Long-Read Structural and Epigenetic Profiling of a Kidney Tumor-matched Sample with Nanopore Sequencing and Optical Genome Mapping

Sapir Margalit[1†], Zuzana Tulpová[1,2 †], Tahir Detinis Zur[1], Yael Michaeli[1], Jasline Deek[1], Gil Nifker[1], Rita Haldar[1], Yehudit Gnatek[3], Dorit Omer[3], Benjamin Dekel[3,4,5] Assaf Grunwald[1*],and Yuval Ebenstein[1,6*]

1 Department of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 6997801 Tel Aviv, Israel; 2 Institute of Experimental Botany of the Czech Academy of Sciences, Olomouc, Czech Republic; 3 Pediatric Stem Cell Research Institute, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, 52621 Ramat Gan, Israel; 4 Pediatric Nephrology Unit, The Edmond and Lily Safra Children's Hospital, Sheba Medical Center, 52621 Ramat Gan, Israel; 5 Sackler Faculty of Medicine, Tel-Aviv University, 6997801 Tel Aviv, Israel; 6 Department of Biomedical Engineering, Tel Aviv University, 6997801 Tel Aviv, Israel.

† These authors contributed equally.

* To whom correspondence should be addressed. Email: uv@tauex.tau.ac.il, assafgrw@gmail.com.

Keywords

Long reads, Optical genome mapping, Nanopore sequencing, Oxford Nanopore Technologies, Bionano Genomics, clear cell renal cell carcinoma, kidney cancer

Abstract

Carcinogenesis often involves significant alterations in the cancer genome architecture, marked by large structural and copy number variations (SVs and CNVs) that are difficult to capture with short-read sequencing. Traditionally, cytogenetic techniques are applied to detect such aberrations, but they are limited in resolution and do not cover features smaller than several hundred kilobases. Optical genome mapping and nanopore sequencing are attractive technologies that bridge this resolution gap and offer enhanced performance for cytogenetic applications. These methods profile native, individual DNA molecules, thus capturing epigenetic information. We applied both techniques to characterize a clear cell renal cell carcinoma (ccRCC) tumor's structural and copy number landscape, highlighting the relative strengths of each method in the context of variant size and average read length. Additionally, we assessed their utility for methylome and hydroxymethylome profiling, emphasizing differences in epigenetic analysis applicability.

Introduction

One of the most prominent signs of carcinogenesis is the structural deviation of the cancer genome from that of the parent cell, which often involves large structural and copy number variations (SVs and CNVs, respectively). Such variations are annotated as structural features (deletions, insertions, duplications, inversions, and translocations) differing from the human genome reference or a matched sample for comparison (Cosenza et al. 2022). Traditionally, cytogenetic techniques such as karyotyping are used to detect large structural variations ranging from whole chromosome duplications, chromosome arm deletions, and down to SVs of ~ 5-10 Mbp (Ozkan and Lacerda 2023; Lixing Yang 2020). Applying fluorescence in-situ hybridization (FISH) techniques may bring the resolution down to several hundred kbp (Cui et al. 2016), but a critical resolution gap remains between cytogenetics and short-read sequencing. The lack of access to genomic variation on the scales of 1-500 kbp has nourished the development of long-read technologies that can address this need. Various long-read methods have been introduced in recent years, including SMRT sequencing commercialized by PacBio, which routinely provides high-quality reads on the 10 kbp scale (Amarasinghe et al. 2020). Another concept involves a library preparation technique that allows linking proximal DNA fragments computationally by sequence barcode ligation (linked reads-10x genomics (Marks et al. 2019)/ TELseq (Chen et al. 2020)). Here, we utilized two techniques, Optical genome mapping and ONT sequencing, that stand out in their ability to cover the full gap in mapping ability between short-read sequencing and karyotyping, offering an enhanced alternative to traditional cytogenetic analysis.

Optical genome mapping (OGM), commercialized by Bionano Genomics Inc. (BNG), has already gained clinical utility and is emerging as an alternative to cytogenetics by mapping the coarse grain structure of unamplified genomic fragments hundreds of Kbp in length (Mathew et al. 2024; Pang et al. 2023). The molecules are labeled at a specific sequence motif (CTTAAG) by a methyltransferase enzyme that transfers a fluorescent molecule to the labeling site from a synthetic cofactor analog. Every molecule acquires a sequence-specific fluorescent pattern along the DNA backbone during this process. The labeled DNA sample is applied to a silicon chip, where the molecules are electrophoretically extended in an array of parallel nanochannels. Millions of long, extended DNA molecules with their overlaying fluorescent barcode are imaged in the channels at high throughput. Once the images are digitized, DNA molecules may be mapped to their genomic location according to the pattern of fluorescent spots along the DNA and its matching to the expected pattern on the genome reference. Alternatively, the patterns may be stitched and assembled to build the whole genome structure *de-novo* (Jeffet et al. 2021).

Oxford Nanopore Technologies (ONT) is another prominent player in the long-read mapping and sequencing space. In recent years, ONT sequencing reads have gotten longer, error rates have diminished, throughput has increased, and prices per genome have dropped to levels that may justify clinical utility (Wang et al. 2021; Mantere et al. 2019). For sequencing, DNA molecules are translocated through protein pores while measuring the electric ionic current flowing through the pore. Different sequence compositions generate various degrees of current attenuation, which is then computationally interpreted to generate the base sequence of the translocated DNA molecules. While offering single-base resolution, ONT provides shorter median read lengths compared to OGM. Both methods may be applied to native DNA that still carries chemical DNA modifications such as DNA methylation or DNA damage adducts. This gives rise to another beneficial feature: the acquisition of epigenetic information during genetic analysis. In OGM, an additional color may be used to chemically tag modifications of interest and create a hybrid genetic/epigenetic physical map of the molecules (Jeffet et al. 2021; Heck et al. 2019; Sharim et al. 2019; Gabrieli et al. 2018; Margalit et al. 2021). ONT, on the other hand, does not require any additional preparative steps for calling epigenetic modifications as it relies solely on the electrical contrast generated by the native chemical structure of the modified base (White and Hesselberth 2022). Nevertheless, accurate modification calling requires a complete training set, which is not trivial for most base modifications.

Herein, we used both methods in order to characterize the structural and copy number landscape of a ccRCC tumor, pinpointing the specific strengths of each technique. Additionally, we characterized the methylome and hydroxymethylome of the sample and highlighted differences in practical utility for epigenetic analysis.

Results

We began by analyzing the genetic makeup of a stage 3 ccRCC tumor, a common type of kidney cancer known for characteristic structural abnormalities (Moore et al. 2012; Quddus et al. 2019). Our workflow consisted of extracting high molecular weight DNA from a tumor and a normal adjacent tissue, followed by per-protocol OGM and ONT analyses (**Fig. 1**).

