

Published in final edited form as:

*Nano Lett.* 2012 March 14; 12(3): 1722–1728. doi:10.1021/nl300372a.

## Electronic Barcoding of a Viral Gene at the Single-Molecule Level

Alon Singer<sup>1</sup>, Srinivas Rapireddy<sup>2</sup>, Danith H. Ly<sup>2</sup>, and Amit Meller<sup>1,\*</sup>

<sup>1</sup>Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

<sup>2</sup>Department of Chemistry, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

### Abstract

A new single-molecule approach for rapid and purely electronic discrimination among similar genes is presented. Combining solid-state nanopores and  $\gamma$ -modified synthetic peptide nucleic acid ( $\gamma$ PNA) probes, we accurately barcode genes by counting the number of probes attached to each gene, and measuring their relative spacing. We illustrate our method by sensing individual genes from two highly similar human immunodeficiency virus (HIV) subtypes, demonstrating feasibility of a novel, single-molecule diagnostic platform for rapid pathogen classification.

### Keywords

nanopore;  $\gamma$ PNA; sequence detection; molecular diagnostics; single-molecule; genotyping

Molecular assays for infectious disease diagnosis have revolutionized the speed and accuracy with which viral diseases can be identified.<sup>1</sup> Parallel advances in antiviral therapies require increasingly accurate and rapid pathogen classification on the genomic level.<sup>2</sup> However, current molecular diagnostic platforms are based on processes so elaborate and time-consuming that they typically must be performed in centralized labs. The introduction of single-molecule based diagnostic sensing platforms could simplify these processes, resulting in shorter turnaround times, and potentially allowing the platform to be used in resource limited arenas.<sup>3</sup> In this study, we report the application of a new type of single-molecule sensor, solid-state nanopores, to perform rapid molecular identification of viral genes and gene variants. This was achieved by combining highly sequence-specific  $\gamma$ -modified synthetic peptide nucleic acid ( $\gamma$ PNA) probes<sup>4-7</sup> with electrical sensing via solid-state nanopores, enabling rapid characterization of two nearly identical genes from two human immunodeficiency virus (HIV) subtypes (>92% similarity), which were originally extracted from unmarked human samples.

A solid-state nanopore is a synthetic device composed of a nanometer-scale pore fabricated in an ultra-thin membrane. Nanopores utilize an extremely simple electrical sensing principle: when biopolymers such as DNA or RNA are electrophoretically threaded across a pore, the ion current flowing through the pore is partially, and measurably, blocked.<sup>8,9</sup> It has recently been demonstrated that solid-state nanopores may be used to detect sub-molecular structural alterations in both DNA and RNA without the need for fluorescent/radio labels.<sup>10-12</sup> Another practical feature of solid-state nanopores is their ability to actively funnel and subsequently capture extremely small copy numbers of DNA molecules, simply

\*Corresponding author. ameller@bu.edu.

Supporting Information Available: Explanation of sample preparation protocols, control experiments, and additional data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

by establishing a salt gradient across the ultra-thin membrane.<sup>13</sup> The DNA funneling and the highly sensitive electrical detection capabilities place nanopores among the most promising candidate platforms for molecular diagnostics, particularly for those involving limited sample quantities.<sup>10, 14</sup> But despite these attributes, to date solid-state nanopores have seldom been utilized for molecular diagnostic sensing applications, primarily due to their inherent lack of DNA or RNA sequence specificity.

Recently, we reported that *bis*-PNA/DNA invasion sites can be detected with a nanopore, providing a strong basis for development of a nanopore-based genomic profiling method.<sup>15</sup> *bis*-PNA, which can bind and intercalate into target regions of DNA, alters the local structure of the DNA at the point of binding.<sup>4</sup> This structural change induces a local, detectable change in a solid-state nanopore's ionic conductance when that region passes through the pore, allowing it to be readily discriminated from bare DNA regions. However, there are a number of major drawbacks for the utilization of *bis*-PNA for nanopore-based molecular diagnostic applications: First, *bis*-PNA forms structural isomers during the invasion process,<sup>16</sup> leading to a noisy ion-current blockade signal. Second, *bis*-PNAs are restricted to purely homopurine target regions, severely limiting their general applicability for sequence-specific genotyping.<sup>4</sup>

