NAR Genomics and Bioinformatics, 2025, **7**, lqae190
https://doi.org/10.1093/nargab/lqae190
Advance access publication date: 7 January 2025
High Throughput Sequencing Methods

Long-read structural and epigenetic profiling of a kidney tumor-matched sample with nanopore sequencing and optical genome mapping

Sapir Margalit[1,2,†], Zuzana Tulpová[1,2,3,†], Tahir Detinis Zur[1,2], Yael Michaeli[1,2], Jasline Deek[1,2], Gil Nifker[1,2], Rita Haldar[1,2], Yehudit Gnatek[4], Dorit Omer[4], Benjamin Dekel[4,5,6], Hagit Baris Feldman[6,7], Assaf Grunwald[1,2,*] and Yuval Ebenstein[1,2,*]

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

[*]To whom correspondence should be addressed. Tel: +972 73 3804352; Email: uv@post.tau.ac.il
Correspondence may also be addressed to Assaf Grunwald. Tel: +972 73 3804352; Email: assafgru@mail.tau.ac.il
[†]The first two authors should be regarded as Joint First Authors.

Abstract

Carcinogenesis often involves significant alterations in the cancer genome, marked by large structural variants (SVs) and copy number variations (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 (OGM) and nanopore sequencing [Oxford Nanopore Technologies (ONT)] bridge this resolution gap and offer enhanced performance for cytogenetic applications. Additionally, both methods can capture epigenetic information as they profile native, individual DNA molecules. We compared the effectiveness of the two methods in characterizing the structural, copy number and epigenetic landscape of a clear cell renal cell carcinoma tumor. Both methods provided comparable results for basic karyotyping and CNVs, but differed in their ability to detect SVs of different sizes and types. ONT outperformed OGM in detecting small SVs, while OGM excelled in detecting larger SVs, including translocations. Differences were also observed among various ONT SV callers. Additionally, both methods provided insights into the tumor's methylome and hydroxymethylome. While ONT was superior in methylation calling, hydroxymethylation reports can be further optimized. Our findings underscore the importance of carefully selecting the most appropriate platform based on specific research questions.

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 variants (SVs) and copy number variations (CNVs). 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 (1). Traditionally, cytogenetic techniques such as karyotyping are used to detect large SVs ranging from whole chromosome duplications, chromosome arm deletions, and down to SVs of ∼5–10 mega basepairs (Mb) (2,3). Applying fluorescence *in-situ* hybridization techniques may bring the resolution down to several hundred kilo basepairs (kb) (4), 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 kb 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 kb scale (5). Another concept involves a library preparation technique that allows linking proximal DNA fragments computationally by sequence barcode ligation [linked reads-10× genomics (6)/TELseq (7)]. Here, we utilized two techniques, optical genome mapping (OGM) and nanopore 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.

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 kb in length (8,9). 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

Received: July 28, 2024. **Revised:** December 12, 2024. **Editorial Decision:** December 23, 2024. **Accepted:** December 31, 2024

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* (10).

Oxford Nanopore Technologies (ONT) is another prominent player in the long-read mapping and sequencing space. Recent advancements in flow cell chemistries, purification kits for ultra-high molecular weight (UHMW) DNA, and basecalling algorithms have significantly extended read lengths and reduced error rates. Additionally, throughput has increased, and the cost per genome has decreased to levels that may justify clinical applications (11–13). 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 (10,14–17). Commercially-supported OGM only provides tools for analyzing genomic structure for cytogenetic applications. However, some of the BNG Saphyr systems contain three laser colors, 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. 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 (18). Nevertheless, accurate modification calling requires a complete training set, which is not trivial for most base modifications. The current recommended basecaller for ONT, Dorado (https://github.com/nanoporetech/dorado), can call modified bases, and has ready-to-use models to call 5-methyl cytosine (5mC) and 5-hydroxymethyl cytosine (5hmC), on top of the four canonical bases.

Unmethylated CpG sites, complementary to methylated (5mC) sites, can be specifically labeled by methyltransferase enzymes for optical mapping. Our group has recently applied engineered CpG methyltransferases to address all unmethylated CpGs (19,20). However, the method was not yet validated for human methylome profiling and thus we selected the previously validated reduced representation optical methylation mapping (ROM) (15,17,21). 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 (15,17).

5hmC, the first oxidation product of 5mC, is another important modification that was linked to gene regulation, development and disease, predominantly cancer (22,23). Optical mapping of 5hmC was introduced several years ago based on the fluorescent labeling of 5hmC residues (16,21).

Identification of modifications in ONT data relies on machine learning techniques. This process involves training and validating models using reference data encompassing the modification across diverse sequence contexts. Such reference data can be obtained by identifying the modification using established methods or *in-situ* approaches (18). However, obtaining high-quality, genome-wide reference data specifically for 5hmC modifications remains a significant challenge due to its cost and complexity. This, in turn, limits the ability to comprehensively train and assess the performance of 5hmC callers for ONT data, and it has not been benchmarked and peer-reviewed to date.

Herein, we compared the ability of both methods to characterize the structural, copy number and epigenetic landscape of a clear cell renal cell carcinoma (ccRCC) tumor and a matched normal adjacent sample. 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 (24–26). Other frequently observed CNVs 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 (24,27,28). Epigenetic alterations, including aberrant levels of 5mC and 5hmC, are also commonly observed in ccRCC tumors (29–31). In light of these hallmarks, this technical comparison identifies the specific strengths and limitations of each technique, highlighting practical differences that should be considered when selecting the appropriate method for addressing specific research questions.

Materials and 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 (ONT)

Samples were prepared for sequencing using Ligation Sequencing Kit V14 (SQK-LSK114, ONT, 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, ONT).

Basecalling of raw POD5 files was performed using the ONT proprietary software Dorado (v 0.3.2, ONT; https://github.com/nanoporetech/dorado) 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 (32) (v.2.24). Bam output files were then merged, sorted and indexed using samtools (33) (v1.16.1). SVs, CNVs and methylation and hydroxymethylation locations were called by the 'wf-human-variation' pipeline (https://github.com/epi2me-labs/wf-human-variation) via EPI2ME software (34) (ONT) with minimum bam coverage set to 5. The default behavior of the pipeline is to report methylation and hydroxymethylation per CpG positions and with combined strands. Analyses were performed on a Linux operating system (Ubuntu 22.04.3) with Nvidia's RTX 6000 GPU.

Optical genome mapping

Samples were labeled by Direct Label and Stain (DLS) chemistry (DLE-1 enzyme, Bionano Genomics), creating a genetic barcode (CTTAAG motive). To create the genetic barcode for 5hmC analysis, 750 ng of UHMW DNA in two reaction tubes were each mixed with 5× DLE-buffer (to a final concentration of 1×), 1.5 μl of 20× 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 h at 37°C. Then, 5hmC sites were labeled by the enzyme β-glucosyltransferase from T4 phage (T4-BGT) (16). 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 10× 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 ultrapure water in a final volume of 45 μl. The reaction mixture was incubated overnight at 37°C. The following day, dibenzocyclooctyl (DBCO)-ATTO643 (21) 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 min at 50°C. After the Proteinase K treatment, the two identical reaction tubes were merged and drop-dialyzed as one against 20 ml of 1× TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) with 0.1 μm dialysis membrane for a total of 2 h. Finally, 300 ng recovered dual-color DNA was stained to visualize DNA backbone, by mixing it with 4× Flow Buffer (Bionano Genomics) to a final concentration of 1×, 1 M DTT (DL-Dithiothreitol; 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, ethylenediaminetetraacetic acid (EDTA) to a final concentration of 0.008–0.01 M, DNA Stain (Bionano Genomics) to a final *v/v* ratio of 8%, and ultrapure water. The reaction mixture was shaken horizontally on a HulaMixer for 1 h and then incubated overnight at 4°C.

To create the genetic barcode for unmethylation analysis, 1 μg of HMW DNA was mixed with 5× DLE-buffer (to a final concentration of 1×), 2 μl of 20× DL-Green and 2 μl of DLE-1 enzyme (Bionano Genomics) in a total reaction volume of 30 μl for 4 h at 37°C, immediately followed by heat inactivation at 80°C for 20 min. 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 (15,35)]. Two 500 ng reaction tubes of DLE1-labeled DNA were each mixed with 4 μl of 10× CutSmart buffer (New England Biolabs), 60 μM of lab-made synthetic AdoYnATTO643 (21), 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 h at 65°C. Then, 5 μl of Puregene Proteinase K (Qiagen) were added and the reaction tube was incubated for additional 2 h 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 1× TE buffer (pH 8) with 0.1 μm dialysis membrane for a total of 2 h. Finally, 300 ng recovered dual-color DNA were stained to visualize DNA backbone by mixing it with 15 μl of 4× Flow Buffer (Bionano Genomics), 6 μl of 1 M DTT (Bionano Genomics), 3 μl of 0.5 M Tris (pH 8), 3 μl of 0.5 M 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.

