Evrim Atas[1]
Alon Singer[1]
Amit Meller[1,2]

[1]Department of Biomedical Engineering, Boston University, Boston, MA, USA
[2]Faculty of Biomedical Engineering, The Technion, Haifa, Israel

Received May 10, 2012
Revised June 29, 2012
Accepted June 29, 2012

Review

DNA sequencing and bar-coding using solid-state nanopores

Nanopores have emerged as a prominent single-molecule analytic tool with particular promise for genomic applications. In this review, we discuss two potential applications of the nanopore sensors: First, we present a nanopore-based single-molecule DNA sequencing method that utilizes optical detection for massively parallel throughput. Second, we describe a method by which nanopores can be used as single-molecule genotyping tools. For DNA sequencing, the distinction among the four types of DNA nucleobases is achieved by employing a biochemical procedure for DNA expansion. In this approach, each nucleobase in each DNA strand is converted into one of four predefined unique 16-mers in a process that preserves the nucleobase sequence. The resulting converted strands are then hybridized to a library of four molecular beacons, each carrying a unique fluorophore tag, that are perfect complements to the 16-mers used for conversion. Solid-state nanopores are then used to sequentially remove these beacons, one after the other, leading to a series of photon bursts in four colors that can be optically detected. Single-molecule genotyping is achieved by tagging the DNA fragments with γ-modified synthetic peptide nucleic acid probes coupled to an electronic characterization of the complexes using solid-state nanopores. This method can be used to identify and differentiate genes with a high level of sequence similarity at the single-molecule level, but different pathology or response to treatment. We will illustrate this method by differentiating the *pol* gene for two highly similar human immunodeficiency virus subtypes, paving the way for a novel diagnostics platform for viral classification.

Keywords:
DNA sequencing / Genotyping / Nanopore / Optical readout / γPNA
DOI 10.1002/elps.201200266

1 Introduction

The development of single-molecule techniques for biomedical applications, such as DNA sequencing, genotyping, and genome mapping, have recently gained considerable momentum [1–8]. In particular, nanopores have recently emerged as a new single-molecule sensing method, holding a great potential for high-throughput genome analyses. A key distinction feature of nanopores is that it does not require immobilization of the probed molecules, since it utilizes the native charge of biomolecules such as nucleic acids and proteins to propel them through an applied electric field. This electrophoretic method may be used to study individual DNA molecules by funneling and threading them through either biological or synthetic nanopores [6, 9–16].

Correspondence: Professor Amit Meller, Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
E-mail: ameller@bu.edu
Fax: +1-617-358-2835

Abbreviations: **CDC**, cyclic DNA conversion; **CMOS**, complementary metal-oxide semiconductor; **EM-CCD**, electron multiplying charge-coupled device; **HIV**, human immunodeficiency virus; **PNA**, peptide nucleic acid

A nanopore is simply a nanometer-scale pore in an ultrathin insulating membrane that separates two chambers containing electrolyte solution. An external electrical field applied across the membrane creates a drop in electric potential local to the nanopore, inducing ionic flow through the channel and creating a measurable current. This electric field first attracts randomly diffusing charged biopolymers, such as DNA, and then threads them through the nanopore, one molecule at a time [13, 17]. As a polymer passes (or "translocates") through the pore, it physically displaces a fraction of the electrolytes in the channel, causing an immediate drop in the pore's conductivity that can be measured with a high-bandwidth electrometer.

Another key feature of the nanopore method is its ability to linearize or unfold and detect long biopolymer coils through size constriction. When a long biopolymer is threaded through a nanopore only slightly larger than the biopolymer's cross-section, its coil structure must be sequentially deformed and linearized as the polymer passes through the narrow constriction [18]. Local linearization is a feature unique to the nanopore method, setting it apart from

Colour Online: See the article online to view Figs. 1–6 in colour.

most other sensing methods. It is particularly useful for analyzing local structure along a long biopolymer, positioning nanopores as an ideal platform for the development of genomic profiling tools and for DNA sequencing [5–7, 19–22].

The ability to determine the exact sequence of genomic DNA is critical to develop a better understanding of disease, disease evolution, and drug design. Thus, high-throughput DNA sequencing technologies are profoundly impacting comparative genomics, biomedical research, and personalized medicine [1]. Fast and cost-effective DNA sequencing technology will enable the rapid analysis of genotypes and identification of pathogens. However from a clinical standpoint, multiple strategies require only the identification of certain known genomic differences, such as those which present resistance to antibiotics or to a proposed line of therapy [23, 24]. Therefore rather than sequencing, genotyping provides an attractive and potentially cost-effective alternative [25]. Clinical genotyping would benefit greatly from a method that does not require extensive sample preparation steps and length enzymatic reaction times, and in addition can be readily multiplexed [26, 27]. In order to achieve this goal, a novel detection mechanism is required, capable of detecting the presence of specific genomic variants of human cells or pathogens quickly. Results reviewed here indicate that peptide nucleic acid (PNA) based nanopore detection has the potential to attain this goal.

In this paper, we present two examples in which nanopores are applied to address biomedical challenges. We first describe a method for high-speed DNA sequencing using optical probes [6]. Here, nanopores are used to directly induce a series of photon bursts in four different colors that correspond to the sequence of the translocating DNA. As a second example, we present a "bar-coding" method that utilizes electrical signals attained during the translocation process of tagged DNA molecules through a pore to discriminate between two highly similar genes [5]. Both examples involve probing of either "optical" or "electrical" probes, which are hybridized to the DNA molecules in a sequence-specific manner. But these two examples are complementary in nature. The aim of DNA sequencing is to read the nucleobase sequence of an unknown DNA molecule, whereas DNA barcoding is used to discriminate among a number of variants of DNAs having known sequence.

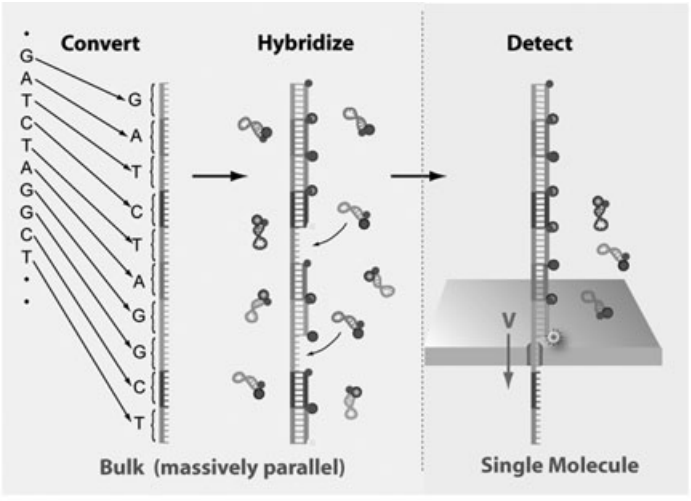
2 Nanopore-based sequencing

Nanopore-based DNA sequencing is widely accepted as one of the most promising next generation sequencing platforms [8]. Two main features of the nanopore method make it remarkably useful for single-molecule-based genome analyses: First, long DNA molecules are electrophoretically focused and threaded from the bulk solution into the pore, making it possible to analyze minute DNA sample quantities. Second, sub 5 nm pores are now routinely used to linearize long DNA coils, so in principle, nanopores can be used to effectively "scan" information along a genome [5, 28, 29]. These features, as well as the fact that solid-state nanopores can be rapidly fabricated in highly dense arrays [30, 31] are the foundation for development of methods for highly parallel detection, and are crucial for the realization of low-cost and high-throughput DNA sequencing [1–3, 32, 33].

