

# Targeted long-read sequencing allows for rapid identification of pathogenic disease-causing variants in retinoblastoma

Kenji Nakamichi, Andrew Stacey & Debarshi Mustafi

To cite this article: Kenji Nakamichi, Andrew Stacey & Debarshi Mustafi (2022) Targeted long-read sequencing allows for rapid identification of pathogenic disease-causing variants in retinoblastoma, *Ophthalmic Genetics*, 43:6, 762-770, DOI: [10.1080/13816810.2022.2141797](https://doi.org/10.1080/13816810.2022.2141797)

To link to this article: <https://doi.org/10.1080/13816810.2022.2141797>



Published online: 03 Nov 2022.



Submit your article to this journal 



Article views: 178



View related articles 



View Crossmark data 

## Targeted long-read sequencing allows for rapid identification of pathogenic disease-causing variants in retinoblastoma

Kenji Nakamichi<sup>a</sup>, Andrew Stacey<sup>a,b</sup>, and Debarshi Mustafi<sup>a,c</sup> 

<sup>a</sup>Department of Ophthalmology and Roger and Karalis Johnson Retina Center, University of Washington, Seattle, WA, USA; <sup>b</sup>Department of Ophthalmology, Seattle Children's Hospital, Seattle, WA, USA; <sup>c</sup>Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

### ABSTRACT

**Background:** Identification of disease-causing variants of the retinoblastoma gene (*RB1*), the predominant cause of retinoblastoma, is challenging. Targeted long-read genome sequencing offers a novel approach to resolve the diverse range of pathogenic variants in *RB1* and provides haplotype information rapidly.

**Materials and Methods:** Genomic DNA was isolated from a venipuncture blood draw of a retinoblastoma patient. Whole genome sequencing (WGS) was carried out using the short-read Illumina platform. WGS and targeted sequencing of *RB1* was accomplished using the long-read Oxford Nanopore Technologies (ONT) platform. Deep-learning frameworks allowed haplotagging, variant calling, and variant annotation of both short- and long-read data.

**Results:** Targeted long-read sequencing of the *RB1* gene allowed for enhanced depth of read coverage for discovery of rare variants and haplotype analysis. A duplication leading to a frameshift and early termination in *RB1* was identified as the most deleterious variant by all sequencing methods, with long-read technology providing additional information of methylation signal and haplotype information. More importantly, there was greater than 98% concordance of *RB1* variants identified between short-read and targeted long-read sequencing modalities.

**Conclusions:** Targeted long-read technology allows for focused sequencing effort for variant discovery. Application of this for the first time in a retinoblastoma patient allowed haplotagged variant identification and demonstrated excellent concordance with benchmark short-read sequencing. The added benefit of targeted long-read sequencing to resolve disease-causing genomic variation in *RB1* rapidly from a blood draw will provide a more definitive diagnosis of heritable RB and guide management decisions for patients and their families.

### ARTICLE HISTORY

Received October 3, 2022

Revised October 11, 2022

Accepted October 20, 2022

### KEYWORDS

*RB1*; retinoblastoma;  
long-read sequencing;  
adaptive sampling; targeted  
sequencing; haplotyping;  
methylation

## Introduction

Retinoblastoma (RB) is the most common pediatric intraocular neoplasm that occurs in 1 in 14,000 births (1). Genomic studies have indicated that inactivation of the tumor suppressor retinoblastoma gene (*RB1*) (2,3) is necessary but not sufficient for tumorigenesis (4) and *RB1* is the cause of nearly all cases of retinoblastoma (5). More importantly, patients with germline *RB1* pathogenic variants are at risk of subsequent malignant neoplasms such as skin melanoma and carcinoma, pinealoblastoma, osteosarcoma, and other sarcomas that are associated with significant morbidity and mortality (6). The main global burden of RB lies in lower and middle income countries (7) where patient survival rate is less than 30% (8). In contrast, disease-free survival rates are as high as 97% in higher income countries where early detection and multiple treatment modalities are available (9). Thus, early screening and genetic diagnosis is paramount in achieving optimal care delivery and determining the disease risk for the extended family.

Mutational analysis of the *RB1* gene is a challenging task, owing to its large size and heterogeneity of over 200 pathogenic disease-causing variants. Consistent with other pediatric cancer types (10), the overall tumor mutational burden is low in

RB with lack of mutational hotspots in *RB1* (11), thus requiring comprehensive coverage of this gene with sequencing reads for accurate detection of disease-causing variants. The *RB1* locus is located on chromosome 13 (12) and spans 27 exons spread over 180,000 base pairs (bp). *RB1* constitutes less than 0.003% of total genomic DNA, meaning that 99.997% of sequencing effort with whole genome approaches is not informative in identifying pathogenic *RB1* variants. Capture-based NGS DNA sequencing panels targeting exons and non-coding regions of suspected cancer-causing genes can reduce the sequencing footprint to a few megabases (13,14), but *RB1* still only constitutes less than 5% of that search space, indicating 95% of sequencing effort from these panels are still not targeting *RB1*. Furthermore, target panel enrichment methods cannot be altered easily and PCR-based methods cannot capture disease-specific epigenetic changes such as methylation information without additional processing.

Whereas the majority of *RB1* mutations can be detected by clinical exome sequencing using targeted panels, genome sequencing (GS) is essential to uncover structural variants in coding portions of *RB1* as well as variants in non-coding portions of the *RB1* gene (15). Next generation sequencing (NGS)

can detect relevant mutations (16) and can help correlate with potentially aggressive histopathologic features (17). However, there are potential disease-causing variants in *RB1* that can be missed with more commonly used short-read genome sequencing approaches. Furthermore, short-read sequencing methods usually only span a single variant and thus cannot provide genetic phase information. Phasing genetic variation is essential to understand on which chromosome a variant lies, in order to elucidate potential inheritance patterns. Targeted enrichment for the *RB1* gene locus from genomic DNA with long-read sequencing technology would avoid wasting sequencing bandwidth on uninformative reads and allow much deeper coverage from the same sequencing effort. Furthermore, since native-base modifications are preserved with long-read sequencing, DNA methylation signals can be discerned. Taken together, targeted long-read sequencing offers an approach to resolve the full complexity of the genomic rearrangements in *RB1* and provide genetic phase information more rapidly.

