

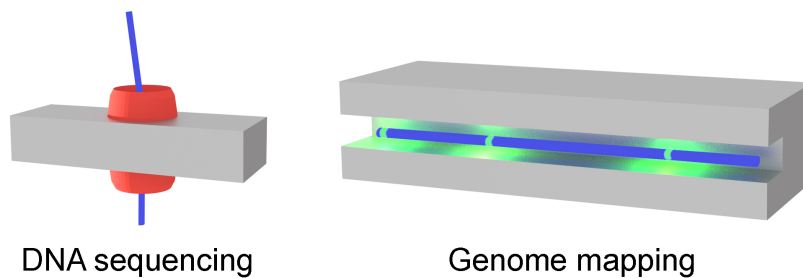
## Nanopores and Nanochannels: From Gene Sequencing to Genome Mapping

Stefan Howorka<sup>1</sup>, Zuzanna Siwy<sup>2</sup>

<sup>1</sup> Department of Chemistry, Institute of Structural Molecular Biology, University College London, London, WC1H 0AJ, UK.

<sup>2</sup> Department of Physics and Astronomy, University of California at Irvine, 210G Rowland Hall, Irvine, CA 92697, USA.

ToC graphic



DNA strands can be analyzed at the single-molecule level by passing them through nanoscale holes. The strategy is used for the label-free and portable sequencing. Nanochannels can also be applied to map genomes with high resolution, as shown by Jeff et al. in this issue of ACS Nano. Here we compare the two strategies in terms of biophysical similarities and differences, and describe that both are complementary and can improve the DNA analysis for genomic research and diagnostics.

Nanopores are nanoscale holes formed inside biological or synthetic membranes and of considerable biotechnological interests as they provide powerful new ways to analyze DNA at the single-molecule level. The most prominent example is nanopore sequencing.<sup>1</sup> Therein, individual DNA strands are threaded through arrays of 10-nm sized nanopores.<sup>2</sup> This label-free sequencing principle has been commercially launched as a \$1000-priced palm-sized device that is expected to help democratize sequencing.<sup>3</sup> Yet, nanopores enable another important way to analyze DNA. As highlighted by Jeff et al. on page XXX of this issue of ACS Nano,<sup>4</sup> genomes can also be mapped. In particular, chromosomal DNA strands are placed inside narrow nanochannels of microscale length to check for genetic rearrangements such as insertions and deletions. This Perspective describes the principle, advantages and challenges of genome mapping with nanochannels, and relates it to DNA sequencing. As singulating DNA strands with nanopores was pioneered with sequencing, the Perspective first introduces this approach and then covers mapping to delineate how both routes differ in terms of how and which genetic information is accessed.

The characteristic feature of nanopores and nanochannels is their narrow lumen of several nanometers in diameter. Biological protein pores used in sequencing have a limiting diameter of around 1 to 1.5 nm. As protein pores are inserted inside a lipid bilayer, they form water-filled conduits for transport of molecular cargo across the membrane. Reflecting their small lumen, only molecules below the threshold size can pass.

Size-selection by protein nanopores is exploited in sequencing. Only one single stranded DNA molecule with a width of around 1 nm can pass the protein hole of matching width at a time (Figure 1).<sup>5</sup> By contrast, the pore is too narrow for two strands or a DNA duplex. Analyzing individual DNA strands has practical benefits. The main advantage is that amplification of DNA samples is not necessary. Enzymatically copying DNA strands is, however, required in several classical sequencing approaches in order to meet the threshold of detecting the bases. Amplification usually translates into higher costs, longer handling time, and potential bias or errors in the final DNA composition. A good route to avoid amplification is the use of methods with single-molecule sensitivity, such as nanopore sensing.

To obtain sequence information with nanopores, individual strands are moved along an internal reading head (Figure 1). The reading head is a narrow and thin molecular constriction of around 1.2 nm composed of several amino acid residues.<sup>5</sup> When the DNA strand passes this constriction, bases can be identified by monitoring the ionic current flowing through the channel. The read-out signal shows characteristic discrete stepwise current levels that reflect the differently sized bases at the reading head and how they interact with this narrow constriction. The signals are stepped because DNA translocates through the pore aided by an attached motor protein that moves the strand base-by-base. The direct electrical detection of bases with nanopores is an advantage compared to other indirect label-based sequencing approaches because no extra handling step for fluorophore tagging is necessary. Another important feature of nanopore sequencing is that it can cover read lengths of up to 200,000 base pairs (bp) which compares to a range of up to 1000 bp of existing ensemble sequencing methods. A challenge is, however, that strings of identical bases with different repeat numbers cannot be well discriminated.

