Cancer Letters 598 (2024) 217121

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Original Articles

Comprehensive identification of pathogenic variants in retinoblastoma by long- and short-read sequencing

Jingjing Zheng [1], Tong Li [1], Huijing Ye, Zehang Jiang, Wenbing Jiang, Huasheng Yang [***], Zhikun Wu [**], Zhi Xie [*]

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

ARTICLE INFO

Keywords:
Retinoblastoma
Structural variant
Long-read sequencing

ABSTRACT

Retinoblastoma (RB) is the most common intraocular malignancy in childhood. The causal variants in RB are mostly characterized by previously used short-read sequencing (SRS) analysis, which has technical limitations in identifying structural variants (SVs) and phasing information. Long-read sequencing (LRS) technology has advantages over SRS in detecting SVs, phased genetic variants, and methylation. In this study, we comprehensively characterized the genetic landscape of RB using combinatorial LRS and SRS of 16 RB tumors and 16 matched blood samples. We detected a total of 232 somatic SVs, with an average of 14.5 SVs per sample across the whole genome in our cohort. We identified 20 distinct pathogenic variants disrupting *RB1* gene, including three novel small variants and five somatic SVs. We found more somatic SVs were detected from LRS than SRS (140 vs. 122) in RB samples with WGS data, particularly the insertions (18 vs. 1). Furthermore, our analysis shows that, with the exception of one sample who lacked the methylation data, all samples presented biallelic inactivation of *RB1* in various forms, including two cases with the biallelic hypermethylated promoter and four cases with compound heterozygous mutations which were missing in SRS analysis. By inferring relative timing of somatic events, we reveal the genetic progression that *RB1* disruption early and followed by copy number changes, including amplifications of Chr2p and deletions of Chr16q, during RB tumorigenesis. Altogether, we characterize the comprehensive genetic landscape of RB, providing novel insights into the genetic alterations and mechanisms contributing to RB initiation and development. Our work also establishes a framework to analyze genomic landscape of cancers based on LRS data.

Retinoblastoma (RB) is the most prevalent eye cancer in children [1]. Without timely diagnosis and treatment, metastasis will happen and RB will be fatal [2]. RB is generally thought to be caused by the inactivation of RB transcriptional corepressor 1 (*RB1*) gene [1]. A comprehensive understanding of the genetic basis and the underlying mutational mechanisms is essential for precise diagnosis and effective clinical treatment.

Previous studies have utilized various approaches, including quantitative multiplex PCR (QM-PCR), Sanger sequencing, whole genome/exome short-read sequencing (WGS/WES) , as well as cloning and sequencing *RB1* cDNAs, to identify germline or somatic single nucleotide variants (SNVs) and small insertions/deletions (InDels), with the purpose to detect disease causal variants in RB [3–6]. However, previous technologies have limitations in identifying structural variants (SVs), methylation, and phasing information, leading to some samples exhibiting only one or even no detectable causal mutation in RB [5,7–11]. It is worth noting that SVs affect larger regions of cancer genomes than any other type of genetic variants [12], therefore, causative variants may be underestimated in RB [3].

Unlike traditional methods, long-read sequencing (LRS) technologies, such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), offer the distinct advantage of directly sequencing native genome DNA. LRS produces reads up to two megabases (Mb) in length [13], allowing efficient detection of SVs and providing an insight into the haplotype context of a given variant. Moreover, ONT platform enables directly detect DNA methylation during sequencing, which

* Corresponding author. No. 54 Xianlie South Road, Yuexiu District, Guangzhou, 510060, China.
** Corresponding author. No. 54 Xianlie South Road, Yuexiu District, Guangzhou, 510060, China.
*** Corresponding author. No. 54 Xianlie South Road, Yuexiu District, Guangzhou, 510060, China.
E-mail addresses: yanghs64@126.com (H. Yang), wuzhikun86@163.com (Z. Wu), xiezhi@gmail.com (Z. Xie).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.canlet.2024.217121
Received 18 October 2023; Received in revised form 16 June 2024; Accepted 11 July 2024
Available online 14 July 2024

overcomes the shortcoming from traditional measurement, such as damaging DNA during bisulfite conversion and the restricted detection of short-range methylation patterns using NGS platform [14]. The increased power of LRS for detecting germline and somatic SVs, as well as methylation patterns, has been widely demonstrated [15–20]. Despite the potential use of LRS to study RB, only two studies have been conducted to date, where each study sequenced only one sample. One study utilized PacBio sequencing to uncover the complex mechanism of *RB1* inactivation through the *RB1-SIAH3* fusion [21]. The other study used targeted ONT sequencing to identify missing disease-causing variants in *RB1* [22].