Labeled samples were loaded on Saphyr chips (G1.2) and run on a Saphyr instrument (Bionano Genomics) to generate single molecule maps. Optical mapping data from several runs were merged to a single dataset using Bionano Access (v1.6.1) and Bionano Solve (v3.6.1) (Bionano Genomics). The assigned channels for genetic and epigenetic labels in the molecules (.BNX) files were swapped with Bionano Solve (v3.6.1) according to manufacturer's advice. *De novo* assemblies and 'variant annotation pipeline' (single sample mode) for SV annotation were generated from 5hmC-labeled data with default parameters for human genomes 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. For epigenetic data processing, molecules spanning over 150 kb were aligned to the *in silico* human genome reference GRCh38, based on DLE-1 recognition sites (hg38_DLE1_0kb_0labels.cmap) using Bionano Access (v1.6.1) and Bionano Solve (v3.6.1), 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 ($P \leq 10^{-15}$) 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.* (16) and Sharim *et al.* (15) and in ebensteinLab/Irys-data-analysis on Github. Only regions covered by at least 20 molecules were considered.

CNV analysis

In order to generate CNV plots of OGM data, 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 (36) *map* (v2.26.0). 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. $\log_2$ of the copy ratio in 500 000 bp bins along ONT data was inferred by employing the EPI2ME workflow 'wf-human-variation' to each sample. A running median over 10 bins was calculated to plot a smooth red line across the $\log_2$ of the copy ratio dots of both methods.

SV analysis

Genomic coordinates, SV type and size of annotated OGM SVs from 'variant annotation Pipeline' with confidence score meeting or exceeding the SV-specific BNG thresholds (0 for insertions and deletions, 0.7 for inversions, 0.3 for intra chromosomal translocations and 0.65 for inter chromosomal translocations. All duplications were considered, as they have undefined confidence scores), were extracted from output smap file and converted to bed format for downstream analysis. In case no end coordinate was supplied, it was taken as start + 1. Both translocation breakpoints were considered for overlap with ONT SVs, but were counted as one event of a large SV (>10 kb). Only unique SVs were kept. 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 [Bedtools *intersect -v* (v2.26.0)]. Coordinates of OGM-detected SVs were extended by 500 bp up and downstream [Bedtools *slop* (v2.26.0)] to account for possible differences in SV resolution between OGM and ONT. This extension was not considered to determine SV size.

Genomic coordinates, SV type and size of 'passed' ONT SVs called by the EPI2ME workflow 'wf-human-variation', with minimum five reads supporting them, residing on canonical chromosomes, were extracted from the Sniffles2 VCF output file and converted to bed format for downstream analysis. End coordinate was taken as end + 1 to avoid cases of 0 difference. Only unique SVs were kept. 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 [Bedtools *intersect -v* (v2.26.0)].

Overlap between SVs of the same type detected by OGM and ONT was calculated with Bedtools *intersect* (v2.26.0). The number and sizes of overlapping SVs is reported based on the ONT calls. Overlapping and non-overlapping SVs were then divided based on their size (absolute value).

AnnotSV (37,38) was used to annotate and interpret the clinical significance of ONT SVs.

Alternative ONT SV callers

SV calling with SVIM (39) (https://github.com/eldariont/svim) was conducted with parameter: –min_sv_size 50.

SV calling with CuteSV (40) (https://github.com/tjiangHIT/cuteSV) was conducted with parameters recommended for ONT data: –max_cluster_bias_INS 100; –max_cluster_bias_DEL 100; –diff_ratio_merging_INS 0.3; –diff_ratio_merging_DEL 0.3; and additional parameters: -l 50; –min_support 5.

SV calling with NanoVar (41) (https://github.com/benoukraflab/NanoVar) was conducted with parameters: –mincov 5; –minlen 50.

The output VCF files of all callers were processed as described for the EPI2ME (Sniffles2) VCF. Only SVs of at least 500 bp were considered for analysis.

Overlap between SVs detected by the different ONT callers was calculated as follows:

To assess the concordance of ONT SV callers, SVs of the same type called by any SV caller were combined. Overlapping SVs were merged [Bedtools *merge* (v2.26.0)], and the merged SVs were intersected with the individual caller's results [Bedtools *intersect –wa –wb –names* (v2.26.0)]. Results of all SV types were then combined and an UpSet plot was generated to visualize overlaps. The merged SVs were also intersected with the OGM SVs, by SV type.

Manual inspection of ONT inter-chromosomal translocation detected by OGM

To find ONT reads that support an inter-chromosomal translocation discovered by OGM, we listed all ONT reads aligned to extended regions around each of the OGM breakpoints (chr3:84 351 000–84 362 000; chr5:53 644 000–53 663 000). Then, we looked for read IDs that aligned to both regions.

Global epigenetic levels

Due to resolution differences between OGM and ONT, the mean epigenetic levels in non-overlapping 1000 bp genomic windows (generated using Bedtools *makewindow* (v2.26.0) was calculated. Only windows on canonical chromosomes that contain at least one relevant recognition site were considered (CpG for 5hmC and ONT mC, TCGA for OGM unmethylation; sites loci were extracted using the R package BSgenome (https://bioconductor.org/packages/release/bioc/html/BSgenome.html). To match the reported measure between OGM and ONT in methylation calling, the unmethylation level (1 − methylation level) was calculated from ONT methylation level. The weighted mean of all ONT epigenetic signals and ONT unmethylation signals in TCGA sites only [crossed with Bedtools *intersect* (v2.26.0)] in these genomic windows was calculated using Bedops *bedmap* (42) (v2.4.41). The number of OGM epigenetic labels and molecules covering each genomic window were counted using Bedtools *intersect* (v2.30.0). The average labels-to-molecules ratio across all windows was reported as the global epigenetic level for OGM.

To create bedgraphs of OGM signals, epigenetic labels and molecules were extended by 500 bp upstream and downstream to account for optical resolution [(Bedtools *slop* (v2.26.0)], prior to calculating the labels to molecules ratio in each genomic location. To reduce the resolution of OGM and ONT bedgraphs to 1 kb windows, the weighted mean of signal in these windows was calculated with Bedops *bedmap* (v2.4.41).

Gene aggregation according to gene expression data

Publicly available RNA-seq data (RPKM scores) of ccRCC tumors ($N = 419$) and normal adjacent tissues ($N = 31$) were adapted from The Cancer Genome Atlas Research Network (43). Entrez gene_id and hgnc gene symbol were used to as-

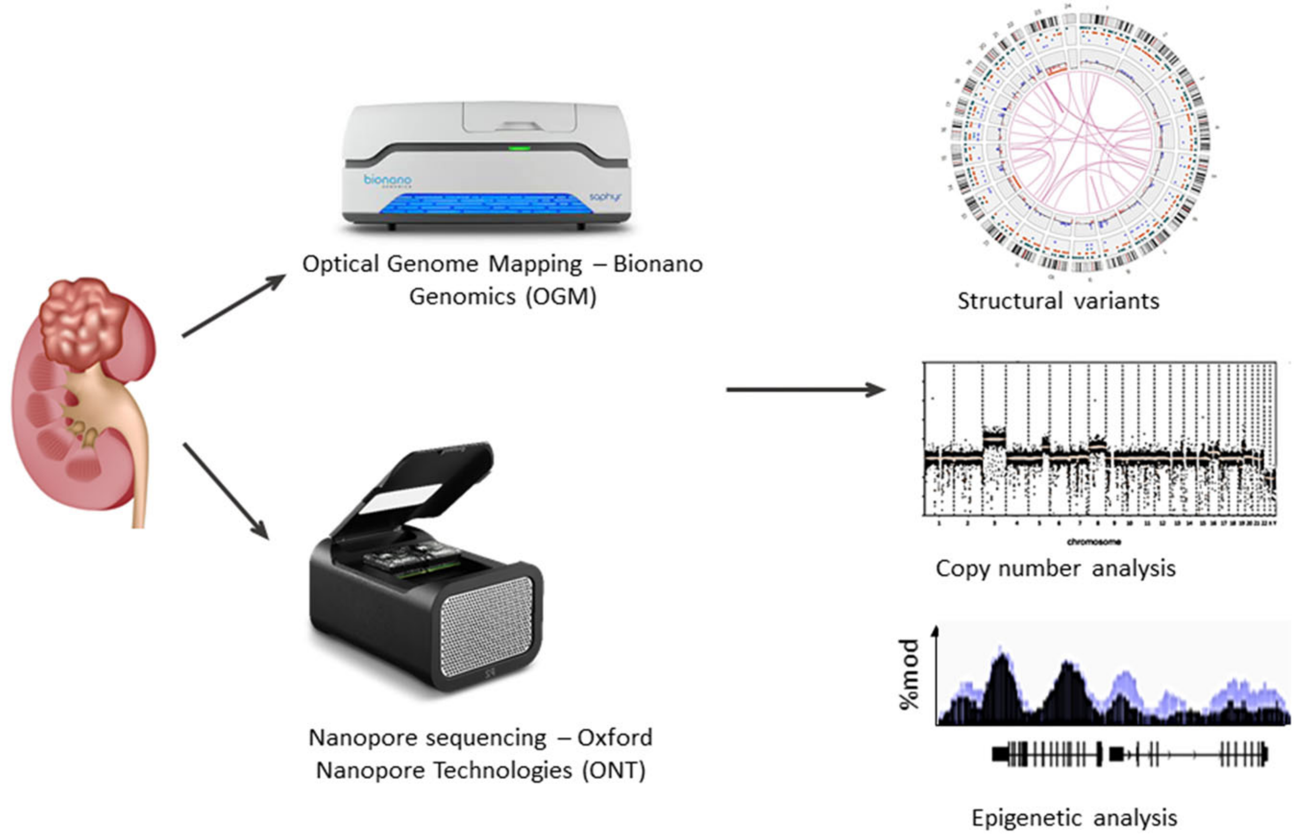
Optical Genome Mapping – Bionano Genomics (OGM)
Nanopore sequencing – Oxford Nanopore Technologies (ONT)
Structural variants
chromosome
Copy number analysis
%mod
Epigenetic analysis