In contrast, recently developed  $\gamma$ PNA probes exhibit an extremely-high affinity for DNA, and can invade with high specificity without any sequence restrictions.<sup>5-7, 17</sup> Moreover,  $\gamma$ PNA forms a more streamlined duplex structure as compared to the *bis*-PNA triplex structure (Figure 1a). We speculated that the more compact  $\gamma$ PNA /DNA duplex structure would allow us to utilize slightly smaller nanopores (as compared with those used for *bis*-PNA detection). By using slightly smaller nanopores, DNA translocation velocity is reduced,<sup>18</sup> substantially improving the temporal resolution for multiple  $\gamma$ PNA molecules bound to the same DNA target. To demonstrate the feasibility of this concept we first evaluated the ion current signals recorded for a ~ 3kbp DNA molecule, which includes three binding sites (15 bp long) for  $\gamma$ PNA molecules. When the bare DNA molecule was threaded through a 3.7 nm pore crafted in a 30 nm thick silicon nitride film we obtained a single blocked-level ion current traces, as shown in Figure 1c. This blocked current level corresponds well to the expected blockade values for dsDNA considering the pore diameter used.<sup>18</sup> Upon invasion of  $\gamma$ PNA probes, three additional blockade episodes are observed during each translocation event, all conforming to a well-defined second blocked ion-current level (see statistical analysis of >1,500 events in Supporting Information). The results shown in Figure 1 demonstrate that in principle we can target and detect short sequences (i.e. 15 bp) in any DNA molecule of interest, forming the basis for the proposed single-molecule gene identification method.

Ultimately, a nanopore/ $\gamma$ PNA-based genotyping method will both count the number of  $\gamma$ PNA/DNA sites per DNA molecule and also *localize* their positions along a DNA molecule, effectively barcoding the target. To demonstrate that nanopores can not only count bound  $\gamma$ PNA probes, but can also be used to accurately determine the distance between two  $\gamma$ PNA probes along a DNA molecule, we designed six model DNA molecules, each containing two identical  $\gamma$ PNA invasion sites to be used as 'molecular rulers' (see Supporting Information for the sample preparation procedure). These molecular rulers are used to calibrate the nanopore system, enabling a simple transformation from the nanopore's time domain data (the lag time between two  $\gamma$ PNA pulses,  $\delta t$ ) to the DNA's spatial coordinates (the gap in basepairs between the two  $\gamma$ PNA sites,  $\Delta n$ ). For these model molecules,  $\Delta n$  ranged from 100 to 1000 bp, while the two outer flanking regions of DNA were kept identical (~1,200 bp) for consistency, as shown in Figure 2a. Nanopore assays were conducted for each molecule to measure the average  $\delta t$  values for a statistical population of at least 1,000 molecules in each case. Typical examples of events for each

value of  $\Delta n$  are shown in Figure 2b (see Supporting Information for statistical analysis). Figure 2c depicts the average value of  $\delta t$  (error bars represent STD of the data), plotted as a function of the actual distance  $\Delta n$  between  $\gamma$ PNA sites. Figure 2c shows that  $\delta t$  can be accurately measured for even the shortest  $\Delta n$  tested (100 bp), corresponding to roughly a  $\sim 34$  nm spacing. The relationship between  $\delta t$  and  $\Delta n$  shown in Figure 2c can be used empirically to estimate the spacing between two unknown  $\gamma$ PNAs binding sites. We note that this relationship is well fitted by a power law with an exponent of  $1.39 \pm 0.09$ , in agreement with previous experiments that measured the translocation time as a function of length for untagged DNA molecules of similar length.<sup>18</sup>

Since target identification is established through a combination of both *probe counting* and *probe spacing*, it should be possible to design relatively small probe libraries that can uniquely identify a multitude of genomic targets by probing the incidence and spatial distribution of a small set of sequences. To evaluate the feasibility of this approach, we leveraged the high sequence specificity of 15-mer  $\gamma$ PNA probes<sup>17</sup> in a simple barcoding strategy: When a probe library set is added to the DNA sample of interest, only those probes which 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*, which are specific to a given library of probes. The pattern of  $\gamma$ PNA delay times acquired in a nanopore assay can then be directly compared against these maps to identify the presence (or absence) of DNA molecules of interest.

