NAR Cancer, 2025, **7**, zcaf008
https://doi.org/10.1093/narcan/zcaf008
Cancer Gene Regulation, Chromatin, and Epigenetics

Optical genome and epigenome mapping of clear cell renal cell carcinoma

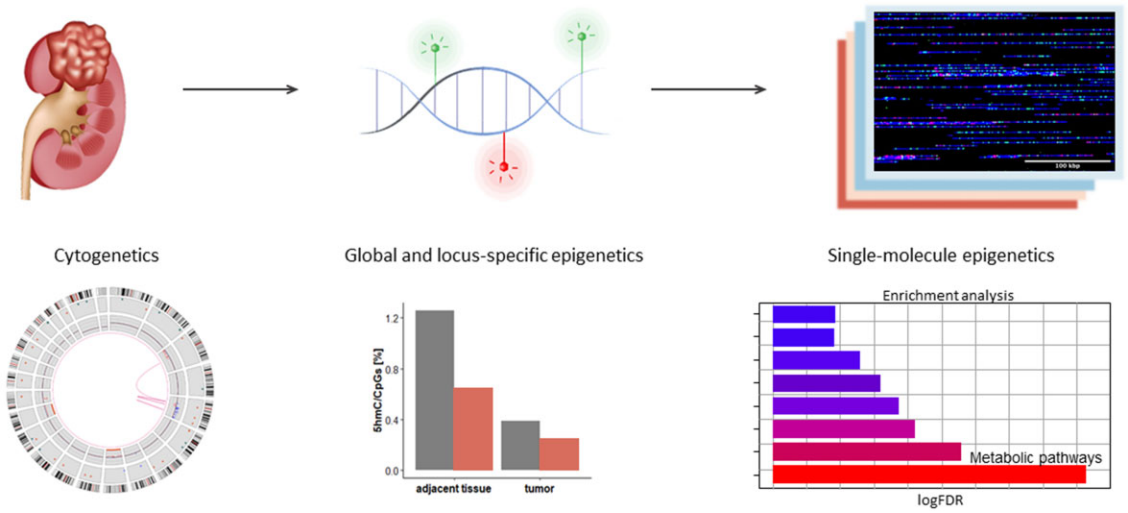
Sapir Margalit[1], Zuzana Tulpová[1,2], Yael Michaeli[1], Tahir Detinis Zur[1], Jasline Deek[1], Sivan Louzoun-Zada[1], Gil Nifker[1], Assaf Grunwald[1], Yuval Scher[1], Leonie Schütz[3], Elmar Weinhold[3], Yehudit Gnatek[4], Dorit Omer[4], Benjamin Dekel[4,5,6], Eitan Friedman[6,7], Yuval Ebenstein[1,8,*]

[1]School 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, 77900, Olomouc, Czech Republic
[3]Institute of Organic Chemistry, RWTH Aachen University, D-52056 Aachen, Germany
[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 Susanne Levy Gertner Oncogenetics Unit, The Danek Gertner Institute of Human Genetics, Sheba Medical Center, 52621 Ramat Gan, Israel
[8]Department of Biomedical Engineering, Tel Aviv University, 6997801 Tel Aviv, Israel
[*]To whom correspondence should be addressed. Email: uv@tauex.tau.ac.il

Abstract

Cancer cells display complex genomic aberrations that include large-scale genetic rearrangements and epigenetic modulation that are not easily captured by short-read sequencing. This study presents a novel approach for simultaneous profiling of long-range genetic and epigenetic changes in matched cancer samples, focusing on clear cell renal cell carcinoma (ccRCC). ccRCC is a common kidney cancer subtype frequently characterized by a 3p deletion and the inactivation of the von Hippel–Lindau (*VHL*) gene. We performed integrated genetic, cytogenetic, and epigenetic analyses on paired tumor and adjacent nontumorous tissue samples. Optical genome mapping identified genomic aberrations as structural and copy number variations, complementing exome-sequencing findings. Single-molecule methylome and hydroxymethylome mapping revealed a significant global reduction in 5hmC level in both sample pairs, and a correlation between both epigenetic signals and gene expression was observed. The single-molecule epigenetic analysis identified numerous differentially modified regions, some implicated in ccRCC pathogenesis, including the genes *VHL*, *PRCC*, and *PBRM1*. Notably, pathways related to metabolism and cancer development were significantly enriched among these differential regions. This study demonstrates the feasibility of integrating optical genome and epigenome mapping for comprehensive characterization of matched tumor and adjacent tissue, uncovering both established and novel somatic aberrations.

Graphical abstract



Received: September 12, 2024. **Revised:** February 18, 2025. **Editorial Decision:** February 20, 2025. **Accepted:** March 5, 2025

Introduction

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 exhibit alterations affecting the short arm of chromosome 3 (3p), ranging from translocations and deletions to the loss of the entire chromosomal arm. The von Hippel–Lindau (*VHL*) gene, located on this arm, is mutated in 30%–56% of sporadic clear cell carcinomas, and is silenced by promoter hypermethylation in up to 19% of cases [1]. In these cases, the inactivation of *VHL* has been identified as the earliest event driving the disease. *VHL* loss in ccRCC affects multiple cellular processes as angiogenesis, cell cycle, cell growth, and metabolism [1, 2]. However, biallelic *VHL* inactivation alone is not sufficient to induce ccRCC [3, 4]. Pathogenic sequence variants affecting other 3p-residing tumor suppressor genes are also frequently observed in ccRCC. Genes such as *PBRM1*, *SETD2*, and *BAP1*, encode for chromatin and histone modifiers and are often mutated in ccRCC, suggesting a possible role of epigenetic dysregulation in ccRCC tumorigenesis [5]. Additionally, DNA copy number variations (CNVs) affecting other chromosomes (e.g., 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), are very common in ccRCC, and some were correlated with prognosis [6–8].

Epigenetics in ccRCC

One of the fundamental epigenetic mechanisms directly affecting gene expression is DNA methylation of cytosine (5mC) in the dinucleotide sequence CpG. Aberrant methylation is common in ccRCC. Hypermethylation at promoter regions often results in the silencing or inactivation of tumor suppressor genes [9, 10]. RCC tissues often exhibit widespread DNA hypermethylation in gene bodies and kidney-specific enhancer regions. This aberrant hypermethylation has been reported to correlate with stage, grade and aggressiveness of RCC, with enhancer hypermethylation being particularly predictive of adverse prognosis [9,10–12]. These and other studies prompted discovery and application of specific prognostic methylation markers in ccRCC [13, 14]. In addition to cytosine methylation, 5-hydroxymethylcytosine (5hmC), the oxidation product of 5mC, has gained attention as a modifier of gene regulation, development, and disease. Some suggested mechanisms for the regulatory action of 5hmC include binding to transcription factors, altering chromatin structure through association with histone modifications, modulating alternative splicing via binding to related proteins, and involvement in miRNA pathways [15]. 5hmC is globally reduced in multiple human cancers [16, 17], including ccRCC [18], and lower 5hmC levels in ccRCC are reportedly associated with poorer prognosis [18, 19]. 5hmC data have only recently become available due to the fact that popular methods, such as bisulfite sequencing or methylation arrays, do not distinguish between DNA methylation and hydroxymethylation, and report on their cumulative presence. In order to differentiate these two marks, Tet-assisted bisulfite sequencing (TAB-seq) [20], oxidative bisulfite sequencing (OxBS-seq) [21], or specific enzymatic labeling such as presented here [17, 22–25] have to be employed.

Optical genome mapping— the full picture

Optical genome mapping (OGM) in nanochannels is a high-throughput, single-molecule technique that captures ultra-long genomic fragments and may uncover genomic information that is mostly inaccessible by sequencing [26]. The method is based on sequence-specific fluorescent labeling of up to mega-base pairs (Mbp) long chromosomal DNA molecules. The labeled DNA molecules are linearized and stretched in nanochannels, allowing for imaging via fluorescence microscopy. The fluorescent marker patterns on each DNA molecule are used to create unique barcodes that identify the genomic origin of the molecules, either by alignment to a reference map or by their assembly to create long consensus contiguous maps *denovo* [26]. This method enables the construction of complex genomes such as those common in many cancers [26–29]. It is used clinically as a modern replacement for traditional cytogenetic testing, and a complementing method to next-generation sequencing (NGS). Additionally, unlike NGS methods, where ensemble averages mask cellular variability, OGM provides information at the single cell level, as each mapped DNA molecule originates from a different cell, allowing a high-throughput characterization of cellular heterogeneity [26, 30]. Using fluorescence microscopy and designated chemistries, OGM can provide multilayered information from individual DNA molecules [22, 26, 27, 31, 32]. Fluorescent labeling of different genomic features with different colors allows studying multiple epigenetic marks on the single-molecule level, creating a hybrid genetic/epigenetic map for every DNA molecule.

Here, we utilize a novel approach to complement single-base resolution exome-sequencing with single-molecule optical genome/epigenome mapping. We comprehensively analyze somatic alterations in two matched ccRCC samples (early and advanced stage) as a demonstration of the ability to apply genome/epigenome mapping to comparative genomics and epigenomics studies.