Figure 1. Experimental workflow. U/HMW DNA was extracted from a ccRCC tumor and a normal adjacent kidney tissue. Samples were analyzed by OGM and ONT to detect SVs and CNVs, and epigenetic modifications. Results from both methods were compared.

sign gene attributes for hg38 using the R package biomaRt (44). Only genes on canonical chromosomes were taken. The mean RPKM level of each gene in the two groups was calculated and the genes were divided into three equal quantiles based on the mean. Mean 5-hmC and unmethylation signals along aggregated genes were calculated using DeepTools (45) *computeMatrix* (v3.5.4) in scale-regions mode, where each gene [from transcription start site (TSS) to end site (TES)] was scaled to 15 kb and divided into 300 bp bins. Compressed matrix output was summarized by DeepTools *plotProfile*. The average signal intensities for both markers were then plotted as a function of the scaled distance relative to the TSS.

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 (24,27), and its normal adjacent tissue. Our workflow consisted of extracting U/HMW DNA, followed by per-protocol OGM and ONT analyses (Figure 1).

By generating long reads, both methods unlock access to intricate areas of the genome, enabling the study of diverse SVs, CNVs 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.

To compare the efficacy of the structural profiling and data analysis processes offered by each method, we applied CNV and SV analyses on data generated by both methods, adhering to the manufacturer's recommended pipelines unless specified otherwise (see methods section). Herein, DNA from a ccRCC tumor and a normal adjacent tissue was analyzed using both ONT and BNG platforms. Table 2 summarizes the resulting N50 and average coverage. Supplementary Figure S1 shows read length histograms.

Strong agreement in CNV landscapes detected by OGM and ONT

First, genome-wide copy number, calculated in 500 kb bins, was compared. Tumor plots are shown in Figure 2 and normal adjacent tissue plots are shown in Supplementary Figure S2. A running median over 10 bins was calculated to plot a smooth red line across the copy number dots. 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 on chromosomes 7 and 12. OGM spotted a small DNA loss on chromosome 9 not reported by ONT. The normal adjacent sample did not exhibit any large CNV in both methods, suggesting somatic aberrations. The loss of 3p, gain in 5q and trisomy of chromosome 7 are well-documented genetic characteristics of ccRCC (27,28). The number of data points sampled in each bin in OGM is determined by the label density of the DLE motif (14–17 labels

Table 1. ONT and OGM specifications

	Bionano Genomics Saphyr	ONT PromethION
Resolution	Optical resolution: 500–1500 bp (10) SV detection resolution: starting from 500 bp for diploid genomes' insertions and deletions. inversions/duplications: >30 kb. translocations: >50 kb (46).	Single bp
Molecules N50*	250–400 kb, when including only molecules exceeding 150 kb (47)	10–50 kb for high throughput (48), and up to 150 kb with ultra-long sequencing (49)
Average human genome coverage per cell	80–300× (effective coverage of filtered (≥150 kb) and aligned molecules) (47)	16–66× (raw coverage) (48)
Price per sample (including cell, reagents and device rental)	550$ (450$), when buying a package for 120 (240) experiments (50)	1010$ (720$), when buying 'project pack' for 96 experiments with PromethION 2 Integrated ('project pack' for 1024 experiments with PromethION 24) (48)
Price per 1× human genome coverage	1.5$–7$	11$–65$
Methylation calling	Labeling of unmodified cytosines in CpG (19,20) or TCGA (15) sites can be added (unsupported)	Integrated (https://github.com/nanoporetech/dorado)
5hmC calling	Direct labeling of 5hmC can be added (unsupported) (16)	Integrated (https://github.com/nanoporetech/dorado)

*Molecules N50 is a measure of reads length indicating that half of the genetic data recorded came from reads longer or equal to this value.

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

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

per 100 kb on average, for human genomes), which is lower than the number of points in ONT (single bp resolution). Consequently, the OGM plot exhibits higher noise.

SV concordance depends on size and type

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 (in the tumor sample in Figure 3A and the normal adjacent tissue in Supplementary Figure S3A). The results reveal a clear trend: ONT detected smaller SVs, while OGM was more effective at detecting larger SVs. Specifically, in the tumor sample, only ONT reported SVs between 50 and 500 bp, with ∼11% of these overlapping with larger SVs detected by OGM. In the next size group (500–1000 bp), both technologies detected SVs, with a significant overlap: 60% (1143 SVs) were identified by both methods. However, ONT detected more unique SVs (565) compared to OGM (202). For SVs sized 1–5 kb, the overlap increased to 74% (1487 SVs), with ONT still detecting more unique SVs (334) than OGM (192). The larger SV groups contained fewer SVs overall. For SVs sized 5–10 kb, the overlap remained substantial at 65% (251 SVs), but OGM detected more unique SVs (97) compared to ONT (37). For SVs larger than 10 kb, OGM was the dominant technology, detecting 192 unique SVs compared to only 8 by ONT, with 40 SVs (17%) detected by both. The observed differences in SV detection can be attributed to the inherent strengths of each technology. OGM's larger N50 and higher coverage make it more effective for detecting larger SVs, while ONT's higher resolution allows it to better characterize smaller SVs.

Out of the OGM-detected SVs, 54 are not present in BNG's database of healthy controls, indicating possibly pathogenic SVs: 9 SVs between 500 bp and 1 kb, 19 SVs between 1 and 5 kb, 8 SVs between 5 and 10 kb and 18 SVs larger than 10 kb. A total of 25 of these rare SVs are not present in the normal adjacent tissue. Out of the ONT-detected SVs, 6 were classified as possibly pathogenic by the ACMG classification [using AnnotSV (37,38)]: one between 500 bp and 1 kb, and five are shorter, between 50 and 500 bp. The reported phenotype for these SVs is unrelated to ccRCC. These SVs are also present in the normal adjacent tissue. Five other SVs, sized between 50 and 500 bp and classified as 'variant of unknown significance' or 'NA' were annotated with a renal cell carcinoma OMIM phenotype. These SVs, or similar SVs within 100 bp proximity, are also present in the normal adjacent tissue.

Our analysis revealed differences in the types of SVs detected by the ONT pipeline and the OGM method (Figure 3B – in the tumor sample, Supplementary Figure S3B – in the normal adjacent sample). While the ONT pipeline, using Sniffles2 (51), only identified deletions (1364) and insertions (2501) in the tumor sample, OGM detected additional SV types, translocations (4) inversions (48) and duplications (35), alongside deletions (1221) and insertions (3229). This observation suggests potential limitations of Sniffles2 for certain SV types. Recent publications evaluating various SV callers within the ONT framework (52,53), suggest that alternative SV callers like SVIM (39), CuteSV (40) or NanoVar (41) may outperform Sniffles2 in detecting such SV types, when used after the same aligner (minimap2). Based on these findings, we employed SVIM, CuteSV and NanoVar for SV detection in the tumor sample and compared the results of all four ONT callers to the OGM results (Figure 3B and Supplementary Figure S4). Notably, NanoVar was the only SV caller that revealed translocations, inversions and duplications, as well as transpositions, not reported by OGM. Figure 3C shows the overlap

