

# Single-Molecule DNA-Mapping and Whole-Genome Sequencing of Individual Cells

Rodolphe Marie<sup>a,1</sup>, Jonas N. Pedersen<sup>a</sup>, Loic Baerlocher<sup>b</sup>, Kamila Koprowska<sup>c</sup>, Marie Pødenphant<sup>a</sup>, Celine Sabatel<sup>d</sup>, Maksim Zalkovskij<sup>e</sup>, Andrej Mironov<sup>e</sup>, Brian Billenberg<sup>e</sup>, Neil Ashley<sup>c</sup>, Henrik Flyvbjerg<sup>a</sup>, Walter Bodmer<sup>c</sup>, Anders Kristensen<sup>a</sup>, and Kalim U. Mir<sup>f</sup>

<sup>a</sup>Department of Micro- and Nanotechnology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark; <sup>b</sup>Fasteris, Ch. du Pont-du-Centenaire 109, CH-1228 Plan-les-Ouates, Switzerland; <sup>c</sup>Department of Oncology, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom; <sup>d</sup>Diagenode, Liege Science Park, Rue du Bois Saint-Jean 3, 4102 Liege, Belgium; <sup>e</sup>NIL Technology ApS, Diplomvej 381, DK-2800 Kongens Lyngby, Denmark; <sup>f</sup>XGenomes, Pagliuca Harvard Life Lab, 127 Western Ave, MA 02134, Boston, USA

This manuscript was compiled on March 15, 2018

To elucidate cellular diversity and clonal evolution in tissues and tumors, one must resolve genomic heterogeneity in single cells at all length scales. We have developed micro-/nanofluidic chips in mass-producible injection-molded thermoplastic comprising a common architecture for extracting DNA from individual cells and modules that either map genomic structure directly, or collect genomic DNA for sequencing. To facilitate mapping we developed a method for light-directed fragmentation of the DNA to  $\sim 1$  Mb lengths inside nuclei and then to extract and pattern the fragments for the Denaturation-Renaturation (D-R) optical mapping method (Marie *et al.* PNAS 2013). Processing of single cells from the LS147T colorectal cancer cell line showed that D-R mapping of single molecules can reveal structural variation in the genome of single cells. In the case of one megabase fragment, one half aligned well to chromosome 19 while the other half showed variable alignment, containing a 50 kb deletion and a region of complexity; paired-end single cell sequencing supported this finding. However, sequencing struggled to uncover a 20 kb gap compared to the reference, that was clearly seen in a megabase fragment that otherwise mapped well to the pericentromeric region of chromosome 4. The pericentromeric regions are complex and there is substantial sequence homology between different chromosomes over these regions, which makes mapping of sequence reads ambiguous. This suggests that D-R mapping of a single molecule can directly reveal characteristics of the single cell genome that are difficult to reconstruct from short-read sequencing data.

Nanofluidics | Single Cell | DNA

Sequencing of the genomes of individual isolated cells or nuclei has become well established (1, 2) and is used to resolve intercellular variations in a heterogeneous population of cells from tissues to tumours (3–5). However, (6) because of short read-lengths and variant detection based on alignment to a reference which discards a substantial proportion of reads, it is difficult to accurately delineate the long-range structure of the genome by sequencing alone. By contrast, single-molecule optical mapping (7–9) of DNA provides a sweeping view of megabase-lengths of the genome (10–13). Several schemes exist to locate sequence information along the length of stretched DNA, and some have been combined with super-resolution microscopy (14, 15). In early realizations of optical mapping, DNA molecules were stretched on a surface (7, 16). More recently, nanoconfinement has emerged as a powerful technique to stretch DNA molecules uniformly and in a controlled manner without attaching them to a surface (17). Das *et al.* were able to image sequence-specific fluorescent optical maps in nanochannels with a cross-section below the persis-

tence length of DNA (18). The combination of nanoconfinement with additional mechanisms such as entropic recoil (19–21), hydrodynamic drag (22, 23), mechanical confinement (24) or thermophoresis (25) can enhance the degree of stretching.

Although genomic DNA extracted from individual cells has been stretched in micro- (26) and nanofluidic (27) devices, single-molecule optical mapping of DNA from individual cells has not previously been demonstrated. The critical challenge is to couple single-cell genome extraction with a means of handling the DNA at the appropriate length-scales and performing the mapping reaction in an integrated device. Stained single genomic DNA molecules can become patterned according to their AT/GC content by partial denaturation and renaturation (D-R) (11). As the DNA double helix opens up, the intercalating dye leaves the AT-rich regions thus creating dark spots along the DNA strand (see SI Appendix and Fig. 1D). A high degree of stretching, typically above 90%, provided by a flow-stretch device allows imaging of D-R patterns at diffraction-limited resolution using epifluorescence microscopy. D-R mapping identifies structural variations relative to a human reference genome in single copies of the genome (11), and this makes it a unique candidate for optical mapping of single cells.

