Toward Single-Molecule Optical Mapping of the Epigenome

18 AUTHORS, INCLUDING:

Natalie R Gassman
National Institute of Environmental Health S…
22 PUBLICATIONS **310** CITATIONS

SEE PROFILE

Ron Lin
University of California, Los Angeles
10 PUBLICATIONS **81** CITATIONS

SEE PROFILE

Shimon Weiss
University of California, Los Angeles
243 PUBLICATIONS **24,264** CITATIONS

SEE PROFILE

Yuval Ebenstein
Tel Aviv University
41 PUBLICATIONS **1,044** CITATIONS

SEE PROFILE

Toward Single-Molecule Optical Mapping of the Epigenome

Michal Levy-Sakin,[†] Assaf Grunwald,[†] Soohong Kim,[‡] Natalie R. Gassman,[‡] Anna Gottfried,[§] Josh Antelman,[‡] Younggyu Kim,[‡] Sam O. Ho,[‡] Robin Samuel,[‡] Xavier Michalet,[‡] Ron R. Lin,[‡] Thomas Dertinger,[‡] Andrew S. Kim,[‡] Sangyoon Chung,[‡] Ryan A. Colyer,[‡] Elmar Weinhold,[§] Shimon Weiss,[‡,*] and Yuval Ebenstein[†,*]

[†]Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry, Tel Aviv University, Tel Aviv 69978, Israel, [‡]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States, and [§]Institute of Organic Chemistry, RWTH Aachen University, Aachen 52062, Germany

ABSTRACT The past decade has seen an explosive growth in the utilization of single-molecule techniques for the study of complex systems. The ability to resolve phenomena otherwise masked by ensemble averaging has made these approaches especially attractive for the study of biological systems, where stochastic events lead to inherent inhomogeneity at the population level. The complex composition of the genome has made it an ideal system to study at the single-molecule level, and methods aimed at resolving genetic information from long, individual, genomic DNA molecules have been in use for the last 30 years. These methods, and particularly optical-based mapping of DNA, have been instrumental in highlighting genomic variation and contributed significantly to the assembly of many genomes including the human genome. Nanotechnology and nanoscopy have been a strong driving force for advancing genomic mapping approaches, allowing both better manipulation of DNA on the nanoscale and enhanced optical resolving power for analysis of genomic information. During the past few years, these developments have been adopted also for epigenetic studies. The common principle for these studies is the use of advanced optical microscopy for the detection of fluorescently labeled epigenetic marks on long, extended DNA molecules. Here we will discuss recent single-molecule studies for the mapping of chromatin composition and epigenetic DNA modifications, such as DNA methylation.
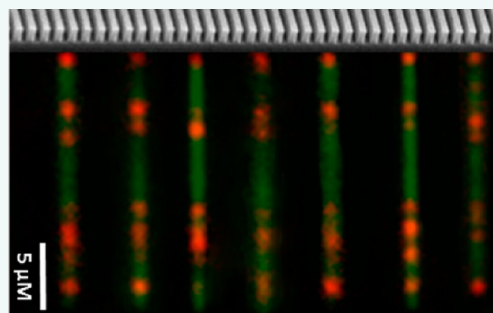
KEYWORDS: nanotechnology · single molecule · epigenetics · chromatin · methylation · fluorescence microscopy · nanoscopy · optical mapping

D ifferences in the genetic and epigenetic composition of genomes are the basis for phenotypic variation. Since the beginning of the Human Genome Project in 1989,[1] our knowledge-base of genomic information increased tremendously. This was achieved thanks to the emergence of the Sanger method for DNA sequencing[2,3] followed by next-generation sequencing methods (NGS).[4−6] However, the sequence layout of the genome is annotated by a plethora of epigenetic marks such as chemical modifications to the DNA bases or the association with specific DNA-binding proteins. These changes dramatically affect the structure and function of the genome without changing the underlying genomic sequence. At any given time, the epigenome of a cell is defined by the pattern of DNA modifications such as DNA methylation and the distribution of DNA-binding proteins, mainly transcription factors (TF) and histones.[7] The detailed composition of the epigenome serves to regulate the execution of the underlying genetic code and defines a specific gene expression profile that sets the phenotype for each cell.

The dynamic nature and high variability of epigenetic signatures limit the information accessible by bulk sequencing techniques. This limitation calls for alternative methodologies for studying the epigenome. Advances in our ability to manipulate and detect biomolecules at the nanoscale offer exciting new approaches to genomic analysis. Here we discuss the physical mapping of genomic and epigenomic content from the single-molecule perspective with emphasis on optical approaches.

DNA Sequencing and Optical Mapping. High-throughput sequencing technologies are all based on assembly of numerous short

sequence reads into long-range sequence contigs.[8] In order to achieve sufficient overlap between the short reads, a genomic region must be sampled multiple times (the sequencing "depth" which defines the reliability of the sequence). This implies that large pools of DNA must be used in order to reliably represent the genome. The use of short reads sampled from a large population leads to two fundamental limitations: difficulty resolving variations and small subpopulations that are masked by population averaging, and loss of long-range information in the context of the individual genome. This is especially relevant to genomic regions that include structural variations (SVs), copy number variations (CNVs), and repetitive elements, which account for large fractions of most genomes.

These limitations are the driving force for developing new DNA mapping approaches that are able to extract high-resolution data from individual chromosomes. Two recent sequencing technologies promise to deliver extremely long reads from single DNA molecules; single-molecule real-time (SMRT) sequencing is a method, developed by Pacific Bioscience (Inc.), in which single DNA molecules are sequenced by measuring the fluorescent signal of nucleotides being incorporated into a DNA template by DNA polymerase. Measuring the elongation rate also enables the detection of chemical modifications of the DNA bases.[9] Another emerging single-molecule DNA sequencing method is nanopore sequencing. Here, DNA is threaded through a small pore, and the ionic current across the pore is measured during DNA translocation. This measure gives information about translocated base, including identification of different DNA modifications.[10] These methods, if they deliver on their theoretical potential, will provide long reads from single DNA molecules with the potential of addressing several epigenetic modifications in conjunction with sequence data. However, it is not foreseen that multiple layers of information, including protein occupancy, will be accessible simultaneously. Such multiplexing may be possible using optical mapping techniques which rely on the imaging of individual, long (50–1000 kbp) DNA molecules.[11−19] In optical mapping methods, the extraction of genomic information is mediated by fluorescent labeling of the DNA[20] and optical detection of these labels along single DNA molecules. Super-resolution localization techniques may be used to enhance mapping precision.[19,21−24] The data acquired using these techniques lack the high resolution of DNA sequencing but offer genomic context and are therefore ideal for aiding sequence assembly when used in combination with DNA sequencing[25−30] as well as for analysis of genomic structural variations at the individual chromosome level.[31,32]

The Complexity of the Genome. The basic nucleic DNA sequence is only one layer of information embedded in the genome. Additional genomic content resides in

VOCABULARY: FISH - fluorescence *in situ* hybridization is a technique for the detection of DNA sequences on chromosomes using fluorescently labeled probes; **Fiber FISH** - a modified FISH method in which the studied DNA is linearized on a surface; **DNA extension** - the process by which coiled DNA is transformed to a linear conformation; this can be achieved by stretching the DNA on a surface or in suspension; **epigenetics** - all inherited DNA and chromatin modifications that are not encoded in the DNA sequence; **chromatin** - the composition of DNA and its associated proteins; **chemical DNA modifications** - chemical modifications of any of the four DNA building blocks, A, C, G, and T; C methylation is the most common of these modifications in mammalian genomes; **ChIP** - chromatin immunoprecipitation is a method for capturing protein-associated DNA by use of chromatin-specific antibodies;

modifications such as DNA methylation and DNA-binding proteins, including the histone code, RNA polymerases (RNAPs), TFs, and many other DNA-binding proteins that control genomic structure and function and contribute to a highly diverse genomic content. For example, as reviewed by Xie *et al.*,[33] it is estimated that one *Escherichia coli* cell contains on average 4.6 Mbp of chromosomal DNA, 10−20 units of DNA polymerase III, 50 units of DnaG primase, 200−2000 actively transcribing RNAPs, 1000−7000 single-strand DNA-binding proteins, and a total of 50 000−200 000 units of various nucleotide-related proteins. The complexity of DNA−protein interactions stem from both the high number of DNA-binding proteins as well as the fact that many proteins bind DNA at multiple sites. For example, Bulyk and co-workers studied the diversity and complexity of 104 mouse DNA-binding proteins and found that about half of the studied TFs could bind multiple binding sites.[34] Nevertheless, each protein had a unique DNA-binding preference, suggesting that predicting protein-binding profiles according to DNA recognition sequences alone is far from being enough for elucidating the DNA−protein network.