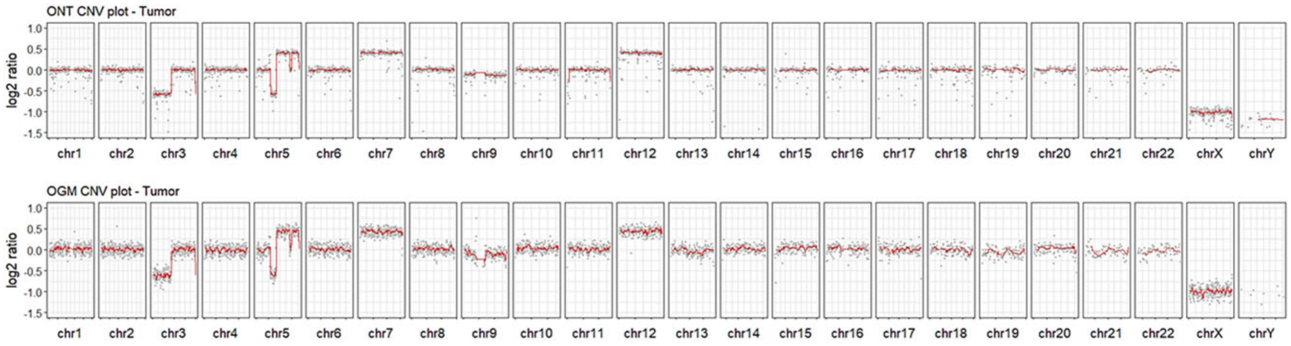

Figure 2. Comparative analysis of CNV in a ccRCC tumor, as detected by ONT and OGM. The plots show $\log_2$ of the copy ratio generated from ONT (top) and OGM (bottom) data. Data illustrate highly similar findings, pinpointing a significant DNA loss on chromosome 3 and various losses and gains on chromosomes 5, 7 and 12.

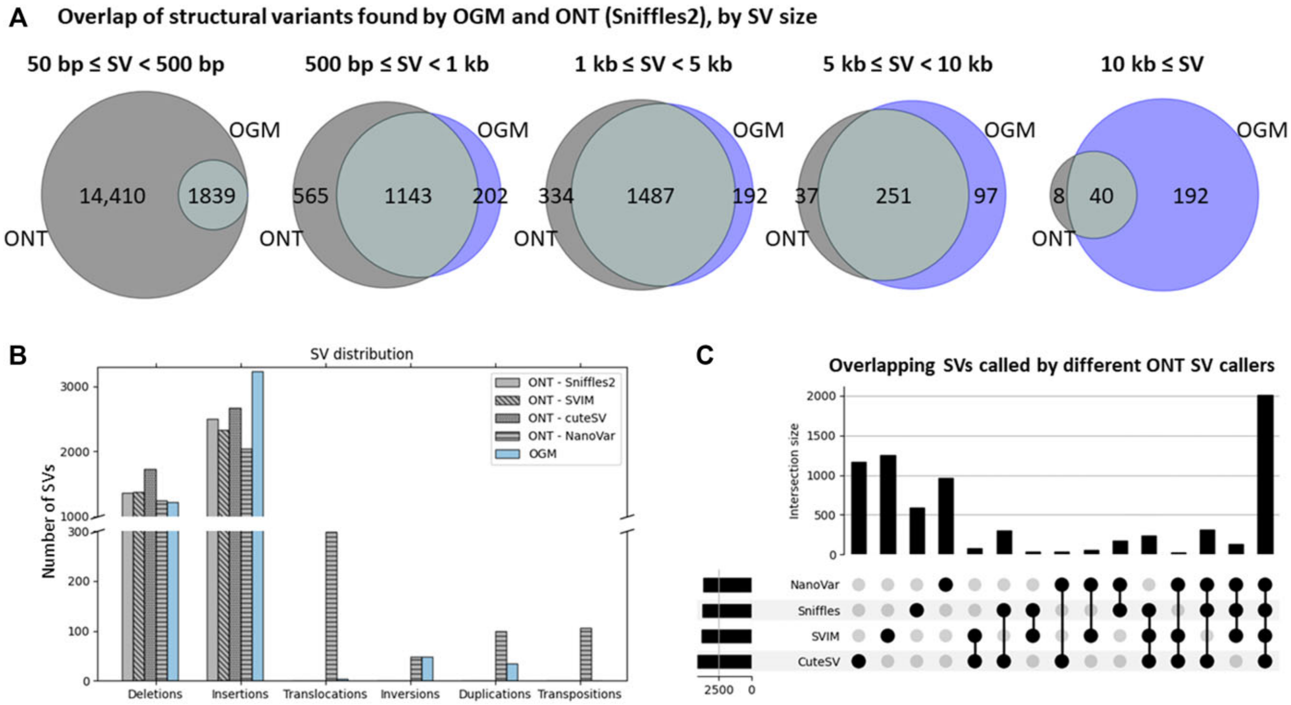

Figure 3. Comparative analysis of SVs in a ccRCC tumor, as detected by ONT and OGM. (**A**) Venn diagrams displaying common and unique SVs to OGM and ONT (Sniffles2), in five size ranges: 50–500 bp, 500 bp–1 kb, 1–5 kb, 5–10 kb and above 10 kb. (**B**) Comparison of number of SVs (≥500 bp) by type, detected by OGM and four SV callers for ONT – Sniffles2, SVIM, CuteSV and NanoVar. (**C**) UpSet plot displaying the overlap between SVs called by the different ONT SV callers.

between the SVs called by the different ONT callers. A total of 27% of the SVs were detected by all the four callers, and Sniffles2 detected less unique SVs than the other callers. Supplementary Figure S4 shows the overlap between OGM SVs and ONT SVs called by any of the SV callers in the tumor sample, by SV type.

Chromosomes 3 and 5 are frequently disrupted in ccRCC. Figure 4A illustrates these two chromosomes with marks indicating the relative positions of SVs and CNVs larger than 5 kb identified by ONT and OGM. 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 for both methods. Out of the OGM-detected SVs, uncommon SVs not present in BNG's database of healthy controls are separately plotted on the bottom. As seen also in Figure 2, the two methods detected DNA gain/loss events in these chromosomes and exhibited a high degree of concordance for insertions and deletions. OGM detected six possible inversions (similar locus), two duplications, three intrachromosomal and one inter-chromosomal (Figure 4B) translocation events. Notably, only 13 SVs identified by OGM in these chromosomes did not appear in BNG's database of mapped healthy controls, indicating possible pathogenic SVs. Two of them were also found by ONT, and 11 of them are potential somatic variants not found in the normal sample adjacent to the tumor (of these, none were found by ONT). ONT SVs found in all chromosomes in the ccRCC tumor and the normal adjacent tissue are shown in Supplementary Figure S5 (raw, unfiltered karyogram) and S6 (processed circos plots, including somatic SVs present in the tumor sample and not in the normal adjacent tissue). OGM SVs found in all chromosomes in the ccRCC tumor and the normal adjacent tissue are shown