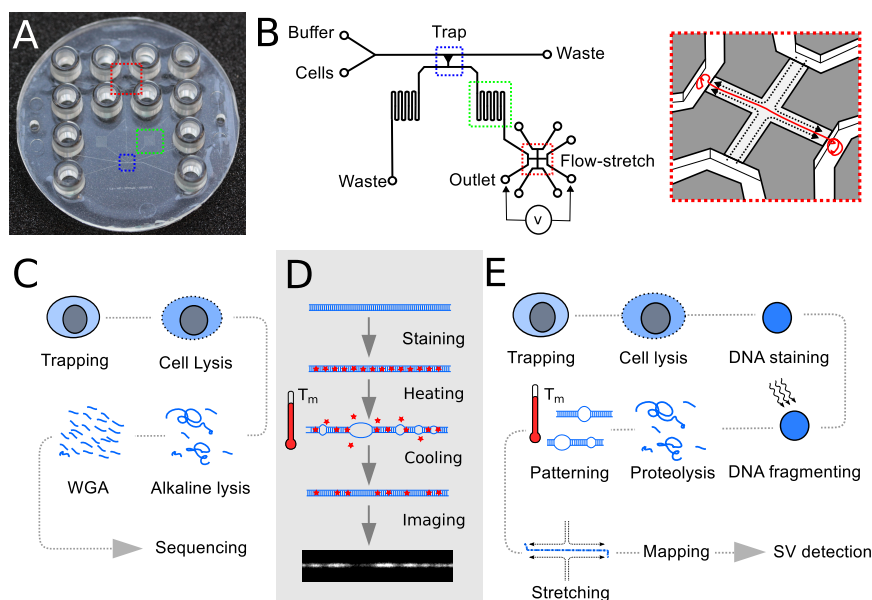
## Significance Statement

We report the first example of optical mapping of DNA from a single cell. We show isolation of single cells, DNA extraction and single molecule genome analysis within a single integrated micro-/nano-fluidic device. Moreover, we show that although the data quality is high when DNA extracted from single cells isolated in a similar microfluidic device is sequenced *off-chip*, it does not detect variation in complex genomic regions as easily as optical mapping conducted *on-chip*. This points to the relative merits of different approaches for genome analysis and may have implications for which genomic profiling approaches are used for the selection of cancer therapies especially as single cell analysis will allow genomic heterogeneity within a tumor to also be addressed.

R.M., A.K. and K.M. conceived the experiment. R.M., M.P., M.Z., A.M., B.B. and K.M. designed and fabricated the devices. R.M. carried out optical mapping; K.K., N.A. and C.S. carried out sequencing experiments. R.M., J.N.P., L.B., H.F., W.B. and K.M. carried out the data analysis; R.M., J.N.P. and K.M. wrote the first draft of the paper, all other authors contributed.

R.M., J.N.P., A.K. and K.M. filed a patent. K.M. declares that XGenomes is developing nucleic acid sequencing technologies. Other authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. E-mail: rodolphe.marie@nanotech.dtu.dk



**Fig. 1.** Architecture of single cell processing device and workflow for single cell D-R mapping and whole-genome sequencing. (A) An all-polymer lab-on-a-chip device with 12 connectors comprises a cell trap (blue), a meandering channel (green) and a flow-stretch device (red). (B) A single cell is captured by hydrodynamic trapping, DNA is extracted and patterned according to AT/GC composition by a heating-cooling cycle, and genomic DNA is stretched and visualized. (C) Workflow using device shown in SI Appendix, Fig. S1 for single-cell trapping and extracting and amplifying DNA before sequencing. (D) Principle of D-R pattern generation: the genomic DNA is homogeneously stained with YOYO-1. During partial denaturation at a temperature  $T_m$  between the melting temperatures of AT and GC bonds,  $T_m^{AT} < T_m < T_m^{GC}$ , the double helix opens up in AT-rich regions and dye leaves these regions. After renaturation, dark and light segments along the DNA constitute the D-R pattern used for mapping. The D-R map is compared with an *in silico* map of the reference genome (see SI Appendix). (E) Workflow using device in (A-B) to prepare genomic DNA for D-R mapping for structural variation detection.

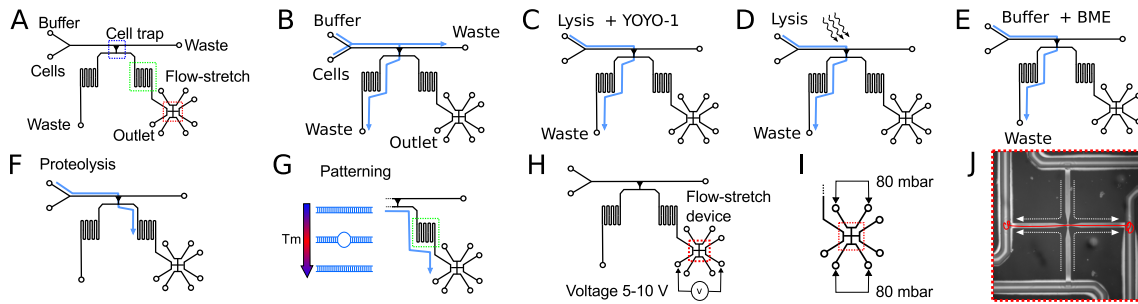
## Results

Using novel valve-less polymer devices that can be easily mass-produced by injection moulding (e.g. Fig. 1A), we have implemented a microfluidic architecture for trapping and isolating single cells from a population (Fig. 1B and SI Appendix, Fig. S1) and extracting their genomic DNA. We show that this trap design can be aligned within the device to an architecture for harvesting the genomic DNA from a single cell for sequencing outside of the device (Fig. 1C) or, alternatively, to an architecture for D-R mapping of the genomic DNA from a single cell within the device (Fig. 1D). Sequencing of DNA from single cells was enabled by the design shown in SI Appendix, Fig. S1. We fed a population of cells from the LS174T colorectal cancer cell line (SI Appendix, Fig. S2) into the polymer device with a flow design that steered cells towards the side of a microchannel having a series of funnel-shape traps. We stopped the feeding when each individual cell occupied a separate trap. The DNA was then extracted by alkaline lysis and collected in the outlet port for multiple displacement amplification (MDA) and Illumina sequencing. The sequence coverage from single cells was comparable to that obtained from a population of cells (SI Appendix, Fig. S3). As with commercial single cell processing devices (e.g. Fluidigm C1), using our device for isolating single cells, DNA extraction and DNA collection in the outlets gave superior whole-genome single-cell sequencing results compared to processing single cells without the use of the microfluidic environment: library preparation was successful for 97% of single cell samples; there was less than 2% contamination from non-human DNA; >87% of bases were represented at a sequencing depth as low as 10x; allelic dropout in the single cells across all high quality heterozygous variants found in bulk was as low as 30% at 15-20x coverage (28).