In our study, we utilized both LRS and SRS to sequence 16 pairs of tumor and matched blood samples to identify various genomic alterations, including SNVs, InDels, SVs and methylations in RB. We detected more SVs by LRS data, particularly the type of INSs, when compared with SRS. In addition, LRS data allowed us to simultaneously detect SVs and methylations to have a comprehensive examination of genomic variants and discover novel pathogenic variants in RB. Furthermore, using LRS, we efficiently phased the genomic alterations to determine the diverse forms of biallelic *RB1* loss. Our work will not only help us comprehensively understand multiple genetic variations and epigenetic patterns of the causative gene (*RB1*) of RB, but also provide a framework to analyze genomic landscapes of cancers based on LRS data.

1. Materials and methods

1.1. Patient recruitment

The tumor tissues and matched blood were collected from the eyes of 16 patients with RB in Sun Yat-sen University Eye Hospital in Guangzhou city of China. The study was conducted according to the guidelines of the Declaration of Helsinki. All the enrolled patients had signed the consent forms and the proposed studies were approved by Ethics Committee (2016KYPJ028). The median age at diagnosis was 21.3 months (ranging from 2.5 to 84 months). The detail information is provided in Table S1.

1.2. Genome sequencing

The genomic DNA was extracted from RB tissues and corresponding blood samples using Qiagen DNAeasy kits. Nanopore whole-genome sequencing was performed for all samples. Eleven pairs of tumor and blood samples were performed by paired-end whole-genome sequencing on Illumina NovaSeq platform, and the other five pairs those overlapped with the samples in published study were performed by paired-end whole-exome sequencing [23]. In order to obtain high quality data, fastp (v0.20.1) [24] was performed on NGS data with default parameters, and the quality control process for ONT data was according to the previous study [25]. On average, we obtained 76 Gb per tumor sample with a median length of 6.9 kb and 50 Gb per blood sample with a median length of 9.5 kb of ONT long-reads after quality control. Meanwhile, an average of 134 Gb per tumor coupled with 87 Gb per blood of NGS short-reads (paired-end 150 bp) were generated for each sample (Table S2). To orthogonally validate the SVs detected by ONT, we additionally used PacBio HiFi to sequence the tumor and blood samples of one patient with RB.

1.3. Small variant calling

For NGS data, the clean reads were aligned to GRCh38 reference genome using BWA (v0.7.17-r1188) [26]. Duplicates were removed using Picard MarkDuplicates (v2.23.9) (https://broadinstitute.github.io/picard/) with default settings. Germline SNVs and InDels were identified using HaplotypeCaller (GATK v4.1.9) [27] and FreeBayes (v1.2.0) with default parameters. Only germline variants called by both tools were retained. The types of loss-of-function (pLoF) variants include nonsense, frameshift and splice site variants. Somatic SNVs and InDels were detected using MuTect2 (GATK v4.1.9) [28], Strelka2 (v 2.9.10) [29] and Neusomatic (v0.2.1) [30] on tumor-normal pairs, separately. Somatic mutations called by more than two of the three methods were kept. Based on MuTect2, Strelka2, Neusomatic variant caller, we used multiple filtering standards to filter SNV/InDels. These criteria included read depth, alternate allele observation count, mapping quality, base quality, strand bias, etc. Detailed filtering standards and thresholds are listed in the table (Table S3). Finally, Annovar [31] was used to annotate these mutations.