Jeffet *et al.*'s genome mapping approach analyzes much longer chromosomal DNA in the range of million bp length<sup>4</sup>. Genome mapping aims to identify the location of genes on chromosomes and their relative distance. It was developed prior to the use of nanopores to assemble the shorter DNA sequencing reads into larger genetic maps. Detailed sequence information is, however, not a requirement to obtain genetic maps. Rather, it is often sufficient to obtain patterns of genetic markers to yield a genetic fingerprint or barcodes that are characteristic of the DNA region of interest.

One simple route to obtain maps with genetic fingerprints is to elongate individual DNA strands, stain them, and detect genetic marks along the sequence via fluorescence microscopy. This optical genomic mapping approach has been pioneered by stretching DNA strands on the surface of a glass slide.<sup>6-7</sup> However, the method can suffer from nonuniform DNA stretching and hence imprecise DNA length measurement, and low through-put. Alternatively, DNA can also be stretched in a microfluidic flow<sup>8</sup> but this does not allow simple microscopy analysis.

Genome mapping with nanopores, by comparison, can be of higher throughput by elongating individual chromosomal fragments inside arrays of nanochannels.<sup>9-10</sup> The nanochannels for the genome mapping are hence long to accommodate the million bp-long double stranded molecules. These channels can be fabricated via lithography. For example, the silicon dioxide channels with a transparent glass top as used by Jeffet *et al.*<sup>4</sup> and others<sup>10</sup> feature a 45 nm opening and a length of 400  $\mu\text{m}$  (Figure 2A), and are available from a commercial vendor.<sup>11</sup> The long nanoscopic structures not only singulate individual DNA molecules but also provide nanoscale confinement to unfold them along their entire length (Figure 2A) as required in genomic mapping, thereby exploiting the unique physics in nanochannels<sup>12-13</sup>. Genome mapping hence shares with nanopores sequencing

the same principle of singulating DNA molecules via size-exclusion. However, genome mapping is conceptually different to sequencing as the former relies on analyzing quasi-static strands while the latter obtains genetic information by directionally threading DNA through a much shorter pore (Figure 1, Figure 2A).

To acquire genetic fingerprints in nanochannel-based mapping, the linearized DNA molecules are stained with fluorescent probes and imaged. The resulting barcode-like fluorescence patterns are characteristic for the DNA sequences but can also reveal genomic rearrangements such as insertions or deletions. The genetic maps obtained via nanochannels have a resolution of around 1500 bp (1.5 kbp).<sup>10</sup> This is a significant improvement over other optical mapping methods that extend DNA on a glass, or the stretch it in a fluid flow. Nevertheless, improving the resolution further is very important to achieve higher mapping accuracy.

The work by Jeffet *et al.* makes an important step in this direction as it tackles the issue of insufficient resolution via a strategy that achieves a 15-fold improvement. The strategy addresses the main culprit of low resolution, namely thermal fluctuations of DNA molecules which is in turn a consequence of analyzing DNA strands in solution as opposed to surface-bound DNA. As the configuration and location of the DNA in the channel is transient, the authors show that a given microscopic snapshot of molecules does not represent the equilibrium position, which contributes to the uncertainty of finding label position.

To mitigate thermal fluctuations Jeffet *et al.* firstly employ single-fluorophore imaging of labeled DNA. Single-fluorophore imaging is widely used in the life-sciences<sup>14</sup> but Jeffet *et al.* are the first to successfully apply it for genomic mapping in nanochannels. The approach was demonstrated with a naturally occurring region in human chromosome 4 that contains multiple repeats of a 3.3 kbp length. Each repeat was labeled with a custom-made fluorophore of exceptional photochemical stability, which allowed recording the thermally fluctuating positions of the dye for more than 100 times at 40 ms intervals (Figure 2B). Averaging the label positions over 50 frames increased the localization accuracy to ~300 bp. This is a 5-fold improvement over the 1.5 kbp limit of the conventional approach where multiple clustered fluorophores of lower photostability are tracked without knowing which individual fluorophore is still active or already bleached.<sup>10</sup>

The accuracy of localizing the single labels was further increased by implementing a new approach to averaging the fluorophore positions obtained from multi-frame imaging. In order to filter out collective fluctuations of the whole molecule, the authors calculated the relative distances between the fluorescent spots in each captured frame, rather than the absolute positions. The resulting location maps were therefore based on mean distances which reduced, to a large extent, the influence of collective movements of a large portion of the DNA molecule.