D-R mapping of DNA from single cells was enabled by the design shown in Fig. 1B and SI Appendix, Fig. S4. The operation of this device is depicted in Fig. 2. We fed cells into the polymer device wherein a single cell was captured in a trap. The DNA in the nucleus was then stained uniformly by adding the fluorescent intercalating dye YOYO-1

to the lysis buffer (SI Appendix, Figs. S5-6). A dose of light was used to nick the chromosomal DNA within the nucleus by YOYO-1-induced oxidative damage. Proteolysis was then used to extricate the chromosomal DNA from the nucleus and remove histone proteins (Fig. 3A). This resulted in a significant number of chromosome sub-fragments approximating megabase-lengths. These fragments are easier to manipulate in the microfluidic structure than chromosomal DNA that remains close to its native length. The relationship between dose of light used for photonic nicking and fragment-length is reciprocal because the relationship between dose and nicking probability per kb is linear (SI Appendix, Figs. S7-11). The nuclear lysate was passed through a meandering microchannel (Fig. 3B) in which the DNA was D-R patterned using a cycle of heating and cooling (Fig. 1D). The patterned molecules were then individually threaded into the cross-flow nanoslit using electrophoretic forces, stretched and imaged at 85 to 100 percent stretching (Fig. 3C), which was sufficient for them to be mapped to the reference genome hg18 (Fig. 3D-E). The cross-flow structure (Fig. 1B inset) is particularly appropriate for the purpose because a useable map can readily be obtained by imaging just one patterned molecule (11). This is important because single-cell mapping requires that the map is derived from a unitary molecule, as there are only two homologous but distinct copies of the genome in a normal diploid cell prior to replication.

We sampled 21 molecules of >1 Mb length from one cell and found that in 17 out of 21 cases, the molecule could be mapped to unique coordinates in the hg18 reference genome (see SI Appendix, Table S1). Mapping is achieved by comparing the experimental D-R map and the *in silico* map generated for the entire reference genome by calculating the sum of squared differences  $\chi^2$  (11). The certainty of the match arises from multiple independent barcodes mapping on the genome at predictable distances from each other (*i.e.* they fall on a diagonal in the plots displayed in SI Appendix Fig. S12-13). When mapped against the mouse genome (GRCm38/mm10), we found no such match for the experimental D-R maps. Figure 3D-E shows the D-R map derived from the image of one of the molecules extracted from the single cell and its



**Fig. 2.** Device operation to realise the workflow shown in Fig. 1D. (A) The device comprises a main channel with inlets for cells and buffer, a cell trap connecting the main channel to the outlets. On the outlet side, a flow stretch device is placed. (B) Cells are introduced and a single cell is captured in the trap. The flow is stopped when a cell occupies the trap. (C) Lysis buffer is introduced to remove the cytoplasm. The lysis solution contains YOYO-1 to stain the DNA of the nucleus. (D) The stained nucleus is illuminated with blue light (480 nm) to induce photo nicking. (E) Buffer with BME is introduced to prevent further photo nicking. (F) Same buffer to which Protease K is added is introduced to release the genomic DNA by proteolysis. (G) When the DNA reaches the meandering channel, temperature is raised to partially denature the DNA before the device is cooled again. (H) DNA molecules are introduced to the flow-stretch device by electrophoresis, and (I) stretched by a flow of buffer at a pressure of 80 mbar. (J) Bright field image of the flow-stretch device indicating the buffer flow (dotted line) and the position of the DNA (red line). The nanoslit is 450  $\mu\text{m}$  across.

matching to the *in silico*-generated melting map of a reference genome. Two of the molecules that did not match anywhere along the reference genome lacked a pattern that was rich enough in features to map back to the genome (see SI Appendix Fig. S14). This complication arose because all molecules were denatured at a single temperature found to be informative for the majority but where a uniformly GC-rich region can appear featureless and bright, while a uniformly AT-rich region can appear featureless and dark. The two other molecules that did not find a match in the reference genome did contain sufficient patterning for mapping to the reference; our inability to find a match for these molecules, or even parts of the molecules, indicates that they came from the fraction of the genome that remains uncharted and is therefore not represented in the reference genome (see SI Appendix, 'Unmapped molecules' section).

Significant structural variation was detected on Molecule 17 by D-R mapping (SI Appendix, Table S1); half of the 1 Mb-long molecule mapped poorly to chr19:8.5-9.1 Mb (SI Appendix, Fig. S16 and Fig. 4A). This is consistent with the LS174T whole-genome sequencing data. The read-depth analysis suggests a 50 kb deletion at 8.687-8.737 Mb which is likely homozygous as well as a 50 kb complex region around 8.87 Mb characterised by a higher read-depth but with a poor mapping of the reads (Fig. 4B-C). In this case single cell optical mapping data and bulk sequencing concur and point to a complex region.

