Multiplexed DNA identification using site specific dCas9 barcodes and nanopore sensing

Nicole E. Weckman,[†] Niklas Ermann,[†] Richard Gutierrez,[‡] Kaikai Chen,[†] James Graham,[‡] Ran Tivony,[†] Andrew Heron,[‡] and Ulrich F. Keyser[*,†]

[†]*Cavendish Laboratory, JJ Thomson Ave, University of Cambridge, Cambridge, UK, CB3 0HE*

[‡]*Oxford Nanopore Technologies, Gosling Building, Edmund Halley Road, Oxford Science Park OX4 4DQ*

E-mail: ufk20@cam.ac.uk

Phone: +44 (0)1223 337272

Abstract

Decorating double-stranded DNA with dCas9 barcodes to identify characteristic short sequences provides an alternative to fully sequencing DNA samples for rapid and highly specific analysis of a DNA sample. Solid-state nanopore sensors are especially promising for this type of single-molecule sensing because of the ability to analyze patterns in the ionic current signatures of DNA molecules. Here, we systematically demonstrate the use of highly specific dCas9 probes to create unique barcodes on the DNA that can be read out using nanopore sensors. Single dCas9 probes are targeted to various positions on DNA strands up to 48 kbp long and are effectively measured in high salt conditions typical of nanopore sensing. Multiple probes bound to the same DNA strand at characteristic target sequences create distinct barcodes of double and triple peaks. Finally, double and triple barcodes are used to simultaneously identify

1

two different DNA targets in a background mixture of bacterial DNA. Our method forms the basis of a fast and versatile assay for multiplexed DNA sensing applications in complex samples.

Keywords

Nanopore sensing, Cas9, DNA barcode, single-molecule detection, biosensor

The advent of personalized medicine depends on the development and adoption of improved medical technologies that are able to rapidly diagnose patient illnesses and direct treatments. Nanopore sensors are one such promising technology for high throughput single-molecule analysis[1] often based on biological nanopores used for DNA sequencing.[2] However, DNA sequencing for pathogen identification, for example to diagnose urinary tract infections, still has a time frame on the order of several hours between taking a sample and delivery of a result.[3] The identification of DNA can also be done by labeling of distinct sequences, hence simplifying the process. For example for the diagnosis of a bacterial infection, it is critical to rapidly identify the pathogen causing the infection and screen for potential resistance genes in order to quickly start an appropriate antibiotic course, avoiding the use of ineffective antibiotics.[3] In cases such as these, rapid clinical decision making requires knowledge about the presence or absence of certain characteristic genes rather than the full genome. An alternative to using biological nanopores for full sequencing is to use solid state nanopores which demonstrate potential for measuring and studying DNA, RNA, proteins, and their interactions.[4–6] Mapping characteristic short sequence motifs is a particularly promising alternative to full genetic sequencing, especially when a rapid response on the order of minutes is required.

The CRISPR associated (Cas) protein, Cas9, is a protein ideally suited for labelling target

sequences on a DNA strand to ultimately allow for identification of the DNA. Originating as part of the adaptive immune system of bacteria and archaea, the CRISPR-Cas9 system uses a guide RNA sequence to specifically identify, bind, and cleave a double stranded DNA target sequence.[7,8] A dead or inactive Cas9 protein, dCas9, has been used to maintain the highly specific binding to the designed target sequence without cleaving the DNA.[9]

Recently it has been demonstrated that nanopores can be used to observe and estimate the location of a dCas9 protein bound to a target DNA sequence, both as the DNA freely translocates through the nanopore[10] and while the translocation is controlled using optical tweezers.[11] The basic premise underpinning these nanopore sensing studies is that as double stranded DNA (dsDNA) is electrically driven through a nanopore that is immersed in an electrolyte, bound dCas9 causes an additional drop in the ionic current beyond that caused by dsDNA alone. This system of analyzing secondary current drops or peaks caused by structures bound to DNA carriers as they pass through a nanopore has been used to study RecA-coated dsDNA,[12] to label and map specific sequences on dsDNA,[13–15] to study the binding of small molecules to dsDNA,[16] to measure proteins bound to a specific location on dsDNA,[17–20] and to analyze DNA origami scaffolds capable of detecting proteins or small molecules.[21,22]

This paper demonstrates the design of dCas9 probes to bind to specific dsDNA sequences at several locations along the DNA. Nanopore sensors are used to identify the bound probes based on the resultant secondary peaks in the DNA translocations. Multiple dCas9 probes are then used to create characteristic patterns of structures, or barcodes, on the DNA. These sequence specific barcodes are used to identify two different DNA targets at the same time in a background DNA mixture, demonstrating the initial steps towards a multiplexed sensing system for rapidly identifying DNA based on characteristic sequence motifs.

Results and Discussion

Single dCas9 probes bound to short and long DNA

The dCas9 probes are first bound to a 3.6 kbp fragment of lambda phage DNA, with the probe targeted to a sequence located 218 bp from one end of the fragment, as schematically depicted in Figure 1 a). Detailed experimental methods can be found in the Methods section with additional information in the Supplementary Information (SI). Briefly, DNA-probe translocations through quartz nanopores (average diameter 17 +/- 5 nm) are measured for more than an hour in 2 M, 3 M, and 4 M LiCl salt conditions to assess the binding of the probes and translocation event characteristics.

The 3.6 kbp DNA is driven through the nanopore from the cis to the trans reservoir of the fluid cell causing characteristic ionic current drops. The functionalization of the DNA with the dCas9 probe causes a secondary peak beyond the current drop of a linear DNA translocation through the pore. These peaks in the DNA translocation events arising from the bound probe can be seen in Figure 1 b). Due to the target sequence being located 218 bp from the end of the DNA fragment, secondary peaks in the event occur at either the beginning or the end of the event depending on whether the probe enters the nanopore first or last. This creates distinctive symmetric groupings in normalized peak positions identified over many events in a measurement that are not seen in the control experiment, as shown in Figure S2 in the SI. The presence of the probe bound to the dsDNA has been verified using AFM and electrophoresis. An AFM image of dsDNA with the probe bound at the expected location is shown in Figure 1 c). An additional AFM image showing multiple dsDNA with bound probes as well as the electrophoresis results can be found in Figure S1 in the SI.

