Title
Toward Single-Molecule Optical Mapping of the Epigenome

Permalink
https://escholarship.org/uc/item/2467z7cj

Journal
ACS Nano, 8(1)

ISSN
1936-0851

Authors
Levy-Sakin, Michal
Grunwald, Assaf
Kim, Soohong
et al.

Publication Date
2014-01-28

DOI
10.1021/nn4050694

Peer reviewed

Towards single-molecule optical mapping of the epigenome

Journal:	*ACS Nano*
Manuscript ID:	Draft
Manuscript Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Levy, Michal; Tel Aviv University, Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry Grunwald, Assaf; Tel Aviv University, Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry Kim, Soohong; University of California at Los Angeles, Chemistry Gassman, Natalie; University of California at Los Angeles, Chemistry Antelman, Josh; UCLA, Chem / Biochem Kim, Younggyu; University of California at Los Angeles, Chemistry Ho, Sam; University of California at Los Angeles, Chemistry Samuel, Robin; University of California at Los Angeles, Chemistry Michalet, Xavier; UCLA, Department of Chemistry and Bio Gottfried, Anna; RWTH Aachen, Institut fır Organische Chemie Lin, Ron; UCLA, Department of Chemistry and Bio Dertinger, Thomas; UCLA, Department of Chemistry and Bio Kim, Andrew; UCLA, Department of Chemistry and Bio Chung, Sangyoon; UCLA, Department of Chemistry and Bio Coyler, Ryan; UCLA, Department of Chemistry and Bio Weinhold, Elmar; RWTH Aachen, Institut fır Organische Chemie Weiss, Shimon; University of California at Los Angeles, Chemistry Ebenstein, Yuval; Tel Aviv University, School of Chemistry

Towards single-molecule optical mapping of the epigenome

Michal Levy-Sakin[1], Assaf Grunwald[1], Soohong Kim[2], Natalie Gassman[2], Josh Antelman[2], Younggyu Kim[2], Sam Ho[2], Robin Samuel[2], Xavier Michalet[2], Anna Gottfried[3], Ron R. Lin[2], Thomas Dertinger[2], Andrew S. Kim[2], Sangyoon Chung[2], Ryan A. Colyer[2], Elmar Weinhold[3], Shimon Weiss[2,] and Yuval Ebenstein[1,]**

[1] Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry, Tel Aviv University, Tel Aviv, Israel.

[2] Department of Chemistry and Biochemistry, University of California, Los Angeles, USA.

[3]Institute of Organic Chemistry, RWTH Aachen University, Aachen, Germany.

* Corresponding authors. E-mail addresses: uv@post.tau.ac.il (Y. Ebenstein), sweiss@chem.ucla.edu (S. Weiss).

Abstract

The last 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 on the population level. The complex composition of the genome has made it an ideal system to study on 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 nano-scale and enhanced optical resolving power for analysis of genomic information. In the very last 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

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 in 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 on any one 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 immuno-precipitation is a method for capturing protein associated DNA by use of chromatin specific antibodies.

Differences 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 knowledgebase 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 (TS) 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 to 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 usage of short reads sampled from a large population leads to two fundamental limitations: Difficulty to resolve variations and small sub-populations that are masked by population averaging, and loss of long-range information in the context of the individual genome. Such regions 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 able to extract high resolution data from individual chromosomes. One such approach is optical mapping and its variants,[9–17] which rely on the imaging of individual, long (50 kbp - 1000 kbp) DNA molecules. In optical mapping methods, the extraction of genomic information is mediated by fluorescent labelling of the DNA[18] and optical detection of these labels along single DNA molecules. Superresolution localization techniques may be used to enhance mapping precision.[19–21] The data acquired using these techniques lacks the high resolution of DNA sequencing but offers genomic context and therefore ideal for aiding sequence assembly, when used in combination with DNA sequencing[22–27] as well as analysis of genomic structural variations on the individual chromosome level.[28,29]

The complexity of the genome

As mentioned, the basic nucleic DNA sequence is only one layer of information embedded in the genome. Additional genomic content resides in 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.*[30] it is estimated that one

E. coli cell contains on average 4.6 Mbp 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 proteins. The complexity of DNA-protein interaction stems from both the high number of DNA binding proteins, as well as, from the fact that many can bind DNA at multiple sites. For example, Buly 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.[31] The 104 proteins studied were members of 22 structural families. However, 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-proteins network.

Epigenomic bulk studies

Current knowledge on the protein content of the genome is available largely from gel shift assays, *in vivo* footprinting,[32] chromatin immunoprecipitation (ChIP),[33] ChIP in combination with DNA microarrays (ChIP-chip),[7] protein-binding microarrays,[34] nuclear run-on techniques[35,36] and bioinformatic predictions.[37–39] 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 genomic scale due to their role in gene expression and in chromatin packing.[7]

One of the factors that influence protein binding to DNA is the degree of genome methylation.[40] In mammals, DNA methylation occurs 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.[41] 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 bisulphite 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, hydroxymethyl cytosine (5hmC).[42] Cytosine hydroxymethylation may be a mediator of DNA demethylation pathways[43,44] and was shown to have a tissue specific distribution.[45] Methods for mapping 5hmC sites are mostly based on selective enzymatic glucosylation of 5hmC by the T4 β-glucosyltransferase enzyme,[45] a process that allows for chemical manipulation and capture of hydroxylated DNA molecules for sequencing. A recent chemo-enzymatic approach was able to map 5hmC at single base resolution.[46] 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.[47]

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 neighbouring cells will have different patterns of proteins and methylation sites along their chromosomes.[48] 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.[49–51] The ability to detect sub populations as well to image long range epigenetic patterns such as cooperative binding of proteins to DNA, are some of the major advantages of the single-molecule approach.