In this work we carry out WGS using both short-read (Illumina) and long-read (Oxford Nanopore Technologies (ONT)) of a patient with a clinical diagnosis of RB. We demonstrate that both techniques identify the same clinically disease-relevant variants, but the ONT long-read platform is able to do so by sequencing the native DNA, thus eliminating amplification bias while preserving base modifications (18–20). Aberrant methylation can provide epi-mutational signatures in RB (21,22) and reveal characteristic signals in RB tumors (23). Moreover, we show that the ONT platform can be adapted to selectively enrich for the *RB1* gene locus and its surrounding regions to achieve greater sequencing depth for variant detection and haplotype reconstruction (24). This targeted genomic sequencing, termed ReadUntil or adaptive sampling (25,26), was used to capture information on key genomic targets without the need for custom library preparation. Furthermore, we demonstrate that implementation of a haplotype-aware genotyping pipeline is able to identify clinically relevant genomic variants from targeted long-read sequencing data on par with commonly used short-read methods. More importantly, sequencing, haplotype analysis, and pathogenic variant prioritization of *RB1* can be achieved in days rather than weeks required for clinical testing results to be available to the clinician. This work would have a tremendous impact in providing a more definitive diagnosis of a patient suspected to have heritable RB to better guide clinical care.

## Material and methods

This study was approved by the institutional review board at the University of Washington (STUDY00014158). Written informed consent was obtained from the study subject. Experiments were conducted according to the principles expressed in the Declaration of Helsinki. Clinical diagnosis of RB was based on history and ophthalmologic findings.

### **Extraction of genomic DNA and sequencing library preparation**

A venipuncture blood of 2 mL was obtained from the study subject and genomic DNA (gDNA) was isolated using the

MagAttract High Molecular Weight genomic DNA isolation kit (Qiagen). Approximately 750 nanograms of gDNA was sheared using a Covaris LE220 focused ultrasonicator targeting 380bp inserts and then subjected to a series library construction steps utilizing the Roche KAPA Hyper Prep kit (KR0961 v1.14) for short-read Illumina sequencing. For long-read sequencing, approximately 1.2 µg of gDNA was used to make sequencing libraries using the ONT Ligation Sequencing Kit (SQK-LSK110) following the manufacturer's instructions. The long fragment buffer was used during the clean up to enrich for fragments greater than 3 kilobases in size. After the final elution step, the eluate was removed and 1 mL quantified by Qubit dsDNA HS assay. The appropriate volume to obtain 5–50 femtomole of DNA was diluted to 12 µL with the elution buffer for loading onto a ONT flow cell.

### **Short-read whole genome sequencing, variant calling and variant annotation**

For the short-read data, base calls were generated in real-time on the Illumina NovaSeq6000 instrument. BAM files were aligned to a human reference (GRCh38) using Burrows-Wheeler Aligner; v0.7.15 (27). A Genome Analysis Toolkit (GATK) (28) (v4.2.6.1) based pipeline following the best practices was used. The reads were then filtered for pairing and minimum alignment mapping quality (MAPQ) score of 50, then the supplementary, secondary, and optical duplicates were removed. The filtered BAM file was variant called using GATK HaplotypeCaller, and the output variant call file (VCF) underwent base quality score recalibration (BQSR) using GATK BaseRecalibrator. The recalibration tables were then used with GATK ApplyBQSR to recalibrate the base quality scores, and the recalibrated BAM file then underwent a second round of variant calling using GATK HaplotypeCaller. The resulting variant call files underwent several variant quality score recalibration (VQSR) steps using GATK VariantRecalibrator with parameters tuned for WGS. The resulting recalibration table and tranches files were then applied using GATK ApplyVQSR sequentially in SNP and INDEL modes. The recalibrated VCF file was then split into SNPs and INDELs using GATK SelectVariants, and filtered using GATK VariantFiltration with tuned parameters. The VCF file was then split into passing variants with a minimum allele depth of 15.

### **Long-read whole genome sequencing and targeted enrichment of *RB1***

Long-read whole genome sequencing was carried out using the ONT PromethION platform with real-time basecalling. For targeted long-read sequencing on the MinION, a R9.4.1 flow-cell was used. A graphical processing unit (GPU) accelerated version of guppy (v6.0.7; API version 10.1.0) was used for basecalling in real-time on two NVIDIA RTX A6000 GPUs using the “super-accurate” model parameters. Adaptive sampling (26) was run using the ONT MinION (v22.03.6) software for in silico enrichment of pre-selected target regions of *RB1* and 100 kilobases of DNA surrounding the gene (chr13:48,253,747–48,531,728). The adaptive sampling mode

was set to enrich the *RB1* locus on either 256 of the 512 available channels in each run or on all 512 channels. Reads were mapped using Minimap2 (v2.22-r1101). Sequencing experiments were run for up to 48–72 hours with a nuclease flush and library reload after 24–48 hours to recover maximal pores for continued sequencing.

#### **Sequence haplotagging, variant calling, and variant annotation of long-read data**

FASTQ files were generated using Guppy and aligned to GRCh38 using minimap2 (29). The BAM file was collated, duplicates marked and the reads filtered for a minimum alignment quality score of MAPQ 50 and secondary, supplementary, and optical duplicates were removed using SAMtools (v1.13–5). Variants were called using PEPPER and haplotyping was achieved using Margin. The DeepVariant pipeline was used to generate a phased VCF (24) and haplotagged BAM file. SAMtools was used to isolate reads that mapped to the *RB1* target region. The Combined Annotation-Dependent Depletion (CADD) (30,31) score, which integrates diverse genome annotations and scores any possible human SNV or indel event for their deleterious nature, was generated for each phased VCF, to provide a quantitative prediction of deleteriousness, pathogenicity, and molecular functionality of the identified variants.

## **Results**

We selected a patient with a clinical diagnosis of RB. There was no family history of RB. There was noted leukocoria in infancy and the patient was diagnosed with bilateral RB at 9 months of age. The left eye underwent enucleation and the right eye was treated with external beam radiation. The patient was disease free until the second decade of life when a secondary left parotid tumor was identified and excised. The patient also developed radiation retinopathy in her right eye in the third decade of life that has been well controlled with intravitreal steroid implants. To characterize the genetic etiology of disease, the patient underwent exome sequencing at University of Washington using the BROCA Cancer Risk Panel on the Illumina short-read sequencing platform. The panel revealed a heterozygous duplication (c.2330dup) in the *RB1* gene as the likely disease-causing variant. To carry out genome sequencing experiments on short-read and long-read platforms, high-molecular weight gDNA was isolated from a venipuncture blood draw of just 2 mL, compared to approximately 10 mL required by most clinical sequencing experiments, for sequencing library preparation.

