Single-molecule DNA-mapping and whole-genome sequencing of individual cells

Marie, Rodolphe; Pedersen, Jonas N.; Bærlocher, Loic; Koprowska, Kamila; Pødenphant, Marie; Sabatel, Céline; Zalkovskij, Maksim; Mironov, Andrej; Bilenberg, Brian; Ashley, Neil

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1804194115

Publication date:
2018

Publisher's PDF, also known as Version of record:

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
To elucidate cellular diversity and clonal evolution in tissues and tumors, one must resolve genomic heterogeneity in single cells. To this end, we have developed low-cost, mass-producible micro-/nanofluidic chips for DNA extraction from individual cells. These chips have modules that collect genomic DNA for sequencing or map genomic structure directly, on-chip, with denaturation-renaturation (D-R) optical mapping [Marie R, et al. (2013) Proc Natl Acad Sci USA 110:4893–4898]. Processing of single cells from the LS174T colorectal cancer cell line showed that D-R mapping of single molecules can reveal structural variation (SV) in the genome of single cells. In one experiment, we processed 17 fragments covering 19.8 Mb of the cell's genome. One megabase-large fragment aligned well to chromosome 19 with half its length, while the other half showed variable alignment. Paired-end single-cell sequencing supported this finding, revealing a region of complexity and a 50-kb deletion. Sequencing struggled, however, to detect a 20-kb gap that D-R mapping showed clearly in a megabase fragment that otherwise mapped well to the reference at the pericentromeric region of chromosome 4. Pericentromeric regions are complex and show substantial sequence homology between different chromosomes, making mapping of sequence reads ambiguous. Thus, D-R mapping directly, from a single molecule, revealed characteristics of the single-cell genome that were challenging for short-read sequencing.

Results
Using valve-less polymer devices that can be easily mass-produced by injection moulding (e.g., Fig. 1A), we have single-molecule optical mapping of DNA from individual cells 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. D-R mapping identifies structural variations (SVs) 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
We report optical mapping of DNA from a single cell. Notably, we demonstrate isolation of single cells, DNA extraction, and optical mapping, all within a single integrated micro-/nanofluidic device. Single-cell optical mapping is less complex than sequencing, which we performed after whole-genome amplification of DNA extracted from a single cell isolated on-chip. In some cases, optical mapping was more efficient than sequencing at detecting structural variation. As single-cell analysis can address genomic heterogeneity within a tumor, it may prove useful for the selection of cancer therapies. Thus, optical mapping of the long-range features of single-cell genomes and sequencing of the short-range features may become complementary tools for the analysis of tumors.


Conflict of interest statement: R.M., J.N.P., A.K., and K.U.M. filed patents. K.U.M. declares that XGenomes is developing nucleic acid sequencing technologies. This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: The single-cell sequencing data are available at the European Nucleotide Archive (accession nos. ERS2168085–ERS2168123), https://www.ebi.ac.uk/ena/browse.

1To whom correspondence should be addressed. Email: rxw@nanotech.dtu.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1804194115/-/DCSupplemental.
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 toward 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 with 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 nonhuman DNA; >87% of bases were represented at a sequencing depth as low as 10×; allelic dropout in the single cells across all high-quality heterozygous variants found in bulk was as low as 30% at 15× to 20× 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 and S6). 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 subfragments 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 photonicking and fragment length is reciprocal because the relationship between dose and nicking probability per kilobase is linear (SI Appendix, Figs. S7–S11). 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% stretching (Fig. 3C), which was sufficient for them to be mapped to the reference genome GRCh37 (Fig. 3D and 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 before 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 GRCh37 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 D-R map segments mapping on the genome at predictable distances from each other (i.e., they fall on a diagonal in the plots displayed in SI Appendix, Figs. S12–S20). When mapped against the mouse genome (GRCm38/mm10), we found no such match for the experimental D-R maps. Fig. 3 D and E shows the D-R map derived from the image of one of the
molecules extracted from the single cell and its 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. S21). This complication might arise because all molecules were denatured at a single temperature. This temperature had been found to yield informative D-R maps for most molecules. But a uniformly GC-rich region can appear featureless and bright at this temperature, while a uniformly AT-rich region can appear featureless and dark (see SI Appendix, section 2E). The two other molecules that did not find a match in the reference genome did contain sufficient pattern for mapping to the reference; our inability to find a match for these molecules, or even parts of the molecules, suggests that they came from the fraction of the genome that remains uncharted or is subject to extreme structural change (29), which renders it unmapable and is therefore not represented in the reference genome (SI Appendix, section 2C).