Notwithstanding the recent progress made in the nanopore field, there remain two key challenges that must be addressed in parallel in order to successfully implement a high-throughput solid-state nanopore-based sequencing platform: (i) a method to enable parallel readout and (ii) discrimination among the four individual nucleobases. A single nanopore can only probe one DNA molecule at a time, so given the enormous length of the human genome; a strategy for manufacturing and monitoring a dense array of nanopores is required. While recent developments in protein-based nanopore sequencing have displayed promise through a newly engineered ability to discriminate between nucleobases [34, 35], scaling up or multiplexing this method from a single pore to a sparse array of tens of thousands of pores remains a major challenge. In contrast, solid-state nanopores, fabricated with common semiconductor processing methods, naturally lend themselves to parallelization in extremely dense arrays. However, as of this writing, their ability to discriminate among the four bases in DNA has yet to be demonstrated.

The optipore sequencing method was designed to addresses both of these main challenges through a two-step process. To increase the contrast between individual nucleobases, a highly parallel bulk process is employed to convert the DNA of interest to an expanded form by systematically substituting each nucleobase with a specific 16-mer oligonucleotide. The converted DNA is hybridized to complementary molecular beacons of four colors (representing the four 16-mer oligos). To detect the sequence, nanopores are then used to strip away, or unzip, the beacons sequentially. With each unzipping event, a new fluorophore is unquenched, creating a series of photon bursts in four colors, which are then recorded by a CCD or complementary metal-oxide semiconductor (CMOS) camera (see Fig. 1). The recorded sequence of photon bursts can be used to decipher the original sequence of the DNA strand. Due to the high speeds of DNA translocation through nanopores, we utilize the DNA unzipping process to slow down this translocation to a rate more compatible with single molecule optical detection [36, 37]. The use of optical detection will enable this process to be multiplexed; an array containing thousands of nanopores could, in theory, be probed simultaneously using a single camera (see Fig. 1).

Ideally, nanopore-based sequencing only involves the physical process of a DNA strand sliding through a pore, driven by the electrophoretic force, which can be regulated by adjusting the applied voltage or other easily controlled physical parameters. This idea motivates the development of purely "physical" approaches for DNA sequencing using synthetically made nanopores. The optipore method presented here is an example of this approach. Biochemical preparation of the target DNA molecules converts each base into a form that can be read directly using an unmodified



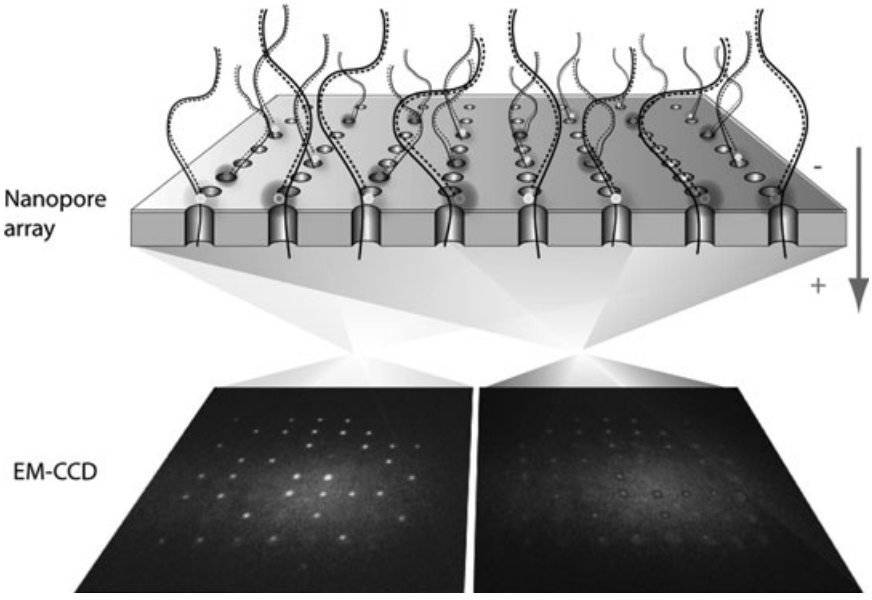


Figure 1. (Top) Schematic of our optical nanopore-sequencing platform. Using the CDC (cyclic DNA conversion) process, each nucleobase in a DNA molecule is converted into 16-mer oligonucleotides of known sequence concatenated to each other. There are four oligonucleotide sequences (color coded in the figure) representing the four DNA nucleobases. Molecular beacons are then hybridized to this converted strand. Passage through a nanopore results in stripping off, or unzipping, of each beacon, which in turn induces a photon burst in one of four colors. (Bottom) Schematic of the parallel detection capability of the optipore setup (only two colors out of four shown for clarity). This illustration highlights the concept of detecting from a large array of pores simultaneously. Figure adapted from McNally et al. [6], with permission. © ACS publications.

solid-state nanopore. Readout speed and length, therefore, are not enzyme limited. The use of optical sensing to detect a DNA sequence rather than using electrical sensing allows a simple transition to a multipore detection scheme. Our technique utilizes a total internal reflection instrument, which permits high spatiotemporal resolution wide-field optical detection of individual DNA molecules translocating through a nanopore [37].

2.1 Cyclic DNA conversion

Cyclic DNA conversion (CDC) is a biochemical process for expanding the "footprint" of individual nucleobases in a DNA molecule. In CDC, each of the nucleobases in a given DNA sequence is substituted with a predesigned oligonucleotide. These predesigned oligonucleotides, which we refer to as "code oligos" are typically 16-mer in length and have a sequence specifically chosen to facilitate optimal conversion in our CDC process. We have designed four "code oligos," each representing one of the four naturally occurring nucleobases. During a single conversion cycle, a single nucleobase or multiple nucleobases are substituted with their corresponding "code oligo." Multiple conversion cycles allow the concatenation of multiple "code oligos" into a long ssDNA molecule that can be read by our nanopore-based DNA sequencer. The converted single strand of DNA is then hybridized with perfectly complementary, predesigned, fluorescently labeled oligonucleotides. These oligonucleotides form the beacons, which enable optical detection. The CDC process does not require the amplification of target DNA, which prevents the errors introduced by polymerase-based amplification steps and eliminates the need for subsequent purification of the amplified products. This method requires only miniscule quantities of

starting DNA material or reagents thus lending itself for a fully integrated lab-on-a-chip approach.

Briefly described, first we prepare the target ssDNA. The isolated genomic DNA is sheared to fragments of roughly a kilobase in length, the ends are enzymatically blunted, and one of the ends is ligated with a short specific adapter (12 bp) that also contains a biotin moiety for immobilization purposes. After immobilization on magnetic beads, the strands are separated by DNA melting. After the initial preparation of the target ssDNA, multiple rounds of "cyclic conversion" are performed as illustrated in Fig. 2. In the first step (i), a hybridization with a library of T-shaped oligonucleotide probes, each of which contains a double-stranded region flanked by two asymmetric single-stranded overhangs, is carried out. The 5′ overhang (red) is designed to be perfectly complementary to the aforementioned tag at the 5′ end of the target ssDNA, and one position at the 5′ end (yellow, labeled with *n*) contains all four possible bases. The red portion of the double-stranded region indicates the "code oligo" sequence for the specific base converted. The black portion in the 3′ single-stranded overhang indicates a single base (A, G, C, or T), the identity of which is encoded by the code oligo. In the example shown, an adenine base is displayed. The other five positions in the 3′ overhang (indicated with *n*) are represented in the library by all possible combinations of four nucleotides. The high hybridization fidelity of the probe library is ensured by the ligase reaction performed in step (ii), which was shown to be extremely reliable as compared with other enzymatic processes [38]. In step (iii), an off-recognition-site cleavage is performed by the type IIS restriction enzyme recognizing the *R* site. In step (iv), the top portion of the probe is removed after a brief temperature-induced melting.

