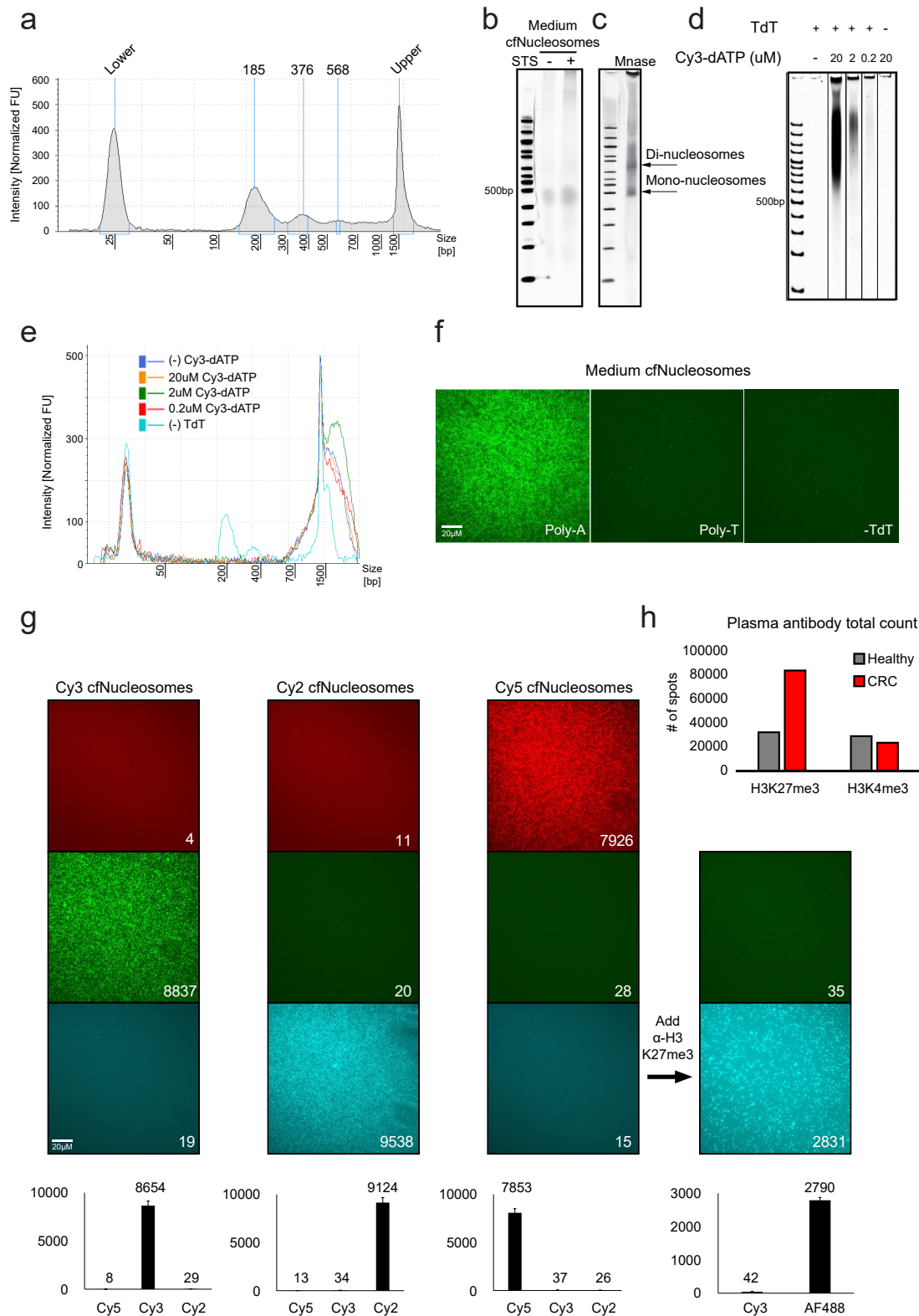

Supplementary information

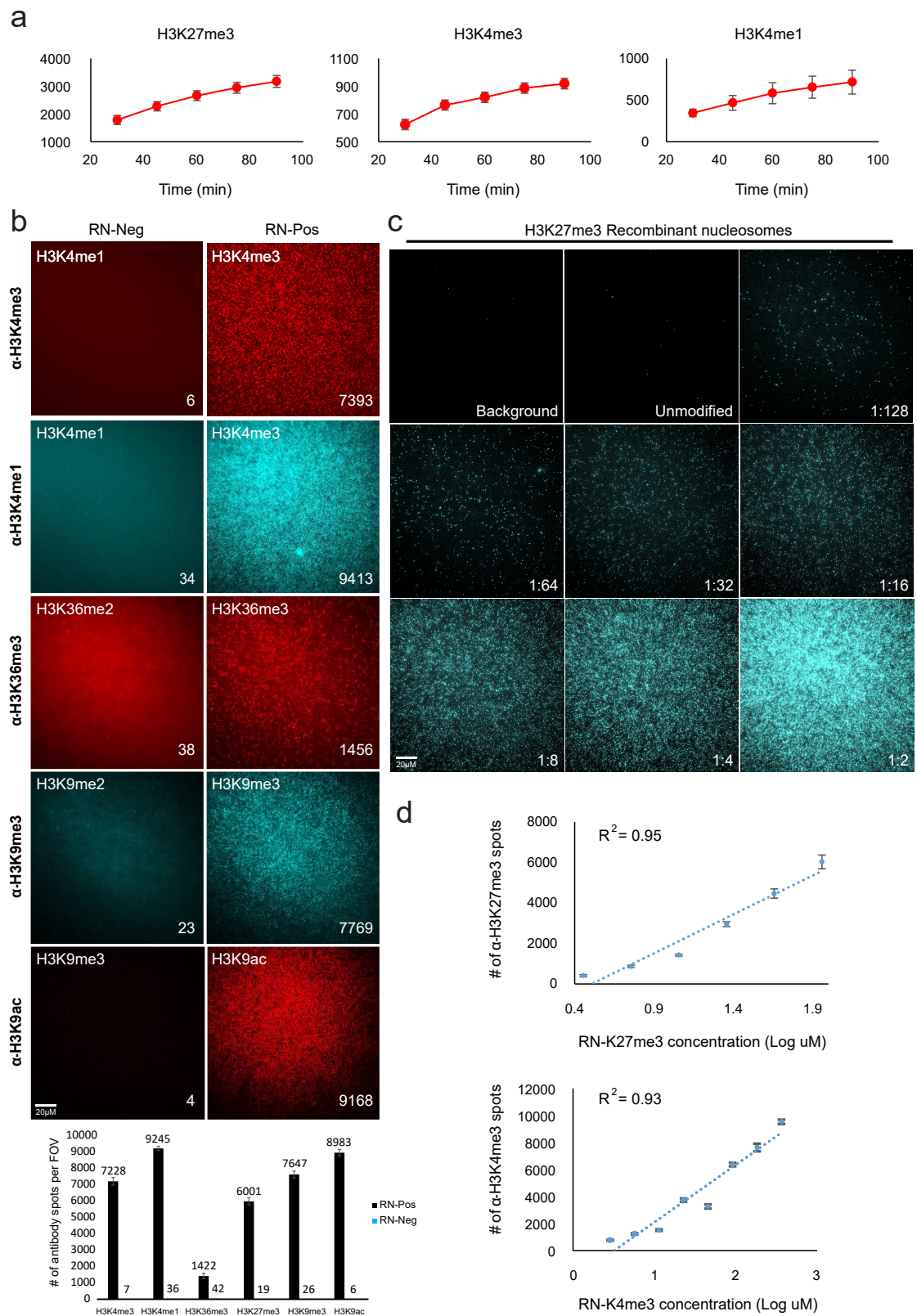
Multiplexed, single-molecule, epigenetic analysis of plasma-isolated nucleosomes for cancer diagnostics