Recording a series of DNA positions also allowed the authors to quantify inaccuracies associated with single-snapshot mapping. By calculating distances between labels in subsequent frames, they found a fraction of fluctuations that reached even 1.2 kbp deviation from the 3.3 kbp equilibrium distance. On average, the maximum displacement from the equilibrium distance was  $\pm 0.7$  kbp.

Jeffet *et al.*'s final step towards higher accuracy was to exploit the highly photostable fluorophores in combination with less stable, blinking dyes. The multiple fluorophores could be resolved even though the distance between the dyes was smaller than the diffraction limit. The spot of the photostable dye was used as a reference point informing on local DNA fluctuations. The reference dye was separated by 3.3 kbp to two other target dyes that were distant from each other by 676 bp. Blinking or bleaching of the two dyes produced a clear change in the recorded distance between the reference and target fluorophores, measured as  $670 \pm 130$  bp (Figure 2C,D). It is predicted that distances as small as 100 bp could be resolved by this approach. By harnessing the instability of fluorophores with very stable reference dyes, the authors achieve the 15-fold enhancement in resolution compared to the conventional route of analysis. Jeffet *et al.* name their combined approach super-resolution mapping in analogy to super-resolution microscopy. In the latter case, the resolution is, however, in the lower Angstrom range as the molecules are immobilized or more tractable positions than the fluorophores in the thermally fluctuating DNA strands within nanochannels.<sup>14</sup>

What are the directions of future research? One of the first aims can be to apply the high-resolution method developed by Jeffet *et al.* to DNA samples where the gain in mapping accuracy translates into better genetic analysis and a benefit in diagnostics or related applications. Specific applications could be the major histocompatibility region, which is important in infectious and autoimmune diseases.

A valuable target for the technological development of genomic mapping could be to improve the resolution by tuning the channel properties. As discussed by Jeffet *et al.*, a further reduction of thermal fluctuations could be achieved by using nanopores whose opening is significantly smaller than the DNA persistence length ( $\sim 50$  nm). The channels that were used in the experiments were 45 nm in width. Fabrication of narrower channels should be accessible by lithography approaches but the threading of the DNA into those tiny channels might be slow.

It would also be exciting to tune the chemical properties of the channel walls, which up to now have not been yet optimized for genetic mapping. In particular, negative charges could be added to the wall to force DNA into the center of the nanochannel by electrostatic repulsion that re-enforces steric confinement. Experiments could also be performed at solutions of higher viscosity or lower

temperature; both factors would reduce thermal fluctuations of DNA molecules. It is also known that fluctuations of DNA molecules becomes more significant for larger distances between labels. Hence, different sets of conditions might need to be used to achieve the same resolution with different samples. Better DNA stains are another route to improved resolution. Currently, intercalator dyes are used in addition to the site-specific fluorophores label in order to track the DNA strands in the channel before single-fluorophore analysis. But the intercalator dyes have different labeling densities along single DNA strands or different DNA molecules which leads to a broader spread of actual DNA distances between the important single-fluorophore markers. This heterogeneity could be addressed by minor groove binding dyes that do not change the length of the DNA duplex as much.

Anchoring the DNA strands at one end of a nanochannel could also diminish the effects of thermal motion on the analysis. Applying voltage across the channel is expected to extend the molecule either further or might facilitate the label localization. The disadvantage of this approach is however increased complexity of the device.

It would be of great advantage to simplify the design of the current analysis system to make it fully portable. Hand-held sequencing devices allow performing DNA analysis at remote locations and independent of a centralized facility.<sup>2</sup> Portable DNA sequencing and mapping have hence the potential to greatly facilitate personalized medicine as well as agriculture and homeland security. A wider adoption of nanopore and nanochannel-based devices would also blur the historic distinction between sequencing and mapping; the very long sequencing reads from nanopores help map larger genomic regions while higher-resolution genomic mapping could provide sufficient detailed sequence information required for some diagnostic applications. These exciting prospects will undoubtedly fuel more research into the fundamental science and applications of nanopore and nanochannel-based DNA analysis.

#### Acknowledgements:

S. H. is financially supported by UK EPSRC grant EP/N009282/1, BBSRC grants BB/M025373/1 and BB/N017331/1, and Oxford Nanopore Technologies. Z.S. acknowledges funding by the NSF (CHE 1306058). We thank Dr. Jonathan R. Burns for creating the ToC graphic.