Epigenomic Bulk Studies. Current knowledge on the protein content of the genome is available largely from gel shift assays, *in vivo* footprinting,[35] chromatin immunoprecipitation (ChIP),[36] ChIP in combination with DNA microarrays (ChIP-chip),[7] protein-binding microarrays,[37] nuclear run-on techniques,[38,39] and bioinformatic predictions.[40−42] Recent advances in array and sequencing technologies allow genome-wide studies of chromatin modifications. In particular, histones and their post-translational modifications serve as key epigenetic marks that are extensively mapped on the genomic scale due to their role in gene expression and in chromatin packaging.[7] The dynamic nature of chromatin structure serves as an important genomic regulator, where active genes are exposed for transcription and inactive genes are concealed within the

chromatin bundle. The use of digestion enzymes such as DNase I, which digest the active exposed regions in live cells, followed by DNA analysis, allows studying the dynamics of chromatin structure and gene regulation.[43]

One of the factors that influence protein binding to DNA is the degree of genome methylation.[44] In mammals, DNA methylation occurs mainly on cytosines in CpG dinucleotides. CG-rich areas of the genome, which are called CpG islands, are usually unmethylated. DNA methylation is generally associated with transcriptional repression mediated by methyl-binding proteins.[45] Mapping of methylation sites can be done using restriction enzymes that are sensitive to methylation state, by affinity purification using methylcytosine DNA-binding domain (MBD) proteins, by immunoprecipitation using anti-methylcytosine antibodies or by bisulfite-based techniques, a chemical that converts cytosines to uracils but does not react with methylcytosine.[7] Recently, a new DNA modification was discovered in mammalian genomes, hydroxymethylcytosine (5hmC).[46] Cytosine hydroxymethylation may be a mediator of DNA demethylation pathways[47,48] and was shown to have a tissue-specific distribution.[49] Methods for mapping 5hmC sites are mostly based on selective enzymatic glucosylation of 5hmC by the T4 β-glucosyltransferase enzyme,[49] a process that allows for chemical manipulation and capture of hydroxylated DNA molecules for sequencing. A recent chemoenzymatic approach was able to map 5hmC with single-base resolution.[50] Despite the wealth of information generated by these techniques, they suffer from the same drawbacks that limit genetic analysis and provide an averaged view of the epigenome.[51]

The decoration of DNA with DNA-binding proteins and DNA methylation is a dynamic process evolving through the differentiation and growth of cells and the exposure to changes in external stimuli. Thus, it is likely that neighboring cells will have different patterns of proteins and methylation sites along their chromosomes.[52] In order to reveal the composite heterogeneity and to overcome the averaging effect of ensemble methods, a single-molecule approach is needed. The long-range data offered by optical mapping may provide access to information such as the distribution of DNA-binding proteins along the genome and methylation patterns. Moreover, a single-molecule approach enables multiplex detection of a number of genetic or epigenetic markers simultaneously. Multiplexed measurements are only rarely applicable in bulk studies, and usually, no more than two observables can be studied simultaneously.[53−55] The ability to detect subpopulations and to image long-range epigenetic patterns such as cooperative binding of proteins to DNA is a major advantage of the single-molecule approach.

Epigenomic Single-Molecule Studies. *Imaging of Single-Molecule Protein−DNA Complexes.* Single-molecule studies of DNA−protein interactions are mainly devoted to
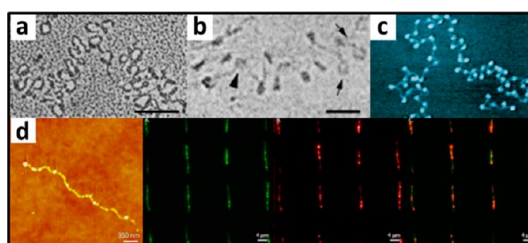

Figure 1. Single-molecule imaging of nucleosomes. (a) Chromatin containing histone H1 was imaged by EM. Adapted from Thoma et al.,[66] with permission. Copyright 1979 Rockefeller University Press. (b) Chromatin as seen by cryo-EM. Reprinted with permission from ref 69. Copyright 1998 National Academy of Sciences. (c) Cryo-AFM image of chicken erythrocyte chromatin fiber on mica. Reprinted with permission from ref 68. Copyright 2002 Springer. (d) Chromatin array: left panel, HeLa chromatin as imaged by AFM. The chromatin was labeled with Alexa-Fluor 647 histone H3 antibodies and YOYO-1. In green, fluorescence micrograph taken at 475 nm excitation; in red, fluorescence micrograph from the same area taken at 620 nm. Last panel shows the overlay of the two fluorescence micrographs, demonstrating that histones H3 are colocalized with DNA. Adapted from ref 70. Copyright 2012 American Chemical Society.

two main themes: (1) revealing the mechanism and dynamics of protein−DNA interactions and (2) mapping the occupancy and distribution of proteins along the studied DNA molecule. The first includes the characterization of protein diffusion along DNA molecules (sliding, hopping, intersegmental transfer, rotation around the helix) and measuring the association rates, step size, processivity, and efficiency of enzymes associated with DNA.[56,57] The main methods used for this purpose are atomic force microscopy, optical tweezers,[58,59] magnetic tweezers,[58] DNA curtains,[60,61] microfluidic devices,[62] molecular combing, and glass microneedles (micropipet).[63] This review focuses on static protein−DNA interaction studies which are more suitable for understanding *where* proteins are bound along DNA rather than *how* they are bound.

The motivation to understand chromatin structure of nucleosomal DNA−histone complexes led to the first single-molecule studies using electron microscopy (EM).[64−66] Advanced attempts for better visualization of nucleosomes were achieved using atomic force microscopy (AFM), cryo-AFM,[67,68] and electron cryo-microscopy.[69] Craighead and co-workers have recently demonstrated a method to form an ordered array of stretched chromatin molecules. They used both AFM and fluorescence imaging to detect the presence of histones bound to genomic DNA[70] (Figure 1). Although chromatin was imaged almost 40 years earlier to this work, their new approach presents new opportunities for studying chromatin. About 250 000 genomic fragments (from HeLa or M091 cells) were stretched and aligned using a combination of soft lithography and capillary force to pattern DNA on (3-Aminopropyl)-triethoxysilane (APTES)-coated coverslips. The fact that the chromatin is spread in an ordered array rather than

distributed randomly on the surface opens the way for high-throughput automatic imaging and processing. In order to image the aligned nucleosomes, the DNA was stained with the intercalating dye YOYO-1 and histones were labeled with specific antibodies conjugated to the organic dye Alexa-Fluor 647.

The multitude of DNA-binding proteins and the structural complexity of the genome render chromatin analysis difficult both experimentally and computationally. Methods for stretching DNA, labeling of desired elements, and data analysis are all important aspects of single-molecule mapping of DNA modifications and DNA—protein interactions. We will first discuss current approaches to these challenges followed by a survey of recent applications.

DNA Extension. Extending genomic fragments into a linear form is essential for the optical detection and localization of tags along the DNA molecule. This experimental approach was first introduced in the 1990s when chromosome stretching was used for fluorescence *in situ* hybridization (FISH) with a method known as "fiber-FISH".[71] However, accurate measurement of the DNA length and precise localization of protein positions on the DNA require reproducible and uniform stretching. Several methods for DNA extension were developed, each bearing its pros and cons, as reviewed by Dorfman *et al.*[72] In general, DNA is either stretched on a solid support or kept stretched in solution. Stretching by deposition on a surface includes the following methods: (1) attaching the DNA to a glass surface functionalized with positive charges. In this case, stretching is induced by applying flow, and DNA is fixed to the surface *via* electrostatic interactions (*e.g.*, with positively charged amines from polylysine[12] or APTES).[73] Here, only partial extension is achieved (about 85%) leading to nonuniformity in the extension factor along the molecules. (2) DNA molecular combing.[13,74] In this method, a hydrophobic surface is brought in contact with a solution containing DNA molecules (for example, by dipping a silanized glass coverslip into a DNA solution). The surface attracts DNA extremities through hydrophobic interactions with the exposed bases, and the rest of the DNA molecule can be extended by pulling the surface out of solution. Stretching forces from the air—water interface contact line cause the DNA to extend uniformly across the substrate. This approach yields very uniform stretching, in which the DNA length is extended up to 1.6 times its B-form DNA length. A recent development of molecular combing uses a microneedle to pull out DNA molecules from solution into air and stretch them mechanically on a desired surface. Despite the low-throughput of the method, it facilitates imaging on clean surfaces isolated from the sample and allows one to choose the properties of the surface.[75] Approaches involving DNA stretching without fixation include (1) DNA stretching in nanochannels, driven by confinement due to the small dimensions of the channels;[76,77] (2) stretching by confinement in nanoslits;[78–80] and (3) stretching by stagnation point flow.[81]

The two main difficulties in the preparation of extended DNA samples are the handling of long DNA molecules prior to the extension process and achieving uniform stretching. Long DNA molecules are very fragile and should be handled with special care. Methods for automated manipulation of genomic DNA may aid in that regard in the future. In addition, all DNA extension methods may suffer from nonuniform stretching along the molecule, and stretching factors may vary between different experiments. Future developments would wishfully allow uniform and robust extension of DNA molecules.