Figure 1 b) shows unfolded events with a bound dCas9 probe clearly visible at all salt concentrations. An unfolded event occurs when the nanopore captures the end of the dsDNA and it translocates linearly through the pore whereas a folded event occurs when the dsDNA is captured further along its length causing it to fold back on itself. The secondary peak
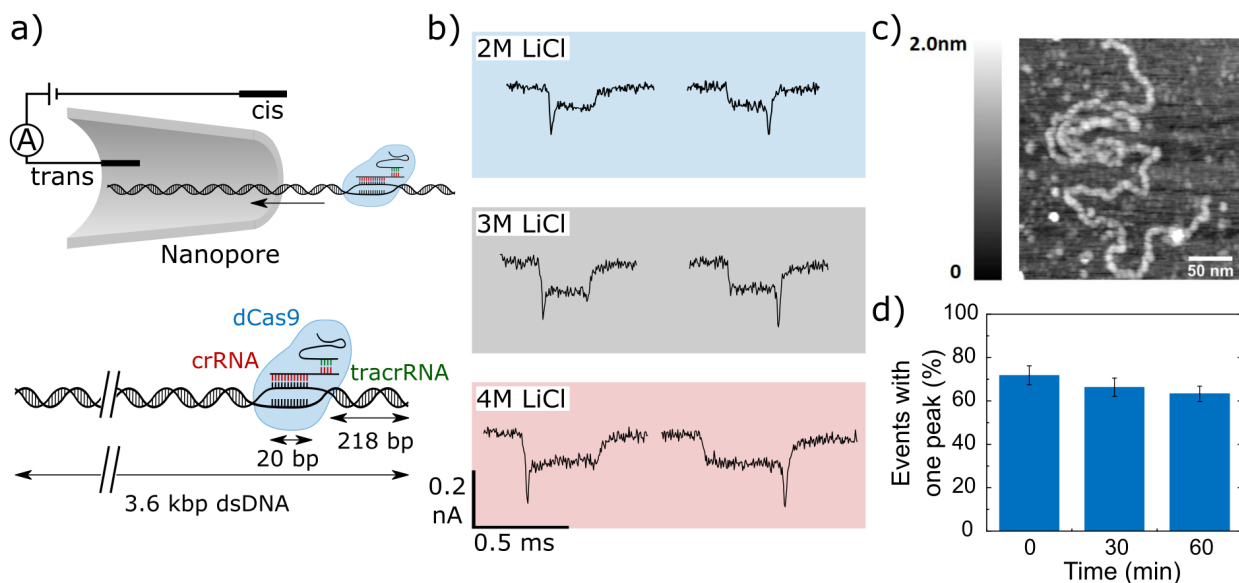
Figure 1: Single dCas9 probes bound to 3.6 kbp DNA measured in various salt conditions. (a) The dCas9-labelled DNA translocating through a nanopore. The probe is bound to a 3.6 kbp fragment of lambda DNA, 218 bp from one end of the DNA. The dCas9 complex has a complementary RNA (crRNA) that contains a 20 bp recognition sequence that specifically targets the DNA fragment. (b) Example events with a bound dCas9 probe measured in 2 M, 3 M, and 4 M LiCl salt conditions. Example events with the DNA translocating either with the probe at the beginning of the DNA translocation (left) or at the end (right) are shown. As the salt concentration increases, the event durations increase. (c) An AFM image of a 3.6 kbp double-stranded DNA with bound dCas9 located 218 bp from one end of the DNA, as schematically illustrated in Figure 1 a). The image was taken in air in non-contact mode, following the stripping of mica and subsequent coating with poly-L-lysine. (d) The percentage of events having a single peak in 2 M LiCl are analyzed from 5 min of data measured at t=0 min, t=30 min, and t=60 min. Over 60% of events are identified to have a probe at all relevant measurement times indicating no substantial loss of bound probes during the experiments. Error bars are the standard deviation over three independent measurements using different nanopores.

5

caused by the probe can be more difficult to identify in events due to folded DNA and hence such events are excluded prior to analysis. Examples of unfolded and folded DNA translocation events and more details about the screening and analysis process can be found in Section S4 and Figure S2 of the SI. As expected, lowering the salt concentration serves to decrease the percentage of folded events[23] and decrease the event durations.[24] The impact of salt concentration on the percentage of unfolded events and on the event durations is shown in Figure S3 of the SI.

The number of events available for further analysis is maximized by using 2M LiCl for the remainder of the study. The stability of the dCas9 binding in 2M LiCl is tested by collecting events for 5 min time periods each starting at t=0 min, t=30 min, and t=60 min. More than 60% of the events exhibit the secondary peak at all 3 of these time points, as seen in Figure 1 d). This indicates that our probes remain bound in the 2 M LiCl salt conditions over the course of an hour long measurement. The binding yield of greater than 60% measured for this dCas9 probe is better than values of 30-35% cited in the literature,[10] however some variability is observed in the binding yields of different probes, which is discussed further in Section S7 of the SI.

Additional unique dCas9 probes are designed with software design tools[25,26] to bind at various locations along the full length of lambda DNA while also minimizing predicted off-target binding. The full length lambda DNA is used so that probes can be targeted to several different sequences along the lambda DNA to demonstrate the specificity of the probe targeting. Placing probe binding sites away from the end of the DNA also reduces the interference of small folds often observed in translocation events.[27] This is further discussed in Section S4 in the SI. Probes were designed to bind to sequences at normalized positions of 0.19, 0.7 and 0.6 along the length of the 48.502 kbp lambda DNA, as shown in Figure 2 a). Similarly to the 3.6 kbp DNA, secondary current peaks identified on events indicate the binding of the dCas9 probes.

