

Electronic Mapping of a Bacterial Genome with Dual Solid-State Nanopores and Active Single-Molecule Control

Arthur Rand, Philip Zimny, Roland Nagel, Chaitra Telang, Justin Mollison, Aaron Bruns, Emily Leff, Walter W. Reisner, and William B. Dunbar*



Cite This: *ACS Nano* 2022, 16, 5258–5273



Read Online

ACCESS |



Metrics & More



Article Recommendations

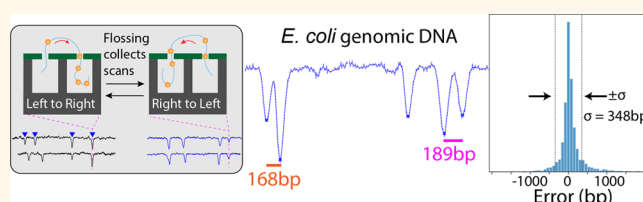


Supporting Information

ABSTRACT: We present an electronic mapping of a bacterial genome using solid-state nanopore technology. A dual-nanopore architecture and active control logic are used to produce single-molecule data that enables estimation of distances between physical tags installed at sequence motifs within double-stranded DNA. Previously developed “DNA flossing” control logic generates multiple scans of each captured DNA.

We extended this logic in two ways: first, to automate “zooming out” on each molecule to progressively increase the number of tags scanned during flossing, and second, to automate recapture of a molecule that exited flossing to enable interrogation of the same and/or different regions of the molecule. Custom analysis methods were developed to produce consensus alignments from each multiscan event. The combined multiscanning and multicapture method was applied to the challenge of mapping from a heterogeneous mixture of single-molecule fragments that make up the *Escherichia coli* (*E. coli*) chromosome. Coverage of 3.1× across 2355 resolvable sites of the *E. coli* genome was achieved after 5.6 h of recording time. The recapture method showed a 38% increase in the merged-event alignment length compared to single-scan alignments. The observed intertag resolution was 150 bp in engineered DNA molecules and 166 bp natively within fragments of *E. coli* DNA, with detection of 133 intersite intervals shorter than 200 bp in the *E. coli* reference map. We present results on estimating distances in repetitive regions of the *E. coli* genome. With an appropriately designed array, higher throughput implementations could enable human-sized genome and epigenome mapping applications.

KEYWORDS: nanopore sensing, nanotechnology, DNA barcode, dual-nanopore, genomics, bioinformatics



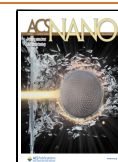
Precisely mapping the location of sequence motifs within individual double-stranded DNA (dsDNA) molecules in heterogeneous samples is central to a wide range of genomics applications.¹ One candidate approach for multiplexed molecular feature mapping measures ionic current modulations that arise when a dsDNA is electrically driven through a solid-state nanopore. Solid-state nanopores have sufficient sensitivity to detect a wide-range of features bound to translocating dsDNA and ssDNA, including anti-DNA antibodies,² streptavidin,^{3,4} transcription factors,⁵ histones,⁶ and peptide nucleic acids.^{7,8} They can also detect structural features involving nucleic acids such as ssDNA versus dsDNA regions,⁹ DNA-hairpins,^{10,11} multiway DNA junctions,¹² and aptamers.^{13,14} Compared to protein pores, solid-state pores can sense a wider range of analytes due to their configurable pore diameter, which can be tuned to optimize sensing of a particular bound DNA feature.¹⁵ While robust technology platforms already exist for high-throughput optical mapping based on detection of fluorescent labels,^{1,16} solid-state nanopores offer an electrical readout with potentially higher

resolution than the 1000 bp resolution limit of optics. Advances in super-resolution microscopy have pushed the resolution down below ~1000 bp resolving fluorophores separated by ~676 bp by averaging the distances between adjacent detected fluorescent tags.¹⁷ Time-averaging images is similar in spirit to the averaging technique we describe here; however, it requires specialized chemicals or else the number of frames is limited by photobleaching. The data presented by Jeffett *et al.* is from a bacterial artificial chromosome, and it is not clear how the method would perform on a genomic sample with chimeric molecules (which can account for 20% of the imaged molecules¹⁸). On the other hand, nanopore resolution

Received: October 28, 2021

Accepted: March 11, 2022

Published: March 18, 2022



is limited by the nanopore-spanning membrane thickness, which in principle can achieve <1 nm when implemented with atomically thin materials.¹⁹

Solid-state nanopore feature mapping along dsDNA has so far been demonstrated only in experiments that contain pools of identical molecules or simple mixtures of known sequence,^{3,10,11} including in our own work.^{20,21} Exploiting solid-state nanopores in medical or industrial genomics applications will require a nontrivial scaling of current solid-state nanopore sensing to analyze complex samples. Such samples consist of heterogeneous mixtures of ~10–100 kbp dsDNA fragments drawn from random locations on genomes larger than one million base pairs (Mbp) and up to human-genome scale (3200 Mbp). Scaling of the current construct-level experiments with solid-state nanopores to larger genomes will require an enabling methodology that is capable of aligning random molecular fragments to a reference genome or to other reads for constructing contigs for genome-wide assembly.

There are three fundamental obstacles to achieving high-quality genome-scale alignment of solid-state nanopore data. First, molecular folding during dsDNA translocation interrupts the linear ordering of molecular features, so a clear map cannot be established. For example, more than 60% of 48 kb λ DNA passes through a ~10 nm diameter pore in a folded configuration.²² While smaller pores that promote dsDNA linearization can be made *in situ* with additional circuitry and logic,²³ features bound to the dsDNA would not routinely pass through such pores. Second, high molecular fluctuations during translocation introduce significant random error which inhibits detection of features and alignment. Construct level barcoding experiments performed with solid-state nanopores suggest a broad distribution in translocation times;^{3,10} this is believed to arise from both fluctuations in initial molecule configuration and diffusion processes arising during translocation.²⁴ Third, genomic alignment requires converting barcodes from the translocation time domain data (microseconds to milliseconds) to units of genomic distance (bp). This conversion is nontrivial because it requires knowledge of the translocation speed. While the translocation speed can be obtained by assessing the translocation time between labels of known separation (e.g., using customized dsDNA calibration molecules with known label patterns²⁵), this is problematic when working with complex samples containing fragments of varying length and sequence motif label patterns. In particular, there is evidence from experiments,^{25,26} Brownian dynamics (BD) simulation and tension-propagation models²⁷ that the translocation velocity is nonuniform and can depend in a complex way on molecule size. Therefore, molecule basis is required to enable mapping distance estimation for *a priori* unknown barcode patterns, but single solid-state nanopore methods cannot meet this need.

We have demonstrated that dual-pore devices combined with active control logic implemented on a Field Programmable Gate Array (FPGA) can systematically address these challenges using ~20 nm diameter pores. Such pores are compatible with scalable lithography fabrication methods^{28,29} but are too large to prevent folding in single-pore configurations. A dual-nanopore device features not one but two pores.^{20,30–33} To realize greater functionality, dual-nanopore devices have been developed that permit independent biasing and current sensing at each pore.^{32,33} In such dual-pore devices, if the two pores are collocated within ~2 μ m or less, it is possible to achieve a dual-capture event where

different regions of a single dsDNA molecule simultaneously translocates through both pores.^{30–33} With the addition of active control logic, the two pores can exert opposing electrophoretic forces on the DNA, leading to a tug-of-war state²⁰ that achieves a controlled and reduced-speed translocation. In addition, exploiting independent sensing at the pores, the time-of-flight (TOF) of a molecular feature translocating between the pores can be accessed.²⁰ With the TOF and the pore-to-pore spacing can be combined to calculate the local translocation velocity, which in turn can be used to convert barcodes from translocation time domain data to units of genomic distance. The tug-of-war control was recently extended to permit active bidirectional control.²¹ Specifically, using FPGA logic to repeatedly change the direction of molecule motion during a tug-of-war event, we can induce a back-and-forth multiscanning or “flossing” of each cocaptured DNA. Multiscanning enables acquisition of multiple reads of the same molecule so that random errors can be minimized through aggregation, while also removing folds to linearize the molecule.

Here, we enhance our dual-pore multiscanning platform and develop accompanying bioinformatics algorithms to enable solid-state nanopore technology to analyze genomic DNA samples of realistic complexity. Our approach is applied to map and align genome-scale pools of heterogeneous DNA fragments drawn from the 4.6 Mbp genome of *Escherichia coli* (*E. coli*). To achieve this, we first use established techniques to decorate high molecular weight input dsDNA with molecular features formed from incorporating oligodeoxynucleotide overhangs³⁴ at sites established by a nicking endonuclease. These installed features, which we refer to as “tags”, give rise to a strong localized current blockade while minimizing pore interaction for facile tag detection during DNA translocation. Next, we introduce two key technology innovations: zoom-based multiscanning and automated recapture. With zoom-based multiscanning, we progressively increase the length of the region of the molecule interrogated during flossing, achieving high scan statistics over long genomic regions. With automated recapture, we repeatedly recapture a molecule after it exits the tug-of-war arrangement, further increasing statistics and length of the region of the molecule scanned (as distinct regions of the molecule are often interrogated upon recapture). These technology innovations lead directly to high quality single molecule scans of sufficient length and statistics to attempt genome-scale alignment, motivating development of bioinformatics algorithms customized for our dual nanopore system. The first computational processing step is to exploit local tag TOF from the dual-pores to calibrate our scans in units of physical distance in nanometers. Then we develop computational tools to align single molecule scans to a reference genome while optimally utilizing our high scan statistics and estimate the genomic distance, in base pairs, between tags. In particular, single-scans from a given multiscanning event are assembled into a consensus barcode that is aligned to a given genomic region. This approach, applied to *E. coli*, yields coverage of 3.1 \times across 2355 resolvable sites (68% of reference sites) after 5.6 h of recording time. These mapping results contain genomically significant information on structural variants, for example, enabling us to identify 93 regions of tandem repeated sequence.