Labeling Agents. Optical visualization of information along the DNA requires a detection method with high optical contrast. Fluorescent probes are the immediate candidates for labeling in this case. In general, since mapping experiments usually require a single "snapshot" of the sample, the desired probes should emit the maximum number of photons in the shortest amount of time, and photostability is only required for the duration of a single shot of the imaging camera (as opposed to dynamic studies which require tracking fluorescence for extended periods). The desired probe ought to have a high extinction coefficient, high quantum yield, short fluorescence lifetime, and narrow emission bands. Such combined properties allow for rapid acquisition of multiple fields of view for high-throughput analysis. High photon flux is also desirable for super-resolution localization, which is only limited in resolution by the number of detected photons. However, if multiple fluorophores are positioned in close proximity (smaller than the diffraction limit), resulting in overlapping fluorescence signals, then photoswitching or blinking of the probes is also required.

Three main classes of fluorescent probes are fluorescent proteins, organic dyes, and quantum dots (QDs). A detailed review on fluorescence probes can be found at Martin-Fernandez and Clarke.[82] Fluorescent proteins (such as GFP) are large (~30 kDa), have poor brightness, and tend to bleach faster than organic dyes and QDs and therefore are not ideal for single-molecule optical imaging. In contrast, both organic dyes and QDs are more promising as labeling reagents. Numerous photostable bright organic dyes with diverse excitation and emission wavelength, ranging between 400 and 800 nm, are commercially available (reviewed by Solomatin and Herschlag).[83] Two properties of organic dyes that make them specifically attractive for labeling are their small size and the variety of their available forms, including diverse functional groups (*e.g*, amino-reactive dyes and sulfhydryl reactive dyes). Organic dyes can be used as single

molecules attached directly to a studied target or by using nanoparticles that encapsulate up to hundreds of dye molecules. QDs are fluorescent semiconductor nanocrystals with tunable emission color controlled by the dimensions of the particle through quantum confinement. QDs are characterized by relatively narrow emission bands[84,85] and are therefore useful in multicolor imaging experiments. Moreover, QDs are also remarkably bright and photostable. The main drawback of commercially available QDs is their relatively large size, about 20 nm in diameter. The large size reduces the mobility of QDs and reduces binding efficiency. In addition, a bound QD masks a region of a few tens of base pairs along the DNA, thus limiting the labeling density.

Future development in organic chemistry and dye engineering would hopefully further modify the properties of currently used dyes, especially in terms of high and stable photon flux and in the ability to make photophysical manipulations, such as photoswitching and blinking. Such improvements will serve to allow super-resolution imaging and improved resolving power between proximal labels. Multiplexing of several genomic observables by the use of multiple colors is one of the attractive features of optical mapping. However, the available optical spectral window is limited by the intrinsic spectral width of available fluorophore emission. In practice, only about four labels can be used simultaneously in the same experiment due to spectral overlap. Future development of dyes in the near-IR region as well as the use of physical phenomena such as energy transfer would allow increasing the number of colors used simultaneously.

Data Analysis. The linear extension of DNA simplifies the localization of molecular entities along DNA strands and lends itself to automated image analysis for large-scale, high-throughput measurements.[86] When analyzing short genomes such as bacteriophage genomes, DNA-bound proteins may be mapped by determining their distance from the DNA terminals. First, the overall size of the DNA should be measured, and accordingly, the degree of extension can be calculated. The measured distance between each labeled protein and the DNA extremity can be calibrated according to the calculated stretching factor. Another aspect of data analysis involves the orientation of the mapped objects; should the map be built from 5′ to 3′ or *vice versa*? In cases where the observed experimental pattern is compared to a theoretical reference, one orientation can be chosen over the other based on the expected positions. Preferably, a sequence-specific marker may be designed to identify the underlying DNA molecule and its orientation. Furthermore, the incorporation of multiple sequence-specific tags at known positions may contribute to a more precise localization of the mapped object by providing better evaluation of the DNA stretching factor.

Recent years have seen great progress in algorithmic approaches to both image analysis and sequencing related data analysis. Similar approaches should be adapted for data analysis of optical mapping experiments. There is a need for algorithms dedicated to extraction of genomic information from visual data as well as for performing complex alignment of such data.

Mapping of DNA-Binding Proteins. To date, only a fairly small number of single-molecule protein—DNA-binding studies have been conducted on extended DNA. Among these studies are imaging of C1 complex proteins bound to the T4 bacteriophage genome[87] and binding of GINS complex proteins to genomic DNA.[88] The latter demonstrated the detection of up to three proteins simultaneously; however, mapping was not conducted as part of these studies. The ability to pinpoint the location of a bound protein in the context of its genomic template is essential for our understanding of epigenetic function. The challenge of relating the location of detected proteins to the underlying genetic code is complex and was addressed by several single-molecule mapping reports.

Li and Yeung reported on the visualization of DNA restriction enzyme complexes in which the protein was bound to one expected locus.[89] *Lambda* phage DNA (48 510 bp) was stained with YOYO-1 in order to visualize the DNA backbone. The restriction enzyme ApaI was labeled with Alexa-Fluor 532 emitting at a separate spectral window. DNA extension on an untreated surface was induced by moderate flow (this was feasible owing to very specific buffer conditions). Each bound enzyme was localized at approximately 1/5 of the DNA contour, in agreement with the known restriction site (10 087 bp) of ApaI (Figure 2a). DNA digestion was avoided due to lack of Mg^{2+} ions. Here, mapping is relatively simple as the 50 kbp genome is imaged intact, and detected fluorescence may be localized relative to the DNA extremities. The fact that the expected binding site forms an asymmetric pattern allows mapping even with low-resolution data.

Taylor *et al.* used fluorescent nanoparticles (latex nanobeads) in order to detect DNA-binding proteins on *Lambda* DNA. These 20 nm wide beads emit bright and stable fluorescence since each nanoparticle contains about 100—200 molecules of dye. The dye is protected from the outside environment, and is thus highly resistant to bleaching. Histone proteins or *Eco*RI restriction enzymes were covalently attached to the nanobeads. Using inverted fluorescence microscopy, the beads could be detected along stretched DNA molecules, demonstrating the nonspecific binding of histone—bead conjugates to *Lambda* phage DNA and specific binding of *Eco*RI—bead conjugates at expected positions along the DNA.[90] The addition of EDTA to the solution allowed binding of *Eco*RI to its recognition sites but inhibited its catalytic activity. In order to determine the position of each bound
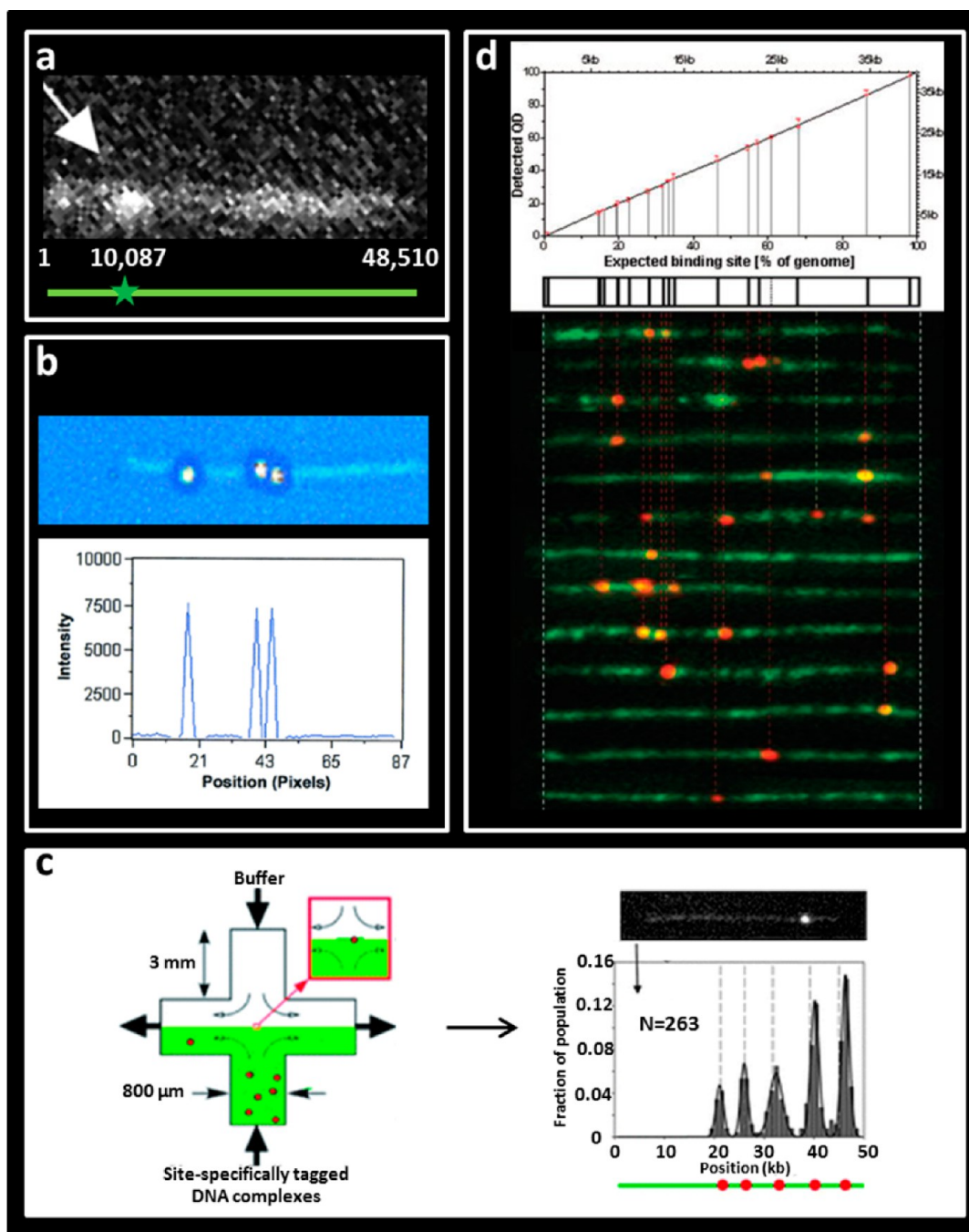
Figure 2. Single-molecule mapping of DNA-binding proteins. (a) ApaI restriction enzyme, labeled with Alexa-Fluor 532, bound to a single recognition site on the *Lambda* phage genome. Adapted from ref 89. Copyright 2005 American Chemical Society. **(b)** Mapping of five *Eco*RI binding sites using *Lambda* DNA extended on polylysine surface. Reprinted from ref 90. Copyright 2000 American Chemical Society. **(c)** At a stagnation point using a microfluidic device. Reprinted with permission from ref 81. Copyright 2010 Royal Society of Chemistry. **(d)** RNAPs mapped on T7 genome. RNAPs were conjugated to QDs, and DNA molecules were extended on polylysine surface. Reprinted from ref 22. Copyright 2009 American Chemical Society.