1.4. Structural variant calling

For ONT data, the clean reads were aligned to GRCh38 reference genome using minimap2 (v2.17-r941) [32] and ngmlr (v0.2.7) [33]. Sniffles (v1.0.12) [33] and cuteSV (v1.0.11) [34] were performed using force call mode to get somatic SV datasets. Since Nanomonsv (v0.4.0) [35] preferred corrected read, the ONT data was firstly corrected using Fmlrc2 (v1.0.0) [36] before being processed by minimap2 and Nanomonsv. Jasmine (v1.1.5) [37,38] was then used to merge the five call sets and an SV called by at least three of the above five methods was retained. Then, the in-house filtering was performed according to 1) SV length $\geq$100 bp; 2) sequencing depth $\geq$6. And further manual checking was performed using IGV (v2.12.2) [39] to ensure accuracy. Additionally, to validate these somatic SVs, we applied a pairwise alignment method to 10 translocations (TRAs), 10 large SVs (>30 M) and 10 SVs detected only from LRS data. We extracted the flanking sequences on both sides of the SV breakpoint and reconstructed the putative reference sequence. The flanking sequence were 10 kb each, resulting in a putative reference sequence of 20 kb. We used BLASTN for pairwise alignment of the supporting LRS reads with the putative sequence. Snapshots of the figures have been uploaded to Zenodo (https://zenodo.org/records/11504094). Finally, AnnotSV [40] was used to annotate these SVs. PacBio HiFi reads were aligned to GRCh38 using pbmm2 (v1.3.0) (https://github.com/PacificBiosciences/pbmm2) with default parameters, followed by pbsv (https://github.com/PacificBiosciences/pbsv) joint calling for tumor-normal pairs. False discovery rate (FDR) was defined as the number of somatic SVs of ONT data that not overlapped with HiFi data divided by the total somatic SVs number that we identified from ONT data. The FDR value greater than 0.1 is considered high. To further validate our pipeline, we applied it to the pair ONT sequencing data of publicly available COLO928 tumor-normal cell line [41], whose somatic SVs had been curated and validated. We focused on the standard somatic SVs whose read length is longer than 100 bps. Precision is defined as the number of overlapped somatic SVs divided by the total number we identified. Sensitivity is defined as the number of overlapped somatic SVs divided by the total number of standard somatic SVs. F1-score is computed according to 2 $\times$ precision*sensitivity/(precision + sensitivity).

For NGS data, Manta (v1.6.0) [42], Delly (v1.1.6) [43], Smoove (v0.2.8) (https://github.com/brentp/smoove) and SvABA (v1.1.0) [44] were used to call somatic SVs. The parameter of SvABA is -L 6, and other SV callers' parameters are default. Then Jasmine (v1.1.5) was used to merge theses four call sets and an SV called by at least two methods was retained. Manual checking was performed using IGV as mentioned in the process of ONT data analysis.

1.5. DNA methylation analysis

For the Nanopore reads (fast5 files), Guppy (v4.0.15) was used for basecalling with the parameters "-x 'auto' –config dna_r9.4.1_450bps_fast.cfg". The basecalled reads were then aligned to the GRCh38 reference genome using minimap2 (v2.17-r941). To determine the methylation status of the *RB1* promoter, Nanopolish (v0.13.2) [45] was employed with default parameters to identify 5mCs in the CpG context of Chr13, then PEPPER-Margin-DeepVariant pipeline (v0.6) [46] was

applied to the Binary Alignment Map (BAM) file with methylation tags to obtain the phasing information. The BAM files were converted by nanopore-methylation-utilities [47] in IGV to the bisulfite mode using the Nanopolish results.

1.6. Mutation driver analysis

To identify cancer driver genes, dNdScv (v0.0.1.0) [48] and WITER (v1.2) [49] were used to analyze somatic SNVs and InDels within exons. The results of dNdScv showed that *RB1* was the only gene with q-value less than 0.1. The WITER analysis was performed on the Java platform using recommended parameters, which showed that six genes passed the Benjamini Hochberg FDR q-value cutoff of 0.1. To obtain a confident result, we only retained the genes identified by both methods with an FDR q-value less than 0.1.

1.7. Haplotype phasing analysis

The phasing information for the *RB1* region on tumor or normal genomes was obtained using the PEPPER-Margin-DeepVariant pipeline (v0.6) [46]. Germline variants and long-read BAM files of each sample were used to create phasing information. Finally, IGV was used for visualization and manual checking.

