA showed higher levels in CRC patients (Fig. 3a
139 and Supplementary Fig. 6a,b), and a reduction in patients after resection. Interestingly, high
140 CEA levels were also observed in a few healthy individuals, generating a 'false positive' signal
141 (for example, samples 6, 9 and 19, marked by * in Supplementary Fig. 6a). Simultaneous
142 probing of MST1, an anti-proliferative factor, allowed us to derive the CEA/MST1 ratio,
143 resulting in better classification of samples and highlighting an advantage of combinatorial
144 biomarker detection (Fig. 3b,c and Supplementary Fig. 6a). Of note, plasma from CRC patients
145 following resection exhibited altered CEA/MST1 ratio compared to non-resected patients,
146 showing higher similarity to healthy individuals (Fig. 3a,b). This demonstrates the potential
147 applicability of our technology to monitor treatment, while underlying the need to collect
148 additional information from each sample to allow correct sample classification.

149 EPINUC also provides quantitative measurements of the total number of cfNucleosomes, six
150 histone PTMs, their pairwise combinations and ratios per plasma sample (Fig. 1). In agreement
151 with the literature, CRC patients had higher cfNucleosomes in their plasma compared to
152 healthy controls[33] (Supplementary Fig. 6c). While most epigenetic parameters did not change,
153 several showed significant differences: CRC patients had higher levels of H3K27me3-,
154 H3K9me3- , H3K9ac- and H3K4me1-modified nucleosomes, and higher ratio of H3K9ac to
155 H3K4me1 (Fig. 3d and Supplementary Fig. 6c,d). Interestingly, the combinatorial pattern of
156 H3K9me3+H3K36me3-modified nucleosomes decreased in CRC, concomitant with an
157 increase in 'bivalent' nucleosomes marked by H3K4me3+H3K27me3 (Fig. 3d). As bivalent
158 chromatin is strongly implicated in many types of cancers[21,34], this result further confirms the
159 diagnostic value of single-molecule quantification of combinatorial histone marks. DNA
160 methylation was reduced in CRC samples, in agreement with previous studies[35,36] (Fig. 3e).

161 The identification of epigenetic and biomarkers alterations in late stage CRC motivated us to
162 apply EPINUC to 17 plasma samples from individuals diagnosed with early stage CRC (stages
163 I,II). As in the later stage, the levels of DNA methylation, CEA and CEA/MST1 ratio
164 significantly differed in early stage cancer patients versus healthy (Fig. 3f and Supplementary
165 Fig. 6e). Interestingly, TIMP1, whose levels did not alter between our cohort of healthy and

late stage CRC, was elevated at the early stage (Supplementary Fig. 6e,f). This might be due to chemotherapy treatment administrated to most of the late stage CRC patients, which was reported to downregulate TIMP1 levels[37]. On the contrary, early stage CRC patients did not receive chemotherapy, and indeed showed elevated TIMP1, rendering it a significant biomarker only for early stage. Of note, plasma from stages I,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. 6e). 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). These results indicate various epigenetic alterations already occurring at early stage CRC.

As proof-of-concept that EPINUC could be generalized and applied to study diverse types of cancers, we obtained and analyzed 10 plasma samples from patients diagnosed with pancreatic ductal adenocarcinoma (PDAC), a devastating disease with poor prognosis and rising incidence[38]. The data revealed multiple epigenetic parameters, as well as protein biomarkers, which differ significantly from healthy subjects (Supplementary Fig. 7a). Interestingly, PDAC patients showed very high levels of H3K4me3- and H3K27me3-modified nucleosomes, compared to both groups of healthy and CRC patients (Supplementary Fig. 7a,b). Moreover, the percentage of bivalent nucleosomes marked by both of these modifications was surprisingly high, clearly differentiating these samples from plasmas obtained from CRC patients. Finally, as opposed to CRC, in PDAC we did not observe a significant change in H3K9me3-modified nucleosomes. Additional parameters that were altered in CRC compared to healthy, such as the combination of H3K9me3+H3K36me3 and the ratio of H3K9ac to H3K4me1, were not affected in patients diagnosed with PDAC (Supplementary Fig. 7c). Overall, while the number of samples analyzed is limited, the data suggests cancer-type specific alterations to various epigenetic parameters. Finally, we calculated the predictive score of each parameter alone to discriminate between the healthy, the distinct groups of CRC patients, and PDAC 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 cancer 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

200 the late-stage CRC, potentially reflecting a transition stage (Fig. 3h). While samples from
201 healthy individuals formed a tight cluster, the cancer samples showed greater variability, likely
202 due to inherent heterogeneity between tumors. CRC patients who underwent resection also
203 exhibited a high heterogeneity; interestingly, patients who received both primary tumor
204 resection and metastectomy were positioned closer to the healthy group (samples 4118, 4211,
205 4050 and 4493). Strikingly, the PDAC samples clustered separately from the CRC samples,
206 pointing to the power of EPINUC in differentiating cancer types solely based on multiplexed
207 measurements of epigenetic parameters.

208 A few CRC patients in our cohort were sampled twice along the course of the study, allowing
209 us to examine the projection of samples taken from the same individual in the PCA plot (Fig.
210 3h, marked in *, #, and †). Sample 4090 was collected two months following sample 4075 from
211 a CRC patient who underwent tumor resection and extensive treatments. Unfortunately, her
212 condition did not improve and she passed away a month later; indeed the later sample projects
213 further from healthy on both principle components. A similar trend can be seen for samples
214 3488 and 4059, taken 6.5 months apart. These results highlight the potential of EPINUC to
215 monitor patients' positive or negative response to treatment, and the power of collecting
216 multiple layers of information from each sample.

217 Finally, in order to integrate all measurements and fine-tune the discrimination between healthy
218 and CRC samples, we employed machine-learning classification (Fig. 4a, Methods). The best
219 predictive model displayed high diagnostic potential by generating a 0.96 AUC [95%
220 confidence interval (CI) 0.945 - 0.975], and sensitivity of 92% [95% CI 89.3 – 94.7] at 85%
221 specificity [95% CI 80.2 – 89.8] and 92% precision [95% CI 89.7 – 94.3], outperforming
222 predictive models relying solely on protein biomarkers, protein biomarkers coupled with DNA
223 methylation, or protein biomarkers coupled with histone PTMs (Fig. 4a and Supplementary
224 Fig. 8a). Intriguingly, this high discrimination power is achieved without including DNA
225 sequencing. This is mainly due to the combination of multiple parameters spanning various
226 cellular pathways into a single assay, and the high accuracy of the single-molecule technology
227 that allows for digital counting of molecules.

228 We hypothesized that introducing a sequencing feature for samples that were classified as
229 cancerous by the machine-learning algorithm would provide yet another layer of specificity
230 and sensitivity. As different tissues vary in their epigenetic modifications, it may allow
231 detection of the tissue-of-origin of the circulating nucleosomes, thus revealing the origin of the
232 cancer. To that end, we coupled the epigenetic analysis with single-molecule DNA

233 sequencing[21] (Fig. 4b and Supplementary Fig. 8b). Briefly, following detection of histone
234 PTMs on cfNucleosomes, the histone proteins are evicted, and the DNA is subjected to
235 repeated cycles of sequencing-by-synthesis using an automated fluidics system. Each cycle
236 consists of incorporation of A, C, T or G by DNA polymerase and imaging; following 120
237 cycles, the data is integrated to build a strand that can be aligned to the genome, corresponding
238 to the position of the modified nucleosome.

239 As proof-of-concept, we applied EPINUC followed by sequencing (EPINUC-seq) to three
240 plasma samples of late stage CRC probed for H3K4me3, H3K27me3 or H3K9ac (Fig. 4c,d and
241 Supplementary Fig. 8c-e). Single-molecule mapped reads, corresponding to modified
242 nucleosomes, were intersected with unique antibody peak signatures generated from ENCODE
243 ChIP-seq data for various tissues and primary cell lines, followed by bootstrapping simulations
244 to calculate significance (Supplementary Fig. 8c, Methods). Reinforcing our hypothesis, we
245 found that plasma samples showed significant overlap with colon-specific H3K4me3,
246 H3K27me3 and H3K9ac peaks, indicating colon as the main tissue-of-origin (Fig. 4c,d and
247 Supplementary Fig. 8c-e). Moreover, comparing our data to a recent ChIP-seq study of
248 H3K4me3 in plasma showed significant overlap with profiles obtained from CRC patients[13],
249 but not with healthy plasma (Fig. 4e). H3K27me3 mapped reads showed a broader pattern
250 compared to H3K4me3 and H3K9ac, overlapping with peaks corresponding to hematopoietic
251 lineage as well as colon. Interestingly, for CRC patients 4044 and 3821 we also observed a
252 significant overlap with liver-specific H3K27me3 and H3K9ac peaks, in agreement with
253 clinical data indicating these patients had liver metastases. This is consistent with recent studies
254 reporting liver damage and release of liver-specific DNA due to tumor cells metastasizing to
255 the liver[39,40]. Yet, the liver signal may also originate from chemotherapy-induced liver
256 damage[41]. Validating our EPINUC-seq approach for tissue-of-origin detection, analysis of
257 plasma samples from PDAC, lung and breast cancer patients revealed pancreas, lung and breast
258 tissues as the main contributors, respectively (Fig. 4f and Supplementary Fig. 8f,g).

259 Finally, we combined a complementary single-molecule approach to identify the tumor tissue-
260 of-origin, by single-molecule profiling of the DNA modification 5-Hydroxymethylcytosine
261 (5hmC). 5hmC is known to play important roles in gene regulation and cancer[42,43]. We captured
262 5hmC-enriched DNA from plasma as previously described[44], followed by single-molecule
263 DNA sequencing (Supplementary Fig. 9a, Methods). In agreement with previous reports[45], we
264 found 5hmC in cfDNA to be enriched at gene bodies and promoter proximal regions
265 (Supplementary Fig. 9b), which are also known to be marked with H3K36me3. Thus, we
266 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. 4g and Supplementary Fig. 9c). We next aimed to explore whether we can correlate the 5hmC signatures with gene expression data from primary CRC tumors. Thus, we generated a list of genes found to be associated with 5hmc signal for each group of healthy and CRC patients. Interestingly, the expression levels of these CRC-specific genes (identified based on the 5hmc signal) was significantly higher in CRC tumors versus the expression levels of the group of healthy-specific genes (Supplementary Fig. 9d). Finally, we showed correct identification of colon origin also for early stage CRC patients, and identification of pancreas origin for the patient diagnosed with PDAC (Fig. 4h and Supplementary Fig. 9e,f).

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[13,46]; (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. For quantification of global DNA methylation levels, several alternative approaches have been developed[47,48]. The methylation status of repeat elements in the genome, measured by bisulfite conversion followed by PCR, is frequently used as proxy for global genomic methylation. Yet, it is uncertain to what extent this method provides accurate representation of global DNA methylation in diverse biological and pathological conditions. Alternative approaches utilize mass spectrometry to quantify methylated versus unmethylated cytosine; while these methods are quantitative and accurate, they require relatively high amounts of DNA. Our approach complements these methodologies by providing quantitative single-molecule measurements of global DNA methylation, which is

300 independent on chemical conversion and is not limited to analysis of specific genomic regions,
301 and requires very low input material.

302 In addition to the unique epigenetic analysis, the single-molecule system outperforms the
303 classical ELISA assay for measuring protein biomarkers. ELISA is of relatively low sensitivity
304 and is therefore limited to proteins that are present at high levels, has lower dynamic range in
305 quantifying proteins, and is not amenable to multiplexed detection of several proteins[27,28]. We
306 showed that the single-molecule system is capable of detecting the mutant form of p53, which
307 is a non-secreted protein that originates directly from the tumor cells. The main advantages of
308 the single-molecule imaging system is its unique ability to follow binding of antibodies over
309 time, thus allowing for accumulation of signal. Interestingly, the accumulation kinetics for each
310 antibody is highly reproducible across different experiments, likely representing the antibody's
311 affinity and avidity to its target epitope. Importantly, the system is straightforward to adapt for
312 detection of additional proteins, thus increasing sensitivity and enabling disease-specific
313 biomarkers analysis.

314 EPINUC is built on the idea that integration of multiple parameters would provide specific
315 diagnosis, which is independent on DNA sequencing. Indeed, we show that the few pancreatic
316 cancer samples tested within this study cluster separately from CRC patients, despite the
317 heterogeneity within each group. This would potentially render the EPINUC approach fast and
318 inexpensive, paving the way to a mass screening method for a variety of cancers. Moreover,
319 the integration of multiple parameters may overcome potential confounding effects generating
320 heterogeneity in the cohort, such as chemotherapy treatments administrated to some of the
321 patients, or varied composition of blood cells between individuals. Expanding the protein
322 biomarkers panel to include cancer-type specific indicators may be instrumental in allowing
323 differential diagnosis, independent of sequencing data. Nevertheless, further studies, analyzing
324 larger cohorts of patients, are needed to test whether this technology can differentiate between
325 various cancer types based solely on the epigenetic and biomarker profiles. Such studies may
326 also provide insights on whether the epigenetic differences detected by EPINUC originate
327 solely from the tumor cells, or may represent, at least in part, epigenetic alterations occurring
328 in non-cancer cells in the tumor microenvironment, or generated by chemotherapy or other
329 treatments. Importantly, we showed that the combination of EPINUC with single-molecule
330 DNA sequencing provides unequivocal evidence of the tumor tissue-of-origin, thus further
331 elucidating the type of cancer of each individual. In future studies, it would be of special interest
332 to apply EPINUC-seq to cancer of unknown primary, potentially identifying its origin. Further
333 efforts in clinical use of EPINUC would require a system that offers streamlined sample

workflow, integrated instrumentation, and robust data processing pipelines and interpretation algorithms. EPINUC 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 thank Lior Segev for his help with writing and integrating the µManager scripts for performing EPINUC-Seq. 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), Emerson Collective, 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.

364 **Methods**

365 Patients

366 All clinical studies were approved by the local ethics committees (Helsinki applications 091-
367 2020 and 0198-14-HMO). Informed consent was obtained from all individuals before blood
368 sampling.

369 Plasma collection

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

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

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

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

386 cfDNA was extracted from 4 ml of healthy human blood plasma, or from 0.5 ml of plasma
387 from CRC patients, using the Mag-Bind cfDNA Kit (Omega Bio-Tek, M3298-01). For
388 optimized yield, protocol was modified by increasing elution time to 20 minutes on a
389 thermomixer, at 1,600 rpm, in 15 µl elution buffer at room temperature. Sample concentration
390 was measured using Qubit Fluorometer (Thermo Fisher Scientific). For fluorescent labeling of
391 plasma isolated DNA, 10 µl of cfDNA was incubated at 37°C for 1 hour with the following
392 reaction mixture: NEBuffer™ 2 (NEB, B7202), 0.25 mM MnCl2 (SIGMA, M1787), 33 µM
393 fluorescently labeled dATP (Jena Bioscience, NU-1611-Cy3), 1.5 µl of Klenow Fragment
394 (3'→5' exo-, NEB, M0212S) and 1.5 µl of T4 Polynucleotide Kinase (NEB, M0201L).
395 Following incubation, samples were inactivated by addition of EDTA (Invitrogen, 15575-038)

396 at a final concentration of 20mM. Next, DNA was purified by AMPure SPRI beads (Beckman
397 Coulter, A63881), and quantified by Qubit (Thermo Fisher Scientific).