Imaging of single-molecule protein-DNA complexes

Single-molecule studies of DNA-protein interactions are mainly devoted to two main themes: 1) revealing the mechanism and dynamics of protein-DNA interactions and 2) mapping binding sites 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.[52,53] The main methods used for this purpose are: atomic force microscopy, optical tweezers,[54,55] magnetic tweezers,[54] DNA curtains,[56,57] microfluidic devices,[58] molecular combing and glass microneedles (micropipette).[59] This review focuses on static protein-DNA interaction studies which are more suitable for understanding where along DNA proteins are bound, rather than how they are bound.

The motivation to understand chromatin structure of nucleosomal DNA-histones complexes led to the first single-molecule studies using electron microscopy (EM).[60–62] Advanced attempts for better visualization of nucleosomes were achieved using atomic force microscopy (AFM), cryo-AFM[63,64] and electron cryo-microscopy.[65] 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[66] (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 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 labelled 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, labelling 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 for these demands and then will discuss some of their 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'.[67] 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*.[68] 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 approach stretching is induced by applying flow and DNA is fixed to the surface *via* electrostatic interactions (*e.g.* with positively charged amines from polylysine[10] or APTES[69]). Here, only partial extension is achieved (about 85%) leading to non-uniformity in the extension factor along the molecules. 2) DNA molecular combing.[11,70] 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 preferably 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 1.5 times of its B-form DNA length. Approaches involving DNA stretching without fixation include: 1) DNA stretching in nanochannels, driven by confinement due to the small dimensions of the channels,[71,72] 2) stretching by confinement in nanoslits,[73–75] and 3) stretching by stagnation point flow.[76]

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 “snap shot” 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 life-time 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 (QD). A detailed review on fluorescence probes can be found at Martin-Fernandez and Clarke.[77] 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 nm to 800 nm, are commercially available (reviewed by Solomatin and Herschlag).[78] 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 nano-particles 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,[79,80] 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.

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.[81] When analysing 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 labelled protein and the DNA extremity can be corrected according to the calculated degree of stretching. 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 a more precise localization of the mapped object by providing better evaluation of the DNA extension factor.

Mapping of DNA binding proteins

To date, only a fairly low number of single molecule protein-DNA interaction studies were conducted on extended DNA. Among these studies are imaging of C1 complex proteins bound to the T4 bacteriophage genome[82] and binding of GINS complex proteins to genomic DNA.[83] 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 epigenetic understanding. The challenge of relating the location of detected proteins to the underlying genetic code is more 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.[84] *Lambda* phage DNA (48,510 bp) was stained with YOYO-1 in order to visualize the DNA backbone. The restriction enzyme ApaI was labelled with alexa-fluor 532 emitting at a separate spectral window. A moderate flow induced DNA extension on an untreated surface (this was feasible owing to very specific buffer conditions). Each bound enzyme was localized at approximately one fifth of the DNA contour. This is 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 extremiries. 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 nano beads) 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 thus it is highly resistant to bleaching. Histone proteins or EcoRI restriction enzymes were covalently attached to the nano-beads. Using inverted fluorescence microscopy the beads could be detected along stretched DNA molecules, demonstrating the non-specific binding of histone-bead conjugates to *Lambda* phage DNA and specific binding of EcoRI-bead conjugates at expected positions along the DNA.[85] The addition of EDTA to the solution allowed binding of EcoRI to its recognition sites but inhibited its catalytic activity. Stained DNA was stretched on polylysine-coated coverslips. In order to determine the position of each bound particle a normalization procedure was used: Instead of using the absolute distance between EcoRI-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 EcoRI 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 non-accurate measurements (Figure 2b).

Muller and co-workers also mapped EcoRI binding sites on the *Lambda* genome.[76] 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 < 1 kbp. Similar to Taylor et al, they also remarked that localizations of sites near the end of the DNA were less precise, as re-demonstrated by Muller's group using polylysine surfaces.[86]

Together, the two studies provide us with 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[87] and transcription was mapped by visualization of fluorescent RNA synthesized by T7 RNAP on the T7 genome.[88] 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.[19–21] RNAPs with biotin tags were labeled with streptavidin-QDs. Stable DNA-RNAP complexes were achieved using stalled transcription via lack of ATP. 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-RNAPs recognition sites, each 23 bp long.[89] Three times more binding events were detected in regions corresponding to the consensus binding sites relative to binding sites with non-consensus sequnce.[90] 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 a non-uniform stretching on data analysis could be by introducing sequence-specific reference tags (RefTags). RefTags with defined spacing can serve as internal calibration which can be used for better fitting of the data. In addition, sequence specific RefTags can be also useful for the analysis of longer genomes such as bacterial or mammalian