1.8. Copy number analysis and ploidy estimation

In order to evaluate the reliability of copy number alterations (CNAs) and ploidy of ONT data, we applied Mosdepth (v0.2.9) [50] with parameter "-n -b region. bed" to calculate the genome coverage from ONT and NGS data; the correlation coefficient ($R \approx 0.87$, Pearson test, $p < 2.2e\text{-}16$) between the two platforms indicated it was accurate enough to obtain CNAs from ONT data. Somatic CNA analysis was performed by NanoGLADIATOR [51] using ONT data. BAF was obtained through HATCHet (v0.4.9) [52] using NGS data as input, of which the process and parameter were referred to the author's recommended. For ONT data, SNPs were detected using NanoCaller (v3.0.1) [53], and then BAF values were calculated using custom script. Per-chromosome ploidy, namely whole-chromosome copy number, was estimated using read depth in 1 Mb sequential bins, excluding the *N*-masked heterochromatic regions.

1.9. Chromothripsis analysis

The presence of chromothripsis was determined based on the breakpoints of somatic SVs and CNAs in the chromosome regions [54]. If a chromosome arm showed more than 6 breakpoints, high breakpoint densities (>2σ above the average breakpoint burden of each sample) and CNAs, it was considered a chromothripsis event. In order to obtain an accurate result, we also applied ShatterSeek (v1.1) [55] to detect chromothripsis events using SVs and CNVs as input files, then selected high confidence candidates with p value less than 0.05. The visualization of chromothripsis was created using ShatterSeek and RCircos (v1.2.2) [56].

1.10. Relative timing of somatic events analysis

Relative timing of somatic events was analyzed using a pipeline from previous research [57]. The subclonal architectures were reconstructed using PyClone-VI (v0.1.1) [58] with the parameters "fit -c 40 -d beta-binomial -r 10″ based on the copy number profile, purity, and ploidy data. Purity was estimated from HATCHet [52] or the VAF of *RB1*. SNVs, InDels and SVs were assigned to mutation clusters using mutationtimeR (v1.00.0) [57] with default parameters to determine the timing of the driver event during tumor development. We referred to the thresholds of the previous research and only retained events that occurred at least three times to ensure the accuracy of the analysis. On this basis, we drew the inferred trajectories of 16 samples, and used Chi-square test to detect the statistical differences of different events. For example, *RB1* has a chi-square value of 10.23 and a significant *P*-value of 0.006 compared to −11q (Table S4). For the drawing of preferential ordering diagrams, we used PhylogicNDT LeagueModel (v1.0) [59] (parameters: -n_seasons 20 –n_perms 50) to integrate 16 samples.

1.11. Experimental validation of RB1 mutations

For further validation of the somatic SV, PCR assays were carried out for a deletion (DEL) SV (DEL:13_48,380,287–13_48,381,346) on both the tumor and blood samples of RB12, as well as on RB10 which did not show any SV in the region. The breakpoint of the predicted SV was covered by the PCR primers designed using Primer-BLAST [60]. The primer sequence was listed in Table S5. The target region was then amplified using 2*x* Phanta® Max Master Mix (vazyme, #P515) according to the manufacturer's instructions.

2. Results

2.1. Study design and analytic approach

To comprehensively understand the genetic basis of RB, we characterized the full spectrum of genetic alterations, including SNVs, InDels and SVs, across the whole or exome genome of RB. Firstly, using the ONT platform, we obtained whole genome LRS data of 16 primary RB and 16 matched blood samples. Meanwhile, using the Illumina platform, we obtained whole genome SRS data of 11 primary RB and 11 matched blood samples. The whole exome SRS data of the other five pairs was obtained from the previous study [23]. Then, the SRS data was used to identify SNVs, InDels and SVs, while the LRS data was used to detect SVs and methylation (Fig. 1a).

To obtain somatic small variants with high-confidence, we retained the call set overlapped by at least two of three callers from the SRS data (**Methods**). To obtain somatic SVs, we developed a pipeline to identify high-confident somatic SVs. The FDR of SV calling for the RB08 ONT datasets using our pipeline was 10 % (Fig. 1b–c, Fig. S1-3, Methods). The FDRs of the two tools specifically designed for somatic SV detection in LRS data were 88.5 % and 99.5 % [35,61] (Table S6-7), which were much higher than the value of our pipeline. To further validate our pipeline, we applied it to publicly available COLO928 tumor-normal cell line pair ONT data [62], whose somatic SVs had been curated and validated. The SV number of the standard dataset was 58. We detected a total of 53 SVs after performing our pipeline to the same ONT data, of which 49 SVs are overlapped with the standard dataset (Fig. 1d). Thus, our pipeline achieved a precision of 92.5 % (49/53), a sensitivity of 84.5 % (49/58) and an F1-score of 88.3 % (2 × precision*sensitivity/(precision + sensitivity)), demonstrating that it is both sensitive and reliable in detecting somatic SVs in tumor tissue. What's more, to validate called somatic SVs in RB, we applied BLASTN, a pairwise alignment method, to 10 translocations (TRAs), 10 large SVs (>30 M) and 10 SVs that only detected from LRS data. Dotplots showed that the LRS reads supported these SVs well with the reconstructed sequence (Methods, Table S8).