398 Cell culture and apoptosis

399 Cell lines were maintained at 37˚C with 5% CO2. HEK-293 cells were cultured in 150 cm
400 plates ($10\times10^6$ cells in 20 ml of media) in DMEM supplemented with 10% FBS and 1% P/S,
401 and passaged every week. For induction of apoptosis, 6 µM of Staurosporine (STS, Holland-
402 Moran, 62996-74-1.25) was added to medium of confluent cells. 72 hours later, medium was
403 collected and immediately processed. To verify fragment sizes along with nucleosome labeling,
404 10 ul of the nucleosomes and 10ul of AMPure extracted DNA (either directly from
405 concentrated medium or after the tagging and tailing reaction), were loaded on High Sensitivity
406 D1000 ScreenTapes (Agilent, 5067-5584) and 6% TBE gel (ThermoFisher Scientific,
407 EC62655BOX), and imaged with 4200 TapeStation (Agilent) or Typhoon imager (Amersham
408 Biosciences), respectively. Apoptotic medium cfNucleosomes were concentrated and
409 recovered using Centricon Plus-70 centrifugation filteres (Merck, UFC710008) according to
410 the manufacture protocol. PI was supplemented 1:100 following concentration.

411 Surface preparation for single-molecule imaging

412 PEGylated-Biotin and PEGylated-poly T coated microscope slides were prepared based on the
413 protocol described by Chandradoss et al[49]. Ibidi glass coverslips (25 mm x 75 mm, IBIDI, IBD-
414 10812) were cleaned with (1) MilliQ H2O (3X washes, 5 minutes sonication, 3X washes); (2)
415 2% Alconox (SIGMA, 242985) (20 min sonication followed by 5X washes with MilliQ H2O);
416 and (3) 100% Acetone (20 min sonication followed by 3X washes with MilliQ H2O). Slides
417 were further cleaned and functionalized (Hydroxylated) by incubation in 1 M KOH (SIGMA,
418 484016) solution for 20 minutes while sonicated, followed by 3X washes with MilliQ H2O.
419 Slides were sonicated twice for 10 minutes in 100% HPLC ethanol (J.T baker 8462-25) prior
420 to applying amino-silanization chemistry. Next, slides were incubated for 24 minutes in a
421 mixture of 3% 3-Aminopropyltriethoxysilane (ACROS Organics, 430941000) and 5% acetic
422 acid in HPLC EtOH, with 1 minute sonication in the middle. Slides were then washed with
423 HPLC EtOH (3X) and MilliQ H2O (3X) and dried with N2. Surface functionalization along
424 with first passivation step was performed by applying mPEG: PEGylated-Biotin/PEG-Azide
425 solution [20 mg PEGylated-Biotin (Laysan, Biotin-PEG-SVA-5000), 180 mg mPEG (Laysan,
426 MPEG-SVA-5000) or 20 mg PEG-Azide (JenKem, A5088), 180 mg mPEG (Laysan, MPEG-
427 SVA-5000)] dissolved in 1560 ul 0.1 M Sodium Bicarbonate (SIGMA, S6297) and degassed
428 (centrifugation at 1 minute at 16,000*g*). Next, 140 µl of solution was applied on one surface,

429 followed by immediate assembly of another surface on top. Each pair of assembled surfaces
430 were incubated overnight in a dark humid environment.

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

438 For PEGylated-poly T surfaces, following PEG-Azide coating, surfaces were washed with
439 MilliQ H2O and dried with N2. To enable anchoring of dT50 to surface via click chemistry,
440 10 µM of 5'heyxynyl-dT50 (IDT) were mixed with 2 mM of CuSO4 (SIGMA, C1297) and
441 DDW. Next, 100 µl of the mixture was applied on one surface, followed by immediate
442 assembly of another surface on top. Each pair of assembled surfaces was incubated overnight
443 in a dark humid environment. In the next day, a second passivation step [MS(PEG)4] was
444 carried out, similarly to PEGylated-Biotin preparation. Surfaces were stored in -20°C post
445 nitrogen flush in a similar fashion.

446 Antibody labeling

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

452 TIMP-1 siRNA transfections

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

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

463 Synthetic DNA preparation for DNA methylation assay

464 DNA fragments were generated by conventional PCR (primer sites underlined) supplementing
465 the reaction with either methylated (NEB, N0356S) or un-methylated cytosine (Thermo Fisher
466 Scientific, R0181), followed by purification with AMPure SPRI beads. The size (~200 bp) was
467 chosen to mimic the size of mono-nucleosomal DNA fragments previously identified in blood
468 plasma[50]. Fragment labeling, purification and quantification was performed as described for
469 plasma cfDNA.

470 Sequence:
471 CATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGT
472 AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
473 AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC
474 TAACTCACA

475 5-Aza-2-Deoxycytidine Treatment

476 HEK-293 cells were plated in 150 cm plates ($10\times10^6$ cells in 20 ml of media) overnight, then
477 treated with 1 uM of 5-Aza-2-deoxycytidine (5 –Aza, SIGMA, A3656) or PBS for 4 days.
478 Next, $5\times10^6$ cells were collected and washed with PBS supplemented with PI (1:100), followed
479 by centrifugation at 3000 rpm for 3 minutes. Cell pellet was resuspended with 1 ml of 0.05%
480 Igepal (SIGMA, I8896) diluted in PBS (supplemented with PI as mentioned above) and
481 centrifuged again at 3000 rpm for 3 minutes. Next, the pellet was resuspended in Lysis buffer
482 [100 mM Tris-HCl pH 7.5 (Gibco, 115567-027), 300 mM NaCl (J.T Baker, 7647145), 2%
483 Triton® X-100 (SIGMA, 9002 93-1), 0.2% sodium deoxycholate (SIGMA, D6750), 10 mM
484 CaCl2 (SIGMA, 21115)] supplemented with PI and Micrococcal Nuclease (ThermoFisher
485 Scientific, 88216). The reaction mixture was incubated at 37°C for 10 minutes and then
486 inactivated by addition of EGTA at a final concentration of 20mM. Then, lysate was
487 centrifuged for 10 minutes at max speed and supernatant was transferred to a new tube. DNA
488 extraction, fluorescent labeling and quantification was performed as described for plasma
489 cfDNA.

490 Single-molecule imaging

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

493 which enables imaging of six different samples or various combinations of antibodies on a
494 single surface. For PEGylated-Biotin flowcells, Streptavidin (SIGMA, S4762) was added to a
495 final concentration of 0.2 mg/ml followed by 10 minutes incubation and washing with imaging
496 buffer [IMB: 12 mM HEPES pH 8 (Thermo Fisher Scientific, 15630056), 40 mM TRIS pH 7.5
497 (Gibco, 115567-027) 60 mM KCL (SIGMA, 60142), 0.32 mM EDTA (Invitrogen, 15575-038),
498 3 mM MgCl2 (SIGMA, 63069), 10% glycerol (Bio-Lab, 56815) , 0.1 mg/ml BSA (SIGMA,
499 A7906) and 0.02% Igepal (SIGMA, I8896)]. For time-lapse imaging experiments (Histone
500 PTMs, p53), prior to sample application, TetraSpeck beads (ThermoFisher Scientific, T7279)
501 diluted in PBS were added and incubated on surface for at least 10 minutes to allow correction
502 for stage drift in image analysis. Imaging was performed on a total internal reflection (TIRF)
503 microscope manufactured by Nikon (ECLIPSE Ti2-E LU-N4 TIRF) with CFI Apochromat
504 TIRF 60X Objective lens and TRF49904, TRF49909, and TRF89902 filter cubes (CHROMA)
505 for the 488, 561, and 647 lasers, respectively. Images were taken with 1.5X magnification
506 setting resulting in FOV of 148x148um, using ANDOR ZYLA 4.2 PLUS camera. At least 50
507 FOVs were imaged per lane.

508 **Histone PTMs analysis**

509 PEGylated-poly T coated coverslips were assembled as described and further passivated with
510 5% BSA (Merck, A7906) for 30 minutes followed by wash with IMB. Next, plasma sample
511 containing tailed and fluorescently labeled cfNucleosomes was incubated with antibodies
512 (diluted 1:60) for 30 minutes at room temperature (RT), to allow formation of antibody-
513 cfNucleosomes complexes. Next, samples were loaded on the surface and incubated for 15
514 minutes to allow hybridization. Flowcell was washed (X3) with IMB, followed by time lapse
515 imaging every 15 minutes, with the three laser channels, across all positions (50 Fields of View
516 (FOVs, 148$\mu m^2$) per experiment). Removing antibodies from the surface (to achieve
517 multiplexed PTM profiling) can be achieved either by multiple washes, or simply by using a
518 reducing agent (TCEP) that disrupts the disulfide bonds between the heavy and light chains of
519 antibodies but maintains nucleosomes intact.

520 **Protein biomarkers analysis**

521 Single-molecule analysis of protein biomarkers requires the use of different capture and
522 detection antibodies that would bind to different epitopes on the target protein. This would
523 allow capture of target protein to the surface, while exposing a different epitope for the binding
524 of the detection antibody. Thus, the methodology relies on the availability of two distinct

525 antibodies for each target protein. The specific antibodies we used for each protein biomarker
526 are listed in Supplementary Table 2.

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

534 **Global DNA methylation analysis**

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

540 **DNA Hydroxymethylation analysis**

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

548 <u>Single-molecule DNA sequencing</u>

549 For single-molecule DNA sequencing of cfNucleosomes, PEGylated-poly T surface was
550 blocked with BSA as described above. Poly-A tailed FluoSpheres (described below) along with
551 TetraSpeck beads and cfNucleosomes were applied to the surface. PTMs of plasma
552 cfNucleosomes were imaged over time for 169 FOV, as described above. Then flowcell was
553 washed with Wash A buffer [150 mM HEPES (KOH, pH 7.0), 1×SSC, 0.1% SDS] and Wash
554 B buffer [150 mM HEPES (KOH, pH 7.0), 150 mM NaCl] to evict histones and antibodies.
555 During single-molecule sequencing, temperature was maintained at 37°C by UNO-T stage top
556 incubator (Okolab), and Luer adapter was used to connect the flowcell to IDEX 1/4-28 flat

557 bottom fittings connecting to 1/16" OD tubing. Sequencing was performed as described
558 previously, using Helicos True Single Molecule Sequencing (tSMS) (http://seqll.com/)[21,51], by
559 repurposing a microfluidics system used in the HeliScope single molecule sequencer to deliver
560 tSMS chemistry to the flowcell (Supplementary Fig. 10). A similar setup was previously used
561 to enable single molecule sequencing[52].

562 The microfluidics system features the storage compartment for tSMS chemistry reagents,
563 connected to a set of syringe pumps and mixing valves that deliver tSMS chemistry to the
564 attached flowcell. During sequencing, imaging on microscope and chemistry on microfluidics
565 system was automatically controlled by using µManager (https://micro-manager.org/) software
566 with custom scripts to enable tSMS sequencing-by-synthesis method.

567 FluoSpheres preparation: FluoSpheres (Carboxylate-Modified Microspheres, Thermo Fisher
568 Scientific, F8789) were conjugated to dA50-amine (IDT), tailed as previously described, and
569 hybridized to the surface to serve as reference points for stage drift correction during alignment
570 of sequencing images.

571 Single-Molecule Hydroxymethylation sequencing

572 2.5 ng of plasma cfDNA was added to a 25 µl solution containing 50 mM HEPES buffer (pH
573 8), 25 mM $MgCl_2$, 60 µM UDP-6-$N_3$-Glc (Jena Bioscience, CLK-076) and 12.5 U T4 beta-
574 glucosyltransferase (Thermo Fisher Scientific, EO0831), and incubated for 2 hours at 37°C.
575 Next, 5 µl DBCO-S-S-biotin (Click Chemistry Tools, 10 mM stock in DMSO) was directly
576 added to the reaction mixture and incubated overnight at 37°C. For 5hmC DNA pulldown,
577 samples were incubated at RT with 15 µl of Streptavidin beads (Thermo Fisher Scientific, Dy-
578 11205D) for 1 hour, followed by 3 washes with 1X wash buffer, and elution in 20 µl of 125
579 mM TCEP (Thermo Fisher Scientific, TS-77720). 5hmC eluted DNA was poly-adenylated and
580 sequenced on a PEGylated-poly T surface as described above. Metagene profile was generated
581 using ngs.plot.

582 Image analysis

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

589 Predictive Power Score (PPS)

590 PPS analysis on the data was conducted using a previously published algorithm
591 (https://github.com/8080labs/ppscore). Briefly, by calculating a cross-validated decisions tree
592 (repeated [10K] 4-fold cross-validation) for the target variable (e.g., diagnosis) using only one
593 of the markers, it is possible to determine which of the markers in the datasets contributes most
594 to the target variable. The PPS is normalized to the most common assignment in order to
595 provide a baseline for comparison. Using the PPS rather than a simple correlation measure
596 allows us to account for non-linear effects and provides an alternative formulation for
597 correlation which also treats categorical variables (e.g., diagnosis, or disease state – see
598 Supplementary Table 3).

599 Machine learning model for sample classification

600 For sample binary classification, various machine-learning algorithms were trained on the
601 features that showed significant differences between healthy and CRC (Fig. 3c,d,e,
602 Supplementary Fig. 6b,c), and evaluated for their performance using a four-fold cross-
603 validation across all samples. The best predictive performance was achieved by a Logistic
604 Regression classifier. To improve classifier performance, we conducted additional feature
605 selection by training the classifier on all possible feature combinations out of the significant
606 features aforementioned. Evaluating the resulting Area Under the Curve (AUC) values of
607 repeated (500 iterations) four-fold cross-validation for each combination revealed an optimal
608 cumulative performance of a five feature combination: H3K27me3/Nuc, H3K9me3/Nuc,
609 CEA, CEA/MST1, and global DNA methylation. To evaluate the classifier overall
610 performance using the selected features, we performed repeated (10K) 4-fold cross-validation
611 across all samples. For each iteration the sensitivity, specificity, accuracy, precision, negative
612 predictive value and the AUC value were calculated and averaged over all iterations. R caret
613 and RWeka packages were used for machine-learning modeling.

614 Tissue and plasma signatures

615 We downloaded and combined two independent *Homo sapiens* based ChIP-seq tracks for each
616 tissue from the Encyclopedia of DNA Elements (ENCODE, **Supplementary Table 4**). To
617 generate a unique antibody peak profile for a given tissue, we discarded peaks found to overlap
618 with at least one of the other eight tissues tested, retaining only tissue specific peaks. Of note,
619 brain H3K36me3 peaks were available only for embryonic tissue, therefore were replaced with
620 spleen tissue H3K36me3 peaks. Similarly, heart H3K9ac peaks were available only for
621 embryonic tissue, therefore were replaced with skeletal muscle tissue H3K9ac peaks. To

622 generate unique plasma H3K4me3 ChIP-seq peaks, we obtained data from healthy (H) and
623 CRC (C) plasma (n=3 for each) produced by Sadeh et. al[13] (Supplementary Table 4). For each
624 group, reads were intersected and only shared reads across all samples were kept for further
625 analysis. Non-overlapping reads between the overlapping healthy and overlapping CRC reads
626 were defined as the unique plasma signature.

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

641 5hmC gene signature expression in CRC

642 To explore the correlation between 5hmC signatures and gene expression in CRC, we
643 generated annotated gene lists for healthy and CRC plasma (n=3 for each) enriched for 5hmC
644 DNA sequenced reads via Clusterpofiler R package. Next, for each group, genes were
645 intersected and only shared genes across all samples within each group were kept for further
646 analysis. Then, in order to generate unique 5hmC gene signature for each group, both datasets
647 were intersected and overlapping genes were discarded. Each unique 5hmC gene signature was
648 intersected with RNA expression datasets of CRC primary tumors, obtained from the
649 cBioportal database (TCGA, Firehose Legacy). Finally, both unique 5hmC gene signatures
650 were compared by their logarithmic mean expression levels (two samples, Welch's t-test).