#### **Short- and long-read whole genome sequencing studies reveal agreement in identified *RB1* variants**

Short-read WGS was carried out using the Illumina NovaSeq6000 platform. Short-read sequencing is the industry benchmark for variant identification in RB patients. To establish the feasibility of long-read technology in variant detection from gDNA, WGS was also carried out using long-read technology. Long-read WGS was carried out on the ONT PromethION platform. Sequencing read outputs from each platform were aligned to the human reference genome (GRCh38). The mean read depth of the *RB1* region was approximately 30X for both short-read and long-read sequencing experiments. The average read length for reads with MAPQ score greater than 50 was 151 base pairs (bp) and the average quality score was 29.2 from the NovaSeq6000 compared to 11,030 bp and average quality score of 23.9 from the PromethION (Table 1). The short-read data was analyzed using GATK for variant calling and annotation. Long-read data was analyzed using the PEPPER-Margin-DeepVariant pipeline for haplotagged variant analysis. The variant call files from both short- and long-read sequencing were then subjected to CADD analysis for ranked scoring of the deleteriousness of each of the identified variants. Analysis of the CADD identified variants revealed that the first pathogenic variant in chromosome 13 detected by both modalities was in *RB1* and corresponded to the variant that had previously been identified by exome sequencing. Overall, the variant in *RB1* was the 9th ranked variant by CADD from both sequencing modalities in terms of deleteriousness in the genome. Examination of the 8 preceding variants revealed that the variants resided in genes not known to be associated with RB (*ANKDD1B*, *DNAH11*, *FUT2*, *PKD1L2*, *LPL*, *TTC4*, *WDR31*). Moreover, only the variant in WD Repeat-Containing Protein 31 (*WDR31*), a gene encoding a proposed ciliary protein, had a rare allelic frequency ( $6 \times 10^{-4}$ ) that would meet the criterion to be considered pathogenic (32). This demonstrated that long-read sequencing provides the same essential results as short-read sequencing. The next step was to determine if targeted long-read sequencing could provide further enrichment of the data for *RB1* in a more rapid timeframe, while preserving the base-level accuracy. To address this we carried out bioinformatic targeted sequencing of the *RB1* locus on the ONT platform.

#### **Targeted long-read sequencing allows for selective enrichment of the *RB1* locus**

Targeted sequencing of the *RB1* gene locus was successfully implemented using adaptive sampling for the ONT platform by rapidly resetting pores processing sequences bioinformatically that were not contained in the *RB1* locus. To demonstrate

**Table 1.** Summary of sequencing parameters across the different platforms.

Sequencing modality	Read depth	Average read length (MAPQ score>50)	Average Base Quality Score
Short-read WGS	30	151 bp	29.2
Long-read WGS	30	11,030 bp	23.9
Long-read targeted sequencing	35	9,721 bp	24.1

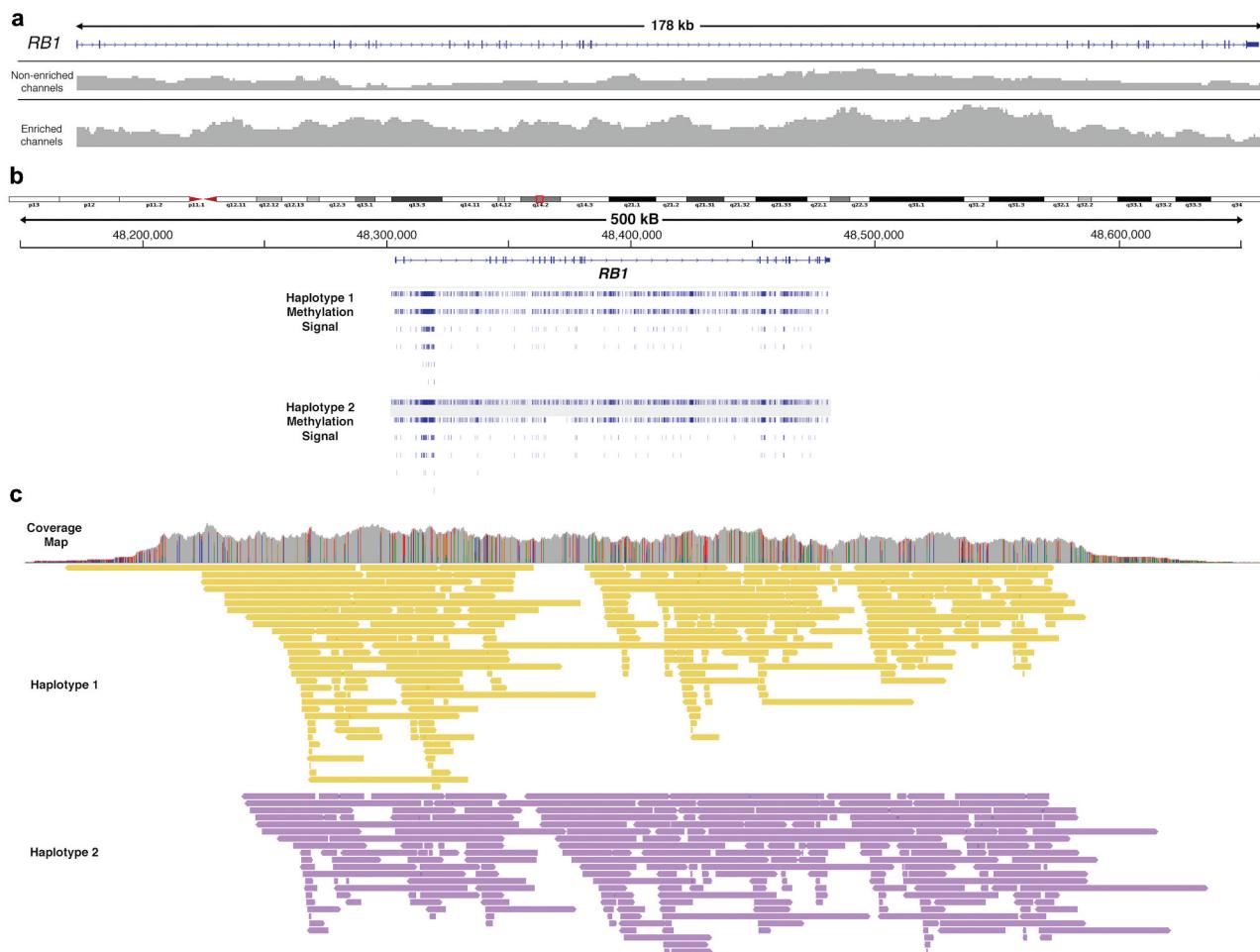
the enrichment methodology of this approach, half of the 512 sequencing pores on the flow cell were subjected to targeted sequencing whereas the other half of the sequencing pores had no target selection to enable calculation of fold enrichment using targeted sequencing. This resulted in a 5-fold relative enrichment of *RB1* compared to background and provided superior base coverage of *RB1* on the enriched channels (Figure 1a). Native DNA was subjected to library preparation and sequencing, which allowed calling of 5-methylcytosine signal across the *RB1* locus (Figure 1b). Targeted sequencing of the *RB1* locus from this patient resulted in 100% base coverage of the *RB1* region with mean sequencing depth of 34.8 with MAPQ scores greater than 50 and mean quality score of 24.1 (Table 1). More importantly, 99.93% of the targeted region of *RB1* had a depth of coverage greater than 15 with MAPQ scores greater than 50, which allowed for haplotyping and variant identification (Figure 1c).