In the format provided by the
authors and unedited



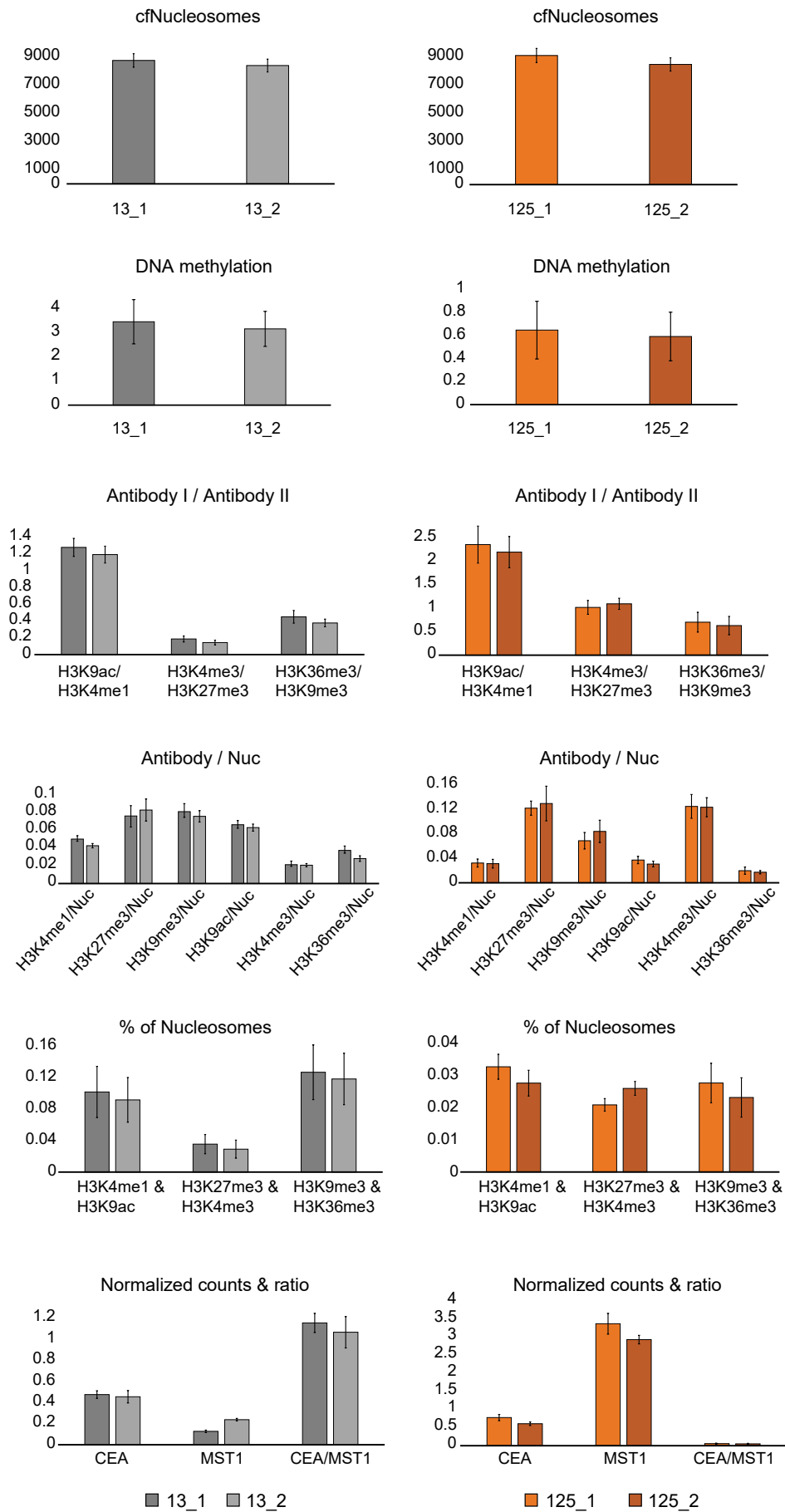
Supplementary Figure 1: In-vitro system for sample preparation of cfNucleosomes.