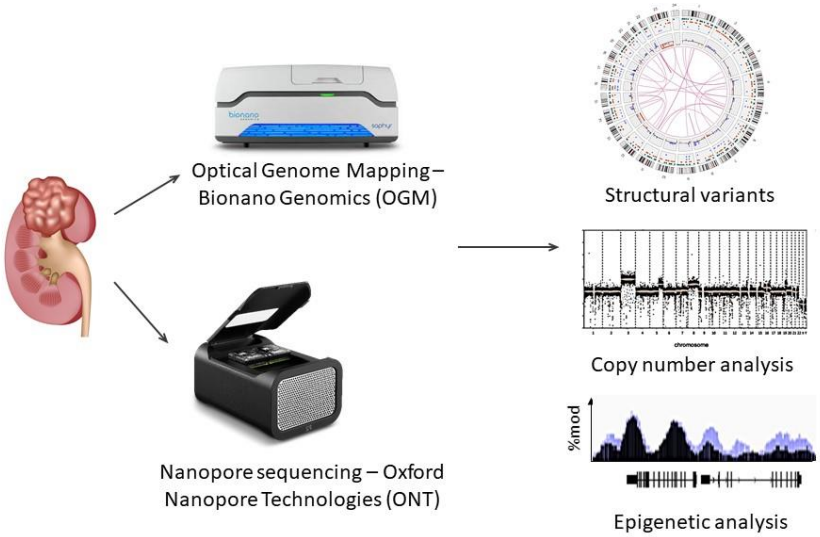

Figure 1. Experimental workflow. High molecular weight DNA was extracted from a ccRCC tumor and a normal adjacent tissue. Samples were analyzed by OGM and ONT to detect structural and copy number variations, and epigenetic modifications. Results from both methods were compared.

By generating long reads, both methods unlock access to intricate areas of the genome, enabling the study of diverse structural variations, copy number variations and repetitive elements. Additionally, as both methods read native, unamplified DNA, they are able to detect epigenetic modifications. The different attributes of each technique affect their performance in the aforementioned analyses. **Table 1** is based on public company material and summarizes some of the performance specifications, pointing to advantages and limitations of the two methods and indicating their compatibility of use, depending on research goals and budget.

	Bionano Genomics Saphyr	**Oxford Nanopore Technologies PromethION 2 Integrated**
Resolution	Optical resolution: ~1500 bp (Jeffet et al. 2021) SV detection resolution: starting from 500 bp for diploid genomes' insertions and deletions. inversions/ duplications: >30 kbp. translocations: >50 kbp (Bionano Genomics 2023c).	Single bp

Molecules N50	250 - 400 kbp, for molecules exceeding 150 kbp only (Bionano Genomics 2023b)	10–50 kbp (ONT 2024a)
Throughput per cell	80-300x (Bionano Genomics 2023b)	16-66x (ONT 2024a)
Price per sample (including cell, reagents and device rental)	550$, when buying a package for 120 experiments) (Bionano Genomics 2023a)	PromethION 2 Integrated: 1010$, when buying “project pack” for 96 experiments (ONT 2024a)
Price per 1x human genome coverage	1.8-7	15$–65$
Methylation calling	Labeling of reduced representation unmodified cytosines can be added (unsupported) (Sharim et al. 2019)	Integrated (ONT 2024a, 2024b)
5hmC calling	Direct labeling of 5hmC can be added (unsupported) (Gabrieli et al. 2018)	Integrated (ONT 2024a, 2024b)

Table 1. ONT and OGM specifications.

SV and CNV analysis of ccRCC tumor

Clear cell renal cell carcinoma (ccRCC) is the most common type of renal carcinoma, and its incidence has been increasing in recent years. Over 90% of ccRCC cases demonstrate distinctive changes to the short arm of chromosome 3 (3p), from translocations and deletions to the loss of the entire chromosomal arm. Most cases involve the genetic or epigenetic inactivation of the von Hippel–Lindau (*VHL*) gene, located on this arm (Moore et al. 2012; Jonasch et al. 2021; Le and Hsieh 2018). Other frequently observed copy number variations and cytogenetic abnormalities in ccRCC include a gain of chromosome 5q, loss of 14q, trisomy of chromosome 7, loss of 8p, loss of 6q, loss of 9p, loss of 4p and loss of chromosome Y in men. Some CNVs were correlated with prognosis (Moore et al. 2012; Quddus et al. 2019; Klatte et al. 2009).

Herein, DNA from a clear cell renal cell carcinoma (ccRCC) tumor and a normal adjacent tissue were analyzed using both ONT and BNG platforms. **Table 2** summarizes the resulting N50 and average coverage.

	Effective genome coverage of aligned molecules	**N50 of aligned molecules**
OGM	Tumor sample: 133X Normal adjacent sample: 123X	Tumor sample: 291 kbp Normal adjacent sample: 233 kbp
ONT	Tumor sample: 36X Normal adjacent sample: 19X	Tumor sample: 18 kbp Normal adjacent sample: 15 kbp

Table 2. Coverage and N50 of OGM and ONT genetic experiments.

To compare the efficacy of the sequencing and data analysis processes offered by each method, we applied the manufacturer's recommended pipelines to our generated data, adhering to default parameters unless specified otherwise (see methods section). Genome-wide copy number and SVs in ONT data were inferred by employing the EPI2ME workflow "wf-human-variation" to each sample. Genome-wide copy number and SVs in OGM data were obtained by generating a *de novo* assembly and "variant annotation pipeline" of each sample as a single sample. Copy number in both methods was calculated in 500 kbp bins. A running median over 10 bins was calculated to plot a smooth red line across the copy number dots and call comparable CNVs (tumor plots are shown in **Figure 2.A**). As expected, both methods produced highly similar CNV plots, identifying the loss of one copy of the entire 3p chromosomal arm, as well as a large DNA gain in 5q, and a smaller DNA loss in the same arm. Aneuploidies were found by both methods in chromosomes 7 and 12. The loss of 3p, gain in 5q and trisomy in chromosome 7 are well-documented genetic characteristics of ccRCC (Quddus et al. 2019; Klatte et al. 2009). The normal adjacent sample did not exhibit any large copy number variation in both methods.

Clearly, both methods are adequate for basic Karyotyping; however, SV detection exhibited less congruence between the two methods. To facilitate a comprehensive comparison, we categorized detected SVs based on their size (**Figure 2.B**). As anticipated, for short SVs (500 bp - 1 kbp), ONT detected more SVs, leveraging its superior resolution at the base-pair level. In the intermediate SV length range (1 to 10 kbp), both methods demonstrated a comparable number of detections, with substantial overlap. Notably, ONT detected 28,089 SVs smaller

than 500 bp while OGM exhibited an advantage in the detection of larger SVs (more than 10 kbp) owing to the larger N50 it generates.

