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

Supplemental Information

Alon Singer,[1] Srinivas Rapireddy,[2] Danith H. Ly[2] and Amit Meller[1,*]

1. Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215
2. Department of Chemistry, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

* Corresponding author. E-mail: ameller@bu.edu

1. General

Oligonucleotides were purchased from Integrated DNA Technologies (IA, USA). Plasmid pBR322. All ladders and enzymes used in this study were purchased from New England Biolabs (MA, USA). Plasmid pET-Duet1 was purchased from EMD Biosciences (NJ, USA). Sephadex G-100 was purchased from GE Healthcare (NJ, USA). All PCR fragments and gel extractions were purified using Invitrogen PureLink kits (CA, USA). All plasmid purifications were performed using Invitrogen PureLink Miniprep kits (CA, USA). All DNA sequencing reactions were processed through Genewiz (MA, USA).

Gel shift assays were used to confirm binding of a γPNA oligomer to a DNA fragment. Samples (typically ~40 ng/lane) were resolved using 8% native polyacrylamide gels (29:1 w/w acrylamide:N,N'-methylene-bis-acrylamide) which were run for ~3 h at 12 V/cm in 1x TBE buffer. Gels were post-stained using SYBR Green I nucleic acid intercalating staining dye (Invitrogen, CA, USA) and imaged using a PharosFX molecular imaging scanner (Bio-Rad, CA, USA).

2. γPNA binding procedure

γPNA oligomers were bound to the target dsDNA in a solution containing TE (pH 7.4), 5% (w/w) polyethylene glycol (PEG) 6000, and 30 mM NaCl. The mixture was incubated at 65°C for 60 min. Post-incubation, PEG and excess γPNAs were removed using TE pre-equilibrated Sephadex G-100

columns. Under these conditions, binding saturation (as determined via quantitative image analysis of gel shift assays) occurred within ~15 min, with a binding yield of > 90%.

3. Molecular ruler preparation

Figure S1a illustrates the construction of the six calibration DNA molecules. The molecules only differed in the inter-site spacing, which was varied from 100-1000 bp. The flanking regions located before the first γPNA site and after the second γPNA site were identical in length for all samples. The γPNA sequence used in this study was: H-Lys-TXGTAAATGXCGCXC-Lys-NH_2, where 'X' denotes a synthetic nucleobase 9-(2-guanidinoethoxy) phenoxazine or 'G-clamp'.[2] All other nucleobases are natural. The γPNA oligomers (which contained (*R*)-diethylene glycol unit at the γ-backbone position) were synthesized and purified as previously described.[3]

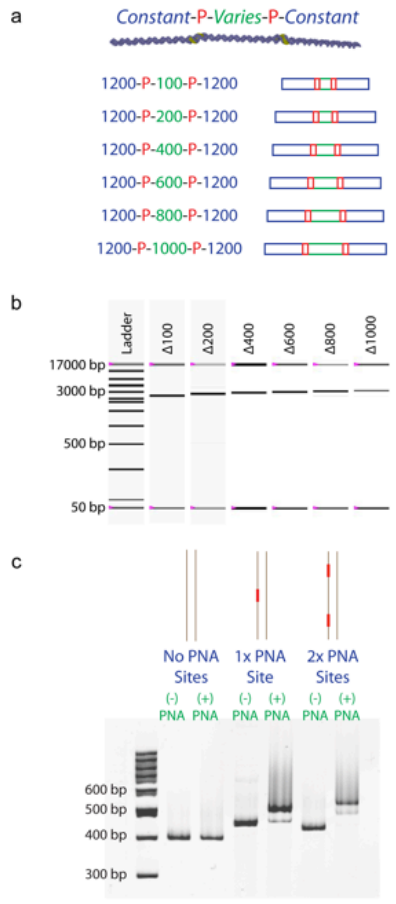

Figure S1: (a) Illustration of the six nearly identical sample molecules with varying distance between two identical γPNA sites (100-1000 bp apart). In all cases, the region upstream of the first γPNA site and downstream of the second γPNA site was held constant at 1,200 bp. These samples were designed to help illuminate the relationship between δ*t* and the actual spacing between γPNA sites. (b) Gel analysis of these six molecules, showing total sample lengths ranging from 2500 bp (100 bp spacing) to 3400 bp (1000 bp spacing). (c) Gel shift analysis indicating binding of the γPNA oligomer to a γPNA target site on sample DNA. Gel shift analysis confirms that when no sites are present, binding does not occur. However, when one or more sites are present, the γPNA probe binds in a stoichiometric fashion to the number of sites present.