To illustrate this approach using genomic samples that naturally require the design and detection of multiple  $\gamma$ PNA sequences, we chose HIV genomes as a test system. Two nearly-identical DNA fragments containing the sequences for the *pol* gene (3,050 bp), extracted from unmarked HIV-1/B and HIV-1/C subtypes, were prepared using standard methods.<sup>19</sup> Four different  $\gamma$ PNA oligomers were synthesized in parallel, designed so that only two of the PNA oligomers would be a perfect match for *both* subtypes ( $\gamma$ PNA<sub>1</sub> and  $\gamma$ PNA<sub>2</sub>), while  $\gamma$ PNA<sub>3</sub> would bind to a sequence present only in subtype B and  $\gamma$ PNA<sub>4</sub> would bind to a sequence present only in subtype C (see Supporting Information). Binding the PNA probe libraries to each sample creates two distinct barcodes:  $\gamma$ PNA<sub>1</sub> and  $\gamma$ PNA<sub>2</sub> are spaced  $\sim 850$  bp apart and bind to the same locations on both subtypes,  $\gamma$ PNA<sub>3</sub> binds  $\sim 850$  bp after  $\gamma$ PNA<sub>2</sub> on subtype B only, and  $\gamma$ PNA<sub>4</sub> binds  $\sim 450$  bp after  $\gamma$ PNA<sub>2</sub> on subtype C only (Figure 3). Correct  $\gamma$ PNA binding patterns were confirmed by cutting each gene into ten fragments, each 300-400 bp, used for gel-shift binding assays to confirm binding specificity. These correct binding patterns were first demonstrated using individual types of  $\gamma$ PNA oligos, then confirmed using the complete library set. Figure 3 shows that  $\gamma$ PNA<sub>1</sub> and  $\gamma$ PNA<sub>2</sub> bind only to fragments VIII and V, respectively, in both *pol* subtypes. However,  $\gamma$ PNA<sub>3</sub> binds to fragment II *only* in subtype B (and not to the corresponding fragment II from subtype C). Likewise,  $\gamma$ PNA<sub>4</sub> binds *only* to fragment IV in subtype C (and not to fragment IV from subtype B). Furthermore, the full library containing all four  $\gamma$ PNA probes displayed identical binding patterns to those four probes bound separately. These gel-shift binding experiments confirm the sequence specificity of the  $\gamma$ PNA probes, since the nonbinding fragments of  $\gamma$ PNA<sub>3</sub> and  $\gamma$ PNA<sub>4</sub> were nearly identical in sequence, with only a few mismatches (see Supporting Information).

The nanopore method proposed here should provide a simple and efficient alternative to the laborious and time-consuming gel-based analysis described in Figure 3, only requiring a few minutes to make a single measurement for each of the two HIV genes subtypes. As expected, nanopore analysis of the two *untagged* subtypes produced practically indistinguishable data sets, due to the fact that the two DNA molecules are highly similar in length and sequence (see Supporting Information). In contrast, the two types of  $\gamma$ PNA-tagged samples each produced a unique ion current signal, which could be visually

distinguished: subtype B tagged molecules produced two nearly identical  $\delta t$  values ( $1.9 \pm 0.2$  ms and  $2.1 \pm 0.2$  ms), whereas subtype C produced two different  $\delta t$  values ( $0.79 \pm 0.10$  ms and  $2.1 \pm 0.3$  ms), corresponding to the locations of the bound PNA along the length of the DNA (see Supporting Information).

To simplify subtype classification, we defined the normalized first and second time delays 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. This normalization facilitated the analysis of hundreds of single-molecule events, allowing for machine-based discrimination between the two subtypes. Figure 4 displays the resulting distributions of  $\tau_1$  and  $\tau_2$  for subtype B and subtype C (red and blue, respectively, 750 events in each case). The symmetrically-tagged subtype B variant yields two nearly identical Gaussian distributions with a *single* peak at  $0.50 \pm 0.02$  (error represents STD), reflecting nearly equal delay times between the PNA probes, as expected. In contrast, the subtype C variant yielded *double* peak Gaussian distributions with peaks centered at  $0.27 \pm 0.02$  and  $0.75 \pm 0.01$  for  $\tau_1$ , and  $0.25 \pm 0.01$  and  $0.73 \pm 0.02$  for  $\tau_2$ , respectively. Referring back to our calibration curve (Figure 2) we note that the expected values of the delay time associated with the 450 bp and 850 bp gaps (0.61 ms and 1.85 ms, respectively) coincide with the observed mean values of  $\tau_1$  and  $\tau_2$ , since  $0.61/(0.61+1.85) = 0.25$  and  $1.85/(0.61+1.85) = 0.75$ , reflecting consistency with the calibration experiment. The appearance of two nearly identical double-peaked distributions for subtype C may be explained simply by the asymmetry of its corresponding barcode and the fact that either end of the DNA can enter the nanopore first. We do note that the absolute values of  $\delta t$  in the calibration assay are slightly shorter than the values obtained in the current experiment. These small variations can be explained by slight differences in pore geometries used in these experiments. Nevertheless, the internal normalization of the delay times appears to nullify slight pore-to-pore variations, enabling a more straightforward and simpler analysis.

The data in Figure 4 demonstrates that our method can readily discriminate between these two highly similar HIV subtypes. In the case of HIV subtype B and subtype C, which display <8% sequence variance, identification is confirmed here with just four unique  $\gamma$ PNA probes. Nanopore sensors are easily extensible to characterization of much longer DNA molecules, representing significantly more complex genomes. The method we describe here can thus be used to detect and discriminate among a wide range of highly similar genomes based on subtle variations in sequence. Improved sensitivity or broader applicability of this technique can be achieved by introducing more  $\gamma$ PNA probes into the library, utilizing the ability to accurately resolve the spatial pattern of  $\gamma$ PNA sites along the DNA molecule screened without significantly complicating the readout process. The effect of small pore-to-pore variations, which might slightly affect translocation velocity (and therefore  $\delta t$ ) of tagged DNA, can be compensated for by employing the simple normalization method introduced in Figure 4.