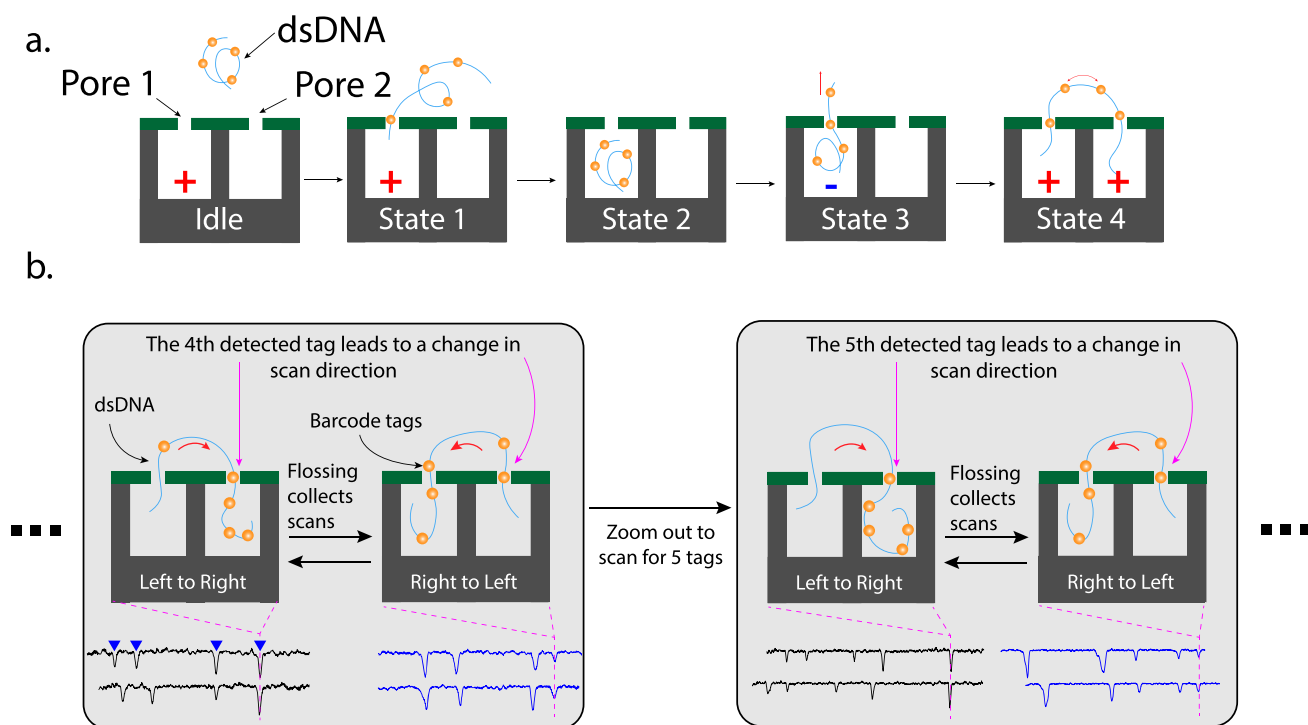


Figure 1. Cartoon descriptions of DNA flossing and “zoom-out” control with an example from recorded data. (a) Cartoon showing the process of capturing DNA for flossing. The dsDNA with barcode tags is depicted as a blue line with orange circles, respectively. Molecules are introduced to the “common chamber” above the nanopores (Idle). Positive voltage at pore 1 captures a single DNA strand in the channel below pore 1 (State 1 and State 2). At this point, a negative voltage at pore 1 threads a portion of the DNA into the common chamber to be captured by a positive voltage at pore 2 (State 3). Once DNA is detected in pore 2, the voltage at pore 1 is modulated as described in the text to move the DNA molecule repeatedly in the left to right (L–R) and right to left (R–L) direction (State 4). (b) Cartoon of logic counting tags in real time to “zoom out” on the DNA strand with individual scans shown below. Black scans are L–R and blue scans are R–L. In the example shown, the tag-counting limit starts at 4 (left shaded box) and a captured molecule is scanned L–R until the 4 tags are detected in pore 2 (blue triangles indicate detected tags on a single scan), and then R–L until the 4th tag is detected in pore 2, magenta line. Once a preset number of scans are collected, the controller zooms out by performing the same logic but waiting for the 5th tag before changing direction (right shaded box). The process is repeated with the tag-counting limit increased up to a user-defined limit (e.g., 8 tags) and/or until the molecule exits cocapture. The pore 2 signal traces are representative scans from a cocaptured fragment from *E. coli*.

RESULTS AND DISCUSSION

Experimental Setup. Dual-pore DNA capture and multiscanning experiments are performed using our previously described setup and devices.^{20,21,33} The double-stranded DNA (dsDNA) molecules were first labeled by conjugating 60 nucleotide long oligodeoxynucleotide (OdN) tags at the recognition sites established by nicking enzymes (see [Conjugation of Oligodeoxynucleotide Tags](#)). As the dsDNA molecules are electrophoretically driven through the nanopores, the OdN tags create sharp current blockade signals that can be distinguished from the baseline DNA blockade level during cocapture events. At the start of an experiment the DNA is then introduced in the top reservoir referred to as the “common chamber” of the borosilicate glass chip containing the nanopores and two opposing nanofluidic channels. A dual-pore flow cell uses 2 M LiCl running buffer dispersed symmetrically on both sides of the nanopores. An experiment requires 7 μL of 1 ng/ μL DNA sample where the DNA is longer than ~ 10 kb to ensure reliable cocapture and active control capabilities.²⁰ The nanopores are 20–30 nm in diameter and are formed *via* focused gallium ion beam milling in a ~ 30 nm thick silicon nitride membrane at the point of closest approach of the two channels; the nanopores thus form the fluidic gate between the nanofluidic channels and the common chamber.³³ The pores are placed ~ 500 nm apart so

that a single dsDNA molecule greater than a few kbp in size can span the interpore distance and simultaneously thread through both pores. Critically, our design permits independent control of voltage biasing as well as independent acquisition of ionic current signal at each pore. This allows simultaneous detection of the DNA ionic current blockade in each pore while separately adjusting the voltages applied across each pore. Active logic implemented on an FPGA adjusts the voltages in response to changes in the ionic current signal at either or both pores, leading to flexibility in control protocol design.^{20,21} The instrumentation and alignment algorithms explored here were first benchmarked using λ -DNA with OdN flaps placed at a superposition of sites for the enzymes Nt.BbvC1 and Nb.BssSI. As a benchmarking exercise compatible with the throughput of the current non-arrayed dual-pore device implementation, genome scaling of the technology was then demonstrated using DNA extracted from *E. coli* with OdN flaps placed at Nb.BsrDI sites (see [Isolation of Genomic *E. coli* DNA](#)).

Flossing with Zoom-Out Mode and Repeated DNA Captures. Tagged DNA samples are introduced to the “common chamber” above the two nanopores (Figure 1a, idle). The FPGA is configured to enhance the probability that a molecule transiting the dual-pores will be captured in a tug-of-war state (Figure 1a, State 4).²⁰ The control logic used to

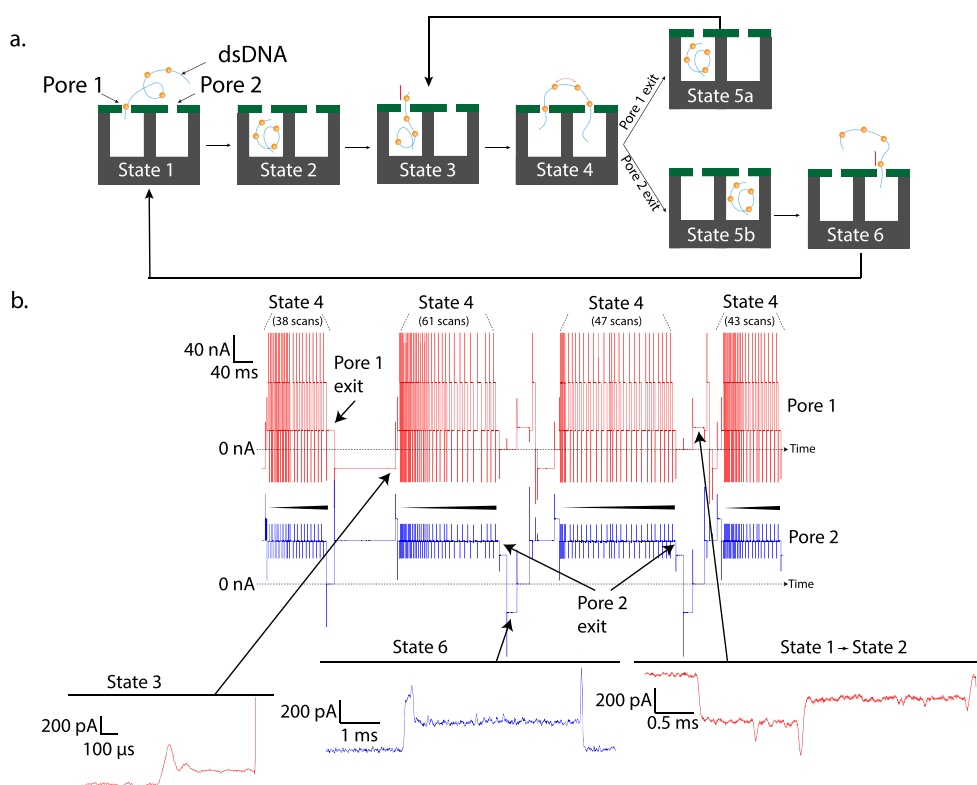


Figure 2. Cartoon depiction and recorded data showing the collective set of single-molecule data that DNA flopping with recaptures can produce. (a) The cartoon shows the recapture procedure, with state 4 comprising the flopping with zoom routine. When the DNA is lost to channel 1 (pore 1 exit, state 5a) the controller triggers tug-of-war and flopping without searching for another molecule from the bulk common chamber. When the molecule is lost to channel 2 (pore 2 exit, state 5b), the controller moves the DNA from channel 2 into the area directly above the two nanopores, priming it for recapture (state 1) and restoration of flopping. (b) Ionic current recording of 4 captures of the same molecule consisting of flopping (State 4) pore 1 exit and pore 2 exit. In pore 1 exit, the DNA is partially threaded through pore 1 with a negative driving voltage. In pore 2 exit and subsequent pore 1 translocation the DNA is completely driven through the respective nanopore.

achieve tug-of-war is the same as previously described (Figure S1, see [Achieving DNA Flopping](#)).²¹ Once captured in tug of war, the biasing at the pores is set to move the DNA in a controllable direction, *i.e.*, with DNA moving in the direction of the pore with the larger channel-side positive voltage bias. The FPGA is also used to dynamically change the voltage bias magnitude so that the molecules' direction is changed, with multiple sequential changes producing multiple back-and-forth interrogations of the same molecule in what is termed "DNA flopping" (Figure 1a and Figure 2, State 4).²¹ During DNA flopping, the voltage bias at one nanopore is instantly changed as a single step to trigger a change in direction, while holding the bias at the second pore constant. We will refer to the pore where the bias changes as pore 1 and the pore with constant voltage as pore 2. DNA motion from pore 1 to pore 2 is referred to as "left to right" motion (L–R); DNA motion in the opposite sense from pore 2 to pore 1 is referred to as "right to left" motion (R–L). In this study, flopping voltages at pore 1 are 150 mV for L–R motion and 650 mV for R–L motion and remain 300 mV at pore 2.

In our previous implementation,²¹ the control reversed direction after a preset number of tags were detected. This approach, while providing consistent performance, introduces a set of challenges when working with DNA fragments of unknown length and that contain an unknown number of tags. For such DNA fragments, if the preset number of tags is too

small, we will achieve a large number of scans but fail to scan other portions of the molecule while also spending too much measurement time on a small region of one molecule. On the other hand, if we set the preset too large then we risk having the molecule disengage from dual-pore capture during the first scan event, preventing the cyclic flopping function altogether.

In order to achieve a balance between maximizing the scan number and scanning a sufficiently large portion of the barcode, we have developed an adaptive strategy in which the controller will "zoom out", iteratively increasing the preset number of tags after a finite number of successful flopping cycles (Figure 1b).

For a newly recaptured molecule from the top chamber (Figure 1a, idle), the zoom-out controller starts by scanning for two tags to trigger changes in scan direction, as previously described and demonstrated.²¹ After a preset number of scans (nominally 4 scans L–R and 4 scans R–L), the tag count is increased by one and the multiscanning continues. A depiction of this zoom out process with representative data is shown in Figure 1b for the 4-tag-count and 5-tag-count stages, showing two representative scans for each scan direction. The zoom out process continues until the tag count reaches a maximum that is set by the user, nominally set to 8 or more tags depending on the anticipated tag density range. The molecule eventually exits capture and flopping for a variety of reasons, including those previously described²¹ such as drift in the motion of the

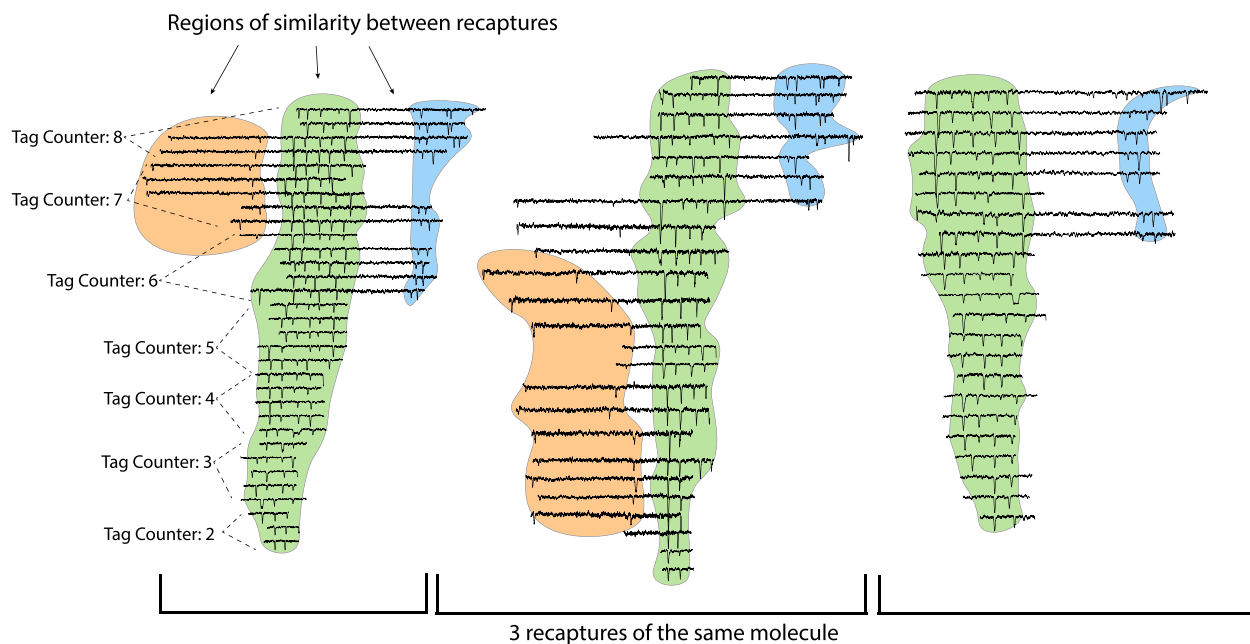


Figure 3. Representative traces from flossing with zoom on a single molecule (each vertical panel) comprising a total of three captures. The tag-counting limit used during zoom-out is reported for the left-most flossing event. The molecule was a fragment from *E. coli*, and signals are as measured at pore 2 in the L–R direction. The scans were produced chronologically from bottom to top in each panel. Drift in the molecule during zoom out can occur due to missing tags in the counting logic, which creates a frame shift in the region of the molecule being scanned.

molecule during flossing and tags that go undetected and thus uncounted (see [Tag Calling](#)). Molecules can also exit and recapture when the total tag density is lower than the maximum tag count.