Target binding sites for the γPNA were constructed by cloning the appropriate oligonucleotides nd/or constructed PCR fragments into the EcoRI/HindII and the BglII/AvrII sites of pET-Duet1 using standard methods.[4] Constructs were incorporated and amplified using Agilent XL10 Gold ultra-competent cells. Cells were grown selectively on 100 μg/ml Ampicilin, plasmids were isolated and purified.

Extracted plasmids were sequenced to ensure γPNA target binding site presence and spacing. For each spacing (100-1000 bp), forward primers were designed to be ~1200 bp upstream from the first γPNA target site and reverse primers was designed to be ~1200 bp downstream from the second γPNA target site. PCR constructs were purified and sequenced for validity, then run on a 12 kbp Experion DNA chip using an Experion automated electrophoresis system (Bio-Rad, CA, USA). Results are displayed in Figure S1b. Total length for the six different molecules ranged from ~2500 bp to ~3400 bp.

Because the mobility shift caused by the binding of a small γPNA oligomer onto a very long dsDNA molecule is relatively small, all gel shift assays were conducted using PCR fragments that contained the same plasmid templates described above, but with shorter "flanking region" i.e. shorter distances up- and downstream from the target sites (~100 bp). Figure S1c displays the binding signatures for a DNA fragments that contain zero binding sites, one binding site, and two binding sites for the γPNA oligomer (1:60 DNA/γPNA ratio). We show here that if a binding site is not present, the γPNA does not bind, in excellent agreement with previous studies[5], whereas if a single binding site is present on the DNA, we can resolve a second, lower-mobility band indicative of a DNA fragment with a bound γPNA probe. When two sites are present on the DNA, we find that an additional band with even lower mobility appears, indicating the presence of DNA fragments with two bound γPNA probes.

4. Nanopore signal analysis

Nanopore measurements were conducted as described in the main text. Within a translocation event, we define the two secondary blockade episodes by denoting their individual blockade levels and dwell times as $\Delta I_{PNA,1}$, $t_{PNA,1}$ and $\Delta I_{PNA,2}$, $t_{PNA,2}$ for the first and second secondary blockade episodes, respectively. Moreover, we define $\delta t_{1\text{-}2}$ as the delay time between two γPNA sites, as measured from the

midpoint of the first γPNA/DNA-to-DNA level transition to the midpoint of the second DNA-to-γPNA/DNA level transition (see Figure S2a). For the calibration measurements presented in Figure 2 of the main text, we analyzed ~1,000 events for each of the six molecules. For all the molecules the normalized blockade levels for γPNA sites, $\Delta I_{PNA}/\Delta I_{DNA}$, are approximately equal to 0.45±0.04. (see Figure S2b). Furthermore, we find that across all six DNA molecules, t_{PNA} times remained constant having a typical value of 0.55±0.06 ms (see Figure S2c). These results show that varying the distance between the two γPNA sites clearly leads to a shift in the $\delta t_{1\text{-}2}$ times. Even at the smallest distance measured here (100 bp or ~34 nm) the signals arising from the two γPNA sites are independent of each other.