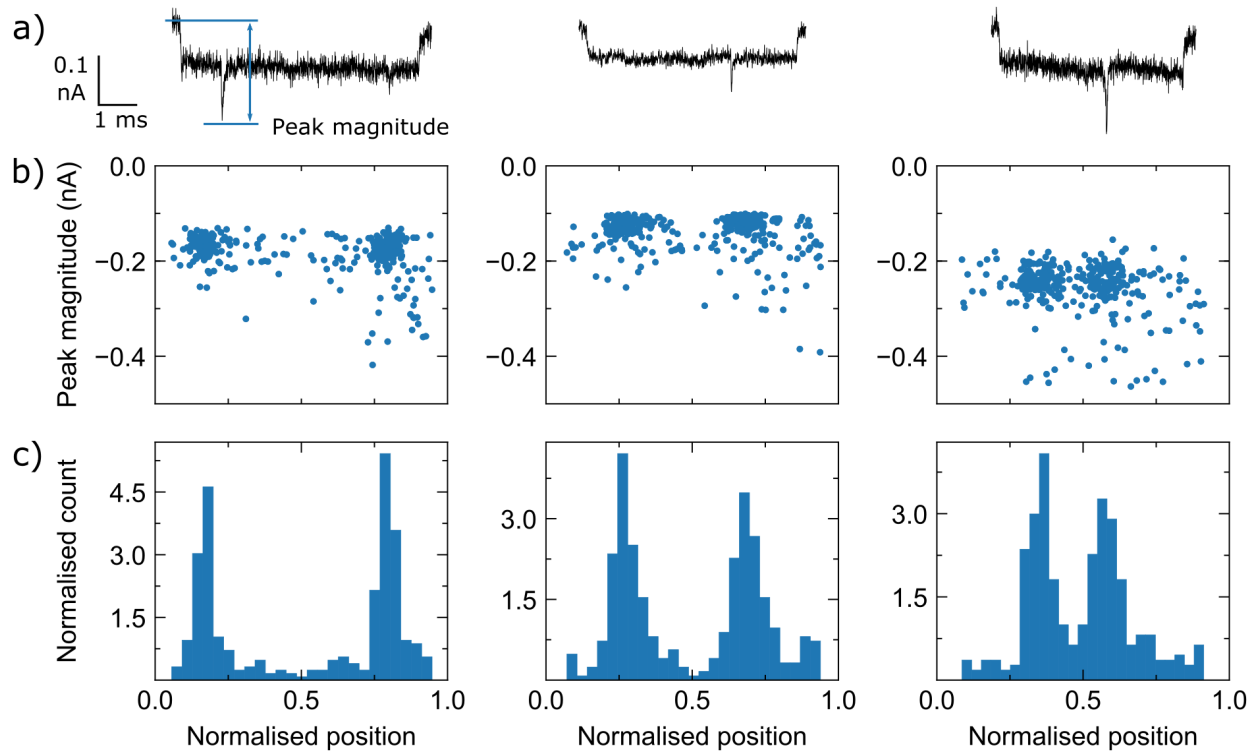As the targeted probe sequence is changed, peaks are identified at different normalized

Figure 2: Single probes targeting specific sequences at various positions along the full length of lambda DNA (48.502 kbp). Results are shown for three probes targeting sequences at normalized positions of (left) 0.19, (center) 0.70, and (right) 0.60 of the length of lambda DNA. a) Example events demonstrating the probe bound at different positions on the lambda DNA. Events shown are translocating in the forward direction and are measured with three different nanopores. b) Peak magnitude and normalized peak position for each peak on an event. Symmetric groupings of peaks occur due to DNA translocating through the nanopore in both directions. c) Normalized position of the peaks on events.

7

peak positions, as seen in Figure 2 b). The two symmetric groups in each measurement arise due to DNA translocations occurring with the end of the DNA with the probe entering the nanopore either first or last. The variable peak magnitude is due to the three different nanopores used for measurement. As the designed location of the targeted probe sequence moves towards the centre of the DNA, the maxima in the distributions of the normalized peak positions approach a normalized position of 0.5 as expected, Figure 2 c).

There is a background of peaks that are identified at all locations along the lambda DNA in each experiment as can be seen by the random occurrence of peaks outside of the main symmetric groups in Figure 2 b). These are also present in the control measurement of full length lambda DNA with no dCas9 probes, where up to 43% of events are found to have a single peak. These spurious peaks are attributed to knots or folds in the DNA, as has been previously studied.[27] The event selection procedure and probe design is further discussed in Sections S1 and S4 in the SI, where the control measurement of the lambda DNA without any dCas9 probes is also presented.

The incidence of knots and folds, particularly with long DNA, can make it difficult to differentiate conclusively between a single peak due to a specifically bound probe or a single peak due to a knot or fold in the DNA itself for any single event. For sensing applications of particularly rare target molecules, it is necessary to design a set of multiple probes to create a distinct peak pattern, or barcode, in order to accurately identify target DNA sequences as well as to enable multiplexed sensing.