(a) TapeStation fragment size analysis of DNA isolated from medium of HEK293 cells treated with the apoptosis-inducing factor Staurosporine (STS). DNA exhibits canonical apoptotic DNA fragmentation pattern. (b) Nucleosomes from medium of HEK293 cells (described in a) treated with STS (+) or with PBS (-) were resolved on a 6% TBE gel and visualized by Typhoon laser scanner. (c) Nucleosomes were extracted from HEK293 cells by digestion with MNase21, resolved on a 6% TBE gel and visualized by Typhoon laser scanner. Bands correspond to mono-nucleosomes and di-nucleosomes. (d) Nucleosomes from medium of HEK293 cells (described in a) were subjected to tagging and tailing reaction (see Methods), resolved on a 6% TBE gel and visualized by Typhoon laser scanner. Image depicts fluorescent labeling via incorporation of Cy3-dATP during polyadenylation reaction. Incorporation is dependent on the presence of TdT in the reaction, and level of florescent signal correlates with the concentration of Cy3-dATP (Left). Samples were ran on the same gel with non-adjacent lanes marked by black rectangles. (e) TapeStation fragment size analysis of the samples in (d). (f) Poly-dA tailed and Cy3-dATP labeled cfNucleosomes show specific anchoring to PEGylated poly-dT surfaces through A:T hybridization. (g) Images and quantification of poly-dA tailed cfNucleosomes, labeled with either Cy2-dATP, Cy3-dATP or Cy5-dATP. Data indicates very low crosstalk between the three fluorescent channels. Further incubation of Cy-5-labeled cfNucleosomes with α -H3K27me3- AF488 antibody reinforces low spectral overlap. Data is presented as the mean $\pm$ s.d. of 50 FOVs for each channel. (h) Total counts for H3K27me3 and H3K4me3 in the plasma samples analyzed in Fig. 1d,e. Scale bar = 20µm for all images.



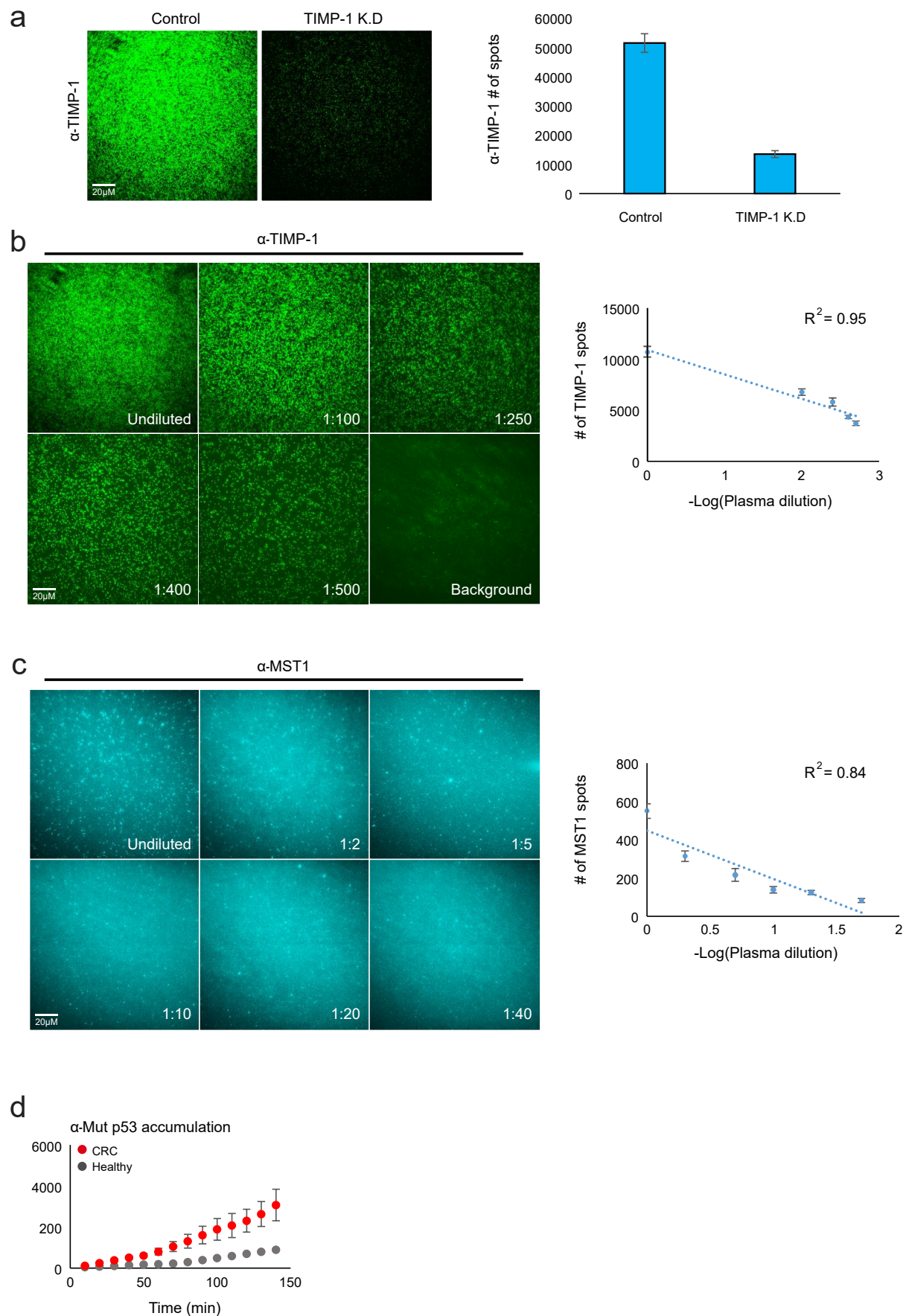
Supplementary Figure 2 : Single-molecule measurements of antibodies' specificity and dynamics.

(a) Accumulation of unique antibody binding events over time. Data is presented as the mean $\pm$ standard deviation (s.d.) of 50 FOVs for each time point. (b) Top: Representative TIRF images demonstrate antibodies binding to various recombinant nucleosomes (RN) carrying different histone PTMs. Left images: negative control, showing very low un-specific binding of fluorescent antibodies to recombinant nucleosomes that do not carry the target PTM. Right: binding of antibodies to recombinant nucleosomes that carry the target modification. The type of modification is indicated at the bottom of the images. Numbers within images represent counted spots. Bottom: Quantification of the averaged counts of the indicated antibodies per 50 FOVs. (c) Representative TIRF images depicting binding of the antibody targeting H3K27me3 to an empty surface (Background), unmodified recombinant nucleosomes (Unmodified), and H3K27me3-modified recombinant nucleosomes diluted as indicated. (d) Regression analysis of number of spots per FOV as a function of target recombinant nucleosomes' concentration demonstrates low variability along the regression line, supporting linearity of binding. 1:2 dilution was excluded from analysis due to high density. Data is presented as the mean $\pm$ s.d. of 50 FOVs for each concentration. Scale bar = 20 μ m for all images.