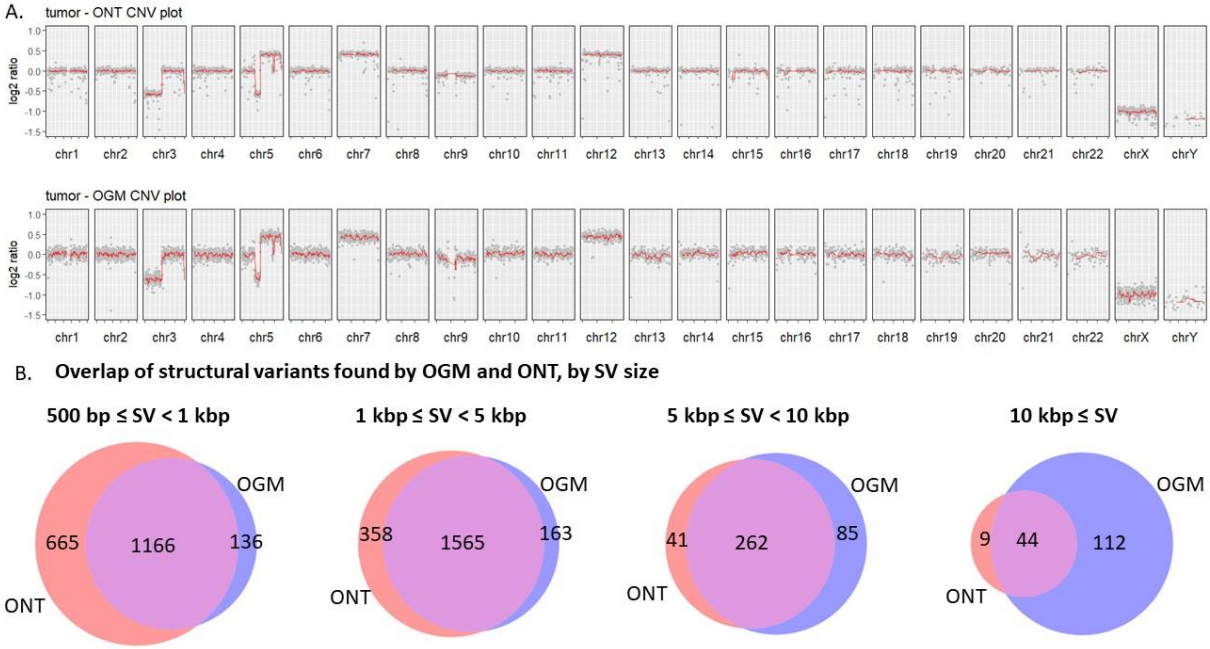

Figure 2. Comparative analysis of structural variations (SVs) and copy number variations (CNVs) in a ccRCC tumor, as detected by ONT and OGM. **A.** CNV plots (log_2 of the copy ratio) generated from ONT (top) and OGM (bottom). Data illustrate highly similar findings, pinpointing a significant DNA loss on chromosome 3 and various losses and gains on chromosomes 5, 7 and 12. **B.** Venn diagrams displaying common and unique SVs to OGM and ONT, in four size ranges: 500 bp -1 kbp, 1-5 kbp, 5-10 kbp and above 10 kbp.

Figure 3 depicts chromosomes 3 and 5, two chromosomes that are frequently disrupted in ccRCC, with marks indicating the relative positions of SVs and CNVs larger than 5 kbp identified by OGM and ONT. SVs overlapping BNG's list of N-base gaps in the reference or putative false positive translocation breakpoints (for de novo assembly, Solve 3.6.1) were masked. As seen also in **Figure 2.A,** the two methods detected DNA gain/loss events in these chromosomes. Additionally, both methods exhibited a high degree of concordance for insertions and deletions. **Figure 3.B** showcases a representative deletion identified by ONT and OGM. More complex SVs, as translocations and inversions, were reported solely by OGM. In chromosomes 3 and 5, OGM detected an inversion, 3 intrachromosomal and 1 interchromosomal (**Figure 3.C**) events. Notably, only 8 SVs identified by OGM in these chromosomes did not appear in BNG's database of mapped healthy controls. 2 of them were also found by ONT, and 6 of them are potential somatic variants, absent from the normal sample adjacent to the tumor (of these, one was also found by ONT).

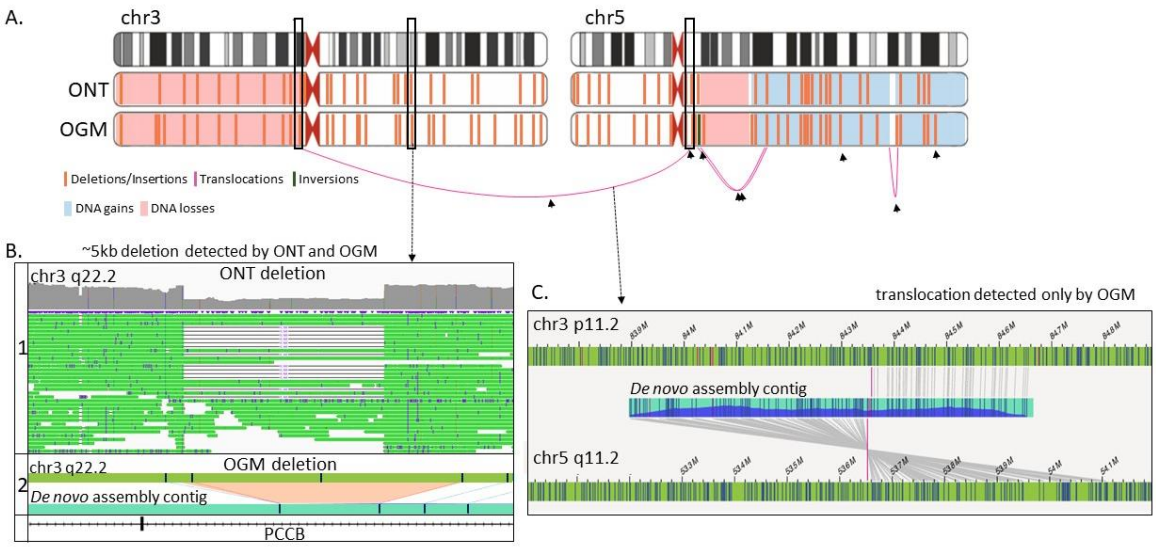

Figure 3. Large SVs detected on chromosomes 3 and 5 by ONT and OGM. **A.** Schematic representation of chromosomes 3 and 5 (cytobands are displayed on top). Relative locations of SVs larger than 4.9 kbp detected by ONT (middle panel) and OGM (bottom panel) are presented and color-coded as detailed in the figure. Small arrowheads indicate OGM SVs that do not appear in BNG's dataset of healthy controls. **B.** A ~5000 bp deletion identified by ONT and OGM in the PCCB gene region (chr3 q22.2). 1. ONT coverage and reads plotted in Integrative Genome Viewer (Robinson et al. 2011). 2. OGM reference chromosome and *de novo* assembled contig showing the deletion. Blue lines indicate CTTAAG barcode labels in the contig and reference. **C.** Interchromosomal translocation detected only by OGM. The green strips at the top and bottom represent the reference chromosomes 3 (top) and 5 (bottom). The middle strip is the assembled contig from the ccRCC tumor (with marked coverage). Blue lines indicate CTTAAG barcode labels in the contig and reference. Pink lines indicate translocation breakpoints.

Epigenetic Analyses

Methylation profiling

Modification calling is an attractive feature of ONT, that sequences native, unamplified DNA. Accordingly, it can call modifications directly, without needing chemical conversions (like in bisulfite sequencing, for example) (Yuen et al. 2021). Recently, ONT incorporated the calling of the two most prominent epigenetic marks, cytosine methylation and hydroxymethylation, to the recommended workflow (wf-human-variation, (ONT 2024b). Methylation calling is performed by the Remora framework (https://github.com/nanoporetech/remora), based on high-quality training data created by *in-situ* enzymatic CpG methylation.

Commercially supported OGM only provides tools for analyzing genomic structure for cytogenetic applications. However, more layers of information may be multiplexed with the use of colors. Some of the BNG Saphyr systems contain three color lasers, two of them are for generating the genetic barcode and DNA backbone, and the third can be used with orthogonal chemistries to tag genomic features of choice, including epigenetic marks.

Unmethylated CpG sites, complementary to methylated (5mC) sites, can be specifically labeled by methyltransferase enzymes. We have recently applied an engineered CpG methyltransferase to address all unmethylated CpGs (Gabrieli et al. 2022; Avraham et al. 2023). However, the method was not yet validated for human methylome profiling and thus we used the previously validated reduced representation optical methylation mapping (ROM; (Sharim et al. 2019; Margalit et al. 2021, 2022). This method uses the methyltransferase M.TaqI, which directly transfers a fluorescent tag from a synthetic cofactor to an adenine base in the enzyme's recognition sequence TCGA. However, if the CpG nested in this sequence motif is methylated or modified, the labeling reaction is blocked (**Fig 4.A**). Consequently, the DNA is labeled in all unmodified CpGs that are within TCGA sites. This reduced representation of the human methylome encompasses only ~ 6% of the total CpGs but coincidently captures the majority of regulatory sites in the genome and has been shown to present a cell-type specific pattern (Sharim et al. 2019; Margalit et al. 2021).