In addition to the zoom out function, we further enhanced the logic to automatically recapture the molecule after exiting from the tug-of-war state ([Figure 2a](#)). Exit in either scan direction leads to the molecule being located in one of the nanofluidic channels, and the recapture logic is different depending on the direction of DNA motion when exit occurs. In the event of a R–L exit, the molecule is in the channel below pore 1 ([Figure 2a](#), State 5a), and the logic restarts the process of achieving a tug-of-war state using precisely the same logic sequence when a molecule is captured initially from the top chamber. In the event of a L–R exit, the molecule is in the channel below pore 2 ([Figure 2a](#), State 5b) and the logic implemented is modestly more complex. First, pore 1 voltage is set to 0 mV and the voltage is reversed at pore 2 to drive the DNA from the channel below pore 2 back through pore 2 into the common chamber. Note that recapture into a pore from the channel has a high probability due to the influence of voltage along the length of the nanofluidic channel.²⁰ Next, after the DNA is fully ejected through pore 2 and into the common chamber, pore 2 voltage is set to 0 mV and the pore 1 voltage is turned back on to recapture the DNA through pore 1 and to return to State 1 in [Figure 2a](#). The time scales of time-of-flight recapture (pore 2 to pore 1) are fast when recapture occurs³³ and are thus time bounded in the logic by a waiting period of 1 s to maximize the probability of capturing the same molecule in pore 1 that exited pore 2, rather than a new molecule from the common chamber. In contrast, the time to capture a new molecule in pore 1 from the common chamber occurs on a time scale that is comparatively slower than the

pore 2-to-pore 1 recapture process, *i.e.*, at the 1 ng/ μ L concentrations used. For the merged *E. coli* experiment data, for example, the exponentially distributed time-to-capture of new molecules from the common chamber had a mean 10.6 s and 10th percentile of 1.1 s, all of which are slower than the maximum wait time of 1 s for recapture, while the time-to-recapture at pore 1 after exit from pore 2 had a mean of 0.2 s, median 0.1 s, and 99th percentile of 0.6 s. [Figure 2b](#) shows a full dual-pore recording of DNA flossing including voltage switches and recapture strategies of four recaptures of the same molecule. In this example with one pore 1 exit and 2 pore 2 exits, as the zooming logic progresses the individual scans increase in time until the molecule is lost.

[Figure 3](#) gives an example of a subset of scans from three recaptures of the same molecule. Each flossing event was recorded from bottom to top, with the first flossing event annotated (left side) with the increasing tag-counting limit that was used during zoom-out control ([Figure S4](#)). Recapture events commonly show similarity in the barcode pattern, as observed for the shaded regions of the signal traces in [Figure 3](#). Also observable are the stochastic variations in the signal in the form of variable tag amplitudes and variable intertag durations. Variations in tag amplitude can lead to imperfect detection and counting of tags during flossing, which in turn contributes to drift in the DNA motion during flossing and thus drift in the scanned region of the molecule over sequential scans. The variation in intertag durations is visibly larger where the tags are spaced farther apart, *e.g.*, for the time gap between the right most tag in the green shading and the left most tag in the blue shading in [Figure 3](#). Variable intertag durations can contribute to larger spread in predicted distances. However, distance estimation variation between *a priori* unknown tag locations is ameliorated here by using direct measurement of molecule

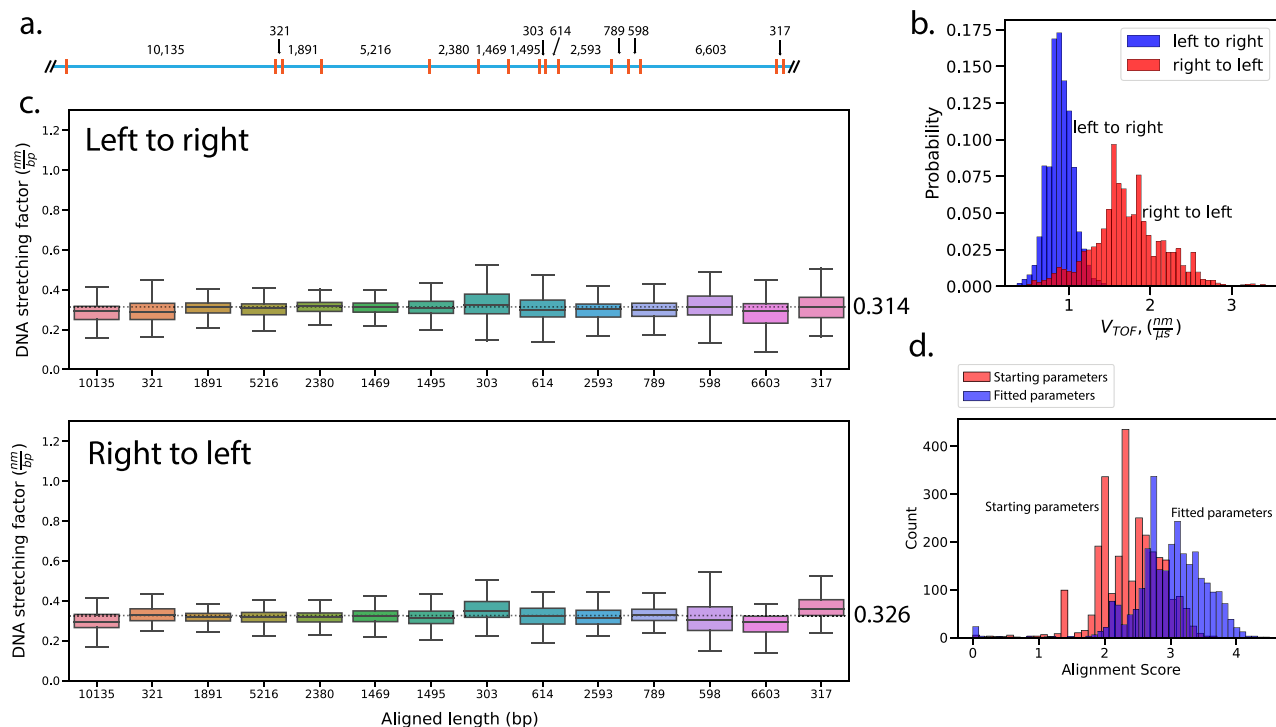


Figure 4. Performance of alignment of individual scans from flossing data using a model 15-tag λ -DNA. (a) Visual description of the 15-tag phage λ -DNA test molecule. Orange hashes indicate nicking sites from the Nt.BbvC1 and Nb.BssSI enzymes. (b) Histogram of the measured V_{TOF} for L–R (blue) and R–L (red) scanning directions, with data drawn from 7711 tags within 1724 scans L–R and 3716 tags within 1074 scans R–L. (c) DNA stretching factor (nm/bp) for the different intervals on the λ -DNA test molecule for L–R (top) and R–L (bottom) for the same scans used in (b). We observe small variations around the weighted average, 0.314 ± 0.05 nm/bp and 0.326 ± 0.05 nm/bp for L–R and R–L scans, respectively (black dotted line, annotated to the right). (d) Histogram of single-scan alignment scores for the initialized and fitted model parameter. The scores improved by using the direction-specific models and the optimized probability distributions.

velocity. Specifically, a differentiating advantage of the dual-pore system is that the velocity of individual tags can be directly measured on a per-scan basis,²¹ which enables intertag distance estimation for each scan as described in the next section. Our direct velocity measurement requires no calibration and can be done directly for any tag detected at both pores. This advantage means that while regions can have variable scan speeds that create variable intertag times, direct knowledge of velocity can be exploited to produce more consistent distance estimates, as suggested in related studies.³⁵

Alignment of Single Scans to a Reference Map. A scan is a single pass of a region of a dsDNA molecule moving through the dual-pore sensors in either the L–R or the R–L direction. For a scan to be aligned to a reference map, it must first have electronically detectable tags. A reference map is a list of genomic positions generated by digesting the reference genome *in silico* by one or more nicking endonucleases (Figure 4a). Our objective is to align the sequence of tags observed in the scan to the positions in the reference map by matching the distances between observed tags to the known number of base pairs between nicking sites. This requires estimating the distances between observed tags in genomic coordinates. To estimate distance in genomic coordinates requires two conversions: (1) converting the time period observed between any two tags into a distance estimate in spatial linear coordinates (units of nanometers) and (2) converting the spatial distance estimate into a distance estimate in genomic coordinates (units of base pairs). The first conversion from intertag time to intertag linear distance utilizes the known linear distance between the dual nanopores (see Estimating

Scan Velocity and Linear Distance between Adjacent Tags).

The second conversion requires consideration of the stretching behavior that occurs between the pores as a result of the tug-of-war forces, which is expected to be asymmetric in the two scan direction since higher tug-of-war voltages are used R–L than L–R. The details of these two sequential conversions are described next.

To estimate the spatial linear distance $D_{n,m}$ between the n th and m th tags detected in a scan, we first identify the positions of the n th and m th tags in the time-domain, which is accomplished by identifying the times associated with the peak current attenuations for the pair. To convert from the time-domain to units of linear distance, we measure the velocity of tags during flossing and create a velocity profile that linearly interpolates between these values. Briefly, when a tag is detected at each pore, the time required for the tag to traverse the interpore distance is obtained and is referred to by the tag time-of-flight (TOF). The interpore distance divided by the tag TOF yields the tags time-of-flight velocity V_{TOF} , and a velocity profile of the DNA chain itself during the scan is generated by linearly interpolating between sequential V_{TOF} values (see Estimating Scan Velocity and Linear Distance between Adjacent Tags). The linear distance $D_{n,m}$ between any two sequentially detected tags n and m is then computed by simply integrating the scan velocity profile over the intertag time period.