The average read length was 9721 bp with a few reads stretching greater than 140,000 bp, which accounts for the majority of the length of the *RB1* gene. Notably for aligned

reads with MAPQ scores greater than 50, the targeted approach covered an additional 7624 bp, whereas the short-read data only covered 79 bp not captured by the targeted sequencing data. Deliberately rejected reads had an average read length of 455 bases, illustrating how this technology allows one to quickly reject a read and move to another to capture target DNA. More importantly, since the region is targeted bioinformatically, no special processing of the gDNA library is required. The specificity of targeted sequencing is clearly evidenced by the low sequencing coverage outside of the boundaries of the pre-designated *RB1* locus (Figure 1). The targeted sequencing and base-calling were all completed between 48 to 72 hours.

### **Targeted long-read sequencing demonstrates clinical overlap of identified variants from whole genome short-read sequencing**

Based on the assumption that genomic variation occurs 1 in every 1000 base pairs (33), we expected to uncover 200 variants



**Figure 1.** Targeted sequencing of the *RB1* loci from blood of a patient with RB provides increased depth of sequencing coverage for methylation calling, haplotagging and variant identification. (a) adaptive sampling leads to a 5-fold increase in sequencing coverage of the *RB1* locus on channels targeting selectively *RB1* from human blood compared to the control, non-enriched channels. furthermore, compared to the non-enriched channels which do not display complete base coverage of the *RB1* gene locus, the enriched channels provide 100% base coverage with superior depth of sequencing necessary for variant identification. (b) sequencing of native DNA allowed detection of 5-methylcytosine signals across the *RB1* gene locus. (c) the colored coverage map at the top shows 100% base coverage of targeted *RB1* gene locus at a sequencing depth of 36X. Targeted sequencing allows robust coverage in the region indicated by the black arrows with low base coverage outside of this selected genomic region (less than 1X coverage). The haploid genome assemblies for this patient are shown colored in the bottom panels (haplotype 1 in yellow and haplotype 2 in purple).

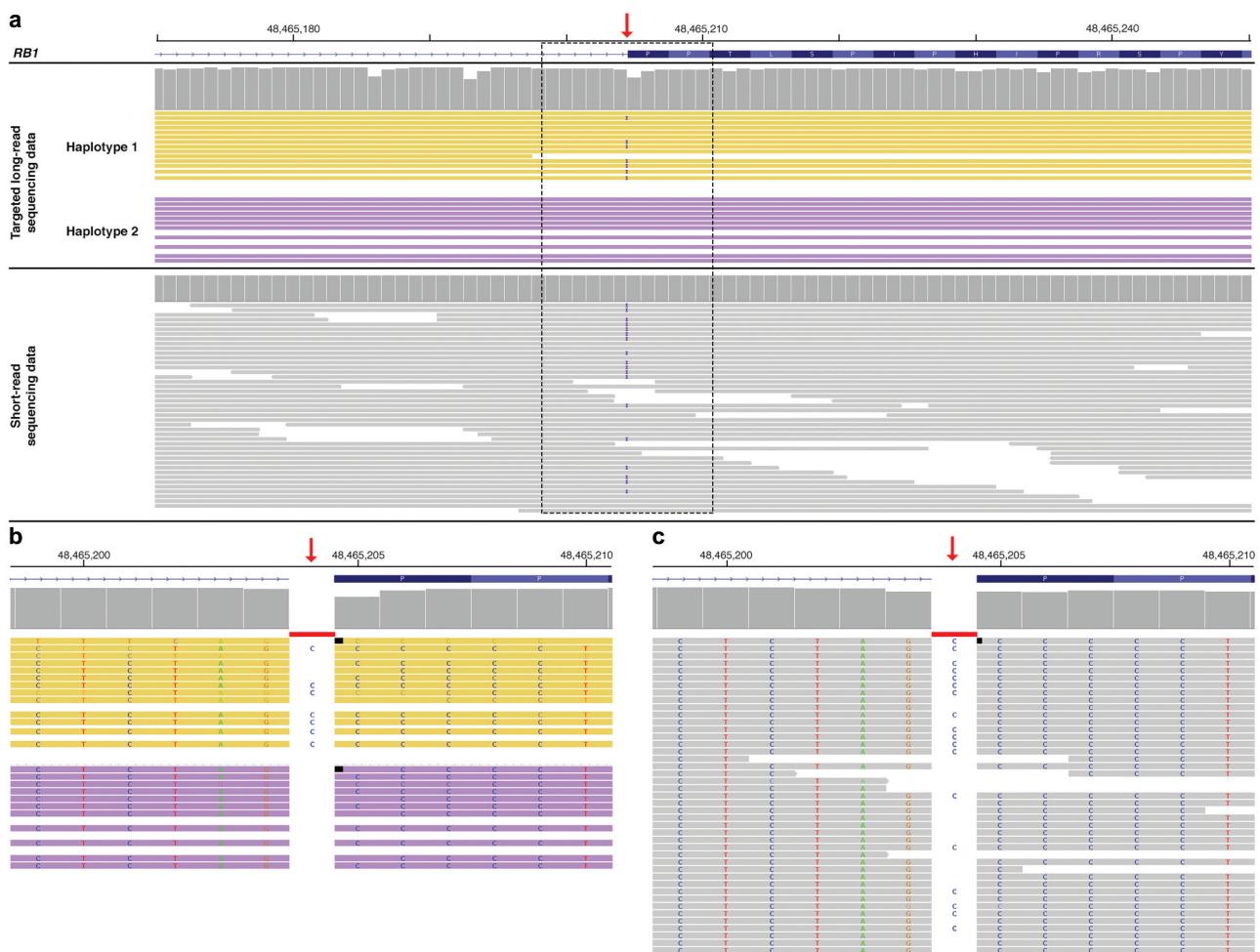
across the *RB1* gene targeted by long-read sequencing. Implementation of a haplotype-aware genotyping pipeline (24) revealed 205 variants of the *RB1* gene locus using targeted long-read sequencing. The short-read pipeline identified 186 variants in the same genomic region. Analysis of these variants with CADD (30,31) revealed that except for one coding variant, all of the rest of the variants were located in non-coding regions such as intronic or intergenic regions either upstream or downstream of the *RB1* gene transcript. The variant deemed to be most deleterious by CADD resulted in duplication of cytosine in exon 23 of *RB1* (c.2330dup). This resulted in a frameshift leading to premature termination codon at position 17 (p.Thr778TyrTer \*17). This corresponded to the same variant identified by WGS using short-read sequencing (Figure 2a). The targeted long-read sequencing was able to demonstrate the duplication was on one haplotype (Figure 2b), which could not be discerned from the short-read sequencing data (Figure 2c).