Double and Triple dCas9 Barcodes

The simplest barcodes consist of binding two or three dCas9 probes closely spaced together on the target DNA to create a cluster of double or triple peaks. The cluster barcodes can then be distinguished based on the number of peaks in a single event as well as the characteristic spacing between these peaks. Examples of the distinctive signals generated by double and triple dCas9 probe clusters bound to full length lambda DNA are demonstrated in Figure 3
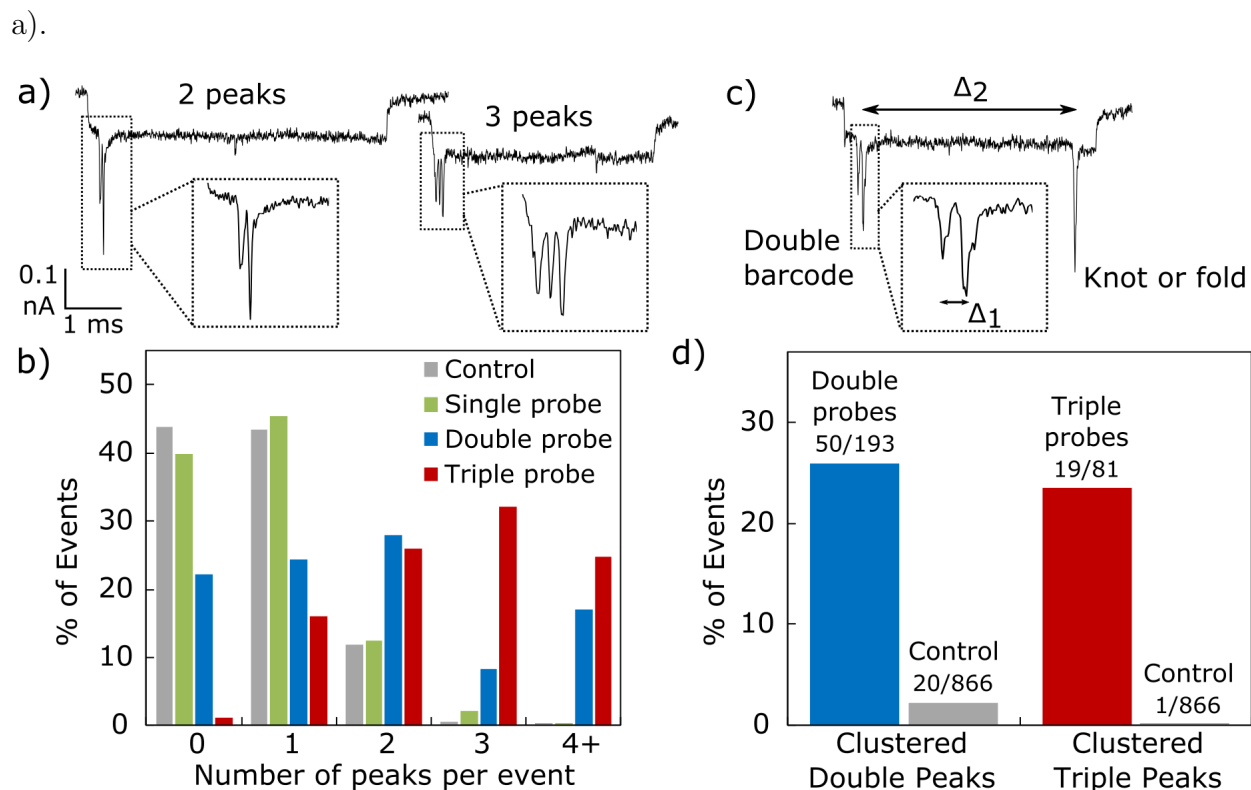
8

a).

Figure 3: Double and triple dCas9 probe barcodes on full length lambda DNA. (a) Example events with 2 and 3 peaks due to binding of the double probe barcode and triple probe barcode, respectively. The events are measured with two different pores. (b) Raw data comparing the number of peaks counted per event after the addition of a single probe, double probe and triple probe. (c) An example event with clustered double peaks (Δ_1 <threshold) that corresponds to the double probe barcode and a third unclustered peak (Δ_2 >threshold) often due to a fold or knot in the DNA. (d) Data for the double and triple probes from b) after clustering analysis to remove false positives. Percentage of events with clustered double peaks after the addition of the double probe barcode (blue) and clustered triple peaks after the addition of the triple probe barcode (red). The corresponding control of unlabelled lambda DNA is assessed using the same clustering thresholds as the barcoded samples. The data labels give the number of events identified with the correct barcode out of the total number of events.

Comparing the percentage of events that are identified with no peaks, one peak, two peaks, three peaks, or more peaks allows for assessment of how many probes are specifically binding to a DNA strand for any given barcode design. Figure 3 b) shows the number of peaks counted on events for samples where a single dCas9 probe, a double probe, or a triple probe have been used. It demonstrates the increase in the number of peaks identified on events resulting from increasing the number of probes used. Some events have fewer peaks

9

than expected due to one or more missing probes. For example, about 26% of events in the DNA sample with 3 probes (red) have only 2 peaks. Events with more peaks than the number of probes used are likely to have a small fold or a knot in addition to the barcode. However, spurious peaks are less problematic with the use of barcodes for labelling, as it becomes increasingly unlikely that 2 or 3 peaks on an event would be due to multiple knots in the DNA. About 28% of events are identified to have 2 peaks in the sample labelled with 2 probes as compared to about 12% in the control measurement; this difference increases further for the sample with three probes, with about 32% of the labelled sample having 3 peaks as compared to about 0.5% of events in the control having three peaks. Thus, through the use of the double and triple probe barcodes, ionic current signatures that are clearly distinct from patterns seen in the control can be created allowing identification of the target DNA.

The identification of the double and triple dCas9 barcodes can be further enhanced by considering not just the number of peaks on any given event, but also whether the peaks occur in a closely spaced cluster. By comparing the difference between the peak positions of consecutive peaks on an event to a cutoff threshold, clustered double and triple peaks corresponding to the barcodes can be identified as demonstrated in Figure 3 c). The first two peaks seen in the event are from bound probes that are designed to be clustered close together (Δ_1 <threshold), whereas the last peak is spaced much further away from the first two peaks (Δ_2 >threshold) and is not part of the designed barcode. Despite having 3 peaks in total, the event is correctly identified as being labelled with the clustered double barcode. The details of the clustering analysis and threshold are discussed further in Section S6 in the SI. For the data sets where the DNA samples are labelled with two (blue) or three probes (red) shown in Figure 3 b), the events are reanalyzed using the clustering analysis to assess what percentage of events have clustered double or triple peaks. Figure 3 d) shows that about 25% of the events are labelled with the correct double or triple clustered barcode. This is at least an order of magnitude higher than the percentage of events in the control

identified with the double or triple barcode when analyzed using the same clustering criteria. In particular, the percentage of events identified with the distinct triple probe barcode is about two orders of magnitude higher than in the control sample where only a single false positive event was found in over 800 events.