Recent technological advances have allowed disease identification techniques to shift from antibody- 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.<sup>1</sup> Further advancements in molecular diagnostics will allow sub-classification of diseases with certain shared pathophysiological features but different responses to treatment.<sup>20</sup> The results presented here support a new nanopore-based genomic barcoding and electronic readout method that can quickly identify genes from an unknown sample using only a small  $\gamma$ PNA library. The sequence sensitivity of this approach is achieved through the use of specific  $\gamma$ PNA probes, which are far more sensitive to sequence mismatches than DNA probes, and which can be incorporated into thoughtfully designed  $\gamma$ PNA libraries to expand the range of identifiable target molecules.<sup>21</sup> The high throughput, minute sample requirements, and spatial resolution of this approach are all achieved through

the use of solid-state nanopores. By combining nanopores with  $\gamma$ PNA probes, we were able to detect and classify genes of closely related viral subtypes, demonstrating the potential of this novel diagnostic platform.

## Methods

$\gamma$ PNA oligomers were synthesized according to published procedures.<sup>5, 6</sup> Upon cleavage from the resin, oligomers were purified by reverse-phase HPLC to homogeneity and verified by MALDI-TOF mass spectrometry.  $3.7 \pm 0.3$  nm single nanopores were fabricated in 30 nm thick, low-stress LPCVD deposited SiN membranes ( $5 \times 5 \mu\text{m}^2$ ), supported by a 380  $\mu\text{m}$  thick Si chip with a 1  $\mu\text{m}$  layer of SiO<sub>2</sub>, using a focused electron beam as previously described.<sup>22</sup> Prior to use, nanopore chips were boiled for 15 min in a Piranha solution (2 parts 95% H<sub>2</sub>SO<sub>4</sub>, 1 part 35% H<sub>2</sub>O<sub>2</sub>) and assembled in a custom-designed cell under controlled atmospheric conditions as previously described.<sup>23</sup> After assembly, nanopores were hydrated with a degassed and filtered KCl electrolyte solution (adjusted to pH 7.5 using 10 mM Tris-HCl). The nanopore cell was then placed in a custom-designed chamber, featuring thermoelectric regulation within  $\pm 0.1^\circ\text{C}$ , rapid thermal equilibration ( $< 5$  min), and electromagnetic shielding. Homemade Ag/AgCl electrodes were immersed into each chamber of the cell and connected to an Axon 200B (Molecular Devices, CA, USA) headstage. All measurements were performed inside a dark Faraday cage.

All experiments used 0.2M/1M KCl solutions (*cis* and *trans* chambers, respectively) and an applied voltage of 300 mV. Asymmetric electrolyte conditions were employed to yield increased dsDNA capture rates and extended dwell times, as compared with symmetric salt concentration solutions.<sup>13</sup> In a typical experiment, DNA solutions at 0.1 fmole/ $\mu\text{l}$  were added to the *cis* chamber, yielding a DNA capture rate of  $\sim 1,000$  events/10 min. Translocation data was acquired using custom-built LabVIEW (National Instruments, TX, USA) software, which collects both continuous current data as well as individual ion current blockade pulses (“translocation events”) detected in real-time. Translocation events were identified based on two criteria: 1) A minimum dwell time of 100  $\mu\text{s}$ ; and 2) A minimum current blockade amplitude of 0.35 nA before returning to the original “open pore” state. The analog signal from the Axon amplifier was first low-pass filtered with a 50 kHz Butterworth filter (typical RMS noise values of 40-50 pA) before being fed into a DAQ card (PCI-6230, National Instruments, TX, USA), which sampled the data at 250 kHz/16 bit. Data was exported to Igor Pro software, where all data was analyzed. DNA and PNA/DNA blockade levels were determined by creating an all-points-histogram for each translocation event and identifying peaks for each blockade level. PNA/DNA complex sites were identified by a characteristic blockade level of  $> 2.5 \times \text{RMS}$ .