Materials and methods

Patient clinically relevant information

Tumor and normal adjacent tissue were obtained in the course of partial nephrectomy performed in a 66-year-old female. Tumor was diagnosed histologically as ccRCC at pT1a stage (<4 cm in the greatest dimension) with cystic degenerative changes ("early stage ccRCC"). Tumor and normal adjacent tissue were also 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 ("advanced stage ccRCC"). Tissues were stored from the time of surgery to analysis at −80°C.

Samples collection and handling were approved by the institutional review board of Sheba Medical Center, Tel Hashomer, Israel, in accordance with the declaration of Helsinki.

Extraction of high-molecular weight DNA

Ultra-high molecular weight (UHMW) DNA was extracted using *SP Tissue and Tumor DNA Isolation Kit* (Bionano Genomics), according to the manufacturer's protocol. High molecular weight (HMW) DNA was extracted using Bionano Prep *Animal Tissue DNA Isolation Kit* (Bionano Genomics),

according to *Fibrous Tissue Protocol* for the tumor and *Soft Tissue* protocol for the normal adjacent tissue.

Whole-exome sequencing and analysis

Exome sequencing was provided as a service (CD Genomics). 500 ng DNA were used for library construction. Sequencing libraries were generated using Agilent SureSelect Human All Exon kit (Agilent Technologies) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample (experimental details provided in the Supplementary Data).

Raw sequencing reads were filtered by Trim Galore software (v0.6.7, 10.5281/zenodo.5127898) to remove reads containing adapters or reads of low quality, so that downstream analyses are based on clean reads. Mapping of paired-end clean reads to the human reference genome (hg38) was performed with Bowtie2 software (v2.2.5, [33]). Following alignment, a pipeline by Genome Analysis Toolkit (GATK, v4.2.2.0, [34]) was followed, including using Samtools (v0.1.19, [35]) for sorting and Picard (https://github.com/broadinstitute/picard/) for marking duplicated reads. Base Quality Recalibration process was applied using standard hg38 reference variants. For variant calling, the alignment files of both samples were first merged for efficient simultaneous variant calling. Then, small variants, including single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) located in exon regions, were called by GATK standard genotype pipeline. The called variants of the different samples were then separated. SnpSift program (v4.3t, [36]) was used to add NCBI dbSNP information (v146, [37]), and SnpEff program (v4.3t, [36]) was used to annotate the variants and determine the effect of each variant. Full pipeline and parameters used can be found in the Supplementary Data.

DNA barcoding and staining for optical genome mapping

All samples were labeled by Direct Label and Stain (DLS) chemistry (DLE-1 enzyme, Bionano Genomics, kit part number: 80005), creating a genetic barcode (CTTAAG motif). Single color labeling was created according to a protocol by Bionano Genomics (https://bionanogenomics.com/wp-content/uploads/2018/04/30206-Bionano-Prep-Direct-Label-and-Stain-DLS-Protocol.pdf).

Dual color labeling for optical epigenome mapping

Samples were subjected to two types of epigenetic labeling procedures to generate a comprehensive optical epigenome map. In order to distinguish the epigenetic marks from the green fluorescent DLE-1 marks, we used the red fluorophore ATTO643 (ATTO-Tech), which was found to perform well under our experimental conditions. Synthetic protocols for the ATTO643 labeling reagents prepared for this study are presented in the Supplementary Data.

Labeling reduced representation of unmodified cytosines in CpG context

To create the genetic barcode, 1 μg of U/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) [27, 38]. 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 (see synthesis and mass spectrum in the Supplementary Data, Supplementary Figs S1 and S2), 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) was 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 was stained to visualize DNA backbone by mixing it with 15 μl of 4× Flow Buffer (Bionano Genomics), 6 μl of 1 M DTT (DL-Dithiothreitol; 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. The orthogonality of the two consecutive reactions was confirmed by no observed increase in false DLE-1 labels.

Labeling 5hmC sites

To create the genetic barcode, 580–750 ng of U/HMW 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 30–35 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) [22]. 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) and uridine diphosphate-6-azideglucose (UDP-6-N3-Glu, lab-made) [24] 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 (see synthesis in the Supplementary Data) 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 (pH 8) with 0.1 μm dialysis membrane for a total of 2–2.5 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 (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 ratio of 8% (v/v), and ultrapure water. The reaction mixture was shaken horizontally on a HulaMixer for an hour and then incubated overnight at 4°C.

Optical mapping

Labeled samples were loaded on Saphyr chips (G1.2) and run on a Saphyr instrument (Bionano Genomics) to generate single molecule maps.

Structural variant and copy number variation calling

De novo assemblies of single-color data were generated by Bionano Access (v1.6.1) with Bionano Solve (v3.6.1). The set parameters were "haplotype with extend and split" and "cut CMPR." The *in silico* digested human genome GRCh38.p13 (*hg38_DLE1_0kb_0labels.cmap*) was used as the reference.

Structural variants (SVs) were called using variant annotation pipeline (VAP), performed both as single sample analysis and as dual samples analysis (tumor versus matched normal tissue) in Bionano Access combined with Bionano Tools (v1.6.1) and Bionano Solve (v3.6.1), with default filters. Only SVs that are not present in the Bionano controls dataset were considered. CNV analysis was performed using the same tools. This analysis allows the detection of large, unbalanced aberrations based on normalized molecule coverage, and was performed with default parameters as a part of the *de novo* assembly.

Optical epigenome mapping analysis

Optical mapping data for each sample were merged to a single dataset using Bionano Access (v1.6.1) and Bionano Solve (v3.6.1). Genetic and epigenetic channels were swapped in these files with Bionano Solve before the alignment of the molecules files to the reference, as instructed by the company. Molecules spanning over 150 kb were then aligned to the *in silico* human genome reference GRCh38.p13, based on DLE-1 recognition sites (*hg38_DLE1_0kb_0labels.cmap*), with default parameters matching the following combination of arguments: haplotype, human, DLE-1, and Saphyr. Only molecules with an alignment confidence ≥ 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) and to single-molecule-level epigenetic maps, according to the pipeline described by Gabrieli *et al.* and Sharim *et al.* [22, 27]. For more information, see 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. The number of epigenetic labels per 100 kb in an experiment was calculated as the total number of labels in mapped and filtered reads divided by the total corrected length of the mapped and filtered reads. Average coverage of the hg38 genome was calculated for a genome size of 3.1 Gb. Positions of sequence motifs in the reference were obtained using the R package *BSgenome* (https://bioconductor.org/packages/release/bioc/html/BSgenome.html).

Definition of annotated genomic regions

Gene bodies were defined as spanning from the transcription start site (TSS) to the transcription end site (TES) annotated by GENCODE (v34, [39]). Promoters were defined as ranging from 1000 bp upstream to 500 bp downstream from the GENCODE TSS. General predicted enhancers were mapped to gene targets by JEME and adapted from Cao *et al.* [40]. Genomic coordinates of enhancers were converted from the human genome build hg19 to hg38 using UCSC liftOver [41]. Enhancers overlapping ambiguous genomic regions [42] were discarded, as well as pairs of enhancers and gene targets that are overlapping or in close proximity (up to 5 kb). ccRCC-related enhancers were adapted from Yao *et al.* [43] based on differential H3K27ac and H3K4me1 scores not overlapping with promoters in histone chromatin immunoprecipitation sequencing (ChIP-seq) of 10 primary tumor/normal pairs, 5 patient-matched tumor-derived cell lines, 2 commercially available ccRCC lines (786-O and A-498), and 2 normal kidney cell lines (HK2 and PCS-400). Some of the enhancers were assigned to target genes. We adapted assignments made by correlations between H3K27ac signals and expression of genes within the same topologically associating domain (TAD) and by a capture-C experiment in 786-O cells. Genomic coordinates of enhancers were converted from the human genome build hg19 to hg38 using UCSC liftOver. ccRCC-related "super-enhancers," regions comprising dense clusters of enhancers located near known regulators of cell identity and disease, were also adapted from Yao *et al.* and converted to hg38 coordinates. Nonoverlapping genomic windows of 1, 5, and 50 kb of hg38 were generated using Bedtools *makewindow* (v2.26.0, [44]).

Epigenetic ideograms

The weighted mean of epigenetic signals in 50 kb genomic windows was calculated using Bedops *bedmap* (v2.4.35, [45]). Ideograms displaying the density of epigenetic labels were created with the R package Rideogram [46] with a minor modification: the values were scaled between 1 and the maximal value in the dataset, times 10 000. The darkest color in a pair of whole-genome ideograms was determined according to the highest value in the adjacent and tumor samples.

Gene expression data

Publicly available RNA-seq data of three tumor-matched pairs of ccRCC (stage 3) patients (PRJNA396588, GEO accessions: pair 1: GSM2723919, GSM2723920; pair 2: GSM2723927, GSM2723928; pair 3: GSM2723929, GSM2723930; [43]) were aligned to the human genome (hg38) using TopHat (v2.1.0, [47]) with default parameters and library-type and fr-firststrand flags, after retrieving the raw files with NCBI SRA toolkit [48]. Only uniquely mapped reads were analyzed (minimal mapping quality of 30). Gene counts were obtained using HTSeq (htseq-count, v0.11.3, [49]) against the GENCODE v34 [39] reference gene models. Transcripts per million (TPM) scores were calculated.