651 Statistical analysis

652 All statistical analysis was conducted using the statistical programming language R. Multiple
653 comparison (Supplementary Fig. 6d) was calculated using Asymptotic K-Sample Brown-Mood
654 Median Test.

655 Data availability

656 Datasets generated and analyzed during this study are summarized in **Supplementary Table**
657 **4**, BED files of plasma sequenced reads are available upon request.

658 Code availability

659 Code for performing overlap analysis is available at https://github.com/Vadim-Fed/EPINUC-
660 overlap.

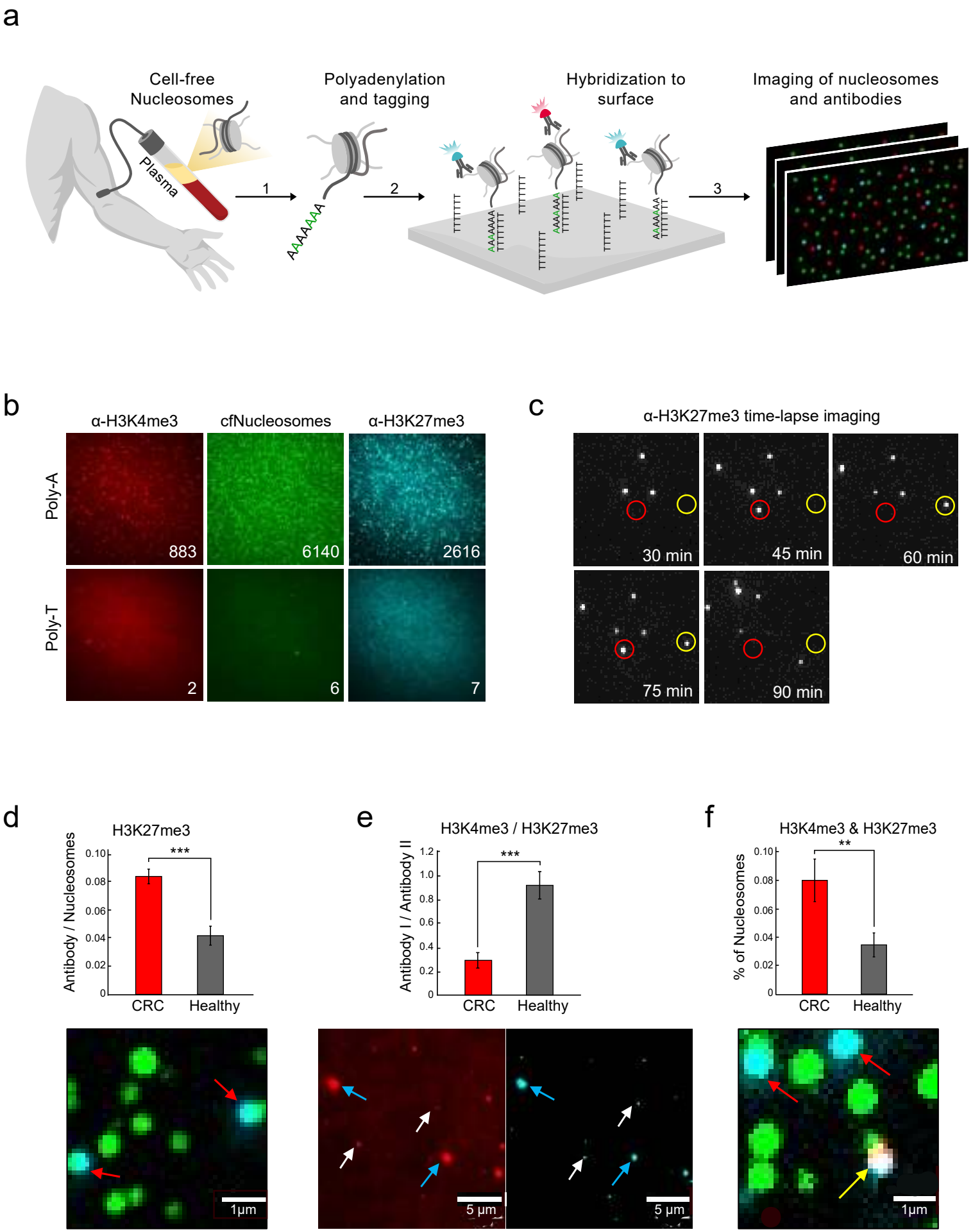
661

662

663

664

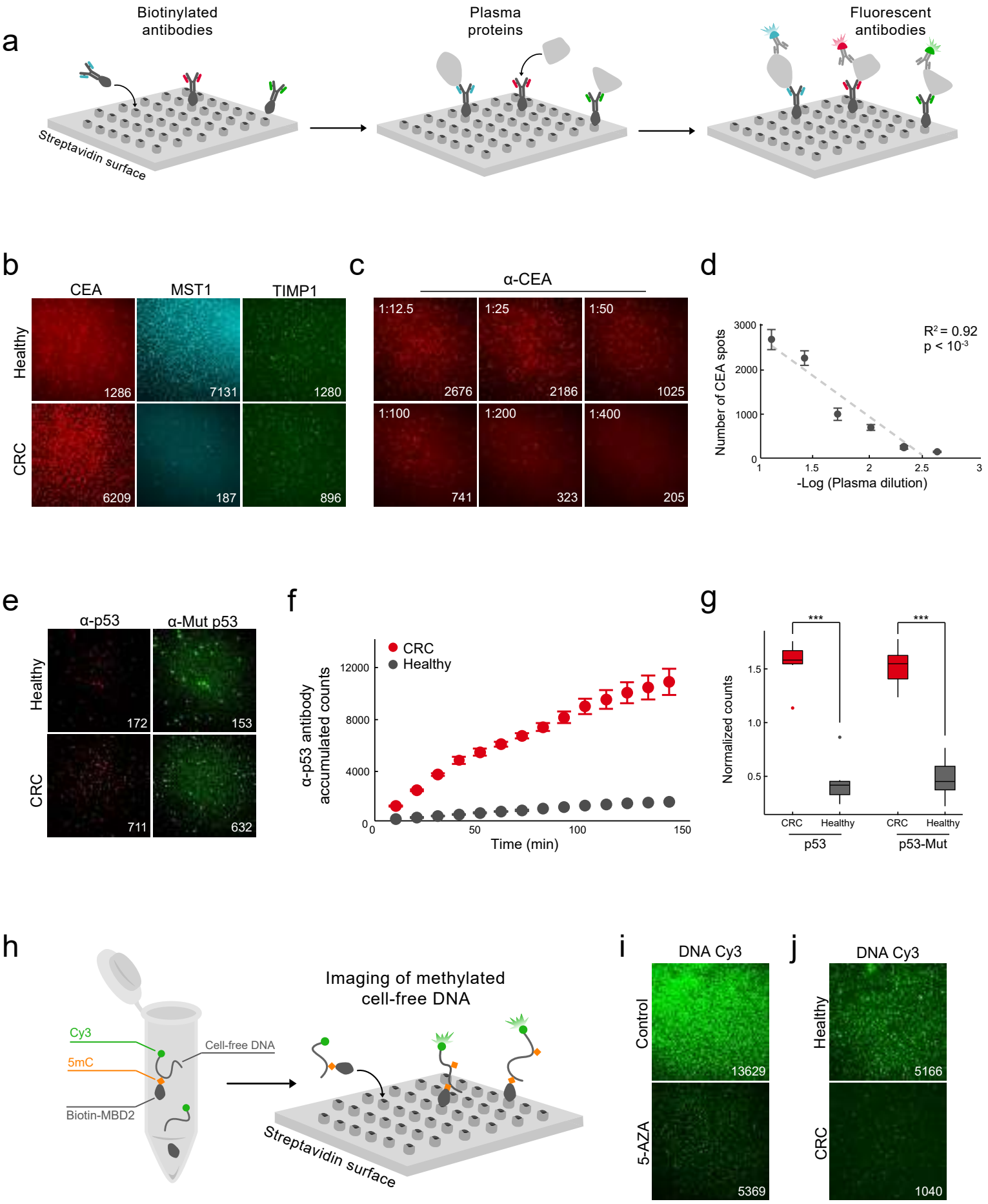
665


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

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

668 **a,** Experimental scheme: (1) Sample preparation procedure of cfNucleosomes is carried out in
669 one-step and consists of two enzymatic processes: repair of DNA ends by Klenow polymerase,
670 and addition of a poly A tail by Terminal Transferase (TdT). The reaction contains a mixture
671 of natural dATPs and fluorescently labeled dATPs (Cy3-dATP) to label nucleosomes. (2)
672 cfNucleosmes are captured on a PEGylated-poly T surface via dA:dT hybridization.
673 Immobilized nucleosomes are incubated with fluorescently labeled antibodies targeting
674 different histone modifications. (3) TIRF microscopy is applied to record nucleosome positions
675 and generate time-lapse imaging of antibodies' binding events. **b,** Representative images of
676 plasma-derived cfNucleosomes and the corresponding H3K4me3 and H3K27me3 signal
677 (numbers within images represent counted spots). Each spot corresponds to a single
678 nucleosome. Nucleosomes anchor to the surface specifically via hybridization, as evident from
679 the lack of signal when tailed with dTTP (Poly-dT) rather than dATP (Poly-dA). **c,**
680 Representative images of antibodies' binding and dissociation events over time from individual
681 target molecules (marked by red/yellow circles). **d,e,f** Example for quantification and
682 representative images of the various parameters measured by EPINUC in plasma samples from
683 one healthy subject and one CRC patient. Zoomed-in image segments of entire field of view
684 (148µm$^2$). **d,** The percentage of cfNucleosomes (green spots, Cy3) that are modified by
685 H3K27me3 (cyan spots, AF488). Red arrows indicate co-localization events. Scale bar = 1µm
686 **e,** Ratio between H3K4me3 (red, AF647) and H3K27me3 (cyan) antibodies. White arrows
687 indicate antibody spots, blue arrows indicate TetraSpeck beads that are used for alignment.
688 Scale bar = 5µm. **f,** cfNucleosomes marked by the combinatorial pattern of both H3K27me3
689 and H3K4me3. Red arrows indicate co-localization events of H3K27me3 only, yellow arrow
690 indicates a combinatorially modified nucleosome. Scale bar = 1µm.

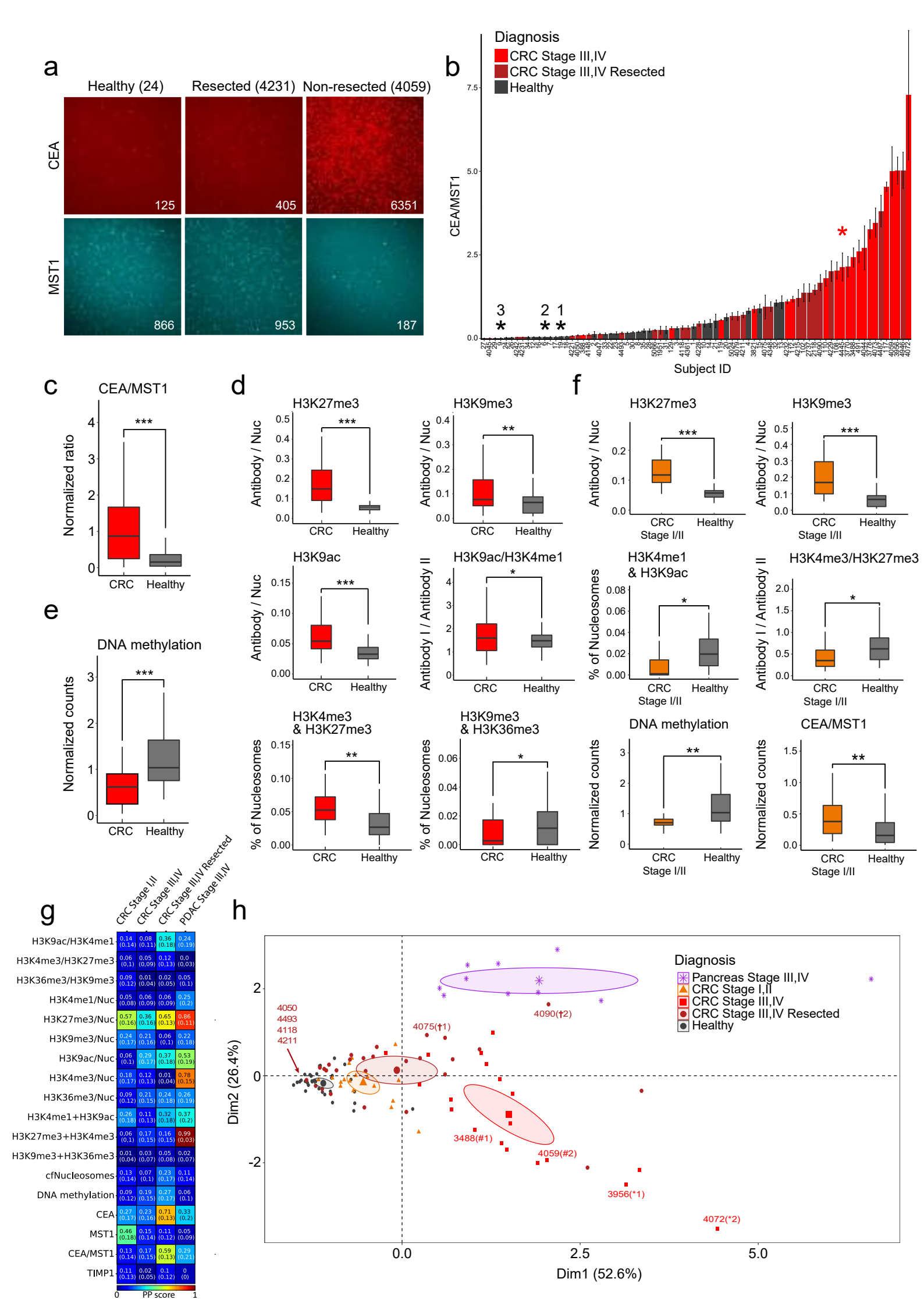
691


a
Biotinylated antibodies
Plasma proteins
Fluorescent antibodies
Streptavidin surface
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