Near-full length bacterial clones of both HIV subtype variants were obtained from the AIDS Research and Reference Reagent program, NIAID, NIH. Subtypes HIV-1/B and HIV-1/C, reagent numbers p98CN009.8 and p98IS002.5, respectively, were received from Drs. Cynthia M. Rodenburg, Beatrice H. Hahn, Feng Gao, and the UNAIDS Network for HIV Isolation and Characterization.<sup>19</sup> Clones were selectively grown on either 50  $\mu\text{g}/\text{ml}$  Ampicillin or 50  $\mu\text{g}/\text{ml}$  Kanamycin, and plasmids were isolated, purified, and sequenced to ensure that plasmid sequences matched those obtained from Genebank.<sup>19</sup> Isolation of the *pol* gene from the two HIV subtypes was achieved through standard PCR reactions. Given the near-identical nature of the *pol* sequence for the two subtypes used here, identical primer sequences and PCR conditions were employed to amplify this gene. PCR amplicons were purified and sequenced to ensure sequence validity and analyzed on a 12 kbp Experion DNA chip (BioRad) (see Figure S3a in Supporting Information).

Based on our sequencing results for the two *pol* genes, four  $\gamma$ PNA target regions were selected to barcode the *pol* gene. The first target site was placed ~800 bp from one end of the dsDNA, ensuring that the DNA blockade level would be recognizable prior to that of the first observed  $\gamma$ PNA/DNA blockade level. The first two targets, Site 1 and Site 2, are common to both HIV-1 subtypes in both sequence and location, and are spaced ~850 bp apart. All target site sequences and their corresponding  $\gamma$ PNA probe sequences are presented in Table 1 found in the Supporting Information.

The target sequence that is perfectly complementary to  $\gamma$ PNA<sub>3</sub> is located *only* on HIV-1/B, and is located ~850 bp upstream from Site 2. The sequence found in the same location on subtype C is shown in Table 1 as well, with the mismatched bases in red. The target sequence that is perfectly complementary to  $\gamma$ PNA<sub>4</sub> is located *only* on HIV-1/C and is located ~450 bp upstream from Site 2. Similarly, Table 1 shows the perfect complementarity of Site 4 with  $\gamma$ PNA<sub>4</sub>, while the mismatches that prevent binding of  $\gamma$ PNA<sub>4</sub> to subtype B are shown in red. In all cases, 'X' denotes the synthetic nucleobase 9-(2-guanidinoethoxy) phenoxazine or 'G-clamp' as described previously.<sup>6</sup> All other nucleobases are natural, and the  $\gamma$ PNA oligomer was synthesized and purified as previously described.<sup>6,7</sup>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

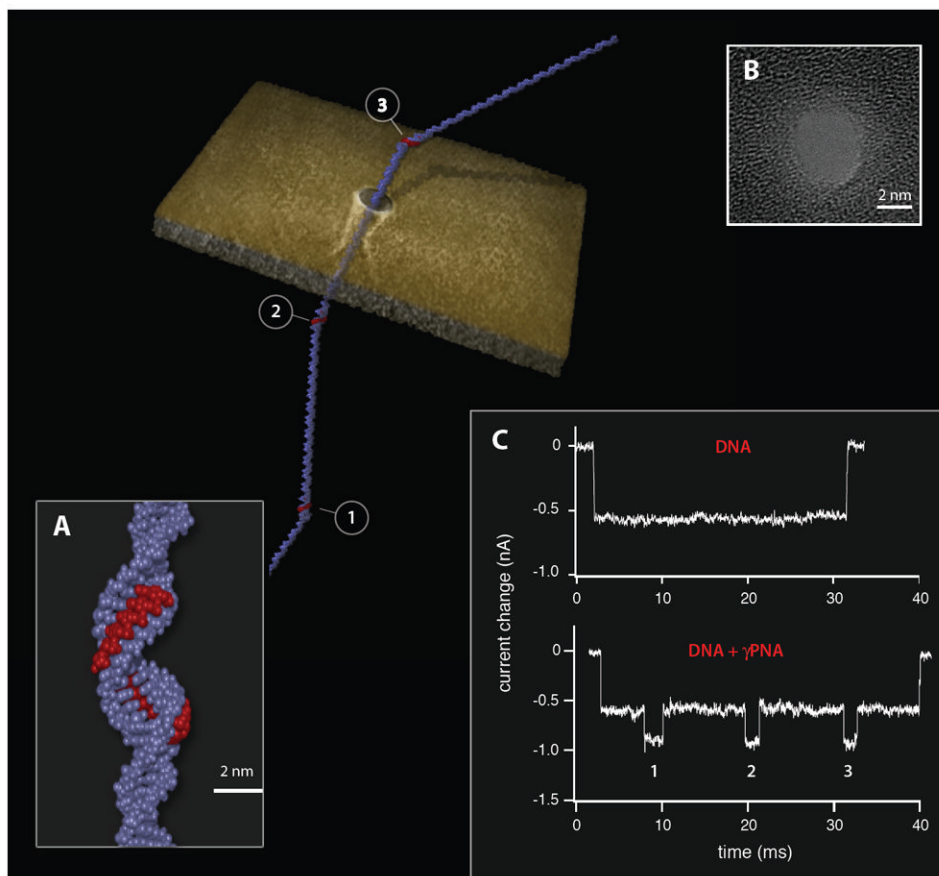
## Acknowledgments

The authors acknowledge support from Harvard's Center for Nanoscale Systems (CNS) and financial support from National Institutes of Health award R01 HG-005871 (A.M.) and National Science Foundation awards PHY-0646637 (A.M.) and CHE-1012467 (D.H.L.), as well as the DSF Charitable Foundation (D.H.L.). The authors thank A. Squires for editing the manuscript.