particle, a normalization procedure was used. Instead of using the absolute distance between *Eco*RI—nanobead and the DNA extremities, this distance was divided by the total DNA length, accounting for variation in stretching factor between DNA molecules. Under the assumption of uniform stretching along the DNA molecule, the normalized values should remain constant. Indeed, the measured locations of the five *Eco*RI binding sites were in good agreement with the theoretical positions. One exception for this observation was the mapping of site number, one which is adjacent to the DNA terminus and was poorly mapped due to the tendency of the ends of the DNA to coil, causing inaccurate measurements (Figure 2b).

Muller and co-workers also mapped *Eco*RI binding sites on the *Lambda* genome.[81,91] They used a

microfluidic device to extend DNA molecules at a stagnation point by applying equal flow in two opposite directions (see Figure 2c). Here, they used a biotinylated EcoRI to conjugate the protein to avidin-coated fluorescent 40 nm spheres. After staining with YOYO-1, DNA−protein complexes were imaged using a fluorescence microscope. Analysis of DNA images resolved all five known restriction sites of EcoRI with an average accuracy of <1 kbp. Mapping results indicated that the localizations of sites near the end of the DNA were less precise, as was redemonstrated by Muller's group also for surface-stretched DNA.[91]

Together, the two studies convey an important message: the extension of DNA is not uniform and is a critical factor determining the mapping precision, especially near the DNA extremities. Improvements in DNA extension methods and the addition of internal calibration markers that report on local stretching parameters are two of the directions taken to enable more precise localization of genomic information.

Due to their key role in gene expression, RNAPs were the subject of several single-molecule mapping studies. For example, E. coli RNAP was studied interacting with DNA curtains of Lambda genomes,[92] and transcription was mapped by visualization of fluorescent RNA synthesized by T7 RNAP on the T7 genome.[93,94] A series of studies from our lab aimed to precisely and directly map the positions and occupancy of T7−RNAP binding in a genomic context.[21−23] RNAPs with biotin tags were labeled with streptavidin−QDs. Stable DNA−RNAP complexes were achieved using stalled transcription via lack of dATP. A sample containing DNA−RNAP−QD complexes was stained with YOYO-1 and stretched on a polylysine-functionalized surface, revealing stretched DNA molecules decorated with fluorescence spots from RNAP−QD bound to the DNA. To demonstrate the mapping accuracy of the optical measurement, the mean position and standard deviation of detected QDs were plotted against the known promoter sites, as shown in Figure 2c. QD mapping was very accurate: 87% (N = 199) of QDs were found to be within 1 kbp, 50% within 398 bp, and 25% within 174 bp of a promoter.

One advantage of this single-molecule approach is the ability to directly detect the relative occupancy of binding sites under various conditions. The T7 genome (40 kbp) contains 17 T7−RNAP recognition sites, each 23 bp long.[95] Three times more binding events were detected in regions corresponding to the consensus binding sites relative to binding sites with nonconsensus sequences.[96] Three sites had remarkable occupancy: the promoter located at 86% of the full genome length, at 46.4 and at 61%, the latter known to be a strong terminator. Review of the literature did not yield any reported explanation for this higher occupancy, suggesting that this observation may be of novel biological significance. This simple experiment thus indicates

that the single-molecule approach may yield insightful results even in relatively well-known systems such as T7.

Improving Mapping Performance Using Genomic Tags. Despite the use of super-resolution localization, offering localization of protein−QD signals to within 30 bp, the overall mapping precision was far poorer, on the order of 1 kbp. This again emphasizes the crucial role of DNA extension in these experiments. A possible strategy to decrease the influence of nonuniform stretching on data analysis could be by introducing sequence-specific reference tags (RefTags). RefTags with defined spacing can serve as internal calibration marks which can be used for better fitting of the data. In addition, sequence-specific RefTags can also be useful for the analysis of longer genomes such as bacterial or mammalian DNA by providing a unique fluorescent "barcode" along the DNA.[23]

In recent years, several approaches for genome tagging have been developed; Das et al.[16] used nick translation to incorporate fluorescent nucleotides and to create a sequence-specific optical barcode along stretched DNA.[15] A second approach for the incorporation of RefTags uses methyltransferase (MTase)-modified enzymes. The modified enzymes can use synthetic cofactors for sequence-specific DNA labeling (SMILing DNA),[97] leading to a unique optical pattern.[19] We used the SMILing DNA method to incorporate three biotin tags and create an asymmetric pattern on T7 genomes using M.BseCI MTases.[98,99] The biotin moieties were further labeled with streptavidin−QDs. A schematic representation of the experimental concept is depicted in Figure 3a. Fluorescence imaging of stretched DNA molecules labeled with RefTags can be found in Figure 3b.

Following the formation of the unique barcode on T7 genomes, the labeled DNA molecules were incubated with RNAPs to form DNA−RNAP−QD complexes. Figure 3c shows a color overlay image of a T7 genome carrying both RNAPs (green) and RefTags (red). RefTags were used to identify the orientation of the DNA and to calculate the local DNA stretching factor. The mapping precision was improved only for T7−RNAPs detected between two RefTags such as those bound to promoter Φ13. Figure 3d shows position histograms for RNAP detected on this promoter. In comparison, histograms generated from the same data without using the RefTags, but relying on the DNA extremities for mapping, are presented (left). RefTags indeed improved the precision of QD localization; the width of the distribution was significantly reduced, and the precision was improved 5-fold, from ∼1.5 kbp to ∼310 bp. This precision compares favorably to the precision of ChIP-chip data.