The average of each scan's velocity profile can be examined to assess the DNA chain speed distribution across scans and in each scan direction. In our experiments on λ -DNA, we estimated the mean L–R velocity to be 0.89 nm/ μ s, which is

$\sim 2\times$ slower than the mean R–L velocity of 1.73 nm/ μ s (Figure 4b). Larger voltages create larger DNA velocities through a nanopore, and in the dual-pore setup the net voltage difference defines the net force on the molecule. Therefore, the $2\times$ higher observed velocity going R–L is likely a consequence of the $2\times$ larger net force: 300 mV R–L (600–300 mV) vs 150 mV L–R (300–150 mV). A detailed analysis of how tag peak widths change as a function of the speed differences in the two directions is explored in another article in preparation.⁴⁹

Once linear distance estimates between tags are computed for a given scan, the second conversion requires converting the spatial distance estimates (nm) into a genomic distance estimates (bp). The core of our alignment method is a probability distribution modeling the stretching of DNA under forces exerted by the dual-pore system. Let μ be the DNA stretching factor (nm/bp), and let G_{ij} be a genomic interval distance (bp) found in the reference and corresponding to nicking site indices i and j . With a given constant stretching factor μ presumed for a given scan, the alignment problem is to find the optimal placement of the scan data $\{D_{n,m}\}$ to the reference data $\{G_{ij}\}$ such that $\mu \approx \frac{D_{n,m}}{G_{ij}}$ and while searching across the entire reference and accommodating for false-negative and false-positive tags. To achieve this, we first model the probability of observing a given DNA stretching factor by a normal distribution $P\left(\frac{D_{n,m}}{G_{ij}}\right) \sim \mathcal{N}(\bar{\mu}, \sigma)$ with mean $\bar{\mu}$ and standard deviation σ . The probability density is

$$P\left(\frac{D_{n,m}}{G_{ij}}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\bar{\mu} - \frac{D_{n,m}}{G_{ij}})^2}{2\sigma^2}\right\} \quad (1)$$

and we convert this into a score model by taking the log of the probability density. The score model allows us to quantify how well a measured set of distances between tags matches a corresponding set of genomic intervals between nicking sites on the reference. Previous studies have shown B-form DNA to have a stacking height of approximately 0.34 nm/bp.^{36,37} In order to make an initial estimate of $\bar{\mu}$ and σ we initialized our model with $\bar{\mu}_0 = 0.34$ nm/bp and $\sigma_0 = 0.12$ (see [Scan-Space Alignment](#)). This score model is then built into an alignment algorithm similar to Smith–Waterman local alignment³⁸ that accounts for false-negative and false-positive tags by the tag calling algorithm (see [Description of Alignment Algorithm](#)).

The alignment method was applied to experimental data and a reference map based on λ -DNA with incorporated tags at 15 nicking sites (see [Conjugation of Oligodeoxynucleotide Tags](#)). This relatively simple reference map permitted alignments to be manually inspected for correctness. Scans collected from 27 dual-pore chips were filtered (see [Data Processing and Filtering](#)), resulting in 2897 individual scans within 889 cocapture events composed of 409 individual molecules. The distribution of fitted stretch factors was obtained separately for L–R and R–L directions and stayed consistently near their averages of $\bar{\mu} = 0.314$ nm/bp for L–R direction and $\bar{\mu} = 0.326$ nm/bp for R–L direction across the length and varying intertag distances of the λ -DNA molecule (Figure 4c). These values indicate that the DNA, while strongly extended, is not 100% stretched as a semiflexible chain at the tug-of-war voltage values used.²⁰ The modestly higher stretching coefficient for the R–L direction is presumed due to the higher tensile stretching forces than L–R.

The stretching factors showed high variability on a per site basis (Figure 4c), as indicated by high standard deviations: $\sigma = 0.25$ nm/bp for L–R and $\sigma = 0.34$ nm/bp for R–L. To obtain an improved estimate, we performed a weighted average where the weights are the probability of the aligned segment using the initial values of $\bar{\mu}_0$ and σ_0 . The probability-weighted average suppressed the influence of low probability outlying pair measurements and improved the per-site estimate variability. The weighted estimates yielded $\bar{\mu} \pm \sigma = 0.314 \pm 0.055$ nm/bp for L–R direction and $\bar{\mu} \pm \sigma = 0.326 \pm 0.051$ nm/bp for R–L direction (Table 1). Upon iterative realignment of the λ -DNA

Table 1. Parameters in Model Equation 1 after Weighted Averaging

	no. of scans	$\bar{\mu}$	σ
left to right	1724	0.314	0.055
right to left	1074	0.326	0.051

results using these refined values of $\bar{\mu}$ and σ for L–R and R–L directions, we observed an increase in alignment score of 6.1% (Figure 4d).

Tag Resolution Is at Least 150 Base Pairs. One advantage of electronic nanopore-based measurement over optical detection is the potential to resolve smaller distances.³ To test the resolution of our dual-pore instrument, we engineered specific λ -DNA reagents possessing closely spaced tags. Cas9 nickase was used to install 60 nucleotide ssDNA tags separated by 150 bp (see [Construction of 150 Base Pair Tag-Pair DNA](#)). We observed 23 molecules with two distinct spikes due to the two tags (Figure 5). We calculated the tag-pair resolution T_{res} using a formula from liquid chromatography

$$T_{\text{res}} = \frac{2(t_2 - t_1)}{1.7(w_2 + w_1)} \quad (2)$$

where the first and second tags have peak minima at times t_1 and t_2 and peak time widths at half-minimum depth of w_1 and w_2 , respectively. We also measured the percentage of the restoration of the signal between the two peaks. This “restoration percentage” is the percentage that the signal achieves from the minima current attenuation up to the tag-free baseline signal. Statistics from the tag-pair resolution values from eq 2 and the corresponding restoration percentage values are reported in Table 2. While the signals for tag pairs are resolvable at 150 bp here, optical mapping has a 50% chance of resolving two fluorescent reporters within 1000 bp.³⁹ Additionally, future instrumentation at $2\times$ higher bandwidth than the 10 kHz low-pass Bessel filter used here should provide $2\times$ sharper peak resolution with an acceptable increase in high frequency noise, which will further increase the resolution score and permit exploring the detection of tag pairs closer than 150 bp.

Generating Consensus Alignments from Multiscan Flossing Events. A flossing event produces a significant quantity of information distributed over multiple scans. This information needs to be aggregated into a best estimate of the overall pattern of tags present on the analyzed fragment. To this end, we developed an algorithm that takes as input a set of single scan alignments to the reference map and assembles these alignments into an overall consensus alignment by using the highest scoring portions of each individual alignment. The overall goal of this reference-guided algorithm is 2-fold: (1)

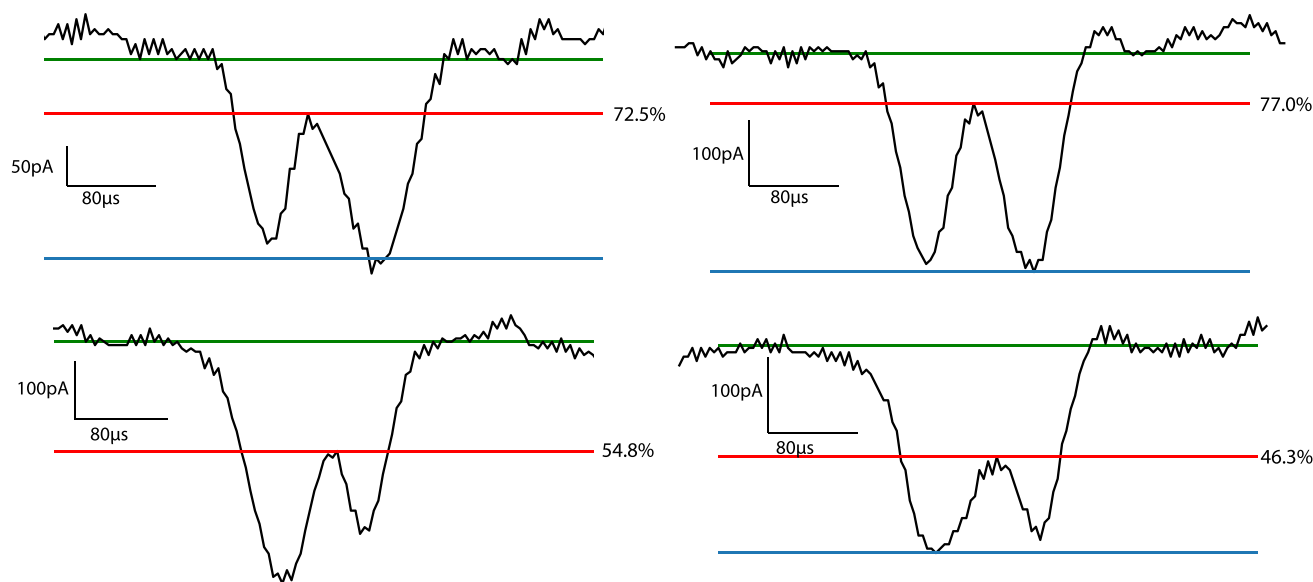


Figure 5. Representative current traces of 60 nucleotide tags separated by 150 bp. Green, blue, and red lines show the estimated tag-free DNA baseline current, the minima current attenuation created by one of the two tags, and the restoration percentage line (with value reported), respectively.

Table 2. Tag-Pair Resolution and Restoration Percent for 23 Molecules

	restoration percentage	tag-pair resolution
mean	58.59	0.77
std	10.74	0.23
min	34.34	0.48
max	77.06	1.17

remove erroneous portions of individual alignments and (2) increase the accuracy of distance predictions by utilizing information from multiple scans.

Our algorithm first aligns the individual scans to the reference map using our alignment algorithm with a tuned Gaussian scoring function. To assemble the alignments into a consensus we consider the problem in a graph theoretic sense. Let \mathcal{G} be a graph with vertices $v_i \in \mathcal{V}$ and edges $e_{ij} \in \mathcal{E}$ connecting vertices v_i and v_j . Each vertex, v_i , represents a nicking site at position i in the reference map (in bp). Each alignment corresponds to a sequence of tuples $A = [(a_0, s_0, d_0), (a_1, s_1, d_1), \dots, (a_k, s_k, d_k)]$. The quantity a_k represents an aligned pair $a_k = (n, m) \diamond (i, j)$, the notation indicating that the n th tag in the scan is aligned to the nicking site at position i in the reference and likewise that the m th tag in the scan is aligned to position j in the reference. In addition, each aligned pair a_k has an associated score s_k and measured distance d_k . The score is calculated in eq 1 and uses the parameters $\bar{\mu} \pm \sigma = 0.314 \pm 0.055$ for L–R data and $\bar{\mu} \pm \sigma = 0.326 \pm 0.051$ for R–L data (Table 1). We start by initializing $\mathcal{E} = \emptyset$ (i.e., no alignments in the set) and proceed with graph construction by iteratively adding each alignment to the graph. For each aligned pair a_k , we add an edge connecting vertex v_i and v_j with the edge weight equal to s_k . If an edge already is present from a previous added alignment, we simply add s_k to the weight corresponding to that edge. We also maintain a mapping, $e_{ij} \mapsto$

D_{ij} , of the measured distances of the aligned pairs for each edge in the graph. Following graph construction, the graph is pruned to remove edges with scores below zero. The consensus alignment, \mathcal{A} , is reported as the maximum scoring path from a start vertex to an end vertex. Lastly, the measured distance for a given interval in the consensus alignment is the average of the measured distances for the edge connecting the interval \bar{D}_{ij} .

To convey the process of generating consensus alignments from multiscan flossing events conceptually, a synthetic reference and examples of synthetic scans are shown in Figure 6a. The concept shows the need for at least two scans to detect conflicts and at least three scans to resolve conflicts (the resulting consensus chooses to align to site 3 instead of site 4). The concept also shows that the resulting consensus can have missed tags (site 3) and skips erroneous alignments to distal regions (sites 8–10). The results of applying the process to one representative multiscan flossing data set from λ -DNA in which all 15 tags were detected is shown in Figure 6b. After removal of 1% of outliers, the performance of the consensus estimates from all λ -DNA flossing data is shown in an error histogram in Figure 6c.