Epigenetic signals along aggregated genes

Transcription start and end sites (TSS and TES) of protein-coding genes were defined according to GENCODE annotation (v34, [39]). Protein-coding genes were divided into four groups based on their average normalized TPM score in the RNA-seq of three tumor/matched ccRCC samples. Unexpressed genes were defined as genes with TPM value $\leq$ 0.01 ($\sim$3000 genes). The other expression groups are three equal quantiles of the expressed protein-coding genes ($\sim$6000 genes per group). Mean epigenetic signals along genes were calculated using DeepTools *computeMatrix* (v3.4.1, [50]) in scale-regions mode, where each gene was scaled to 15 kb and

divided into 300 bp bins. Compressed matrix output was summarized by DeepTools *plotProfile*.

Finding differentially modified regions

The number of epigenetic labels along annotated genomic regions (gene bodies, promoters, general predicted enhancers, ccRCC-related enhancers, and super-enhancers; see "Definition of annotated genomic regions" section) and along 1 kb genomic windows in individual molecules was counted. Only regions entirely covered by over 20 molecules were regarded. To identify differentially modified regions, Welch's *t*-test was applied to the populations sampled by independent molecules fully covering the regions from each sample. Regions with a zero in the *t*-test's denominator (when the estimated standard deviations in both samples were zero) where discarded from the analysis.

$$T_{\text{region}} = \frac{avg\,(\text{adjacent}) - avg\,(\text{tumor})}{\sqrt{\frac{S^2_{\text{adjacent}}}{N_{\text{adjacent}}} + \frac{S^2_{\text{tumor}}}{N_{\text{tumor}}}}};$$

$$d.f = \frac{\left(\frac{S^2_{\text{adjacent}}}{N_{\text{adjacent}}} + \frac{S^2_{\text{tumor}}}{N_{\text{tumor}}}\right)^2}{\frac{\left(\frac{S^2_{\text{adjacent}}}{N_{\text{adjacent}}}\right)^2}{(N_{\text{adjacent}} - 1)} + \frac{\left(\frac{S^2_{\text{tumor}}}{N_{\text{tumor}}}\right)^2}{(N_{\text{tumor}} - 1)}}$$

Equation (1): Welch's *T*-statistics and degrees of freedom (d.f). avg (adjacent) and avg (tumor) are the average number of labels in the region of molecules covering it, in each of the samples. S_{adjacent} and S_{tumor} are the estimates for the standard deviations in each sample. N_{adjacent} and N_{tumor} are the number of molecules covering the region in each of the samples.

A *P*-value was then calculated and a Benjamini–Hochberg false discovery rate (FDR) [51] correction was applied. Regions with *q-value* < 0.1 were considered differentially modified.

Calculating fold change of epigenetic signals between adjacent and tumor samples

A continuous genome-wide track of fold change (ratio between the epigenetic signal in the adjacent tissue and the tumor) was calculated as follows: bedgraph files containing epigenetic signals in both samples were combined for direct comparison using Bedtools *unionbedg* (v2.26.0, [44]). A pseudo signal of 0.01 was added to each position in either sample to avoid division by 0. Then, the signal in the adjacent tissue was divided by the signal in the tumor to generate the fold change track.

Fold change in discrete regions was calculated as follows: the average number of epigenetic labels in molecules covering the region in each sample was calculated (only regions covered by at least 20 molecules were considered). A pseudo signal of 0.005 was added to each region in either sample to avoid division by 0. Then, the signal in the adjacent tissue was divided by the signal in the tumor. Log_2 of this ratio was then calculated.

Enrichment analysis

The clusterProfiler R package [52] was used to find Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that are significantly enriched among unique genes associated with the differentially modified annotated genomic regions (gene bodies, promoters, general predicted enhancers, ccRCC-related enhancers, and super-enhancers; see "Definition of annotated genomic regions" section). Unique genes associated with all elements of the same type covered in the experiment served as background lists. Pathway enrichment was assessed using the Benjamini–Hochberg method for *p*-value adjustment to control the FDR (*q*-value and *p*-value < 0.05).

Results

We present a comprehensive analysis of a human tumor and a matched tissue by optical genome/epigenome mapping, revealing disease-relevant and differential SVs, CNVs, and epigenetic modifications. Pathology-classified ccRCC tumor and an adjacent normal kidney tissues (Fig. 1A) were sequenced to detect genetic disease signatures (Fig. 1B). OGM of the samples was then performed on the Bionano Genomics Saphyr instrument to provide next-generation cytogenetics (Fig. 1C). For this purpose, Mbp-long DNA molecules were extracted from each sample. Subsequently, a methyltransferase enzyme attached a fluorescent tag to a specific sequence motif (CTTAAG), generating unique genetic barcodes. These barcodes enabled the alignment of molecules to the reference genome and their assembly into consensus maps, revealing the unique cytogenetic landscape of each sample. To integrate epigenetic information, custom labeling chemistries for unmodified cytosines in TCGA motif or for 5hmC were employed on top of the genetic barcode labels (Fig. 1D). Labeled DNA molecules were then confined and stretched within nanochannels and imaged in three colors: genetic barcode, epigenetic marks, and molecule contour, enabling optical epigenome mapping (OEM, Fig. 1E). The resulting epigenetic maps were analyzed at the genome-wide, locus-specific, and single-molecule level (Fig. 1F).

Optical genome mapping detects SVs and CNVs of ccRCC tumor and adjacent tissue

Initial exome-sequencing revealed several genetic aberrations in genes associated with ccRCC [9, 14, 43], including variants with a well-accepted and proven clinical impact in both the early and the advanced stage ccRCC tumors. Noteworthy are variants in *VHL* and *PBRM1* that are known to be highly associated with ccRCC [9, 14, 43]. More details about SNPs and InDels discovered by this analysis can be found in Supplementary Fig. S3 and in Supplementary Table S1.

To further investigate the genetic structure of these ccRCC samples, next-generation cytogenetics was employed using OGM. Genetic single-molecule data of early and advanced stage ccRCC tumors and adjacent tissues were generated. An average of 744.3 Gb (±52.9 Gb) of size-filtered (>150 kb) single-molecule data was generated per sample, with an average molecule N50 of 270.8 kb (±24.5 kb). The single-molecule data served to construct an annotated and phased *de novo* assembly for each sample. The average N50 of contigs in all assemblies was 59.1 Mbp (±0.7 Mbp) (full details can be found in Supplementary Table S2).

Bionano Genomics' VAP for normal-tumor pairs revealed 5666 SVs in the early stage tumor sample and 5658 SVs in its adjacent normal tissue, 5842 SVs in the advanced stage tumor, and 5829 SVs in its adjacent normal tissue. These SVs include insertions, deletions, inversions, and duplications, and the dis-

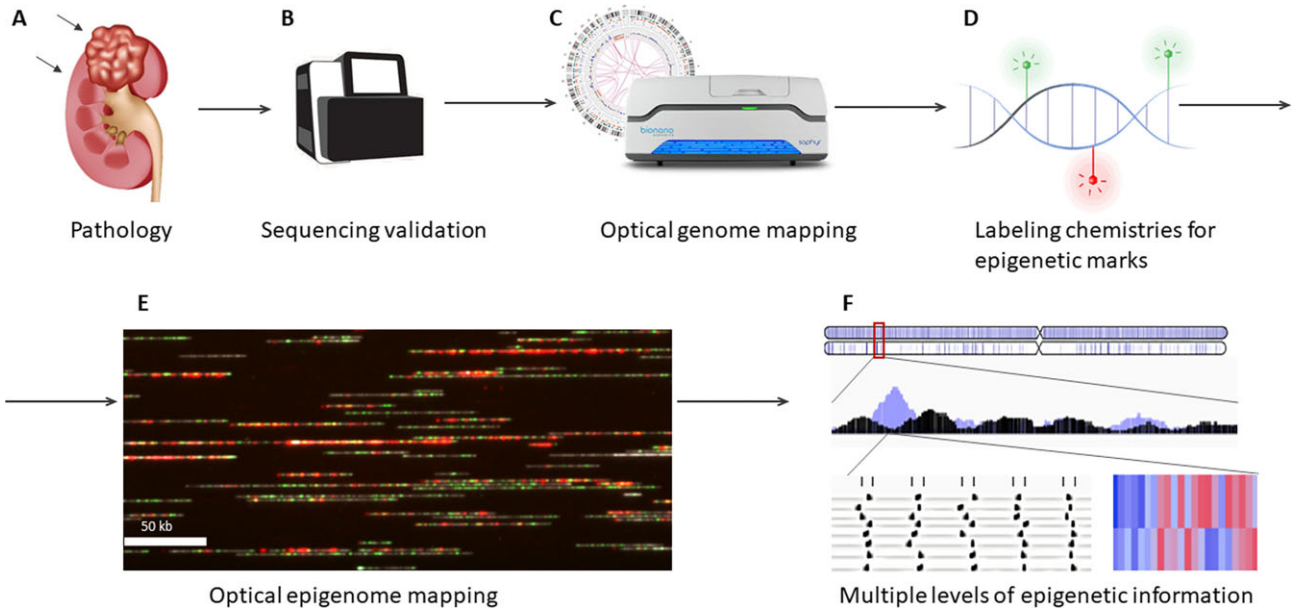


Figure 1. Assay workflow. (**A**) Two matched pairs of a ccRCC tumor and a normal adjacent tissue were pathologically classified and isolated. (**B**) Whole-exome of the tissues was sequenced. (**C**) OGM was applied to detect SVs and CNVs. (**D**) Additional labeling schemes for epigenetic features, developed in our lab, were applied to the long DNA molecules. (**E**) Labeled DNA molecules were imaged in three colors (genetic barcode, epigenetic marks, and molecule contour) for OEM. (**F**) Multiple levels of epigenetic information were extracted— genome view, chromosome level, locus-specific average, and single-molecule level, in coding and noncoding genomic regions.