Single-Molecule Mapping of DNA Methylation Sites. First attempts to build a single-molecule optical methylation map have been made using ordered restriction
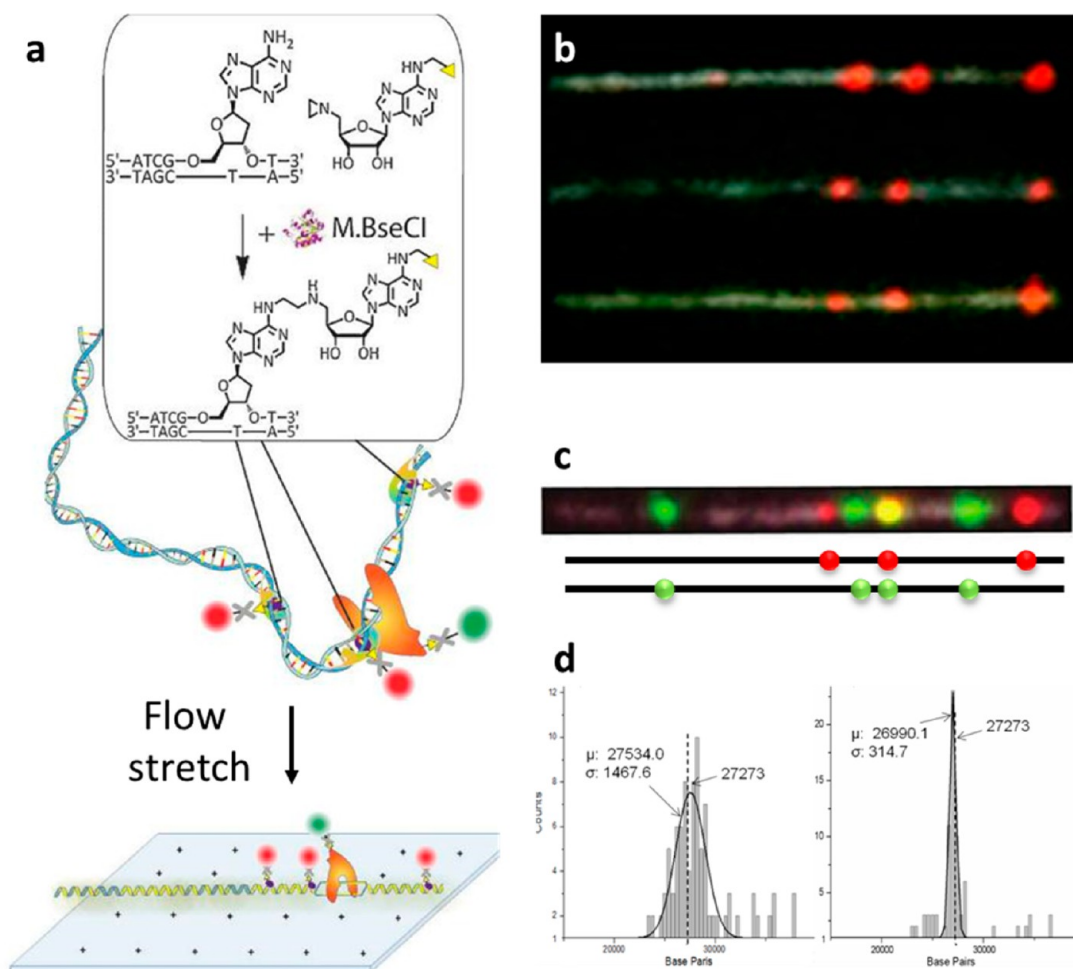
Figure 3. Incorporation of RefTags for improving protein mapping performance. (a) Schematic representation of QD-labeled RNAP bound to sequence-specific-labeled T7 bacteriophage DNA. **(b)** Image of flow-stretched, YOYO-1-stained T7 bacteriophage DNA (gray) with QD-labeled M.BseCl RefTags (red) and **(c)** T7 genome (green) with RefTags (red) and RNAPs labeled with spectrally distinct QDs (green). Overlapping red and green signals are shown in yellow. **(d)** Histograms of the detected locations of RNAP bound to promoter Φ13. Gaussian fit for localized RNAP on T7 bacteriophage using distance measurement to the DNA ends (left) *versus* localization by RefTags (right). Dotted lines represent the expected position of the Φ13 promoter. Histograms using RefTags yield a ∼5-fold increase in accuracy as evidenced by a sharp reduction in the width of the promoter localization distributions. Sigma units are in bp. Modified with permission from ref 23. Copyright 2012 John Wiley & Sons.

mapping (Figure 4a). In this method, DNA is extended on a surface and then digested using sequence-specific restriction enzymes. Enzymatic restriction mapping is a powerful method, developed by Schwartz and co-workers, and was already used to aid in *de novo* sequencing of full genomes, including the recently published goat genome.[28,30] A modified version of this approach established it also as a potent tool for epigenetic analysis of DNA methylation.[100] Here, ordered restriction maps were built using methylation-sensitive restriction enzymes, and methylation was detected as the absence of an expected cut. This method was used to map methylation sites in specific loci of human embryonic stem cells.

Another approach is to utilize the specific molecular recognition of some proteins to DNA modifications such as methylcytosine. Riehn and co-workers used Alexa-Fluor-568-labeled MBD proteins to detect regions of DNA methylation (Figure 4b). They used a mixture of methylated *Lambda* DNA with unmethylated *Lambda* DNA to form hybrid concatamers.[101] The hybrid concatamers were imaged using a fluorescence microscope that allowed detection of MBD binding patterns and DNA fluorescence in two separate emission channels, revealing the pattern of methylated *versus* unmethylated segments. In another report, QD-immobilized MBD was used for single-molecule mapping of methylated DNA (Figure 4c).[102] In their work, Baba and co-workers incorporated five methylation sites onto unmethylated *Lambda* genomes using *Bam*HI MTase. They were able to resolve four out of the five methylation sites by detecting MBD-QD fluorescence along the DNA molecules. These experiments represent a major step toward single-molecule mapping of methylation patterns; however, since large fractions of many bacterial, plant, and mammalian
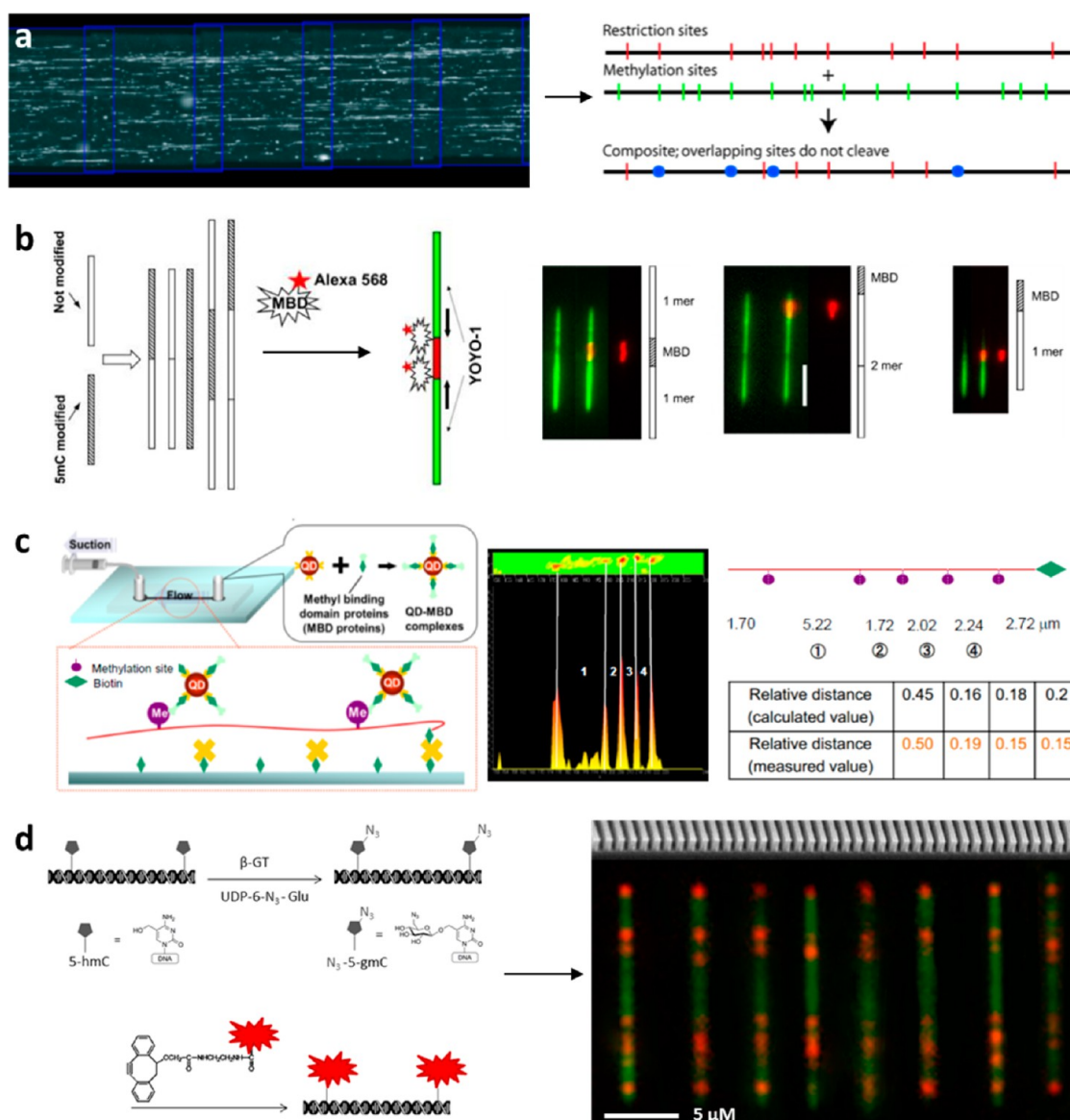
Figure 4. Single-molecule mapping of DNA methylation sites. (a) Enzymatic restriction mapping of methylation sites. The applicability of the method was demonstrated on human embryonic stem cell DNA. Reprinted from ref 100. (b) Mapping of methylation sites using MBD proteins labeled with Alexa-Fluor 568 dye, as seen on *Lambda* DNA concatamers. In green, DNA stained with YOYO-1; in red, methylated *Lambda* DNA bound to labeled MBD proteins. Reprinted with permission from ref 101. Copyright 2011 American Institute of Physics. (c) Mapping of methylation sites using MBD proteins labeled with QDs. Methylcytosines were incorporated onto *Lambda* genome at known sites using a Mtase enzyme and were detected using QD-labeled MBD proteins. Adapted from Okamoto *et al.*,[102] with permission. (d) Covalent labeling of 5hmC sites with a fluorescent dye for single-molecule mapping in nanochannels. 5hmC sites were incorporated into the *Lambda* genome at known sites using nick translation. A glucosyltransferase enzyme was used to attach an azido-modified sugar at each 5hmC site that was further labeled with an alkyne-modified Alexa-Fluor dye by a click reaction (Ebenstein lab, unpublished).

genomes are methylated, it is expected that a large amount of MBD proteins will be required for optical mapping.