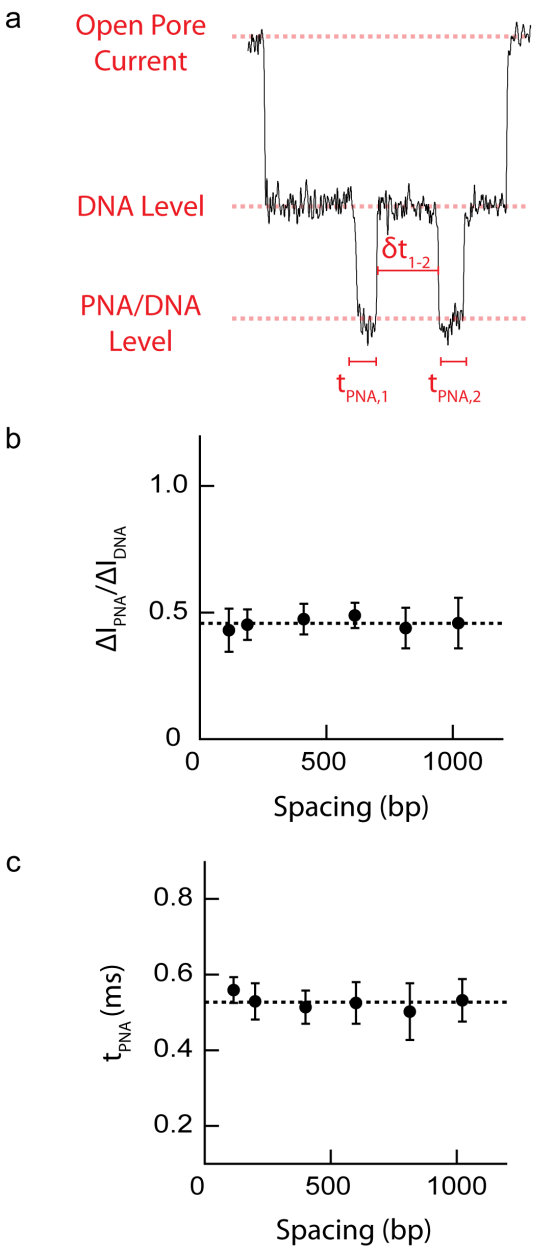

Figure S2. (a) Analyzing a DNA molecule tagged with two γPNAs was conducted in a manner similar to that which was described in previously.[1] Here, however, these analysis methods have been applied to the two secondary blockade episodes, which can be characterized individually by their blockade and dwell times. Additionally, here it was also necessary to include the delay-time ($\delta t_{1\text{-}2}$) between the secondary blockages in our analysis. Through statistical analysis of thousands for events of each of the different sample PNA target spacings, we have found that (b) $\Delta I_{PNA}/\Delta I_{DNA} = 0.45 \pm 0.04$, and that (c) $t_{PNA} = 0.55 \pm 0.06$ ms, independent of the probe-to-probe spacing. 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.

5. Nanopore assay of untagged *pol* gene

As a control for further nanopore experiments, we conducted a nanopore-based analysis of both subtypes, *untagged*. Here, as expected, we found that the minute sequence differences present between the two subtypes were insufficient to allow distinction via the nanopore method alone (see 2D histograms in Figure S3b and S3c).