To facilitate a direct comparison of methylation signals between the two methods, we transformed the ONT methylation values into unmethylation signals by presenting the complement to 1 of the calculated methylation level. We applied a minimum coverage threshold, requiring at least 5 reads for ONT and 20 reads for OGM. In order to account for the lower resolution of OGM, we calculated the average unmethylation signals in non-overlapping 1 kbp genomic windows. **Fig. 4.B** shows the average genome-wide unmethylation signal in the ccRCC tumor sample and the normal adjacent tissue, as was detected by both methods. The plot shows that ONT signals (tumor: 21.06%; normal adjacent: 22.15%) are higher than OGM signals (tumor: 11.96%; normal adjacent: 11.55%), as expected from whole methylome vs. reduced representation methods. Additionally, both methods show that the global unmethylation level in the ccRCC tumor and normal adjacent tissue are very similar. The resolution and methylation representation differences become more apparent when characterizing smaller regions of the genome. **Fig 4.C** shows the unmethylation profile of a ~400 kb region in chromosome 22q11.21. Three representative examples for methylation comparison are marked in red boxes that contain variable TCGA content (seen in the lower panel). The leftmost box showcases a region lacking any TCGA sites in the reference genome. Consequently, the ONT plot exhibits a high unmethylation signal (indicating unmethylated

CpGs), while the OGM profile shows no signal. The middle box highlights two adjacent bins with relatively high TCGA density, resulting in signal peaks by both methods. The rightmost box depicts a region lacking a reference TCGA site, yet the OGM profile displays a peak. Intriguingly, investigation of the corresponding ONT sequence revealed an A-to-G single nucleotide polymorphism (SNP), creating a new TCGA site recognizable by the M.TaqI enzyme, thus explaining the observed OGM signal. **Figure 4.D** shows that despite absolute intensity differences, similar trends are seen in the unmethylation profile generated along genes when grouped by their gene expression score in normal tissues adjacent to ccRCC tumors ((Yao et al. 2017); see methods). Both methods display the high unmethylation signal around the transcription start site (TSS), which increases with gene expression. In contrast, the level in gene bodies is much lower and similar among all expression levels.

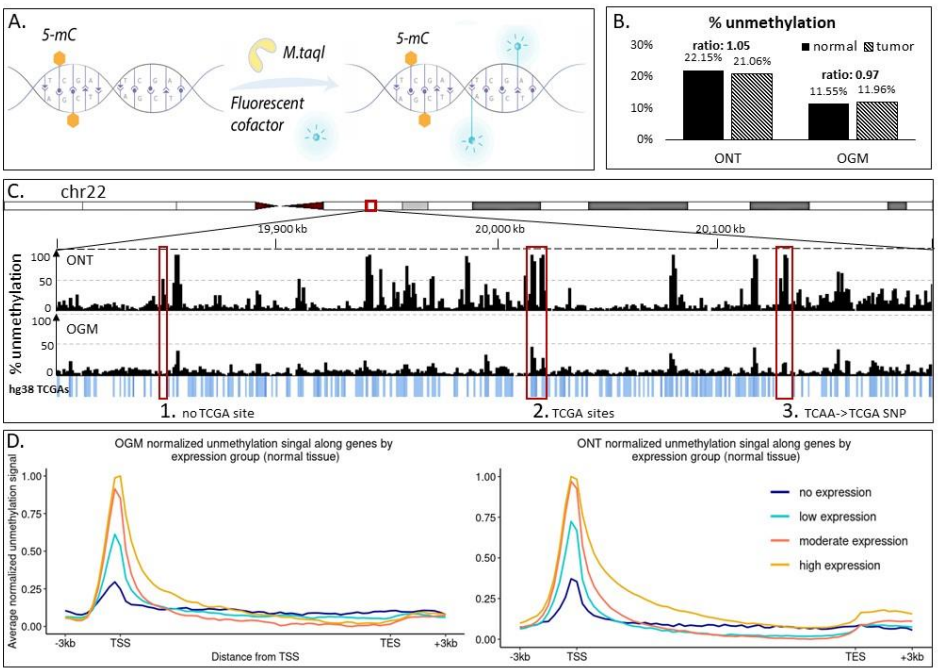

Figure 4. Unmethylation analysis. **A.** Fluorescent labeling scheme for unmodified CpGs embedded in TCGA motif. **B.** Global unmethylation levels of a ccRCC tumor and normal adjacent tissue, as detected by OGM and ONT. **C.** Unmethylation profiles by ONT and OGM along a region in chromosome 22, and the corresponding density of TCGA motif in the hg38 reference. Three red boxes mark three regions that differ in TCGA content: 1: no TCGAs, there is a peak in ONT signal and not in OGM. 2: TCGAs are present, peaks in both methods. 3: no TCGAs in the reference, peaks in both methods due to a SNP. Positions of the sequence motif TCGA in the reference were obtained using the R package BSgenome (https://bioconductor.org/packages/release/bioc/html/BSgenome.html). **D.** ONT and OGM unmethylation signal, each normalized between 0 and 1, of the normal adjacent kidney tissue

along aggregated genes. Genes were grouped based on their expression in kidney tissues adjacent to ccRCC tumor.

hydroxymethylation profiling

5-hydroxymethyl cytosine (5hmC), the first oxidation product of methylcytosine, is another important modification that was linked to gene regulation, development and disease, predominantly cancer (Shi et al. 2017; Jin et al. 2011). Accurately detecting DNA modifications with ONT requires training data encompassing the modification across diverse sequence contexts. Such datasets can be generated through experimental validation with established methods or in-situ approaches (as for 5mC) (White and Hesselberth 2022). Generating in-situ 5hmC residues is chemically challenging compared to 5mC, and the Remora-based 5hmC caller hasn't been benchmarked and peer-reviewed to date.

Optical mapping of 5hmC was introduced several years ago based on the fluorescent labeling of 5hmC residues (Gabrieli et al. 2018; Margalit et al. 2022). 5hmC is directly labeled in a process that involves the attachment of an azide-modified glucose moiety from a synthetic cofactor (UDP-6-N3-Glu; (Nifker et al. 2015; Song et al. 2011)) to the hydroxyl group of 5-hmC, followed by a click reaction that connects a fluorophore-bound alkyne to the azide-labeled 5-hmC (**Fig 5.A**; (Michaeli et al. 2013; Shahal et al. 2014)). **Fig. 5.B** shows the average genome-wide 5hmC signal in the ccRCC tumor sample and the adjacent tissue, as was detected by both methods. Consistent with published reports indicating a global reduction of 5hmC in various cancers (Jin et al. 2011; Margalit et al. 2020; Chen et al. 2016), both methods revealed a 3-fold decrease in 5hmC levels in the tumor compared to the adjacent normal tissue. This time, OGM detected higher absolute levels of 5hmC compared to ONT. We assume that the reason is an underestimation of 5hmC calls by the ONT model due to incomplete training sets and challenging sequence context. Similar conclusions are drawn from the 5hmC profile generated by both methods and displayed in Fig. 5.C. Albeit showing different absolute levels of 5hmC, the modulation of 5hmC level along gene bodies, as well as the increase in single as a function of gene expression, can be seen by both methods (**Figure 5.D**). **Figure 5.E.** shows an example of a large repetitive element containing a group of genes from the *GAGE* family, poorly represented in the hg38 reference (the entire array spans ~190 kbp in the reference, with a gap within these coordinates) (Nifker et al. 2023). Long molecules spanning the entire uncharacterized region in OGM aided in assembling a contig of the full repetitive element, and the 5hmC tags on these molecules provided the 5hmC profile along the unknown region. The panel also depicts a 5hmC-containing single molecule (digitized) and

the average 5hmC signal along the contig. Epigenetic characterization of this region by ONT was not possible due to the shorter molecules that could only penetrate several thousand bases into the ENCODE blacklist-masked *GAGE12* region (Amemiya et al. 2019).