#### References:

1. Wanunu, M., Nanopores: A Journey Towards DNA Sequencing. *Phys. Life Rev.* **2012**, 9, 125-158.
2. Quick, J.; Loman, N. J.; Duraffour, S.; Simpson, J. T.; Severi, E.; Cowley, L.; Bore, J. A.; Koundouno, R.; Dudas, G.; Mikhail, A.; Ouedraogo, N.; Afrough, B.; Bah, A.; Baum, J. H.; Becker-Ziaja, B.; Boettcher, J. P.; Cabeza-Cabrerizo, M.; Camino-Sanchez, A.; Carter, L. L.; Doerrbecker, J.; Enkirch, T.; Garcia-Dorival, I.; Hetzelt, N.; Hinzmann, J.; Holm, T.; Kafetzopoulou, L. E.; Koropogui,

M.; Kosgey, A.; Kuisma, E.; Logue, C. H.; Mazzarelli, A.; Meisel, S.; Mertens, M.; Michel, J.; Ngabo, D.; Nitzsche, K.; Pallasch, E.; Patrono, L. V.; Portmann, J.; Repits, J. G.; Rickett, N. Y.; Sachse, A.; Singethan, K.; Vitoriano, I.; Yemanaberhan, R. L.; Zekeng, E. G.; Racine, T.; Bello, A.; Sall, A. A.; Faye, O.; Faye, O.; Magassouba, N.; Williams, C. V.; Amburgey, V.; Winona, L.; Davis, E.; Gerlach, J.; Washington, F.; Monteil, V.; Jourdain, M.; Bererd, M.; Camara, A.; Somlare, H.; Camara, A.; Gerard, M.; Bado, G.; Baillet, B.; Delaune, D.; Nebie, K. Y.; Diarra, A.; Savane, Y.; Pallawo, R. B.; Gutierrez, G. J.; Milhano, N.; Roger, I.; Williams, C. J.; Yattara, F.; Lewandowski, K.; Taylor, J.; Rachwal, P.; Turner, D. J.; Pollakis, G.; Hiscox, J. A.; Matthews, D. A.; O'Shea, M. K.; Johnston, A. M.; Wilson, D.; Hutley, E.; Smit, E.; Di Caro, A.; Wolfel, R.; Stoecker, K.; Fleischmann, E.; Gabriel, M.; Weller, S. A.; Koivogui, L.; Diallo, B.; Keita, S.; Rambaut, A.; Formenty, P.; Gunther, S.; Carroll, M. W., Real-Time, Portable Genome Sequencing for Ebola Surveillance. *Nature* **2016**, *530*, 228-232.

3. <http://www.nanoporetech.com>.

4. Jeffet, J.; Kobo, A.; Su, T.; Grunwald, A.; Green, O.; Nilsson, A.; Eisenberg, E.; Ambjornsson, T.; Westerlund, E.; Shabat, D.; Purohit, P.; Ebenstein, Y., Super-Resolution Genome Mapping in Silicon Nanochannels. *ACS Nano* **2016**.

5. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M., Automated Forward and Reverse Ratcheting of DNA in a Nanopore at 5-Angstrom Precision. *Nat. Biotechnol.* **2012**, *30*, 344-348.

6. Michalet, X.; Ekong, R.; Fougerousse, F.; Rousseaux, S.; Schurra, C.; Hornigold, N.; van Slegtenhorst, M.; Wolfe, J.; Povey, S.; Beckmann, J. S.; Bensimon, A., Dynamic Molecular Combing: Stretching the Whole Human Genome for High-Resolution Studies. *Science* **1997**, *277*, 1518-1523.

7. Jing, J.; Reed, J.; Huang, J.; Hu, X.; Clarke, V.; Edington, J.; Housman, D.; Anantharaman, T. S.; Huff, E. J.; Mishra, B.; Porter, B.; Shenker, A.; Wolfson, E.; Hiort, C.; Kantor, R.; Aston, C.; Schwartz, D. C., Automated High Resolution Optical Mapping Using Arrayed, Fluid-Fixed DNA Molecules. *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 8046-8051.

8. Chan, E. Y.; Goncalves, N. M.; Haeusler, R. A.; Hatch, A. J.; Larson, J. W.; Maletta, A. M.; Yantz, G. R.; Carstea, E. D.; Fuchs, M.; Wong, G. G.; Gullans, S. R.; Gilmanshin, R., DNA Mapping Using Microfluidic Stretching and Single-Molecule Detection of Fluorescent Site-Specific Tags. *Genome Res.* **2004**, *14*, 1137-1146.