Identifying Target DNA Using dCas9 Barcodes in a DNA Mixture

The clustered double and triple barcodes create distinctive current traces when patterned on the full length lambda DNA, however, ultimately the aim is to use these barcodes to identify different target DNA sequences in a background. To create a well defined background of DNA of homogeneous strand length, lambda DNA and T7 phage DNA are individually sheared into shorter fragments. The long DNA is centrifuged to push it through a precise orifice creating shearing forces that cut the DNA into many fragments, with the fragment size roughly controlled by the centrifuge RPM (g-TUBE, Covaris). The shearing also serves to reduce any problems with nanopore clogging and knots that can occur more frequently with longer DNA. In the sheared sample of lambda DNA, a triple barcode is targeted to one fragment and in the sheared sample of T7 phage DNA, a double barcode is targeted to one fragment as is described in Figure 4. Thus, each barcode is used to identify a sheared target DNA fragment in a background of itself. For example, for full length 48 kbp lambda DNA sheared to a median of 6 kbp we expect approximately 1 in 8 fragments to contain the triple cluster of bound probes.

The triple dCas9 probe barcode binds to one fragment of the lambda DNA with the other 7 sheared fragments remaining unlabelled. For perfectly labelled samples, 1 in 8 or roughly 12% of events are tagged, assuming that the DNA shears into consistent fragments. Our data yields a percentage of events with a clustered triple barcode at approximately 2.3%, as shown in Figure 4. This indicates 2.3% instead of an expected 12% of events are correctly labelled with the triple barcode which gives a labelling efficiency of approximately 2.3%/12% = 19%. This is slightly smaller albeit comparable to the triple barcode labelling efficiency
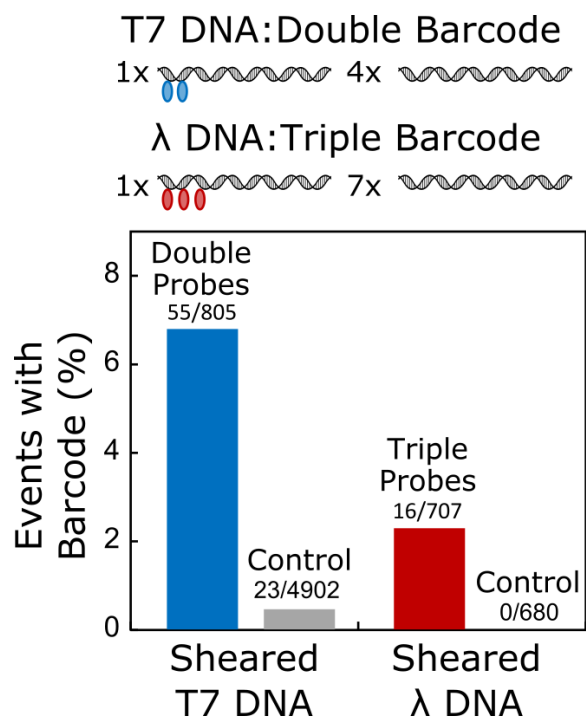
Figure 4: Detecting one fragment of sheared DNA in a background of unlabelled fragments using double and triple barcodes. T7 DNA is sheared into roughly 5 fragments with a target size of 9 kbp. Double dCas9 probes target one of the fragments, labelling it with a double peak barcode. Lambda DNA is sheared into roughly 8 fragments with a target size of 6 kbp. Triple dCas9 probes target one of the fragments, labelling it with a triple peak barcode. Percentage of events with clustered peaks in a double barcode for the T7 DNA or a triple barcode for the lambda DNA. A control sample of sheared T7 DNA or sheared lambda DNA is analysed for clustered peaks using the same clustering threshold as the respective barcoded sample. The data labels give the number of events identified with the correct barcode out of the total number of events.

of 23% seen in Figure 3 d) when bound to the full length lambda DNA. Similarly, as seen in Figure 4, the double probe on the sheared T7 phage DNA demonstrates 6.8% of events correctly labelled as compared to approximately 20% of events expected to be labelled, or a 6.8%/20% = 34% efficiency of labelling the desired T7 fragment with a clustered double barcode. Our distinct double and triple dCas9 barcodes allow the clear differentiation of a labelled DNA sample from background DNA fragments.

The above results demonstrate the use of double and triple barcodes to identify a sheared target DNA fragment in a background of itself. However, the aim of the barcodes is to label a target sequence motif in a mixture of many unknown sources of DNA, paving the way to diagnostic applications. Towards this end, Figure 5 shows a proof of concept, individually using the double and triple barcodes targeted to the T7 and lambda DNA sequences to label their target sequence in a background of sheared *E. coli* DNA. The sequence rich *E. coli* DNA background provides a challenge to the specificity of the probes to the exact target sequences of interest.

The T7, lambda, and *E. coli* DNA are all sheared to an approximate fragment length of 9 kbp. This length was chosen based on the commercially recommended shearing protocols (g-TUBE, Covaris). A mixture of the sheared DNA is made that includes 0.2 nM of the T7 DNA target fragment, 0.2 nM of the lambda DNA target fragment and *E. coli* DNA fragments to a DNA concentration of 4 nM. This mixture was then diluted into 2 M LiCl to a final concentration of 1 nM total DNA for nanopore measurements. A double barcode targeting the T7 DNA is used along with a triple barcode targeting the lambda DNA, with the target fragments of both lambda and T7 present at an equimolar concentration that corresponds to a ratio of about 1 of each target fragment in 20 fragments (18 unlabelled background fragments). In the first instance, the double and triple dCas9 barcode probes are each tested independently in the full mixture of DNA.