For the λ -DNA results, the consensus alignment error distribution appears normally distributed with 30 bp mean and 349 bp standard deviation. A mean error near zero implies that the dual-pore does not systematically under or overestimate the distance between tags. A more meaningful measure of the inaccuracy of the system is the absolute error, which we calculated to have a mean of 192 bp and median of 84 bp. An approximation for the 95th percentile is the mean absolute error plus two standard deviations of the error, which is 890 bp, suggesting most of the consensus errors are below the resolution limit of optical mapping (1 kb).³⁹ Potential sources of random error are DNA stretching and temporary arrest of DNA motion. We filtered the raw DNA flossing scans using heuristics described in Data Processing and Filtering, and the

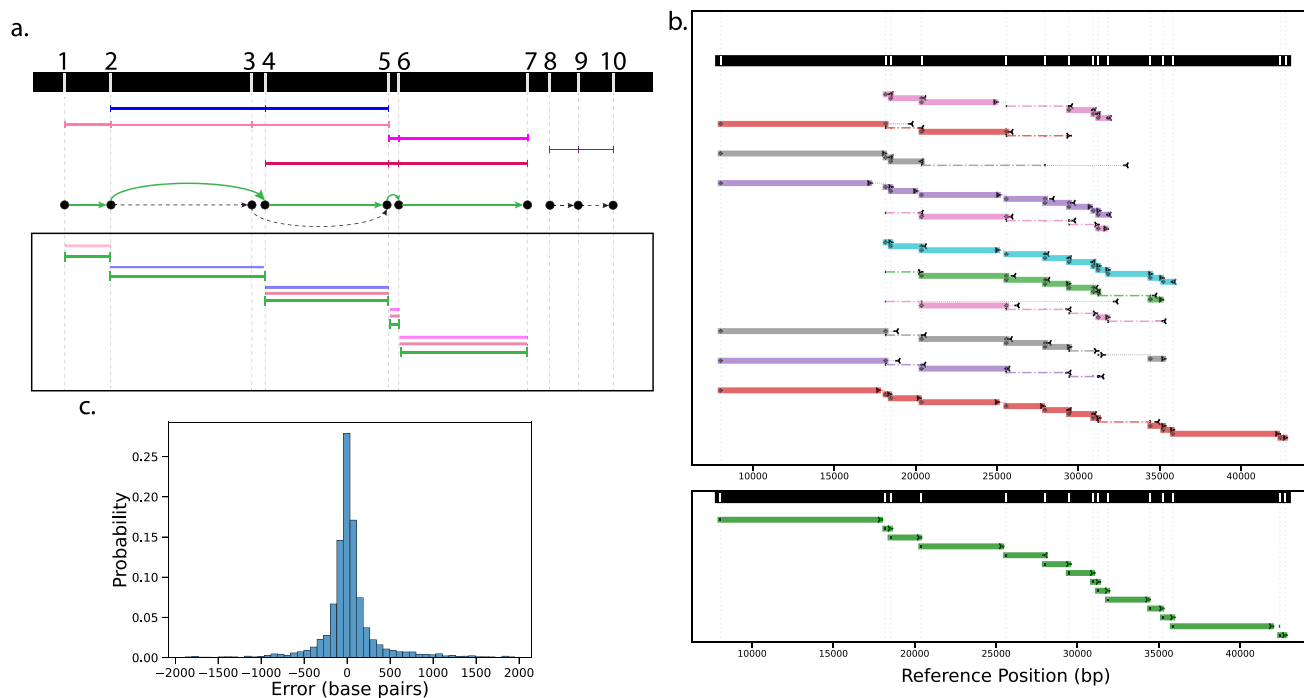


Figure 6. Concept and recorded data show how consensus alignments are produced from flossing event data. (a) A fictional reference map with 10 tagged sites and five conceptual aligned scans with line thickness proportional to the relative fitting score. The four scans spanning sites 1–7 (colors: blue, pink, magenta, and red) are used to identify the highest scoring directed graph path (1, 2, 4, 5, 6, 7) (solid green lines with arrows). The scan spanning sites 8–10 (purple) is low scoring and not connected to the rest of the graph and is thus ignored. An alternative path (1, 2, 3, 5, 6, 7) (inserted dashed lines) would require ignoring the more probable and scan-supported 4–5 path transition and thus produce a lower score. The consensus alignment (box, green lines) shows the supporting scan intervals by color that are averaged to generate the consensus distance estimates. (b) Interval plots of individual scan alignments (upper box) and resulting consensus alignment (lower box) from a single recorded λ -DNA flossing event. Each aligned pair of tags are shown as colored bars if sufficiently high scoring and thus used in the consensus or as dashdotted lines if low scoring and not used in the consensus. Bars and lines of the same color are from the same scan. The bar or line length indicates the estimated number of base pairs for that interval with a vertical offset of adjacent intervals used to help visualization. Dotted lines are inserted where length estimates appreciably deviate from their assigned reference values, with arrows to demark the end point of estimates and the direction of the assigned reference end point. The consensus (bottom) produced estimates for all 15 sites from the first to the last tag. (c) Consensus intertag alignment length error histogram (trimming 1% outliers with absolute error >2 kb), comprising 3268 consensus tag-pair distance predictions across 607 captured λ -DNA molecules, resulting in 30 bp mean and 349 bp standard deviation for the distribution plotted and with 192 bp mean and 84 bp median absolute error.

methods described therein would not remove scans exhibiting these pathologies. Filtering data based on alignment scores or calculating weighted estimates would likely improve these measures at the expense of removing data.

The accuracy improvement when using consensus estimates is apparent by considering the relative performance to that of the single longest scan from each flossing event, as a proxy for single read data. We removed 1% of outliers, resulting in 3156 estimates from the set of longest scans that produced an error distribution with mean 153 bp and 1060 bp standard deviation and 425 bp mean and 112bp median for the absolute error. Again using the mean absolute error plus 2 standard deviations of the error results in a 95th percentile error estimate of 2500 bp, nearly three times higher than the multiscan consensus accuracy. Thus, the consensus multiscan map estimates improve accuracy and reduce the variance of distance estimates compared to single-scan nanopore data. It is possible that multiple tags are incorporated in close proximity (<50 bp) due to the polymerase adding multiple dUTP-azide residues during nick translation (see [Conjugation of Oligodeoxynucleotide Tags](#)). The resulting peak in the ionic current would be the convolution of the influence of the two closely spaced tags and could cause a distance prediction error.

The process of generating consensus alignments using a multiscan flossing molecule from *E. coli* is shown in [Figure 7](#). These same data also annotated in [Figure 3](#). The molecule had a total of four captures with scans organized into tracks (i–iv) in [Figure 7a](#). For each capture, the initial scans with fewer tags align to multiple sites in the reference genome, as shown in the top lines for each track. After zooming out, subsequent scans have more tags and thus align less ambiguously to the reference, converging in their support to a locus near the midpoint of the reference genome ([Figure 7b,c](#)).

Physical Genome Mapping with *E. coli*. We collected 5.6 h of dual-pore recording data from genomic *E. coli* DNA across three devices ([Table 3](#)). This data includes 979 single-molecule flossing events composed of 564 individual molecules. Based on our results indicating a resolution of at least 150 bp ([Tag Resolution Is at Least 150 Base Pairs](#)), we modified the Nb.BsrDI reference map by merging sites that were within 150 bp of each other. We generated consensus alignments as described in [Generating Consensus Alignments from Multiscan Flossing Events](#), requiring that the consensus alignment contain at least three aligned intervals, resulting in 767 consensus alignments. The data contained 247 consensus alignments with more than one scan support. The consensus alignments resolved 2,355 (68%) of the sites in the reference

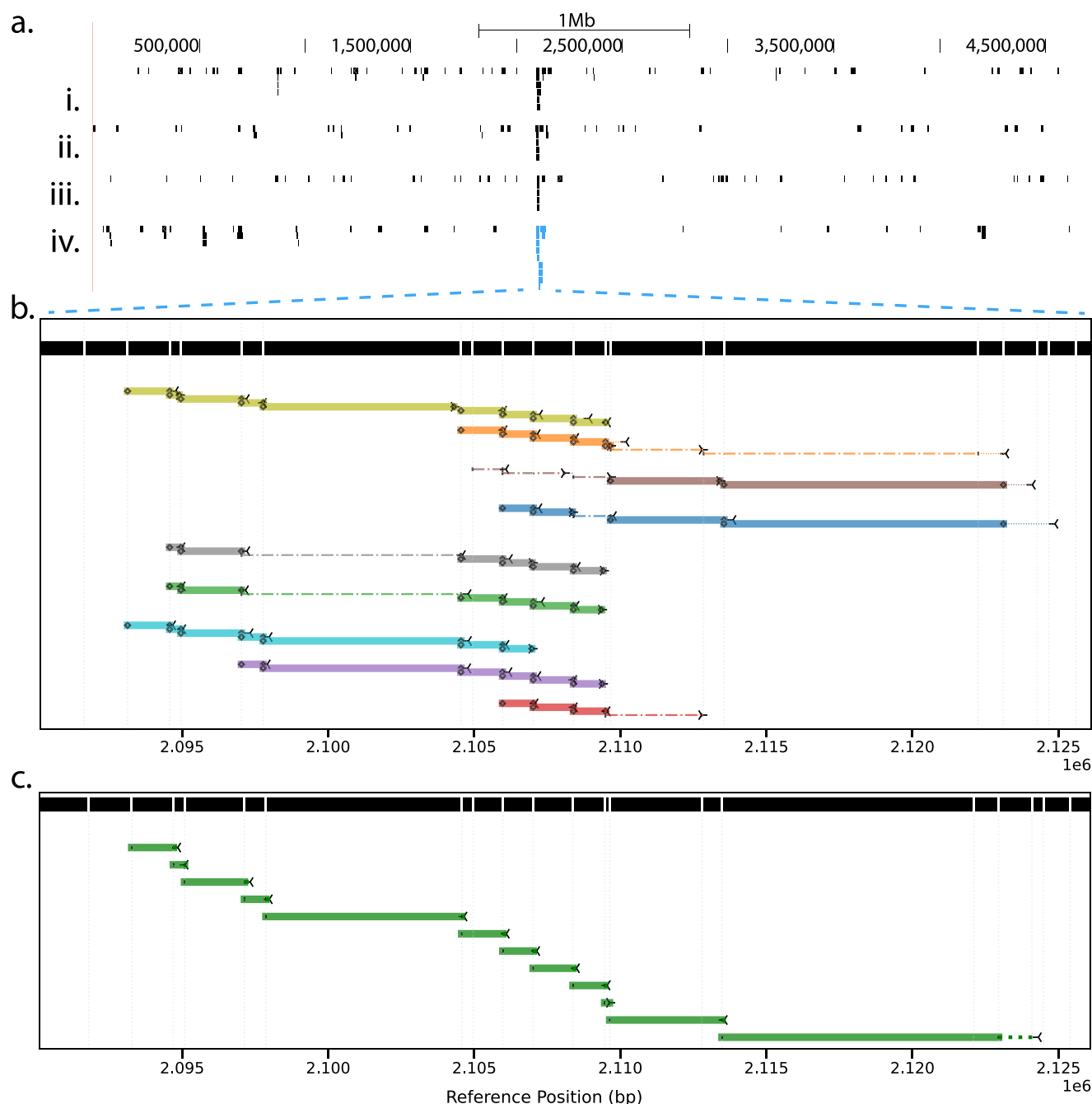


Figure 7. Representative consensus single-molecule alignment produced from an *E. coli* experiment. (a) Data are from four DNA flossing captures from the same molecule. Each bar is a single scan, and each track (i–iv) is a recapture of the same molecule. Scan alignments for the four events (i–iv) with fewer tags are spread across the *E. coli* reference genome with multiple prospective alignments, while scans with the highest tag counts achieve alignment with strong support at a unique common locus near the midpoint of the genome. (b) Interval plots of individual scan alignments for the highlighted multiscan data within recapture (iv). Annotations of scan details are consistent with the description in Figure 6b caption. (c) The consensus produced comprised a total of 13 sites across the 2,075,645 to 2,178,824 locus.

Table 3. Throughput for Three Dual-Pore Devices Used to Generate *E. coli* Data

data set no.	scan count	minutes recording	flossing events	molecule count
1	590	73	208	143
2	474	65	136	89
3	18,181	196	711	378

map including 133 intersite intervals of less than 200 bp, with an average coverage of 3.1 \times over all sites (Figure 8a).