After these four steps are completed, a single base has been transferred from the 5′ end to the 3′ end of the target ssDNA, and the code associated with this specific base has been added, along with a few more bases that are the recognition site for the type IIS restriction enzyme (blue). The system is now ready for the next cycle. This process can be repeated, in principle, as many times as needed to convert long reads. We note that by design, our CDC is Markovian in nature, as each successive cycle does not depend on the outcome of previous cycles. Incorrect hybridizations will not be ligated and will be cleared during the next cycle. The conversion process of different templates is completely asynchronous. We note that the converted DNA is not expected to have stable secondary structure, as it is concatenated from "bits" designed to exhibit minimal cross-hybridization. We predict that the conversion of long DNA strands is feasible, limited only by the time required for conversion.

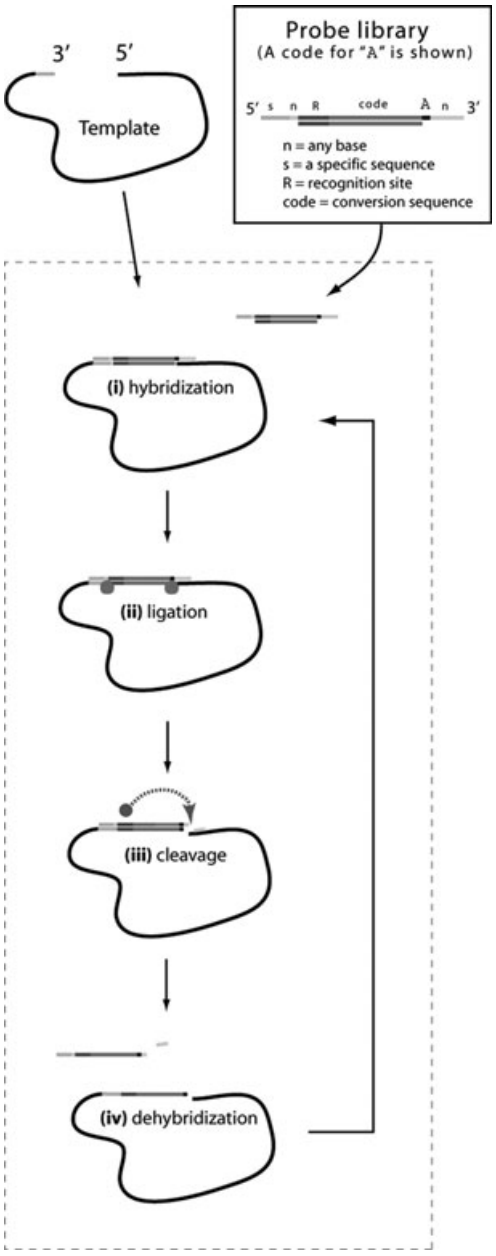


Figure 2. The CDC method involves four steps, which are repeated as necessary to convert any number of bases per cycle. This scheme was employed in McNally et al. [6], with permission. © ACS publications.

2.2 Optical detection

Proof-of-principle for the optical detection scheme was initially carried out using two colors for simplicity. Here, the four nucleobases are represented by two-bit combinations of two predefined unique sequences, bit "0" and bit "1." Using this approach, we designated an A as "1 1," a G as 1 0," a T as "0 1," and finally a C as "0 0" (see Fig. 3A). Two different molecular beacons carrying two types of fluorophores are complementary to and hybridize specifically to the "0" and "1" sequences. After converting the target DNA, the digitized DNA molecules are hybridized to the two different molecular beacons. After hybridization, the dsDNA molecules are electrophoretically threaded through an approximately 5 nm solid-state pore, where the beacons are sequentially stripped off. As previously described, each time a molecular beacon is

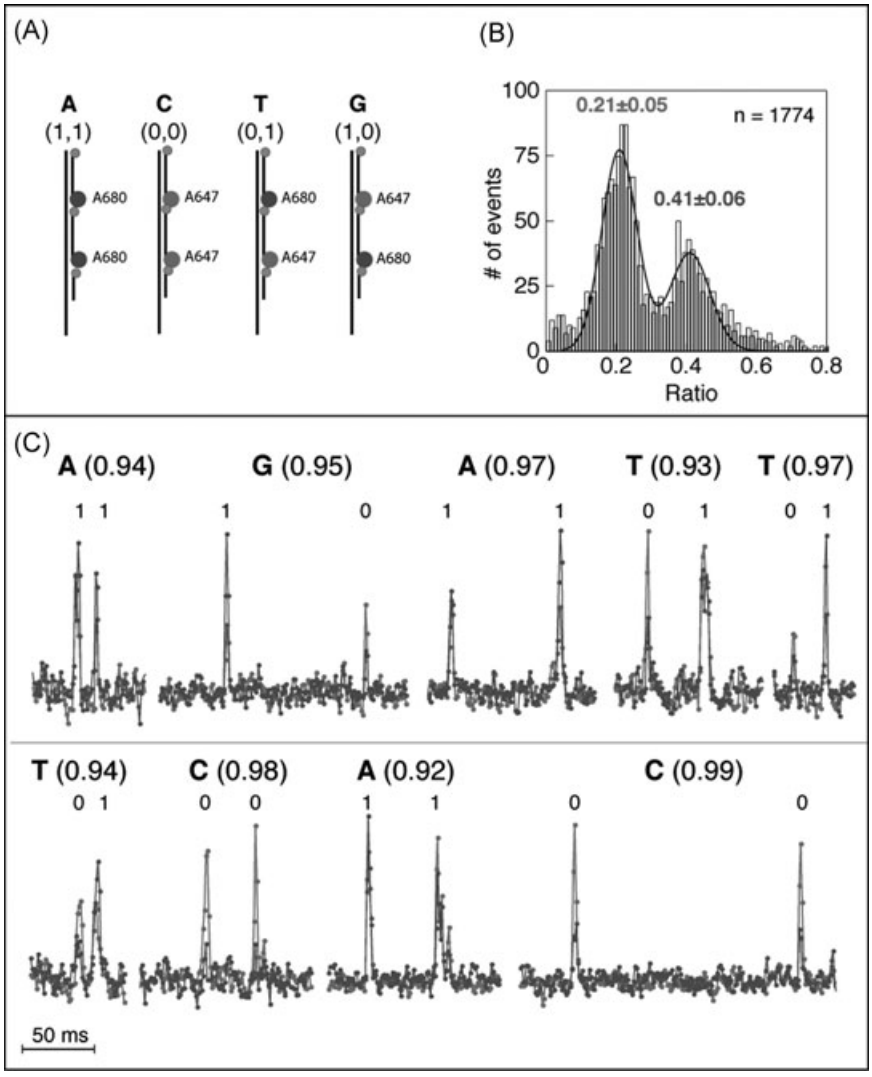


Figure 3. Nanopore-induced photon emission enables nucleobase identification. (A) Using two different fluorophores (Atto647 and Atto680) enables construction of distinct samples, which correspond to all four DNA nucleobases. (B) The distribution of *R*, generated with thousands of events, reveals two means at 0.21 ± 0.05 and 0.41 ± 0.06, which correspond to the A647 and A680 fluorophores, respectively, in excellent agreement with control studies. Solid line represents a double Gaussian fit to the histogram. (C) Representative traces of individual 2-color optical unzipping events, with the corresponding beacon identified, base called and certainty score indicated above the event. This figure was reproduced from McNally et al. [6], with permission. © ACS publications.

stripped off, a new fluorophore is unquenched, giving rise to a burst of photons, recorded at the location of the pore using an electron multiplying CCD (EM-CCD). The sequence of the two-color photon bursts at the pore location (colored blue and red in Fig. 3B) is the binary code of the target DNA sequence.