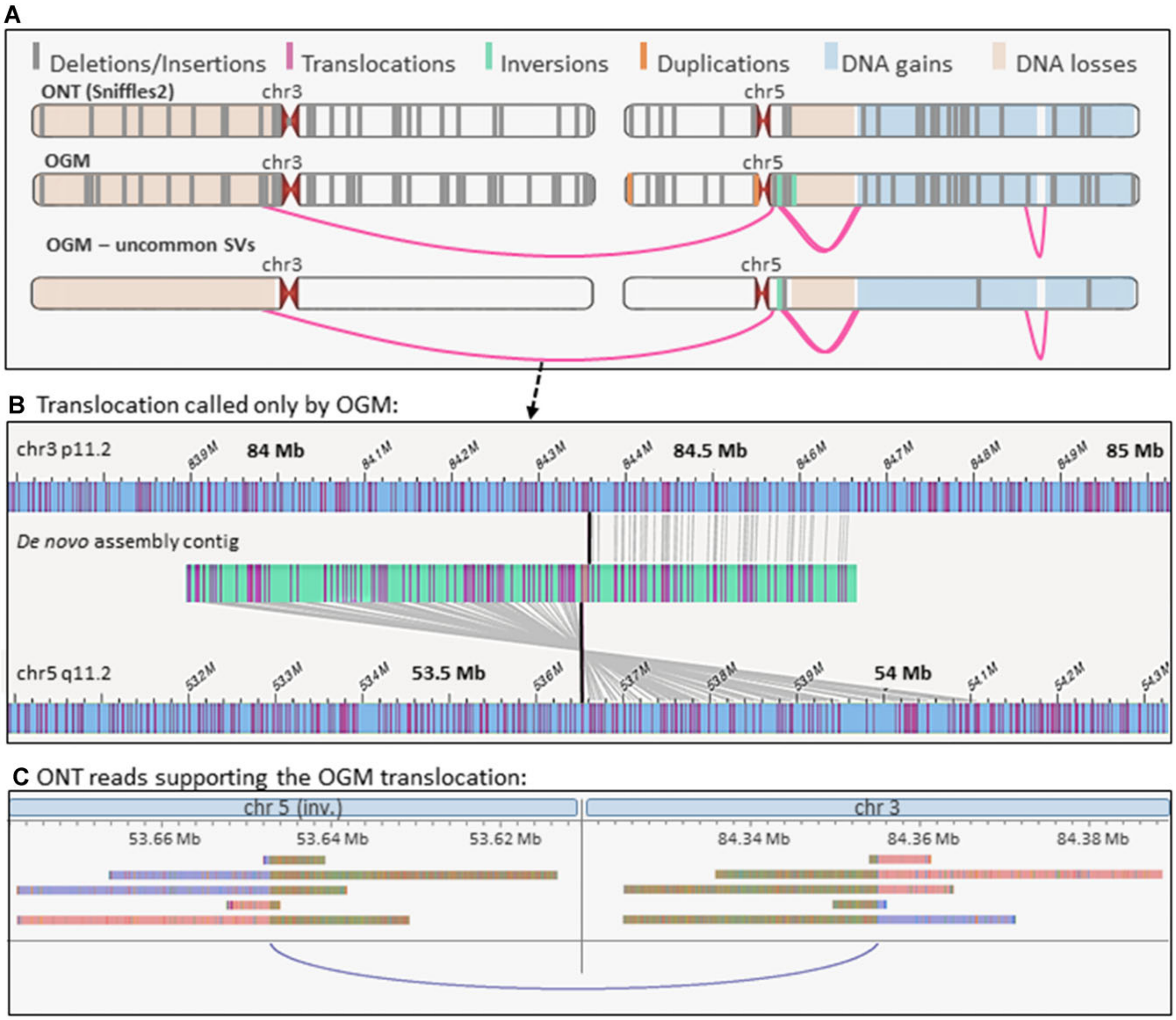

Figure 4. SVs detected by ONT and OGM. (**A**) Illustration of SVs larger than 5 kb on chromosomes 3 and 5, as detected by ONT (top panel) and OGM (middle panel). Bottom panel shows OGM SVs that do not appear in BNG's dataset of healthy controls, hence potentially pathogenic. (**B**) Inter-chromosomal translocation detected by OGM, and not reported by any of the tested ONT SV callers. The light blue strips at the top and bottom represent the reference chromosomes 3 (top) and 5 (bottom), and the middle strip is a *de novo* assembled contig, composed of fragments mapped to chromosome 3 and inverted chromosome 5. Pink lines indicate CTTAAG barcode labels in the contig and reference. Bold black lines indicate translocation breakpoints (gap is due to the method's resolution). (**C**) Manual investigation of ONT reads revealed five reads supporting the OGM-detected translocation, and identified the true breakpoints. The panel shows the breakpoint loci on chromosome 3 and on inverted chromosome 5, with five reads that aligned to both. Each read is shown twice in the same row, under each chromosome. The part of the read that aligns to the above chromosome is colored according to strand (blue or red) and displays only sparse mismatches labels (colorful lines). The arc below marks the translocation breakpoint revealed by these reads. Reads were visualized in Jbrowse2 (54).

in Supplementary Figure S7 (processed circos plots of all SVs, including somatic SVs present in the tumor sample and not in the normal adjacent tissue), S8 (processed circos plots of SVs not present in BNG's database of healthy controls, including somatic SVs present in the tumor sample and not in the normal adjacent tissue), and S9 (BNG default display summarizing all results. SVs plotted are not present in BNG's database of healthy controls).

The inter-chromosomal translocation discovered by OGM, depicted in Figure 4B, connects chromosome 3 and inverted chromosome 5. In OGM, it was supported by multiple molecules spanning the breakpoint (at least 20) and received a high confidence score (at least 0.98). The reported variant allele frequency for it is 0.28. However, none of the four inspected ONT SV callers confirmed it. Manual investigation of the ONT reads revealed five reads spanning the breakpoint and supporting it (Figure 4C). We note that a minimum threshold of five reads was selected for the SV callers' analysis, which means this SV could have been detected. These ONT reads, all sharing the same breakpoint, refine the 7.8 kb resolution gap around the OGM translocation breakpoint (seen between the two black lines in Figure 4B) to a bp level breakpoint at chr3:84 354 973; chr5:53 647 008. The calculated translocation frequency is 0.07 (five reads divided by the total number of reads aligned to the two breakpoint loci). The higher frequency reported by OGM can be attributed to the larger portion of long molecules in OGM (higher N50 and higher coverage).

Epigenetic analyses

Comparative analysis of methylation calls shows superior performance of ONT over OGM

While ONT can directly call methylation signals from native sequences using the appropriate Dorado basecalling model, OGM requires enzymatic attachment of a fluorophore to specific sites for detection. We employed reduced ROM (15,17,21) to tag unmodified CpG sites within specific sequence contexts. 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 (Figure 5A). Consequently, the DNA is labeled in all unmodified CpGs within TCGA sites.

To facilitate a direct comparison of methylation signals between ONT's direct methylation calling and ROM, 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 per site, requiring at least five reads for ONT (Supplementary Figure S10A) 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 kb genomic windows. We compared the entire ONT methylome (all CpG sites) as well a reduced ONT methylome (TCGA motif) to the reduced representation OGM signals (Figure 5B). Our analysis revealed a higher unmethylation signal in ONT compared to OGM. Interestingly, the difference persisted, and even slightly increased when we specifically analyzed TCGA-embedded CpG sites in ONT data. This suggests a potential underestimation of unmethylation by OGM, likely attributable to its lower optical resolution, rather than to the reduced representation approach. Consequently, multiple closely spaced TCGA sites might be erroneously merged into a single unit by OGM, leading to an underestimation of the overall unmethylation signal. Plots showing distances between adjacent TCGA sites, the number of TCGA sites in 1 kb windows, and this number versus the number of CpG sites in the same bins, are presented in Supplementary Figure S10B–D. Figure 5C shows that despite absolute intensity differences, similar trends are seen in the normalized unmethylation profile generated along genes when grouped by their gene expression score in ccRCC tumors (55). Both methods display the higher unmethylation signal around the TSS, which increases with gene expression. In contrast, the level in gene bodies is much lower and more similar among all expression levels, indicating highly modified gene bodies regardless of expression level. Un-normalized OGM and ONT levels of the tumor and normal adjacent tissue along genes are presented in Supplementary Figure S11. The relationship between the tumor-to-normal fold-change in genes unmethylation signal and the corresponding changes in gene expression is presented in Supplementary Figure S12. The resolution and methylation representation differences become more apparent when zooming in to smaller regions of the genome. Figure 5D shows the unmethylation profile of a ∼400 kb region on chromosome 22q11.21. Three representative examples for methylation comparison are marked in red boxes that contain variable TCGA content (shown in blue 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 SNP, creating a new TCGA site recognizable by the M.TaqI enzyme, thus explaining the observed OGM signal.

Hydroxymethylation profiling: similar trends, lower absolute levels in ONT compared to OGM

Optical mapping of 5hmC was compared to ONT 5hmC calls made using the appropriate Dorado basecalling model. For optical mapping, 5hmC is labeled through a process involving the enzymatic attachment of an azide-modified glucose moiety from a synthetic cofactor (56,57) (UDP-6-N3-Glu) to the hydroxyl group of 5-hmC, followed by a click reaction that connects a fluorophore-bound alkyne to the azide-labeled 5-hmC (58,59) (Figure 6A). Figure 6B shows the average genome-wide 5hmC signal in the ccRCC tumor sample and the normal adjacent tissue, as was detected by both methods. Consistent with published reports indicating a global reduction of 5hmC in various cancers (23,60,61), 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. As the labeling scheme used to tag 5hmC residues in the OGM experiment has no false positives, and was validated with liquid chromatography-tandem mass spectrometry (LC-MS/MS) in previous work (16), we hypothesize that there is an underestimation of 5hmC calls by the ONT model due to incomplete training sets and challenging sequence contexts. Albeit showing different absolute levels of 5hmC, the modulation of 5hmC level along gene bodies, as well as the increase in signal as a function of gene expression, can be seen by both methods (Figure 6C and Supplementary Figure S13). The relationship between the tumor-to-normal 5hmC signal fold-change in genes and the corresponding changes in gene expression is presented in Supplementary Figure S14. The 5hmC profile generated by both methods and displayed in Figure 6D reveals a broadly correlated profile, but with distinct amplitude variations between the different datasets, in line with the average global levels. Figure 6E 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 kb in the reference, with a gap within these coordinates) (62). 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 (63).