Overall, there was excellent agreement in the identified variants between targeted long-read and whole genome short-read sequencing modalities. The single largest source of discordance was due to four genomic regions (totaling 7624 bp)

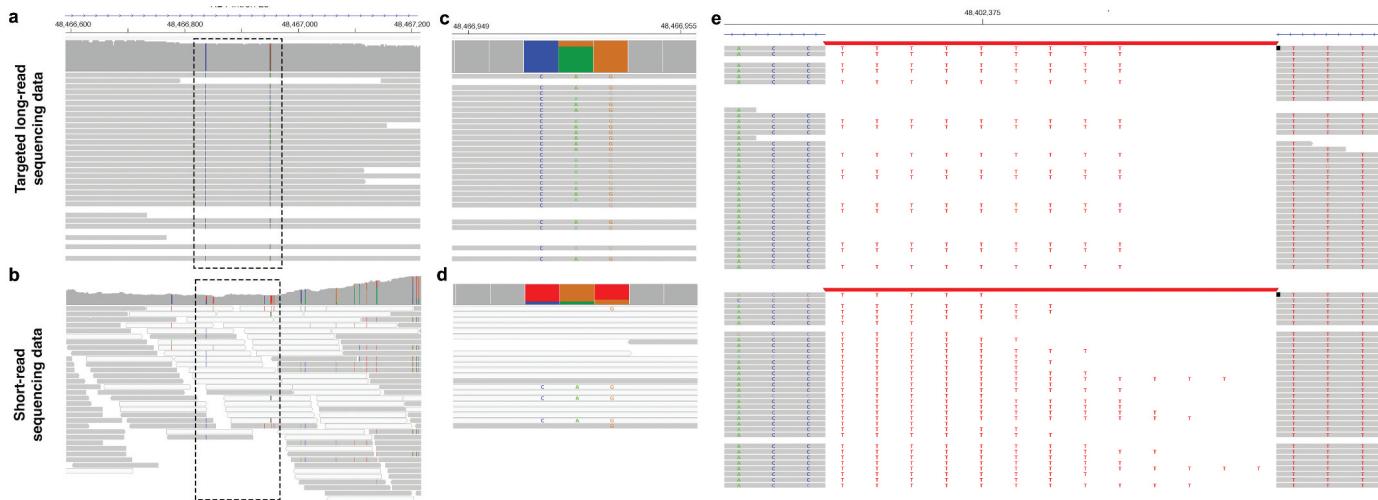
that did not have high-quality short-read data mapped to them, and whose boundaries had a higher density of low-quality variants than found in the targeted long-read data. This accounted for 30 of the 37 (81%) variant locations only identified in the targeted long-read data and not found in the short-read sequencing reads (Figure 3a-d). Examination of the variants unique to the short-read data revealed that 10 of the 16 (63%) variants not found by long-read sequencing, were homopolymeric sequences that did not pass quality control but were flagged by the pipeline (Figure 3e). Notably, when phased variants called in the intersection of regions sequenced through both technologies with depth at least 15 with MAPQ scores greater than 50 (172,437 bp) were compared, only 2 of 168 variant locations (1.2%) disagreed. The remaining 166 variants (98.8%) in this region spanning 96% of the target locus, agreed on both phase and call across both sequencing modalities.

## Discussion

In this study we demonstrate that targeted long-read sequencing of genomic DNA from a patient with RB can allow



**Figure 2.** Calling of a pathogenic variant in *RB1* by the different sequencing modalities. (a) examination of the region between intron 22 and exon 23 reveals a variant at the beginning of the exon (c.2330dup) (red arrow) in both the targeted long-read sequencing data and short-read sequencing data (blue insertion label). In the long-read data the duplication segregates with one haplotype confirming its heterozygous inheritance pattern. A zoomed in view of the boxed section reveals (b) duplication of a cytosine base (agccccc>agccccc) at the beginning of exon 23, which leads to a frameshift and early termination codon at position 17 (p.Thr778TyrTer \*17). The duplication is only found in one of the two haplotypes. (c) the same duplication is found in genome sequencing data from illumina short-reads.



**Figure 3.** Discordance in variant calling between short-read and targeted long-read sequencing modalities. The source of discordance in variants exclusively called by long-read sequencing were found in regions, such as intron 23 of *RB1*, where there is good depth of coverage in (a) targeted long-read sequencing but (b) poor coverage with short-read sequencing. There is an abundance of low MAPQ reads in the short-read data (denoted as white bars). A zoomed in view of that region denoted by the dashed box of the sequencing data demonstrates accurate calling of the variants uniformly in the (c) long-read data compared to the (d) short-read data. Comparatively, the errors in long-read data stems from homopolymer regions, as shown in intron 17 of *RB1*. (e) Compared to short-read sequencing that uniformly calls an insertion of a string of nine thymidine (T) bases, the long-read sequencing calls anywhere from 5 to 13 thymidine bases.

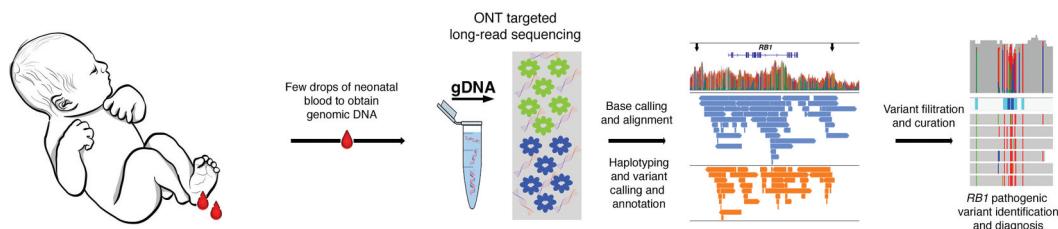
haplotagging for the accurate identification of the chromosomal architecture of identified disease variants. To then narrow down causal variants we demonstrated that application of deep learning frameworks can rank variants obtained from the haplotype-aware genotyping pipeline to provide tiers of potential pathogenic candidates without *a priori* knowledge of the disease-causing variants. In this work, the top ranked deleterious variant by CADD, was consistent with previous research exome sequencing. In contrast to short-read data, the long-read approaches provided haplotype and methylation data all from one sequencing run with results obtained in a matter of days after sample collection.