DNA by providing a unique fluorescent "barcode" along the DNA.[21]

In the last years several approaches for genome tagging were developed; Das *et al.*[14] used the nick translation method to incorporate fluorescent nucleotides and to create a sequence specific optical barcode along stretched DNA.[13] A second approach was developed for the incorporation of RefTags, using methyltransferase (MTase) modified enzymes. The modified enzymes can use synthetic cofactors for sequence-specific DNA labelling (SMILind DNA)[91] leading to a unique optical pattern.[17] We used the SMILing DNA method to incorporate three biotin tags with asymmetric pattern on T7 genomes using M.BseCI MTases.[92,93] The biotin moieties were further labelled with streptavidin-QDs. A schematic representation of the experimental concept is depicted in Figure 3a. Fluorescence imaging of stretched DNA molecules labelled with RefTags can be found in Figure 3b.

Following the formation of the unique barcode on T7 genomes, the labelled DNA molecules were incubated with RNAPs to form DNA-T7-RNAP-QD complexes. Figure 3c shows color overlay images of two genomes carrying both RNAPs (white) 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. Therefore, we focused our analysis on one RNAPs binding site, the promoter Φ13 that lies between two RefTags. Figure 3d shows position histograms for RNAP detected on promoter Φ13. For 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 favourably to the precision of ChIP data.

Single molecule mapping of DNA methylation sites

First efforts to build single molecule optical methylation map was done using ordered restriction 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.[25,27] A modified version of this approach established it also as a potent tool for epigenetic analysis of DNA methylation.[94] Here, ordered restriction maps were built using methylation-sensitive restriction enzymes and methylation was detected as the absence of an expected cut. The 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 labelled MBD protein with alexa-fluor 568 to detect regions of DNA methylation (Figure 4b). They used a mixture of methylated *Lambda* DNA with un-methylated *Lambda* DNA to form hybrid concatamers.[95] 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 vs. unmethylated segments. In another report, QD immobilized MBD was used for single molecule mapping of methylated DNA (Figure 4c).[96] In their work, Baba and coworkers incorporated five methylation sites onto unmethylated *Lambda* genome using BamHI 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 towards single-molecule mapping of methylation patterns; however, since large fractions of many bacterial, plant and mammalian genomes are methylated, it is expected that a large amount of

MBD proteins will be required for optical mapping.

Recently, we demonstrated single molecule mapping of 5hmC sites by covalent chemical labelling 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 ten 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 click chemistry reaction. YOYO-1 stained DNA was extended on a modified cover slip 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 labelling.

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 extension 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.[26] 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, single 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. Such analysis may elucidate the presence of rare sub-populations 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.

Standard optical mapping approaches yield resolution of about 1 kb, superresolution methods can improve the resolution to about 100 bp. This can be achieved using blinking probes or opticalswithching. Such resolution is highly relevant for visualization of many epigenetic markers, such as transcription factors bound to gene promoter and the methylation state of the promoter and its surrounding.

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, order and occupancy). Nevertheless, progress in nanofabrication and optical imaging promises to boost research in this direction towards a high-resolution view of the genome and its composition on the single molecule level

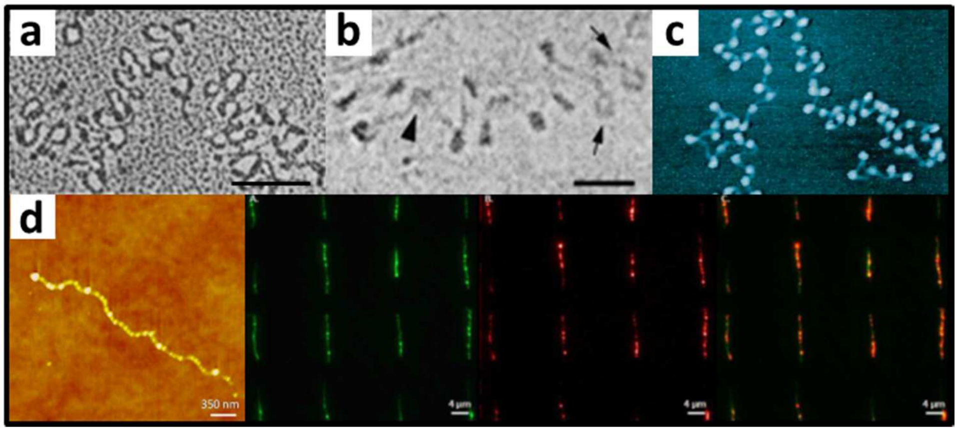


Figure 1. Single molecule imaging of nucleosomes. a) Chromatin containing histone H1 was

imaged by EM, (adapted from Thoma *et al.*,[62] with permission). b) Chromatin as seen by Cryo-EM (reprinted with permission,[65] copyright, 1998, National Academy of Sciences, U.S.A.) c) Cryo-AFM image of chicken erythrocyte chromatin fiber on mica (reprinted with permission,[64] copyright, 2002, American Chemical Soiciety). d) Chromatin array - in the left panel, HeLa chromatin as imaged by AFM. The chromatin was labelled 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 with permission*.,*[66] copyright, 2012, American Chemical Society).