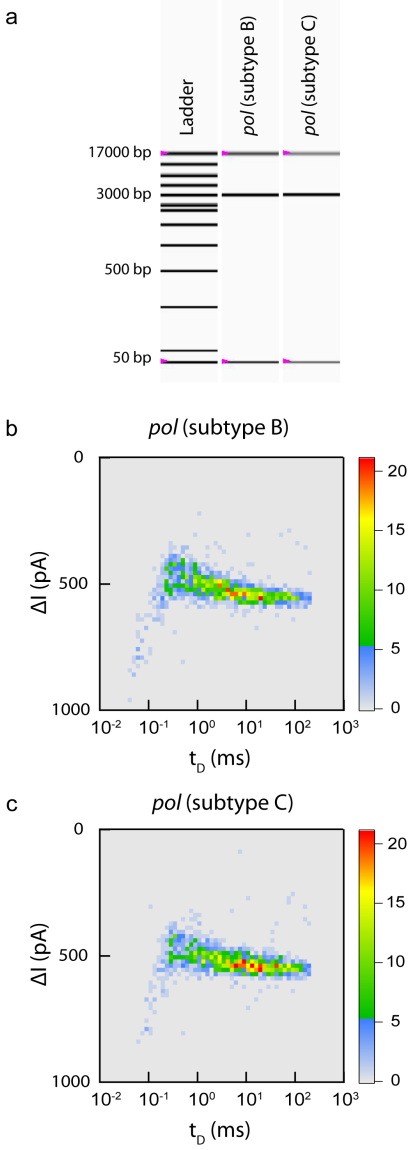

Figure S3. (a) Gel analysis of the *pol* gene for both subtypes as indicated (~3,050 bp). 2D histogram of nanopore-based analysis (change in current versus dwell time) for both (b) subtype B and (c) subtype C. In their untagged form, given the similar size and sequence of subtypes B and C, the nanopore is unable to distinguish between the two subtypes. Color represents the density of events for the 2-D histograms.

6. HIV *pol* mapping via gel shift assays

Previous γPNA studies have incorporated no more than a single target site per DNA fragment. Since we aim to target each DNA fragment with three different probes simultaneously, while maintaining specificity against a fourth probe, standard gel shift assays are required to prove that each probe binds only to its pre-designed target site.

Site/Probe Number	**Target Sequence** (5' – 3')	***γPNA***
1	TTTAATCGTCCACCC	H-Lys-TTTAATXGTXCAXCC-Lys-NH_2
2	GACCCATCAAAAGAC	H-Lys-GAXCXATXAAAAGAC-Lys-NH_2
3	Subtype B: GCATTAACAGCAATT Subtype C: GCATTAATGGAGATT	H-Lys-GXATTAAXAGXAATT-Lys-NH_2
4	Subtype B: CTTTTAGAAAACAAA Subtype C: CCTTCAGGGCACAAA	H-Lys-CXTTXAGGGXACAAA-Lys-NH_2

Table 1. The four target site sequences for the *pol* gene of HIV subtypes B and C and their corresponding γPNA probe sequences. Mismatches are marked in red.

Ten sequential regions from each of the *pol* genes from subtypes B and C were PCR amplified using standard methods (see Figure S4). Because both subtypes B and C are extremely similar at the genomic level, particularly the *pol* gene (<8% variance), we were able to use the exact same primer pairs and PCR conditions for both subtypes for all fragments. Each fragment measured 300-400 bp in length. Where possible, the forward primer sequence for fragment 'x' was designed to be exactly complementary to the reverse primer sequence used on the previous fragment 'x-1'. All ten fragments contain overlapping sequences at the very ends, ensuring that the entirety of the *pol* gene was tested for both subtypes. All DNA fragments were sequenced to verify identity.

Gel shift assays for each of the ten fragments for both subtypes were conducted using similar conditions to those indicated previously. Initially, only one γPNA probe was introduced to sample fragments at a time. Our results showed that the individual γPNA probes are indeed specific and display an affinity only for the fragment which contained its corresponding exact target sequence. Next, a "library mixture" of all four γPNA probes was created to test whether the binding competition or self-binding of different γPNA probes together would reduce the binding affinity or sequence specificity of probe invasion. High concentrations of γPNA are known to aggregate, but our optimized conditions prevented aggregation and yielded DNA/*correct*-γPNA ratios of ~1:30, allowing us to attain the correct binding pattern with all targets bound.