When only the double probes targeting the T7 fragment are added to the full mixture, around 2% of the events are identified with the clustered double barcode and almost no
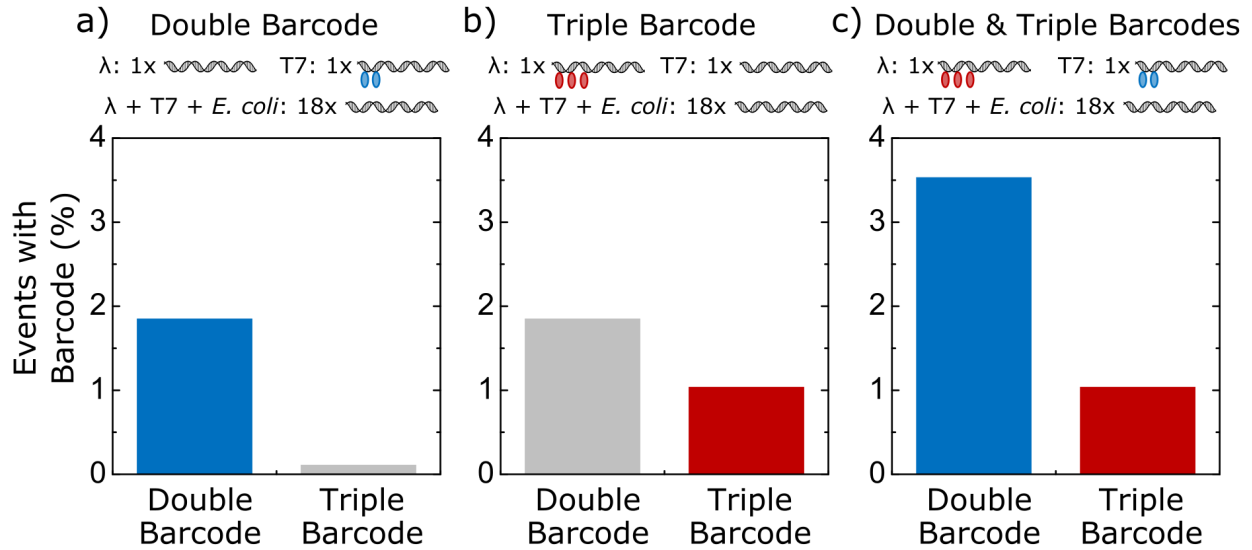
Figure 5: Identifying T7 and lambda DNA fragments using double and triple barcodes in a background of mixed sheared DNA from *E. coli*. a) A fragment of T7 DNA is labelled with a double probe barcode in the mixture. The percentage of events identified with the clustered double barcode (blue) is compared to the low incidence of a false positive of a clustered triple barcode (grey) being identified in the same sample. b) A fragment of lambda DNA is labelled with a triple probe barcode. The percentage of events with the clustered triple barcode (red) is compared to the high incidence of a false positive of a clustered double barcode (grey) being identified. If one of the probes does not bind correctly in the triple barcode, it is identified as a double barcode. c) The T7 fragment and lambda DNA fragment are labelled concurrently using both the double and triple barcodes, respectively. Using both the double and triple probe barcodes in the same mixture, both T7 and lambda DNA can be identified at the same time in a background of *E. coli* DNA.

false positive events are identified with a signature resembling the triple barcode, as seen in Figure 5 a). However, in the experiment with only the triple probes added targeting the lambda DNA, approximately 1% of the events are correctly identified with the clustered triple barcode whereas approaching 2% of the events are identified with a false positive signature resembling the double barcode, as seen in Figure 5 b). This crosstalk between the two barcode signatures arises due to the fact that any one of the 3 probes from the triple probe barcode may not bind correctly to the target. The potential for crosstalk between barcodes is an important consideration in multiplexed barcode design and can be avoided in the future by redesigning the spacing between the probes to make the barcodes distinct from each other. A control is also performed with the double probes in the mixed DNA without T7 and the triple probes in the mixture without lambda DNA which showed that the percentage of events identified with the correct barcode in the control mixture is over an order of magnitude less than when the target is present. These results demonstrate that when the target is not present there is a very low incidence of false positives despite the variety of genetic sequences in the mixture arising from the *E. coli* background.

Finally, the ability to use all 5 probes to simultaneously identify both target sequences from T7 and lambda DNA in the *E. coli* background is explored in the results shown in Figure 5 c). Using all five probes simultaneously yields comparable results to using only the double or the triple barcode independently. Approaching 4% of the events in the mixture with all 5 probes are identified as having a double barcode and about 1% of the events are identified as having the triple barcode. The expected ratio is the same as the independent barcode experiments, with 1 double and 1 triple barcode each expected to occur once in approximately 20 events. The percentage of events with the double barcode in the experiment with all 5 probes is higher than when using only the double probe due to the crosstalk with the triple probe sample as previously discussed. However the percentage of events labelled with the triple probe in the experiment with all 5 probes in the mixture almost exactly replicates the results of using the triple probe alone, showing that the triple probe binding in this sample

is reproducible.

This final experiment shows the use of two barcodes to successfully identify two different species of DNA concurrently in a background mixture of *E. coli* DNA. Importantly, while off target dCas9 binding can be problematic in certain assays, with the use of the clustering analysis to identify the distinct barcodes in our assay the false positive responses resulting from probes binding off-target in the control experiments is found to be negligible. A better understanding of binding efficiency combined with improvements to the design of the barcodes to reduce cross-talk between barcode signals are needed. However, the final experiment demonstrates the potential of a multiplexed dCas9 barcoding system to be used to simultaneously identify multiple target sequences in a complex mixture. Finally, the dCas9-nanopore assay has a relatively fast time-to-result, currently on the order of a couple of hours in a research setting. The current time-to-result is mainly limited by manual sample handling. It is important to note that the binding kinetics of dCas9 and measurements of thousands of events per nanopore should allow for results in under 30min. An optimized and automated dCas9-nanopore assay has the potential for a time-to-result on the order of tens of minutes but is already comparable to sequencing for pathogen identification[3] and is much faster than standard bacterial culture methods. Further improvements in time-to-result are possible by scaling the number of targets in the assay, increasing the number of nanopores measuring in parallel, and analysing events in real-time. The advantages of fast time-to-result and highly multiplexed detection would open the door to a range of clinical applications requiring rapid identification of many different DNA sequences in parallel.