We demonstrated the ability to differentiate among the four DNA nucleotides using our custom-built total internal reflection imaging system coupled with a solid-state nanopore system, demonstrating the feasibility of the optipore method. In order to distinguish among the four nucleotides, a two-color coding scheme was deigned, employing two high-quantum yield fluorophores, Atto647 (A647) and Atto680 (A680). The optical emission signal was split into two channels using a dichroic mirror, which were then imaged side-by-side on the EM-CCD camera. Due to the spectral overlap of emission from the two fluorophores, a fraction of the A647 emission "leaks" into the A680 channel (channel 2) and a fraction of A680 "leaks" to the A647 channel (channel 1). We thus found that the ratio (R) of the fluorescent intensities in channel 2 versus channel 1 can be used to reliably discriminate between the two fluorescent molecular beacons, rather than the absolute intensity in each channel. Calibration unzipping measurements conducted with each of the fluorophores separately yielded R values of 0.20 ± 0.06 and 0.40 ± 0.05 for the A647 and A680, respectively.

Using the calibration distributions of the intensity ratio, we identified the four different two-beacon combinations, namely 11 (A), 00 (C), 01 (T), and 10 (G), where "0" and "1" correspond to the A647 and A680 beacons, respectively. The bimodal distribution of R displays two modes at 0.21 ± 0.05 and 0.41 ± 0.06, in complete agreement with the calibration measurements. Base identification was achieved through the classification of all photon bursts with $R < 0.30$ as a "0," and those with $R > 0.30$ as a "1." The distribution of R was also used to compute the probability of correct classification. This provides us with a statistical means to calibrate the two channels for optimal discrimination between the two fluorophores. Representative optical unzipping events ($n >$ 2000 photon bursts) are depicted in Fig. 3, showing the

single-molecule identification of each of the four DNA bases. The strength of our two-color identification is attributed primarily to the excellent signal/background levels of the photon bursts and the significant separation between the fluorophore intensity ratios (R) for the two channels. A custom algorithm to automatically identify photon bursts and determine beacon identity using the calibration distributions was developed. This algorithm was used to perform base calling (with corresponding certainty scores). Typical results are shown in Fig. 3. We note that even with the simple experimental setup used here, employing only a single excitation line and off-the-shelf agents, the classification error per nucleobase is less than 10% (per single read).

These results show that nucleobase identification of converted DNA is feasible with solid-state nanopores. The raw error in base calling (for a single read) is relatively low, and it is expected that as the method is further improved this error will be reduced substantially. Specifically, we expect that with introduction of four different color beacons the uncertainty in base calling will drop significantly below 1%, thereby doubling the sequencing speed.

The optically based nanopore DNA sequencing offer several advantages over alternative methods. First, the system is extremely robust, as it does not involve moving parts (such as stages) or a cyclic flow of analytes during the readout process. Moreover, the nanopore chip itself is a simple array of solid-state pores, not involving enzymes at the readout stage. The speed of readout is completely controllable, and is limited only by the temporal resolution of the detection modality. Currently, the system can read 50–250 bases per second per nanopore, which compares favorably with other single-molecule approaches [2, 3, 20, 21]. We anticipate that a straightforward adaptation for 4-color and the use of optimized reagents will enable us to achieve >500 bases per second per nanopore with a raw base calling error rate of less than 1%. Most importantly, optical readout enables straightforward readout from hundreds of pores using commercial high sensitivity cameras.

3 Nanopore-based genotyping

A remarkable attribute of solid-state nanopores, differentiating them from other single-molecule techniques, lies in their ability to rapidly and robustly probe individual DNA or RNA molecules extracted from an extremely small sample pool, down to a few attomoles ($\sim 10^5$ copies) [17]. These unique features position solid-state nanopores as prime candidates for molecular diagnostics applications, specifically those involving a minute batch of cells. To date, despite these promising features, nanopores have seldom been applied in molecular diagnostic sensing applications, primarily due to their lack of DNA or RNA sequence specificity. However, by using γ-modified synthetic PNA (γPNA) probes [39–42] to tag DNA and detecting these complexes with solid-state nanopores, it becomes possible to positively identify and discriminate between genes bearing a high level of sequence similarity. This method has recently been demonstrated by sensing and differentiating two versions of the same gene from two highly similar human immunodeficiency virus (HIV) subtypes, paving the way for a novel diagnostics platform for viral classification. Specifically, it was shown that probe sites as close as only a few tens of nanometers apart could be readily resolved on submillisecond timescales. The intrinsic advantages of solid-state nanopores as a single-molecule technique are well complemented by the addition of γPNA probe libraries to enable an unprecedented rapid electronic, nucleic acid based diagnostic platform.

γPNA probes are newly developed synthetic oligomers, which due to their extremely high affinity to DNA, can displace one strand of a DNA molecule while binding to the other, invading the helix with remarkable specificity but without sequence restrictions [40–44]. Strikingly, when γPNA-tagged DNA is threaded through a nanopore, clear and well-defined ion current blockades are detected due to the tagged regions of larger cross-section. This observation is illustrated in Fig. 4, which depicts a DNA molecule with three preselected PNA binding sites. In absence of the γPNA probes, the DNA molecules translocations show distinct current blockades of given dwell times and amplitudes, with a single blocked current level that corresponds well to the dsDNA mean cross-section [18]. Upon invasion of the γPNA probes, three additional blockade episodes for each translocation event are observed, all conforming to a well-defined and uniform ion-current state. This finding allows us, in principle, to target and detect short sequences (i.e. 15 bp) in any DNA molecule of interest, forming the basis for a novel single-molecule gene identification method.

Ultimately, nanopore/γPNA-based genotyping will require not only counting the number of γPNA/DNA sites per DNA molecule, but also the ability to localize their positions along the DNA molecule, effectively bar-coding the target. In other words, an unequivocal identification of a certain gene requires that both the number and the gaps between probe locations match a predesigned pattern for a target gene. To evaluate the nanopore's ability to accurately determine the distance between two γPNA probes along a DNA molecule, six similar DNA molecules were constructed, each containing two identical γPNA sites. The intersite spacing, Δn, between the two γPNA binding sites ranged from 100 to 1000 bp, while the sequence and length of the two flanking regions was identical for all samples ($\sim$1200 bp; Fig. 5A).