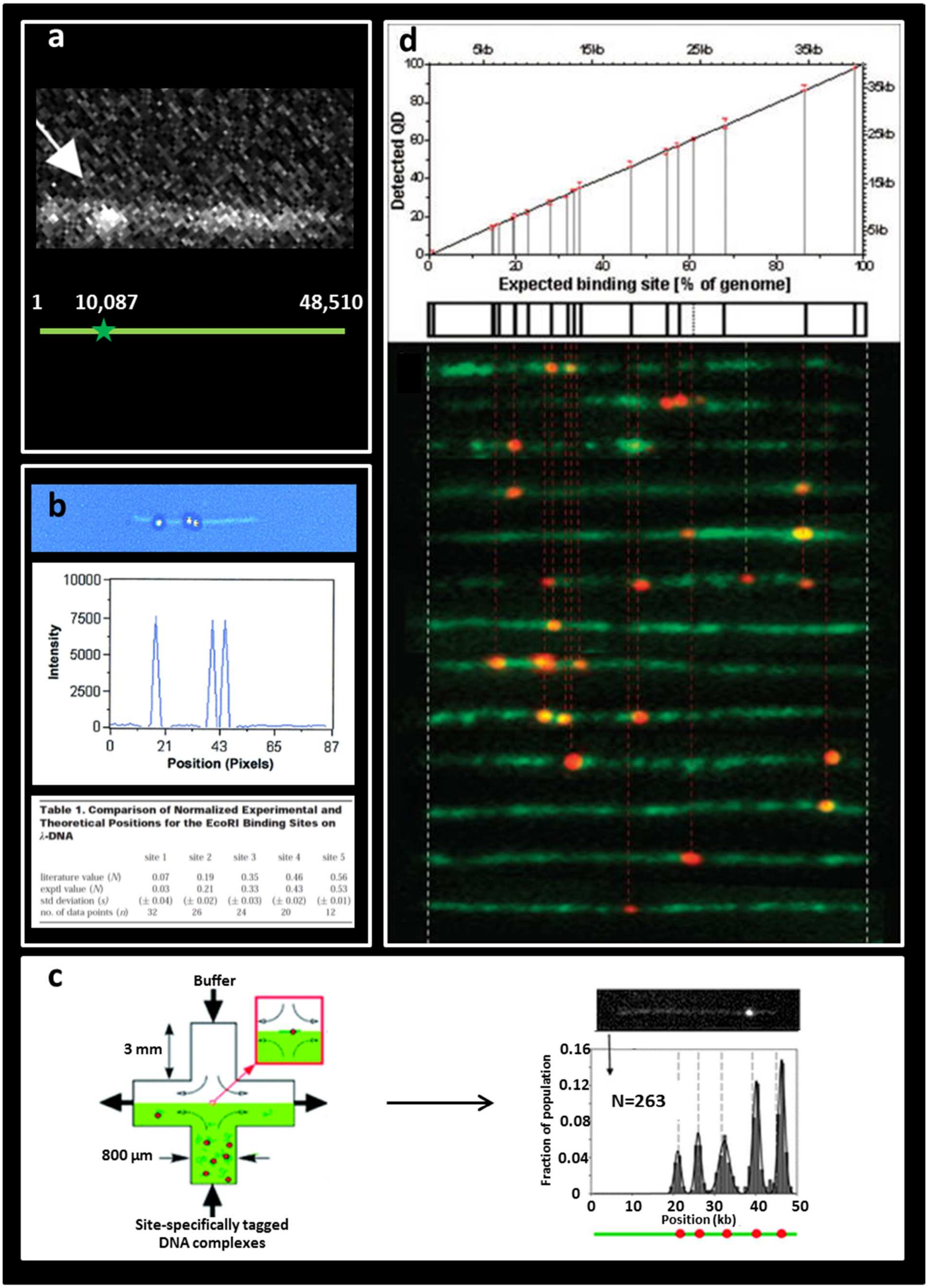


Table 1. Comparison of Normalized Experimental and Theoretical Positions for the EcoRI Binding Sites on λ-DNA

	site 1	site 2	site 3	site 4	site 5
literature value (*N*)	0.07	0.19	0.35	0.46	0.56
exptl value (*N*)	0.03	0.21	0.33	0.43	0.53
std deviation (*s*)	(± 0.04)	(± 0.02)	(± 0.03)	(± 0.02)	(± 0.01)
no. of data points (*n*)	32	26	24	20	12

Figure 2. Single molecule mapping of DNA binding proteins. a) ApaI restriction enzyme, labelled with alexa-fluor 532, bound to a single recognition site on *Lambd*a genome (adapted with permission,[84] copyright, 2005, American Chemical Society). b) Mapping of five EcoRI binding sites using *Lambda* DNA extended on polylysine surface, (reprinted with permission,[85] copyright, 2000, American Chemical Society), or c) at a stagnation point using a microfluidic device.[76] d) RNAPs mapping on T7 genome, RNAPs were conjugated to QD and DNA molecules were extended on polylysine surface (reprinted with permission,[20] copyright, 2009, American Chemical Society).