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

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

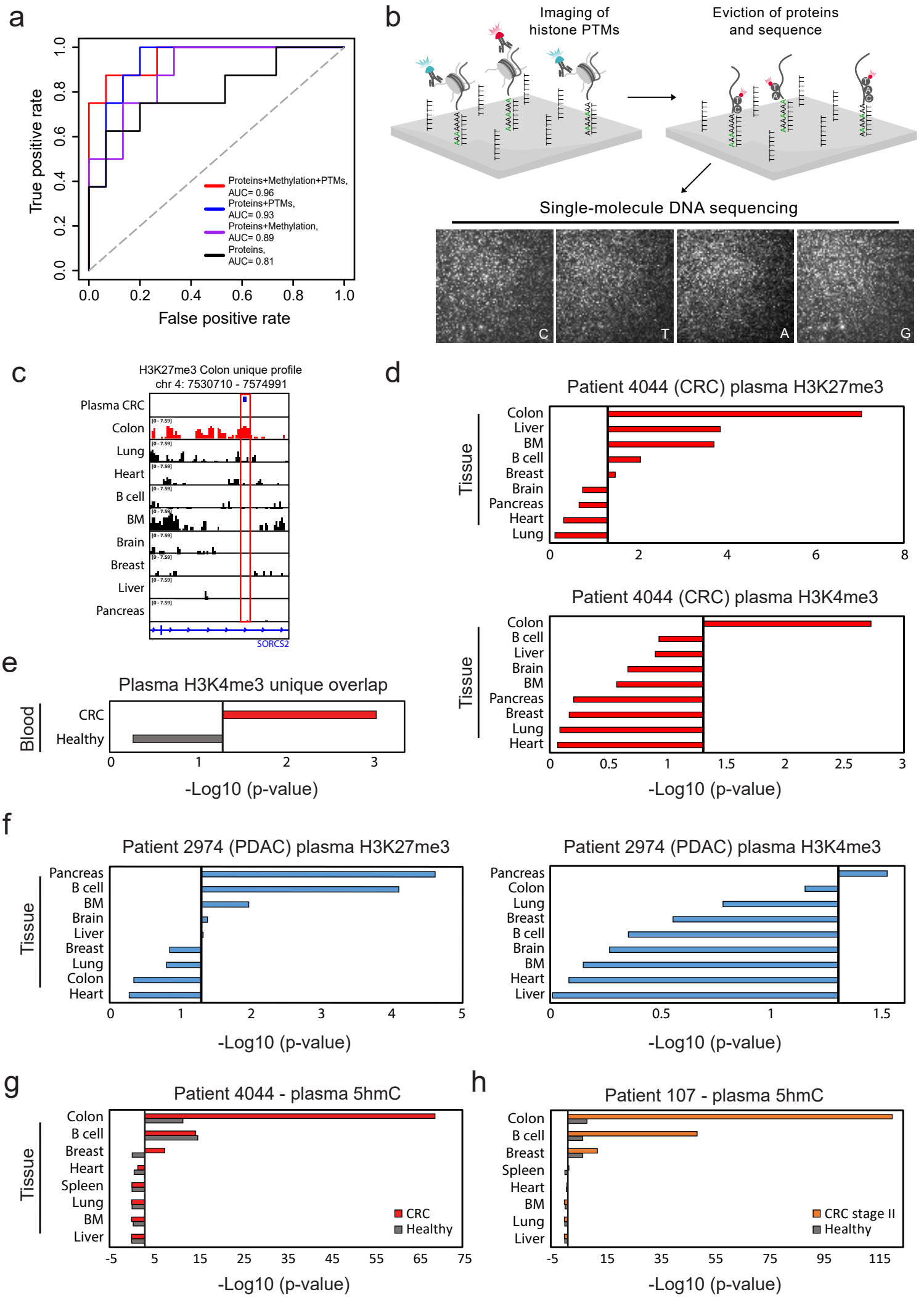
724


a
Healthy (24)
Resected (4231)
Non-resected (4059)
CEA
MST1
125
405
6351
866
953
187
b
Diagnosis
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
DNA methylation
Normalized counts
f
H3K4me1 & H3K9ac
H3K4me3/H3K27me3
Stage I/II
g
CRC Stage I,II
CRC Stage III,IV
CRC Stage III,IV Resected
PDAC Stage III,IV
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
Pancreas Stage III,IV
CRC Stage I,II
CRC Stage III,IV
CRC Stage III,IV Resected
Healthy
Dim1 (52.6%)
Dim2 (26.4%)
4050
4493
4118
4211
4075(†1)
4090(†2)
3488(#1)
4059(#2)
3956(*1)
4072(*2)

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

727 **a,** Representative TIRF images depict changes in protein biomarker levels in the plasma of
728 healthy, CRC patient, and CRC following tumor resection. Numbers within images represent
729 counted spots. **b,** CEA/MST1 normalized levels in the plasma of CRC patients and healthy
730 individuals. Each bar represents a subject; data is presented as the mean +/- s.d. of 50 fields of
731 view per sample. *1-3 correspond to healthy samples 19, 6, and 9, respectively. Sample 4445
732 (CRC, red) is denoted by *. **c,** Box plot representation of the data in B (healthy = 33, CRC =
733 46). Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest
734 (smallest) value no further than 1.5× interquartile range from the hinge. P values were
735 calculated by Welch's t-test. *** P value < 0.001. **d,** Histone PTMs, ratios and combinations
736 (as indicated on the graphs) that significantly differ between healthy and CRC late stage
737 samples (healthy = 33, CRC = 46). P values were calculated by Welch's t-test. * P value < 0.05
738 ** P value < 0.01. *** P value < 0.001. **e,** Global DNA methylation levels, measured as in Fig.
739 2J, in the same cohort as (d). **f,** EPINUC measurements (Histone PTMs, DNA methylation and
740 protein biomarkers) that significantly differ between healthy and early stage CRC patients
741 (healthy = 33, early CRC = 17). P values were calculated by Wilcoxon rank sum exact test. *
742 P value < 0.05 ** P value < 0.01. *** P value < 0.001. **g,** Individual parameters predictive
743 power score (PPS) analysis for the various subgroups (see Methods). Color scale represents
744 PPS value. **h,** Principal Component Analysis (PCA) with input parameters of H3K27me3/Nuc,
745 H3K4me3 & H3K27me3, CEA/MST1 and CEA. Sample groups are color-coded as indicated,
746 each dot represents a plasma sample. Ellipse represents 95% confidence interval for the
747 barycenter of each group.

748


a
True positive rate
False positive rate
Proteins+Methylation+PTMs, AUC= 0.96
Proteins+PTMs, AUC= 0.93
Proteins+Methylation, AUC= 0.89
Proteins, AUC= 0.81
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
Pancreas
SORCS2
d
Patient 4044 (CRC) plasma H3K27me3
Tissue
Colon
Liver
BM
B cell
Breast
Brain
Pancreas
Heart
Lung
Patient 4044 (CRC) plasma H3K4me3
-Log10 (p-value)
e
Plasma H3K4me3 unique overlap
Blood
CRC
Healthy
-Log10 (p-value)
f
Patient 2974 (PDAC) plasma H3K27me3
Patient 2974 (PDAC) plasma H3K4me3
Tissue
Pancreas
B cell
BM
Brain
Liver
Breast
Lung
Colon
Heart
-Log10 (p-value)
g
Patient 4044 - plasma 5hmC
Tissue
Colon
B cell
Breast
Heart
Spleen
Lung
BM
Liver
CRC
Healthy
-Log10 (p-value)
h
Patient 107 - plasma 5hmC
Colon
B cell
Breast
Spleen
Heart
BM
Lung
Liver
CRC stage II
Healthy
-Log10 (p-value)

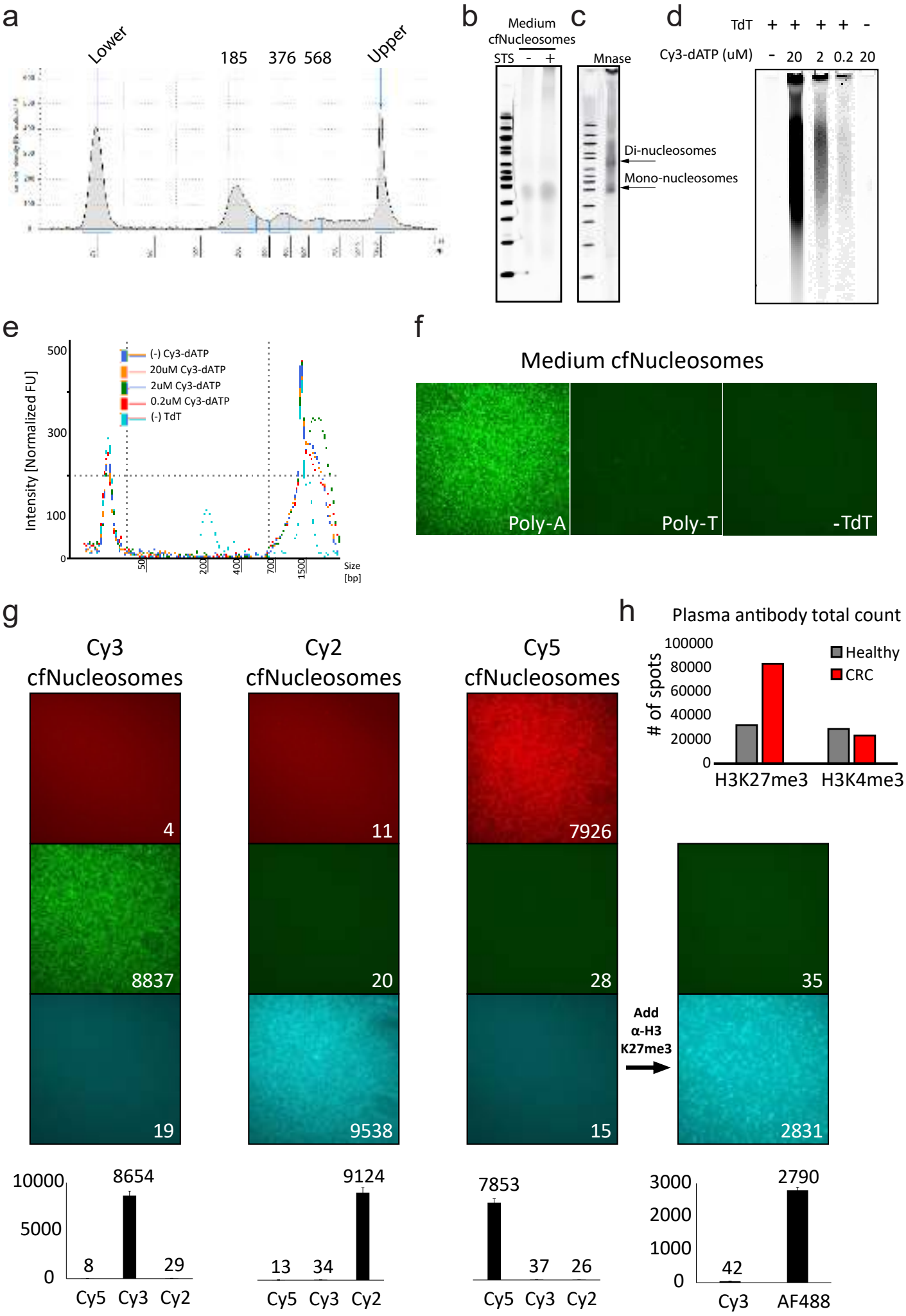
749 **Fig.4: EPINUC differentiates healthy versus CRC patients and informs the tumor tissue-**
750 **of-origin.**

751 **a,** ROC curves discriminate between healthy (n=33) and all CRC samples (n=63) using a
752 logistic regression model. The area under the curve (AUC) is shown for protein biomarkers
753 only (Proteins; CEA + CEA/MST1, black line), protein biomarkers in combination with global
754 DNA methylation levels (Proteins + Methylation, purple line), protein biomarkers with histone
755 PTMs (Proteins + PTMs, blue line) and the complete dataset generated by EPINUC (Proteins
756 + Methylation + PTMs, Red line). Gray diagonal line indicates expected curve for random
757 classification. **b,** Experimental scheme for EPINUC-seq. Histone PTMs are decoded as shown
758 in Fig. 1A. Next, histones are evicted by increasing salt concentration, retaining DNA strands
759 at identical positions. Single-molecule DNA sequencing-by-synthesis is performed by cycles
760 of incorporation of fluorescently labeled nucleotides and TIRF imaging[21]. Images represent
761 four sequencing cycles, showing incorporation of cytosine (C), thymine (T), adenosine (A) and
762 Guanine (G). For each x,y coordinate on the surface, sequence data is analyzed and integrated
763 with the initial images that registered histone PTMs, revealing the modification state of the
764 corresponding nucleosome. **c,** Single-molecule reads of H3K27me3 (blue) from a CRC patient
765 overlap with ChIP–seq profile of H3K27me3 in the colon, but not the other indicated tissues.
766 BM=bone marrow. **d,** EPINUC-seq analysis of plasma from stage IV CRC patient (patient
767 4044). Tissues and primary cell lines ranked by overlap significance with single-molecule
768 plasma H3K4me3 (top) or H3K27me3 (bottom) positive reads. Black line corresponds to P
769 value of 0.05. P values were determined by Z-test. **e,** Overlap significance of H3K4me3 single-
770 molecule reads with ChIP-seq data[13] performed in the plasma of healthy and CRC patients. **f**,
771 EPINUC-seq analysis of plasma from stage IV PDAC patient (patient 2974). Tissues and
772 primary cell lines ranked by overlap significance with single-molecule H3K27me3 (Left) or
773 H3K4me3 (Right) positive reads. **g,h,** Overlap significance of tissues and primary cell lines
774 unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus late stage
775 CRC (**g**) or early stage CRC (**h**), calculated as in (e). Black line corresponds to P value of 0.05.
776 P values were determined by Z-test.

777

778

779

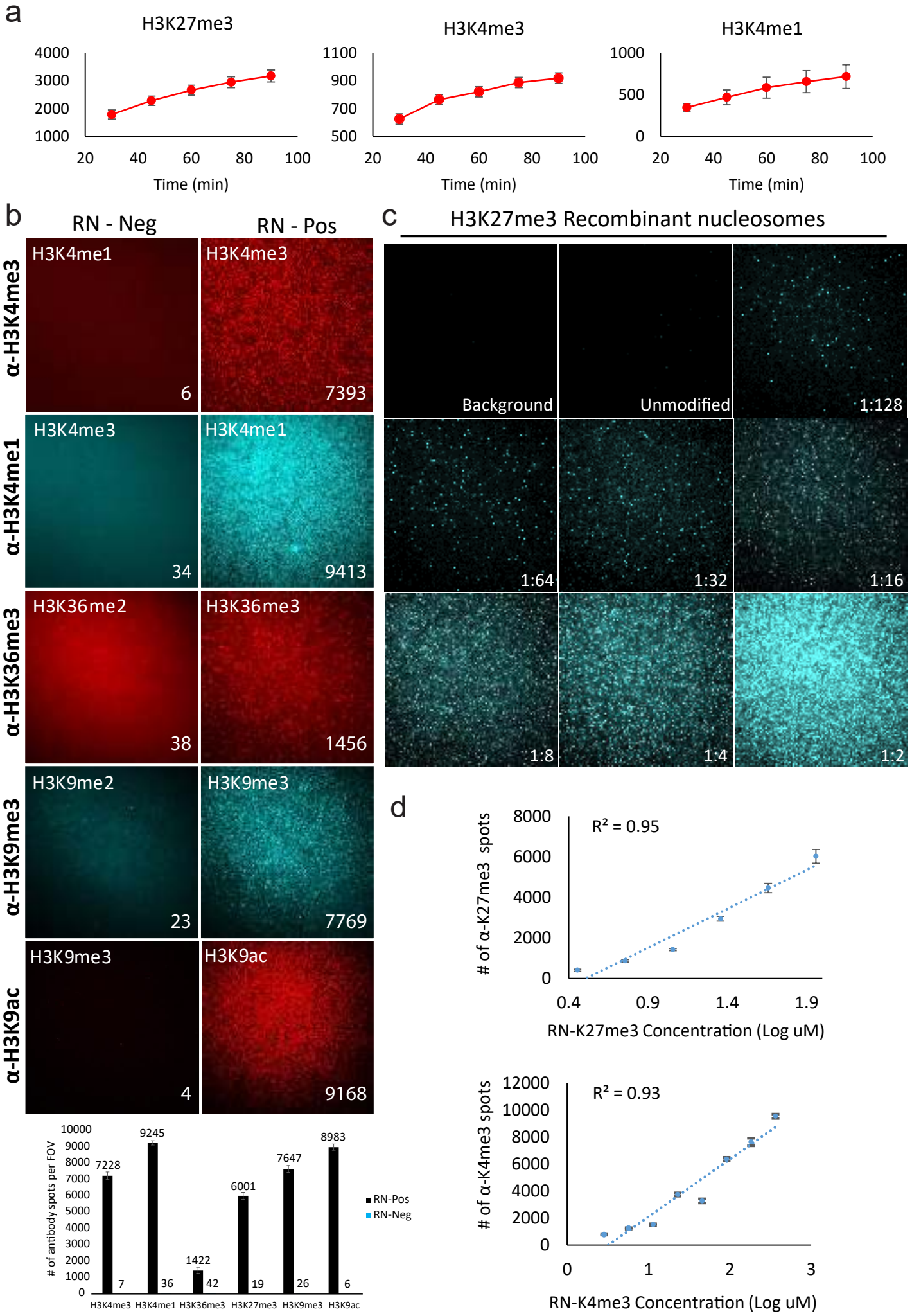

a
Lower
185
376
568
Upper
b
Medium cfNucleosomes
STS - +
c
Mnase
Di-nucleosomes
Mono-nucleosomes
d
TdT + + + + -
Cy3-dATP (uM) - 20 2 0.2 20
e
Intensity [Normalized FU]
500
300
100
0
(-) Cy3-dATP
20uM Cy3-dATP
2uM Cy3-dATP
0.2uM Cy3-dATP
(-) TdT
50
200
400
700
1500
Size [bp]
f
Medium cfNucleosomes
Poly-A
Poly-T
-TdT
g
Cy3 cfNucleosomes
Cy2 cfNucleosomes
Cy5 cfNucleosomes
4
8837
19
11
20
9538
7926
28
15
Add α-H3 K27me3
35
2831
10000
5000
0
8
8654
29
Cy5 Cy3 Cy2
13
34
9124
Cy5 Cy3 Cy2
7853
37
26
Cy5 Cy3 Cy2
3000
2000
1000
0
42
2790
Cy3 AF488
h
Plasma antibody total count
# of spots
100000
80000
60000
40000
20000
0
Healthy
CRC
H3K27me3
H3K4me3