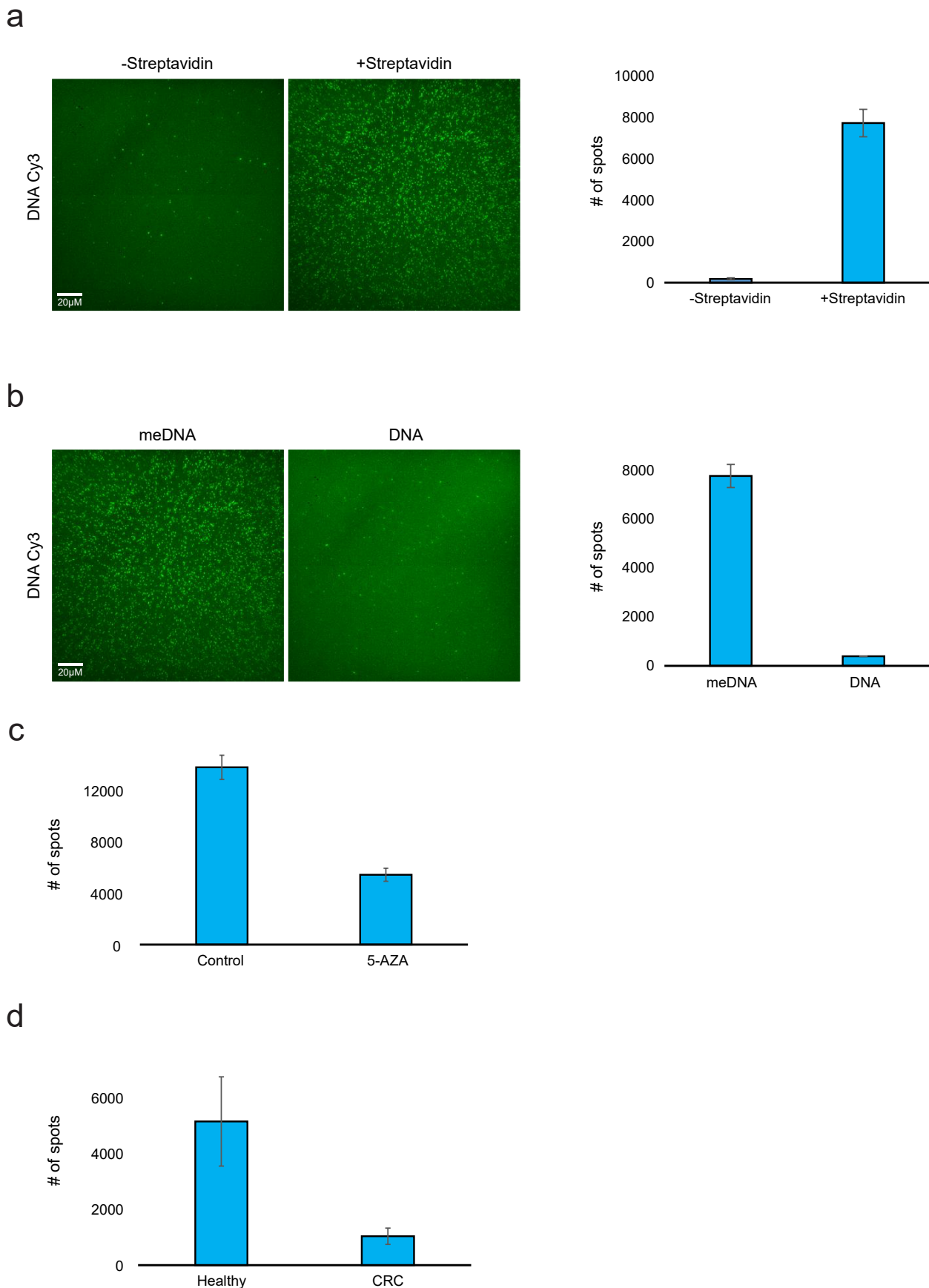
Supplementary Figure 3: Reproducibility of EPINUC measurements.