Discussion

BNG and ONT now offer tools that aim to unveil the complexity of aberrant genomes and replace many cytogenetic workflows. Both companies have developed dedicated toolkits for variant calling. To navigate this evolving landscape, this

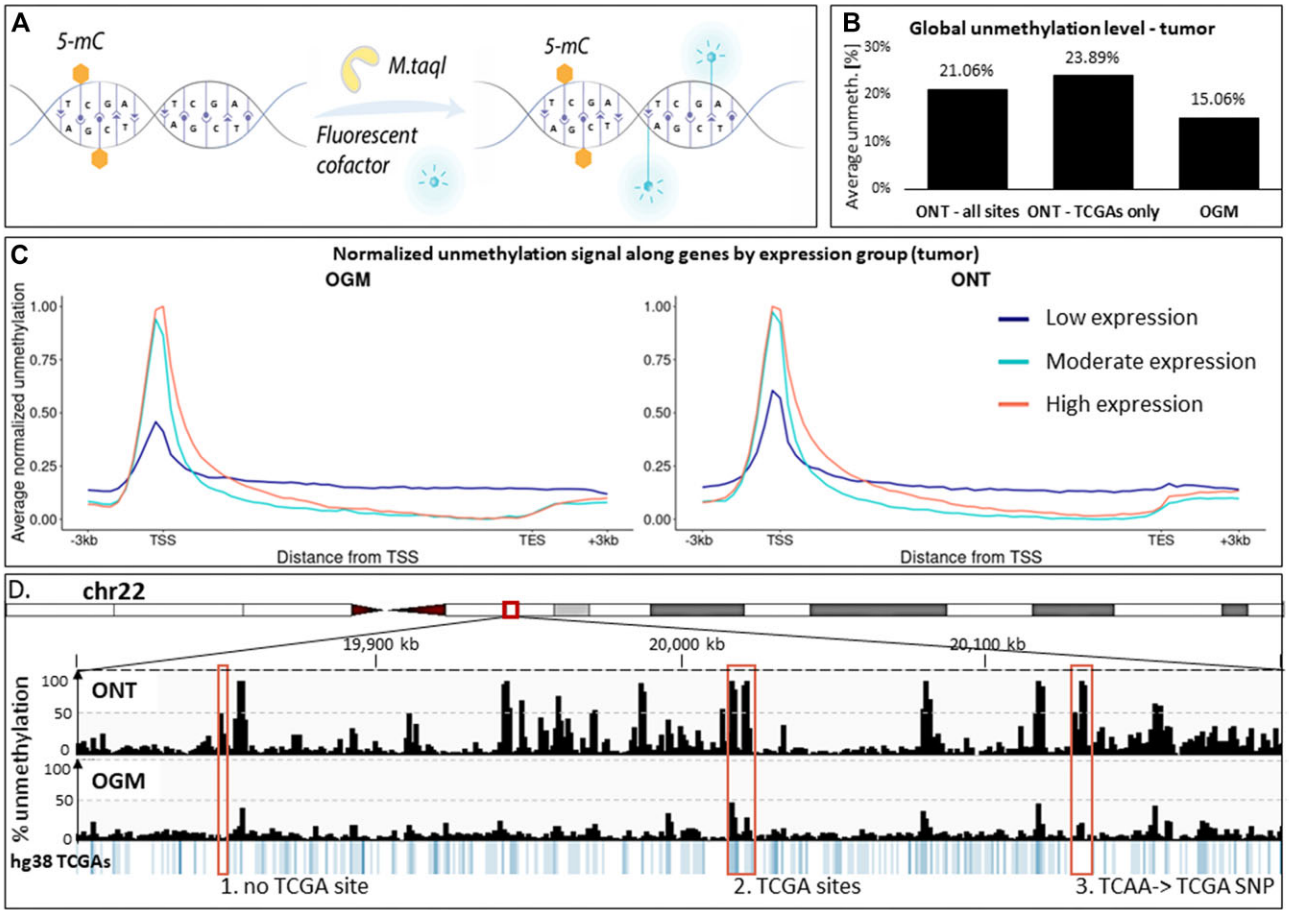

Figure 5. Unmethylation analysis. (**A**) Fluorescent labeling scheme for unmodified CpGs embedded in TCGA motifs for ROM. (**B**) Average global unmethylation levels of a ccRCC tumor, as detected by ONT in all CpG sites, by ONT when restricted to TCGA-embedded CpG sites only, and by OGM (inherently marking only TCGA sites). (**C**) ONT and OGM unmethylation signal, each normalized between 0 and 1, of the ccRCC tumor along aggregated genes. Genes were grouped based on their expression in ccRCC tumors. (**D**) Unmethylation profiles of the ccRCC tumor by ONT and OGM along a region on chromosome 22, and the corresponding density of TCGA motifs in the hg38 reference. Three boxes mark three regions that differ in TCGA content (all regions contain CpG sites): 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 single nucleotide polymorphism (SNP).

report joins earlier research comparing SV identification by OGM to ONT (64,65) and serves as one of the first benchmarks for real data comparisons between these methods. We expand the scope by considering different size ranges, testing multiple SV callers for ONT, and examining two key epigenetic marks, 5mC and 5hmC, for the first time. This thorough evaluation of the two methods offers an objective comparison, delving into the data types accessible with each technology, and the capabilities of their respective analytical tools. We recognize 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. BNG additionally compiled a substantial reference database of healthy controls. This enables the filtering of non-pathogenic findings.

At the karyotype level, the methods conform, and both are capable of providing reliable copy number evaluations. Nevertheless, slight differences in copy number can be observed and are attributed to the higher resolution of ONT.

A trade-off between resolution, read-length and coverage was observed in detecting structural variants of varying sizes. ONT's single bp resolution gives access to short insertions and deletions that OGM is blind to. Additionally, the high resolution also enables more accurate reporting of the SVs' breakpoints. On the other hand, large or complex SVs are challenging for ONT under the experimental N50 and coverage. ONT excelled in detecting small SVs, uniquely identifying ∼89% of SVs in the 50–500 bp range. This percentage gradually declines with increasing SV size, down to only 3% for SVs above 10 kb. Conversely, OGM demonstrated a clear advantage in detecting large SVs, uniquely identifying 80% of SVs above 10 kb, with this percentage decreasing for smaller SV sizes. Both methods detected high percentages (60–74%) of common SVs in the 500 bp–10 kb range.

Regarding SV types, both insertions and deletions were detected by OGM and all four evaluated ONT SV callers: Sniffles2, SVIM, CuteSV, and NanoVar. However, translocations, inversions and duplications were identified only by NanoVar and OGM, with minimal overlap between their findings. Notably, an inter-chromosomal translocation detected by OGM was manually identified in ONT reads but was missed by all SV callers. This highlights the potency of the technology and suggests that analysis pipelines can be further optimized to fully leverage its capabilities.