780 **Supplementary Fig.1**

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

782 **(a)** TapeStation fragment size analysis of DNA isolated from medium of HEK293 cells treated
783 with the apoptosis-inducing factor Staurosporine (STS). DNA exhibits canonical apoptotic
784 DNA fragmentation pattern. **(b)** Nucleosomes from medium of HEK293 cells (described in a)
785 treated with STS (+) or with PBS (-) were resolved on a 6% TBE gel and visualized by Typhoon
786 laser scanner. **(c)** Nucleosomes were extracted from HEK293 cells by digestion with MNase[21],
787 resolved on a 6% TBE gel and visualized by Typhoon laser scanner. Bands correspond to
788 mono-nucleosomes and di-nucleosomes. **(d)** Nucleosomes from medium of HEK293 cells
789 (described in a) were subjected to tagging and tailing reaction (see Methods), resolved on a 6%
790 TBE gel and visualized by Typhoon laser scanner. Image depicts fluorescent labeling via
791 incorporation of Cy3-dATP during polyadenylation reaction. Incorporation is dependent on the
792 presence of TdT in the reaction, and level of florescent signal correlates with the concentration
793 of Cy3-dATP (Left). **(e)** TapeStation fragment size analysis of the samples in (d). **(f)** Poly-dA
794 tailed and Cy3-dATP labeled cfNucleosomes show specific anchoring to PEGylated poly-dT
795 surfaces through A:T hybridization. **(g)** Images and quantification of poly-dA tailed
796 cfNucleosomes, labeled with either Cy2-dATP, Cy3-dATP or Cy5-dATP. Data indicates very
797 low crosstalk between the three fluorescent channels. Further incubation of Cy-5-labeled
798 cfNucleosomes with α-H3K27me3- AF488 antibody reinforces low spectral overlap. Data is
799 presented as the mean +/- s.d. of 50 FOVs for each channel. **(h)** Total counts for H3K27me3
800 and H3K4me3 in the plasma samples analyzed in Fig. 1d,e.

801

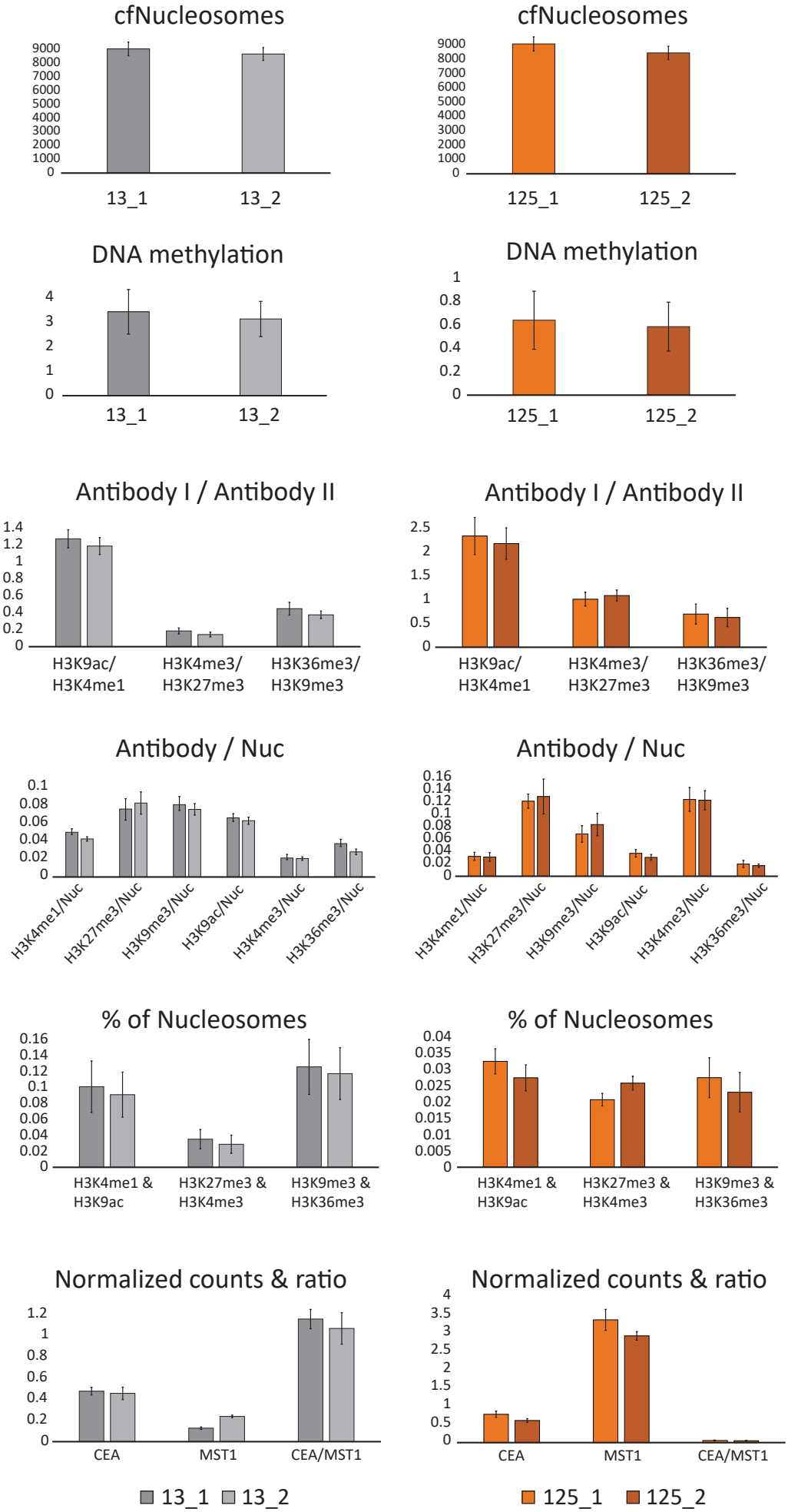

a
H3K27me3
4000
3000
2000
1000
20
40
60
80
100
Time (min)
H3K4me3
1100
900
700
500
20
40
60
80
100
Time (min)
H3K4me1
1000
500
0
20
40
60
80
100
Time (min)
b
RN - Neg
RN - Pos
α-H3K4me3
H3K4me1
6
H3K4me3
7393
α-H3K4me1
H3K4me3
34
H3K4me1
9413
α-H3K36me3
H3K36me2
38
H3K36me3
1456
α-H3K9me3
H3K9me2
23
H3K9me3
7769
α-H3K9ac
H3K9me3
4
H3K9ac
9168
# of antibody spots per FOV
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
7228
7
9245
36
1422
42
6001
19
7647
26
8983
6
H3K4me3
H3K4me1
H3K36me3
H3K27me3
H3K9me3
H3K9ac
RN-Pos
RN-Neg
c
H3K27me3 Recombinant nucleosomes
Background
Unmodified
1:128
1:64
1:32
1:16
1:8
1:4
1:2
d
# of α-K27me3 spots
8000
6000
4000
2000
0
R² = 0.95
0.4
0.9
1.4
1.9
RN-K27me3 Concentration (Log uM)
# of α-K4me3 spots
12000
10000
8000
6000
4000
2000
0
R² = 0.93
0
1
2
3
RN-K4me3 Concentration (Log uM)

802 **Supplementary Fig.2**

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

804 **(a)** Accumulation of unique antibody binding events over time. Data is presented as the mean
805 +/- standard deviation (s.d.) of 50 FOVs for each time point. **(b)** Top: Representative TIRF
806 images demonstrate antibodies binding to various recombinant nucleosomes (RN) carrying
807 different histone PTMs. Left images: negative control, showing very low un-specific binding
808 of fluorescent antibodies to recombinant nucleosomes that do not carry the target PTM. Right:
809 binding of antibodies to recombinant nucleosomes that carry the target modification. The type
810 of modification is indicated at the bottom of the images. Numbers within images represent
811 counted spots. Bottom: Quantification of the averaged counts of the indicated antibodies per
812 50 FOVs. **(c)** Representative TIRF images depicting binding of the antibody targeting
813 H3K27me3 to an empty surface (Background), unmodified recombinant nucleosomes
814 (Unmodified), and H3K27me3-modified recombinant nucleosomes diluted as indicated. **(d)**
815 Regression analysis of number of spots per FOV as a function of target recombinant
816 nucleosomes' concentration demonstrates low variability along the regression line, supporting
817 linearity of binding. 1:2 dilution was excluded from analysis due to high density. Data is
818 presented as the mean +/- s.d. of 50 FOVs for each concentration.

819

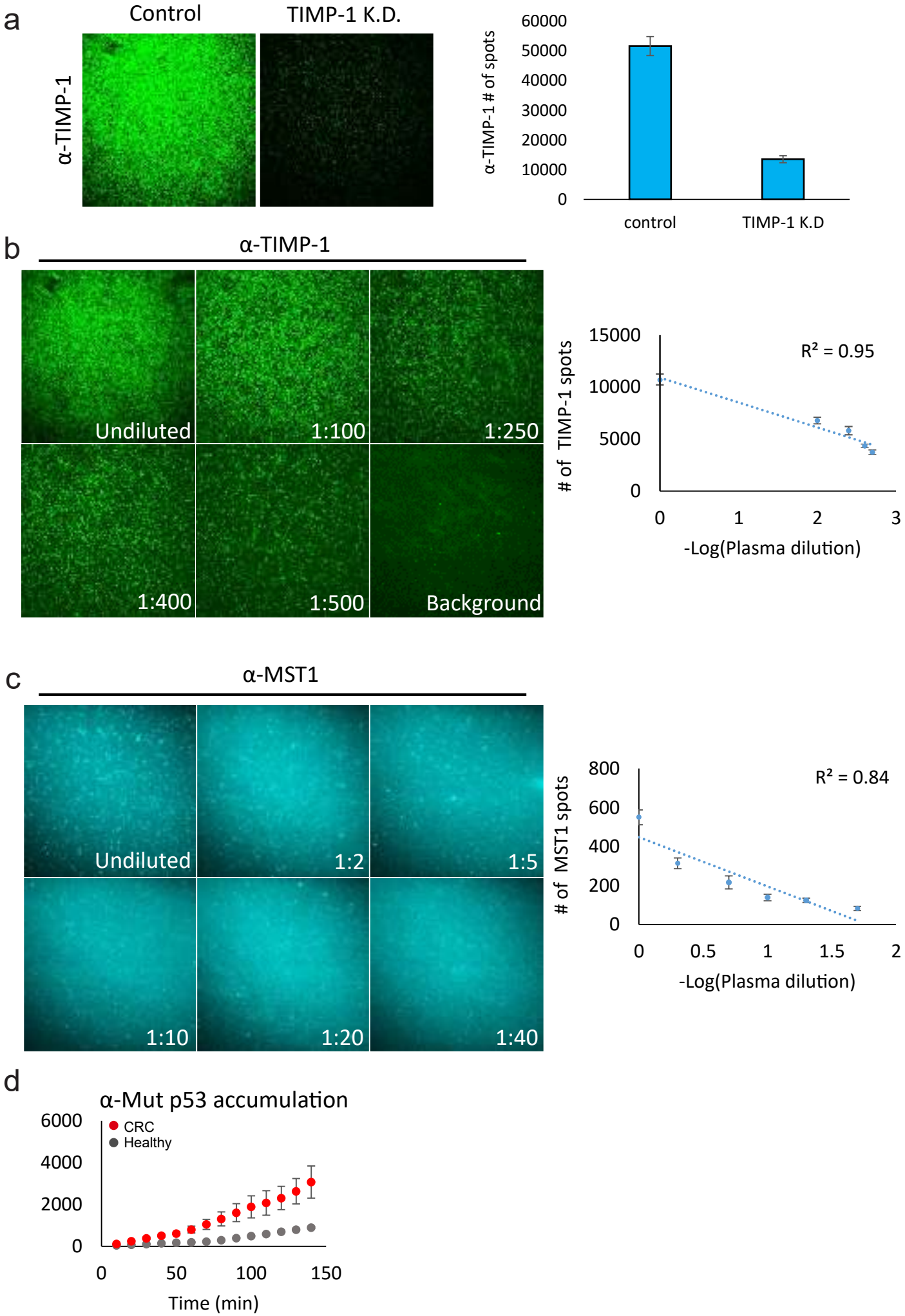

cfNucleosomes
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
13_1
13_2
cfNucleosomes
125_1
125_2
DNA methylation
4
3
2
1
0
13_1
13_2
DNA methylation
1
0.8
0.6
0.4
0.2
0
125_1
125_2
Antibody I / Antibody II
1.4
1.2
1
0.8
0.6
0.4
0.2
0
H3K9ac/ H3K4me1
H3K4me3/ H3K27me3
H3K36me3/ H3K9me3
Antibody I / Antibody II
2.5
2
1.5
1
0.5
0
H3K9ac/ H3K4me1
H3K4me3/ H3K27me3
H3K36me3/ H3K9me3
Antibody / Nuc
0.1
0.08
0.06
0.04
0.02
0
H3K4me1/Nuc
H3K27me3/Nuc
H3K9me3/Nuc
H3K9ac/Nuc
H3K4me3/Nuc
H3K36me3/Nuc
Antibody / Nuc
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
H3K4me1/Nuc
H3K27me3/Nuc
H3K9me3/Nuc
H3K9ac/Nuc
H3K4me3/Nuc
H3K36me3/Nuc
% of Nucleosomes
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
H3K4me1 & H3K9ac
H3K27me3 & H3K4me3
H3K9me3 & H3K36me3
% of Nucleosomes
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
H3K4me1 & H3K9ac
H3K27me3 & H3K4me3
H3K9me3 & H3K36me3
Normalized counts & ratio
1.2
1
0.8
0.6
0.4
0.2
0
CEA
MST1
CEA/MST1
13_1
13_2
Normalized counts & ratio
4
3.5
3
2.5
2
1.5
1
0.5
0
CEA
MST1
CEA/MST1
125_1
125_2

820 **Supplementary Fig.3**

821 **Reproducibility of EPINUC measurements.**

822 Technical repeats (n=2) of all EPINUC measurements were conducted for healthy (subject 13)
823 and early stage (subject 125) plasma samples. Results indicate high reproducibility of
824 EPINUC's measured parameters, with low variation between repetitions. Of note, these
825 technical repeats were carried out six months apart, using different batches of surfaces and
826 different aliquots of samples. Data is presented as the mean +/- s.d. of 50 FOV for each
827 repetition.