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



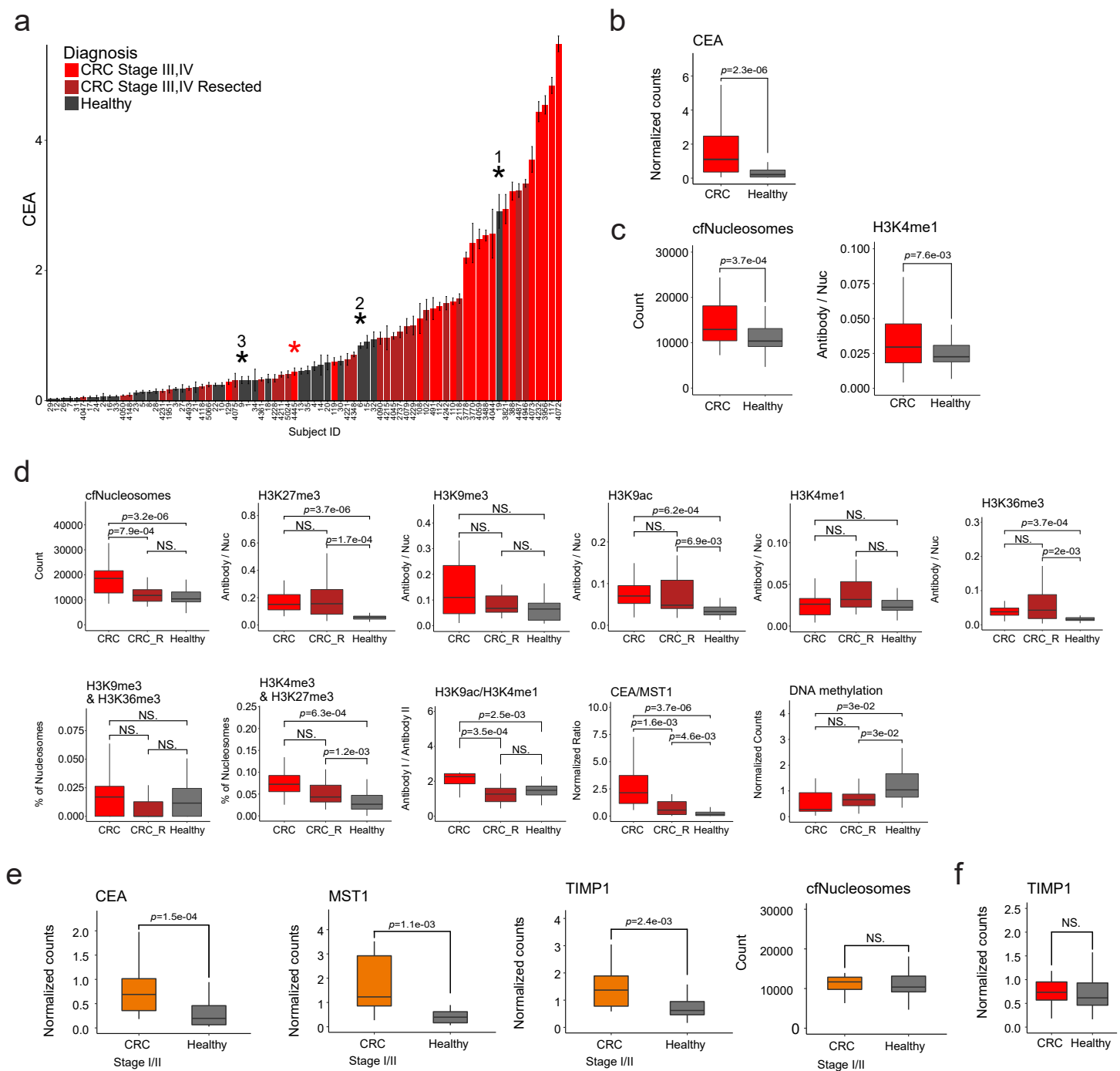
Supplementary Figure 4: Single-molecule imaging of MST1, TIMP-1 and mutant p53.

(a) Representative TIRF images (Left) and quantification (Right) of TIMP-1 protein levels in SW480 medium, following TIMP-1 knock-down versus control cells. Data is presented as the mean $\pm$ s.d. of 50 FOVs for each treatment. (b, c) Representative TIRF images and standard curves of antibodies targeting MST1 and TIMP-1 on serial plasma dilutions, depicting linear detection of molecules within this concentration range. Data is presented as the mean $\pm$ s.d. of 50 FOVs for each concentration. (d) Signal accumulation of mutant p53 over time for late stage CRC and healthy plasma samples. Each time point is presented as the mean $\pm$ s.d. of 50 FOVs. Scale bar = 20 μ m for all images.



Supplementary Figure 5: Single-molecule detection of DNA methylation levels.

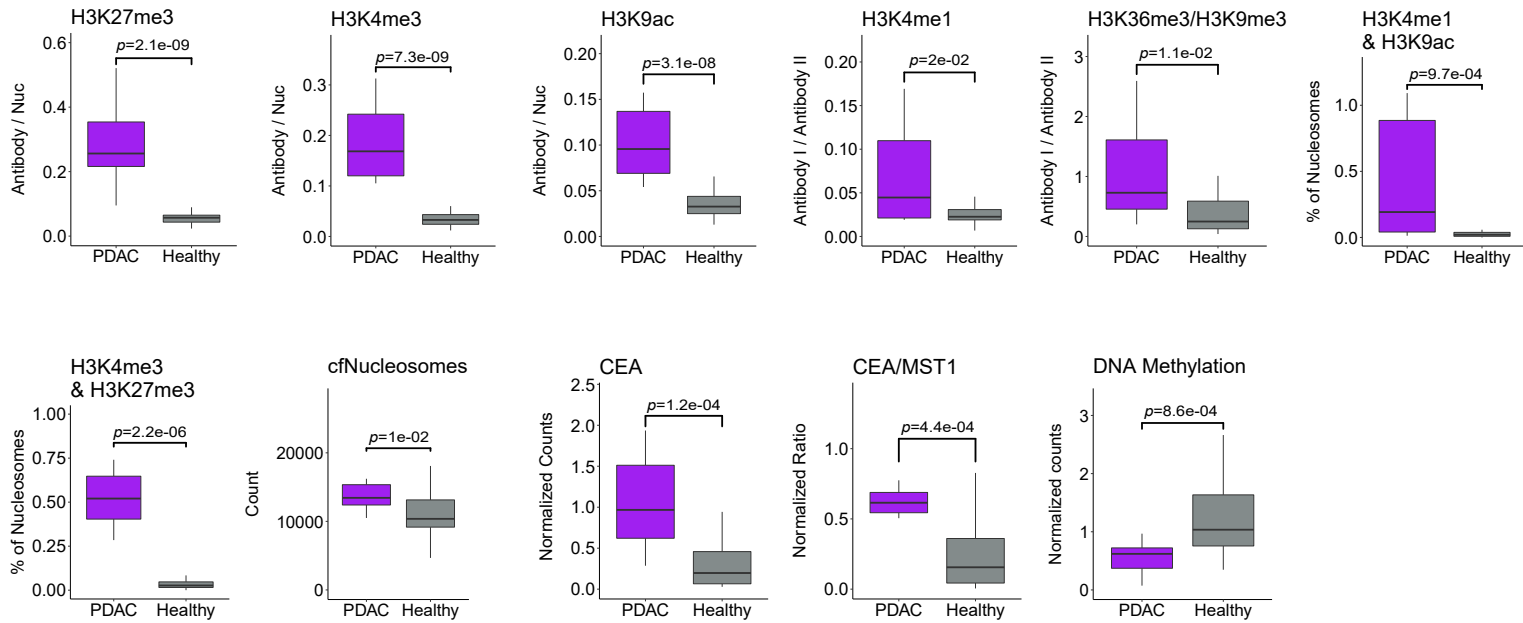
(a) Representative TIRF images (Left) and quantification (Right) indicating specific anchoring of Biotin-MBD-meDNA complexes to streptavidin-coated surfaces. Data is presented as the mean $\pm$ s.d. of 50 FOV for each treatment. (b) Biotin-MBD specifically binds methylated DNA. Representative TIRF images (Left) and quantification (Right) of global DNA methylation levels following incubation of MBD2-biotin with Cy3-labeled (green) methylated (meDNA) or un-methylated synthetic DNA fragments. Data is presented as the mean $\pm$ s.d. of 50 FOV for each treatment. (c, d) Quantification of global DNA methylation levels presented in Fig. 2i (c) and Fig. 2j (d). Data is presented as the mean $\pm$ s.d. of 50 FOV for each sample. Scale bar = 20 μ m for all images.