Conclusions

Highly specific dCas9 probes are demonstrated to bind to target sequences on DNA and can be used to create unique barcodes that can be read out using nanopore sensors. Single dCas9 probes are targeted to various positions on DNA strands and, notably, can be designed to

remain tightly bound during relevant experimental times in high salt conditions typical of nanopore sensing. Multiple probes are bound to the same DNA strand in clusters of two or three probes to create current patterns that are clearly distinct from knots or folds in the DNA. These double and triple barcodes create distinct ionic current patterns, and can be used to rapidly identify a target DNA sequence in a background sample. Finally, the barcodes can be used simultaneously to identify two different DNA targets in a background mixture of bacterial DNA laying the groundwork for a fast and versatile assay for multiplexed DNA sensing applications in mixed samples. In an era of increasing antibiotic resistance, culture-free multiplexed identification of pathogens would be an invaluable tool to facilitate faster and more informed clinical decision making. Furthermore, identification of resistance genes via SNP dependent Cas9 probes would allow for more efficient targeted antibiotic prescription. Beyond this example, the sensing platform is promising for a wide range of future applications needing rapid and highly multiplexed DNA sensing in complex mixtures.

Methods

Labelling DNA with dCas9 Probes

The *S. pyogenes* double nuclease knockout mutant Cas9 D10A/H10A (dCas9) probes bind specifically to the target double stranded DNA via a sequence specific guide RNA (cr-RNA). The sequences for the crRNA for the probes were designed using online software (http://chopchop.cbu.uib.no/),[25,26] and can be found in Section S1 of the SI along with the targeted position along the DNA. The tracrRNA(200nM, IDT), crRNA (250nM, IDT Alt-R), and dCas9 (100nM) were incubated in low salt buffer (25 mM HEPES-NaOH (pH 8.0), 150 mM NaCl, 1 mM MgCl2) to assemble the probes before adding the target DNA. The target DNA was then added to an excess of probes at a typical ratio of 10-15 dCas9 probes per DNA and was incubated for at least 20 min at room temperature to bind the probes to their target DNA sequence. Probes were bound to a 3.6 kbp fragment of lambda DNA, full

length lambda DNA (Thermo Scientific, 48.502 kbp) and T7 phage DNA (Thistle Scientific, 39.937 kbp), as well as various mixtures of sheared lambda, T7, and *E. coli* DNA fragments at approximate lengths of 6 kbp-10 kbp (g-tubes, Covaris). The labelled DNA sample is stored in the low salt buffer until immediately before nanopore measurements when it is diluted to the desired concentration in the measurement electrolyte.

Nanopore Fabrication and Measurements

Conical nanopores with an average diameter of 17 +/- 5 nm were fabricated by pulling quartz capillaries (Sutter Instrument) using a laser-assisted puller.[15,21,28] The nanopores are embedded in a PDMS fluid cell with cis and trans reservoirs filled with 2 M LiCl, 1xTE (pH 8.0), unless otherwise specified. Current-voltage curves of each pore are measured in the electrolyte to allow for approximation of the pore size, further details of which are found in Section S2 of the SI. The DNA sample is then added to the cis reservoir and a bias voltage of 600 mV is applied to drive the DNA through the nanopore.

Current measurements were carried out using an Axopatch 200B patch-clamp amplifier (Molecular Devices, CA, USA) with an internal filter setting of 100 kHz. An external 8-pole analog low-pass Bessel filter with a cut-off frequency of 50 kHz (900CT, Frequency Devices, IL, USA or 3382, Krohn-Hite, MA, USA) further reduced the noise before recording with a data acquisition card (PCIe-6351, National Instruments, TX, USA) and a custom-written LabView program at 250 kHz.

Analysis of DNA translocations

Translocation events were identified as deviations from the baseline current greater than a set threshold. The DNA folding state and positions of secondary peaks are detected following a previously published procedure.[23] In order to filter out spurious peaks in the case of 2 or more dCas9 probes, the time differences between consecutive peaks are calculated. A threshold is then set at 1.5 times the standard deviation from the mean of the difference distribution.

Peaks that lie further than this threshold from their next neighbour are discarded. Since the dCas9 probes target clustered positions with similar distances between them, this approach allows the number of bound probes to be correctly counted despite the presence of spurious peaks caused by knots or small folds in the DNA.

Acknowledgement

Notes

The authors declare the following competing financial interest(s): The study was partly supported by Oxford Nanopore Technologies through funding of the postdoc position of Dr. Nicole Weckman.

Supporting Information Available

Supporting Information is available including details on dCas9 probes, nanopore characterization, analysis methods, and binding efficiencies.

References

(1) Wang, H.; Ettedgui, J.; Forstater, J.; Robertson, J. W.; Reiner, J. E.; Zhang, H.; Chen, S.; Kasianowicz, J. J. Determining the physical properties of molecules with nanometer-scale pores. *ACS Sensors* **2018**, *3*, 251–263, DOI: `10.1021/acssensors.7b00680`.

(2) Bayley, H. Nanopore sequencing: From imagination to reality. *Clinical Chemistry* **2015**, *61*, 25–31, DOI: `10.1373/clinchem.2014.223016`.

(3) Schmidt, K.; Mwaigwisya, S.; Crossman, L. C.; Doumith, M.; Munroe, D.; Pires, C.; M. Khan, A.; Woodford, N.; Saunders, N. J.; Wain, J.; O'Grady, J.; Livermore, D. M. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *Journal of Antimicrobial Chemotherapy* **2017**, *72*, 104–114, DOI: `10.1093/jac/dkw397`.

(4) Dekker, C. Solid-state nanopores. *Nature Nanotechnology* **2007**, *2*, 209–215, DOI: `10.1038/nnano.2007.27`.