To investigate whether LRS is better at detecting SVs than SRS, we did a comparison of the SVs and CNAs identified from 11 pairs WGS data by ONT and NGS platform, respectively (Fig. S4-7, Table S9). The results show that ONT has the advantages in detecting the total SV number (140 vs. 122), as well as the certain SV type, such as INSs (18 vs. 1).

2.2. Landscape of genomic alterations in RB

Based on the LRS data of 16 RB samples, we detected 232 somatic SVs, ranging from 3 to 52 per sample with an average number of 14.5 (Table 1, Fig. S8a, Table S9). These SVs included 51 DELs, 71

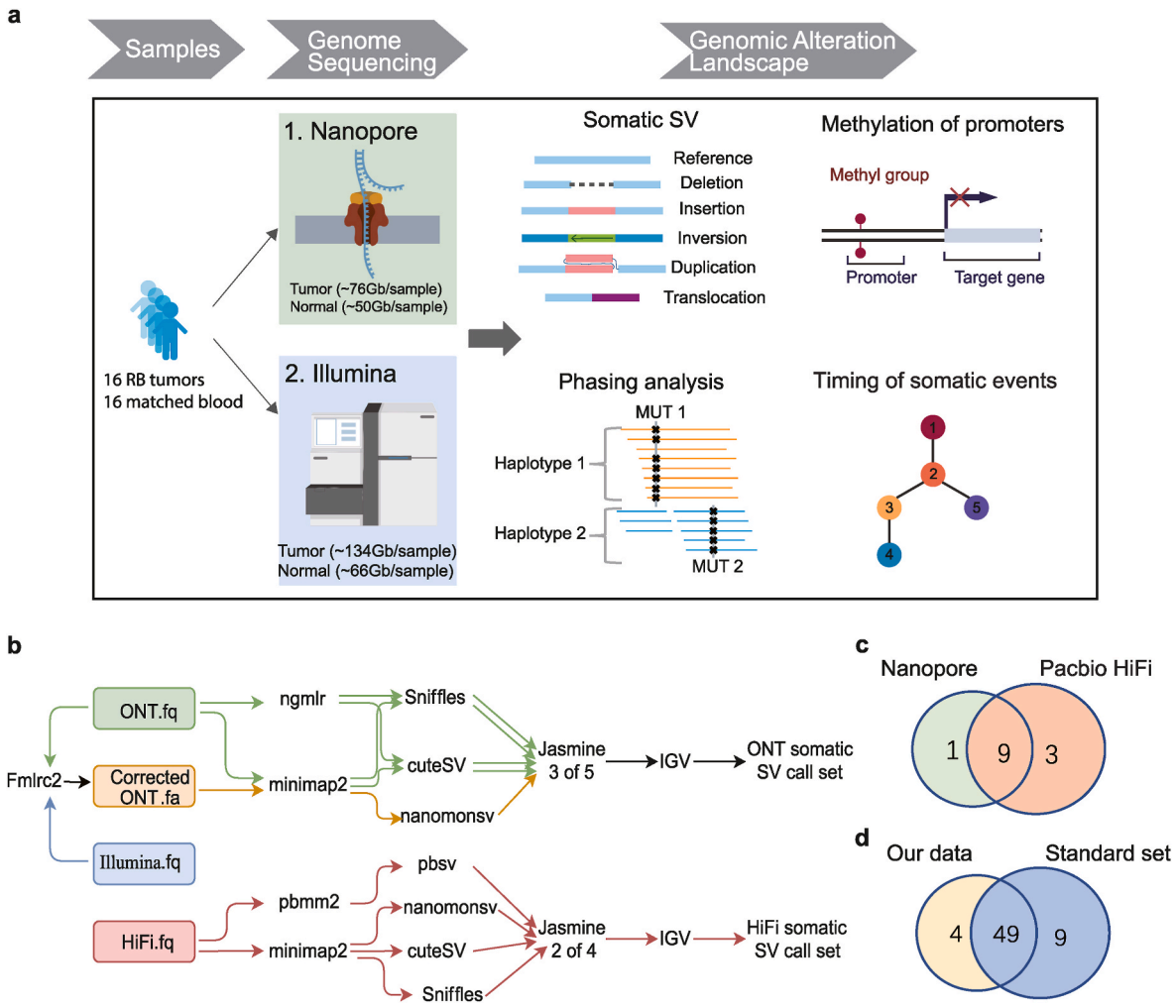

Fig. 1. Overview of this study in RB. **a,** Schematic diagram of study design. Sequencing data from Nanopore and Illumina platforms were used to detect somatic SV and methylation, as well as phase the variants of *RB1* gene. Then the detected variants were used to infer relative timing of somatic events. **b,** Ensemble workflow of somatic SV detection using ONT and PacBio HiFi data. IGV means the manually checking and selecting the correct variants using Integrative Genomics Viewer. **c,** The comparison of somatic SV number of orthogonally validated sample RB08 by ONT and PacBio HiFi data. **d,** The comparison of somatic SV number of validation sample COLO928 identified from our pipeline and those reported in published literature.

Table 1
SV number in RB samples.

Sample	SVnum	pLoF SVnum
RB01	19	9
RB02	3	1
RB03	13	5
RB04	5	1
RB05	52	17
RB06	5	0
RB07	5	0
RB08	10	4
RB09	3	1
RB10	20	8
RB11	23	8
RB12	11	4
RB13	29	13
RB14	4	0
RB15	24	15
RB16	6	5
Average	14.5	5.7