tribution between the different types in all samples is similar. The exception is translocation breakpoints, that were detected only in the advanced stage ccRCC tumor. Full details are given in Supplementary Table S3 and Supplementary Fig. S4. The detected SVs were compared to Bionano genomics' healthy controls database to define rare, possibly pathogenic, SVs. Less than 1% of SVs were not found in the healthy database and defined as rare variants. Figure 2A shows the rare SVs detected in the four samples in chromosomes 3 and 5. Most (60%–70%) of the rare SVs detected overlap with at least one gene. A significantly higher percentage of genes overlapping rare SVs in the advanced stage pair uniquely overlapped tumor-specific SVs (63%) compared to the early stage pair (25%) (Fig. 2B, Supplementary Data and Supplementary Table S4). The *VHL* gene locus did not overlap any rare SV in any sample; however, CNV analysis (Fig. 2C and Supplementary Fig. S5) revealed that one copy of the entire 3p chromosomal arm was lost in the advanced stage tumor (and not in the normal adjacent tissue or the early stage samples). In addition to *VHL*, this aneuploidy covers the genes *BAP1, PBRM1*, and *SETD2* that are known to be associated with ccRCC [5, 53]. Also deleted is the entire 3p26.3 cytoband, which is known to be associated with deletions causing the 3p-deletion syndrome (Del3p), typically characterized by renal and gastrointestinal abnormalities, in addition to growth retardation and developmental delay [54]. Additional aneuploidies, in chromosomes 9, 14, and 5q were detected in this sample, in line with previously reported data for ccRCC. Jonasch *et al.* [53] considered 5q gain to be an alternative ccRCC tumor initiator. Losses of 9p and 14q are considered lethal events followed by metastases, as these chromosomal arms involve genes essential for cell cycle or the metabolism of the *VHL* product, including *CDKN2A* (cyclin-dependent kinase inhibitor 2A) on chromosome 9p [55] or *HIF1A* (hypoxia-inducible factor 1A) on chromosome 14q [5, 53, 56]. Although no such aneuploidies were detected in the early stage samples, several smaller DNA gains and losses were observed, including a DNA gain (segmental duplication) on chromosome 3 spanning the genes *LINC01266* and *CNTN6*. Seven other regions, on chromosomes 3, 5, 10, and X, were found to have DNA gains or losses unique to the tumor sample, ranging between 0.5 and 6.9 Mbp in length (Supplementary Table S5). These CNVs are not known to be associated with ccRCC [7, 8, 53]. A CNV on chromosome 10 overlaps the region encoding microRNA-584, which was shown to have significantly lower expression levels in ccRCC tumors, as well as lower cell viability and motility, and was therefore marked as a tumor suppressor microRNA in ccRCC [57]. A 1.2 Mbp long DNA gain on chromosome 5 with a fractional CNV of ~2.4 copies covers the entire *FOXD1* gene locus. *FOXD*1 encodes a forkhead transcription factor belonging to a family of proteins that act as terminal effectors of several key signaling pathways, such as the mitogen-activated protein kinase (MAPK) pathway. They contribute to the regulation of homeostasis, and their misregulation can induce human genetic diseases including cancer [58]. The longest CNV in the early stage tumor is a ~6.9 Mbp DNA loss located on chromosome X. It spans several genes, including *FMR1* and microRNA genes.

Several putative events of gene fusion, which can potentially form chimeric genes from the concatenation of independent genes as a byproduct of genomic instability, were detected in both sample pairs (see Supplementary Data). One of them is a fusion of the PCDHA gene cluster, involving 14 genes (*PCDHA1-13*, *PCDHAC1*), and *AC011346*, which was caused by an intrachromosomal translocation on chromosome 5 (Fig. 2A).

An important benefit of OGM is the ability to phase structurally complex regions, such as large repetitive arrays. An example in Fig. 2D shows allele-specific copy number of a DNA repeat array in chromosome 4.

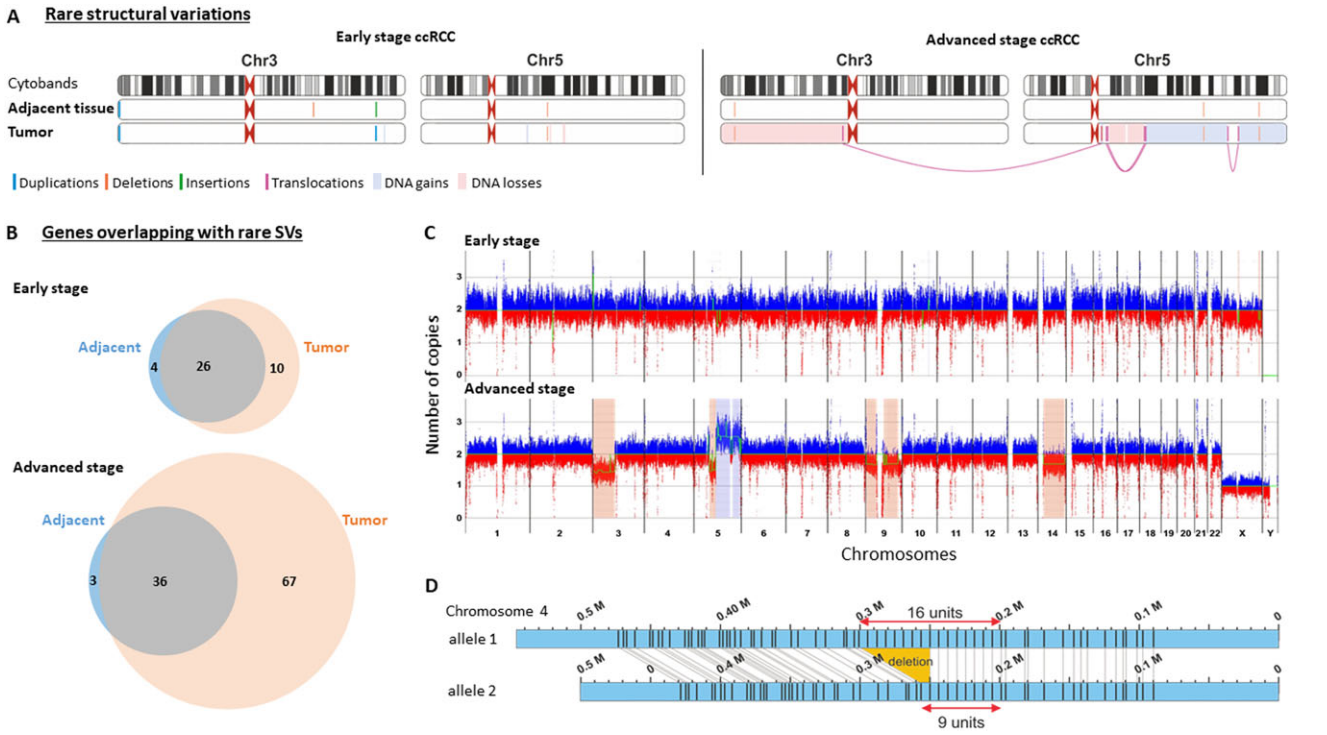


Figure 2. Next-generation cytogenetics with OGM. (**A**) Chromosomes 3 and 5 in early and advanced stage ccRCC tumors and normal adjacent tissues. Marked below the cytobands, are rare, possibly pathogenic structural variants. (**B**) Venn diagrams depicting genes overlapping with rare structural variations discovered in the early and advanced stage ccRCC tumor and normal adjacent tissues. (**C**) CNV profiles of early and advanced stage ccRCC tumors with a baseline set to two copies. DNA gains are colored in blue, DNA losses are colored in red. Aneuploidies are highlighted with colored boxes in the same color scale as DNA gains and losses. (**D**) An example of a repeat array with a different number of repeats in two alleles, discovered by OGM, showing a phased chromosomal segment (Conting IDs: 2281 and 2282).

Genome-wide and locus-specific epigenetic profiles of unmodified CpGs and 5hmC

In order to test the added value of epigenetic profiling, two epigenetic marks were labeled and optically mapped by OEM. For 5hmC, we employed an enzymatic method to attach an azide-modified glucose moiety from a synthetic cofactor (UDP-6-N3-Glu) [24, 59] (UDP-6-N3-Glu) to the hydroxyl group of 5hmC, followed by a click reaction that links a fluorophore-bound alkyne to the azide-labeled 5hmC, as detailed in Gabrieli *et al.* [22]. The reproducibility of the labeling reaction for optical mapping is demonstrated in Supplementary Fig. S6. Methylation was implied by the complementary labeling of unmodified CpG sites within TCGA sequence motifs. These were specifically labeled using the methyltransferase enzyme M.TaqI, which transfers a fluorophore from a synthetic cofactor to the adenine base in the enzyme's recognition sequence TCGA. This reaction is blocked if the CpG within this sequence is methylated or modified, resulting in a reduced representation map of unmodified CpGs (referred to as 'unmethylation'), as detailed in Sharim *et al.* [27]. The reproducibility of this labeling reaction for optical mapping is demonstrated in Supplementary Fig. S7. Our group recently applied engineered CpG methyltransferases to address all unmethylated CpGs [31, 60]. However, since these methods have not yet been validated for human methylome profiling, we opted for the previously validated reduced representation approach. Although this map covers only ~6% of the total CpGs, it captures most regulatory sites in the genome and exhibits a cell-type specific pattern [27, 30]. Since most CpG sites in the human genome are methylated [61], labeling the sparser unmodified sites enhances sensitivity in a low-resolution method as optical mapping. The genomic coverage of the epigenetic datasets is 48–106×, with molecules N50 of 180–240 kb (Table 1).