To quantify alignable molecule lengths, we use the N50 (see Calculating N50 of alignments), noting that the length values are defined as the first-to-last tag distances and do not include the length of the molecules outside of these tags. The N50 of the consensus alignments are longer (17.4 kb) than the N50 of the longest single scan (15.8 kb) from the corresponding flossing event since consensus alignments use the best scoring

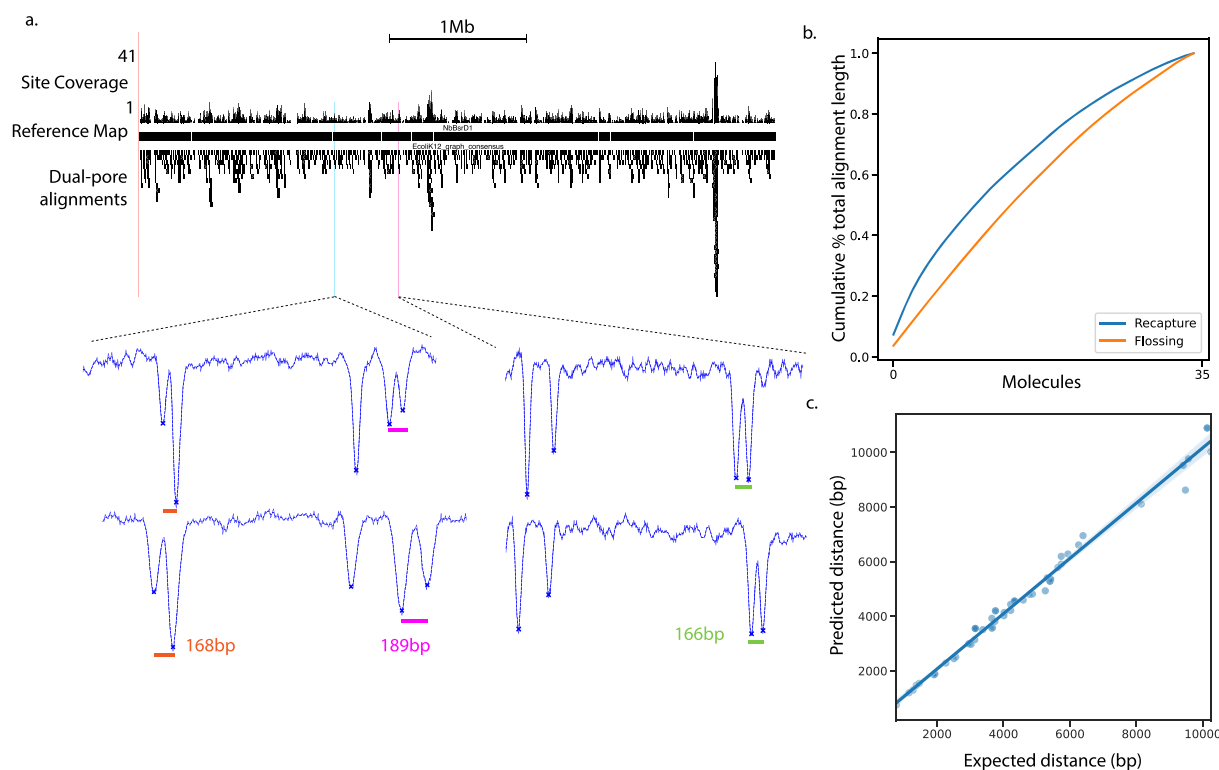


Figure 8. Dual-nanopore physical genome mapping of *E. coli*. (a) (Top) Microbes genome browser⁴¹ of the *E. coli* K12 genome and Nb.BsrDI genome map. The per site coverage, reference map sites, and consensus dual-pore alignments are shown on the top, middle, and bottom tracks, respectively. (Bottom) Representative sequential scans with tags that align to distances below 200 bp in the reference genome and bars showing the aligned distance values. (b) Cumulative alignment length as a percentage of the total length of all alignments for recaptured molecules vs single capture molecules (formula in [Joining Recapture Alignments](#)). (c) Correlation of predicted and expected base pair distances for intertag regions spanning tandem repeats ($R^2 = 0.97$).

regions across all scans including the longest. We can also boost alignment lengths further by joining together, where possible, the consensus alignments resulting from multiple recaptures of the same molecule. Our data set contained 35 molecules that were recaptured at least once after escape from initial capture. Joining recaptured molecules into a single consensus alignment increased the N50 to 21.9 kb, a 38% increase compared to the longest single-scan alignments ([Figure 8b](#); [Joining Recapture Alignments](#)). Length scales of the consensus molecules including the N50 values are reported for *E. coli* and λ -DNA in [Table 4](#). We also quantified the

Table 4. Single-Molecule Consensus Lengths Using Dual-Nanopore DNA Flossing

experiment	molecule count	max length ^a (bp)	median length ^a (bp)	N50 ^a (bp)
λ -DNA	607	34,724	11,477	14,859
<i>E. coli</i> ^b	35	50,957	17,531	21,856

^aLengths are defined by the first-to-last tag distances and do not include the length of the molecules outside of these tags. ^bResults are after merging recaptured data sets.

tagging efficiency as the number of observed aligned tags divided by the number of expected tag sites spanning the net consensus alignment length, resulting in 88% efficiency for the 15-site λ -DNA and 69% efficiency for *E. coli*.

One potential application of physical genome mapping is to detect structural variations, including those in repetitive regions which are often difficult to resolve with short-read

sequencing. We identified 93 regions of tandem repeated sequence using Tandem Repeats Finder⁴⁰ ([Supporting File 1](#)), to use as test regions for the accuracy of our method. We used alignments spanning the repeat region and converted their measured distance (in nm) to bp by dividing by $\bar{\mu}$ for the score model (see [Alignment of Single Scans to a Reference Map](#)). We then averaged their estimated genomic distances. Filtering to sites with at least two aligned molecules, our base pair length predictions showed good correlation with the expected values, $R^2 = 0.97$ ([Figure 8c](#)).

CONCLUSIONS

Our results demonstrate mapping of a heterogeneous mixture of DNA molecules, varying in size and label density, onto a mega-base scale genome using solid-state nanopore technology. We enhanced our previous dual-pore DNA flossing approach by adding the capabilities of zooming out and repeating captures, which when combined with the presented analytical framework increases both the quality and size of consensus single-molecule alignments. We also presented proof-of-concept results estimating genomic distance in repetitive regions showing good correlation with low coverage (minimum of two molecules) and setting the stage for structural variation analysis with the proposed method. Our method uses relatively small amounts of input genomic DNA (7 ng/flow cell) compared to other single molecule approaches. Moreover, the nanopore technology's purely electrical basis confers a small footprint and cost relative to optical mapping approaches. Thus, our approach has potential

to produce high resolution physical genome maps with low sample input and low cost for potential field and clinical research use.

Key remaining technical challenges include data throughput and device fabrication. In particular, only 41% of the *E. coli* single-molecule data generated by the dual-pore passes our aggressive quality filters for consensus generation and alignment. Pathologies for nonpassing include too few tags or too sparse a tag pattern for unambiguous alignment. We expect that improvements in DNA sample preparation with increasing tag density will lead to large increases in this metric. Data throughput can be addressed with a dual-pore arrayed device with commensurate low-noise and multichannel application-specific integrated circuits (ASICs). Recent commercialization of protein nanopores for sequencing by Oxford Nanopore Technologies has shown that many of these challenges are surmountable. Finally, while our “zoom out” control logic allows for interrogation of *a priori* unknown tag numbers and tag patterns within molecules of *a priori* unknown length, we observed that the controller sometimes does not fully explore the molecule’s length. While this was not the case for λ -DNA with the 15-tag experiments, longer molecules may preferentially only be scanned near the end, and a failure to recapture the molecule may mean losing the opportunity to explore the entire tag set. Molecular dynamics simulations of molecule translocation and capture processes in two pore geometries may assist in further optimizing our scanning protocols to increase read-length, coverage and molecule capture efficiency.^{35,42}

We demonstrated resolution of features separated by 150 bp at the construct level and 166 bp vis-a-vis the *E. coli* genome reference map. This result is comparable to the 141 bp separation reported by Chen *et al.* using a 5 nm diameter pore³ and superior to commercialized optical approaches which are limited to 1000 bp.¹ We observed an average of 58.6% return to baseline between the detected peaks, therefore it is reasonable to expect that with sharper peak resolution from higher bandwidth recording the true resolution is less than 150 bp. We expect that optimization of the structure of the OdN tags and reduced DNA speed by tuning the competing voltages during flossing can also improve the spatial resolution of the dual-pore method. nucleosomes (150 bp of wrapped plus linker DNA⁴³), allowing mapping of chromatin accessibility and potentially DNA binding proteins.

In conclusion, we view the dual-pore as a promising technology for genome scaling of solid-state nanopore technology. While genome mapping is a potential application, in the future we envision that the molecular feature used for barcoding (in our case OdN tags) can provide a scaffold for organizing, relative to the genome, the location of additional molecular motifs that might be discriminated based on size differences, such as nucleosomes or regulatory proteins, to provide a functional annotation/overlay on top of the sequence motif map. Labeling methods to differentially tag epigenetic sites can also be explored, *e.g.*, across CpG islands to assay methylation. Notably, while genomic rearrangements are identifiable with sufficiently high throughput physical genome mapping technologies using optics,¹ the resolution at 1000 bp is not high enough to capture the footprint of nucleosomes (~150 bp⁴³) and CpG islands (as small as 300 bp; with 1–10 CpGs/100 bp in mammalian promoters⁴⁴). Meanwhile, Somatic structural variations appear to have a major role in shaping the cancer DNA methylome.⁴⁵ A tool that can capture

genome-wide genomic and epigenetic alterations on single molecules, including changes to chromatin accessibility and methylation, would benefit research in cancers, aging-related diseases, and other conditions driven by such alterations.

EXPERIMENTAL METHODS

Conjugation of Oligodeoxynucleotide Tags. We use two DNA substrates in this study, λ -DNA (New England Biolabs) and *E. coli* (K12 strain). Both are treated with nicking endonucleases as the first step in installing oligodeoxynucleotide (OdN) tags for dual-pore detection. The λ -DNA reagents are prepared starting with 2 μ g of commercially prepared DNA incubated with 25 units of Nt.BbvCI and Nb.BssSI to a final volume of 100 μ L in 1 \times 3.1 buffer (New England Biolabs). Nicking of genomic *E. coli* DNA was performed identically using Nb.BsrDI and 1X CutSmart (New England Biolabs) substituted for the two enzymes used to prepare λ -DNA. In both cases, the nicking reaction is incubated at 37 $^{\circ}$ C for 1 h. Nick translation was initiated by the addition of 5 μ L of 10 μ M dUTP-azide, dATP, dGTP, and dCTP, (ThermoFischer Scientific) and Taq polymerase (New England Biolabs) in 1 \times Standard Taq buffer. The mixture was incubated at 68 $^{\circ}$ C for 1 h at which point 3 μ L of 0.5 μ M EDTA was added to quench the reaction. The DNA was then purified by phenol/chloroform extraction and overnight ethanol precipitation. The dried pellet was washed with 70% ethanol, air-dried, and resuspended in 500 mM NaCl/12.5 mM PO₄ at pH 6. Synthetic OdNs with dibenzocyclooctynes (DBCO) moieties were added to the resuspended DNA to a final concentration of 1 μ M and reacted at 60 $^{\circ}$ C overnight.⁴⁶ Final reaction mixture was used in nanopore experiments directly following dilution in 2 M LiCl to 1 ng/ μ L. The OdN sequences can be found in [Supporting File 2](#). Both sequences had a 5' DBCO moiety.

Achieving DNA Flossing. Flossing tagged dsDNA with a dual-nanopore system has been extensively described by Liu *et al.*²¹ We provide here a brief description. Tagged DNA molecules are introduced to the “common chamber” above the two nanopores ([Figure 1a](#), idle). A single DNA molecule is captured in the channel below pore 1 by a driving positive voltage across pore 1 ([Figure S1](#), State 1 and State 2). After translocation, a portion of the DNA strand is threaded back through pore 1 by a driving negative voltage ([Figure S1](#), State 3). The protruding DNA is captured in pore 2 achieving recapture by applying a positive voltage across pore 2 ([Figure S1](#), State 4).

Construction of 150 Base Pair Tag-Pair DNA. We generated a λ -DNA reagent with only two tags separated by 150 bp to test the resolution of the dual-pore instrument. We prepared the Cas9D10A nickase ribonucleoprotein (RNP) by incubating 1 μ L of 100 μ M annealed guide RNA and tracer DNA with 1 μ L of 10 μ M Cas9D10A (IDT) in NEB 3.1 buffer for 10 min at rt then put on ice until use. The guide RNA sequences were CATT TTT TTT TCGT GAGCAAT and AATTCAGGATAATGTGCAAT for the 5' and 3' cut sites, respectively ([Supporting File 2](#)). Both Cas9 RNPs were combined with the λ -DNA template to a final concentration of 100 nM RNP (each) and 1.56 nM λ -DNA and incubated at 37 $^{\circ}$ C for 1 h. The reaction was purified by phenol/chloroform extraction followed by ethanol precipitation and resuspended in deionized water. We then installed 60 nt tags at the nick sites using the methods described in [Conjugation of Oligodeoxynucleotide Tags](#).