A number of the molecules that mapped to the genome had peaks in their $\chi^2$ values that coincide with line shifts larger than a few kilobases in their plots of the genome versus D-R map position (SI Appendix, Figs. S12–S20). These peaks indicate poor alignment to the reference; however, if no line shift is detected, we cannot pinpoint the nature of the SV (11). Insertions and deletions down to 5 kb can be detected with high confidence (for the SV detection sensitivity of D-R mapping, see SI Appendix, section 3). Significant SV 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 to 9.1 Mb (SI Appendix, Fig. S22 and Fig. S4). This is consistent with the LS174T whole-genome sequencing data. The read-depth analysis suggests a 50-kb deletion at 8.687 to 8.737 Mb, which is likely homozygous, as well as a 50-kb complex region around 8.87 Mb characterized by a higher read depth but with a poor mapping of the reads (Fig. 4 B and C). In this case, single-cell D-R mapping data and bulk sequencing concur and point to a complex region. However, the poor agreement of the D-R map with the reference genome leading to a high $\chi^2$-value background does not allow the SV detection by D-R mapping alone (SI Appendix, section 2B).

We found two segments clearly missing in molecule 12 (8 and 12 kb) at 51.012 and 51.513 Mb. Their absence could not be detected in the bulk or the single-cell data. However, we found a 20-kb region limited by two regions of reads with an incorrect insert size centered at the position of interest, 51.012 Mb. At the other position of interest, in one of the single-cell sequencing data, we see two deletions smaller than a kilobase positioned at 23 kb.
on each side of the position 51,510 Mb. Both regions are shown in SI Appendix, Figs. S23 and S24. In this case, the D-R mapping’s SV detection points to SVs detected by sequencing but without a one-to-one correspondence. This lack of one-to-one correspondence may be due to the fact that it is not the very same cell that we D-R map and sequence.

A molecule from another single cell was found to map to the pericentromeric region of Chr4, but a 20-kb segment was absent between Chr4:49.127 and 49.147 Mb on GRCh37 (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. S25); this intercell variation would be impossible to resolve by bulk sequencing.

To confirm the verity of the 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 ref. 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 did not (Fig. 5D and E and SI Appendix, Fig. S26). This paints a picture of local genome instability. This is unsurprising given the region’s pericentromeric location where poorly delineated tracts of satellite and non-satellite sequences are known to span hundreds of kilobases (30). 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

![Fig. 4.](image-url) 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. (B) The coverage and the visual analysis of the reads in the Integrative Genomics Viewer (IGV) show three normal regions and two ROIs in this region, one is a homzygous deletion and one is a complex region. (C) The coverage of reads over the whole fragment.

![Fig. 5.](image-url) 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 with reference: This is the 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 the bulk sample and a single cell comprising (C) an arbitrary 1-kb region 1 Mb before the ROI, (D) 1 kb at the start of the ROI, and (E) 1 kb at the end of the ROI. For each case, the coverage is shown in gray in the top row. A stack of reads is shown in the rows below (arrows indicating orientation); gray 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.
shows that our approach can provide answers in regions where sequencing does not.

Discussion

A substantial proportion of somatic SVs in cancer genomes are complex events and are rich in interchromosomal rearrangements (29, 31). Short-read sequencing does not easily reveal mid- to long-range complex structural changes without a substantial bioinformatics effort (32). By contrast, the megabase-range views of the genome provided by single-cell D-R optical mapping detect such changes.

We have shown that D-R mapping can determine that a region of the genome (on Chr19) in a LS174T cell has a ~0.6-Mb region that matches poorly to the reference and that our single-cell sequencing data are consistent with this. However, D-R mapping of a megabase fragment from another single cell shows the absence of a 20-kb segment on Chr4, 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 that sequencing has not been able to chart accurately. Computational reconstruction of the region by paired reads is unproductive because the region is 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 uniquely identify its genomic 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 match with high confidence to Chr4, 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. The LS174T cell line has a 45,X karyotype, and spectral karyotyping does not reveal any large-scale rearrangements (33, 34). Moreover, bulk SNP analysis of Chr19 does not find microdeletions (SI Appendix, Table S4). 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.