(5) Waduge, P.; Hu, R.; Bandarkar, P.; Yamazaki, H.; Cressiot, B.; Zhao, Q.; Whitford, P. C.; Wanunu, M. Nanopore-based measurements of protein size, fluctuations, and conformational changes. *ACS Nano* **2017**, *11*, 5706–5716, DOI: `10.1021/acsnano.7b01212`.

(6) Albrecht, T. Single-molecule analysis with solid-state nanopores. *Annual Review of Analytical Chemistry* **2019**, *12*, DOI: `10.1146/annurev-anchem-061417-125903`.

(7) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012**, *482*, 331–338, DOI: `10.1038/nature10886`.

(8) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.;

Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819–823, DOI: `10.1126/science.1231143`.

(9) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152*, 1173–1183, DOI: `10.1016/j.cell.2013.02.022`.

(10) Yang, W.; Restrepo-Pérez, L.; Bengtson, M.; Heerema, S. J.; Birnie, A.; Van Der Torre, J.; Dekker, C. Detection of CRISPR-dCas9 on DNA with solid-state nanopores. *Nano Letters* **2018**, *18*, 6469–6474, DOI: `10.1021/acs.nanolett.8b02968`.

(11) Bulushev, R. D.; Marion, S.; Petrova, E.; Davis, S. J.; Maerkl, S. J.; Radenovic, A. Single molecule localization and discrimination of DNAprotein complexes by controlled translocation through nanocapillaries. *Nano Letters* **2016**, *16*, 7882–7890, DOI: `10.1021/acs.nanolett.6b04165`.

(12) Smeets, R. M. M.; Kowalczyk, S. W.; Hall, A. R.; Dekker, N. H.; Dekker, C. Translocation of RecA-coated double-stranded DNA through solid-state nanopores. *Nano Letters* **2009**, *9*, 3089–3095, DOI: `10.1021/nl803189k`.

(13) Singer, A.; Wanunu, M.; Morrison, W.; Kuhn, H.; Frank-Kamenetskii, M.; Meller, A. Nanopore based sequence specific detection of duplex DNA for genomic profiling. *Nano Letters* **2010**, *10*, 738–742, DOI: `10.1021/nl100058y`.

(14) Singer, A.; Rapireddy, S.; Ly, D. H.; Meller, A. Electronic barcoding of a viral gene at the single-molecule level. *Nano Letters* **2012**, *12*, 1722–1728, DOI: `10.1021/nl300372a`.

(15) 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 Letters* **2017**, *17*, 5199–5205, DOI: `10.1021/acs.nanolett.7b01009`.

(16) Wanunu, M.; Sutin, J.; Meller, A. DNA profiling using solid-state nanopores: detection of DNA-binding molecules. *Nano Letters* **2009**, *9*, 3498–3502, DOI: `10.1021/nl901691v`.

(17) Plesa, C.; Ruitenberg, J. W.; Witteveen, M. J.; Dekker, C. Detection of individual proteins bound along DNA using solid-state nanopores. *Nano Letters* **2015**, *15*, 3153–3158, DOI: `10.1021/acs.nanolett.5b00249`.

(18) Yu, J.-S.; Lim, M.-C.; Huynh, D. T. N.; Kim, H.-J.; Kim, H.-M.; Kim, Y.-R.; Kim, K.-B. Identifying the location of a single protein along the DNA strand using solid-state nanopores. *ACS Nano* **2015**, *9*, 5289–5298, DOI: `10.1021/acsnano.5b00784`.

(19) Carlsen, A. T.; Zahid, O. K.; Ruzicka, J. A.; Taylor, E. W.; Hall, A. R. Selective detection and quantification of modified DNA with solid-state nanopores. *Nano Letters* **2014**, *14*, 5488–5492, DOI: `10.1021/nl501340d`.

(20) 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. *Nature Communications* **2017**, *8*, 1552, DOI: `10.1038/s41467-017-01584-3`.

(21) Bell, N. A. W.; Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nature Nanotechnology* **2016**, *11*, 645–651, DOI: `10.1038/nnano.2016.50`.

(22) Beamish, E.; Tabard-Cossa, V.; Godin, M. Identifying structure in short DNA scaffolds using solid-state nanopores. *ACS Sensors* **2017**, *2*, 1814–1820, DOI: `10.1021/acssensors.7b00628`.

(23) Ermann, N.; Hanikel, N.; Wang, V.; Chen, K.; Weckman, N. E.; Keyser, U. F. Promoting single-file DNA translocations through nanopores using electro-osmotic flow. *The Journal of Chemical Physics* **2018**, *149*, 163311, DOI: `10.1063/1.5031010`.

(24) Kowalczyk, S. W.; Wells, D. B.; Aksimentiev, A.; Dekker, C. Slowing down DNA translocation through a nanopore in Lithium Chloride. *Nano Letters* **2012**, *12*, 1038–1044, DOI: `10.1021/nl204273h`.

(25) Montague, T. G.; Cruz, J. M.; Gagnon, J. A.; Church, G. M.; Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research* **2014**, *42*, W401–W407, DOI: `10.1093/nar/gku410`.

(26) Labun, K.; Montague, T. G.; Gagnon, J. A.; Thyme, S. B.; Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research* **2016**, *44*, W272–W276, DOI: `10.1093/nar/gkw398`.

(27) Plesa, C.; Verschueren, D.; Pud, S.; van der Torre, J.; Ruitenberg, J. W.; Witteveen, M. J.; Jonsson, M. P.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Direct observation of DNA knots using a solid-state nanopore. *Nature Nanotechnology* **2016**, *11*, 1093–1097, DOI: `10.1038/nnano.2016.153`.

(28) Bell, N. A. W.; Thacker, V. V.; Hernández-Ainsa, S.; Fuentes-Perez, M. E.; Moreno-Herrero, F.; Liedl, T.; Keyser, U. F. Multiplexed ionic current sensing with glass nanopores. *Lab on a Chip* **2013**, *13*, 1859–1862, DOI: `10.1039/c3lc50069a`.

Graphical TOC Entry