A molecule from another single cell was found to map to the pericentromeric region of chromosome 4 but a 20 kb segment was absent between chr4:49.124 - 49.144 Mb on GRCh38 (region of interest (ROI)) (Fig. 5A). The sliding-window analysis showed a clear signature for the absence of the segment (11).

Whole genome sequencing indicates that the ROI is situated in a 66 kb-region with high but uneven coverage in both the bulk and single cell sequencing data (Fig. 5B). Single nucleotide variants, indels and gaps were pervasive around the ROI and varied between individual cells (SI Appendix, Fig. S17); this inter-cell variation would be impossible to resolve by bulk sequencing.

To confirm the verity of our single cell sequencing data, we looked at an arbitrary position in the genome (Fig. 5C) and found high coverage, low sequence variation, and alignment of both reads to the expected locations in the reference (see also

(28)). This contrasted to the substantial discordance between read pairs we found in the ROI, with one read mapping to the ROI while the other not (Fig. 5D-E and SI Appendix, Fig. S18). This paints a picture of local genome instability. This is unsurprising, given the pericentromeric location of the region where poorly delineated tracts of satellite and non-satellite sequences are known to span hundreds of kilobases (29). Segmental duplications are enriched in these regions, and may seed copy-number changes through errors in homologous recombination.

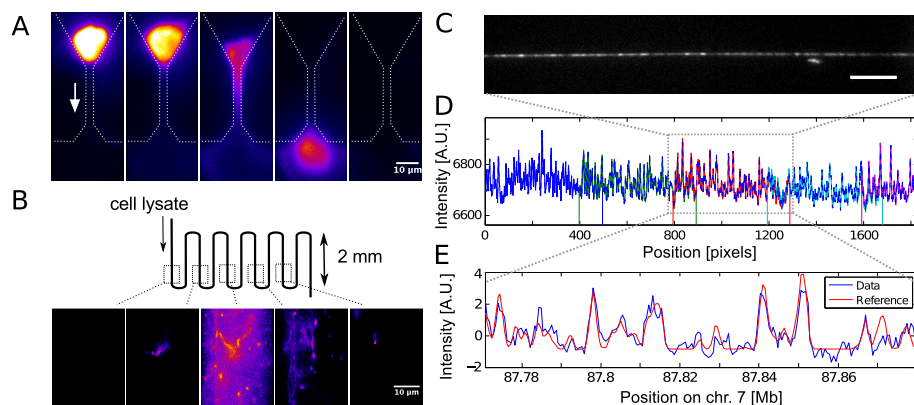
Not unexpectedly, pericentromeric regions are poorly represented in reference genomes, and sequence reads are difficult to align and would typically be discarded. Although we found that some reads in our data could be mapped to this region, the same reads found matches to several other locations in pericentromeric regions of the genome (SI Appendix, Table S3). Clearly, the mapping of the single cell sequencing data for this region could not be easily interpreted. This makes the clear detection of the missing segment by D-R mapping highly significant and shows that our approach can provide answers in regions where sequencing does not.

## Discussion

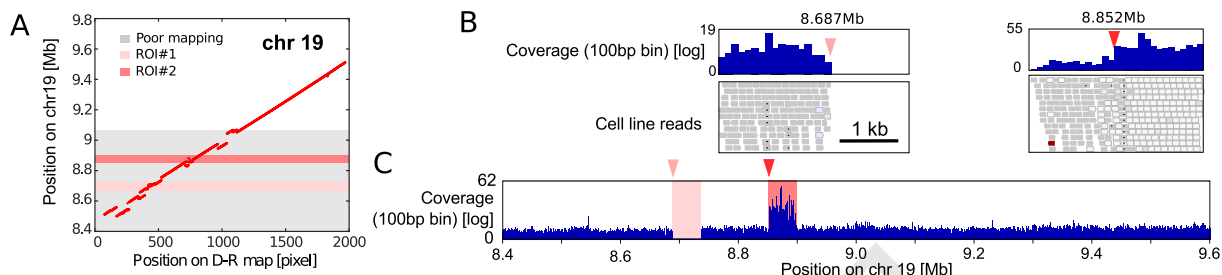
A substantial proportion of somatic structural variations in cancer genomes are complex events and are rich in inter-chromosomal rearrangements (30, 31). Short-read sequencing does not easily reveal mid- to long-range complex structural changes without a substantial bioinformatics effort (32). By contrast, the coarse-grained megabase range views of the genome provided by single-cell D-R optical mapping easily detect such changes.

We have shown that D-R mapping can determine that a region of the genome (on chromosome 19) in a LS147T cell has a 50 kb segment that is missing and one that matches poorly to the reference, and that our single cell sequencing data agrees with this finding. However, D-R mapping of a megabase fragment from another single cell shows the absence of a 20 kb segment on chromosome 4, which is not picked up by single cell read-depth or read pair sequence analysis. However single cell sequencing does suggest that this is a complex, unstable region of the genome which sequencing has not been able to chart accurately. Computational reconstruction of the region by paired reads is unproductive because the region is





**Fig. 3.** Extraction and mapping of DNA from a trapped cell. (A) Proteolysis liberates the genomic DNA from the trapped nucleus under flow (arrow). (B) DNA is collected as a 10 nL plug of solution in a section of the device where denaturation and re-naturation is performed. (C) Single field-of-view of a segment of a stretched DNA molecule (scale bar is 10  $\mu\text{m}$ ). (D) Full D-R map of a molecule is stitched together from five fields-of-view. (E) Match between the D-R map and a position on chromosome 7 of computer-simulated whole-genome melting map.