The predominant approach for variant discovery in RB and other ocular diseases relies on short-read sequencing technology due to the base-level accuracy of Illumina platform sequencing (34). However, the superior coverage of long-reads can allow detection of *de novo* inversions, repeats, and deletions, and phasing of such variants that are not detected in short-read Illumina sequencing data of the same patient genomes (35). We directly compared the Illumina and ONT platforms to demonstrate that both techniques have excellent agreement in detection of variants in the *RB1* gene in a patient with RB. Moreover, targeted long-read sequencing demonstrated in this work for the first time of the *RB1* locus, showed that the depth of genomic coverage can be obtained to identify rare variants that contribute to disease from isolated blood. The differences in variant calls found between ONT and Illumina platforms arise from systematic errors unique to each platform. In the case with Illumina, missed variants found using ONT were due to poorly mapped regions with low coverage or reads with low MAPQ scores, which often are in non-coding regions (Figure 3b–e). On the other hand, the error in calling certain variants on the ONT platform identified from Illumina were in areas of low complexity, such as homopolymer regions (Figure 3a). In these regions, the constant translocation speed of the DNA through the protein nanopore can make the

detection of the exact lengths of homopolymers difficult using ONT (36). Continuing advances in computational error correction methods will look to bridge the gap between the read level accuracy of short-read and long-read approaches.

In diseases like RB where greater than 98% of disease is attributable to pathogenic variants in the single gene (*RB1*), adaptive sampling using the ONT platform provides a targeted approach to selectively sequencing this genomic loci and interrogate potential disease-causing variants. A small minority of RB do show amplification of the *MYCN* gene (37) rather than inactivation of the *RB1* gene. Given the ease of assigning the target region for sequencing bioinformatically without any dependence on sample preparation, long-read sequencing panels can be carried out targeting both the *RB1* and *MYCN* genomic regions. Another important concept in RB progression is somatic copy number alterations. These have been noted with gains in genomic regions 1q, 2p, 6p and losses in genomic regions 13q and 16q from both enucleated eyes (38) and recently from aqueous humor (39). These regions can also be targeted and analyzed for any copy number changes in those regions. The adaptability of this technology holds tremendous promise to examine areas beyond *RB1* to provide diagnostic information of RB patients.

There have been great advances to the molecular diagnosis of RB. In patients with RB, the aqueous humor, an accessible compartment of the eye, has been demonstrated to serve as a surrogate for tumor tissue (40,41). This approach has tremendous utility in treatment response on intraocular RB (42). Emerging long-read consensus sequencing will provide another avenue to examine circulating tumor DNA to assess for tumor burden and treatment response (43). Furthermore, genomic DNA displays excellent concordance of *RB1* pathogenic variants identified from tumor DNA (14). In this work we demonstrate that germline status can be ascertained using targeted *RB1* gene analysis from gDNA isolated from blood. Testing for pathogenic variants in *RB1* is especially critical in



**Figure 4.** A rapid non-invasive approach to examine gDNA for pathogenic variants in *RB1* to expedite the molecular diagnosis of RB. A venipuncture blood draw can be done at birth or shortly after using only a few drops of blood obtained from a heel prick blood draw. Isolated gDNA can then be subjected to targeted long-read sequencing of the *RB1* locus or even expanded to other genomic regions of interest that will provide not only variant analysis but also chromosomal architecture data in a matter of days for rapid diagnosis of pathogenic disease-causing variants contributing to RB.

cases of unilateral disease where a de novo variant may be the cause or in known heritable cases. This approach is minimally invasive and provides adequate depth of coverage to identify rare variants that contribute to disease. Moreover, targeted long-read sequencing offers a rapid and economical approach to variant discovery in RB. After genomic DNA isolation, the time to variant identification was less than 72 hours using the targeted sequencing approach of *RB1* (Figure 4). The current lab setup has a small footprint, requiring a sequencer, which is the size of a USB drive, and a laptop computer for data analysis, which can be purchased for less than \$2000 with price per run equating to around \$500 or even less if samples are multiplexed. Thus, once implemented, this technology would be able to deliver diagnoses in the most endemic areas and would have broad applicability to the global population afflicted with RB in remote settings where clinical infrastructure may not be set up for sequencing patient samples. The ONT technology has been demonstrated to be instrumental in tracking the 2015 Ebola outbreak with sequencers used in the field (44) to deliver a genetic diagnosis in a rapid setting from critically ill patients (45).

Retinoblastoma has a profound effect on the lives of patients and families affected and given the lifelong risk of secondary malignancies, early diagnosis and intervention are an essential step to improve morbidity and mortality. The genetic information obtained from the approaches outlined in this work can potentially lead to fewer invasive procedures, earlier diagnosis and earlier treatment decision making, and improved outcomes in patients at risk of developing RB. Incorporating rapid, non-invasive genetic testing as part of standard RB-care has significant advantages that would revolutionize our clinical approach in the neonatal population. This information is invaluable for not only the patient, but future family planning as it provides inheritance pattern information all from one sequencing run. By genetically phasing an individual genome, long-read data can be partitioned into two parental genome datasets that can be independently assembled to evaluate whether variants are in cis or trans configuration. The ability to produce fast, reliable results of RB in a neonate will greatly impact care in this vulnerable patient population and will provide a more definitive diagnosis of heritable RB and guide management decisions for patients and their families. Application of existing machine learning tools to rank the identified variants (30,31) will allow prioritization of the top candidates for future cell-based stem cell work to develop patient-specific therapeutic interventions (46).

## Acknowledgements

The authors would like to thank Angela Sandt in the Van Gelder lab for technical assistance. We acknowledge the Northwest Genomics Center for providing assistance for Illumina library preparation and performing short-read genome sequencing on the Illumina NovaSeq6000 platform. We thank Dr. Danny Miller and his laboratory at the University of Washington for library preparation and performing long-read genome sequencing on the ONT PromethION platform.

## Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

## Funding

This work was supported by an unrestricted grant from Research to Prevent Blindness to the Department of Ophthalmology at University of Washington; The Sinksey Foundation (DM); and the Violet Sees Foundation (AWS, DM).