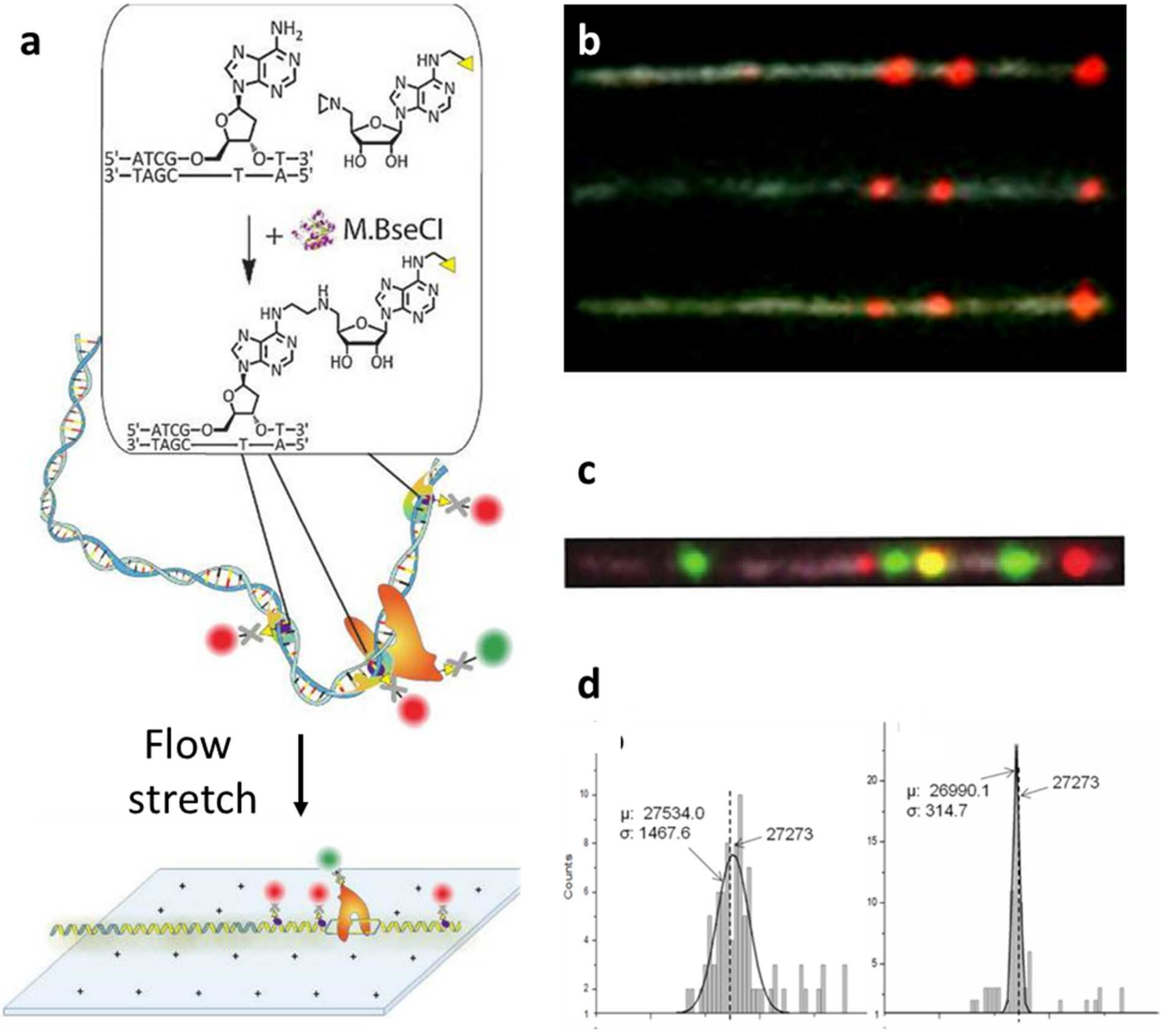
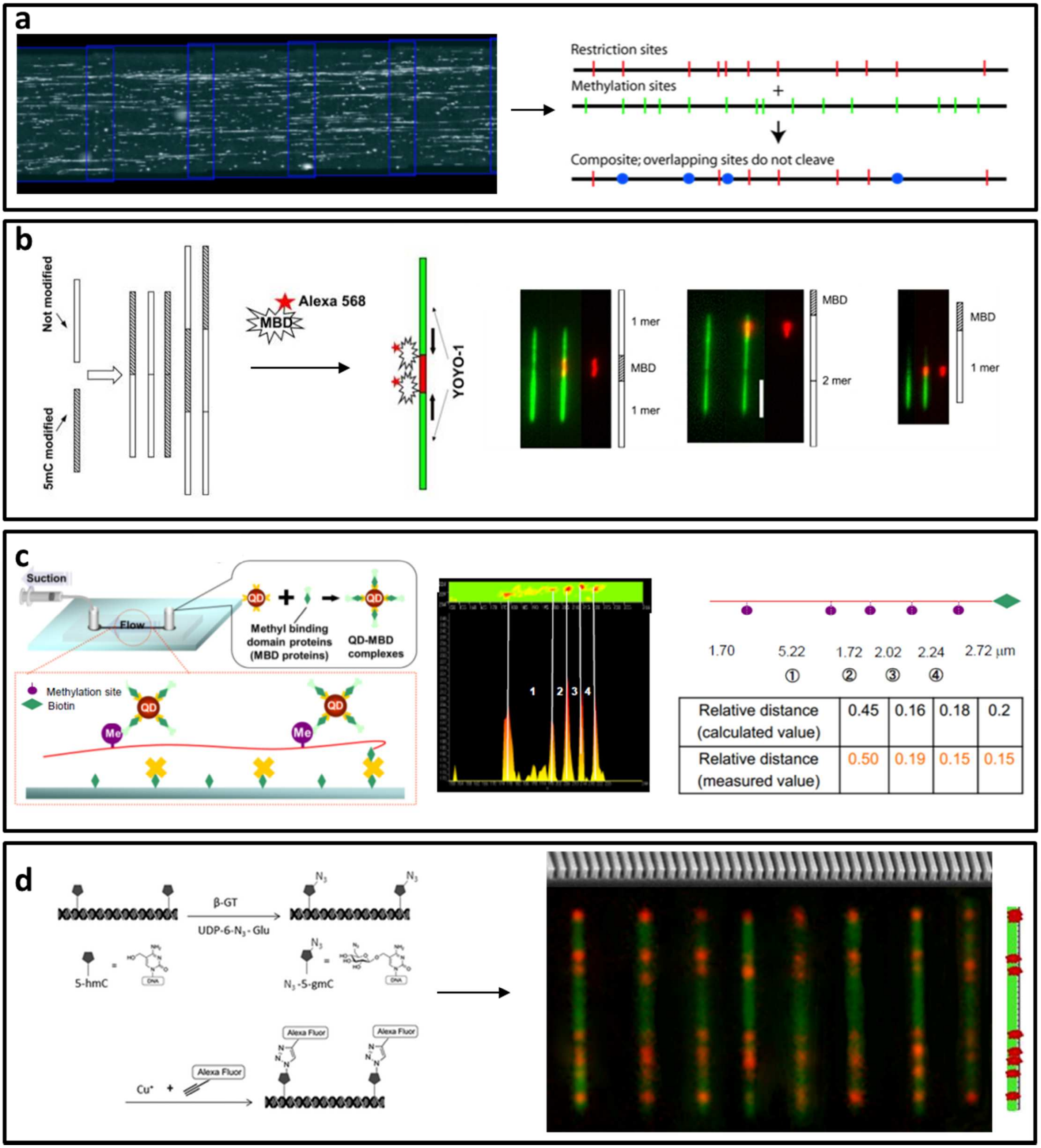