**Fig. 4.** Optical D-R map and paired-end sequencing on Chr 19: (A) Plot of sliding window analysis of experimental versus *in silico* D-R maps, revealing a region where the experimental D-R map aligns poorly to the reference. The coverage (C) and the visual analysis of the reads in the Integrative Genomics Viewer (IGV) (B) show two regions of interest (ROI) in this region, one is a homozygous deletion and one is a complex region.

not well mapped in the reference and has a high degree of similarity with other parts of the genome. By contrast, with D-R mapping we are able to directly visualize the structure of the fragment and unequivocally assign its genome location by comparing its D-R melting map with an *in silico* generated melting map of the human reference genome (SI Appendix, Section 2); most of the megabase fragments matches with high confidence to chromosome 4 but, if the reference genome is an accurate representation of this region, a 20 kb segment is clearly deleted in this fragment or translocated to another location. Because of the visual nature of the D-R experimental data there is little doubt that what we are looking at is real and not an artifact. From LS147T cell line literature, the cell line has a 45,X karyotype, and spectral karyotyping does not reveal any large scale rearrangements (33, 34). Moreover, bulk SNP analysis of chromosome 19 does not find microdeletions (data not shown). The single-cell D-R mapping might therefore be providing information on an uncommon variant. Single cell D-R mapping thus expands the tools available to detect genomic heterogeneity and such a capability might help provide deeper insight into the structure and dynamics of a genomic region and may aid the mapping of disease genes and the finding of drug targets.

The modest throughput of our current device design is balanced by the fact that we obtain a map from just one molecule. While other nanofluidic platforms may currently be able to process more molecules, the optical maps generated rely intrinsically on the data from several overlapping molecules and thus do not have true single molecule sensitivity required to map genomic DNA from single cells. Moreover, analysis based on individual megabase molecules means that we can resolve the long-range structure of haplotypes in a diploid genome. With the application of our full automation method

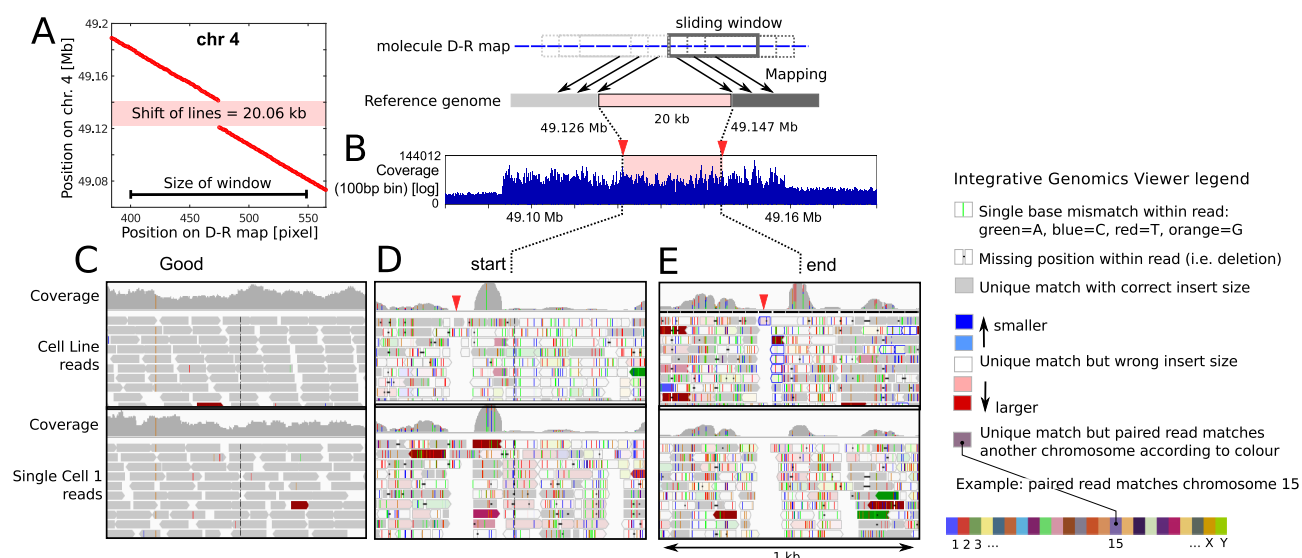
(35) it will be possible to apply D-R optical mapping to whole single cell genomes. Single cell D-R mapping will then have the potential to simply and inexpensively screen structural variation at the whole genome-scale, and between cells as a complement or alternative to sequencing.

## Materials and Methods

**Optical mapping device design.** The device architecture for optical mapping is summarized in Fig. 1B. Microchannels are designed to accommodate cells from a human cancer cell line with a typical diameter of 15  $\mu\text{m}$  (SI Appendix, Fig. S2), so the microfluidic network has minimal depths and widths of 33  $\mu\text{m}$ . The device comprises an inlet for cells and an inlet for buffer (Fig. 1B), which merge into a single channel to feed the single-cell trap (blue box in Fig. 1B). At the intersection between the cell and buffer inlets, cells get aligned along the side-wall of the feeding channel where the trap is located. The trap is a simple constriction dimensioned to capture cells from a human cancer cell line. The constriction for cell trapping has a trapezoidal cross-section: it is 4.3  $\mu\text{m}$ -wide at the bottom, 6  $\mu\text{m}$  at middle depth and 8  $\mu\text{m}$  at the top with a depth of 33  $\mu\text{m}$ . The cell trap connects the feeding channel with the outlet channels, which consist of two meandering channels: one for waste and the other (green box in Fig. 1B) leading to the outlet and the flow-stretch device (red box in Fig. 1B). The flow-stretch device consists of a 20  $\mu\text{m}$ -wide, 450  $\mu\text{m}$ -long, 100 nm-deep cross-shaped nanoslit that connects the outlet channel to three other microchannels (11).