Nanopore assays using these molecules were conducted to measure the delay time between γPNA probe pairs, δt (as shown in Fig. 5B; see Fig. 5D for statistical analysis). Figure 5C depicts the results when the average value of δt, obtained from $N \sim 1000$ events for each of the six molecules, is plotted as a function of the actual distance between the two γPNA sites (Δn). As shown in Fig. 5C, δt could be accurately pinpointed to the shortest Δn we tested (100 bp), corresponding to approximately 34 nm spacing. Additionally, our data are well fitted by a power law relationship with an exponent of 1.39 ± 0.09, in agreement with previous experiments measuring the translocation time of untagged DNA molecules

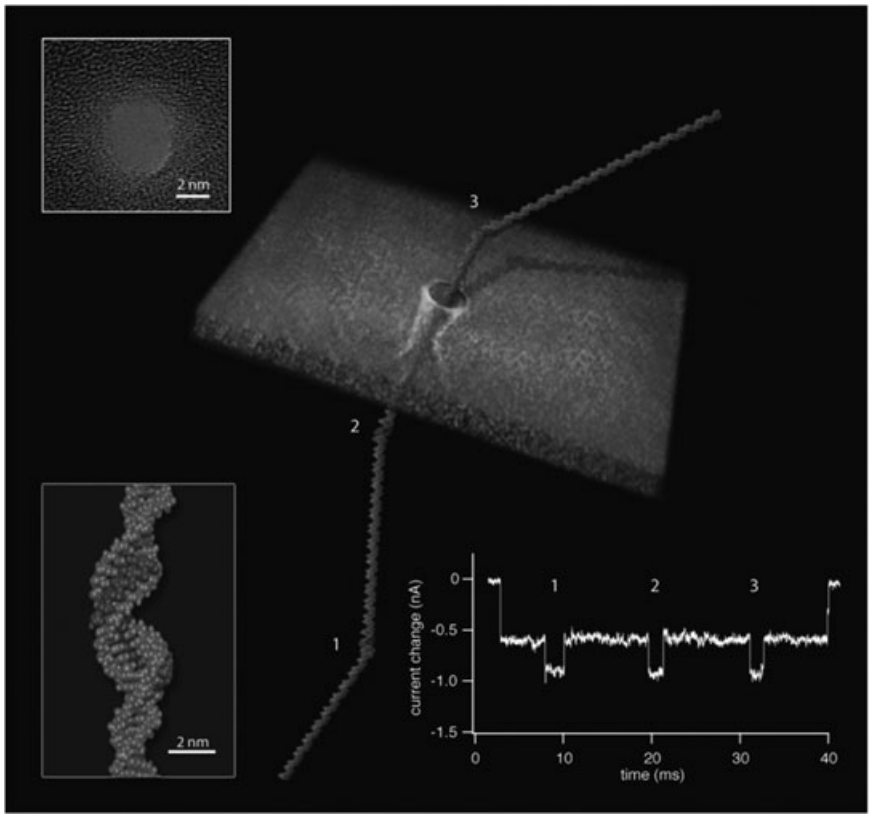


Figure 4. Electronic bar-coding of a single DNA molecule using the nanopore/PNAs (where PNA is peptide nucleic acid) system. Illustration of a DNA molecule tagged with three γPNA probes equally spaced, translocating across a sub 4-nm solid-state nanopore sculptured in a thin silicon nitride membrane. Three distinct secondary blockade episodes, in excess of the typical DNA current blockade level, clearly mark the point in time when the γ-modified synthetic PNA (γPNA) tagged region crossed through the nanopore. Top inset: High-resolution transmission electron microscopy (TEM) image of a 3.7-nm nanopore, made in a thin silicon nitride film. Bottom inset: Close-up approximated structure of a γPNA probe site. Given its high affinity to DNA, the single-stranded γPNA (red) binds to its complementary DNA sequence, locally displacing the second DNA strand. This process leads to a local change in the helical structure, allowing it to be discerned with the solid-state nanopore. Figure adapted from Singer et al. [5], with permission. © ACS publications.

of similar lengths [18]. This calibration process enables the nanopore readout signal to identify a γPNA binding pattern along a DNA strand down to a probe-to-probe distance of approximately 100 bp.

Combining information regarding the number of probes as well as the distance between probes bound can, in principle, be used to create a relatively small probe library appropriate for a multitude of genomic targets. In order to evaluate the feasibility of this approach as a potential molecular diagnostic platform, we leverage the high sequence specificity of 15-mer γPNA probes in combination with the ability to accurately determine relative probe positions along DNA using the nanopore technique in a simple bar-coding strategy. When a DNA sample extracted from a source of interest is combined with a probe library set, only those probes that have sequences complementary to the designed targets will bind to the extracted DNA. In practice, this principle allows us to create maps of possible binding patterns, termed barcodes, specific to any given library of PNA probes. The pattern of γPNA delay times acquired in a nanopore experiment can then be directly compared against these maps to identify the presence (or absence) of DNA molecules of interest.

To test this approach, HIV genomes were used as a model set of samples. HIV is an excellent model organism on which to test this technique due to the virus' inherent ability to rapidly develop resistance to treatment, therefore necessitating the development of a fast, simple, and low-cost method for identifying HIV resistance [45–47]. In particular the HIV *pol* gene, which encodes for protease, integrase, and reverse transcriptase, is an optimal target since it is a key gene related to HIV vaccine development [45]. Furthermore, given the high similarity in *pol* displayed between the subtypes (92% identity following alignment of the selected subtypes), this gene served as a stringent candidate for a proof-of-concept study.

Two nearly identical DNA fragments that contained the sequences for the *pol* gene (3050 bp) from HIV-1/B and HIV-1/C were synthesized using standard methods [48]. Four different γPNA oligomers were also synthesized, of which two oligomers were a perfect match to both subtypes (γPNA_1 and γPNA_2), while the third, γPNA_3, would bind only to the sequence present in subtype B, and γPNA_4 would bind only to the sequence present in subtype C. This should produce two different patterns or barcodes for the two subtypes: γPNA_1 and γPNA_2 would be spaced approximately 850 bp apart on both subtypes, while γPNA_3 would be spaced approximately 850 bp from γPNA_2 on subtype B, and γPNA_4 would be spaced approximately 450 bp from γPNA_2 on subtype C. Correct γPNA binding patterns were confirmed by selectively dividing each gene into ten fragments, each 300–400 bp, followed by gel-shift binding assays for each of the different γPNA oligos mounted separately and also as a complete library of all four probes together [5].