828


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
1:5
1:10
1:20
1:40
# 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)

829 **Supplementary Fig.4**

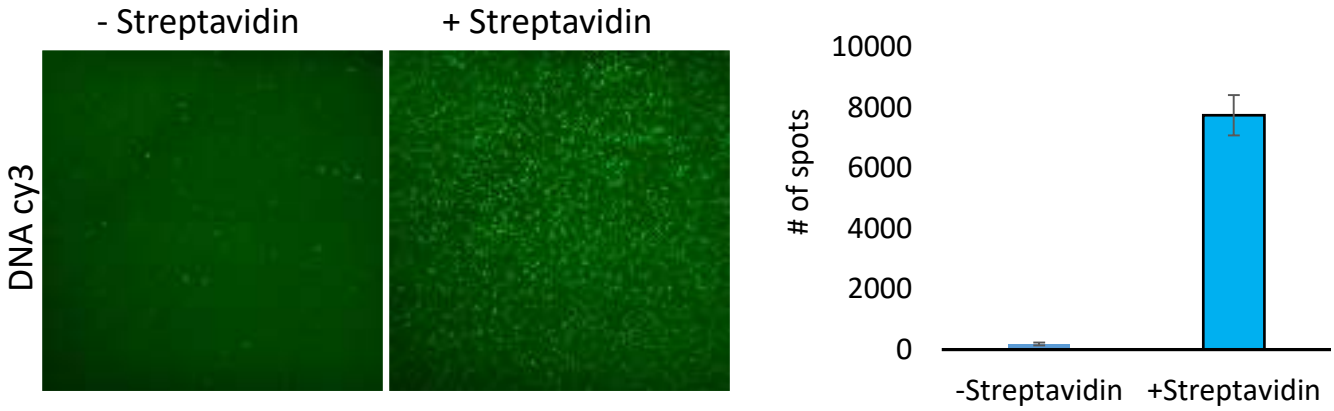
830 **Single-molecule imaging of MST1, TIMP-1 and mutant p53.**

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

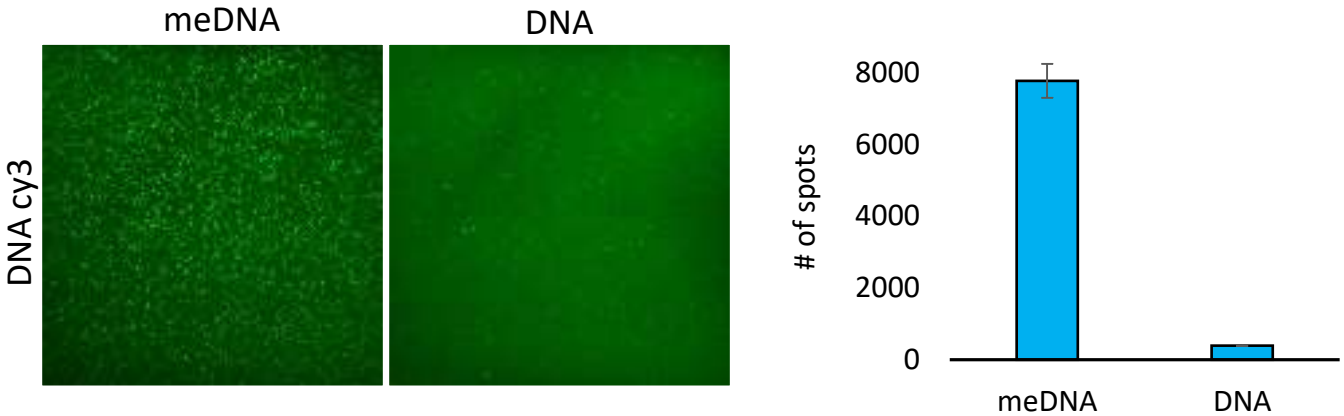
839

840

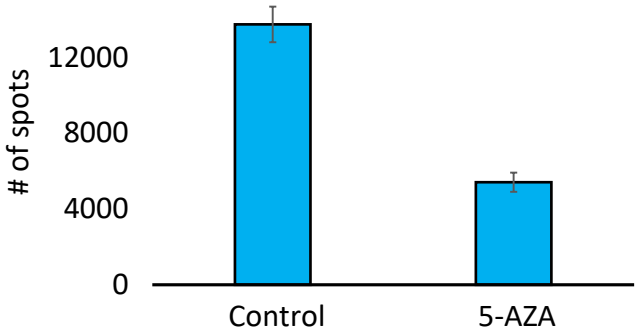
a


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

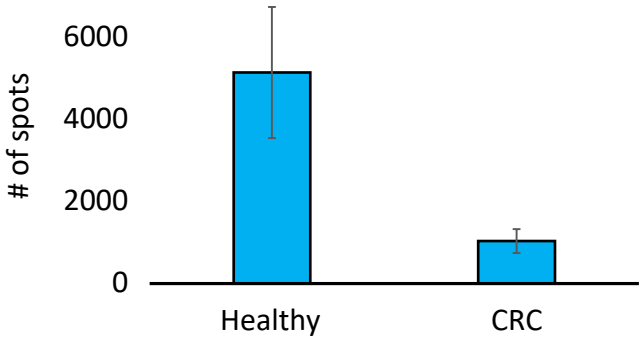
b


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

c


# of spots
12000
8000
4000
0
Control
5-AZA

d

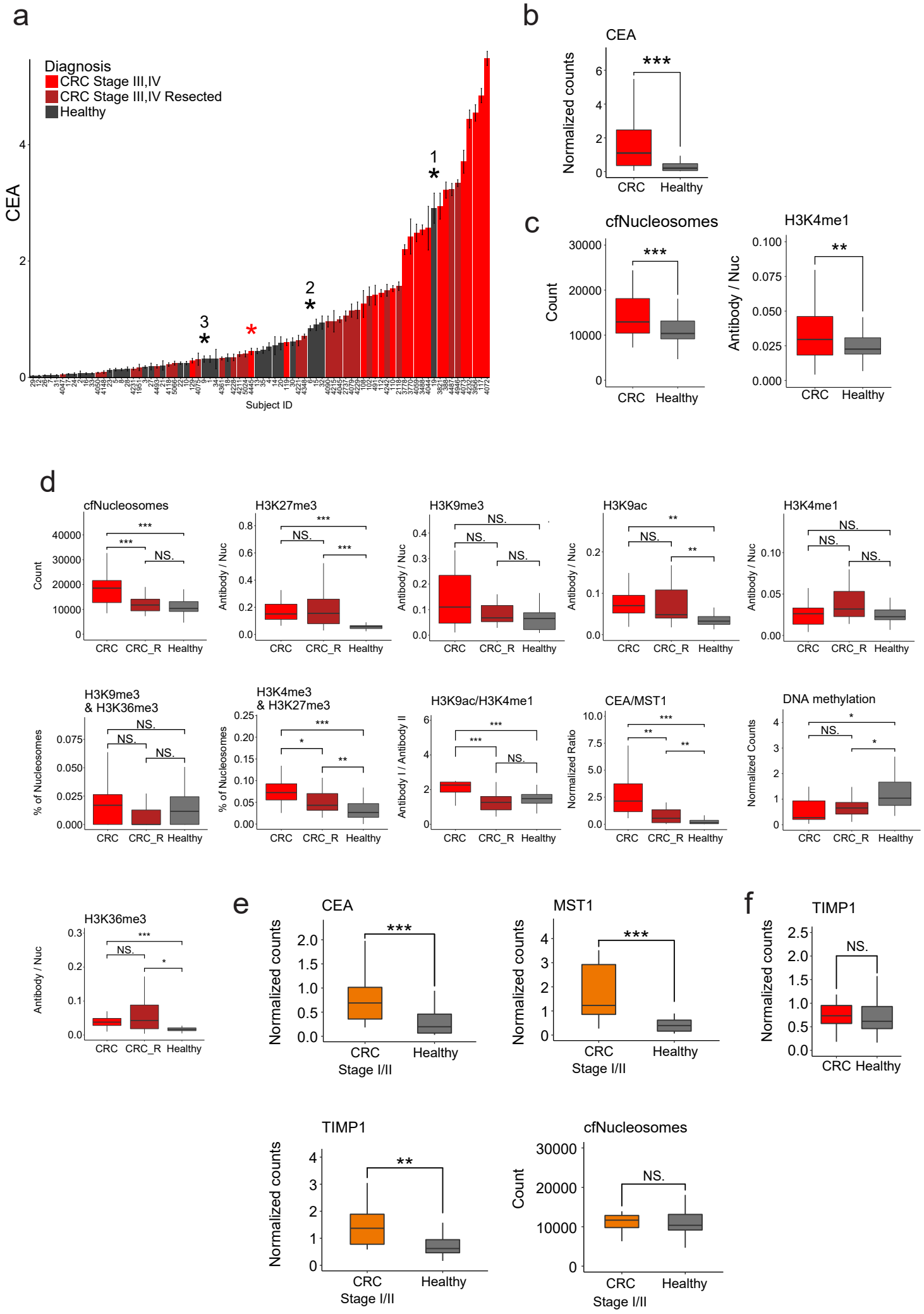

# of spots
6000
4000
2000
0
Healthy
CRC

841 **Supplementary Fig.5**

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

843 **(a)** Representative TIRF images (Left) and quantification (Right) indicating specific anchoring
844 of Biotin-MBD-meDNA complexes to streptavidin-coated surfaces. Data is presented as the
845 mean +/- s.d. of 50 FOV for each treatment. **(b)** Biotin-MBD specifically binds methylated
846 DNA. Representative TIRF images (Left) and quantification (Right) of global DNA
847 methylation levels following incubation of MBD2-biotin with Cy3-labeled (green) methylated
848 (meDNA) or un-methylated synthetic DNA fragments. Data is presented as the mean +/- s.d.
849 of 50 FOV for each treatment. **(c, d)** Quantification of global DNA methylation levels
850 presented in Fig. 2i **(c)** and Fig. 2j **(d)**. Data is presented as the mean +/- s.d. of 50 FOV for
851 each sample.

852


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

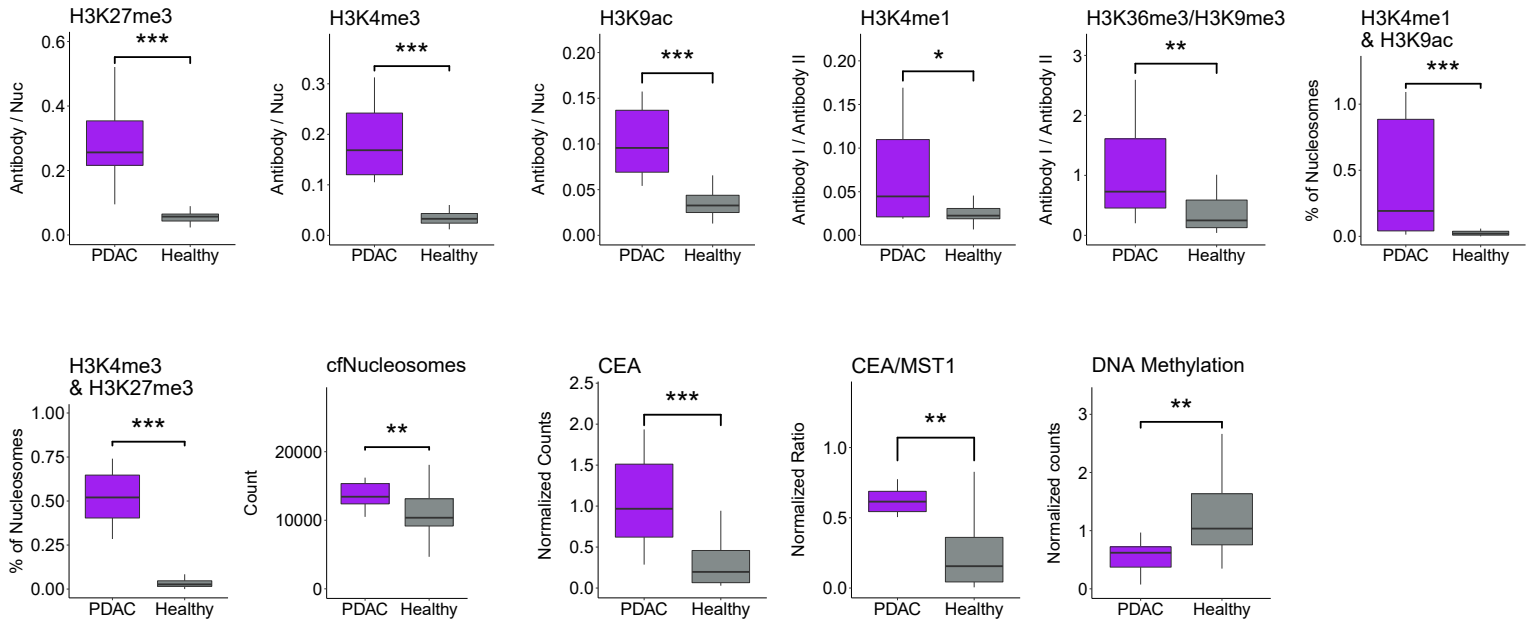
853 **Supplementary Fig.6**

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

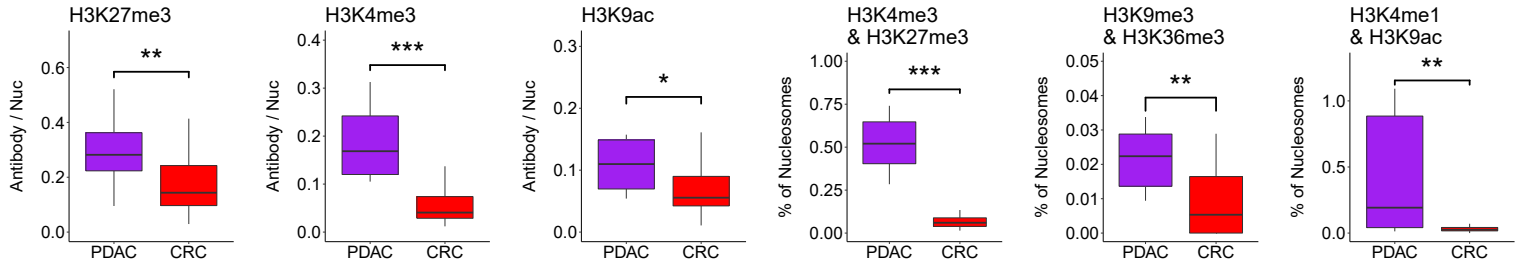
856 **(a)** CEA normalized levels in the plasma of CRC patients and healthy individuals. Each bar
857 represents a subject, data is presented as the mean +/- s.d. of 50 FOVs per sample. *1-3
858 correspond to healthy samples 19, 6, and 9, respectively (as in Fig. 3b). Sample 4445 (CRC,
859 red) is denoted by *. **(b)** Box plot representation of the data in A (healthy = 33, CRC = 46).
860 Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest
861 (smallest) value no further than 1.5× interquartile range from the hinge. P values were
862 calculated by Welch's t-test. *** P value < 0.001. **(c)** Global level of cfNucleosomes, and levels
863 of H3K4me1-modified nucleosomes, significantly differ between healthy and CRC late stage
864 samples (healthy = 33, CRC = 46). P values were calculated by Welch's t-test. * P value < 0.05
865 ** P value < 0.01. *** P value < 0.001. **(d)** Multiple comparison of all significant parameters
866 between CRC, CRC resected (CRC_R) and healthy samples (CRC = 19, CRC_R=27, healthy
867 = 33), corresponding to the data presented in main Fig. 3c-e. Of note, while all of these
868 parameters differ between healthy versus the combined cohort of <u>all CRC</u> patients, this figure
869 shows the differences between CRC patients with/without tumor resection. In some
870 parameters, resected patients show higher similarity to healthy, and in other parameters, they
871 are similar to CRC patients prior to tumor resection. See methods for P value calculation. * P
872 value < 0.05 ** P value < 0.01. *** P value < 0.001. **(e)** Levels of CEA, MST1 and TIMP1
873 significantly differ between healthy and early stage CRC patients (healthy = 33, early CRC =
874 17). Total levels of cfNucleosomes do not show significant difference between the groups,
875 likely due to low tumor burden at early stage patients. P values were calculated by Wilcoxon
876 rank sum exact test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(f)** TIMP1 levels
877 do not significantly differ in the cohort of healthy versus late-stage CRC patients.