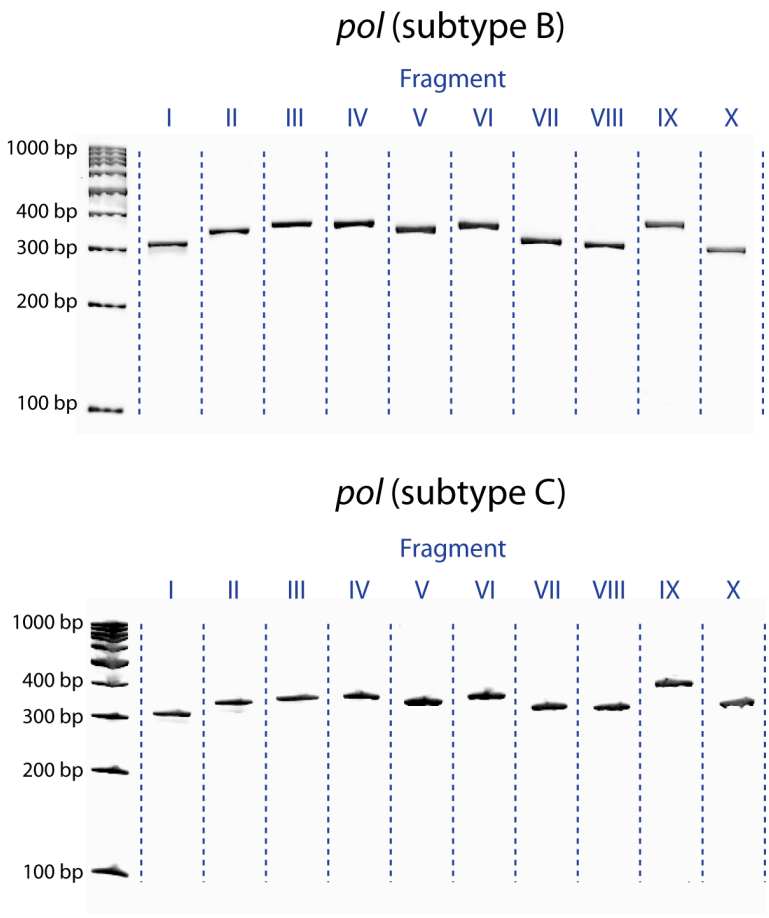

Figure S4. Mapping of the *pol* gene of both (top) subtype B and (bottom) subtype C. The 3,050 bp DNA fragment was divided up into ten fragments, each 300-400 bp in length which facilitated gel shift γPNA binding assays.

7. Nanopore analyses of HIV *pol* gene

Nanopore-based analysis for each of the *pol* fragments was conducted as described above. Because each gene is tagged with three separate γPNA probes, we define each of their associated secondary blockades by their individual blocked current levels ($\Delta I_{PNA,1}$, $\Delta I_{PNA,2}$, and $\Delta I_{PNA,3}$, respectively) and dwell times ($t_{PNA,1}$, $t_{PNA,2}$, and $t_{PNA,3}$, respectively). Additionally, the three distinct secondary blockade episodes per translocation event allow us to define two delay times per event: the delay between the first and second episode, $\delta t_{1\text{-}2}$, and the delay between the second and third episode, $\delta t_{2\text{-}3}$. Each delay time was measured in an identical manner to that previously described (see Figure S5a). Statistical analysis of ~750 events for each of the two subtypes gives the typical normalized blockade levels, $\Delta I_{PNA}/\Delta I_{DNA}$, of 0.425±0.002 and 0.419±0.003 for subtypes B and C, respectively. Additionally, we find typical PNA dwell times, t_{PNA}, of 0.63±0.05 ms and 0.64±0.04 ms for subtypes B and C, respectively (see Figure S5b and S5c for subtypes B and C, respectively).

Because a DNA molecule may enter the pore with either end inserted first, in order to explain the delay times between γPNA sites, it was necessary to compare $\delta t_{1\text{-}2}$ to $\delta t_{2\text{-}3}$ for every translocation event by relabeling the shorter of the two delay times as δt_{short} and the longer of the two as δt_{long}. In the case of the subtype B variant, where both inter-probe distances were nearly identical (~850 bp between Sites 1 and 2 and also between Sites 2 and 3) we find that both δt_{short} and δt_{long} have nearly identical values of 1.9±0.2 ms and 2.2±0.2 ms for the "short" and "long" delay times, respectively (see Figure S6a). However, when the same procedure is completed for the subtype C variant (where the spacings are ~450 and ~850 bp) we find typical values of 0.79±0.1 ms and 2.1±0.3 ms for the "short" and "long" delay times, respectively, indicating markedly different delay times between target sites (see Figure S6b). Given the power law dependence of δt on the distance between tag sites, we would expect a value for $\delta t_{long}/\delta t_{short}$ of $\left(850/450\right)^{1.39}$ or ~2.4. Our experimentally derived results return a strikingly similar ratio of 2.6±0.2. This further supports our findings that the normalized delay times between γPNA detection events serves

as an excellent metric in quantifying the spatial distance along the DNA, thereby minimizing slight variations in absolute time which can occur from pore to pore.