Recently, we have demonstrated single-molecule mapping of 5hmC sites by covalent chemical labeling of a fluorescent reporter molecule to the modified base. Mapping was demonstrated by engineering a specific hydroxymethylation pattern in *Lambda* DNA. 5hmC nucleotides were incorporated into the *Lambda* genome in 10 known sites using nick translation with Nt.BspQI. Next, using T4 β-glucosyltransferase, an

azido-modified glucose was attached to 5hmC sites. The presence of an azide moiety allowed us to label each of the modified sites with an alkyne-modified Alexa-Fluor dye by a copper-free click chemistry reaction. YOYO-1-stained DNA was extended on a modified coverslip and imaged using fluorescence microscopy. Individual fluorescent labels on *Lambda* genomes were mapped at expected positions according to known Nt.BspQI recognition sites. In order to overcome some of the stretching inhomogeneity, similar samples were also extended in silicon nanochannels,

as shown in Figure 4d. In addition, we showed that this method was sufficient to detect natural 5hmC sites on stretched genomic DNA extracted from mouse tissues. These results open up new avenues for single-molecule epigenetic mapping relying on the robustness of covalent labeling.

SUMMARY

Overall, we discussed the basic principles for reliable mapping of epigenetic marks along genomic DNA. We emphasize the importance of the sequence-specific reference tags for internal calibration and genetic barcoding. These may also allow mapping structural variations in genomic DNA by visualizing the physical pattern of short sequence motifs along DNA.[29] When combined with the visualization of an additional layer of information such as protein-binding sites, optical mapping provides the contextual information lacking in bulk assays such as DNA arrays or sequencing. Specifically, by investigating such patterns over long, individual DNA molecules, new information regarding the cooperative nature of certain binding proteins and epigenetic DNA modifications, as well as variations within individual chromosomes, may be examined. Standard optical mapping approaches yield resolution of about 1 kb, limiting the resolvable label density. Super-resolution methods can improve the resolution to about 100 bp,[20] potentially increasing the information content attainable by optical mapping. This may enhance our ability to elucidate the presence of rare subpopulations that are otherwise obscured by ensemble averaging. Early detection of rare events may facilitate targeted and early medical intervention and may prove to be of particular relevance for diagnostic and medical monitoring purposes. Advances in micro- and nanomanipulation bring forward the prospect of single-cell analysis. An exciting future direction for optical mapping is an integrated approach for processing, labeling, and mapping of genomes from individual cells.[103]

The field of single-molecule epigenomics is in its infancy, and further development is needed in order to achieve the goal of resolving the epigenetic composition of the genome (identity, layout, and occupancy). Nevertheless, progress in nanofabrication and optical imaging promises to boost research in this direction toward a high-resolution, high-content view of the genome and its composition on the single-molecule level.

REFERENCES AND NOTES

1. Watson, J. The Human Genome Project: Past, Present, and Future. *Science* **1990**, *248*, 44–49.
2. Sanger, F.; Coulson, A. R. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *J. Mol. Biol.* **1975**, *94*, 441–448.
3. Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463–5467.
4. Ansorge, W. J. Next-Generation DNA Sequencing Techniques. *Nat. Biotechnol.* **2009**, *25*, 195–203.
5. Mardis, E. R. The Impact of Next-Generation Sequencing Technology on Genetics. *Trends Genet.* **2008**, *24*, 133–141.
6. Schuster, S. C. Next-Generation Sequencing Transforms Today's Biology. *Nat. Methods* **2008**, *5*, 8–16.
7. Schones, D. E.; Zhao, K. Genome-Wide Approaches To Studying Chromatin Modifications. *Nat. Rev. Genet.* **2008**, *9*, 179–191.
8. Shendure, J.; Ji, H. Next-Generation DNA Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.
9. Fang, G.; Munera, D.; Friedman, D. I.; Mandlik, A.; Chao, M. C.; Banerjee, O.; Feng, Z.; Losic, B.; Mahajan, M. C.; Jabado, O. J.; *et al.* Genome-Wide Mapping of Methylated Adenine Residues in Pathogenic *Escherichia coli* Using Single-Molecule Real-Time Sequencing. *Nat. Biotechnol.* **2012**, *30*, 1232–1239.
10. Wallace, E. V. B.; Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Donohoe, T. J.; Bayley, H. Identification of Epigenetic DNA Modifications with a Protein Nanopore. *Chem. Commun.* **2010**, *46*, 8195–8197.
11. Neely, R.; Deen, J.; Hofkens, J. Optical Mapping of DNA: Single Molecule Based Methods for Mapping Genomes. *Biopolymers* **2011**, *95*, 298–311.
12. Meng, X.; Benson, K.; Chada, K.; Huff, E. J.; Schwartz, D. C. Optical Mapping of Lambda Bacteriophage Clones Using Restriction Endonucleases. *Nat. Genet.* **1995**, *9*, 432–438.
13. Michalet, X.; Ekong, R.; Fougerousse, F.; Rousseaux, S.; Schurra, C.; Hornigold, N.; van Slegtenhorst, M.; Wolfe, J.; Povey, S.; Beckmann, J. S.; *et al.* Dynamic Molecular Combing: Stretching the Whole Human Genome for High Resolution Studies. *Science* **1997**, *277*, 1518–1523.
14. Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; *et al.* DNA Mapping Using Microfluidic Stretching and Single-Molecule Detection of Fluorescent Site-Specific Tags. *Genome Res.* **2004**, *14*, 1137–1146.
15. Xiao, M.; Phong, A.; Ha, C.; Chan, T. Rapid DNA Mapping by Fluorescent Single Molecule Detection. *Nucleic Acids Res.* **2007**, *35*, e16.
16. Das, S. K.; Austin, M. D.; Akana, M. C.; Deshpande, P.; Cao, H.; Xiao, M. Single Molecule Linear Analysis of DNA in Nano-Channel Labeled with Sequence Specific Fluorescent Probes. *Nucleic Acids Res.* **2010**, *38*, e177.
17. Reisner, W.; Larsen, N. B.; Silahtaroglu, A.; Kristensen, A.; Tommerup, N.; Tegenfeldt, J. O.; Flyvbjerg, H. Single-Molecule Denaturation Mapping of DNA in Nanofluidic Channels. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13294–13299.
18. Zohar, H.; Hetherington, C. L.; Bustamante, C.; Muller, S. J. Peptide Nucleic Acids as Tools for Single-Molecule Sequence Detection and Manipulation. *Nano Lett.* **2010**, *10*, 4697–4701.
19. Neely, R. K.; Dedecker, P.; Hotta, J.; Urbanavičiūtė, G.; Klimašauskas, S.; Hofkens, J. DNA Fluorocode: A Single Molecule, Optical Map of DNA with Nanometre Resolution. *Chem. Sci.* **2010**, *1*, 453–460.
20. Levy-Sakin, M.; Ebenstein, Y. Beyond Sequencing: Optical Mapping of DNA in the Age of Nanotechnology and Nanoscopy. *Curr. Opin. Biotechnol.* **2013**, *24*, 1–9.

21. Ebenstein, Y.; Gassman, N.; Kim, S.; Weiss, S. Combining Atomic Force and Fluorescence Microscopy for Analysis of Quantum-Dot Labeled Protein−DNA Complexes. *J. Mol. Recognit.* **2009**, *22*, 397−402.

22. Ebenstein, Y.; Gassman, N.; Kim, S.; Antelman, J.; Kim, Y.; Ho, S.; Samuel, R.; Michalet, X.; Weiss, S. Lighting Up Individual DNA Binding Proteins with Quantum Dots. *Nano Lett.* **2009**, *9*, 1598−1603.

23. Kim, S.; Gottfried, A.; Lin, R. R.; Dertinger, T.; Kim, A. S.; Chung, S.; Colyer, R. A.; Weinhold, E.; Weiss, S.; Ebenstein, Y. Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. *Angew. Chem., Int. Ed.* **2012**, *51*, 3578−3581.

24. Baday, M.; Cravens, A. J.; Hastie, A.; Kim, H.; Kudeki, D.; Kwok, P.-Y.; Xiao, M.; Selvin, P. R. Multi-color Super-resolution DNA Imaging for Genetic Analysis. *Nano Lett.* **2012**, *12*, 3861−3866.

25. Narzisi, G.; Mishra, B. Comparing *De Novo* Genome Assembly: The Long and Short of It. *PLoS One* **2011**, *6*, e19175.

26. Narzisi, G.; Mishra, B. Scoring-and-Unfolding Trimmed Tree Assembler: Concepts, Constructs and Comparisons. *Bioinformatics* **2011**, *27*, 153−160.

27. Aston, C.; Mishra, B.; Schwartz, D. C. Optical Mapping and Its Potential for Large-Scale Sequencing Projects. *Trends Biotechnol.* **1999**, *17*, 297−302.

28. Chen, S.; Xu, J.; Liu, C.; Zhu, Y.; Nelson, D. R. D.; Zhou, S.; Li, C.; Wang, L.; Guo, X.; Sun, Y.; et al. Genome Sequence of the Model Medicinal Mushroom Ganoderma Lucidum. *Nat. Commun.* **2012**, *3*, 913.