This time-consuming and laborious gel-based analysis can be replaced by a single nanopore measurement for each of the two HIV genes subtypes, requiring only a few minutes. We first note that nanopore analysis of the untagged

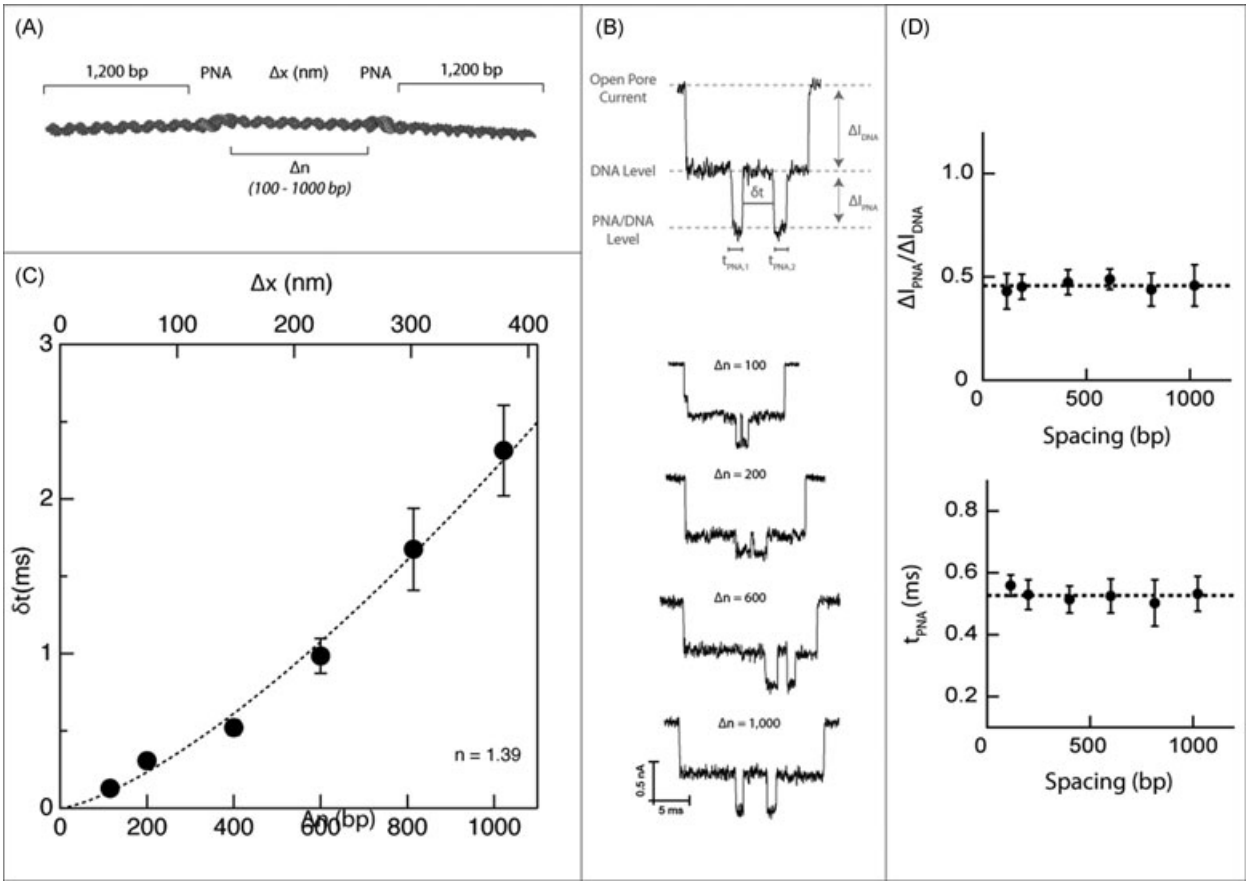


Figure 5. Varying the distance between two γ-modified synthetic peptide nucleic acid (γPNA) sites to create a time (δt) to distance (Δn or Δx) conversion map. (A) Six nearly identical and symmetrical DNA molecules were created, consisting of identical flanking regions (1200 bp before the first γPNA site and 1200 bp after the second γPNA site), with varying spacing between the two γPNA sites. (B) Representative single-molecule traces illustrating the time-delay differences, δt, between the different molecules. (C) Statistical analysis showing the dependence of δt on the γPNA site-to-site distance relationship (as measured in bp). The data are well approximated by a power law dependence with an exponent of 1.39 (dashed line). (D) Through statistical analysis of thousands of events of each of the different spacings, we have found that (top) $\Delta I_{PNA}/\Delta I_{DNA} = 0.45 \pm 0.04$, and that (bottom) $t_{PNA} = 0.55 \pm 0.06$ ms, independent of the probe-to-probe distance. These constant values (slope of best linear fit equal to ~0) indicate a lack of crosstalk between γPNA sites, even at the smallest spacing of 100 bp. Figure adapted from Singer et al. [5], with permission. © ACS publications.

subtypes is insensitive to their slight differences in sequence differences present between the two subtypes. In contrast, the γPNA-tagged DNA produced distinctive ion current signals, enabling a simple identification of both subtypes by comparing the differences in delay times measured between three different γPNA sites (see representative traces in Fig. 6 for both variants).

An in-depth analysis of thousands of molecules was simplified by normalizing the individual delay times between γPNA sites by the sum of all the delay times in each and every event. This procedure eliminated small pore-to-pore variability, allowing a much cleaner and more direct analysis of the data. Specifically, we calculated the delay time between the first and second γPNA sites as δt_{1-2}, and between the second to third γPNA sites as δt_{2-3}. The normalized delay times are defined as: $\tau_1 = \delta t_{1-2}/(\delta t_{1-2} + \delta t_{2-3})$ and $\tau_2 = \delta t_{2-3}/(\delta t_{1-2} + \delta t_{2-3})$, respectively. Figure 6 displays the distributions of τ_1 and τ_2 for subtype B and subtype C (red and blue respectively, $N = 750$ events in each case). As expected, the symmetrically tagged subtype B variant yields two nearly identical distributions with a single peak at 0.500 ± 0.004, reflecting the fact that the delay times between the PNA probes are practically equal (850 bp). In contrast, the subtype C variant yielded double peak distributions centered at 0.269 ± 0.005 and 0.749 ± 0.004 for τ_1 and 0.252 ± 0.004 and 0.731 ± 0.005 for τ_2, respectively. Referring to the calibration curve (Fig. 5), we note that these values coincide precisely with the delay time associated with the 450 bp and 850 bp gaps: 0.61 ms and 1.85 ms, respectively, since $0.61/(0.61 + 1.85) = 0.248$ and $1.85/(0.61 + 1.85) = 0.752$. The appearance of two nearly identical distributions in each case simply reflects the fact that the DNA can traverse the nanopore with either of its two ends entering first.

Figure 6 demonstrates that this technique can readily discriminate between the two HIV subtypes. More importantly, it suggests that the normalized delay times between the PNA