Table 1. Effective genome coverage and N50 of epigenetic experiments

Sample	Modification	Effective genome coverage	Molecules N50* (kb)
Early-stage tumor	5hmC	73×	193
	Unmethylation	99×	231
Early-stage normal adjacent tissue	5hmC	48×	187
	Unmethylation	84×	184
Advanced-stage tumor	5hmC	106×	239
	Unmethylation	83×	206
Advanced-stage normal adjacent tissue	5hmC	95×	195
	Unmethylation	53×	190

*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.

To evaluate the reproducibility of the inspected epigenetic signals across kidney samples, we compared the 5hmC signals in the tumors and normal adjacent tissues of the early and advanced stage ccRCC patients. We also compared the unmethylation signal in these and additional patients (a total of three normal adjacent tissues and four tumors). These comparisons (Supplementary Figs S8 and S9) resulted in high correlation scores (Pearson correlation coefficients: 5hmC: 0.87,

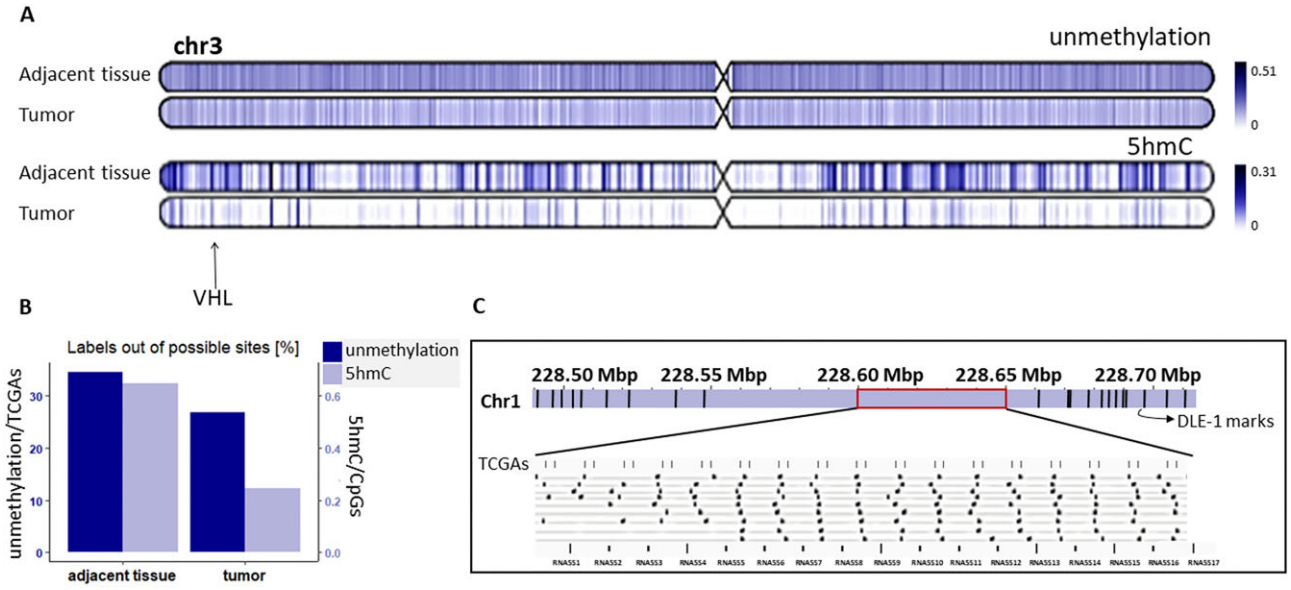


Figure 3. Genome-wide unmethylation and 5hmC profiles. (**A**) Double Ideograms showing the density of unmethylation (top) and 5hmC (bottom) along chromosome 3 in the early stage ccRCC tumor and normal adjacent tissue. (**B**) Percentages of detected unmodified CpG sites in TCGA sequence context out of all appearances of TCGA sites in hg38, and of 5hmC sites detected out of all CpG sites in hg38, in the early stage ccRCC tumor and the normal adjacent tissue. (**C**) An example of a repeat array in chromosome 1, not marked by the genetic barcode, but marked with TCGA sites. The array corresponds to RNA5S genes.

unmethylation: 0.54–0.68). To further validate our results, we have measured the advanced sample pair using Oxford nanopore sequencing (ONT) with direct epigenetic readout and compared it to the OEM profile. The epigenetic profiles acquired by the two methods are highly concordant as can be seen in Supplementary Figs S10 and S11. A comprehensive comparison between the two methods in characterizing the structural, copy number and epigenetic landscape in this sample pair can be found in [62].

Figure 3 and Supplementary Figs S12 and S13 show a chromosomal distribution and genome-wide global levels of unmethylation and 5hmC for the early (Fig. 3A and B, and Supplementary Fig. S12A and B) and advanced (Supplementary Figs S12C and D, and Supplementary Fig. S13) stage ccRCC tumor and the normal adjacent tissues. Most notably, we observed that the 5hmC profile in both tissue samples is sparse, as expected, and a global reduction is observed in the tumor sample compared to the normal adjacent tissue, in accordance with global 5hmC reduction known in many types of cancers [18]. The unmethylation signals are much denser, and the differences between the tumor and the normal adjacent tissue are mild at this perspective.

Epigenetic labeling can also provide complementary genetic information as it addresses regions where the genetic barcode is sparse. This combined approach provides basic genetic information unavailable through other methods, such as repetitive regions. Repetitive elements in DNA are further differentiated by the methylation state of the repeat units, which can affect the function of individual units or even the activity of the entire array. Methylation levels of known repeat arrays were shown to correlate with disease [63–65]. An interesting example is a tandem repeat array in chromosome 1 (1.q42), containing a domain of RNA5S genes, encoding 5S rRNA, composed of 17 units according to the human reference, but the actual cluster size is known to vary among individuals [66]. Since the repetitive region lacks DLE-1 motif, the optical genome map cannot provide an accurate copy number of the repeats. Figure 3C shows that there are 17 double unmethylation recognition sites (TCGA) throughout the array in hg38 and that the presence of these sites in the repetitive units allows both counting the copy number and assessing the methylation state of each unit, simultaneously. Supplementary Fig. S14 shows fluorescence microscopy images of two representative DNA molecules from the digitized pileup shown in Fig. 3C.

5hmC and unmethylation levels correlate with gene expression

An attractive feature of epigenetic mapping is the ability to relate the epigenetic status of genes to gene expression. To examine this aspect, protein-coding genes were divided into four groups based on their average TPM value in publicly available RNA sequencing of three tumor-matched ccRCC pairs by Yao *et al.* ([43]; see "Materials and methods" section). The mean unmethylation and 5hmC signals in each expression group were plotted against the normalized distance from the TSS (early stage pair in Fig. 4, advanced stage pair in Supplementary Fig. S15). In all samples, unmethylation signal around the TSS increases with gene expression (Fig. 4C and Supplementary Fig. S15C). In both sample pairs, the 5hmC level in gene bodies in all expression groups was lower in the tumor than in the adjacent tissue. This level increased with gene expression in the four samples (Fig. 4F and Supplementary Fig. S15F). Such correlations with gene expression suggest that epigenetic signals on DNA may serve as a proxy for gene expression, reducing the need to quantify RNA levels. We observed a similar correlation between gene expression and unmethylation signal in genes for the GM12878 cell line, also supported by whole-genome bisulfite sequencing signal (Supplementary Fig. S16). A similar correlation between gene expression and 5hmC signal in genes was previously observed in human peripheral blood cells [22].