29. Lam, E. T.; Hastie, A.; Lin, C.; Ehrlich, D.; Das, S. K.; Austin, M. D.; Deshpande, P.; Cao, H.; Nagarajan, N.; Xiao, M.; et al. Genome Mapping on Nanochannel Arrays for Structural Variation Analysis and Sequence Assembly. *Nat. Biotechnol.* **2012**, *30*, 771−776.

30. Dong, Y.; Xie, M.; Jiang, Y.; Xiao, N.; Du, X.; Zhang, W.; Tosser-Klopp, G.; Wang, J.; Yang, S.; Liang, J.; et al. Sequencing and Automated Whole-Genome Optical Mapping of the Genome of a Domestic Goat (*Capra hircus*). *Nat. Biotechnol.* **2013**, *31*, 135−141.

31. Caburet, S.; Conti, C.; Schurra, C.; Lebofsky, R.; Edelstein, S. J.; Bensimon, A. Human Ribosomal RNA Gene Arrays Display a Broad Range of Palindromic Structures. *Genome Res.* **2005**, *15*, 1079−1085.

32. Kidd, J. M.; Cooper, G. M.; Donahue, W. F.; Hayden, H. S.; Sampas, N.; Graves, T.; Hansen, N.; Teague, B.; Alkan, C.; Antonacci, F.; et al. Mapping and Sequencing of Structural Variation from Eight Human Genomes. *Nature* **2008**, *453*, 56−64.

33. Xie, X. S.; Choi, P. J.; Li, G.-W.; Lee, N. K.; Lia, G. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annu. Rev. Biophys.* **2008**, *37*, 417−444.

34. Badis, G.; Berger, M. F.; Philippakis, A. A.; Talukder, S.; Gehrke, A. R.; Jaeger, S. A.; Chan, E. T.; Metzler, G.; Vedenko, A.; Chen, X.; et al. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* **2009**, *324*, 1720−1723.

35. Arthur, D.; Riggs, G. P. P. Polymerase Chain Reaction Aided Genomic Footprinting: Principles and Applications. *Adv. Mol. Cell Biol.* **1997**, *21*, 47−72.

36. Zhu, C.; Byers, K. J. R. P.; McCord, R. P.; Shi, Z.; Berger, M. F.; Newburger, D. E.; Saulrieta, K.; Smith, Z.; Shah, M. V; Radhakrishnan, M.; et al. High-Resolution DNA-Binding Specificity Analysis of Yeast Transcription Factors. *Genome Res.* **2009**, *19*, 556−566.

37. Berger, M. F.; Bulyk, M. L. Protein Binding Microarrays (PBMs) for Rapid, High-Throughput Characterization of the Sequence Specificities of DNA Binding Proteins. *Methods Mol. Biol.* **2006**, *338*, 245−260.

38. Hirayoshi, K.; Lis, J. T. Nuclear Run-On Assays: Assessing Transcription by Measuring Density of Engaged RNA Polymerases. *Methods Enzymol.* **1999**, *304*, 351−362.

39. Pelechano, V.; Chávez, S.; Pérez-Ortín, J. A Complete Set of Nascent Transcription Rates for Yeast Genes. *PLoS One* **2010**, *5*, e15422.

40. Chen, C.-Y.; Chien, T.-Y.; Lin, C.-K.; Lin, C.-W.; Weng, Y.-Z.; Chang, D. T.-H. Predicting Target DNA Sequences of DNA-Binding Proteins Based on Unbound Structures. *PLoS One* **2012**, *7*, e30446.

41. Kauffman, C.; Karypis, G. An Analysis of Information Content Present in Protein−DNA Interactions. *Pac. Symp. Biocomput.* **2008**, *13*, 477−488.

42. Ahmad, S.; Gromiha, M. M.; Sarai, A. Analysis and Prediction of DNA-Binding Proteins and Their Binding Residues Based on Composition, Sequence and Structural Information. *Bioinformatics* **2004**, *20*, 477−486.

43. Wang, X.; Simpson, R. T. Chromatin Structure Mapping in *Saccharomyces cerevisiae in Vivo* with DNase I. *Nucleic Acids Res.* **2001**, *29*, 1943−1950.

44. Tate, P. H.; Bird, A. P. Effects of DNA Methylation on DNA-Binding Proteins and Gene Expression. *Curr. Opin. Genet. Dev.* **1993**, *3*, 226−231.

45. Jørgensen, H. F.; Adie, K.; Chaubert, P.; Bird, A. P. Engineering a High-Affinity Methyl-CpG-Binding Protein. *Nucleic Acids Res.* **2006**, *34*, e96.

46. Kriaucionis, S.; Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **2009**, *324*, 929−930.

47. Zhang, P.; Su, L.; Wang, Z.; Zhang, S.; Guan, J.; Chen, Y.; Yin, Y.; Gao, F.; Tang, B.; Li, Z. The Involvement of 5-Hydroxymethylcytosine in Active DNA Demethylation in Mice. *Biol. Reprod.* **2012**, *86*, 104−1049.

48. He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333*, 1303−1307.

49. Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C.; Zhang, W.; Jian, X.; et al. Selective Chemical Labeling Reveals the Genome-Wide Distribution of 5-Hydroxymethylcytosine. *Nat. Biotechnol.* **2011**, *29*, 68−72.

50. Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; et al. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* **2012**, *149*, 1368−1380.

51. Jothi, R.; Cuddapah, S.; Barski, A.; Cui, K.; Zhao, K. Genome-Wide Identification of *In Vivo* Protein−DNA Binding Sites from ChIP-Seq Data. *Nucleic Acids Res.* **2008**, *36*, 5221−5231.

52. Chris, R. Calladine, H. D. *Understanding DNA: The Molecule and How It Works*; Academic Press: New York, 1997; pp 270−294.

53. Darst, R. P.; Pardo, C. E.; Pondugula, S.; Gangaraju, V. K.; Nabilsi, N. H.; Bartholomew, B.; Kladde, M. P. Simultaneous Single-Molecule Detection of Endogenous C-5 DNA Methylation and Chromatin Accessibility Using MAPit. *Methods Mol. Biol.* **2012**, *833*, 125−141.

54. Geisberg, J. V; Struhl, K. Quantitative Sequential Chromatin Immunoprecipitation, a Method for Analyzing Co-occupancy of Proteins at Genomic Regions *in Vivo*. *Nucleic Acids Res.* **2004**, *32*, e151.

55. Scully, K. M. Allosteric Effects of Pit-1 DNA Sites on Long-Term Repression in Cell Type Specification. *Science* **2000**, *290*, 1127−1131.

56. Gorman, J.; Greene, E. C. Visualizing One-Dimensional Diffusion of Proteins Along DNA. *Nat. Struct. Mol. Biol.* **2008**, *15*, 768−774.

57. Allemand, J.-F.; Bensimon, D.; Croquette, V. Stretching DNA and RNA To Probe Their Interactions with Proteins. *Curr. Opin. Struct. Biol.* **2003**, *13*, 266−274.

58. Monico, C.; Capitanio, M.; Belcastro, G.; Vanzi, F.; Pavone, F. S. Optical Methods To Study Protein−DNA Interactions *in Vitro* and in Living Cells at the Single-Molecule Level. *Int. J. Mol. Sci.* **2013**, *14*, 3961−3992.

59. Perkins, T. T.; Li, H.-W.; Dalal, R. V; Gelles, J.; Block, S. M. Forward and Reverse Motion of Single RecBCD Molecules on DNA. *Biophys. J.* **2004**, *86*, 1640−1648.

60. Finkelstein, I. J.; Greene, E. C. Molecular Traffic Jams on DNA. *Annu. Rev. Biophys.* **2013**, *42*, 241−263.

61. Visnapuu, M.-L.; Greene, E. C. Single-Molecule Imaging of DNA Curtains Reveals Intrinsic Energy Landscapes for

Nucleosome Deposition. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1056–1062.

62. Forget, A. L.; Dombrowski, C. C.; Amitani, I.; Kowalczykowski, S. C. Exploring Protein−DNA Interactions in 3D Using *In Situ* Construction, Manipulation and Visualization of Individual DNA Dumbbells with Optical Traps, Microfluidics and Fluorescence Microscopy. *Nat. Protoc.* **2013**, *8*, 525–538.

63. Leger, J. F.; Robert, J.; Bourdieu, L.; Chatenay, D.; Marko, J. F. RecA Binding to a Single Double-Stranded DNA Molecule: A Possible Role of DNA Conformational Fluctuations. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12295–12299.

64. Finch, J. T.; Klug, A. Solenoidal Model for Superstructure in Chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1897–1901.

65. Thoma, F.; Koller, T. Influence of Histone H1 on Chromatin Structure. *Cell* **1977**, *12*, 101–107.

66. Thoma, F.; Koller, T.; Klug, A. Involvement of Histone H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin. *J. Cell Biol.* **1979**, *83*, 403–427.