Figure 3. Incorporation of RefTags for improving protein mapping performance. a) Schematic representation of QD-labeled RNAP bound to sequence-specific-labeled T7 bacteriophage DNA. b) Image of flow-stretched, YOYO-1 stained T7 bacteriophage DNA (green) with QD-labeled M.BseCl refTags (red) and c) with RNAPs labeled with spectrally distinct QD (green), Overlapping red and green signals are shown in yellow. d) Histograms of all analyzed DNA-RNAP complexes with gaussian fit for localized RNAP on T7 bacteriophage using distance measurment to the DNA ends (left) versus localization by refTags (right). Dotted lines represent actual position of Φ13 promoter. Histograms using refTags yield a ~5-fold increase in accuracy as evidenced by sharp reductions in the width of promoter localization distributions. Sigma units are in bp (modified with permission,[21] John Wiley and Sons).



Relative distance (calculated value)	0.45	0.16	0.18	0.2
Relative distance (measured value)	0.50	0.19	0.15	0.15

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 cells DNA[97] b) Mapping of methylation sites using MBD proteins labelled with alexa-fluor 568 dye, as seen on *Lambda* DNA concatemers. In green, DNA stained with YOYO-1, in red, methylated *Lambda* DNA bound to labelled MBD proteins (reprinted with permission,[95] copyright, 2011, American Institute of Physics). c) Mapping of methylation sites using MBD proteins labelled with QDs. Methylcytosines were incorporated onto *Lambda* genome at known sites using a Mtase enzyme and were detected using QD labelled MBD proteins (adapted from Okamoto *et al.,*[96] with permission from Baba Y.) d) Covalent labelling 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 is used to attach an azido modified sugar at each 5hmC site and further labelled with an alkyne modified alexa fluor dye by a click reaction (Ebenstein lab, unpublished).

Acknowledgments

This work was supported by the UCLA-DOE Institute for Genomics and Proteomics (grant DE-FC02-02ER63421) and NIH grant R01-EB000312. YE is supported by the Legacy Heritage fund program of the Israel Science Foundation, the Marie Curie Career Integration Grant and the European Research Council. MLS would like to thank Tel Aviv University Center for Nanoscience and Nanotechnology for financial support.