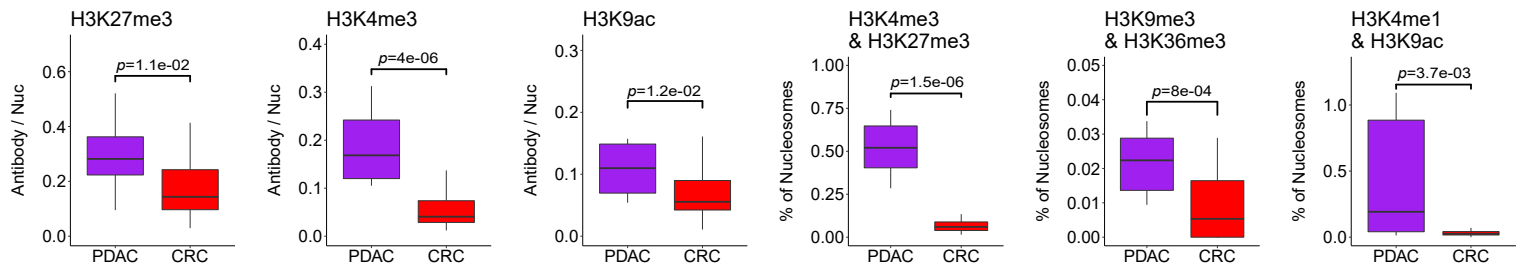
Supplementary Figure 6: Analysis of histone PTMs, protein biomarkers and DNA methylation in the cohort of plasma samples from healthy and CRC subjects.

(a) CEA normalized levels in the plasma of CRC patients and healthy individuals. Each bar represents a subject, data is presented as the mean $\pm$ s.d. of 50 FOVs per sample. *1-3 correspond to healthy samples 19, 6, and 9, respectively (as in Fig. 3b). Sample 4445 (CRC, red) is denoted by *. (b) Box plot representation of the data in A (healthy = 33, CRC = 46). Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than $1.5\times$ interquartile range from the hinge. P values were calculated by two-sided Welch's t-test. (c) Global level of cfNucleosomes, and levels of H3K4me1-modified nucleosomes, significantly differ between healthy and CRC late stage samples (healthy = 33, CRC = 46). P values were calculated by two-sided Welch's t-test. (d) Multiple comparison of all significant parameters between CRC, CRC resected (CRC_R) and healthy samples (CRC = 19, CRC_R=27, healthy = 33), corresponding to the data presented in main Fig. 3c-e. Of note, while all of these parameters differ between healthy versus the combined cohort of all CRC patients, this figure shows the differences between CRC patients with/without tumor resection. In some parameters, resected patients show higher similarity to healthy, and in other parameters, they are similar to CRC patients prior to tumor resection. Multiple comparison corrections were conducted using Asymptotic K-Sample Brown-Mood Median Test (Methods). NS, Not significant. (e) Levels of CEA, MST1 and TIMP1 significantly differ between healthy and early stage CRC patients (healthy = 33, early CRC = 17). Total levels of cfNucleosomes do not show significant difference between the groups, likely due to low tumor burden at early stage patients. P values were calculated by two-sided Wilcoxon rank sum exact test. (f) TIMP1 levels do not significantly differ in the cohort of healthy versus late-stage CRC patients (healthy = 33, CRC = 46). P values were calculated by two-sided Welch's t-test.