Isolation of Genomic *E. coli* DNA. The *E. coli* cells were grown in LB media overnight to stationary phase. The cells were harvested by centrifugation, and genomic DNA was isolated using the Circulomics Nanobind HMW DNA extraction kit as per manufacturer instructions.

Nanopore Measurement. We performed nanopore experiments as previously described by our group.²¹ Briefly, the dual-pore chip was assembled in a custom fabricated flow cell. Our optimized flow cell uses 7 μ L of 1 ng/ μ L of substrate DNA for each experiment. Ag/AgCl electrodes are fabricated for use with the flow cell on a thin PET sheet that is positioned adjacent to the dual-pore chip and with electrodes in the relevant fluidic flow paths. The current and voltage signal was

collected by Molecular Device Multi-Clamp 700B and was digitized by Axon Digidata 1550. The signal is sampled at 250 kHz and filtered at 10 kHz. The tag-sensing and voltage control module was built on National Instruments Field Programmable Gate Array (FPGA) PCIe-7851R and control logic was developed and run on the FPGA through LabView.

Tag Calling. The installed DNA tags cause a characteristic “spike” attenuation in the ionic current as the DNA moves through the nanopore sensors (Figure 1b and Figure S4). In online analysis, the flossing with zoom out logic requires detecting and counting tags. Detection is performed only in pore 2 current and when a tag creates a 70 pA or larger deviation in the 10 kHz signal compared to a moving average (2500 samples) filtered signal that emulates the DNA tag-free baseline signal (Figure S4). Tags that produce a deviation less than 70 pA are missed. In off-line analysis, to perform alignment, we need to determine where these tags translocate through the two nanopores. The process to determine the location of tags is the same for the ionic current from pore 1 and pore 2 and follows a five-step process: (1) the current is inverted (making the downward spikes into upward ones); (2) an exponential fit is performed to detrend the transient in the baseline of the signal (and this fit is subtracted from the signal); (3) the signal is mean-shifted and standardized; (4) a Gaussian filter is applied to the signal resulting in a smoothed representation; and (5) a peak-calling algorithm from the *scipy* library⁴⁷ is applied the tag location and assigned to the maximum of the peak in the Gaussian filtered signal.

Data Processing and Filtering. We filtered the data for downstream use. This was done at the scan level, meaning a single molecule capture may have multiple scans (L–R or R–L recordings of the molecule) which pass or fail the filtering criteria. The process involves three steps, tag-count filtering, scan-length filtering and pore 1/pore 2 tag count imbalance filtering.

Tag Count Filtering. For a molecule to progress to flossing, it must enter a tug-of-war state, automatically filtering out small DNA molecules and material that is not linear dsDNA (for example, DNA with a double-strand break). However, occasionally DNA molecules or other material comes in contact with the dual-pore sensor during flossing that result in high noise signal. The tag detector will overcall the number of tags in these scans (oftentimes detecting 100s of tags erroneously), so we filter out scans with greater than 20 detected tags. For *E. coli*, we remove scans without at least four tags, and for λ -DNA, we remove scans with fewer than three tags.

Scan Length Filtering. Short scans are typically false flossing triggers and are symptomatic of noisy or poor quality data. Similarly, extremely long scans are often due to stalling, molecule stiction in the pore or similar pathologies. We removed scans which are shorter than 1.6 ms and longer than 400 ms.

Tag Count Imbalance Filtering. We rely on detecting tags in both pore 1 and pore 2 to estimate velocity (Estimating Scan Velocity and Linear Distance between Adjacent Tags). We removed scans where the number of tags detected in pore 1 (n_1) appreciably differed from the number of tags detected in pore 2 (n_2) by calculating the relative difference in their counts and requiring:

$$|n_1 - n_2| \leq \frac{1}{4}(n_1 + n_2)$$

Estimating Scan Velocity and Linear Distance between Adjacent Tags. The common chamber is the volume adjacent to both nanopores. We refer to “entering” tags as ones which are coming from the channel into the common chamber and “exiting” tags as ones leaving the common chamber into the channel. In the L–R direction the entering tags are detected in pore 1 and the exiting tags are detected in pore 2, whereas in the R–L direction the entering tags are detected in pore 2 and the exiting tags are detected in pore 1. Agnostic to scan direction, a tag is assigned a time-of-flight velocity V_{TOF} by dividing the tags entry-to-exit transit time into the known distance between the pores. This produces units of linear distance (nanometers) divided by time (microseconds) for V_{TOF} . In detail, for each scan we iterate over the entering tags and look for corresponding exiting tags which would produce a V_{TOF} within the interval (0.4, 3.5)

nm/ μ s given the known pore-to-pore distance. Given multiple candidates, the first one in logical order is taken. Once a tag is paired, it is removed from consideration. For the purpose of generating a scan velocity profile, the time value assigned to each tags V_{TOF} value is at the point of that tags exit time. The scan velocity profile is the piece-wise linear curve that linearly interpolates between the computed V_{TOF} values, and is constant and equal to the first and last V_{TOF} values before and after they occur, respectively. Lastly, with either pore current, the linear distance between two sequentially detected tags is computed by integrating the scan velocity profile over the intertag time period.

Description of Alignment Algorithm. Individual scans are aligned to a reference map using an adapted version of the Smith–Waterman local alignment algorithm.³⁸ There are two adaptations made for dual-pore distance data. First, instead of a substitution matrix, the score model described in eq 1 is used when determining the match score for a pair of tags and a length of nucleotides in the reference. Second, we must allow for incomplete tagging of the DNA molecule as well as spuriously detected tags. At a given step in the dynamic programming we search for the best match among interval covered up to the current position as well as the intervals covered using upstream tags as the start to the interval. More formally, let $s_{[i-a,i],[j-b,j]}$ be the score $s(D_{i-a,i}, G_{j-b,j})$, where $D_{i-a,i}$ is the estimated linear distance between tags indexed at $i - a$ and i , $G_{j-b,j}$ is the genomic interval distance between nicking sites indexed at $j - b$ and j , and indices enumerate candidate restriction sites on the reference map. At each step in the dynamic programming recursion the match score is defined as

$$(a, b) = \arg \max_{a,b} (s_{[i-a,i],[j-b,j]}, \forall a \in [1, A], \forall b \in [1, B])$$

$$M_{i,j} = F_{i-a,j-b} + s_{[i-a,i],[j-b,j]}$$

where A and B are nick-site counting parameters that define the index search size (default 3 for both), F is a dynamic programming matrix recursively defined below, and M is the optimized score. The recursions for “inserting” and “deleting” tags, as part of the dynamic programming search and which correspond to spuriously detecting tags or failing to detect a tag, respectively, use a cost based on the maximum score of the model. The maximum potential value for the score model is defined when the ratio $\frac{D_{i-a,i}}{G_{j-b,j}} = \bar{\mu}$, with $\bar{\mu}$ defined in

Equation 1. Let s_{max} be the maximum score for the model. The recursion for inserting a tag is computed as

$$I_{i,j} = F_{i-1,j} - \lambda s_{\text{max}}$$

and the recursion for deleting a tag is similarly computed as

$$E_{i,j} = F_{i,j-1} - \gamma s_{\text{max}}$$

where λ and γ are parameters to the algorithm to adjust the relative cost of inserting a tag or deleting a tag. The default values are $\lambda = 0.1$, $\gamma = 1.0$. The full recursion for the dynamic programming is

$$F_{i,j} = \max \begin{cases} M_{i,j} \\ I_{i,j} \\ E_{i,j} \end{cases}$$

Joining Recapture Alignments. Consensus alignments from a recaptured molecule may overlap (for example, Figure 7) or be disjointed. The DNA isolation procedure suggests that the fragments are longer than 50 kb (Figure S2). When the consensus alignments do not overlap we follow a procedure of grouping the consensus alignments from a single molecule when they have at least two scans of support and align to the reference within 250 kbp of each other. We then take the window with the highest scoring set of alignments as the final aggregated single molecule alignment. Our data contained 7 and 28 molecules with disjoint and overlapping consensus alignments, respectively. The mean distance between disjoint consensus alignments is 21.9 kb.

Calculating N50 of Alignments. Let $\{\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n\}$ be the set of all $n + 1$ consensus molecule alignments that met acceptance criteria. Let $l(\mathcal{A}_j)$ be the net alignment length (in base pairs) defined as the maximum genomic coordinate minus the minimum genomic coordinate among the set of aligned tags, and computed for each consensus alignment \mathcal{A}_j , $j = 0, \dots, n - 1$. The net alignment length is thus a first-to-last tag distance and does not include the length of the molecules outside of these tags. Next, let $\{\mathcal{A}_{k_0}, \mathcal{A}_{k_1}, \dots, \mathcal{A}_{k_n}\}$ be the original set but now sorted in descending from largest net alignment length $l(\mathcal{A}_{k_0}) = \max_i l(\mathcal{A}_{k_i})$ to the smallest net alignment length $l(\mathcal{A}_{k_n}) = \min_i l(\mathcal{A}_{k_i})$. Define the cumulative sum of net alignment lengths as

$$g(j) = \sum_{i=0}^j l(\mathcal{A}_{k_i})$$

with the total alignment length denoted as $L = g(k_n)$. Note that the cumulative alignment length as a percentage of the total length of all alignments plotted in Figure 8b is $g(j)/L$. The 50% mark of the total alignment length is denoted $L_{50\%} = \lfloor 0.5 \times L \rfloor$. Finally, we calculate the N50 as

$$\text{N50} = \arg \min_j l(\mathcal{A}_{k_j}), \text{ such that } g(j) \geq L_{50\%}$$

The N50 is the smallest net alignment length in the sorted list such that the cumulative sum up to that length is at least as big as the 50% mark of the total alignment length.

Scan-Space Alignment. In order to estimate of the standard deviation for the DNA stretching factor we aligned scans from λ to a reference scan. For each molecule scan in the Phage Lambda data set, we generate a synthetic reference scan by assuming the DNA stretching factor is equal to 0.34 nm/bp and the V_{TOF} for every tag is equal to the average for that scan.

The alignment procedure begins by taking one tag from each scan (called the anchor pair of tags, Figure S3, red line) and moving them into alignment; it then calculates time from the anchor tag t_i to every other distal tag t_j in the scan. The cost C_{t_i, t_j} of aligning two tags is defined as the squared difference in their distance:

$$C_{t_i, t_j} = \begin{cases} \inf & \text{sign}(t_i) \neq \text{sign}(t_j) \\ |t_i - t_j|^2 & \end{cases}$$

If C_{t_i, t_j} exceeds a user-defined value (default $400 \mu\text{s}^2$) the tag is left unpaired. The distal tags are aligned (blue lines) using a dynamic programming algorithm analogous to global alignment⁴⁸ to find the pairings which minimize the error. The first row and column of the dynamic programming matrix are initialized to the gap cost (a user-provided parameter) multiplied by the number of tags skipped on alignment. The recursion proceeds by taking the minimum of the difference in the distance between the two tags to their anchor or incurring a skip cost added to the corresponding prior cell in the matrix. The minimum error over all pairwise combinations is taken from the bottom right cell of the matrix and a traceback is performed. Every pair of tags is attempted as the anchor pair. The alignment with minimum error over all anchor pairs is used as the final alignment. This procedure produces a mapping of scan intervals to genomic positions, which we used to estimate the initial value of σ .

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsnano.1c09575>.