Number of somatic SVs.
pLoF SVnum is the number of somatic SVs directly disrupting coding sequences of genes.

duplications (DUPs), 31 insertions (INSs), 47 translocations (TRAs) and 32 inversions (INVs) (Fig. S8b). Chromosome13 (Chr13) was the predominant one where the SV breakends located, followed by Chr2 and Chr1 (Fig. S8c). Notably, SVs tended to cluster in certain hotspot regions, such as Chr2p and Chr13q, where two known RB-causing genes, *RB1* and *MYCN*, are located (Fig. S8d), suggesting the high levels of genomic instability around susceptibility genes in RB. A total of 91 somatic SVs, ranging from 0 to 17 per sample, were found to directly disrupt gene sequences, which were defined as pLoF somatic SVs (Table 1).

Based on the SRS datasets, we detected 1788 non-redundant somatic SNVs/InDels for the 16 samples, with an average of 112 variants per tumor (Table S10-11). Among these, 95 variants (averaging 6 per sample) located in exonic and splicing regions, which could be predicted as pLoF (Fig. S9). Missense mutation was the predominant type, accounting for 67 % of the total small variants.

With a comprehensive characterization of SNVs, InDels and SVs in our RB cohort, the somatically acquired and whole genomic mutational landscape of RB showed that *RB1* was the most commonly mutated gene, followed by *CNTNAP2* and *KLF12*, which had at least one pLoF variant in more than two patients (Fig. 2). Furthermore, we found alterations that affect multiple chromosomes, such as chromothripsis. The results revealed the presence of chromothripsis in two samples (RB05 and