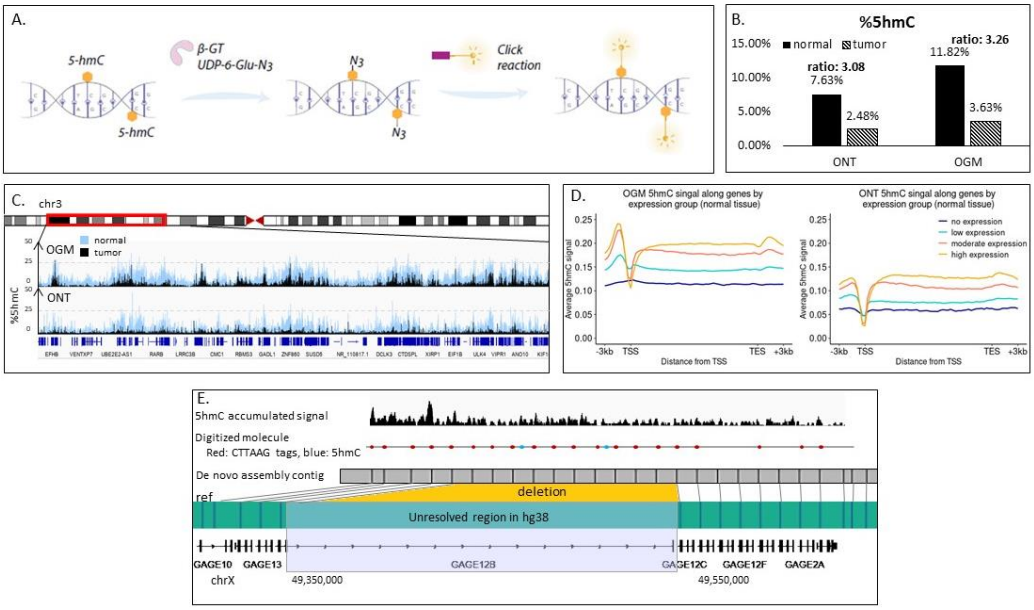

Figure 5. 5hmC analysis. **A.** Direct fluorescent labeling of 5hmC for OGM. **B.** Global 5hmC levels of a ccRCC tumor and a normal adjacent tissue, as detected in OGM and ONT. **C.** 5hmC profiles of a ccRCC tumor and a normal adjacent tissue generated by OGM and ONT along a ~25 Mbp region in chromosome 3. **D.** ONT and OGM 5hmC signal of the normal kidney tissue adjacent to the tumor along aggregated genes. Genes are grouped based on their expression in kidney tissues adjacent to ccRCC tumors. **E.** A repetitive sequence element in chromosome X, poorly characterized in the hg38 reference (green strip. Blue lines on it indicate the genetic barcode labels) above it, a *de novo* assembled OGM contig (gray) spanning the entire repeat array, indicating a deletion compared to hg38. The region spans genes from the *GAGE* family, and the gapped region contains the gene *GAGE12B*. Above it, a digitized single OGM molecule, with genetic barcode labels (red) and 5hmC labels (blue). Above it, is the average 5hmC signal along the contig.

Discussion

BNG and ONT now offer tools that aim to unveil the complexity of aberrated genomes and replace many cytogenetic workflows. Both companies have developed dedicated toolkits for variant calling. To navigate this evolving landscape, this report offers an objective comparison of the two methods, delving into the data types accessible with each technology, and the capabilities of their respective analytical tools, recognizing these tools as crucial for generating

reports with clear clinical relevance. In this respect, BNG is more clinically oriented in the cytogenetic space, with pipelines and reports that are aligned with clinical needs.

At the karyotype level, the methods conform, and both are capable of providing reliable copy number evaluations. As for detecting structural variations in variable sizes, a trade-off between resolution and read-length was observed: both methods called a high percentage of mid-range SVs (1-10 Kbp) ONT, owing to its single-base resolution, performed better in detecting small SVs (500 - 1000 bp), with over 30% of SVs detected only by ONT (and ~30,000 SVs below 500 bp were detected). OGM on the other hand, showed a clear advantage in detecting large SVs (larger than 10 kbp) with ~70% of large SVs, including translocations, not detected by ONT. Despite the low information content of OGM, the high coverage and N50 of these measurements increase the chance of detecting challenging structural aberrations. ONT's resolution advantage is also significant for reporting the SVs' breakpoints at a single bp resolution. This gives access to single nucleotide polymorphisms and short insertions and deletions that OGM is blind to. On the other hand, large or complex SVs are challenging for ONT under the experimental N50 and coverage. Translocations, for example, are not reported by the Sniffles2-based ONT pipeline (Smolka et al. 2024), while OGM called three intrachromosomal translocations on chromosome 5 and one interchromosomal translocation between chromosomes 3 and 5.

Per dollar, the genomic coverage generated by OGM is higher than that of ONTs, potentially opening a window to detect low-frequency variants. However, in this experiment, we did not meet the recommended coverage for running BNG's "Rare variant pipeline*"* (300x is recommended for high sensitivity to low frequency variants (Bionano Genomics 2021)), therefore we performed the pipelines of "*De novo* assembly" and "variant annotation pipeline*"*, instead. The choice of analytical tools significantly influences the insights extracted from data generated by both methods. While we employed recommended tools optimized for our data type and coverage, we note that these tools have inherent limitations that potentially extend beyond purely technological constraints (Savara et al. 2021). As for epigenetics, ONT can now call methylated CpGs from native DNA, together with generation of genetic data, an obvious advantage compared to OGM. OGM users that seek methylation information have to fluorescently tag the epigenetic modifications prior to data acquisition. These additional labeling steps are not commercialized by BNG and are not supported by the company. Methylation mapping extent is confined by the ability of the methyltransferase enzyme selected for this procedure. The enzyme M.TaqI, described here, efficiently labels CpG sites nested within the TCGA motif (Sharim et al. 2019). This provides a reduced representation of the unmethylome. These recognition sites make up ~6% of the CpG sites in the human

reference, with correlating methylation states in many important regions of the genome (Sharim et al. 2019), but inherent reduced representation limitations apply. Additionally, the indirect labeling done by methyltransferase enzymes, pointing unmodified sites, can't distinguish methylation from other cytosine modifications and is subjected to labeling efficiency, thus is inferior to direct methylation calling. Our analysis showed that global trends persisted, while locus specific signals depend on TCGA representation. With 5hmC, the picture is slightly different to date. The fluorescent labeling added to OGM directly labels 5hmC residues and not complementary sites (Gabrieli et al. 2018). Nonetheless, it is also external and not supported by BNG. ONT, on the other hand, needs a comprehensive training model containing the modification in all possible sequence contexts to assure accurate calls. Such training set for 5hmC is challenging for *in-situ* production. In our comparison, we denoted lower 5hmC levels called by ONT compared to OGM, implying that the ONT model currently underestimates density of 5hmC and misses many of them.