Supplementary figures accompanying the Experimental Methods section (PDF)

Archived folder contains the output from running Tandem Repeats Finder (version 4.09) on the *E. coli*

genome. The reference sequence used is included as well as the output containing the parameters for running the program (ZIP)

This FASTA-formatted sequence file contains the nucleotide sequences of the 60nt tags used in this work. The guide RNA sequences were used with the Cas9D10A nickase to install tags separated by 150 bp on λ DNA (ZIP)

AUTHOR INFORMATION

Corresponding Author

William B. Dunbar – Nooma Bio, Santa Cruz, California 95060-5790, United States; orcid.org/0000-0002-0913-318X; Email: bill@nooma.bio

Authors

Arthur Rand – Nooma Bio, Santa Cruz, California 95060-5790, United States

Philip Zimny – Nooma Bio, Santa Cruz, California 95060-5790, United States

Roland Nagel – Nooma Bio, Santa Cruz, California 95060-5790, United States

Chaitra Telang – Nooma Bio, Santa Cruz, California 95060-5790, United States

Justin Mollison – Nooma Bio, Santa Cruz, California 95060-5790, United States

Aaron Bruns – Nooma Bio, Santa Cruz, California 95060-5790, United States

Emily Leff – Nooma Bio, Santa Cruz, California 95060-5790, United States

Walter W. Reisner – Department of Physics, McGill University, Montreal, QC, Canada H3A 2T8; orcid.org/0000-0003-3223-615X

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsnano.1c09575>

Notes

The authors declare the following competing financial interest(s): The authors declare competing financial interests: A.R., P.Z., R.N., C.T., J.M., A.B., E.L., and W.B.D. are employees of Nooma Bio, Inc., which has exclusively licensed the dual-pore device patent from the University of California, Santa Cruz, for commercialization purposes. All work was privately financed by Nooma Bio, with the exception of the tag-pair resolution work that was funded by NIH grant 1R21HG011236-01.

This article was uploaded to bioRxiv prior to publication as Rand, A.; Zimny, P.; Nagel, R.; Telang, C.; Mollison, J.; Bruns, A.; Leff, E.; Reisner, W.; Dunbar, W. B. Electronic Mapping of a Bacterial Genome with Dual Solid-State Nanopores and Active Single-Molecule Control. 2021, <https://www.biorxiv.org/content/10.1101/2021.10.29.466509v1> (March 7, 2022).

ACKNOWLEDGMENTS

The work done in Santa Cruz was performed and financially supported by Nooma Bio, Inc. The contributions to the text by W.R. and the tag-pair regent costs at Nooma Bio were financially supported by the NIH (Grant No. NHGRI R21HG011236).

REFERENCES

- (1) Lam, E. T.; Hastie, A.; Lin, C.; Ehrlich, D.; Das, S. K.; Austin, M. D.; Deshpande, P.; Cao, H.; Nagarajan, N.; Xiao, M.; Kwok, P.-Y. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **2012**, *30*, 771–776.
- (2) Plesa, C.; Ruitenbergh, J. W.; Witteveen, M. J.; Dekker, C. Detection of Individual Proteins Bound along DNA Using Solid-State Nanopores. *Nano Lett.* **2015**, *15*, 3153–3158.
- (3) Chen, K.; Juhasz, M.; Gularek, F.; Weinhold, E.; Tian, Y.; Keyser, U. F.; Bell, N. A. W. Ionic Current-Based Mapping of Short Sequence Motifs in Single DNA Molecules Using Solid-State Nanopores. *Nano Lett.* **2017**, *17*, 5199–5205.
- (4) Kong, J.; Bell, N. A. W.; Keyser, U. F. Quantifying Nanomolar Protein Concentrations Using Designed DNA Carriers and Solid-State Nanopores. *Nano Lett.* **2016**, *16*, 3557–3562.
- (5) Squires, A.; Atas, E.; Meller, A. Nanopore sensing of individual transcription factors bound to DNA. *Sci. Rep.* **2015**, *5*, 11643.
- (6) Soni, G. V.; Dekker, C. Detection of Nucleosomal Substructures using Solid-State Nanopores. *Nano Lett.* **2012**, *12*, 3180–3186.
- (7) Singer, A.; Rapireddy, S.; Ly, D. H.; Meller, A. Electronic Barcoding of a Viral Gene at the Single-Molecule Level. *Nano Lett.* **2012**, *12*, 1722–1728.
- (8) Morin, T. J.; Shropshire, T.; Liu, X.; Briggs, K.; Huynh, C.; Tabard-Cossa, V.; Wang, H.; Dunbar, W. B. Nanopore-Based Target Sequence Detection. *PLoS One* **2016**, *11*, 1–21.
- (9) Liu, K.; Pan, C.; Kuhn, A.; Nievergelt, A. P.; Fantner, G. E.; Milenkovic, O.; Radenovic, A. Detecting topological variations of DNA at single-molecule level. *Nat. Commun.* **2019**, *10*, 3.
- (10) Bell, N. A. W.; Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat. Nanotechnol.* **2016**, *11*, 645–651.
- (11) Chen, K.; Kong, J.; Zhu, J.; Ermann, N.; Predki, P.; Keyser, U. F. Digital Data Storage Using DNA Nanostructures and Solid-State Nanopores. *Nano Lett.* **2019**, *19*, 1210–1215.
- (12) Zhu, J.; Ermann, N.; Chen, K.; F. K. U. Image Encoding Using Multi-Level DNA Barcodes with Nanopore Readout. *Small* **2021**, *17*, 2100711.
- (13) Sze, J. Y. Y.; Ivanov, A. P.; Cass, A. E. G.; Edel, J. B. Single molecule multiplexed nanopore protein screening in human serum using aptamer modified DNA carriers. *Nat. Commun.* **2017**, *8*, 1552.
- (14) Kong, J.; Zhu, J.; Chen, K.; Keyser, U. F. Specific Biosensing Using DNA Aptamers and Nanopores. *Adv. Funct. Mater.* **2019**, *29*, 1807555.
- (15) Albrecht, T. Single Molecule Analysis with Solid-State Nanopores. *Annual Review of Analytical Chemistry* **2019**, *12*, 371.
- (16) Yuan, Y.; Chung, C. Y.-L.; Chan, T.-F. Advances in optical mapping for genomic research. *Computational and Structural Biotechnology Journal* **2020**, *18*, 2051–2062.
- (17) Joffe, J.; Kobo, A.; Su, T.; Grunwald, A.; Green, O.; Nilsson, A. N.; Eisenberg, E.; Ambjörnsson, T.; Westerlund, F.; Weinhold, E.; Shabat, D.; Purohit, P. K.; Ebenstein, Y. Super-Resolution Genome Mapping in Silicon Nanochannels. *ACS Nano* **2016**, *10*, 9823–9830.
- (18) Chen, P.; Jing, X.; Ren, J.; Cao, H.; Hao, P.; Li, X. Modelling BioNano optical data and simulation study of genome map assembly. *Bioinformatics* **2018**, *34*, 3966–3974.
- (19) Farimani, A. B.; Min, K.; Aluru, N. R. DNA Base Detection Using a Single-Layer MoS₂. *ACS Nano* **2014**, *8*, 7914–7922.
- (20) Liu, X.; Zhang, Y.; Nagel, R.; Reisner, W.; Dunbar, W. B. Controlling DNA Tug-of-War in a Dual Nanopore Device. *Small* **2019**, *15*, 1901704.
- (21) Liu, X.; Zimny, P.; Zhang, Y.; Rana, A.; Nagel, R.; Reisner, W.; Dunbar, W. B. Flossing DNA in a Dual Nanopore Device. *Small* **2020**, *16*, 1905379.
- (22) Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. Translocation of double-strand DNA through a silicon oxide nanopore. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* **2005**, *71*, 051903.
- (23) Waugh, M.; Briggs, K.; Gunn, D.; Gibeault, M.; King, S.; Ingram, Q.; Jimenez, A. M.; Berryman, S.; Lomovtsev, D.; Andrzejewski, L.; Tabard-Cossa, V. Solid-state nanopore fabrication by automated controlled breakdown. *Nat. Protoc.* **2020**, *15*, 122–143.
- (24) Lu, B.; Albertorio, F.; Hoogerheide, D. P.; Golovchenko, J. A. Origins and Consequences of Velocity Fluctuations during DNA Passage through a Nanopore. *Biophys. J.* **2011**, *101*, 70–79.
- (25) Plesa, C.; van Loo, N.; Ketterer, P.; Dietz, H.; Dekker, C. Velocity of DNA during Translocation through a Solid-State Nanopore. *Nano Lett.* **2015**, *15*, 732–737.
- (26) Chen, K.; Jou, I.; Ermann, N.; Muthukumar, M.; Keyser, U. F.; Bell, N. A. W. Dynamics of driven polymer transport through a nanopore. *Nat. Phys.* **2021**, *17*, 1043.
- (27) Sarabadani, J.; Ala-Nissila, T. Theory of pore-driven and end-pulled polymer translocation dynamics through a nanopore: an overview. *J. Phys.: Condens. Matter* **2018**, *30*, 274002.
- (28) Verschueren, D. V.; Yang, W.; Dekker, C. Lithography-based fabrication of nanopore arrays in freestanding SiN and graphene membranes. *Nanotechnology* **2018**, *29*, 145302.
- (29) Pearson, M. D.; Nguyen, L.; Zhao, Y.; McKenna, W. L.; Morin, T. J.; Dunbar, W. B. Fast and accurate quantification of insertion-site specific transgene levels from raw seed samples using solid-state nanopore technology. *PLoS One* **2019**, *14*, e0226719.
- (30) Pud, S.; Chao, S.-H.; Belkin, M.; Verschueren, D.; Huijben, T.; van Engelenburg, C.; Dekker, C.; Aksimentiev, A. Mechanical Trapping of DNA in a Double-Nanopore System. *Nano Lett.* **2016**, *16*, 8021–8028.
- (31) Cadinu, P.; Paulose Nadappuram, B.; Lee, D. J.; Sze, J. Y. Y.; Campolo, G.; Zhang, Y.; Shevchuk, A.; Ladame, S.; Albrecht, T.; Korchev, Y.; Ivanov, A. P.; Edel, J. B. Single Molecule Trapping and Sensing Using Dual Nanopores Separated by a Zeptoliter Nanobridge. *Nano Lett.* **2017**, *17*, 6376–6384.
- (32) Cadinu, P.; Campolo, G.; Pud, S.; Yang, W.; Edel, J. B.; Dekker, C.; Ivanov, A. P. Double barrel nanopores as a new tool for controlling single-molecule transport. *Nano Lett.* **2018**, *18*, 2738–2745.
- (33) Zhang, Y.; Liu, X.; Zhao, Y.; Jen-Ken, Y.; Reisner, W.; Dunbar, W. B. Single-Molecule DNA Resensing Using a Two-Pore Device. *Small* **2018**, *14*, 1801890.
- (34) Chen, K.; Gularek, F.; Liu, B.; Weinhold, E.; Keyser, U. F. Electrical DNA Sequence Mapping Using Oligodeoxynucleotide Labels and Nanopores. *ACS Nano* **2021**, *15*, 2679–2685.
- (35) Seth, S.; Bhattacharya, A. DNA barcode by flossing through a cylindrical nanopore. *RSC Adv.* **2021**, *11*, 20781–20787.
- (36) Jiang, Y.; Feldman, T.; Bakx, J. A. M.; Yang, D.; Wong, W. P. Stretching DNA to twice the normal length with single-molecule hydrodynamic trapping. *Lab Chip* **2020**, *20*, 1780–1791.
- (37) Smith, S. B.; Cui, Y.; Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **1996**, *271*, 795–799.
- (38) Smith, T.; Waterman, M. Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197.
- (39) Chen, P.; Jing, X.; Ren, J.; Cao, H.; Hao, P.; Li, X. Modelling BioNano optical data and simulation study of genome map assembly. *Bioinformatics* **2018**, *34*, 3966–3974.
- (40) Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580.
- (41) Chan, P. P.; Holmes, A. D.; Smith, A. M.; Tran, D.; Lowe, T. M. The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res.* **2012**, *40*, D646–D652.
- (42) Seth, S.; Bhattacharya, A. DNA barcodes using a double nanopore system. *Sci. Rep.* **2021**, *11*, 9799.
- (43) Grigoryev, S. A. Nucleosome spacing and chromatin higher-order folding. *Nucleus (Austin, Tex.)* **2012**, *3*, 493–499.
- (44) Bestor, T. H.; Edwards, J. R.; Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6796.
- (45) Zhang, Y.; et al. Global impact of somatic structural variation on the DNA methylome of human cancers. *Genome Biology* **2019**, *20*, 209.

(46) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide—Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

(47) Virtanen, P.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(48) Durbin, R.; Eddy, S. R.; Krogh, A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1998.

(49) Seth, S.; Rand, A.; Reisner, W.; Dunbar, W. B.; Sladec, R.; Bhattacharya, A. Discriminating protein tags on dsDNA constructs using a dual Nanopore device. *arXiv* **2022**, DOI: [10.48550/arXiv:2201.08809](https://doi.org/10.48550/arXiv:2201.08809).