## ORCID

Debarshi Mustafi <http://orcid.org/0000-0002-9164-7488>

## References

- Stacey AW, Bowman R, Foster A, Kivelä TT, Munier FL, Cassoux N, Fabian ID, Harby LA, Portabella SA, Alia DB, et al. Incidence of incidence of retinoblastoma has increased: results from 40 European countries. *Ophthalmology*. 2021;128(9):1369–71. doi:10.1016/j.ophtha.2021.01.024.
- Dimaras H, Corson TW, Cobrinik D, White A, Zhao J, Munier FL, Abramson DH, Shields CL, Chantada GL, Njuguna F, et al. Retinoblastoma. *Nat Rev Dis Primers*. 2015;1(1):15021. doi:10.1038/nrdp.2015.21.
- Hong FD, Huang HJ, To H, Young LJ, Oro A, Bookstein R, Lee EY, Lee WH. Structure of the human retinoblastoma gene. *Proc Natl Acad Sci USA*. 1989;86(14):5502–06. doi:10.1073/pnas.86.14.5502.
- Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, Ulyanov A, Wu G, Wilson M, Wang J, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature*. 2012;481(7381):329–34. doi:10.1038/nature10733.
- Kooi IE, Mol BM, Massink MPG, Ameziane N, Meijers-Heijboer H, Dommering CJ, van Mil SE, de Vries Y, van der Hout AH, Kaspers GJL, et al. Somatic genomic alterations in retinoblastoma beyond *RB1* are rare and limited to copy number changes. *Sci Rep*. 2016;6(1):25264. doi:10.1038/srep25264.
- Kleinerman RA, Yu C-L, Little MP, Li Y, Abramson D, Seddon J, Tucker MA. Variation of second cancer risk by family history of

- retinoblastoma among long-term survivors. *J Clin Oncol.* **2012**;30(9):950–57. doi:[10.1200/JCO.2011.37.0239](https://doi.org/10.1200/JCO.2011.37.0239).
7. Fabian ID, Abdallah E, Abdulla SU, Abdulqader RA, Adamou Boubacar S, Ademola-Popoola DS, Adio A, Afshar AR, Aggarwal P, Aghaji AE, et al. Global retinoblastoma presentation and analysis by national income level. *JAMA Oncol.* **2020**;6(5):685–95. doi:[10.1001/jamaonc.2019.6716](https://doi.org/10.1001/jamaonc.2019.6716).
  8. Ancona-Lezama D, Dalvin LA, Shields CL. Modern treatment of retinoblastoma: a 2020 review. *Indian J Ophthalmol.* **2020**;68(11):2356–65. doi:[10.4103/ijo.IJO\\_721\\_20](https://doi.org/10.4103/ijo.IJO_721_20).
  9. Fernandes AG, Pollock BD, Rabito FA. Retinoblastoma in the United States: a 40-year incidence and survival analysis. *J Pediatr Ophthalmol Strabismus.* **2018**;55(3):182–88. doi:[10.3928/01913913-20171116-03](https://doi.org/10.3928/01913913-20171116-03).
  10. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, et al. The landscape of genomic alterations across childhood cancers. *Nature.* **2018**;555(7696):321–27. doi:[10.1038/nature25480](https://doi.org/10.1038/nature25480).
  11. Mallipatna A, Marino M, Singh AD. Genetics of retinoblastoma. *Asia Pac J Ophthalmol (Phila).* **2016**;5(4):260–64. doi:[10.1097/APO.0000000000000219](https://doi.org/10.1097/APO.0000000000000219).
  12. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature.* **1986**;323(6089):643–46. doi:[10.1038/323643a0](https://doi.org/10.1038/323643a0).
  13. Kline CN, Joseph NM, Grenert JP, van Ziffle J, Talevich E, Onodera C, Aboim M, Cha S, Raleigh DR, Braunstein S, et al. Targeted next-generation sequencing of pediatric neuro-oncology patients improves diagnosis, identifies pathogenic germline mutations, and directs targeted therapy. *Neuro Oncol.* **2017**;19(5):699–709. doi:[10.1093/neuonc/now254](https://doi.org/10.1093/neuonc/now254).
  14. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, et al. Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn.* **2015**;17(3):251–64. doi:[10.1016/j.jmoldx.2014.12.006](https://doi.org/10.1016/j.jmoldx.2014.12.006).
  15. Davies HR, Broad KD, Onadim Z, Price EA, Zou X, Sheriff I, Karra EK, Scheimberg I, Reddy MA, Sagoo MS, et al. Whole-genome sequencing of retinoblastoma reveals the diversity of rearrangements disrupting rb1 and uncovers a treatment-related mutational signature. *Cancers.* **2021**;13(4):754. doi:[10.3390/cancers13040754](https://doi.org/10.3390/cancers13040754).
  16. Francis JH, Richards AL, Mandelker DL, Berger MF, Walsh MF, Dunkel IJ, Donoghue MTA, Abramson DH. Molecular changes in retinoblastoma beyond RB1: findings from next-generation sequencing. *Cancers.* **2021**;13(1):149. doi:[10.3390/cancers13010149](https://doi.org/10.3390/cancers13010149).
  17. Afshar AR, Pekmezci M, Bloomer MM, Cadenas NJ, Stevers M, Banerjee A, Roy R, Olshan AB, Van Ziffle J, Onodera C, et al. Next-generation sequencing of retinoblastoma identifies pathogenic alterations beyond RB1 inactivation that correlate with aggressive histopathologic features. *Ophthalmology.* **2020**;127(6):804–13. doi:[10.1016/j.ophtha.2019.12.005](https://doi.org/10.1016/j.ophtha.2019.12.005).
  18. Vilfan ID, Tsai Y-C, Clark TA, Wegener J, Dai Q, Yi C, Pan T, Turner SW, Korlach J. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnol.* **2013**;11(1):8. doi:[10.1186/1477-3155-11-8](https://doi.org/10.1186/1477-3155-11-8).
  19. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods.* **2017**;14(4):411–13. doi:[10.1038/nmeth.4189](https://doi.org/10.1038/nmeth.4189).
  20. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods.* **2017**;14(4):407–10. doi:[10.1038/nmeth.4184](https://doi.org/10.1038/nmeth.4184).
  21. Gelli E, Pinto AM, Somma S, Imperatore V, Cannone MG, Hadjistilianou T, De Francesco S, Galimberti D, Currò A, Bruttini M, et al. Evidence of predisposing epimutation in retinoblastoma. *Hum Mutat.* **2019**;40(2):201–06. doi:[10.1002/humu.23684](https://doi.org/10.1002/humu.23684).
  22. Quiñonez-Silva G, Dávalos-Salas M, Recillas-Targa F, Ostrosky-Wegman P, Aranda DA, Benítez-Bribiesca L. Monoallelic germline methylation and sequence variant in the promoter of the RB1 gene: a possible constitutive epimutation in hereditary retinoblastoma. *Clin Epigenetics.* **2016**;8(1):1. doi:[10.1186/s13148-015-0167-0](https://doi.org/10.1186/s13148-015-0167-0).
  23. Li H-T, Xu L, Weisenberger DJ, Li M, Zhou W, Peng C-C, Stachelek K, Cobrinik D, Liang G, Berry JL. Characterizing DNA methylation signatures of retinoblastoma using aqueous humor liquid biopsy. *Nat Commun.* **2022**;13(1):5523. doi:[10.1038/s41467-022-33248-2](https://doi.org/10.1038/s41467-022-33248-2).
  24. Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, Baid G, Kolmogorov M, Eizenga JM, Miga KH, et al. Haplotype-aware variant calling with PEPPER-margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods.* **2021**;18(11):1322–32. doi:[10.1038/s41592-021-01299-w](https://doi.org/10.1038/s41592-021-01299-w).
  25. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. *Nat Methods.* **2016**;13(9):751–54. doi:[10.1038/nmeth.3930](https://doi.org/10.1038/nmeth.3930).
  26. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol.* **2021**;39(4):442–50. doi:[10.1038/s41587-020-00746-x](https://doi.org/10.1038/s41587-020-00746-x).
  27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **2009**;25(14):1754–60. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
  28. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* 201178. [accessed 2018 July 24]. doi:[10.1101/201178](https://doi.org/10.1101/201178).
  29. Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **2018**;34(18):3094–100. doi:[10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191).
  30. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* **2014**;46(3):310–15. doi:[10.1038/ng.2892](https://doi.org/10.1038/ng.2892).
  31. Rentszsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**;47(D1):D886–94. doi:[10.1093/nar/gky1016](https://doi.org/10.1093/nar/gky1016).
  32. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* **2015**;17(5):405–23. doi:[10.1038/gim.2015.30](https://doi.org/10.1038/gim.2015.30).
  33. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Eichler EE, et al. A global reference for human genetic variation. *Nature.* **2015**;526(7571):68–74. doi:[10.1038/nature15393](https://doi.org/10.1038/nature15393).
  34. Li W, Freudenberg J. Mappability and read length. *Front Genet.* **2014**;5:381. doi:[10.3389/fgene.2014.00381](https://doi.org/10.3389/fgene.2014.00381).
  35. Stancu MC, Van Roosmalen MJ, Renkens I, Nieboer MM, Middelkamp S, De Ligt J, Pregno G, Giachino D, Mandrile G, Valle-Inclan JE, et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun.* **2017**;8(1):1–13. doi:[10.1038/s41467-017-01343-4](https://doi.org/10.1038/s41467-017-01343-4).
  36. Delahaye C, Nicolas J, Andrés-León E. Sequencing DNA with nanopores: troubles and biases. *Plos One.* **2021**;16(10):e0257521. doi:[10.1371/journal.pone.0257521](https://doi.org/10.1371/journal.pone.0257521).
  37. Lee WH, Murphree AL, Benedict WF. Expression and amplification of the N-myc gene in primary retinoblastoma. *Nature.* **1984**;309(5967):458–60. doi:[10.1038/309458a0](https://doi.org/10.1038/309458a0).