To conclude, selecting the most suitable platform hinges on a clear understanding of the data requirements dictated by the clinical or research question. To this end, a thorough understanding of the data generated by each platform, alongside the strengths and limitations of their respective analytical toolkits is needed.

Methods

Patient clinically relevant information

Tumor and normal adjacent tissue were obtained in the course of radical nephrectomy performed in an 82-year-old male. Tumor was diagnosed histologically as ccRCC with morphological features of eosinophilic variant at pT3a stage. Tissues were stored from the time of surgery to analysis at -80°C (fresh-frozen sample).

Sample collection and handling was approved by institutional review boards in accordance with the declaration of Helsinki.

Extraction of high molecular weight DNA

Ultra-high molecular weight (UHMW) DNA for 5-hmC OGM and ONT analyses was extracted using *SP Tissue and Tumor DNA Isolation kit* (Bionano Genomics), according to the manufacturer's protocol. High molecular weight (HMW) DNA for OGM unmodified CpG analysis was extracted using *Animal Tissue Isolation kit* (Bionano Genomics) according to

Bionano Prep Animal Tissue DNA Isolation Soft Tissue/ Fibrous Tissue Protocol for normal tissue/ tumor, respectively.

Nanopore sequencing (Oxford Nanopore Technologies)

Samples were prepared for sequencing using Ligation Sequencing Kit V14 (SQK-LSK114, Oxford Nanopore Technologies, UK) according to protocol with a starting DNA amount of 1 µg. Whole genome sequencing was performed on a "P2-Solo" device using R10.4.1 Flow cells (FLO-PRO11, Oxford Nanopore Technologies).

Basecalling of raw POD5 files was performed using the ONT proprietary software (Dorado v 0.3.2, Oxford Nanopore Technologies, UK) with the model: "dna_r10.4.1_e8.2_400bps_hac_@v4.0.0_5mCG_5hmCG@v2.cfg". Reads were then aligned to the hg38 human reference genome using minimap2 v.2.24 (Li 2018). Bam output files were then merged, sorted and indexed using samtools version1.16.1 (Li et al. 2009). SVs, CNVs and methylation and hydroxymethylation locations were called by the "wf-human-variation" pipeline via EPI2ME software (Oxford Nanopore Technologies; https://github.com/epi2me-labs/wf-human-variation; (Ewels et al. 2020)). Analyses were performed on a Linux operating system (Ubuntu 22.04.3) with Nvidia's RTX 6000 GPU.

Optical Genome Mapping (OGM)

a. **Labeling DNA for genetic and 5hmC analysis**

To create the genetic barcode, 750 ng of UHMW DNA in two reaction tubes were each mixed with 5X DLE-buffer (to a final concentration of 1X), 1.5 µL of 20X DL-Green and 1.5 µL of DLE-1 enzyme (Bionano Genomics) in a total reaction volume of 30 µL. The reaction was incubated for 4 hours at 37°C. Then, 5hmC sites were labeled by the enzyme β-glucosyltransferase from T4 phage (T4-BGT) (Gabrieli et al. 2018). Magnesium chloride was added to 30 µL of DLE-labeled DNA to a final concentration of 9 mM. Then, the DNA was added to 4.5 µL of 10X NEBuffer 4 (New England Biolabs), uridine diphosphate-6-azideglucose (UDP-6-N3-Glu; (21)) in a final concentration of 50 µM, 30 units of T4 β-glucosyltransferase (New England Biolabs) and ultra-pure water in a final volume of 45 µL. The reaction mixture was incubated overnight at 37°C. The following day, dibenzocyclooctyl (DBCO)-ATTO643 (Margalit et al. 2022) was added to a final concentration of 150 µM and the reaction was incubated again at 37°C overnight. The next day, the reaction tubes were added 5 µL of PureGene Proteinase K (Qiagen) and incubated for additional 30 minutes at 50°C.

After the Proteinase K treatment, the two identical reaction tubes were merged and drop-dialyzed as one against 20 mL of 1X TE buffer (pH 8) with 0.1 µm dialysis membrane for a total of 2 hours. Finally, 300 ng recovered dual-color DNA was stained to visualize DNA backbone, by mixing it with 4X Flow Buffer (Bionano Genomics) to a final concentration of 1X, 1M DTT (Bionano Genomics) to a final concentration of 0.1 M, Tris (pH 8) to a concentration of 25 mM, NaCl, to a concentration of 25 mM, EDTA to a final concentration of 0.008-0.01 M, DNA Stain (Bionano Genomics) to a final vol/vol ratio of 8%, and ultrapure water. The reaction mixture was shaken horizontally on a HulaMixer for an hour and then incubated overnight at 4°C.

b. Labeling DNA for unmethylation analysis (reduced representation of unmodified cytosines in CpG context)

To create the genetic barcode, 1 µg of U/HMW DNA was mixed with 5X DLE-buffer (to a final concentration of 1X), 2 µL of 20X DL-Green and 2 µL of DLE-1 enzyme (Bionano Genomics) in a total reaction volume of 30 µL for 4 hours at 37°C, immediately followed by heat inactivation at 80°C for 20 minutes. Heat inactivation at these conditions degrades over 97% of the DL-Green cofactor, therefore preventing it from being incorporated by M.TaqI in the following reaction, and making the two reactions orthogonal. Then, unmodified cytosines in the recognition sequence TCGA were fluorescently labeled to perform reduced representation optical methylation mapping (ROM) (Sharim et al. 2019; Grunwald et al. 2015). Two 500 ng reaction tubes of DLE1-labeled DNA were each mixed with 4 µL of 10X CutSmart buffer (New England Biolabs), 60 µM of lab-made synthetic AdoYnATTO643 (Margalit et al. 2022), 1 µL of M.TaqI (10 units/µL; New England Biolabs) and ultrapure water in a total volume of 40 µL, and incubated for 5 hours at 65°C. Then, 5 µL of Puregene Proteinase K (Qiagen) were added and the reaction tube was incubated for additional 2 hours at 45°C. After the Proteinase K treatment, the two 500 ng reaction tubes were merged and drop-dialyzed as one against 20 mL of 1X TE buffer (pH 8) with 0.1 µm dialysis membrane for a total of 2 hours. Finally, 300 ng recovered dual-color DNA were stained to visualize DNA backbone by mixing it with 15 µL of 4X Flow Buffer (Bionano Genomics), 6 µL of 1M DTT (Bionano Genomics), 3 µL of 0.5M Tris (pH 8), 3 µL of 0.5M NaCl, 4.8 µL of DNA Stain (Bionano Genomics) and ultrapure water to a total volume of 60 µL, and incubated overnight at 4°C.

c. Running OGM

Labeled samples were loaded on Saphyr chips (G1.2) and run on a Saphyr instrument (Bionano Genomics) to generate single molecule maps. Data were analyzed using Bionano Access software (v1.6.1) with Bionano Solve (v3.6.1) tool. *De novo* assemblies were generated with default parameters for human genomes. Variant annotation pipeline (single sample mode) for SV annotation was performed using Bionano Access v1.7.1 and Bionano Solve v3.7.1. The *in-silico* digested human genome GRCh38 (*hg38_DLE1_0kb_0labels.cmap*) was used as the reference.