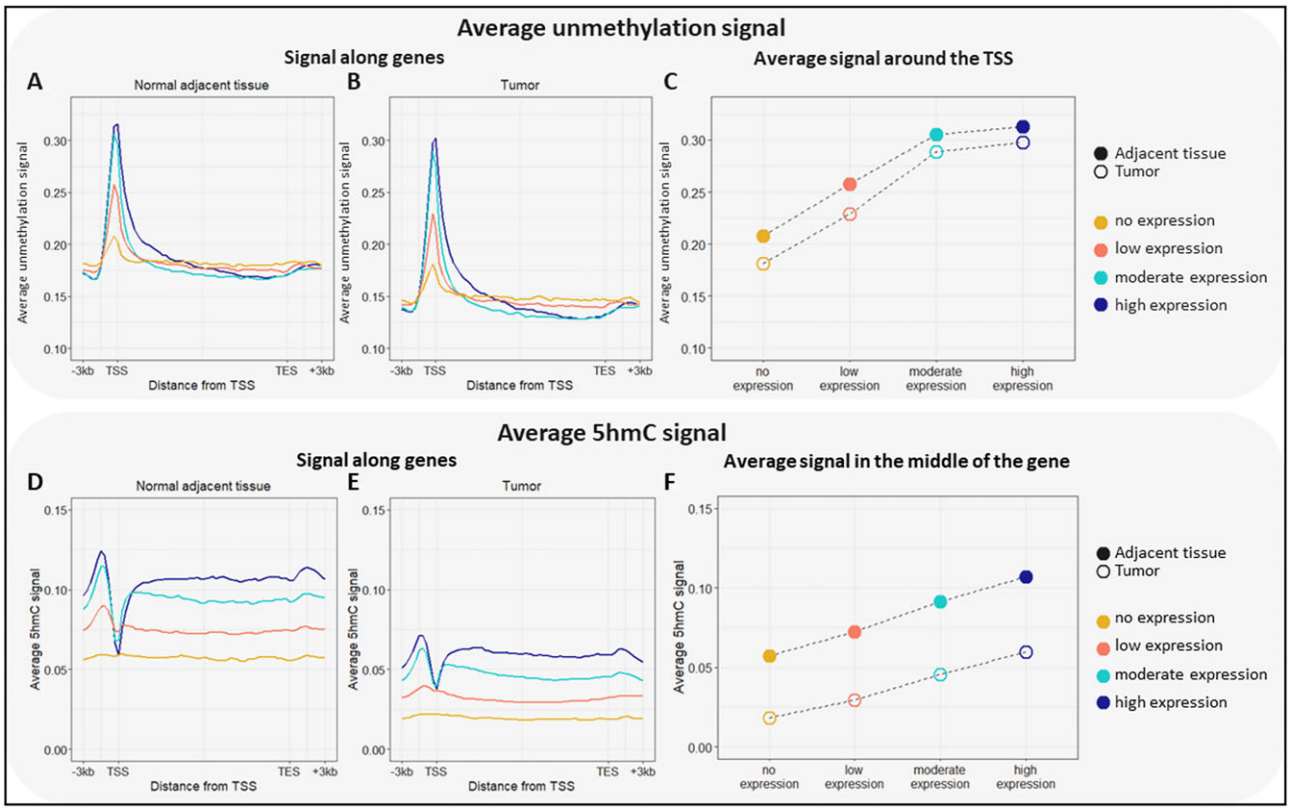


Figure 4. Epigenetic signals along genes as a function of gene expression in the early stage ccRCC sample pair. (**A**) Unmethylation signal in a normal adjacent kidney tissue in four gene groups divided according to their expression score in a publicly available RNA-seq experiment of adjacent kidney tissues. (**B**) Unmethylation signal in a ccRCC tumor in four gene groups divided according to their expression score in a publicly available RNA-seq experiment of ccRCC tumors. (**C**) Average unmethylation signal around the TSS. (**D**) 5hmC signal in a normal adjacent kidney tissue in the four gene groups from panel (A). (**E**) 5hmC signal in a ccRCC tumor in the four gene groups from panel (B). (**F**) Average 5hmC signal at gene–body midpoints.

Differentially modified regions

After characterizing the epigenetic landscape in the genomes of the tumors and the adjacent tissues, we aimed to locate differentially modified regions. Unlike conventional, short-read methods, this analysis can inspect full, long, biologically relevant regions (both annotated and unannotated) covered by individual molecules. We applied differential modification (unmethylation or 5hmC) analysis on gene bodies, promoters, general predicted enhancers, ccRCC-related enhancers, and ccRCC-related super-enhancers. Results of these analyses are visualized as volcano plots in Supplementary Figs S17–S20, while detailed lists of differentially modified annotated elements are provided in Supplementary Tables S6 and S7. Then, we focused on identifying pathways affected by these epigenetic changes by testing the enrichment of such pathways among the genes associated with the differential elements. Additionally, we analyzed unannotated 1 kb genomic windows, with the findings presented in Supplementary Figs S21–S23.

Several ccRCC-related genes or their associated annotated elements were found to be differentially modified. For example, gene bodies with differential hydroxymethylation in both sample pairs include *PRCC*, *BAP1*, and *VHL*. Gene bodies with differential unmethylation signals in the early stage sample pair include *PBRM1*, *TP53*, and *KDM5C*. General predicted enhancers with differential hydroxymethylation in both sample pairs target genes including *PRCC*, *VEGFA*, *SETD2*, and *VHL*. General predicted enhancers with differential unmethylation signals target genes including *PRCC* (both sample pairs), *VHL*, *BAP1*, and *PBRM1* (early stage sample pair). *VEGFA* is targeted by ccRCC-related enhancers and super-enhancers with differential 5hmC levels in both sample pairs and by ccRCC-related enhancers with differential unmethylation levels in the early stage sample pair. Figure 5A highlights the 5hmC and unmethylation signals in the early-stage sample pair along a region in chromosome 3 spanning the VHL gene and two general predicted enhancers targeting it. Figure 5B shows the KEGG pathways enriched among genes associated with differentially modified annotated elements in both sample pairs, including genes, promoters, general predicted enhancers, and ccRCC-related enhancers. No pathways were enriched among genes associated with ccRCC-related super-enhancers.

Many of the enrichment tests resulted in pathways related to metabolism (as metabolic pathways and other pathways). Alterations in metabolites and dysregulated metabolism are known hallmarks of cancer, and changes in several groups of essential metabolites are manifested in ccRCC. One characteristic of ccRCC cells is their morphological burden by lipid and glycogen, suggesting altered fatty acid and glucose metabolism. Additionally, changes in metabolites are associated with tumor progression and metastasis [67]. The enrichment of "Pathways in cancer" among ccRCC-related enhancers exhibiting differential unmethylation signals supports the involvement of mechanisms underlying cancer development in the early stage sample pair. The enrichment of pathways involving chemokines and cytokines among

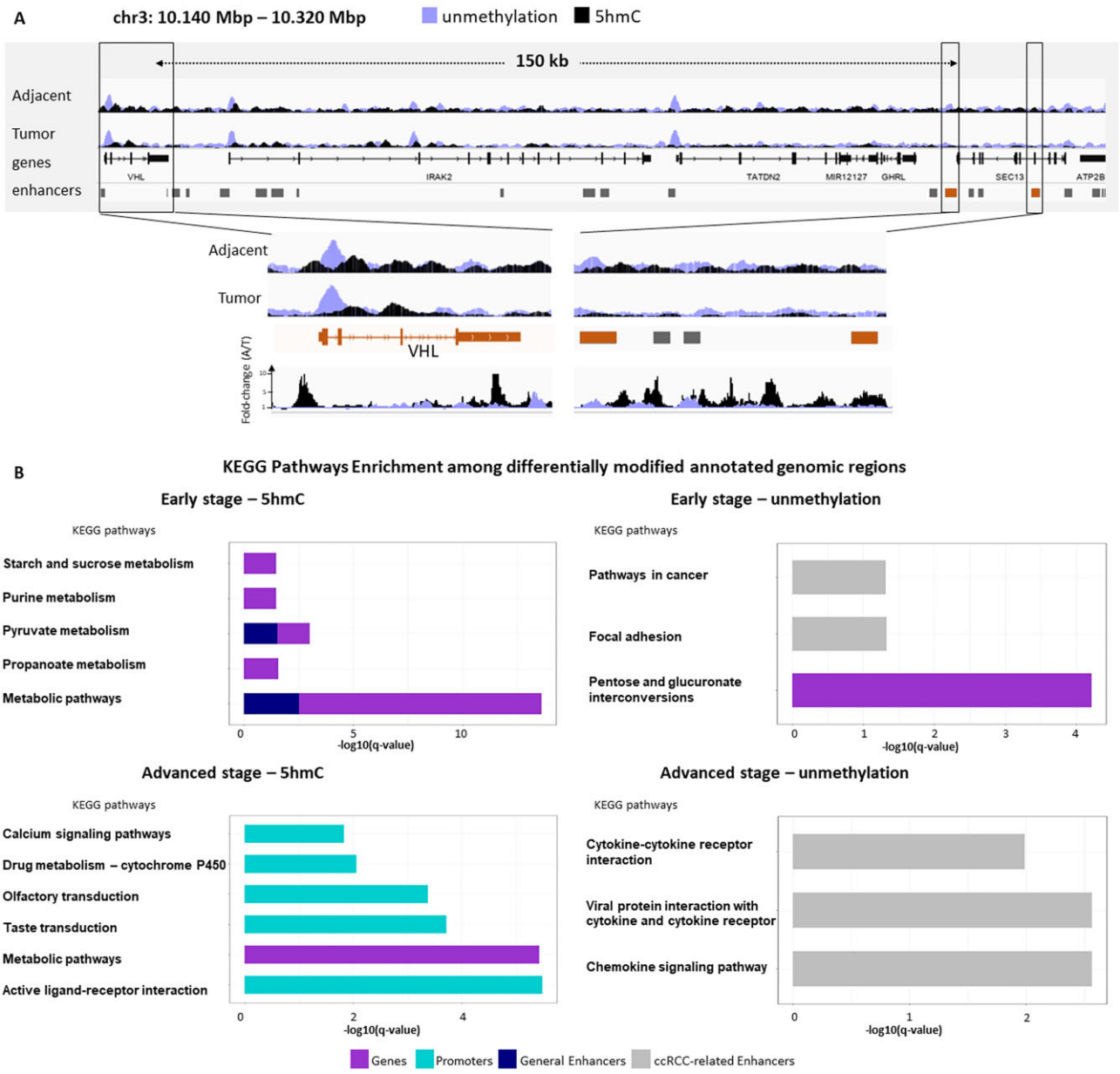


Figure 5. Differentially modified annotated genomic regions. (**A**) Epigenetic signal tracks of unmethylation (violet) and 5hmC (black) signals in the early stage sample pair, along a 180 kb region in chromosome 3 that covers the *VHL* gene and two general predicted enhancers targeting it (orange). Below is a zoomed-in view of the *VHL* gene and the enhancers. The bottom panel shows plots displaying the fold change of the signals, calculated as the signal in the normal adjacent tissue divided by the signal in the tumor tissue. (**B**) KEGG pathways enriched (q-value < 0.05) among genes associated with annotated elements that have differential unmethylation and 5hmC signals (q-value < 0.1).

ccRCC-related enhancers with differential unmethylation signals in the advanced stage sample pair is consistent with the fact that tumors interact with their environment through the secretion of these factors. Chemokines are critically involved in ccRCC progression by modulating immune responses, angiogenesis, and metastasis [68].