## References

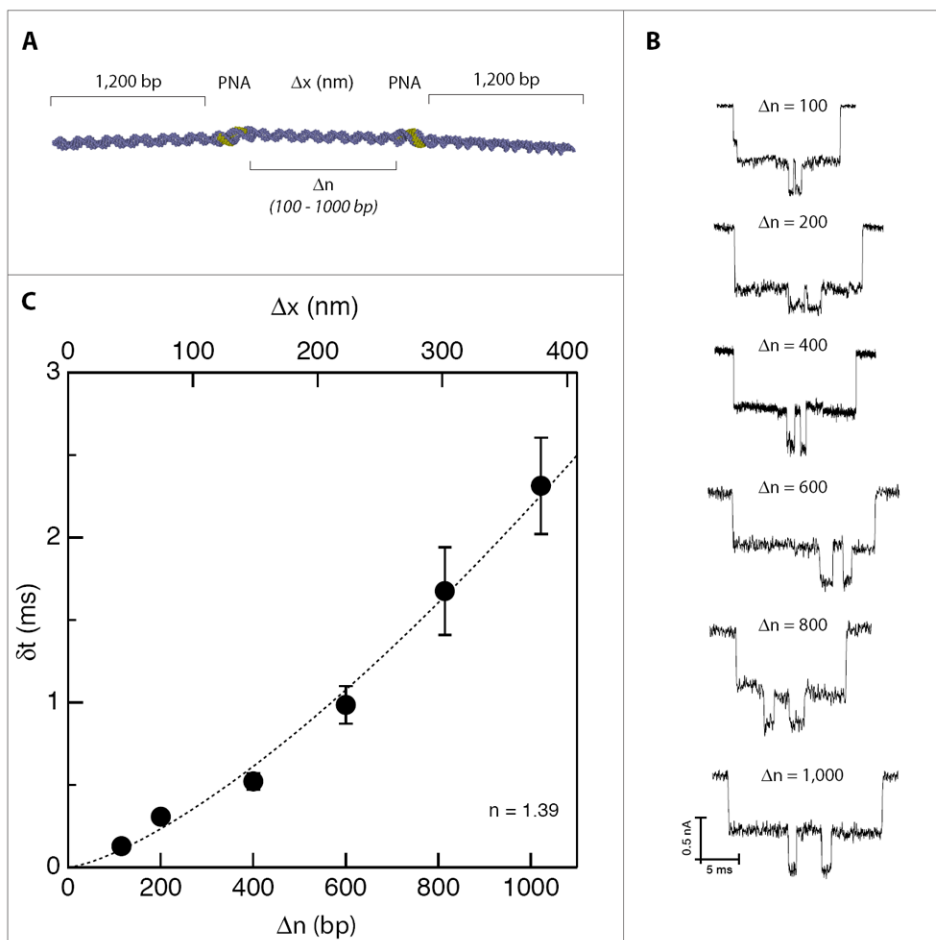
1. Liu Y-T. Infect Disord Drug Targets. 2008; 8(3):183–188. [PubMed: 18782035]
2. Domiati-Saad R, Scheuermann RH. Clin Chim Acta. 2006; 363(1-2):197–205. [PubMed: 16098959]
3. Kaittanis C, Santra S, Perez JM. Adv Drug Deliver Rev. 2010; 62(4-5):408–423.
4. Nielsen PE. Chem Biodivers. 2010; 7(4):786–804. [PubMed: 20397216]
5. Sahu B, Sacui I, Rapireddy S, Zanotti KJ, Bahal R, Armitage BA, Ly DH. J Org Chem. 2011; 76(14):5614–5627. [PubMed: 21619025]
6. Chenna V, Rapireddy S, Sahu B, Ausin C, Pedroso E, Ly DH. Chembiochem. 2008; 9(15):2388–2391. [PubMed: 18816545]
7. Rapireddy S, He G, Roy S, Armitage BA, Ly DH. J Am Chem Soc. 2007; 129(50):15596–15600. [PubMed: 18027941]
8. Deamer DW, Branton D. Accounts Chem Res. 2002; 35(10):817–825.
9. Bayley H, Cremer PS. Nature. 2001; 413(6852):226–230. [PubMed: 11557992]
10. Venkatesan BM, Bashir R. Nat Nano. 2011; 6(10):615–624.
11. Wanunu M, Dadosh T, Ray V, Jin J, McReynolds L, Drndic M. Nat Nano. 2010; 5(11):807–814.
12. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang XH, Jovanovich SB, Krstic PS, Lindsay S, Ling XSS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA. Nat Biotechnol. 2008; 26(10):1146–1153. [PubMed: 18846088]
13. Wanunu M, Morrison W, Rabin Y, Grosberg AY, Meller A. Nat Nano. 2010; 5(2):160–165.
14. Wang Y, Zheng D, Tan Q, Wang MX, Gu L-Q. Nat Nano. 2011; 6(10):668–674.
15. Singer A, Wanunu M, Morrison W, Kuhn H, Frank-Kamenetskii M, Meller A. Nano Lett. 2010; 10(2):738–742. [PubMed: 20088590]