67. Shao, Z. Probing Nanometer Structures with Atomic Force Microscopy. *News Physiol. Sci.* **1999**, *14*, 142–149.

68. Zlatanova, J.; Leuba, S. H. Stretching and Imaging Single DNA Molecules and Chromatin. *J. Muscle Res. Cell Motil.* **2002**, *23*, 377–395.

69. Bednar, J.; Horowitz, R. A.; Grigoryev, S. A.; Carruthers, L. M.; Hansen, J. C.; Koster, A. J.; Woodcock, C. L. Nucleosomes, Linker DNA, and Linker Histone Form a Unique Structural Motif That Directs the Higher-Order Folding and Compaction of Chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14173–14178.

70. Cerf, A.; Tian, H. C.; Craighead, H. G. Ordered Arrays of Native Chromatin Molecules for High-Resolution Imaging and Analysis. *ACS Nano* **2012**, *6*, 7928–7934.

71. Raap, A. K. Advances in Fluorescence *In Situ* Hybridization. *Mutat. Res.* **1998**, *400*, 287–298.

72. Dorfman, K. D. K.; King, S. B. S.; Olson, D. W. D.; Thomas, J. D. P.; Tree, D. R. Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching. *Chem. Rev.* **2013**, *113*, 2584–2667.

73. Cai, W.; Aburatani, H.; Stanton, V. P.; Housman, D. E.; Wang, Y. K.; Schwartz, D. C. Ordered Restriction Endonuclease Maps of Yeast Artificial Chromosomes Created by Optical Mapping on Surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5164–5168.

74. Herrick, J.; Bensimon, A. Introduction to Molecular Combing: Genomics, DNA Replication, and Cancer. *Methods Mol. Biol.* **2009**, *521*, 71–101.

75. Payne, A. C.; Andregg, M.; Kemmish, K.; Hamalainen, M.; Bowell, C.; Bleloch, A.; Klejwa, N.; Lehrach, W.; Schatz, K.; Stark, H.; *et al.* Molecular Threading: Mechanical Extraction, Stretching and Placement of DNA Molecules from a Liquid−Air Interface. *PLoS One* **2013**, *8*, e69058.

76. Riehn, R.; Lu, M.; Wang, Y.-M.; Lim, S. F.; Cox, E. C.; Austin, R. H. Restriction Mapping in Nanofluidic Devices. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10012–10016.

77. Austin, M.; Deshpande, P.; Requa, M.; Kunkel, M.; Sadowski, H.; Bozinov, D.; Sibert, J.; Gallagher, S.; Stedman, W.; Fernandes, N.; *et al.* Single-Molecule Rapid Imaging of Linear Genomes in Nanochannel Array. *Mater. Res. Soc. Symp. Proc.* **2011**, *1346*, 1–9.

78. Reisner, W.; Larsen, N. B.; Flyvbjerg, H.; Tegenfeldt, J. O.; Kristensen, A. Directed Self-Organization of Single DNA Molecules in a Nanoslit *via* Embedded Nanopit Arrays. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 79–84.

79. Jo, K.; Dhingra, D. A Single-Molecule Barcoding System Using Nanoslits for DNA Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2673–2678.

80. Marie, R.; Pedersen, J. N.; Bauer, D. L. V.; Rasmussen, K. H.; Yusuf, M.; Volpi, E.; Flyvbjerg, H.; Kristensen, A.; Mir, K. U. Integrated View of Genome Structure and Sequence of a Single DNA Molecule in a Nanofluidic Device. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4893–4898.

81. Dylla-Spears, R.; Townsend, J. E.; Jen-Jacobson, L.; Sohn, L. L.; Muller, S. J. Single-Molecule Sequence Detection via Microfluidic Planar Extensional Flow at a Stagnation Point. *Lab Chip* **2010**, *10*, 1543–1549.

82. Martin-Fernandez, M. L.; Clarke, D. T. Single Molecule Fluorescence Detection and Tracking in Mammalian Cells: The State-of-the-Art and Future Perspectives. *Int. J. Mol. Sci.* **2012**, *13*, 14742–14765.

83. Solomatin, S.; Herschlag, D. Methods of Site-Specific Labeling of RNA with Fluorescent Dyes. *Methods Enzymol.* **2009**, *469*, 47–68.

84. Thompson, R. E.; Larson, D. R.; Webb, W. W. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophys. J.* **2002**, *82*, 2775–2783.

85. Lagerholm, B. C.; Averett, L.; Weinreb, G. E.; Jacobson, K.; Thompson, N. L. Analysis Method for Measuring Submicroscopic Distances with Blinking Quantum Dots. *Biophys. J.* **2006**, *91*, 3050–3060.

86. Jing, J. P.; Reed, J.; Huang, J.; Hu, X. H.; Clarke, V.; Edington, J.; Housman, D.; Anantharaman, T. S.; Huff, E. J.; Mishra, B.; *et al.* Automated High-Resolution Optical Mapping Using Arrayed, Fluid-Fixed DNA Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8046–8051.

87. Tissot, B.; Daniel, R.; Place, C. Interaction of the C1 Complex of Complement with Sulfated Polysaccharide and DNA Probed by Single Molecule Fluorescence Microscopy. *Eur. J. Biochem.* **2003**, *270*, 4714–4720.

88. Cohen, S. M.; Chastain, P. D.; Cordeiro-Stone, M.; Kaufman, D. G. DNA Replication and the GINS Complex: Localization on Extended Chromatin Fibers. *Epigenet. Chromatin* **2009**, *2*, 6–614.

89. Li, H.-W.; Yeung, E. S. Direct Observation of Anomalous Single-Molecule Enzyme Kinetics. *Anal. Chem.* **2005**, *77*, 4374–4377.

90. Taylor, J. R.; Fang, M. M.; Nie, S. Probing Specific Sequences on Single DNA Molecules with Bioconjugated Fluorescent Nanoparticles. *Anal. Chem.* **2000**, *72*, 1979–1986.

91. Dylla-Spears, R.; Townsend, J. E.; Sohn, L. L.; Jen-Jacobson, L.; Muller, S. J. Fluorescent Marker for Direct Detection of Specific dsDNA Sequences. *Anal. Chem.* **2009**, *81*, 10049–10054.

92. Wang, F.; Redding, S.; Finkelstein, I. J.; Gorman, J.; Reichman, D. R.; Greene, E. C. The Promoter-Search Mechanism of *Escherichia coli* RNA Polymerase Is Dominated by Three-Dimensional Diffusion. *Nat. Struct. Mol. Biol.* **2013**, *20*, 174–181.

93. Gueroui, Z.; Place, C.; Freyssingeas, E.; Berge, B. Observation by Fluorescence Microscopy of Transcription on Single Combed DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6005–6010.

94. Wu, T.; Schwartz, D. C. Transchip: Single-Molecule Detection of Transcriptional Elongation Complexes. *Anal. Biochem.* **2007**, *361*, 31–46.

95. Dunn, J. J.; Studier, F. W. Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements. *J. Mol. Biol.* **1983**, *166*, 477–535.

96. Smeekens, S. P.; Romano, L. J. Promoter and Nonspecific DNA Binding by the T7 RNA Polymerase. *Nucleic Acids Res.* **1986**, *14*, 2811–2827.

97. Pljevaljčić, G.; Schmidt, F.; Weinhold, E. Sequence-Specific Methyltransferase-Induced Labeling of DNA (SMILing DNA). *ChemBioChem* **2004**, *5*, 265–269.

98. Wilkinson, S.; Diechtierow, M.; Estabrook, R. A.; Schmidt, F.; Hüben, M.; Weinhold, E.; Reich, N. O. Molecular Scale Architecture: Engineered Three- and Four-Way Junctions. *Bioconjugate Chem.* **2008**, *19*, 470–475.

99. Braun, G.; Diechtierow, M.; Wilkinson, S.; Schmidt, F.; Hüben, M.; Weinhold, E.; Reich, N. O. Enzyme-Directed Positioning of Nanoparticles on Large DNA Templates. *Bioconjugate Chem.* **2008**, *19*, 476–479.

100. Ananiev, G. E.; Goldstein, S.; Runnheim, R.; Forrest, D. K.; Zhou, S.; Potamousis, K.; Churas, C. P.; Bergendahl, V.; Thomson, J. A.; Schwartz, D. C. Optical Mapping Discerns

Genome Wide DNA Methylation Profiles. *BMC Mol. Biol.* **2008**, *9*, 68–6814.

101. Fang Lim, S.; Karpusenko, A.; Sakon, J. J.; Hook, J. A.; Lamar, T. A.; Riehn, R. DNA Methylation Profiling in Nanochannels. *Biomicrofluidics* **2011**, *5*, 34106–341068.

102. Okamoto, Y.; Sano, T.; Kaji, N.; Tokeshi, M.; Baba, Y. Identification of Single Molecular DNA Methylation Points by Microfluidic DNA Molecule Stretching and Quantum Dot Detection. In *16th International Conference on Miniaturized Systems for Chemistry and Life Sciences*; 2012; Vol. 3, pp 1939—1941.

103. Aguilar, C. A.; Craighead, H. G. Micro- and Nanoscale Devices for the Investigation of Epigenetics and Chromatin Dynamics. *Nat. Nanotechnol.* **2013**, *8*, 709–718.