**Optical mapping device fabrication.** The device is fabricated by replicating a nickel shim using injection molding of TOPAS 5013 (TOPAS) as previously reported (36, 37). Briefly, a silicon master is produced by UV lithography and reactive ion etching. A 100 nm NiV seeding layer is deposited and nickel is electroplated to a final thickness of 330  $\mu\text{m}$ . The Si master is chemically etched away in KOH. Injection molding is performed using a melt temperature of 250°C, a mould temperature of 120°C, a maximum holding pressure of 1500 bar for 2 seconds, and an injection rate varying between





**Fig. 5.** Optical D-R map and paired-end sequencing on Chr 4: (A) plot of sliding window analysis, revealing a 20 kb-discontinuity in the experimental D-R map compared to reference: this is the region of interest (ROI). (B) Coverage plot of the cell line bulk sequencing data displayed with a 100 bp bin around the molecule including the ROI (C-E) IGV-plot of read-pair data for three genomic regions in bulk sample and a single cell comprising: an arbitrary 1 kb region 1 Mb before the ROI (C); 1 kb at the start of the ROI (D); 1 kb at the end of the ROI (E). For each case, the coverage is shown in grey in the top row. A stack of reads is shown in the rows below (arrows indicating orientation); grey indicates that both reads of the pair are consistent with the reference, including orientation and insert size; blue indicates a decrease and red color an increase in insert size between read pairs on the same chromosome; the color palette indicates that one read of the pair is on another chromosome according to the color key provided; vertical colored lines within reads indicate the identity of a single nucleotide polymorphism (with color code: A=green, C=blue, T=red, G=orange); the block dots or short line within reads indicate insertions. The red downward arrowheads point to a gap in reads that is present in all single-cell and bulk data in the vicinity of the start and end of the ROI.

20 cm<sup>3</sup>/s and 45 cm<sup>3</sup>/s. Finally, a 150 μm TOPAS foil is used to seal the device by a combined UV and thermal treatment under a maximum pressure of 0.51 bar. The surface roughness of the foil is reduced by pressing the foil at 140°C and 5.1 bar for 20 minutes between two flat nickel plates electroplated from silicon wafers prior to sealing the device. This ensures that the lid of the device is optically flat allowing for high-NA optical microscopy. The device is mounted on an inverted fluorescence microscope (Nikon T2000) equipped with an air objective (50x/0.8), an oil objective (60x/1.40) and an EMCCD camera (Photometrics cascade II 512). Liquids are driven through the device using a pressure controller (MFCS, Fluigent) at pressures in the 0-10 mbar range. The device is primed with ethanol, then degassed FACSFlow Sheath Fluid (BD Biosciences) is loaded in all microchannels but the microchannel connecting the flow-stretch device. Instead, a buffer suitable for single molecule imaging and electrophoresis (0.5xTBE+0.5 % v/v Triton-X100+1% v/v β-mercaptoethanol (BME)) is loaded in the channels of the flow-stretch device. This buffer prevents DNA sticking in the nanoslit of the flow-stretch device and suppress electroosmotic flow that can counteract the introduction of DNA to the nanoslit later in the experiment.

**Cells.** LS174T colorectal cancer cells are cultured in Dulbecco's Modified Eagle Medium (DMEM; Gibco) with 10% fetal bovine serum (FBS; Autogen-Bioclone UK Ltd) and 1% penicillin/streptomycin (Lonza) before freezing at a concentration of 1.7 10<sup>6</sup> cells/mL in 10% DMSO in FBS. After thawing, cell suspension is mixed 1:1 with FACSFlow buffer, centrifuged at 400 rpm for 5 minutes and resuspended in FACSFlow. Finally, the cells are stained with 1 μM Calcein AM (Invitrogen) and loaded in the chip at 0.35.10<sup>6</sup> cells/mL. So out of the 7000 cells loaded, the first cell trapped is analysed.

**Optical mapping device operation.** The protocol for operating the device is sketched in Fig. 1D and Fig. 2A-I. First, cells and buffer are introduced simultaneously aligning the cells along the side-wall of the micro channel where the trap is located. A cell is captured and kept in the trap for a buffer flow through the trap up to 30 nL/min. Cell doublets are present in the cell culture and may be captured in the cell trap. In this case the experiment is aborted. The lysis buffer composed of 0.5xTBE+0.5% v/v Triton-X100+0.1

μM YOYO-1 (Invitrogen) is loaded in one of the inlets and injected at 10 nL/min through the trap for 10 minutes (see SI Appendix). Then the solution is exchanged to a buffer without YOYO-1 in all wells in order to stop the staining. Next, the cell nucleus is exposed to blue excitation light at a dose of 1 nW/(μm)<sup>2</sup> for up to 300 seconds causing a partial photonic nicking of the DNA (see SI Appendix). Then the buffer is changed to a solution containing BME (0.5xTBE+0.5 % v/v triton-X100+1% v/v BME) and the excitation of the fluorescence lamp is lowered to the minimum intensity that still allows fluorescence imaging. Finally, the temperature is raised to 60°C and the proteolysis solution (0.5xTBE+0.5 % v/v Triton-X100+1% v/v BME+200 g/mL) is introduced, pushing the lysate through the trap. In experiments where a denaturation pattern is created on DNA extracted from a single cell, the DNA staining is modified. Thus the lysis buffer contains 0.04 μM YOYO-1 instead of 0.1 μM. In addition, the device is heated further after the proteolysis for 10 minutes at 80°C before rapid cooling to room temperature by moving the oil immersion objective in place. The temperature at which the D-R pattern is created is highly dependent on the dye loading and the ionic strength of the buffer and should be adjusted by steps of 2°C if necessary. DNA travels through the meandering channel as the temperature is lowered and an oil immersion objective is moved into place for single molecule imaging (60x/1.4). DNA fragments are introduced from the microchannel to the nanoslit of the flow-stretch device using electrophoresis by applying a voltage V of 5-10 Volts across the nanoslit. When a DNA fragment has both ends in opposite microchannels, voltage is turned off and the final stretching of the DNA is obtained by the flow-stretch at a pressure drop of 80 mbar (11).