d. CNV analysis

In Order to generate CNV plots, the coverage of DLE-1 labeling sites was extracted from raw output of CNV analysis (*cnv_rcmap_exp.txt*). Genomic regions with very high variance in coverage across Bionano Genomics' control datasets compared to typical loci (*hg38_cnv_masks.bed*) were subtracted from analysis. Then, the mean coverage of such sites in 500,000 bp bins was calculated using Bedtools *map* (v2.26.0; (Quinlan and Hall 2010)). Then, for each bin, the $\log_2$ of the copy ratio (in a diploid organism, copy number/2) was calculated and plotted along the chromosomes. A running median over 10 bins was calculated to plot a smooth red line across the copy number dots.

e. Epigenetic analysis

Optical mapping data from several runs were merged to a single dataset using Bionano Access (v1.6.1) and Bionano Solve (v3.6.1). Prior to alignment, the assigned channels for genetic and epigenetic labels in the molecules (.BNX) files were swapped with Bionano Solve according to manufacturer's advice. Molecules spanning over 150 kbp were then aligned to the *in silico* human genome reference GRCh38, based on DLE-1 recognition sites (hg38_DLE1_0kb_0labels.cmap), with default parameters according to the following combination of arguments: haplotype, human, DLE-1, Saphyr. Only molecules with an alignment confidence equal to or higher than 15 (P <= 10-15) that at least 60% of their length was aligned to the reference were used for downstream analysis. Alignment outputs were converted to global epigenetic profiles (bedgraph files) according to the pipeline described by (Gabrieli et al. 2018) and (Sharim et al. 2019) and in ebensteinLab/Irys-data-analysis on Github. Only regions covered by at least 20 molecules were considered. The average epigenetic score in each genomic position was calculated as the number of detected epigenetic labels in the position divided by the total number of molecules covering the position.

Global epigenetic levels

The weighted mean of epigenetic signals in non-overlapping 1000 bp genomic windows (generated using Bedtools *makewindow* (v2.26.0; (Quinlan and Hall 2010)) was calculated using Bedops *bedmap* (v2.4.41, (Neph et al. 2012)). The average of all genomic windows was reported as the global epigenetic level. To match the reported measure between OGM and ONT in methylation calling, the unmethylation level (1 – methylation level) was calculated from ONT mC level.

ONT and OGM common and unique SVs

To determine common and unique SVs to each method, the SVs called by both methods were compared. High-confidence SVs (confidence score ≥ 0.5) detected by OGM, and SVs residing on canonical chromosomes called by ONT were included in the comparison. SVs overlapping BNG's list of N-base gaps in the reference or putative false positive translocation breakpoints (for "*de novo* assembly", Solve 3.6.1) were masked from analysis. Coordinates of OGM-detected SVs were extended by 500 bp up and downstream to account for possible differences in SV resolution between the methods (Bedtools *slop* (v2.26.0; (Quinlan and Hall 2010)). This extension was not considered to determine SV size. Overlap between SVs detected by both methods was calculated with Bedtools *intersect*. The number of overlapping SVs is reported based on the ONT calls.

Gene expression data

Publicly available RNA-seq data of normal tissues adjacent to ccRCC tumors (stage 3) from three patients (PRJNA396588, GEO accessions: GSM2723919, GSM2723927, GSM2723929 (Yao et al. 2017)) were aligned to the human genome (hg38) using TopHat (v2.1.0, (Trapnell et al. 2009)) with default parameters and library-type and fr-firststrand flags, after retrieving the raw files with NCBI SRA toolkit (Leinonen et al. 2011). Only uniquely mapped reads were analyzed (minimal mapping quality of 30). Gene counts were obtained using HTSeq (htseq-count, v0.11.3, (Anders et al. 2015)) against the GENCODE v34 (Frankish et al. 2019) reference gene models. Transcripts per million (TPM) scores were calculated.

Epigenetic modifications signals along aggregated genes

Transcription start and end sites (TSS and TES) of protein-coding genes were defined according to GENCODE annotation (v34, (Frankish et al. 2019)). Protein-coding genes were divided into four groups based on their average normalized TPM score in the RNA-seq of three

matched ccRCC pairs. Unexpressed genes were defined as genes with TPM value <= 0.01 (~3000 genes). The other expression groups are three equal quantiles of the expressed protein-coding genes (~6000 genes per group). Mean 5-hmC signals along genes were calculated using DeepTools *computeMatrix* (v3.5.4, (Ramírez et al. 2014)) in scale-regions mode, where each gene was scaled to 15 kbp and divided into 300 bp bins. Compressed matrix output was summarized by DeepTools *plotProfile*.

Funding

This work was supported by the European Research Council consolidator [grant number 817811]; Israel Science Foundation [grant number 771/21]; The National Institute of health/The National Human Genome Research Institute (NIH/NHGRI) [grant number R01HG009190].

References

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**.

Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**: 1–5.

Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Avraham S, Schütz L, Käver L, Dankers A, Margalit S, Michaeli Y, Zirkin S, Torchinsky D, Gilat N, Bahr O, et al. 2023. Chemo-Enzymatic Fluorescence Labeling Of Genomic DNA For Simultaneous Detection Of Global 5-Methylcytosine And 5-Hydroxymethylcytosine**. *ChemBioChem* **24**.

Bionano Genomics. 2023a. Bionano Genomics Ordering Guide 2023. *Bionano Genomics website*.

Bionano Genomics. 2023b. Bionano Genomics website. *Prod Sheet - Optim Sample Prep Opt Genome Mapp Simpl Work Opt Genome Mapp*. https://bionano.com/wp-content/uploads/2023/01/BNG-23-087-Sample-Prep-Product-Sheet-Update_3.0_DIGITAL.pdf.

Bionano Genomics. 2023c. Bionano Genomics Website. *Reveal More Genomic Var That Matters With Opt Genome Mapp*. https://bionano.com/wp-content/uploads/BNG-23-064-Saphyr-Brochure-Update-2023_6.0_DIGITAL.pdf (Accessed March 18, 2024).

Bionano Genomics. 2021. Bionano Solve Theory of Operation: Structural Variant Calling. *Bionano Genomics website* 30110K. https://bionano.wpenginepowered.com/wp-

content/uploads/2022/05/30110_Rev.L_Bionano-Solve-Theory-of-Operation-Structural-Variant-Calling.pdf (Accessed March 18, 2024).

Chen S, Dou Y, Zhao Z, Li F, Su J, Fan C, Song S. 2016. High-Sensitivity and High-Efficiency Detection of DNA Hydroxymethylation in Genomic DNA by Multiplexing Electrochemical Biosensing. *Anal Chem* **88**: 3476–3480.

Chen Z, Pham L, Wu TC, Mo G, Xia Y, Chan PL, Porter D, Phan T, Che H, Tran H, et al. 2020. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome Res* **30**: 898–909.

Cosenza MR, Rodriguez-Martin B, Korbel JO. 2022. Structural Variation in Cancer: Role, Prevalence, and Mechanisms. *Annu Rev Genomics Hum Genet* **23**: 123–152.

Cui C, Shu W, Li P. 2016. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Front Cell Dev Biol* **4**: 89.

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Tommaso P Di, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**: 276–278.

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766–D773.

Gabrieli T, Michaeli Y, Avraham S, Torchinsky D, Margalit S, Sch¨utz L, Juhasz M, Coruh C, Arbib N, Zhou ZS, et al. 2022. Chemoenzymatic labeling of DNA methylation patterns for single-molecule epigenetic mapping. *Nucleic Acids Res*.