Discussion

OGM is a powerful tool that complements DNA sequencing. While NGS identifies short InDels and SNPs, OGM unravels complex genomic structures including SVs and CNVs of medium to large sizes. The long reads in OGM enable *de novo* construction of a sample's genomic structure, which can be particularly complex in cancer. Here, we present a complete structural comparison between tumors and matched samples in two pairs in different stages of disease. Our analysis revealed some well-established as well as some novel somatic aberrations of ccRCC [5, 53–56]. Key genetic alterations identified in this study include high-impact InDels in the tumors of both pairs (exome-seq), and extensive CNVs detected by OGM in the advanced stage tumor, such as the loss of one copy of 3p, a gain in 5q, and losses in chromosomes 9 and 14.

Beyond genetic structure, OEM can be obtained simultaneously with OGM at (almost) no extra cost, and adds significant information to it. With our epigenetic profiling added to commercial OGM, we were able to detect epigenetic changes at all resolutions in one experiment, from global,

through locus-specific, and up to single-molecule. This multi-resolution perspective provides a comprehensive understanding of the epigenetic variation across different scales. On the global level, we observed a significant reduction of 5hmC signal in the tumor sample compared with the matched adjacent tissue in both sample pairs and of unmethylation signal in one pair (early stage). The double reduction in this pair suggests that the tumor sample is both hyper-methylated and hypo-hydroxymethylated, in line with previous studies [8, 15]. The other pair exhibits very close unmethylation levels in both the tumor and adjacent tissue. At the locus-specific level, we found a correlation between gene expression and the unmethylation signal around TSSs and 5hmC signals in gene bodies. Such correlation may imply that gene expression information may be deduced from the same experiment, reducing the need for dedicated RNA sequencing, and adding additional Omics aspects to the same data. OEM provides insights into the epigenetic variability at a single-molecule resolution. Previously, we showed that the simultaneous recording of the methylation status of gene promoters and their distant enhancers enables generation of epigenetic signatures, and used these signatures to de-convolve cell mixtures [30]. In the current study, we leverage the single-molecule-level methylation and hydroxymethylation status, captured along full biologically-relevant genomic regions, to identify regions with significant differences between samples, even when accounting for inter-sample variation. This analysis identified significantly distinct regions between tumor and matched tissues. Some of these regions fall within annotated elements known to be associated with ccRCC (e.g. *PRCC*, *BAP1, VHL*, *PBRM1*; [3, 43]), further validating the results, and some within coding or non-coding regions not previously linked to ccRCC. Among annotated differentially modified regions, we found significant enrichment of metabolic pathways, and other pathways involving metabolites. Metabolism plays an important role in renal health and disease, and especially in renal cancers [67]. Additional enriched pathways align with cancerous processes (like pathways in cancer and pathways related to chemokines and cytokines). As seen here, epigenetic dysregulation of ccRCC-related genes as well as of other genomic regions is highly distinct and may inspire new studies on the mechanism of the disease.

To conclude, this study, using whole-exome sequencing, OGM and OEM, provides a uniquely comprehensive analysis of the spectrum of somatic alteration affecting ccRCC, and demonstrates the potential of combining these methods to understand carcinogenesis. The methodology can be applied to other questions in comparative genomics. Nevertheless, clinical insights will require larger cohort studies.

Acknowledgements

Author contributions: S.M.: Investigation, Formal analysis, Software, Methodology, Writing—Original draft; Z.T.: Investigation, Formal analysis, Methodology; Y.M., T.D.Z., J.D., S.L.Z., G.N., A.G., Y.S., L.S., E.W.: Investigation; Y.G., D.O., B.D: Resources; E.F.: Writing—review & editing; Y.E.: Conceptualization, Funding acquisition, Writing—original draft.

Supplementary data

Supplementary data is available at NAR Cancer online.

Conflict of interest

None declared.

Funding

This work was supported by the European Research Council consolidator [grant number 817811] to Y.E; Israel Science Foundation [grant number 771/21] to Y.E; the National Institute of Health/The National Human Genome Research Institute (NIH/NHGRI) [grant number R01HG009190] to Y.E; and Israel Innovation Authority and German Federal Ministry of Education and Research [NATI 61976 and 13GW0282B] to Y.E and E.W.

Data availability

OGM and OEM data generated in this study are available in Zenodo, with digital object identifiers: 10.5281/zenodo.14760958 for OGM and early-stage sample OEM, and 10.5281/zenodo.14266624 for advanced-stage sample OEM. The exome-seq data generated in this study were deposited to the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA1215807.

References

1. López JI. Renal tumors with clear cells. A review. *Pathology* 2013;**209**:137–46. https://doi.org/10.1016/j.prp.2013.01.007
2. Bacigalupa ZA, Rathmell WK. Beyond glycolysis: hypoxia signaling as a master regulator of alternative metabolic pathways and the implications in clear cell renal cell carcinoma. *Cancer Lett* 2020;**489**:19–28. https://doi.org/10.1016/j.canlet.2020.05.034
3. Gossage L, Eisen T, Maher ER. VHL, the story of a tumour suppressor gene. *Nat Rev Cancer* 2015;**15**:55–64. https://doi.org/10.1038/nrc3844
4. Duns G, Hofstra RMW, Sietzema JG *et al.* Targeted exome sequencing in clear cell renal cell carcinoma tumors suggests aberrant chromatin regulation as a crucial step in ccRCC development. *Hum Mutat* 2012;**33**:1059–62. https://doi.org/10.1002/humu.22090
5. Le VH, Hsieh JJ. Genomics and genetics of clear cell renal cell carcinoma: a mini-review. *JTGG* 2018;**2**:17. https://doi.org/10.20517/jtgg.2018.28
6. Moore LE, Jaeger E, Nickerson ML *et al.* Genomic copy number alterations in clear cell renal carcinoma: associations with case characteristics and mechanisms of VHL gene inactivation. *Oncogenesis* 2012;**1**:e14. https://doi.org/10.1038/oncsis.2012.14
7. Quddus M, Pratt N, Nabi G. Chromosomal aberrations in renal cell carcinoma: an overview with implications for clinical practice. *Urol Ann* 2019;**11**:6–14.
8. Klatte T, Rao PN, De Martino M *et al.* Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *JCO* 2009;**27**:746–53. https://doi.org/10.1200/JCO.2007.15.8345
9. Shenoy N, Vallumsetla N, Zou Y *et al.* Role of DNA methylation in renal cell carcinoma. *J Hematol Oncol* 2015;**8**:88. https://doi.org/10.1186/s13045-015-0180-y
10. Hu CY, Mohtat D, Yu Y *et al.* Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. *Clin Cancer Res* 2014;**20**:4349–60. https://doi.org/10.1158/1078-0432.CCR-14-0494
11. Arai E, Chiku S, Mori T *et al.* Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Carcinogenesis* 2012;**33**:1487–93. https://doi.org/10.1093/carcin/bgs177