**Imaging and mapping.** Imaging and mapping is performed as previously reported and the match on the human reference genome is validated as shown in detail in the Supplementary Information to (11). Briefly, the 450 μm portion of the molecule stretched in the nanoslit is imaged using 50 ms exposures in 5 overlapping fields-of-view. Since the DNA is at a predictable position in the nanoslit, this can be done using simple automation of the motorized stage and image acquisition (Metamorph, Molecular Devices) (11). In each field-of-view, the D-R pattern along the DNA (intensity averaged over 3-5 pixel width) was extracted from the average of twenty frames. Consecutive fields-of-view are stitched together using a

chi-square fit (Matlab) to produce a continuous D-R pattern covering the entire length of the nanoslit. Mapping of the experimental D-R map with the theoretical melting profile for the whole human genome (hg18, build 36) produced with the software Bubblyhelix ([www.bubblyhelix.org](http://www.bubblyhelix.org)) (38) is according to the Supplementary Information to (8). For comparing D-R maps to the sequencing data, the sliding-window analysis of the experimental D-R maps is performed on the GRCh37 genome.

**Device design for DNA extraction.** In experiments, where DNA is prepared for sequencing, LS174T cells are trapped in a separate device fabricated as described above and comprising only cell traps (and not the flow-stretch part) so that it allows to process several cells in parallel. Detailed description of the following can be found in (28). The device consists of a microfluidic network at a single depth (31  $\mu\text{m}$ ) and cell traps have a geometry similar to the trap in the device for optical mapping (i.e. they are 4.5  $\mu\text{m}$ -wide constrictions). The device architecture is shown in SI Appendix, Fig. S1. On this device, eight single cell traps are fed through a common channel and lead to separate outlets. Cells introduced to the common channel are aligned along the side-wall of the channel by a flow of buffer incoming from inlet B1 and B2.

**Device operation and DNA amplification.** The device is wetted with degassed 0.1% Triton-X100 in BD FACSTFlow buffer followed by degassed 0.1 mg/mL BSA in BD FACSTFlow buffer and left in the fridge overnight. Cells are loaded in the cell inlet and BDFACSTFlow is loaded in all other wells. Cells are trapped by applying 4 mbar to the cell, 5 mbar to B1 and B2 inlet until single cells occupy the traps. The trapped cells are lysed by loading the alkaline lysis solution (Repli-g kit) in inlet B1 and B2. Lysis occurs at room temperature by applying 5–10 mbar to B1 and B2 and DNA is pushed in separate outlets of the device. After adding neutralization solution, multiple displacement amplification is performed in the outlet wells using either REPLI-g Mini Kit (Qiagen). MasterMix with Phi29 polymerase was added to the chip outlets containing gDNA, and the device is placed in a thermal cycler for 8 hours at 30°C. Finally, polymerase is inactivated 5 min at 65°C. The amplified DNA is collected and quantified using Qubit dsDNA HS kit (Thermo Fischer). DNA from single cells presented in Fig. 4C–F was extracted on different devices, on different days.

**DNA sequencing.** To confirm the quality of WGA products, multiplex PCR is performed on the amplified DNA. Five primers targeting five chromosomes (2, 4, 12, 13 and 22) are used. PCR products (295 bp Chr13, 235 bp Chr12, 196 bp Chr2, 150 bp Chr22 and 132 bp Chr4) are visualized on a 4% agarose gel. The DNA material that passes quality controls, i.e. five bands on gel after multiplex PCR, was then prepared for sequencing. Library preparation for next-generation sequencing was done using Nextera DNA Library Prep kit (Illumina) with 10 ng of amplified DNA as input. Illumina paired-end sequencing (2x100 bp or 2x125 bp) was performed on HiSeq 2000/2500 instruments using HiSeq SBS and Cluster kits (v3 or v4 chemistry). For comparison we also prepared a library from DNA extracted off-chip from the LS174T cell line. For this sample (Cell line in Fig. 4C–F) the library was prepared using the Illumina TruSeq Genomic kit (this is because, with a larger amount of DNA, this is the best preparation method and represents the gold standard of Illumina sequencing), but was run on four of the Illumina HiSeq high output lanes, with 2x100bp paired end reads, using the v3 chemistry. Read alignment is performed on the GRCh37 genome and visualized in the Integrative Genomics Viewer (39).

**ACKNOWLEDGMENTS.** This work was supported by European Union's Seventh Framework Programme P7/2007–2013 Grant 278204 (CellOMatic) and Danish Council for Strategic Research Grant 10-092322 (PolyNano).

- Navin NE (2015) Delineating cancer evolution with single-cell sequencing. *Science Translational Medicine* 7(296):296s29.
- Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* 17(3):175–188.
- McConnell MJ, et al. (2013) Mosaic copy number variation in human neurons. *Science* 342(6158):632–637.