878

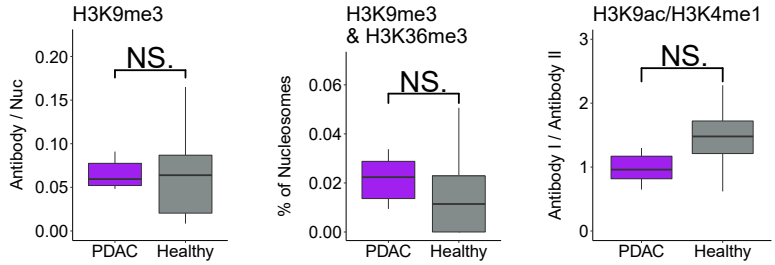
a


H3K27me3
Antibody / Nuc
***
PDAC
Healthy
H3K4me3
Antibody / Nuc
***
H3K9ac
Antibody / Nuc
***
H3K4me1
Antibody I / Antibody II
*
H3K36me3/H3K9me3
Antibody I / Antibody II
**
H3K4me1
& H3K9ac
% of Nucleosomes
***
H3K4me3
& H3K27me3
% of Nucleosomes
***
cfNucleosomes
Count
**
CEA
Normalized Counts
***
CEA/MST1
Normalized Ratio
**
DNA Methylation
Normalized counts
**

b


H3K27me3
Antibody / Nuc
**
PDAC
CRC
H3K4me3
Antibody / Nuc
***
H3K9ac
Antibody / Nuc
*
H3K4me3
& H3K27me3
% of Nucleosomes
***
H3K9me3
& H3K36me3
% of Nucleosomes
**
H3K4me1
& H3K9ac
% of Nucleosomes
**

c

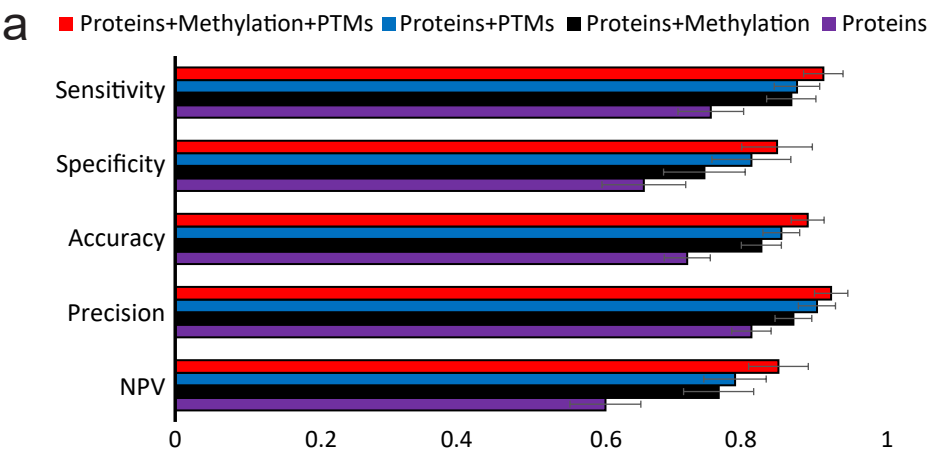

H3K9me3
Antibody / Nuc
NS.
PDAC
Healthy
H3K9me3
& H3K36me3
% of Nucleosomes
NS.
H3K9ac/H3K4me1
Antibody I / Antibody II
NS.

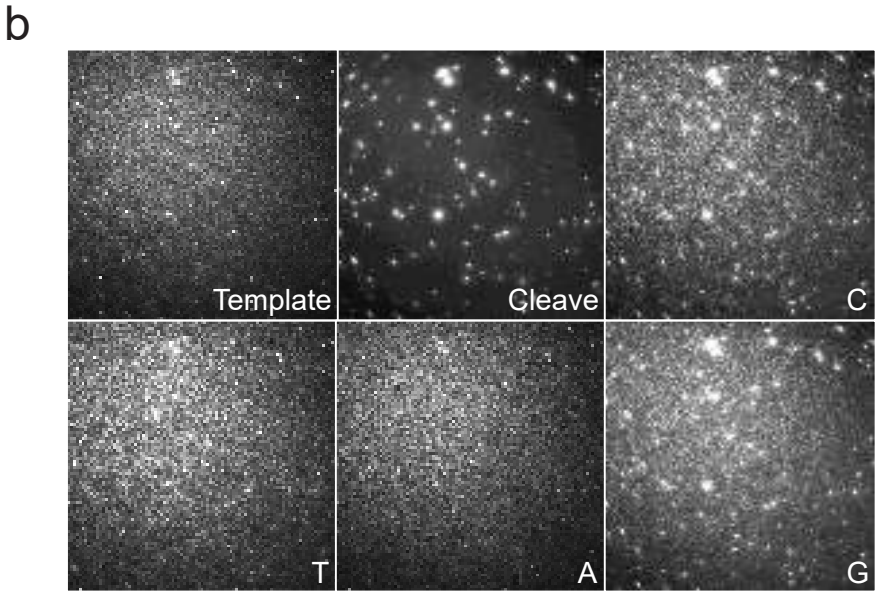
879 **Supplementary Fig.7**

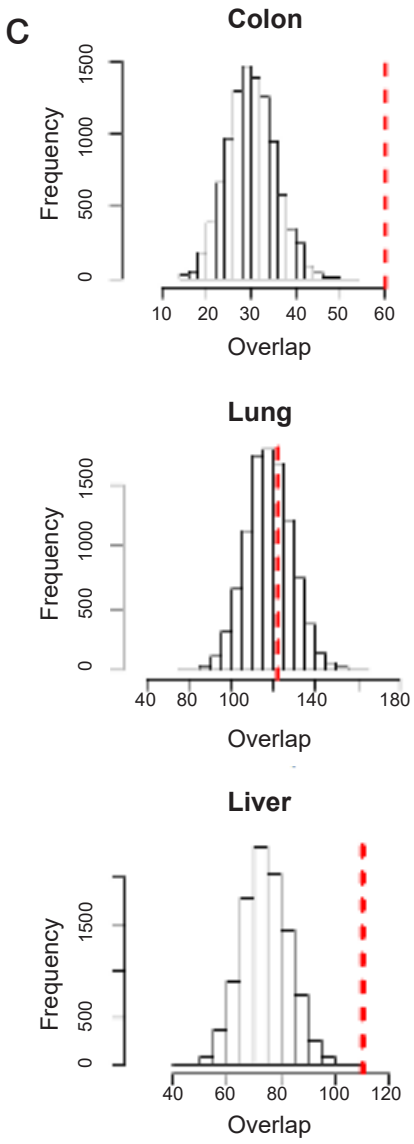
880 **EPINUC reveals significant epigenetic and biomarkers alterations in the plasma of PDAC**
881 **patients.**

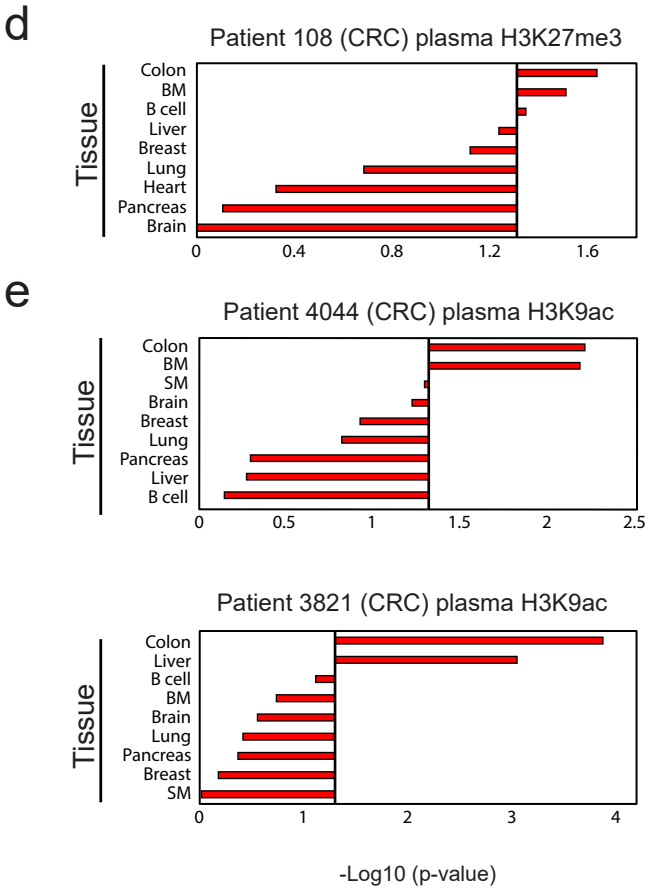
882 **(a)** EPINUC measurements (as indicated on the graphs) that significantly differ between
883 healthy and late stage PDAC samples (healthy = 33, PDAC = 10). P values were calculated by
884 Wilcoxon rank sum exact test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(b)**
885 EPINUC measurements (as indicated on the graphs) that significantly differ between PDAC
886 and CRC late stage plasma samples (CRC = 46, PDAC = 10). P values were calculated by
887 Wilcoxon rank sum exact test. * P value < 0.05 ** P value < 0.01. *** P value < 0.001. **(c)**
888 Parameters that are specifically altered in late stage CRC compared to healthy (see Fig. 3d),
889 but are not altered in patients diagnosed with PDAC. P values were calculated by Wilcoxon
890 rank sum exact test. NS, Not significant.

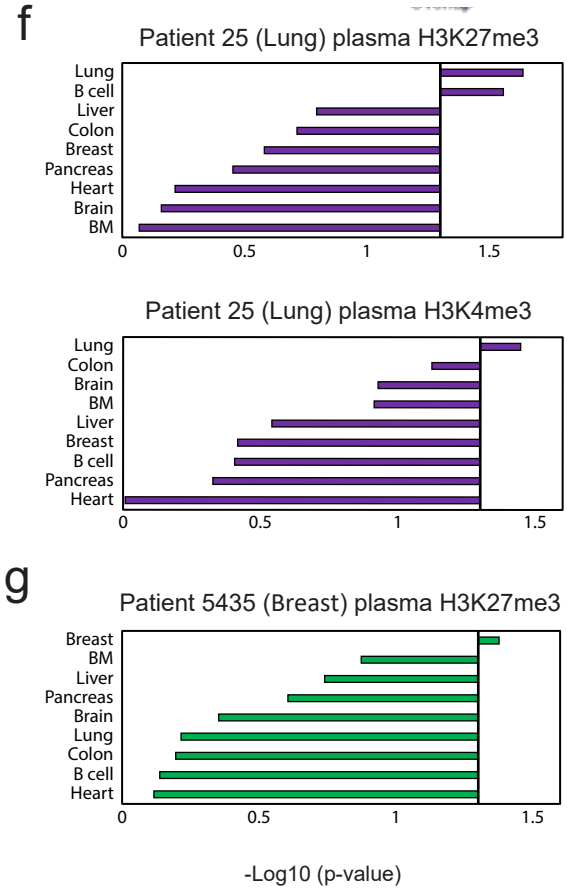
891


a
Proteins+Methylation+PTMs
Proteins+PTMs
Proteins+Methylation
Proteins
Sensitivity
Specificity
Accuracy
Precision
NPV
0
0.2
0.4
0.6
0.8
1


b
Template
Cleave
C
T
A
G


c
Colon
Frequency
0
500
1000
1500
10
20
30
40
50
60
Overlap
Lung
Frequency
0
500
1000
1500
40
80
100
140
180
Overlap
Liver
Frequency
0
500
1000
1500
40
60
80
100
120
Overlap


d
Patient 108 (CRC) plasma H3K27me3
Tissue
Colon
BM
B cell
Liver
Breast
Lung
Heart
Pancreas
Brain
0
0.4
0.8
1.2
1.6
e
Patient 4044 (CRC) plasma H3K9ac
Tissue
Colon
BM
SM
Brain
Breast
Lung
Pancreas
Liver
B cell
0
0.5
1
1.5
2
2.5
Patient 3821 (CRC) plasma H3K9ac
Tissue
Colon
Liver
B cell
BM
Brain
Lung
Pancreas
Breast
SM
0
1
2
3
4
-Log10 (p-value)

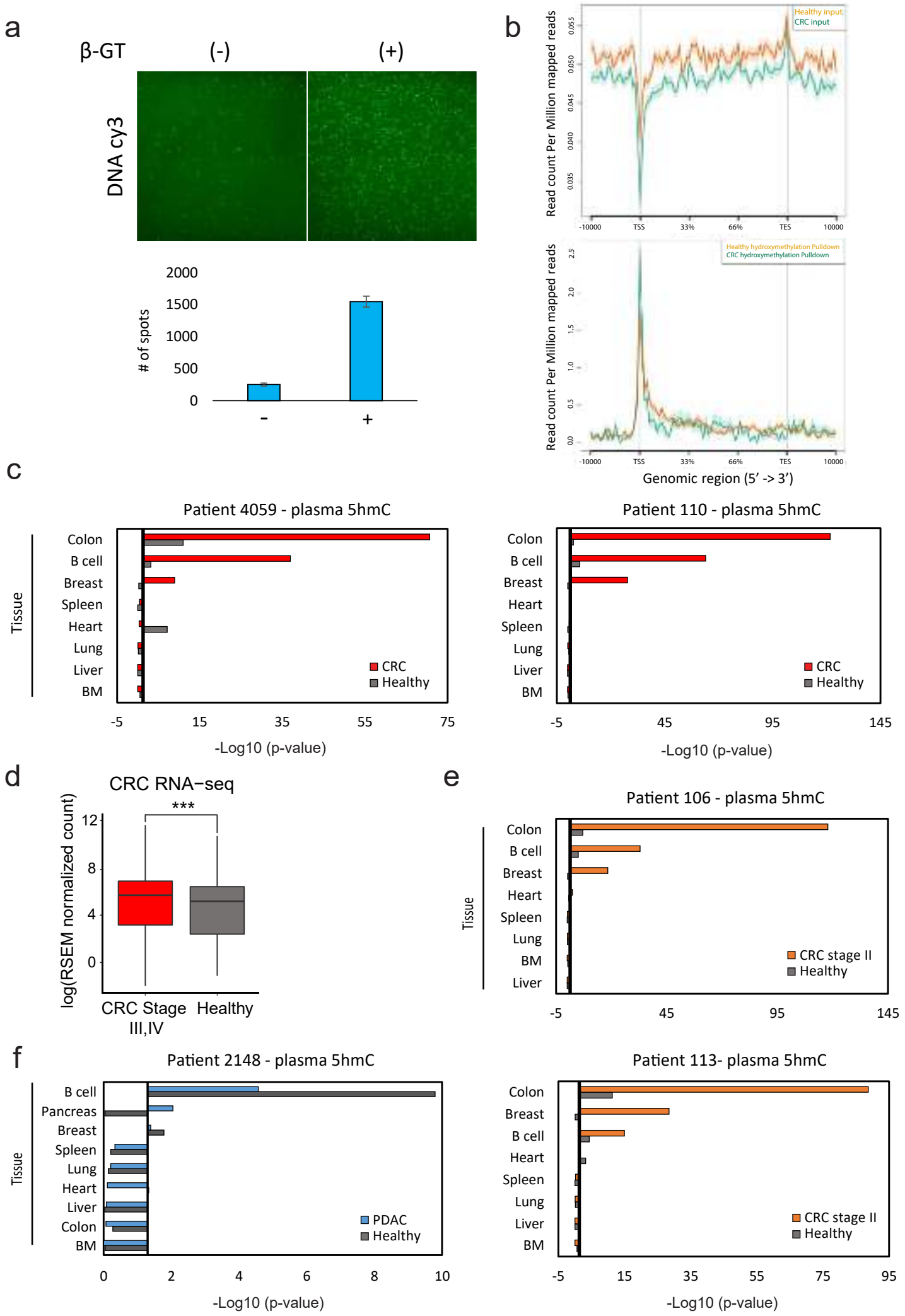

f
Patient 25 (Lung) plasma H3K27me3
Lung
B cell
Liver
Colon
Breast
Pancreas
Heart
Brain
BM
0
0.5
1
1.5
Patient 25 (Lung) plasma H3K4me3
Lung
Colon
Brain
BM
Liver
Breast
B cell
Pancreas
Heart
0
0.5
1
1.5
g
Patient 5435 (Breast) plasma H3K27me3
Breast
BM
Liver
Pancreas
Brain
Lung
Colon
B cell
Heart
0
0.5
1
1.5
-Log10 (p-value)