Gabrieli T, Sharim H, Nifker G, Jeffet J, Shahal T, Arielly R, Levy-sakin M, Hoch L, Arbib N, Michaeli Y, et al. 2018. Epigenetic Optical Mapping of 5- Hydroxymethylcytosine in Nanochannel Arrays. *ACS Nano* **12**: 7148–7158.

Grunwald A, Dahan M, Giesbertz A, Nilsson A, Nyberg LK, Weinhold E, Ambjörnsson T, Westerlund F, Ebenstein Y. 2015. Bacteriophage strain typing by rapid single molecule analysis. *Nucleic Acids Res* **43**: e117–e117.

Heck C, Michaeli Y, Bald I, Ebenstein Y. 2019. Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. *Curr Opin Biotechnol* **55**: 151–158.

Jeffet J, Margalit S, Michaeli Y, Ebenstein Y. 2021. Single-molecule optical genome mapping in nanochannels: multidisciplinarity at the nanoscale Jonathan. *Essays Biochem*.

Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, Krex D, Lu Q, Pfeifer GP. 2011. 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* **71**: 7360–7365.

Jonasch E, Walker CL, Rathmell WK. 2021. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. *Nat Rev Nephrol* **17**: 245–261.

Klatte T, Rao PN, De Martino M, Larochelle J, Shuch B, Zomorodian N, Said J, Kabbinavar FF, Belldegrun AS, Pantuck AJ. 2009. Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J Clin Oncol* **27**: 746–753.

Le VH, Hsieh JJ. 2018. Genomics and genetics of clear cell renal cell carcinoma: a mini-review. *J Transl Genet Genomics*.

Leinonen R, Sugawara H, Shumway M. 2011. The sequence read archive. *Nucleic Acids Res* **39**: 2010–2012.

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lixing Yang. 2020. A practical guide for structural variation detection in human genome. *Curr Protoc Hum Genet* **107**: e103.

Mantere T, Kersten S, Hoischen A. 2019. Long-read sequencing emerging in medical genetics. *Front Genet* **10**: 426.

Margalit S, Abramson Y, Sharim H, Manber Z, Bhattacharya S, Chen Y-W, Vilain E, Barseghyan H, Elkon R, Sharan R, et al. 2021. Long reads capture simultaneous enhancer–promoter methylation status for cell-type deconvolution. *Bioinformatics,* **37**: i327–i333.

Margalit S, Avraham S, Shahal T, Michaeli Y, Gilat N, Magod P, Caspi M, Loewenstein S, Lahat G, Friedmann-Morvinski D, et al. 2020. 5-Hydroxymethylcytosine as a clinical biomarker: Fluorescence-based assay for high-throughput epigenetic quantification in human tissues. *Int J Cancer* **146**: 115–122.

Margalit S, Tulpová Z, Michaeli Y, Zur TD, Deek J, Louzoun-Zada S, Nifker G, Grunwald A, Scher Y, Schütz L, et al. 2022. Optical Genome and Epigenome Mapping of Clear Cell Renal Cell Carcinoma. *bioRxiv*.

Marks P, Garcia S, Barrio AM, Belhocine K, Bernate J, Bharadwaj R, Bjornson K, Catalanotti C, Delaney J, Fehr A, et al. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Res* **29**: 635–645.

Mathew MT, Babcock M, Hou YCC, Hunter JM, Leung ML, Mei H, Schieffer K, Akkari Y. 2024. Clinical Cytogenetics: Current Practices and Beyond. *J Appl Lab Med* **9**: 61–75.

Michaeli Y, Shahal T, Torchinsky D, Grunwald A, Hoch R, Ebenstein Y. 2013. Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem Commun (Camb)* **49**: 8599–601.

Moore LE, Jaeger E, Nickerson ML, Brennan P, De Vries S, Roy R, Toro J, Li H, Karami S, Lenz P, et al. 2012. Genomic copy number alterations in clear cell renal carcinoma:

Associations with case characteristics and mechanisms of VHL gene inactivation. *Oncogenesis* **1**: 1–9.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. 2012. BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**: 1919–1920.

Nifker G, Grunwald A, Margalit S, Tulpova Z, Michaeli Y, Har-Gil H, Maimon N, Roichman E, Schütz L, Weinhold E, et al. 2023. Dam Assisted Fluorescent Tagging of Chromatin Accessibility (DAFCA) for Optical Genome Mapping in Nanochannel Arrays. *ACS Nano* **17**: 9178–9187.

Nifker G, Levy-Sakin M, Berkov-Zrihen Y, Shahal T, Gabrieli T, Fridman M, Ebenstein Y. 2015. One-pot chemoenzymatic cascade for labeling of the epigenetic marker 5-hydroxymethylcytosine. *ChemBioChem* **16**: 1857–1860.

ONT. 2024a. PromethION. *Website, Oxford Nanopore Technol*. https://nanoporetech.com/sites/default/files/s3/literature/BR_1204(EN)_V2_24Jan2024_FAW_DIGITAL-SPREADS.pdf.

ONT. 2024b. WORKFLOW: EPIGENETICS - Calling methylation in the human genome with PCR-free nanopore sequencing. *Oxford Nanopore Technol website*. https://nanoporetech.com/sites/default/files/s3/literature/epigenetics-workflow.pdf (Accessed March 18, 2024).

Ozkan E, Lacerda M. 2023. Genetics, Cytogenetic Testing And Conventional Karyotype. *StatPearls StatPearls Publ Treasure Isl* PMID: 33085440.

Pang AWC, Kosco K, Sahajpal NS, Sridhar A, Hauenstein J, Clifford B, Estabrook J, Chitsazan AD, Sahoo T, Iqbal A, et al. 2023. Analytic Validation of Optical Genome Mapping in Hematological Malignancies. *Biomedicines* **11**.

Quddus M, Pratt N, Nabi G. 2019. Chromosomal aberrations in renal cell carcinoma: An overview with implications for clinical practice. *Urol Ann* **11**: 6–14.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–2.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: 187–191.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genome Viewer. *Nat Biotechnol* **29**: 24–6.

Savara J, Novosád T, Gajdoš P, Kriegová E. 2021. Comparison of structural variants detected by optical mapping with long-read next-generation sequencing. *Bioinformatics* **37**: 3398–3404.

Shahal T, Gilat N, Michaeli Y, Redy-keisar O, Shabat D, Ebenstein Y. 2014. Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA. *Anal Chem* **86**: 8231–8237.

Sharim H, Grunwald A, Gabrieli T, Michaeli Y, Margalit S, Torchinsky D, Arielly R, Nifker G, Juhasz M, Gularek F, et al. 2019. Long-read single-molecule maps of the functional methylome. *Genome Res* **29**: 646–656.

Shi DQ, Ali I, Tang J, Yang W-C. 2017. New insights into 5hmC DNA modification: Generation, distribution and function. *Front Genet* **8**: 100.

Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, Kalef-Ezra E, Gandhi M, Hong K, Pehlivan D, et al. 2024. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol*.

Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C, Zhang W, Jian X, et al. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. **29**: 68–72.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**: 1348–1365.

White LK, Hesselberth JR. 2022. Modification mapping by nanopore sequencing. *Front Genet* **13**: 1–14.

Yao X, Tan J, Lim KJ, Koh J, Ooi WF, Li Z, Huang D, Xing M, Chan YS, Qu JZ, et al. 2017. VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discov* **7**: 1284–1305.

Yuen ZWS, Srivastava A, Daniel R, McNevin D, Jack C, Eyras E. 2021. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat Commun* **12**: 1–12.