38. Rushlow DE, Mol BM, Kennett JY, Yee S, Pajovic S, Thériault BL, Prigoda-Lee NL, Spencer C, Dimaras H, Corson TW, et al. Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies. *Lancet Oncol.* **2013**;14 (4):327–34. doi:[10.1016/S1470-2045\(13\)70045-7](https://doi.org/10.1016/S1470-2045(13)70045-7).
39. Kim ME, Polski A, Xu L, Prabakar RK, Peng C-C, Reid MW, Shah R, Kuhn P, Cobrinik D, Hicks J, et al. Comprehensive somatic copy number analysis using aqueous humor liquid biopsy for retinoblastoma. *Cancers.* **2021**;13(13):3340. doi:[10.3390/cancers1313340](https://doi.org/10.3390/cancers1313340).
40. Berry JL, Xu L, Murphree AL, Krishnan S, Stachelek K, Zolfaghari E, McGovern K, Lee TC, Carlsson A, Kuhn P, et al. Potential of aqueous humor as a surrogate tumor biopsy for retinoblastoma. *JAMA Ophthalmol.* **2017**;135(11):1221–30. doi:[10.1001/jamaophthalmol.2017.4097](https://doi.org/10.1001/jamaophthalmol.2017.4097).
41. Xu L, Shen L, Polski A, Prabakar RK, Shah R, Jubran R, Kim JW, Biegel J, Kuhn P, Cobrinik D, et al. Simultaneous identification of clinically relevant RB1 mutations and copy number alterations in aqueous humor of retinoblastoma eyes. *Ophthalmic Genet.* **2020**;41(6):526–32. doi:[10.1080/13816810.2020.1799417](https://doi.org/10.1080/13816810.2020.1799417).
42. Polski A, Xu L, Prabakar RK, Kim JW, Shah R, Jubran R, Kuhn P, Cobrinik D, Hicks J, Berry JL. Cell-free DNA tumor fraction in the aqueous humor is associated with therapeutic response in retinoblastoma Patients. *Transl Vis Sci Technol.* **2020**;9(10):30. doi:[10.1167/tvst.9.10.30](https://doi.org/10.1167/tvst.9.10.30).
43. Marcozzi A, Jager M, Elferink M, Straver R, van Ginkel JH, Peltenburg B, Chen L-T, Renkens I, van Kuik J, Terhaard C, et al. Accurate detection of circulating tumor DNA using nanopore consensus sequencing. *NPJ Genom Med.* **2021**;6(1):106. doi:[10.1038/s41525-021-00272-y](https://doi.org/10.1038/s41525-021-00272-y).
44. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, et al. Real-time, portable genome sequencing for ebola surveillance. *Nature.* **2016**;530(7589):228–32. doi:[10.1038/nature16996](https://doi.org/10.1038/nature16996).
45. Gorzynski JE, Goenka SD, Shafin K, Jensen TD, Fisk DG, Grove ME, Spiteri E, Pesout T, Monlong J, Baid G, et al. Ultrarapid nanopore genome sequencing in a critical care setting. *N Engl J Med.* **2022**;386(7):700–02. [accessed 2022 Jan 12]. doi:[10.1056/NEJMc2112090](https://doi.org/10.1056/NEJMc2112090).
46. Norrie JL, Nityanandam A, Lai K, Chen X, Wilson M, Stewart E, Griffiths L, Jin H, Wu G, Orr B, et al. Retinoblastoma from human stem cell-derived retinal organoids. *Nat Commun.* **2021**;12 (1):4535. doi:[10.1038/s41467-021-24781-7](https://doi.org/10.1038/s41467-021-24781-7).