892 **Supplementary Fig.8**

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

894 **(a)** Logistic model performance increases with integration of additional layers of information.
895 Corresponds to Fig. 4a. NPV, negative predictive value. **(b)** Representative TIRF images of
896 antibodies' binding (Template) followed by cleavage of fluorophore (Cleave) and sequencing
897 cycles (C,T,A,G). High diameter spots correspond to FluoSpheres used for image alignment.
898 **(c)** Histograms portray colon, lung and liver frequency distributions of random reads
899 (equivalent to the number of H3K27me3 plasma CRC positive reads) that overlap with tissue
900 H3K27me3 unique peaks (bootstrapping simulation, 10K iterations, Methods). Red line
901 corresponds to the number of H3K27me3 unique tissue peaks that overlap with CRC (patient
902 4044) H3K27me3 positive plasma reads. **(d)** EPINUC-seq analysis of an additional CRC
903 patient, similar to the analysis shown in Fig. 4d. **(e)** EPINUC-seq H3K9ac analysis of plasma
904 from two CRC patients (stage IV). Tissues and primary cell lines ranked by overlap
905 significance with single-molecule H3K9ac positive reads, indicating CRC as the tissue-of-
906 origin. Heart H3K9ac peaks were available only for embryonic tissue (ENCODE), therefore
907 were replaced with skeletal muscle tissue H3K9ac peaks. **(f)** EPINUC-seq analysis of plasma
908 from stage IV lung cancer patient (patient 25). Tissues and primary cell lines ranked by overlap
909 significance with single-molecule H3K27me3 (Left) or H3K4me3 (Right) positive reads,
910 indicating lung as the tissue-of-origin. **(g)** EPINUC-seq analysis of plasma from stage IV
911 Breast cancer patient (patient 5435). Tissues and primary cell lines ranked by overlap
912 significance with single-molecule H3K27me3 positive reads, indicating breast as the tissue-of-
913 origin. Black line corresponds to P value of 0.05. P values were determined by two tailed Z-
914 test or Wilcoxon rank sum test.

915

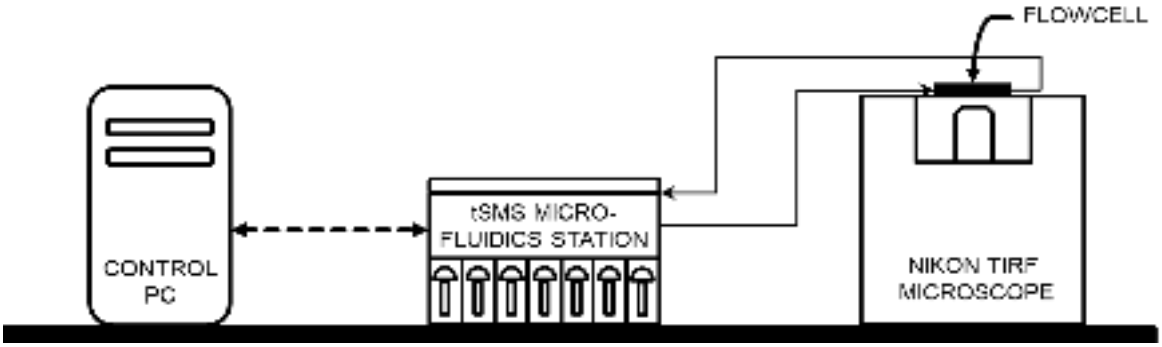

a
β-GT
(-)
(+)
DNA cy3
# of spots
2000
1500
1000
500
0
-
+
b
Healthy input
CRC input
Read count Per Million mapped reads
-10000
TSS
33%
66%
TES
10000
Healthy hydroxymethylation Pulldown
CRC hydroxymethylation Pulldown
Genomic region (5' -> 3')
c
Patient 4059 - plasma 5hmC
Tissue
Colon
B cell
Breast
Spleen
Heart
Lung
Liver
BM
CRC
Healthy
-Log10 (p-value)
Patient 110 - plasma 5hmC
d
CRC RNA−seq
***
log(RSEM normalized count)
CRC Stage III,IV
Healthy
e
Patient 106 - plasma 5hmC
CRC stage II
f
Patient 2148 - plasma 5hmC
Pancreas
PDAC
Patient 113- plasma 5hmC

916 **Supplementary Fig.9**

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

918 **(a)** Representative TIRF images (Top) and quantification (Bottom) of fluorescently labeled
919 nucleosomal DNA (Cy3, green) enriched for 5hmC, with (+) or without (-) the biotin
920 conjugating enzyme beta-glucosyltransferase (β-GT). Data is presented as the mean +/- s.d. of
921 50 FOV for each treatment. **(b)** Metagene profiles of input (Top) and 5hmC enriched (Bottom)
922 cfDNA sequenced from healthy (n=3, green) and CRC (n=3, Orange) samples, exhibiting
923 5hmC enrichment at gene bodies. **(c)** Overlap significance of tissues and primary cell lines
924 unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus late stage
925 CRC samples, similar to the analysis shown in Fig. 4g. **(d)** Boxplot depicts comparison of RNA
926 expression in CRC primary tumor of each group of unique 5hmc gene signatures in healthy vs
927 CRC (see Methods). RSEM, RNA-Seq by expectation-maximization. **(e-f)** Overlap
928 significance of tissues and primary cell lines unique H3K36me3 profiles with single-molecule
929 5hmC reads from healthy versus early stage CRC **(e)** or healthy versus PDAC **(f)**, similar to
930 the analysis shown in Fig. 4h. Each panel represents a different patient. Black line corresponds
931 to P value of 0.05. P values were determined by Z-test.

932


FLOWCELL
CONTROL
PC
tSMS MICRO-
FLUIDICS STATION
NIKON TIRF
MICROSCOPE

933 **Supplementary Fig.10**

934 **Single-molecule DNA sequencing technical setup**

935 Schematic representation of key instrumentation used to perform single-molecule DNA
936 sequencing. Straight and dashed lines represent microfluidic tubing and electric cables,
937 respectively.

938

939

940 **References**


941 1. Wan, J. C. M. *et al.* Liquid biopsies come of age: Towards implementation of
942 circulating tumour DNA. *Nat. Rev. Cancer* **17,** 223–238 (2017).

943 2. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA
944 as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **17,** 100087
945 (2019).

946 3. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future
947 perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20,** 71–
948 88 (2019).

949 4. Lo, Y. M. D., Han, D. S. C., Jiang, P. & Chiu, R. W. K. Epigenetics, fragmentomics,
950 and topology of cell-free DNA in liquid biopsies. *Science (80-. ).* **372,** (2021).

951 5. Xu, R. H. *et al.* Circulating tumour DNA methylation markers for diagnosis and
952 prognosis of hepatocellular carcinoma. *Nat. Mater.* **16,** 1155–1162 (2017).

953 6. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of
954 circulating cell-free DNA in health and disease. *Nat. Commun.* **9,** 5068 (2018).

955 7. Kang, S. *et al.* CancerLocator: Non-invasive cancer diagnosis and tissue-of-origin
956 prediction using methylation profiles of cell-free DNA. *Genome Biol.* **18,** 1–12 (2017).

957 8. Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free
958 DNA methylomes. *Nature* **563,** 579–583 (2018).

959 9. Reinberg, D. & Vales, L. D. Chromatin domains rich in inheritance only certain
960 histone posttranslational modifications qualify as being epigenetic. *Science (80-. ).*
961 **361,** 33–34 (2018).

962 10. Shema, E., Bernstein, B. E. & Buenrostro, J. D. Single-cell and single-molecule
963 epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.*
964 **51,** 19–25 (2019).

965 11. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev.*
966 *Genet.* **17,** 487–500 (2016).

967 12. Mancarella, D. & Plass, C. Epigenetic signatures in cancer: proper controls, current
968 challenges and the potential for clinical translation. *Genome Med. 2021 131* **13,** 1–12
969 (2021).

970 13. Sadeh, R. *et al.* ChIP-seq of plasma cell-free nucleosomes identifies gene expression
971 programs of the cells of origin. *Nat. Biotechnol.* **39,** 586–598 (2021).

972 14. Gezer, U. *et al.* Histone methylation marks on circulating nucleosomes as novel blood-
973 based biomarker in colorectal cancer. *Int. J. Mol. Sci.* **16,** 29654–29662 (2015).

974 15. Van den Ackerveken, P. *et al.* A novel proteomics approach to epigenetic profiling of
975 circulating nucleosomes. *Sci. Rep.* **11,** 1–12 (2021).

976 16. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA
977 Comprises an in Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*
978 **164,** 57–68 (2016).

979 17. Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA.
980 *Nat. Genet.* **48,** 1273–1278 (2016).

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

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

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

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

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

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

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

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

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

999 27. Mandal, S. *et al.* Direct Kinetic Fingerprinting for High-Accuracy Single-Molecule

1000 Counting of Diverse Disease Biomarkers. *Acc. Chem. Res.* (2020).
1001 doi:10.1021/acs.accounts.0c00621

1002 28. Furth, N. *et al.* Unified platform for genetic and serological detection of COVID-19
1003 with single-molecule technology. *PLoS One* **16,** e0255096 (2021).

1004 29. Nakayama, M. & Oshima, M. Mutant p53 in colon cancer. *J. Mol. Cell Biol.* **11,** 267–
1005 276 (2019).

1006 30. Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F. & Goel, A. Epigenetics of
1007 colorectal cancer: biomarker and therapeutic potential. *Nat. Rev. Gastroenterol.*
1008 *Hepatol.* **17,** 111–130 (2020).

1009 31. Dawson, M. A. The cancer epigenome: Concepts, challenges, and therapeutic
1010 opportunities. *Science (80-. ).* **355,** 1147–1152 (2017).

1011 32. Wood, K. H. & Zhou, Z. Emerging molecular and biological functions of MBD2, a
1012 reader of DNA methylation. *Front. Genet.* **7,** (2016).

1013 33. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage
1014 human malignancies. *Sci. Transl. Med.* **6,** 224ra24-224ra24 (2014).

1015 34. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised
1016 epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* **14,** 747–753
1017 (2014).

1018 35. Kerachian, M. A. *et al.* Crosstalk between DNA methylation and gene expression in
1019 colorectal cancer, a potential plasma biomarker for tracing this tumor. *Sci. Rep.* **10,** 1–
1020 13 (2020).

1021 36. King, W. D. *et al.* A cross-sectional study of global DNA methylation and risk of
1022 colorectal adenoma. *BMC Cancer* **14,** 1–9 (2014).

1023 37. Frederiksen, C. *et al.* Plasma TIMP-1 levels and treatment outcome in patients treated
1024 with XELOX for metastatic colorectal cancer. *Ann. Oncol.* **22,** 369–375 (2011).

1025 38. Garrido-Laguna, I. & Hidalgo, M. Pancreatic cancer: from state-of-the-art treatments
1026 to promising novel therapies. *Nat. Rev. Clin. Oncol. 2015 126* **12,** 319–334 (2015).

1027 39. Lubotzky, A. *et al.* Liquid biopsy reveals collateral tissue damage in cancer. *JCI*
1028 *Insight* **7,** (2022).

1029 40. Gai, W. *et al.* Liver- and Colon-Specific DNA Methylation Markers in Plasma for

1030 Investigation of Colorectal Cancers with or without Liver Metastases. *Clin. Chem.* **64,**
1031 1239–1249 (2018).

1032 41. Tannapfel, A. & Reinacher-Schick, A. [Chemotherapy associated hepatotoxicity in the
1033 treatment of advanced colorectal cancer (CRC)]. *Z. Gastroenterol.* **46,** 435–440
1034 (2008).

1035 42. Li, W. *et al.* 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as
1036 diagnostic biomarkers for human cancers. *Cell Res.* **27,** 1243–1257 (2017).

1037 43. Lio, C. W. J., Yuita, H. & Rao, A. Dysregulation of the TET family of epigenetic
1038 regulators in lymphoid and myeloid malignancies. *Blood* **134,** 1487–1497 (2019).

1039 44. Zhang, L. *et al.* Tet-mediated covalent labelling of 5-methylcytosine for its genome-
1040 wide detection and sequencing. *Nat. Commun.* **4,** (2013).

1041 45. Song, C. X. *et al.* 5-Hydroxymethylcytosine signatures in cell-free DNA provide
1042 information about tumor types and stages. *Cell Res.* **27,** 1231–1242 (2017).

1043 46. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of
1044 circulating tumor DNA. *Nat. Biotechnol.* **34,** 547–555 (2016).

1045 47. Lisanti, S. *et al.* Comparison of Methods for Quantification of Global DNA
1046 Methylation in Human Cells and Tissues. *PLoS One* **8,** 79044 (2013).

1047 48. Bock, C. *et al.* Quantitative comparison of DNA methylation assays for biomarker
1048 development and clinical applications. *Nat. Biotechnol. 2016 347* **34,** 726–737 (2016).

1049 49. Chandradoss, S. D. *et al.* Surface passivation for single-molecule protein studies. *J.*
1050 *Vis. Exp.* (2014). doi:10.3791/50549

1051 50. Fleischhacker, M. & Schmidt, B. Circulating nucleic acids (CNAs) and cancer-A
1052 survey. *Biochim. Biophys. Acta - Rev. Cancer* **1775,** 181–232 (2007).

1053 51. Harris, T. D. *et al.* Single-Molecule DNA Sequencing of a Viral Genome. *Science (80-*
1054 *. ).* **320,** 106–109 (2008).

1055 52. Kim, K. L. *et al.* Systematic detection of m6A-modified transcripts at single-molecule
1056 and single-cell resolution. *Cell Reports Methods* **1,** 100061 (2021).

1057