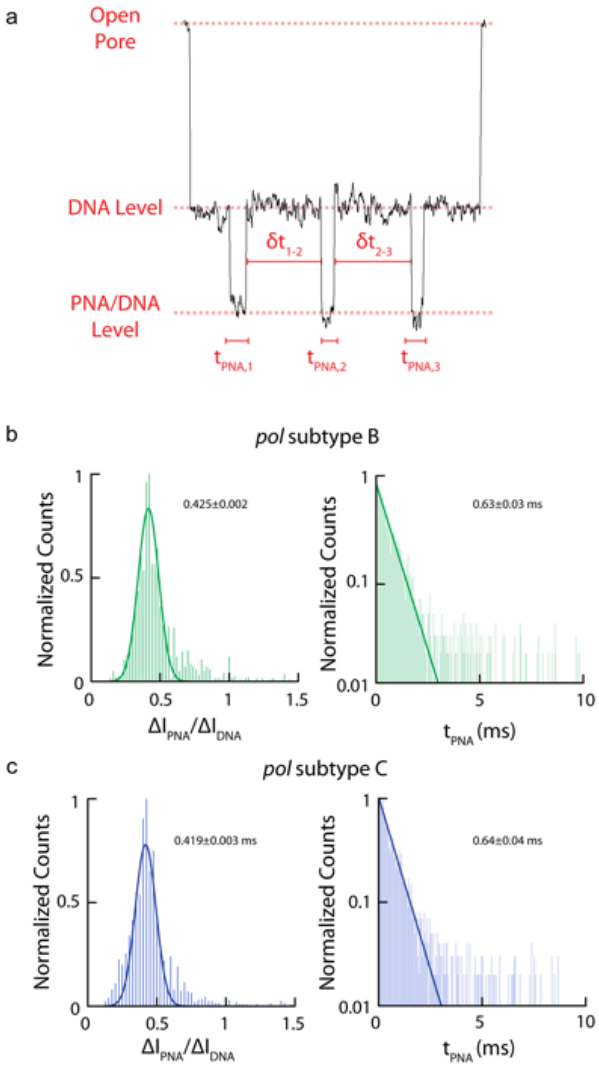

Figure S5 (a) Analysis of a DNA molecule with three γPNAs probes. In addition to the three secondary blockade episodes, we identify two separate time delays $\delta t_{1\text{-}2}$, and $\delta t_{2\text{-}3}$. To establish the typical γPNA induced signal for both subtypes, we look at the ratio of blocked levels, and the characteristic time spent at that level: (b) For subtype B we typically find that $^{\Delta I_{PNA}}/_{\Delta I_{DNA}} = 0.425 \pm 0.002$ with a t_{PNA} value of 0.63±0.03 ms. (c) For subtype C we typically find that $^{\Delta I_{PNA}}/_{\Delta I_{DNA}} = 0.419 \pm 0.003$ with a t_{PNA} value of 0.64±0.04 ms.