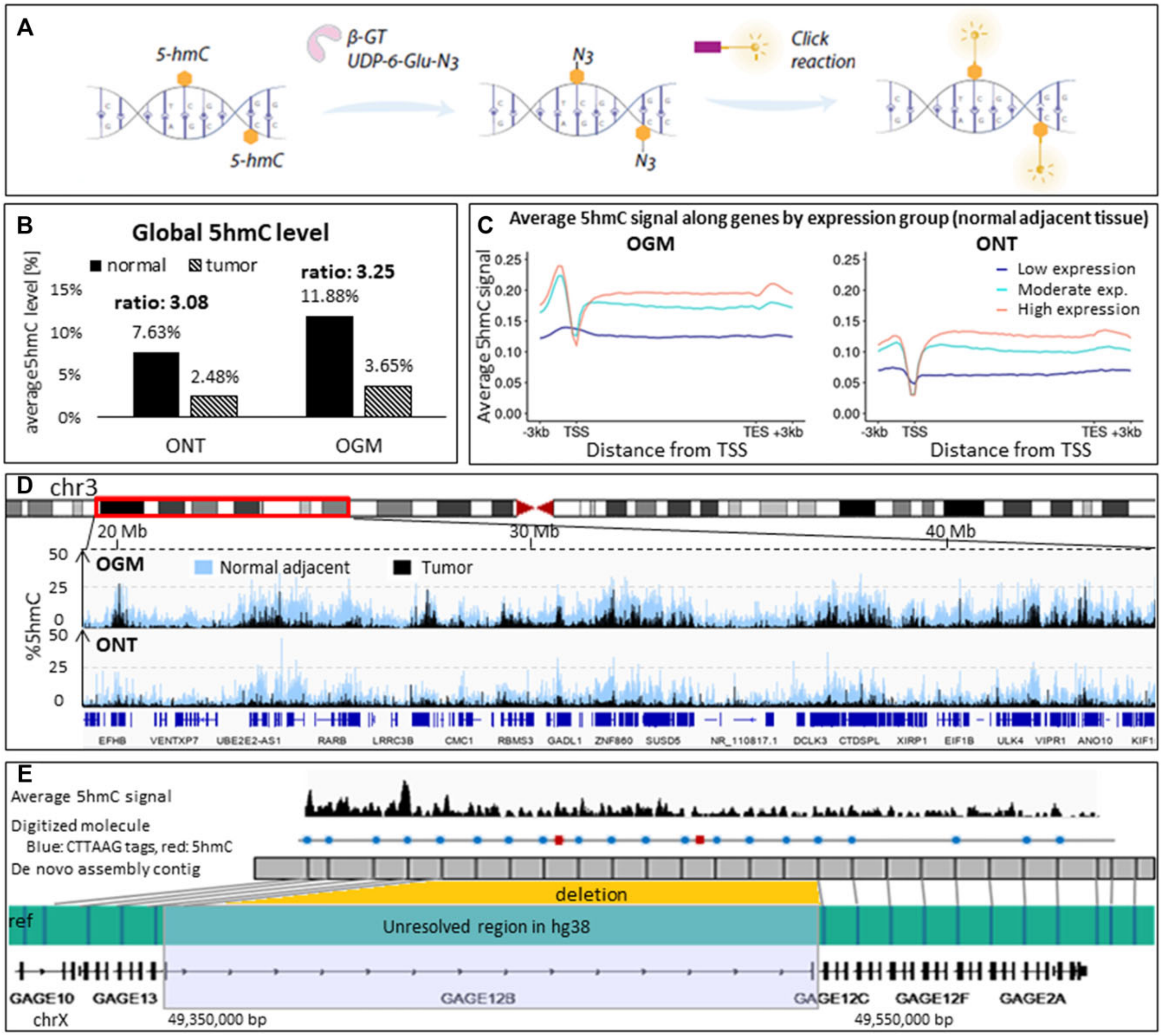

Figure 6. 5hmC analysis. (**A**) Direct fluorescent labeling of 5hmC for OGM. (**B**) Global 5hmC levels of a ccRCC tumor and a normal adjacent tissue, as detected in OGM and ONT. (**C**) OGM and ONT 5hmC signal of a normal kidney tissue adjacent to a ccRCC tumor along aggregated genes. Genes are grouped based on their expression in kidney tissues adjacent to ccRCC tumors. (**D**) 5hmC profiles of a ccRCC tumor and a normal adjacent tissue generated by OGM and ONT along a ∼25 Mb region on chromosome 3. (**E**) A repetitive sequence element on 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 (blue circles) and 5hmC labels (red squares). Above it, is the average 5hmC signal along the contig.

Per dollar, the genomic coverage generated by OGM is higher than that of ONT, opening a window to detect low-frequency variants and more resilience to sample heterogeneity. However, in this experiment, we did not meet the recommended coverage for running Bionano Genomics's 'rare variant pipeline' [300× is recommended for high sensitivity to low frequency variants (66)], 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 (52,53,64).

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 and the density of its recognition sites. The enzyme M.TaqI, described here, efficiently labels CpG sites nested within the TCGA motif (15). 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 (15), but inherent reduced representation limitations apply, in addition to constrains added by the difficulty to resolve adjacent labels due to optical resolution (diffraction limit). Additionally, the indirect labeling done by methyltransferase enzymes, pointing unmodified sites, can not 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, such as correlation of signal with gene expression group, persisted, while locus specific signals depend on TCGA representation. This comparison highlights OGM's limitations in methylation calling compared to ONT. However, to date, the picture for 5hmC presents a different scenario. In this case, the fluorescent labeling added to OGM, while also external and not supported by BNG, directly labels 5hmC residues and not complementary sites (16). Similarly to methylation calling, ONT enables 5hmC identification together with canonical basecalling without additional experimental steps, but some differences have to be considered. As the process of modification calling relies on machine learning, model training is a crucial step for accurate identification of 5hmC. This step requires comprehensive reference data covering the modification in all possible sequence contexts and distinguishing it from other cytosine modifications to assure accurate calls. Unfortunately, unlike for methylation, obtaining high-quality genome-wide reference data for 5hmC is still challenging and expensive, and might limit the comprehensiveness of the training data, thus affecting the performance of 5hmC calling models. This may explain the lower 5hmC levels called by ONT compared to OGM, seen in our comparison, and suggest that the ONT model currently underestimates the density of 5hmC and misses many of the modified bases.

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, and for optimal utilization of resources, a thorough understanding of the data generated by each platform, alongside the strengths and limitations of their respective analytical toolkits is needed.

Data availability

The ONT data generated in this study have been submitted to the NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra under accession number PRJNA1196660. The OGM data generated in this study, as well as processed ONT files, have been submitted to Zenodo with digital object identifier 10.5281/zenodo.14266624.

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

Author contribution: S.M.: Methodology, software, formal analysis, investigation and writing (original draft); Z.T.: Investigation and formal analysis; T.D.Z., Y.M., J.D. and G.N.: Investigation; R.H.: Writing (review and editing); Y.G.: Resources; D.O.: Resources; B.D.: Resources; H.B.F.: Writing (review and editing); A.G.: Conceptualization, methodology, software, formal analysis, investigation and writing (original draft); YE: Conceptualization and writing (original draft).

Funding

European Research Council [817 811]; Israel Science Foundation [771/21]; National Institute of Health/National Human Genome Research Institute (NIH/NHGRI) [R01HG009190]. Funding for open access charge: European Research Council.

Conflict of interest statement

None declared.

References

1. Cosenza,M.R., Rodriguez-Martin,B. and Korbel,J.O. (2022) Structural variation in cancer: role, prevalence, and mechanisms. *Annu. Rev. Genomics Hum. Genet.*, **23**, 123–152.
2. Ozkan,E. and Lacerda,M. (2023) Genetics, cytogenetic testing and conventional karyotype. In: *StatPearls [Internet]*. StatPearls Publishing, Treasure Island (FL).
3. Yang,L. (2020) A practical guide for structural variation detection in human genome. *Curr. Protoc. Hum. Genet.*, **107**, e103.
4. Cui,C., Shu,W. and Li,P. (2016) Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.*, **4**, 89.
5. Amarasinghe,S.L., Su,S., Dong,X., Zappia,L., Ritchie,M.E. and Gouil,Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.*, **21**, 30.
6. Marks,P., Garcia,S., Barrio,A.M., 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.
7. Chen,Z., Pham,L., Wu,T.C., Mo,G., Xia,Y., Chan,P.L., 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.
8. Mathew,M.T., Babcock,M., Hou,Y.C.C., Hunter,J.M., Leung,M.L., Mei,H., Schieffer,K. and Akkari,Y. (2024) Clinical cytogenetics: current practices and beyond. *J. Appl. Lab. Med.*, **9**, 61–75.
9. Pang,A.W.C., Kosco,K., Sahajpal,N.S., Sridhar,A., Hauenstein,J., Clifford,B., Estabrook,J., Chitsazan,A.D., Sahoo,T., Iqbal,A., *et al.* (2023) Analytic validation of optical genome mapping in hematological malignancies. *Biomedicines*, **11**, 3263.
10. Jeffet,J., Margalit,S., Michaeli,Y. and Ebenstein,Y. (2021) Single-molecule optical genome mapping in nanochannels: multidisciplinarity at the nanoscale Jonathan. *Essays Biochem.*, **65**, 51–66.
11. Wang,Y., Zhao,Y., Bollas,A., Wang,Y. and Au,K.F. (2021) Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.*, **39**, 1348–1365.
12. Mantere,T., Kersten,S. and Hoischen,A. (2019) Long-read sequencing emerging in medical genetics. *Front. Genet.*, **10**, 426.
13. Ni,Y., Liu,X., Simeneh,Z.M., Yang,M. and Li,R. (2023) Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput. Struct. Biotechnol. J.*, **21**, 2352–2364.
14. Heck,C., Michaeli,Y., Bald,I. and Ebenstein,Y. (2019) Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. *Curr. Opin. Biotechnol.*, **55**, 151–158.
15. 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.
16. 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.