16. Hansen GI, Bentin T, Larsen HJ, Nielsen PE. *J Mol Biol.* 2001; 307(1):67–74. [PubMed: 11243804]
17. He GF, Rapireddy S, Bahal R, Sahu B, Ly DH. *J Am Chem Soc.* 2009; 131(34):12088–12090. [PubMed: 19663424]
18. Wanunu M, Sutin J, McNally B, Chow A, Meller A. *Biophys J.* 2008; 95(10):4716–4725. [PubMed: 18708467]
19. Rodenburg CM, Li YY, Trask SA, Chen YL, Decker J, Robertson DL, Kalish ML, Shaw GM, Allen S, Hahn BH, Gao F, Isolati UNNH. *AIDS Res Hum Retrov.* 2001; 17(2):161–168.
20. Green ED, Guyer MS, Inst NHGR. *Nature.* 2011; 470(7333):204–213. [PubMed: 21307933]
21. Zhang N, Appella DH. *J Infect Dis.* 2010; 201:S42–S45. [PubMed: 20225945]
22. Kim MJ, Wanunu M, Bell DC, Meller A. *Adv Mater.* 2006; 18(23):3149–3153.
23. Wanunu, M.; Meller, A. Single Molecule Analysis of Nucleic Acids and DNA-protein Interactions using Nanopores. In: Selvin, P.; Ha, T., editors. *Laboratory Manual on Single Molecules.* Cold Spring Harbor Press; 2008.

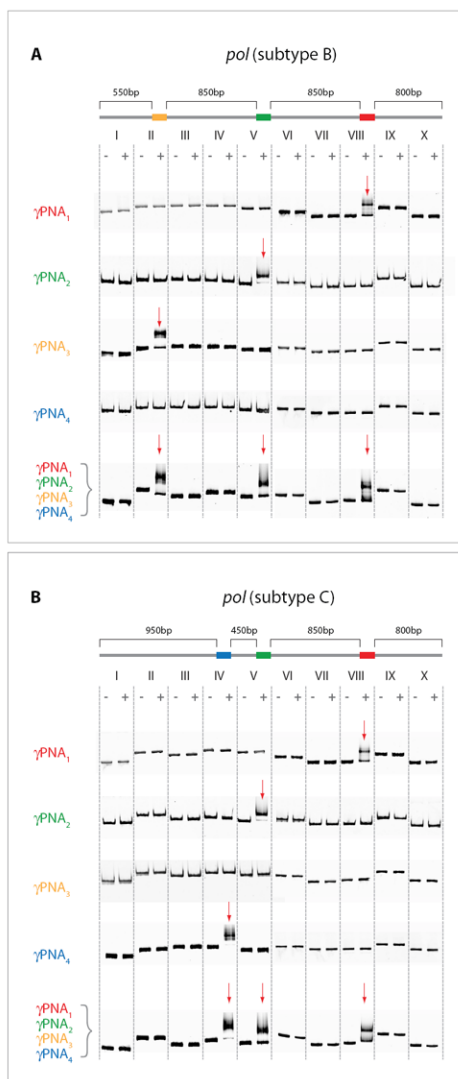


**Figure 1.** Electronic barcoding of a single DNA molecule using the nanopore/PNAs system. Illustration of a DNA molecule tagged with three equally spaced  $\gamma$ PNA probes, translocating across a 3.7 nm solid-state nanopore sculpted in a thin silicon nitride membrane. A) Close-up approximated structure of a  $\gamma$ PNA probe site. Given its high affinity for DNA, the single-stranded  $\gamma$ PNA (red) binds to its complementary DNA sequence, locally displacing the second DNA strand. This process leads to a local change in the helical structure, which is detectable with the solid-state nanopore. B) High-resolution TEM image of a 3.7 nm nanopore, made in a thin silicon nitride film. C) In contrast to the translocation of an untagged or bare DNA molecule depicted by a single blockade episode for the duration of a translocation event, in the case of a  $\gamma$ PNA tagged DNA molecule three distinct secondary blockade episodes are present, clearly marking the points in time when the  $\gamma$ PNA-tagged region crossed through the nanopore.