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 the 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 whole-genome SV 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. 18. Microchannels are designed to accommodate cells from a human cancer cell line with a typical diameter of 15 μm (SI Appendix, Fig. S2), so the microfluidic network has minimal depths and widths of 33 μm. The device comprises an inlet for cells and an inlet for buffer (Fig. 18) that merge into a single channel to feed the single-cell trap (blue box in Fig. 18). 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 μm wide at the bottom, 6 μm at middle depth, and 8 μm at the top with a depth of 33 μm. The cell trap connects the feeding channel with the outlet channels, which consist of two meandering channels: one for waste (red) and the other (green) to the flow-stretch device (red box in Fig. 18). The flow-stretch device consists of a 20-μm-wide, 450-μm-long, 100 nm-deep cross-ship 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-μm NIV seed layer is deposited and nickel is electroplated to a final thickness of 330 μm. 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 1,500 bar for 2 s, and an injection rate varying between 20 cm3/s and 45 cm3/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 MPa. The surface roughness of the foil is reduced by pressing the foil at 140°C and 5.1 MPa for 20 min between two flat nickel plates electroplated from silicon wafers before 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 a 60×/0.8, an oil objective and an EMCCD camera (Photometrics cascade II S2)). Fluids are driven through the device using a pressure controller (MFCS, Fluigent) at pressures in the 0 to 10 mbar range. The device is primed with ethanol, and 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 (0.5× TBE + 0.5% w/v Triton-X100 + 1% w/v β-mercaptoethanol, BME) is loaded in the channel of the flow-stretch device. This buffer prevents DNA sticking in the nanoslit of the flow-stretch device and suppresses 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’s medium (DMEM, Gibco) with 10% fetal bovine serum (FBS; AutogenBio, Bioclear UK Ltd) and 1% penicillin/streptomycin (Lonza) before freezing at a concentration of 1.7×10^5 cells per milliliter in 10% DMSO in FBS. After thawing, cell suspension is mixed 1:1 with FACSFLOW buffer, centrifuged at 28.8×g (A-4-44, Eppendorf) for 5 min, and resuspended in FACSFLOW. Finally, the cells are stained with 1 μM Calcein AM (Invitrogen) and loaded in the chip at 0.35×10^6 cells per milliliter. So out of the 7,000 cells loaded, the first cell trapped is analyzed.

Optical Mapping Device Operation. The protocol for operating the device is sketched in Figs. 1D and 2 A–I. First, cells and buffer are introduced simultaneously, aligning the cells along the side wall of the microchannel 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 0.5× TBE + 0.5% w/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 min (see SI Appendix). Then, the solution is exchanged to a buffer without YOYO-1 in all wells to stop the staining. Next, the cell nucleus is exposed to blue excitation light at a dose of 1 W/m^2 for up to 300 s, causing a partial photoinactivation of the DNA (see SI Appendix). Then, the buffer is changed to a solution containing BME (0.5× TBE + 0.5% w/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 (0.5× TBE + 0.5% w/v Triton-X100 + 1% v/v β-mercaptoethanol, BME) is introduced, pushing the lystate 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 0.4 μM YOYO-1 instead of 0.1 μM. In addition, the device is heated further after the proteolysis for 10 min at 80°C before rapid cooling to room temperature by moving the oil immersion objective in place. In 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 (60×/1.4, with a 1.5× lens giving a 180-nm pixel size). DNA fragments are introduced from the microchannel to the nanoslit of the flow-stretch device using electrophoresis by applying a voltage of 5
to 10 V 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).

DNA Sequencing. To confirm the quality of whole genome amplification (WGA) products, multiplex PCR is performed on the amplified DNA. Five primers targeting five chromosomes (2, 4, 12, 13, 22) were 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—that is, 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 (2 × 100 bp or 2 × 125 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 Figs. 4 B and C and 5 B–E), 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 2 × 100 bp paired-end reads, using the v3 chemistry. Read alignment is performed on the GrCh37 genome and visualized in the IGV (39).

ACKNOWLEDGMENTS. We thank Charles Cantor and Wilhelm Ansorge for advice. This work was supported by European Union’s Seventh Framework Programme FP7/2007-2013 Grant 278204 (CellOmatic) and Danish Council for Strategic Research Grant 10-092322 (PolyNano).