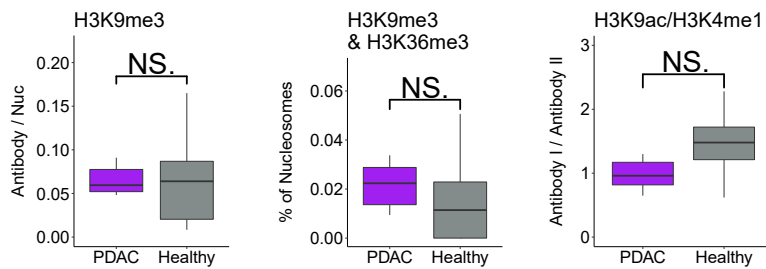
a



b

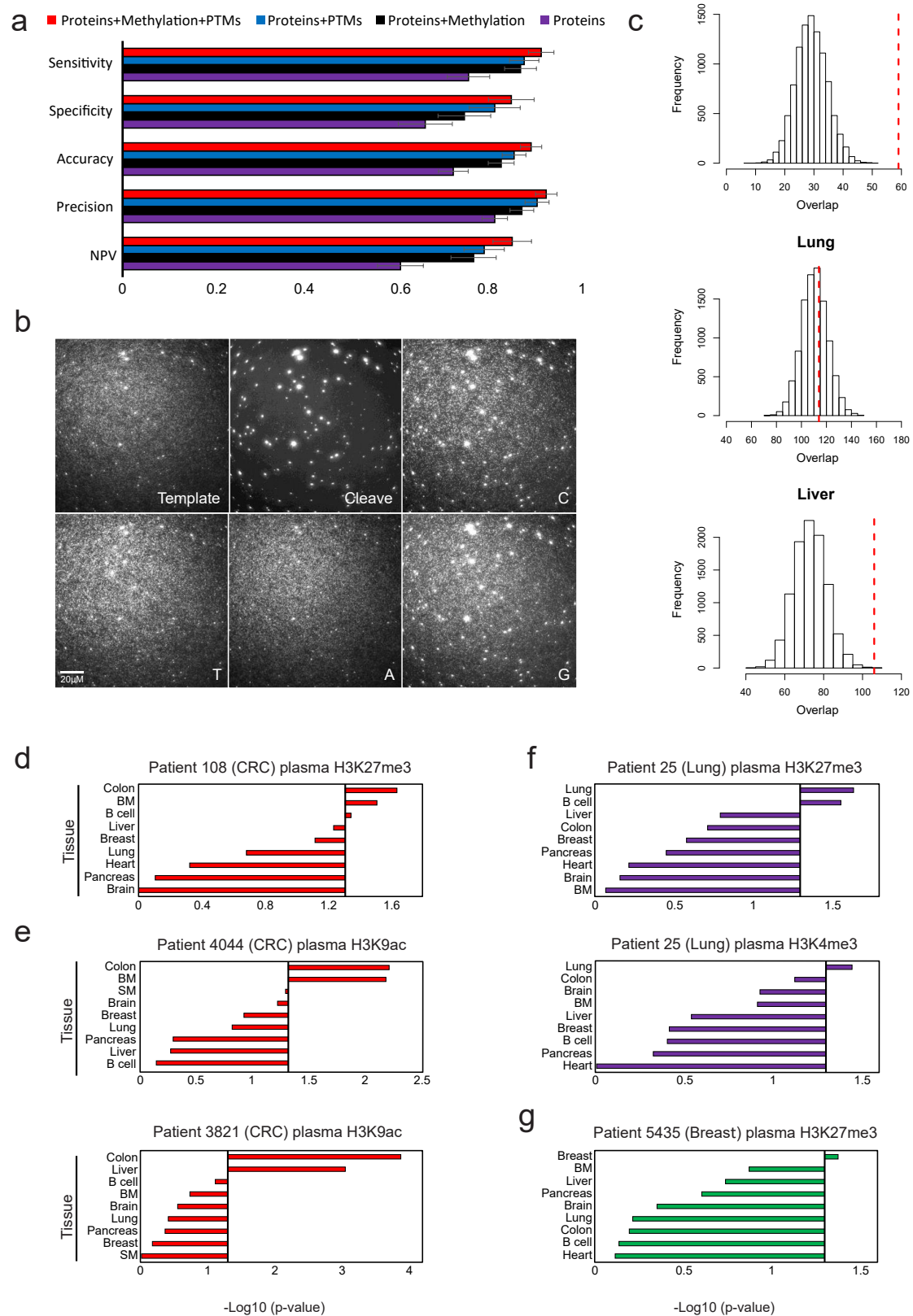


c



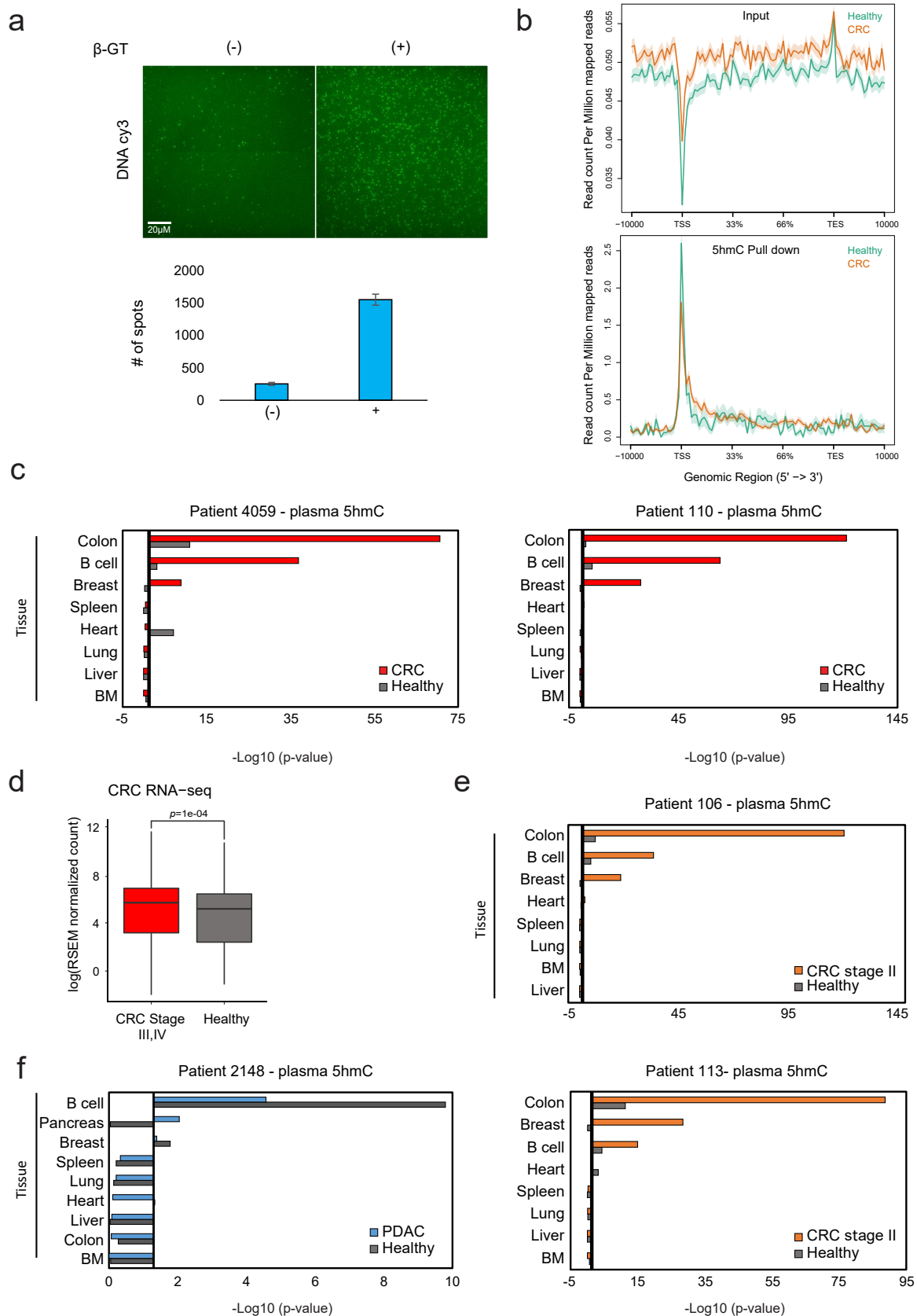
Supplementary Figure 7: EPINUC reveals significant epigenetic and biomarkers alterations in the plasma of PDAC patients.