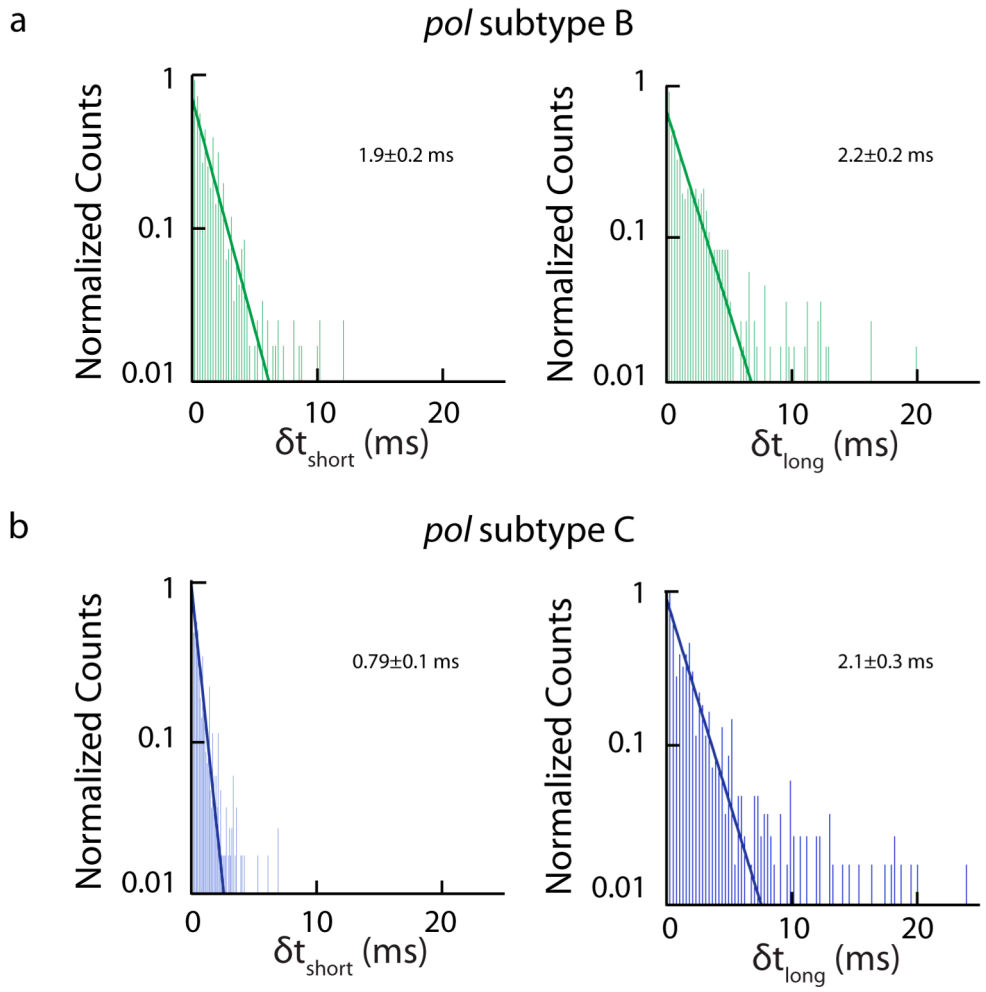

Figure S6 Comparing δt_{long} to δt_{short} (a) In subtype B, we find that the two delay-times are nearly identical, with values of 1.9±0.2 ms and 2.2±0.2 ms for δt_{short} and δt_{long}, respectively. (b) Subtype C, however, displayed markedly different timescales with values of 0.79±0.1 ms and 2.1±0.3 ms for δt_{short} and δt_{long}, respectively. In agreement with earlier results showing the delay time as a function of spacing, here the time-delay differences between the two subtypes may be attributed to their different tag spacings, demonstrating effective "barcoding" of the two subtypes.

References

1. Singer, A.; Wanunu, M.; Morrison, W.; Kuhn, H.; Frank-Kamenetskii, M.; Meller, A. *Nano Lett.* **2010,** 10, (2), 738-742.

2. Chenna, V.; Rapireddy, S.; Sahu, B.; Ausin, C.; Pedroso, E.; Ly, D. H. *Chembiochem* **2008,** 9, (15), 2388-2391.

3. Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. *J. Org. Chem.* **2011,** 76, (14), 5614-5627.

4. Sambrook, J.; Russell, D. W., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Press: NY, 2001.

5. Kuhn, H.; Bichismita, S.; Rapireddy, S.; Ly, D.; Frank-Kamenetskii, M. *Artificial DNA: PNA & XNA* **2010,** 1, (1), 45-53.