- Cheow LF, et al. (2016) Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature Methods* 13(10):833–.
- Bedard PL, Hansen AR, Ratain MJ, Siu LL (2013) Tumour heterogeneity in the clinic. *Nature* 501(7467):355–364.
- Bentley DR, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218):53–59.
- Schwartz DC, et al. (1993) Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science* 262(5130):110–114.
- Reisner W, et al. (2010) Single-molecule denaturation mapping of DNA in nanofluidic channels. *Proceedings of the National Academy of Sciences of the United States of America* 107(30):13294–13299.
- Neely RK, et al. (2010) Dna fluorocode: A single molecule, optical map of dna with nanometre resolution. *Chemical Science* 1(4):453–460.
- Teague B, et al. (2010) High-resolution human genome structure by single-molecule analysis. *Proceedings of the National Academy of Sciences of the United States of America* 107(24):10848–10853.
- Marie R, et al. (2013) Integrated view of genome structure and sequence of a single DNA molecule in a nanofluidic device. *Proceedings of the National Academy of Sciences of the United States of America* 110(13):4893–8.
- Hastie AR, et al. (2013) Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex *Aegilops tauschii* Genome. *PLoS ONE* 8(2):e55864.
- Freitag C, et al. (2015) Visualizing the entire DNA from a chromosome in a single frame. *Biomicrofluidics* 9(4):044114.
- Baday M, et al. (2012) Multicolor Super-Resolution DNA Imaging for Genetic Analysis. *Nano Letters* 12(7):3861–3866.
- Vranken C, et al. (2014) Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. *Nucleic Acids Research* 42(7):e50–e50.
- Bensimon A, et al. (1994) Alignment and sensitive detection of DNA by a moving interface. *Science* 265(5181):2096–2098.
- Tegenfeldt JO, et al. (2004) The dynamics of genomic-length DNA molecules in 100-nm channels. *Proceedings of the National Academy of Sciences of the United States of America* 101(30):10979–10983.
- Das SK, et al. (2010) Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Research* 38(18):Article No.: e177.
- Yeh JW, Taoni A, Chen YL, Chou CF (2012) Entropy-Driven Single Molecule Tug-of-War of DNA at Micro-Nanofluidic Interfaces. *Nano Letters* 12(3):1597–1602.
- Jo K, et al. (2007) A single-molecule barcoding system using nanoslits for DNA analysis. *Proceedings of the National Academy of Sciences of the United States of America* 104(8):2673–2678.
- Kounovsky-Shafer KL, et al. (2013) Presentation of Large DNA Molecules for Analysis as Nanoconfined Dumbbells. *Macromolecules* 46(20):8356–8368.
- Perkins T, Smith D, Chu S (1997) Single polymer dynamics in an elongational flow. *Science* 276(5321):2016–2021.
- Cipriani BR, et al. (2012) Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. *Proceedings of the National Academy of Sciences of the United States of America* 109(22):8477–8482.
- Berard DJ, et al. (2014) Convex lens-induced nanoscale templating. *Proceedings of the National Academy of Sciences of the United States of America* 111(37):13295–13300.
- Pedersen JN, et al. (2014) Thermophoretic Forces on DNA Measured with a Single-Molecule Spring Balance. *Physical Review Letters* 113(26):268301.
- Benitez JJ, et al. (2012) Microfluidic extraction, stretching and analysis of human chromosomal DNA from single cells. *Lab on a Chip* 12(22):4848–4854.
- Mahshid S, et al. (2015) Development of a platform for single cell genomics using convex lens-induced confinement. *Lab on a Chip* 15(14):3013–3020.
- Marie R, et al. (2018) Sequencing metrics of human genomes extracted from single cancer cells individually isolated in a valveless microfluidic device. *bioRxiv*.
- Horvath J, Schwartz S, Eichler E (2000) The mosaic structure of human pericentromeric dna: A strategy for characterizing complex regions of the human genome. *Genome Research* 10(6):839–852.
- Yang L, et al. (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 153(4):919–929.
- Stephens PJ, et al. (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1):27–40.
- Baslan T, et al. (2012) Genome-wide copy number analysis of single cells. *Nature Protocols* 7(6):1024–1041.
- Rutzky L, Kaye C, Siciliano M, Chao M, Kahan B (1980) Longitudinal Karyotype And Genetic Signature Analysis Of Cultured Human-Colon Adenocarcinoma Cell-Lines LS180 And LS174T. *Cancer Research* 40(5):1443–1448.
- Abdel-Rahman WM, et al. (2001) Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. *Proceedings of the National Academy of Sciences of the United States of America* 98(5):2538–2543.
- Sorensen KT, et al. (2015) Automation of a Single-DNA Molecule Stretching Device. *Review of Scientific Instruments* 86.
- Utiko P, Persson F, Kristensen A, Larsen NB (2011) Injection molded nanofluidic chips: fabrication method and functional tests using single-molecule DNA experiments. *Lab on a chip* 11(2):303–8.
- Ostergaard PF, et al. (2015) Optical mapping of single-molecule human DNA in disposable, mass-produced all-polymer. *Journal of Micromechanics and Microengineering* 25:105002.
- Tostesen E, Liu F, Jenssen TK, Hovig E (2003) Speed-up of DNA melting algorithm with complete nearest neighbor properties. *Biopolymers* 70(3):364–376.
- Robinson JT, et al. (2011) Integrative genomics viewer. *Nature Biotechnology* 29(1):24–26.