1. Sawicki, M. P.; Samara, G.; Hurwitz, M.; Passaro, E. Human Genome Project. *The American Journal of Surgery* **1993**, *165*, 258–264.

2. Sanger, F.; Coulson, A. R. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *Journal of Molecular Biology* **1975**, *94*, 441–448.

3. Sanger, F. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences* **1977**, *74*, 5463–5467.

4. Ansorge, W. J. Next-generation DNA Sequencing Techniques. *N Biotechnol* **2009**, *25*, 195–203.

5. Mardis, E. R. The Impact of Next-generation Sequencing Technology on Genetics. *Trends Genet* **2008**, *24*, 133–41.

6. Schuster, S. C. Next-generation Sequencing Transforms Today's Biology. *Nat Meth* **2008**, *5*, 16–18.

7. Schones, D. E.; Zhao, K. Genome-wide Approaches to Studying Chromatin Modifications. *Nat Rev Genet* **2008**, *9*, 179–91.

8. Shendure, J.; Ji, H. Next-generation DNA Sequencing. *Nat Biotechnol* **2008**, *26*, 1135–45.

9. Neely, R. K.; Deen, J.; Hofkens, J. Optical Mapping of DNA: Single-molecule-based Methods for Mapping Genomes. *Biopolymers* **2011**, *95*, 298–311.

10. 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*, 432–8.

11. Michalet, X.; Ekong, R.; Fougerousse, F.; Rousseaux, S.; Schurra, C.; Hornigold, V.; M, W.; J, P.; S Dynamic Molecular Combing: Stretching the Whole Human Genome for High Resolution Studies. *Science (80-)* **1997**, *277*, 1518–1523.

12. 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*, 1137–46.

13. Xiao, M.; Phong, A.; Ha, C.; Chan, T.-F.; Cai, D.; Leung, L.; Wan, E.; Kistler, A. L.; DeRisi, J. L.; Selvin, P. R. *et al.* Rapid DNA Mapping by Fluorescent Single Molecule Detection. *Nucleic Acids Research* **2007**, *35*, e16.

14. 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.

15. 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*, 13294–9.

16. 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*, 4697–701.

17. 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. *Chemical Science* **2010**, *1*, 453.

18. Levy-Sakin, M.; Ebenstein, Y. Beyond Sequencing: Optical Mapping of DNA in the Age of Nanotechnology and Nanoscopy. *Current Opinion in Biotechnology* **2013**, *null*.

19. Ebenstein, Y.; Gassman, N.; Kim, S.; Weiss, S. Combining Atomic Force and Fluorescence Microscopy for Analysis of Quantum-dot Labeled protein-DNA Complexes. *Journal of molecular recognition JMR* **2009**, *22*, 397–402.

20. 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 Letters* **2009**, *9*, 1598–1603.

21. 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 Engl* **2012**, *51*, 3578–81.

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

23. Narzisi, G.; Mishra, B. Scoring-and-unfolding Trimmed Tree Assembler: Concepts, Constructs and Comparisons. *Bioinformatics* **2011**, *27*, 153–60.

24. Aston, C.; Mishra, B.; Schwartz, D. C. Optical Mapping and Its Potential for Large-scale Sequencing Projects. *Trends in Biotechnology* **1999**, *17*, 297–302.

25. Chen, S.; Xu, J.; Liu, C.; Zhu, Y.; Nelson, D. R.; 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.

26. 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*, 771–6.

27. 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). *Nature Biotechnology* **2012**, *advance on*.

28. 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*, 1079–85.

29. 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.

30. 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*, 417–44.

31. 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*, 1720–3.

32. Arthur D. Riggs, G. P. P. Polymerase Chain Reaction Aided Genomic Footprinting: Principles and Applications. *Advances in Molecular and Cell Biology* **1997**, *21*, 47–72.

33. 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*, 556–66.

34. Michael F Berger, M. L. B. Protein Binding Microarrays (PBMs) for the Rapid, HighThroughput Characterization of the Sequence Specificities of DNA Binding Proteins. **2009**, 1–15.

35. Hirayoshi K, L. J. Nuclear Run-on Assays: Assessing Transcription by Measuring Density of Engaged. *Methods Enzymol* **1999**, *304*, 351–362.

36. Pe, E. A Complete Set of Nascent Transcription Rates for Yeast Genes. **2010**, *5*.

37. 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.

38. Kauffman, C.; Karypis, G. An Analysis of Information Content Present in protein-DNA Interactions. *Pacific Symposium on Biocomputing* **2008**, 477–488.

39. 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*, 477–86.

40. Tate, P. H.; Bird, a P. Effects of DNA Methylation on DNA-binding Proteins and Gene Expression. *Curr Opin Genet Dev* **1993**, *3*, 226–31.

41. Jørgensen, H. F.; Adie, K.; Chaubert, P.; Bird, A. P. Engineering a High-affinity methyl-CpG-binding Protein. *Nucleic Acids Research* **2006**, *34*, e96.

42. Kriaucionis, S.; Heintz, N. The Nuclear DNA Base 5-hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **2009**, *324*, 929–30.

43. 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.

44. 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*, 1303–7.

45. Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C.-H.; Zhang, W.; Jian, X. *et al.* Selective Chemical Labeling Reveals the Genome-wide Distribution of 5-hydroxymethylcytosine. *Nature Biotechnology* **2010**, *29*, 68–72.

46. 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*, 1368–80.

47. 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*, 5221–31.

48. Chris R. Calladine, H. D. *Understanding DNA: The Molecule and How It Works*; Academic Press, 1997; p. 283.

49. 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. **2012**, *833*, 125–141.

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

51. Scully, K. M. Allosteric Effects of Pit-1 DNA Sites on Long-Term Repression in Cell Type Specification. *Science (80-)* **2000**, *290*, 1127–1131.

52. Gorman, J.; Greene, E. C. Visualizing One-dimensional Diffusion of Proteins Along DNA. *Nat Struct Mol Biol* **2008**, *15*, 768–74.

53. Allemand, J.-F.; Bensimon, D.; Croquette, V. Stretching DNA and RNA to Probe Their Interactions with Proteins. *Current Opinion in Structural Biology* **2003**, *13*, 266–274.

54. 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*, 3961–92.

55. 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*, 1640–8.

56. Finkelstein, I. J.; Greene, E. C. Molecular Traffic Jams on DNA. *Annu Rev Biophys* **2013**, *42*, 241–63.

57. 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*, 1056–62.

58. 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*, 525–38.

59. 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. *Proceedings of the National Academy of Sciences* **1998**, *95*, 12295–12299.

60. Finch, J. T.; Klug, a Solenoidal Model for Superstructure in Chromatin. *Proc Natl Acad Sci U S A* **1976**, *73*, 1897–901.

61. Thoma, F.; Koller, T. Influence of Histone H1 on Chromatin Structure. *Cell* **1977**, *12*, 101–7.

62. 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*, 403–27.

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

64. Zlatanova, J.; Leuba, S. H. Stretching and Imaging Single DNA Molecules and Chromatin. *J Muscle Res Cell Motil* **2002**, *23*, 377–95.

65. 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. *Proceedings of the National Academy of Sciences* **1998**, *95*, 14173–14178.

66. Cerf, A.; Tian, H. C.; Craighead, H. G. Ordered Arrays of Native Chromatin Molecules for High-resolution Imaging and Analysis. *ACS Nano* **2012**, *6*, 7928–34.

67. Raap, a K. Advances in Fluorescence in Situ Hybridization. *Mutat Res* **1998**, *400*, 287–98.

68. Dorfman, K. D.; King, S. B.; Olson, D. W.; Thomas, J. D. P.; Tree, D. R. Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching. *Chem Rev* **2012**.

69. 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*, 5164–8.

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

71. 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*, 10012–6.

72. Austin, M. D.; 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. *MRS Proceedings* **2011**, *1346*.

73. 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.

74. Jo, K.; Dhingra, D. M.; Odijk, T.; Pablo, J. J. de; Graham, M. D.; Runnheim, R.; Forrest, D.; Schwartz, D. C. A Single-molecule Barcoding System Using Nanoslits for DNA Analysis. *Proceedings of the National Academy of Sciences* **2007**, *104*, 2673–2678.

75. 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*, 4893–8.

76. 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*, 1543–9.

77. 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*, 14742–65.

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

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

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

81. 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. *PNAS* **1998**, *95*, 8046–8051.

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

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

84. Li, H.-W.; Yeung, E. S. Direct Observation of Anomalous Single-molecule Enzyme Kinetics. *Anal Chem* **2005**, *77*, 4374–7.

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

86. 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*, 10049–54.

87. 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*, 174–81.

88. Gueroui, Z.; Place, C.; Freyssingeas, E.; Berge, B. Observation by Fluorescence Microscopy of Transcription on Single Combed DNA. *Proceedings of the National Academy of Sciences* **2002**, *99*, 6005–6010.

89. 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.

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

91. Pljevaljcić, G.; Schmidt, F.; Weinhold, E. Sequence-specific Methyltransferase-induced Labeling of DNA (SMILing DNA). *Chembiochem* **2004**, *5*, 265–9.

92. 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. *Bioconjug Chem* **2008**, *19*, 470–5.

93. Braun, G.; Diechtierow, M.; Wilkinson, S.; Schmidt, F.; Hüben, M.; Weinhold, E.; Reich, N. O. Enzyme-directed Positioning of Nanoparticles on Large DNA Templates. *Bioconjug Chem* **2008**, *19*, 476–9.

94. Ananiev, G.; Goldstein, S.; Runnheim, R.; Forrest, D.; Zhou, S.; Potamousis, K.; Churas, C.; Bergendahl, V.; Thomson, J.; Schwartz, D. C. Optical Mapping Discerns Genome Wide DNA Methylation Profiles. *BMC Molecular Biology* **2008**, *9*, 68.

95. 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.

96. Okamoto, Y.; Sano, T.; Kaji, N.; Tokeshi, M.; Baba, Y. Identification of Single Molecular DNA Methylation Points by Microfluidic DNA Molecule Stretching and Quantom Dot Detection. **2012**, *3*, 1939–1941.

97. 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.