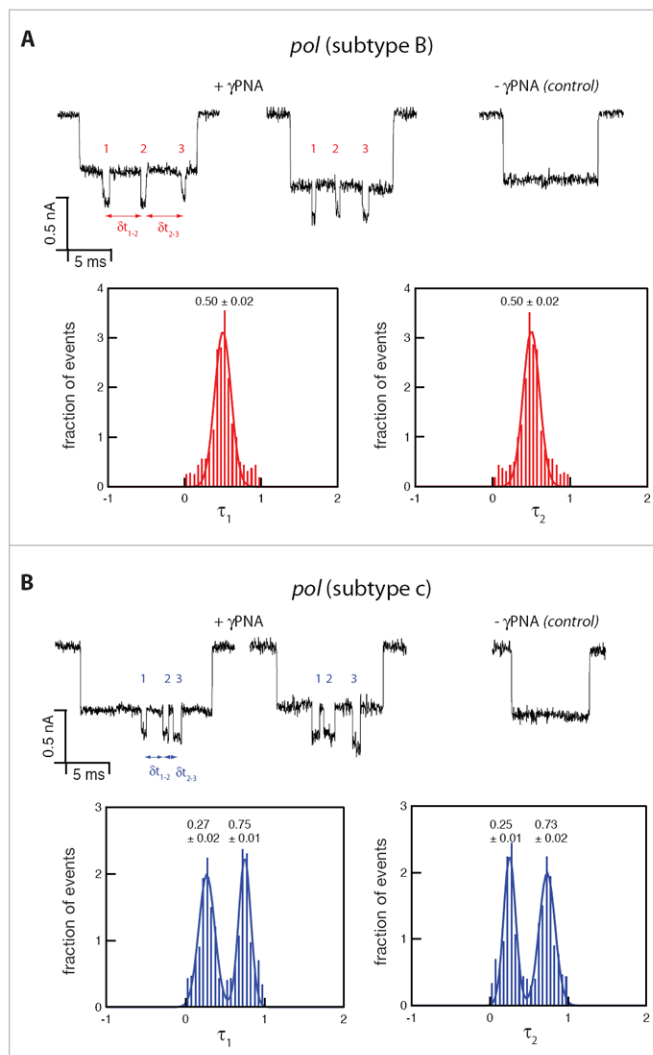




**Figure 2.** Varying the distance between two  $\gamma$ PNA sites to create a time ( $\delta t$ ) to distance ( $\Delta n$  or  $\Delta x$ ) conversion map. A) Six nearly identical and symmetrical DNA molecules were created, consisting of identical flanking regions (1,200 bp before the first  $\gamma$ PNA site and 1,200 bp after the second  $\gamma$ PNA site), and incorporating varied spacing between the two  $\gamma$ PNA sites; B) Representative single-molecule traces illustrating the time-delay differences,  $\delta t$ , among the various molecules; C) Statistical analysis further illuminating the relationship between  $\delta t$  and the  $\gamma$ PNA site-to-site spacing (as measured in bp). The data is well approximated by a power law dependence with an exponent of 1.39 (dashed line).



**Figure 3.** Mapping of the HIV *pol* gene. The 3,050 bp long *pol* gene - A)-subtype B, and B)-subtype C - was split into ten fragments, allowing us to determine onto which fragment a particular  $\gamma$ PNA probe would bind using gel shift assays (see red arrow).  $\gamma$ PNA<sub>1</sub> and  $\gamma$ PNA<sub>2</sub> bound to fragments VIII and V, respectively, on both HIV subtypes.  $\gamma$ PNA<sub>3</sub> bound to fragment II on subtype B only, while  $\gamma$ PNA<sub>4</sub> bound to fragment IV on subtype C only. Mixing all four  $\gamma$ PNA probes displayed the same binding pattern for the individual subtypes as was observed for each probe individually.



**Figure 4.** Nanopore-based analysis of  $\gamma$ PNA-tagged HIV *pol* gene displaying two unique barcodes for each of the variants, while both untagged variants are indistinguishable. The distributions of the normalized delay times,  $\tau_1$  and  $\tau_2$  measured for the two HIV-subtypes (750 events in each case). A) Subtype B variant displays a single distribution for both  $\tau_1$  and  $\tau_2$ , with a mean $\pm$ STD value of  $0.50\pm 0.02$ . B) Subtype C variant, displays two distinct populations with mean $\pm$ STD values of  $0.27\pm 0.02$  and  $0.75\pm 0.01$  for  $\tau_1$ , and  $0.25\pm 0.01$ ,  $0.73\pm 0.02$  for  $\tau_2$ .