(a) EPINUC measurements (as indicated on the graphs) that significantly differ between healthy and late stage PDAC samples (healthy = 33, PDAC = 10). P values were calculated by two-sided Wilcoxon rank sum exact test. (b) EPINUC measurements (as indicated on the graphs) that significantly differ between PDAC and CRC late stage plasma samples (CRC = 46, PDAC = 10). P values were calculated by two-sided Wilcoxon rank sum exact test. (c) Parameters that are specifically altered in late stage CRC compared to healthy (see Fig. 3d), but are not altered in patients diagnosed with PDAC. P values were calculated by two-sided Wilcoxon rank sum exact test. NS, Not significant. For all comparisons, box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than $1.5 \times$ interquartile range from the hinge.



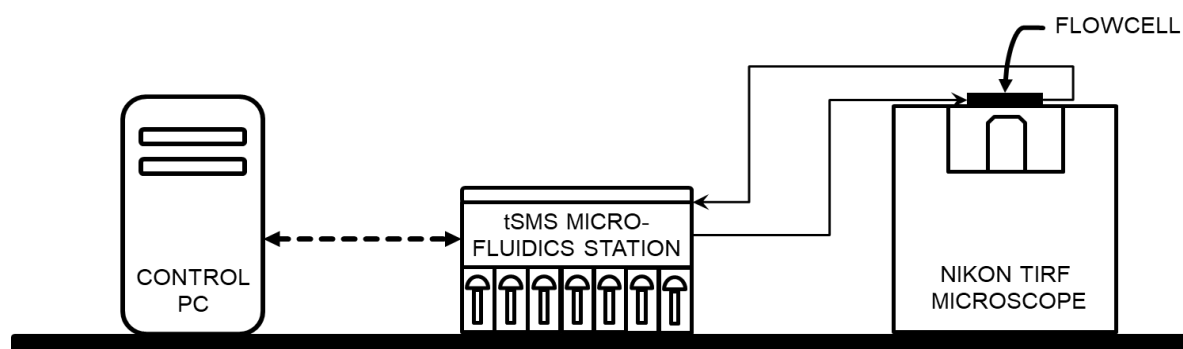
Supplementary Figure 8: Prediction model performance and Tissue-of-origin analysis based on EPINUC-seq.

(a) Logistic model performance increases with integration of additional layers of information. Corresponds to Fig. 4a. Classifier performance parameters mean was estimated for repeated (10K) 4-fold cross-validation across all healthy and CRC samples (Healthy=33, CRC=46). NPV, negative predictive value. Error bars corresponds to the 95% confidence interval for the mean. (b) Representative TIRF images of antibodies' binding (Template) followed by cleavage of fluorophore (Cleave) and sequencing cycles (C,T,A,G). High diameter spots correspond to FluoSpheres used for image alignment. Scale bar = 20µm. (c) Histograms portray colon, lung and liver frequency distributions of random reads (equivalent to the number of H3K27me3 plasma CRC positive reads) that overlap with tissue H3K27me3 unique peaks (bootstrapping simulation, 10K iterations, Methods). Red line corresponds to the number of H3K27me3 unique tissue peaks that overlap with CRC (patient 4044) H3K27me3 positive plasma reads. (d) EPINUC-seq analysis of an additional CRC patient, similar to the analysis shown in Fig. 4d. (e) EPINUC-seq H3K9ac analysis of plasma from two CRC patients (stage IV). Tissues and primary cell lines ranked by overlap significance with single-molecule H3K9ac positive reads, indicating CRC as the tissue-of-origin. Heart H3K9ac peaks were available only for embryonic tissue (ENCODE), therefore were replaced with skeletal muscle tissue H3K9ac peaks. (f) EPINUC-seq analysis of plasma from stage IV lung cancer patient (patient 25). Tissues and primary cell lines ranked by overlap significance with single-molecule H3K27me3 (Left) or H3K4me3 (Right) positive reads, indicating lung as the tissue-of-origin. (g) EPINUC-seq analysis of plasma from stage IV Breast cancer patient (patient 5435). Tissues and primary cell lines ranked by overlap significance with single-molecule H3K27me3 positive reads, indicating breast as the tissue-of-origin. Black line corresponds to P value of 0.05. P values were determined by two tailed Z-test or Wilcoxon rank sum test.



Supplementary Figure 9: Tissue-of-origin analysis based on single-molecule 5hmC sequencing.

(a) Representative TIRF images (Top) and quantification (Bottom) of fluorescently labeled nucleosomal DNA (Cy3, green) enriched for 5hmC, with (+) or without (-) the biotin conjugating enzyme beta-glucosyltransferase (β -GT). Scale bar = 20 μ m. Data is presented as the mean $\pm$ s.d. of 50 FOV for each treatment. (b) Metagene profiles of input (Top) and 5hmC enriched (Bottom) cfDNA sequenced from healthy (n=3, green) and CRC (n=3, Orange) samples, exhibiting 5hmC enrichment at gene bodies. (c) Overlap significance of tissues and primary cell lines unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus late stage CRC samples, similar to the analysis shown in Fig. 4g. (d) Boxplot depicts comparison of RNA expression in CRC primary tumor of each group of unique 5hmC gene signatures in healthy vs CRC (Healthy = 498, CRC=4407, Methods). RSEM, RNA-Seq by expectation-maximization. Box plots limits: 25–75% quantiles, middle: median, upper (lower) whisker to the largest (smallest) value no further than 1.5 $\times$ interquartile range from the hinge. P values were calculated by two-sided Welch's t-test. (e-f) Overlap significance of tissues and primary cell lines unique H3K36me3 profiles with single-molecule 5hmC reads from healthy versus early stage CRC (e) or healthy versus PDAC (f), similar to the analysis shown in Fig. 4h. Each panel represents a different patient. Black line corresponds to P value of 0.05. P values were determined by two-sided Z-test.



Supplementary Figure 10: Single-molecule DNA sequencing technical setup.

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