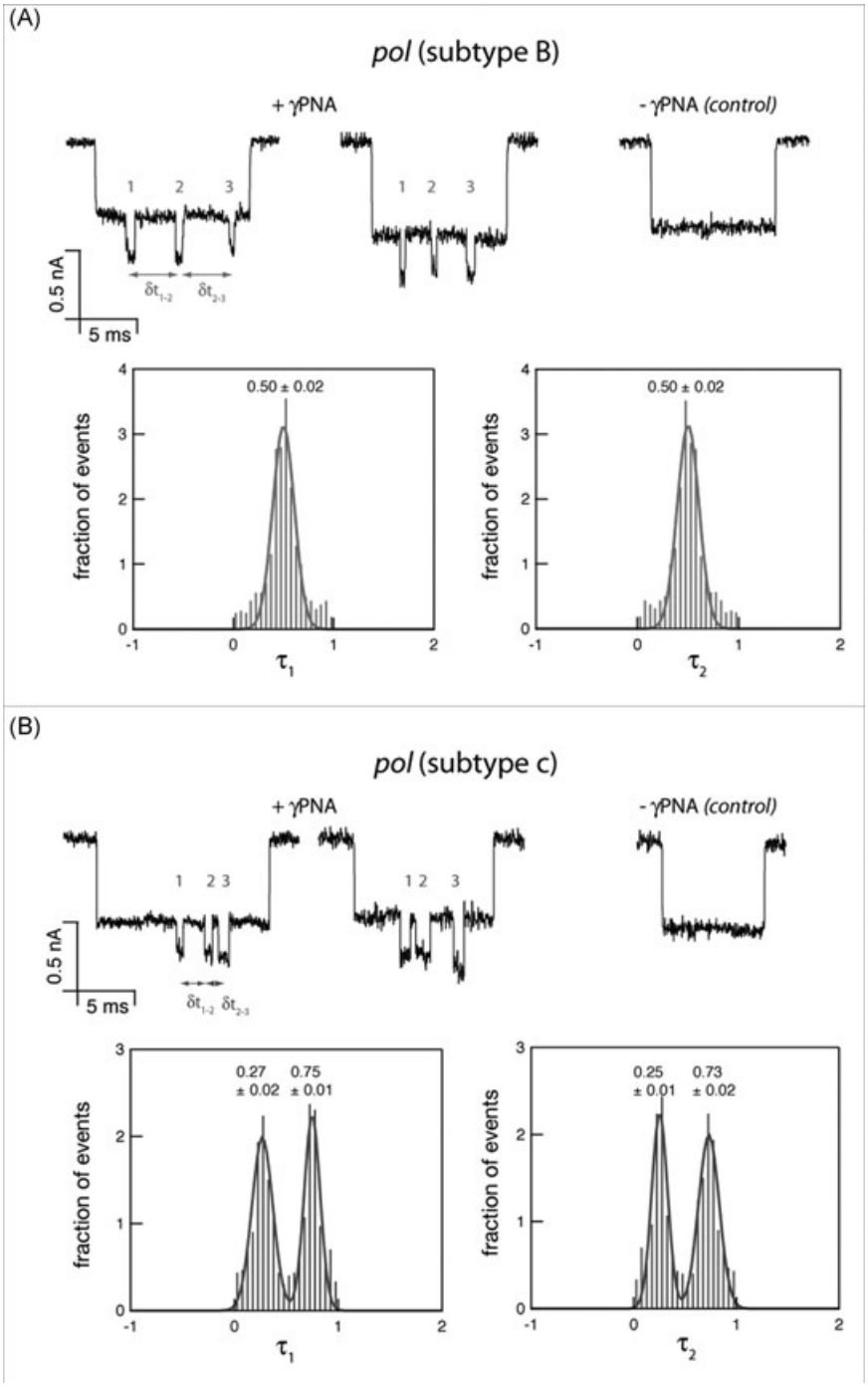


Figure 6. Nanopore-based analysis of γ-modified synthetic peptide nucleic acid (γPNA) tagged *pol* reveals two easily distinguished barcodes for each variant, while both untagged variants are indiscernible threaded through the nanopore. By normalizing the individual delay times between γPNA sites using the sum of all delay times in each event, a simple and robust parameter, τ, is obtained in for comparison of the different nanopore experiments. We find that (A) subtype B variant displays a single distribution for both τ_1 and τ_2, with a mean value of 0.500 $\pm$ 0.004. However, in the case of (B) subtype C variant—both normalized dwell times display two distinct populations with mean values of 0.269 $\pm$ 0.0045 and 0.749 $\pm$ 0.004 for τ_1, and 0.252 $\pm$ 0.004 and 0.731 $\pm$ 0.005 for τ_2, in excellent agreement with the ratio of the physical distances between γPNA probes (450 and 850 bp, respectively) using the calibration curve in Fig. 5. Figure reproduced from Singer et al. [5], with permission. © ACS publications.

signals serve as an excellent metric in quantifying the linear spacing along the DNA between the target sequences. In the case of HIV subtype B and subtype C, which display <8% sequence variance, identification is achieved with as few as four γPNA probes. Nanopore-based sensing, however, could in theory be applied to much longer DNA molecules with more tags, representing much more complex genomes [17]. Furthermore, by adding additional γPNA sites and by utilizing not only the spatial information (as done here) but also taking into account the number of sites, we would be able to substantially reduce the likelihood of incorrect classification. Thus, the method we describe can be used to detect and discriminate between genomes with relatively small sequence variation, simply by creating a spatially resolvable pattern of γPNA sites.

Recent technological advancements have allowed disease identification techniques to shift from antibody-based to nucleic acid based assays, enabling genotypic sensitivity, and will facilitate the continued transformation of these techniques from large laboratory-based sample testing to point-of-care

diagnostics [49]. Further advances in molecular diagnostics will allow subclassification of diseases with shared pathophysiological features but different response to treatment [50]. The results presented here, and the ease with which this process can be scaled up, suggest that a combined genomic bar-coding and purely electronic readout method is feasible. The γPNA/nanopore method presented here provides a way to quickly probe an unknown target and determine its exact nature. The sequence specificity of this approach is achieved by using the highly specific γPNA probes, which are more sensitive to sequence mismatches than DNA probes, and additionally enable multiplexing with γPNA libraries [51, 52]. The high-throughput, minute sample requirements, and resolution of this approach are achieved through the use of solid-state nanopores, one of the most sensitive single-molecule techniques devised to date. Combining these methods provides the foundation for a novel diagnostics platform, which can simplify the process of tailoring a patient's treatment regime to the strain or subtype of their disease. Such a tool could lead to both a reduction in healthcare costs and substantial quality-of-care benefits to patients.

4 Conclusions

Nanopore-based DNA analysis is an extremely attractive area of research, due to the simplicity of the method, and the ability not only to probe individual molecules, but also to detect a very small amount of initial genomic material. This review summarizes the proof-of-concepts for both an optical-based high-throughput nanopore sequencing technique and a solid-state nanopore-based method for purely electrical genotyping. The nanopore-based DNA sequencing method utilizes the DNA unzipping process to both slow down the translocation velocity and also to allow the sequential detection of fluorescently labeled beacons corresponding to DNA sequence. These results demonstrate the first solid-state nanopore-based DNA sequence readout. Further technical refinements will permit a highly parallel nanopore-based DNA sequencing system. The genomic bar-coding method uses γPNA to tag DNA, which can then be detected as it translocates through a nanopore. This novel diagnostic tool promises to improve diagnosis time and specificity of treatment for patients, ultimately bringing healthcare closer to its goals of personalized medicine. Ultrafast and amplification-free DNA sensing methods have numerous applications in biomedical research and in the diagnosis, characterization, and treatment of human diseases. The two nanopore-based approaches described here illustrate the great potential of this new platform in biomedical applications.

The authors acknowledge A. Squires for careful editing and commenting on the manuscript. Financial support from National Institutes of Health award R01 HG-005871 and National Science Foundation awards PHY-0646637 are acknowledged.

The authors have declared no conflict of interest.

5 References

[1] Shendure, J., Mitra, R. D., Varma, C., Church, G. M., *Nat. Rev. Genet.* 2004, *5*, 335–344.

[2] Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J. W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S. R., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H., Xie, Z., *Science* 2008, *320*, 106–109.

[3] Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., *Science* 2009, *323*, 133–138.

[4] Milos, P. M., *Expert Rev. Mol. Diagn.* 2009, *9*, 659–666.

[5] Singer, A., Rapireddy, S., Ly, D. H., Meller, A., *Nano Lett.* 2012, *12*, 1722–1728.

[6] McNally, B., Singer, A., Yu, Z., Sun, Y., Weng, Z., Meller, A., *Nano Lett.* 2010, *10*, 2237–2244.

[7] Rosenstein, J. K., Wanunu, M., Merchant, C. A., Drndic, M., Shepard, K. L., *Nat. Methods* 2012, *9*, 487–492.