17. 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.
18. White,L.K. and Hesselberth,J.R. (2022) Modification mapping by nanopore sequencing. *Front. Genet.*, **13**, 1037134.
19. Gabrieli,T., Michaeli,Y., Avraham,S., Torchinsky,D., Margalit,S., Sch¨utz,L., Juhasz,M., Coruh,C., Arbib,N., Zhou,Z.S., *et al.* (2022) Chemoenzymatic labeling of DNA methylation patterns for single-molecule epigenetic mapping. *Nucleic. Acids. Res.*, **50**, e92.
20. 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**, e202300400.
21. Margalit,S., Tulpová,Z., Michaeli,Y., Zur,T.D., 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 doi: https://doi.org/10.1101/2022.10.11.511152, 12 October 2022, preprint: not peer reviewed.
22. Shi,D.Q., Ali,I., Tang,J. and Yang,W.-C. (2017) New insights into 5hmC DNA modification: generation, distribution and function. *Front. Genet.*, **8**, 100.
23. Jin,S.G., Jiang,Y., Qiu,R., Rauch,T.A., Wang,Y., Schackert,G., Krex,D., Lu,Q. and Pfeifer,G.P. (2011) 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.*, **71**, 7360–7365.
24. Moore,L.E., Jaeger,E., Nickerson,M.L., 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**, e14.
25. Jonasch,E., Walker,C.L. and Rathmell,W.K. (2021) Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. *Nat. Rev. Nephrol.*, **17**, 245–261.
26. Le,V.H. and Hsieh,J.J. (2018) Genomics and genetics of clear cell renal cell carcinoma: a mini-review. *J. Transl. Genet. Genomics*, **2**, 17.
27. Quddus,M., Pratt,N. and Nabi,G. (2019) Chromosomal aberrations in renal cell carcinoma: an overview with implications for clinical practice. *Urol. Ann.*, **11**, 6–14.
28. Klatte,T., Rao,P.N., De Martino,M., Larochelle,J., Shuch,B., Zomorodian,N., Said,J., Kabbinavar,F.F., Belldegrun,A.S. and Pantuck,A.J. (2009) Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J. Clin. Oncol.*, **27**, 746–753.
29. Shenoy,N., Vallumsetla,N., Zou,Y., Galeas,J.N., Shrivastava,M., Hu,C., Susztak,K. and Verma,A. (2015) Role of DNA methylation in renal cell carcinoma. *J. Hematol. Oncol.*, **8**, 88.
30. Hu,C.Y., Mohtat,D., Yu,Y., Ko,Y.A., Shenoy,N., Bhattacharya,S., Izquierdo,M.C., Park,A.S.D., Giricz,O., Vallumsetla,N., *et al.* (2014) Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. *Clin. Cancer Res.*, **20**, 4349–4360.
31. Chen,K., Zhang,J., Guo,Z., Ma,Q., Xu,Z., Zhou,Y., Xu,Z., Li,Z., Liu,Y., Ye,X., *et al.* (2016) Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res.*, **26**, 103–118.
32. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
33. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
34. Ewels,P.A., Peltzer,A., Fillinger,S., Patel,H., Alneberg,J., Wilm,A., Garcia,M.U., Di Tommaso,P. and Nahnsen,S. (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.*, **38**, 276–278.
35. Grunwald,A., Dahan,M., Giesbertz,A., Nilsson,A., Nyberg,L.K., Weinhold,E., Ambjörnsson,T., Westerlund,F. and Ebenstein,Y. (2015) Bacteriophage strain typing by rapid single molecule analysis. *Nucleic. Acids. Res.*, **43**, e117.
36. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
37. Geoffroy,V., Herenger,Y., Kress,A., Stoetzel,C., Piton,A., Dollfus,H. and Muller,J. (2018) AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics*, **34**, 3572–3574.
38. Geoffroy,V., Guignard,T., Kress,A., Gaillard,J.B., Solli-Nowlan,T., Schalk,A., Gatinois,V., Dollfus,H., Scheidecker,S. and Muller,J. (2021) AnnotSV and knotAnnotSV: a web server for human structural variations annotations, ranking and analysis. *Nucleic. Acids. Res.*, **49**, W21–W28.
39. Heller,D. and Vingron,M. (2019) SVIM: structural variant identification using mapped long reads. *Bioinformatics*, **35**, 2907–2915.
40. Jiang,T., Liu,Y., Jiang,Y., Li,J., Gao,Y., Cui,Z., Liu,Y., Liu,B. and Wang,Y. (2020) Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol.*, **21**, 189.
41. Tham,C.Y., Tirado-Magallanes,R., Goh,Y., Fullwood,M.J., Koh,B.T.H., Wang,W., Ng,C.H., Chng,W.J., Thiery,A., Tenen,D.G., *et al.* (2020) NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol.*, **21**, 56.
42. Neph,S., Kuehn,M.S., Reynolds,A.P., Haugen,E., Thurman,R.E., Johnson,A.K., Rynes,E., Maurano,M.T., Vierstra,J., Thomas,S., *et al.* (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics*, **28**, 1919–1920.
43. Creighton,C.J., Morgan,M., Gunaratne,P.H., Wheeler,D.A., Gibbs,R.A., Robertson,G., Chu,A., Beroukhim,R., Cibulskis,K., Signoretti,S., *et al.* (2013) Comprehensivemolecular characterization of clear cell renal cell carcinoma. *Nature*, **499**, 43–49.
44. Durinck,S., Spellman,P.T., Birney,E. and Huber,W. (2009) Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nat. Protoc.*, **4**, 1184–1191.
45. Ramírez,F., Dündar,F., Diehl,S., Grüning,B.A. and Manke,T. (2014) DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic. Acids. Res.*, **42**, 187–191.
46. Bionano Genomics (2023) Bionano Genomics website. *Reveal More Genomic Var. That Matters With Opt. Genome Mapp.*
47. Bionano Genomics (2023) Bionano Genomics website. *Prod. Sheet - Optim. Sample Prep. Opt. Genome Mapp. Simpl. Work. Opt. Genome Mapp.*
48. Oxford Nanopore Technologies (2024) PromethION. Website, Oxford Nanopore Technol.
49. Cahyani,I., Tyson,J., Holmes,N., Quick,J., Moore,C., Loman,N.J. and Loose,M.W. (2024) FindingNemo: a toolkit for DNA extraction, library preparation and purification for ultra long nanopore sequencing. bioRxiv doi: https://doi.org/10.1101/2024.08.16.608306, 18 August 2024, preprint: not peer reviewed.
50. Bionano Genomics (2023) Bionano genomics ordering Guide 2023. *Bionano Genomics Website.*
51. Smolka,M., Paulin,L.F., Grochowski,C.M., Horner,D.W., 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.*, **42**, 1571–1580.
52. Bolognini,D. and Magi,A. (2021) Evaluation of germline structural variant calling methods for nanopore sequencing data. *Front. Genet.*, **12**, 761791.

53. Liu,Y.H., Luo,C., Golding,S.G., Ioffe,J.B. and Zhou,X.M. (2024) Tradeoffs in alignment and assembly-based methods for structural variant detection with long-read sequencing data. *Nat. Commun.*, **15**, 2447.
54. Diesh,C., Stevens,G.J., Xie,P., De Jesus Martinez,T., Hershberg,E.A., Leung,A., Guo,E., Dider,S., Zhang,J., Bridge,C., *et al.* (2023) JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.*, **24**, 74.
55. Yao,X., Tan,J., Lim,K.J., Koh,J., Ooi,W.F., Li,Z., Huang,D., Xing,M., Chan,Y.S., Qu,J.Z., *et al.* (2017) VHL deficiency drives enhancer activation of oncogenes in clear cell renal cell carcinoma. *Cancer Discov.*, **7**, 1284–1305.
56. Nifker,G., Levy-Sakin,M., Berkov-Zrihen,Y., Shahal,T., Gabrieli,T., Fridman,M. and Ebenstein,Y. (2015) One-pot chemoenzymatic cascade for labeling of the epigenetic marker 5-hydroxymethylcytosine. *ChemBioChem*, **16**, 1857–1860.
57. Song,C.-X., Szulwach,K.E., 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. *Nat. Biotechnol.*, **29**, 68–72.
58. Michaeli,Y., Shahal,T., Torchinsky,D., Grunwald,A., Hoch,R. and Ebenstein,Y. (2013) Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem. Commun. (Camb).*, **49**, 8599–8601.
59. Shahal,T., Gilat,N., Michaeli,Y., Redy-keisar,O., Shabat,D. and Ebenstein,Y. (2014) Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA. *Anal. Chem.*, **86**, 8231–8237.
60. 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.
61. Chen,S., Dou,Y., Zhao,Z., Li,F., Su,J., Fan,C. and Song,S. (2016) High-sensitivity and High-efficiency detection of DNA hydroxymethylation in genomic DNA by multiplexing electrochemical biosensing. *Anal. Chem.*, **88**, 3476–3480.
62. 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.
63. Amemiya,H.M., Kundaje,A. and Boyle,A.P. (2019) The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.*, **9**, 9354.
64. Savara,J., Novosád,T., Gajdoš,P. and Kriegová,E. (2021) Comparison of structural variants detected by optical mapping with long-read next-generation sequencing. *Bioinformatics*, **37**, 3398–3404.
65. Pei,Y., Tanguy,M., Giess,A., Dixit,A., Wilson,L.C., Gibbons,R.J., Twigg,S.R.F., Elgar,G. and Wilkie,A.O.M. (2024) A comparison of structural variant calling from short-read and nanopore-based whole-genome sequencing using optical genome mapping as a benchmark. *Genes (Basel).*, **15**, 925.
66. Bionano Genomics (2021) Bionano Solve theory of operation: structural variant calling. *Bionano Genomics Website*.

Received: July 28, 2024. **Revised:** December 12, 2024. **Editorial Decision:** December 23, 2024. **Accepted:** December 31, 2024