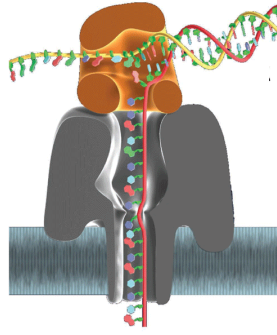
9. Jo, K.; Dhingra, D. M.; Odijk, T.; de Pablo, J. J.; Graham, M. D.; Runnheim, R.; Forrest, D.; Schwartz, D. C., A Single-Molecule Barcoding System Using Nanoslits for DNA Analysis. *Proc. Natl. Acad. Sci. U S A* **2007**, *104*, 2673-2678.

10. 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.

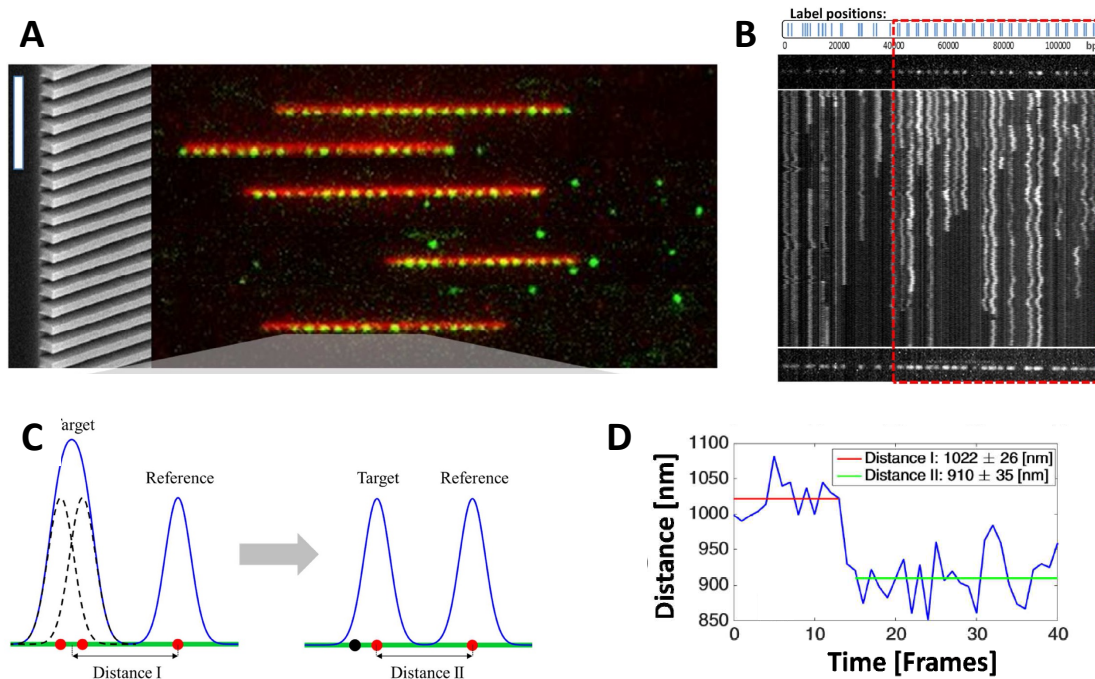
11. <http://www.bionanogenomics.com/>.

12. Schoch, R. B.; Han, J. Y.; Renaud, P., Transport Phenomena in Nanofluidics. *Reviews of Modern Physics* **2008**, *80*, 839-883.
13. Odijk, T., On the Statistics and Dynamics of Confined or Entangled Stiff Polymers. *Macromolecules* **1983**, *16*, 1340-1344.
14. Sauer, M., Localization Microscopy Coming of Age: From Concepts to Biological Impact. *J. Cell Sci.* **2013**, *126*, 3505-3513.





**Figure 1.** DNA sequencing with nanopores. A protein pore is equipped with an enzyme motor, which threads the molecule base-by-base. The readout is based on transmembrane current that is modulated by the presence of one or only few bases in the narrowest part of the pore. Reproduced from ref.<sup>5</sup>



**Figure 2.** Genome mapping with nanochannels. (A) Genomic mapping utilizes nanopores to singulate and extend single DNA molecules labeled with fluorescent dyes. The molecule is analyzed by localizing fluorescent spots informative on genomic repeats and deletions. (B) Accuracy of spots localization in genomic mapping is increased by recording a long series of subsequent DNA positions that result from thermal fluctuations. (C, E) Two fluorescent dyes separated by a distance below diffraction limit can be resolved using a combination of stable and blinking fluorophores. Blinking of the dyes resulted in clear change of the recorded distance from the stable fluorophore. Reproduced from ref.<sup>4</sup>