[8] Branton, D., Deamer, D., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J. A., *Nat. Biotechnol.* 2008, *26*, 1146–1153.

[9] Kasianowicz, J., Brandin, E., Branton, D., Deamer, D., *Proc. Natl. Acad. Sci. U.S.A.* 1996, *93*, 13770–13773.

[10] Akeson, M., Branton, D., Kasianowicz, J., Brandin, E., Deamer, D., *Biophys. J.* 1999, *77*, 3227–3233.

[11] Meller, A., Nivon, L., Branton, D., *Phys. Rev. Lett.* 2001, *86*, 3435–3438.

[12] Bayley, H., Cremer, P., *Nature* 2001, *413*, 226–230.

[13] Deamer, D. W., Branton, D., *Acc. Chem. Res.* 2002, *35*, 817–825.

[14] Li, J., Stein, D., McMullan, C., Branton, D., Aziz, M. J., Golovchenko, J. A., *Nature* 2001, *412*, 166–169.

[15] Storm, A. J., Chen, J. H., Ling, X. S., Zandbergen, H. W., Dekker, C., *Nat. Mater.* 2003, *2*, 537–540.

[16] Kim, M. J., Wanunu, M., Bell, D. C., Meller, A., *Adv. Mater.* 2006, *18*, 3149–3153.

[17] Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y., Meller, A., *Nat. Nanotechnol.* 2009, *5*, 160–165.

[18] Wanunu, M., Sutin, J., McNally, B., Chow, A., Meller, A., *Biophys. J.* 2008, *95*, 4716–4725.

[19] Venkatesan, B. M., Bashir, R., *Nat. Nanotechnol.* 2011, *6*, 615–624.

[20] Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., Pavlenok, M.,

Niederweis, M., Gundlach, J. H., *Nat. Biotech.* 2012, *30*, 349–353.

[21] Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K., Akeson, M., *Nat. Biotechnol.* 2012, *30*, 344–348.

[22] Xie, P., Xiong, Q. H., Fang, Y., Qing, Q., Lieber, C. M., *Nat. Nanotechnol.* 2012, *7*, 119–125.

[23] Liu, Y. T., *Infect. Disord. Drug Targets* 2008, *8*, 183–188.

[24] Gulick, R. M., *J. Acquir. Immune Defic. Syndr.* 2010, *55*(Suppl 1), S43–S48.

[25] Phillips, K. M., Larson, J. W., Yantz, G. R., D'Antoni, C. M., Gallo, M. V., Gillis, K. A., Goncalves, N. M., Neely, L. A., Gullans, S. R., Gilmanshin, R., *Nucleic Acids Res.* 2005, *33*, 5829–5837.

[26] Miller, M. B., Tang, Y. W., *Clin. Microbiol. Rev.* 2009, *22*, 611–633.

[27] Kaittanis, C., Santra, S., Perez, J. M., *Adv. Drug Deliv. Rev.* 2010, *62*, 408–423.

[28] Wanunu, M., Sutin, J., Meller, A., *Nano Lett.* 2009, *9*, 3498–3502.

[29] Singer, A., Wanunu, M., Morrison, W., Kuhn, H., Frank-Kamenetskii, M., Meller, A., *Nano Lett.* 2010, *10*, 738–742.

[30] Nam, S.-W., Rooks, M. J., Kim, K.-B., Rossnagel, S. M., *Nano Lett.* 2009, *9*, 2044–2048.

[31] Lee, H.-S., Kim, B.-S., Kim, H.-M., Wi, J.-S., Nam, S.-W., Jin, K.-B., Arai, Y., Kim, K.-B., *Adv. Mater.* 2007, *19*, 4189–4193.

[32] Healy, K., *Nanomedicine* 2007, *2*, 459–481.

[33] Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., Huang, X., Jovanovich, S. B., Nelson, J. R., Schloss, J. A., Schwartz, D. C., Vezenov, D. V., *Nat. Biotechnol.* 2009, *27*, 1013–1023.

[34] Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., Pavlenok, M., Niederweis, M., Gundlach, J. H., *Nat. Biotechnol.* 2012, *30*, 349–353.

[35] Cherf, G. M., Lieberman, K. R., Rashid, H., Lam, C. E., Karplus, K., Akeson, M., *Nat. Biotechnol.* 2012, *30*, 344–348.

[36] McNally, B., Wanunu, M., Meller, A., *Nano Lett.* 2008, *8*, 3418–3422.

[37] Soni, V. G., Singer, A., Yu, Z., Sun, Y., McNally, B., Meller, A., *Rev. Sci. Instru.* 2010, *81*, 014301.

[38] Shendure, J., Porreca, G., Reppas, N., Lin, X., McCutcheon, J., Rosenbaum, A., Wang, M., Zhang, K., Mitra, R., Church, G., *Science* 2005, *309*, 1728–1732.

[39] Nielsen, P. E., *Chem. Biodivers.* 2010, *7*, 786–804.

[40] Sahu, B., Sacui, I., Rapireddy, S., Zanotti, K. J., Bahal, R., Armitage, B. A., Ly, D. H., *J. Org. Chem.* 2011, *76*, 5614–5627.

[41] Chenna, V., Rapireddy, S., Sahu, B., Ausin, C., Pedroso, E., Ly, D. H., *Chembiochem* 2008, *9*, 2388–2391.

[42] Rapireddy, S., He, G., Roy, S., Armitage, B. A., Ly, D. H., *J. Am. Chem. Soc.* 2007, *129*, 15596–15600.

[43] Kuhn, H., Bichismita, S., Rapireddy, S., Ly, D., Frank-Kamenetskii, M., *Artif. DNA PNA XNA* 2010, *1*, 45–53.

[44] He, G. F., Rapireddy, S., Bahal, R., Sahu, B., Ly, D. H., *J. Am. Chem. Soc.* 2009, *131*, 12088–12090.

[45] Gulick, R. M., *J. Acquir. Immune Defic. Syndr.* 2010, *55*, S43–S48.

[46] Mabey, D., Peeling, R. W., Ustianowski, A., Perkins, M. D., *Nat. Rev. Microbiol.* 2004, *2*, 231–240.

[47] Thompson, M. A., Aberg, J. A., Cahn, P., Montaner, J. S. G., Rizzardini, G., Telenti, A., Gatell, J. M., Gunthard, H. F., Hammer, S. M., Hirsch, M. S., Jacobsen, D. M., Reiss, P., Richman, D. D., Volberding, P. A., Yeni, P., Schooley, R. T., *JAMA* 2010, *304*, 321–333.

[48] Rodenburg, C. M., Li, Y. Y., Trask, S. A., Chen, Y. L., Decker, J., Robertson, D. L., Kalish, M. L., Shaw, G. M., Allen, S., Hahn, B. H., Gao, F., Isolati, U. N. N. H., *AIDS Res. Hum. Retroviruses* 2001, *17*, 161–168.

[49] Liu, Y.-T., *Infect. Disord. Drug Targets* 2008, *8*, 183–188.

[50] Green, E. D., Guyer, M. S., Inst, N. H. G. R., *Nature* 2011, *470*, 204–213.

[51] Zhang, N., Appella, D. H., *J. Infect. Dis.* 2010, *201*, S42–S45.

[52] Karkare, S., Bhatnagar, D., *Appl. Microbiol. Biotechnol.* 2006, *71*, 575–586.