12. Lasseigne BN, Brooks JD. The role of DNA methylation in renal cell carcinoma. *Mol Diagn Ther* 2018;**22**:431–42. https://doi.org/10.1007/s40291-018-0337-9
13. Joosten SC, Odeh SNO, Koch A *et al.* Development of a prognostic risk model for clear cell renal cell carcinoma by systematic evaluation of DNA methylation markers. *Clin Epigenet* 2021;**13**:103. https://doi.org/10.1186/s13148-021-01084-8
14. Wang J, Zhang Q, Zhu Q *et al.* Identification of methylation-driven genes related to prognosis in clear-cell renal cell carcinoma. *J Cell Physiol* 2020;**235**:1296–308. https://doi.org/10.1002/jcp.29046
15. Shi DQ, Ali I, Tang J *et al.* New insights into 5hmC DNA modification: generation, distribution and function. *Front Genet* 2017;**8**:100. https://doi.org/10.3389/fgene.2017.00100
16. Jin SG, Jiang Y, Qiu R *et al.* 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 2011;**71**:7360–5. https://doi.org/10.1158/0008-5472.CAN-11-2023
17. Margalit S, Avraham S, Shahal T *et al.* 5-Hydroxymethylcytosine as a clinical biomarker: fluorescence-based assay for high-throughput epigenetic quantification in human tissues. *Intl J Cancer* 2020;**146**:115–22. https://doi.org/10.1002/ijc.32519
18. Chen K, Zhang J, Guo Z *et al.* Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res* 2016;**26**:103–18. https://doi.org/10.1038/cr.2015.150
19. Chen S, Zhou Q, Liu T *et al.* Prognostic value of downregulated 5-hydroxymethyl-cytosine expression in renal cell carcinoma: a 10 year follow-up retrospective study. *J Cancer* 2020;**11**:1212–22. https://doi.org/10.7150/jca.38283
20. Yu M, Han D, Hon GC *et al.* Tet-assisted bisulfite sequencing (TAB-seq). *Methods Mol Biol* 2018;**1708**:645–63. https://doi.org/10.1007/978-1-4939-7481-8_33
21. Booth MJ, Ost TWB, Beraldi D *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* 2013;**8**:1841–51. https://doi.org/10.1038/nprot.2013.115
22. Gabrieli T, Sharim H, Nifker G *et al.* Epigenetic optical mapping of 5-hydroxymethylcytosine in nanochannel arrays. *ACS Nano* 2018;**12**:7148–58. https://doi.org/10.1021/acsnano.8b03023
23. Michaeli Y, Shahal T, Torchinsky D *et al.* Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem Commun* 2013;**49**:8599–601. https://doi.org/10.1039/c3cc42543f
24. Nifker G, Levy-Sakin M, Berkov-Zrihen Y *et al.* One-pot chemoenzymatic cascade for labeling of the epigenetic marker 5-hydroxymethylcytosine. *ChemBioChem* 2015;**16**:1857–60. https://doi.org/10.1002/cbic.201500329
25. Shahal T, Gilat N, Michaeli Y *et al.* Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA. *Anal Chem* 2014;**86**:8231–7. https://doi.org/10.1021/ac501609d
26. Jeffet J, Margalit S, Michaeli Y *et al.* Single-molecule optical genome mapping in nanochannels: multidisciplinarity at the nanoscale Jonathan. *Essays Biochem* 2021;**65**:51–66. https://doi.org/10.1042/EBC20200021
27. Sharim H, Grunwald A, Gabrieli T *et al.* Long-read single-molecule maps of the functional methylome. *Genome Res* 2019;**29**:646–56. https://doi.org/10.1101/gr.240739.118
28. Mak ACY, Lai YYY, Lam ET *et al.* Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics* 2016;**202**:351–62. https://doi.org/10.1534/genetics.115.183483
29. Detinis ZurT, Margalit S, Jeffet J *et al.* Single-molecule toxicogenomics: optical genome mapping of DNA-damage in nanochannel arrays. *DNA Repair* 2025;**146**:103808. https://doi.org/10.1016/j.dnarep.2025.103808
30. Margalit S, Abramson Y, Sharim H *et al.* Long reads capture simultaneous enhancer–promoter methylation status for cell-type deconvolution. *Bioinformatics* 2021;**37**:i327–33. https://doi.org/10.1093/bioinformatics/btab306
31. Gabrieli T, Michaeli Y, Avraham S *et al.* Chemoenzymatic labeling of DNA methylation patterns for single-molecule epigenetic mapping. *Nucleic Acids Res* 2022;**50**:e92. https://doi.org/10.1093/nar/gkac460
32. Uppuluri L, Jadhav T, Wang Y *et al.* Multicolor whole-genome mapping in nanochannels for genetic analysis. *Anal Chem* 2021;**93**:9808–16. https://doi.org/10.1021/acs.analchem.1c01373
33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
34. Van der Auwera GA, O'connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st Edition) O'Reilly Media. 2020.
35. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008
36. Cingolani P, Platts A, Wang LL *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 2012;**6**:80–92. https://doi.org/10.4161/fly.19695
37. Sherry ST, Ward MH, Kholodov M *et al.* DbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;**29**:308–11. https://doi.org/10.1093/nar/29.1.308
38. Grunwald A, Dahan M, Giesbertz A *et al.* Bacteriophage strain typing by rapid single molecule analysis. *Nucleic Acids Res* 2015;**43**:e117. https://doi.org/10.1093/nar/gkv563
39. Frankish A, Diekhans M, Ferreira AM *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;**47**:D766–73. https://doi.org/10.1093/nar/gky955
40. Cao Q, Anyansi C, Hu X *et al.* Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat Genet* 2017;**49**:1428–36. https://doi.org/10.1038/ng.3950
41. Haeussler M, Zweig AS, Tyner C *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* 2019;**47**:D853–8. https://doi.org/10.1093/nar/gky1095
42. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;**9**:9354. https://doi.org/10.1038/s41598-019-45839-z
43. Yao X, Tan J, Lim KJ *et al.* VHL deficiency drives enhancer activation of oncogenes in clear cell renal cell carcinoma. *Cancer Discov* 2017;**7**:1284–305. https://doi.org/10.1158/2159-8290.CD-17-0375
44. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033
45. Neph S, Kuehn MS, Reynolds AP *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* 2012;**28**:1919–20. https://doi.org/10.1093/bioinformatics/bts277
46. Hao Z, Lv D, Ge Y *et al.* RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* 2020;**6**:e251. https://doi.org/10.7717/peerj-cs.251
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;**25**:1105–11. https://doi.org/10.1093/bioinformatics/btp120
48. Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res* 2011;**39**:D19–21. https://doi.org/10.1093/nar/gkq1019
49. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**:166–9. https://doi.org/10.1093/bioinformatics/btu638
50. Ramírez F, Dündar F, Diehl S *et al.* DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;**42**:W187–91. https://doi.org/10.1093/nar/gku365
51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;**57**:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

52. Yu G, Wang LG, Han Y *et al.* ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7. https://doi.org/10.1089/omi.2011.0118
53. Jonasch E, Walker CL, Rathmell WK. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. *Nat Rev Nephrol* 2021;**17**:245–61. https://doi.org/10.1038/s41581-020-00359-2
54. Gandawijaya J, Bamford RA, Burbach JPH *et al.* Cell adhesion molecules involved in neurodevelopmental pathways implicated in 3p-deletion syndrome and autism spectrum disorder. *Front Cell Neurosci* 2021;**14**:611379. https://doi.org/10.3389/fncel.2020.611379
55. Di Nunno V, Mollica V, Brunelli M *et al.* A meta-analysis evaluating clinical outcomes of patients with renal cell carcinoma harboring chromosome 9P loss. *Mol Diagn Ther* 2019;**23**:569–77. https://doi.org/10.1007/s40291-019-00414-0
56. Monzon FA, Alvarez K, Peterson L *et al.* Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Mod Pathol* 2011;**24**:1470–9. https://doi.org/10.1038/modpathol.2011.107
57. Ueno K, Hirata H, Shahryari V *et al.* Tumour suppressor microRNA-584 directly targets oncogene Rock-1 and decreases invasion ability in human clear cell renal cell carcinoma. *Br J Cancer* 2011;**104**:308–15. https://doi.org/10.1038/sj.bjc.6606028
58. Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: key players in health and disease. *Trends Genet* 2011;**27**:224–32. https://doi.org/10.1016/j.tig.2011.03.003
59. Song C-X, Szulwach KE, Fu Y *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 2011;**29**:68–72. https://doi.org/10.1038/nbt.1732
60. Avraham S, Schütz L, Käver L *et al.* Chemo-enzymatic fluorescence labeling of genomic DNA for simultaneous detection of global 5-methylcytosine and 5-hydroxymethylcytosine. *ChemBioChem* 2023;**24**:e202300400. https://doi.org/10.1002/cbic.202300400
61. Uroshlev LA, Abdullaev ET, Umarova IR *et al.* A method for identification of the methylation level of CpG islands from NGS data. *Sci Rep* 2020;**10**:8635. https://doi.org/10.1038/s41598-020-65406-1
62. Margalit S, Tulpová Z, Zur TD *et al.* Long-read structural and epigenetic profiling of a kidney tumor-matched sample with nanopore sequencing and optical genome mapping. *NAR Genomics Bioinform* 2025;**7**:lqae190. https://doi.org/10.1093/nargab/lqae190
63. Balog J, Miller D, Sanchez-Curtailles E *et al.* Epigenetic regulation of the X-chromosomal macrosatellite repeat encoding for the cancer/testis gene CT47. *Eur J Hum Genet* 2012;**20**:185–91. https://doi.org/10.1038/ejhg.2011.150
64. Pook MA. DNA methylation and trinucleotide repeat expansion diseases. In: *DNA Methylation - From Genomics to Technology*. InTech, 2012, 193–208.
65. Hansen KD, Timp W, Bravo HC *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 2011;**43**:768–75. https://doi.org/10.1038/ng.865
66. Stults DM, Killen MW, Pierce HH *et al.* Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res* 2008;**18**:13–8. https://doi.org/10.1101/gr.6858507
67. Hakimi AA, Reznik E, Lee C-H *et al.* An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell* 2016;**29**:104–16. https://doi.org/10.1016/j.ccell.2015.12.004
68. Monjaras-Avila CU, Lorenzo-Leal AC, Luque-Badillo AC *et al.* The tumor immune microenvironment in clear cell renal cell carcinoma. *Int J Mol Sci* 2023;**24**:7946. https://doi.org/10.3390/ijms24097946

Received: September 12, 2024. **Revised:** February 18, 2025. **Editorial Decision:** February 20, 2025. **Accepted:** March 5, 2025