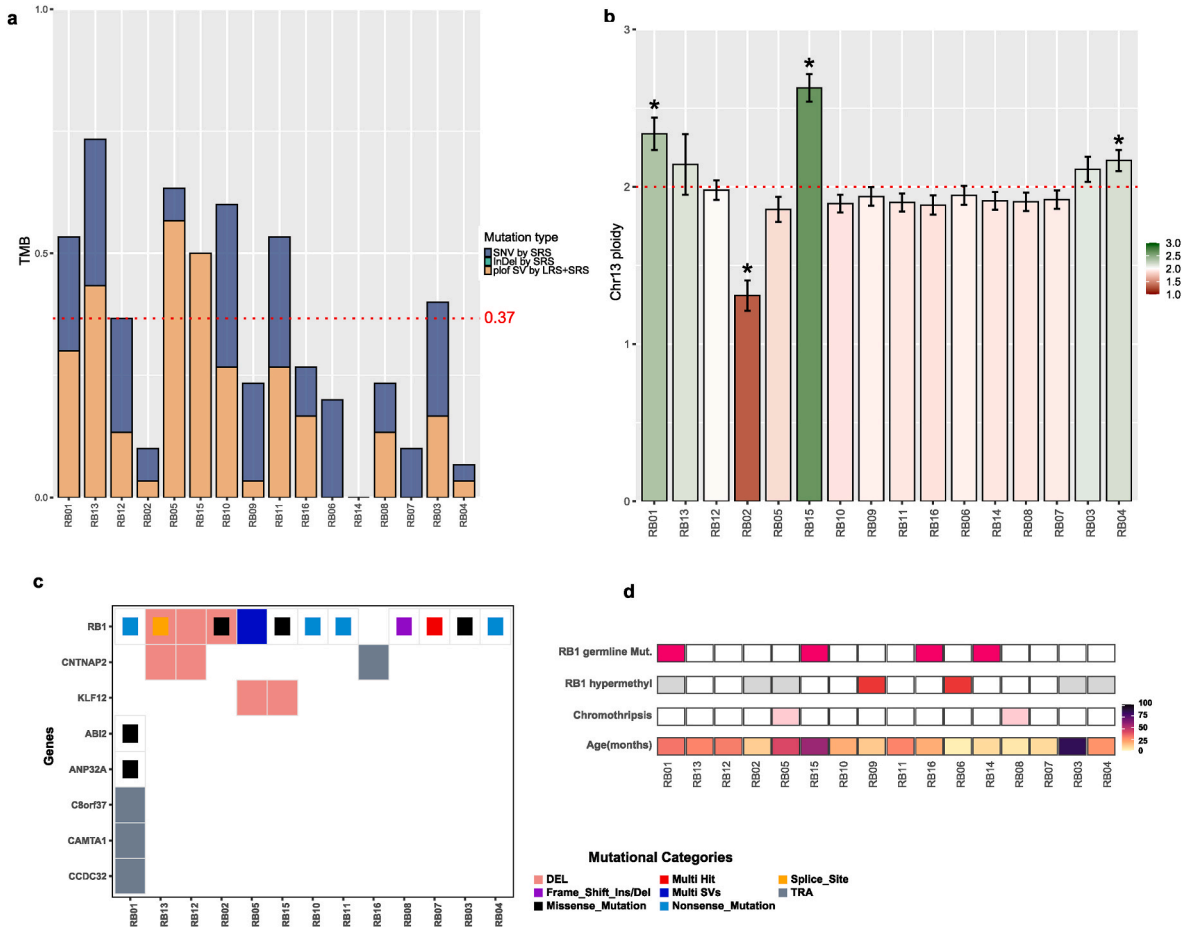

Fig. 2. The landscape of genomic alterations in RB. **a,** Number of somatic variants (SNV/InDel/pLoF SV) per Mb over the coding regions from LRS or SRS data. **b,** Mean ploidy of Chr13 from LRS data, ranging from 0 (red) to 3 (green; triploid). Diploidy is shown in white. **c,** Overview of somatic mutated genes per sample, background of each cell depicts SVs and the inner square depicts small (coding) variants. **d,** From top to bottom: presence of germline variants in *RB1*; presence of biallelic hypermethylation in *RB1* promoter; presence of chromothripsis; age of the patients at diagnosis. The samples with any nonsynonymous germline variant in *RB1* are shown in aubergine. The tumors with and without biallelic hypermethylation in *RB1* promoter are shown in red and white, respectively. The samples without methylation data are shown in gray. The tumors with chromothripsis are shown in pink.

RB08), affecting *RB1* and *MYCN*, respectively (Fig. S10, Table S12-13). Taken together, our findings suggested that both small variants and large SVs played significant roles in the somatic genetic variants of the disease.

2.3. Pathogenic alterations of RB1 in RB

Genes whose mutation facilitates tumor growth are called cancer driver genes [49,63,64]. By using two driver gene detection tools, WITER [49] and dNdScv [48], we identified *RB1* as the most potential driver gene in RB tumorigenesis (Fig. S11a), which was consistent with previous findings [2]. As a driver gene, both germline and somatic pLoF variants in *RB1* contributed to the tumorigenesis of RB. We identified a total of 20 non-redundant genomic variants spreading over *RB1* region in 16 samples, where three small variants and five somatic SVs were not reported in previous studies (Table 2, Table S14). In the blood of four patients, we detected five small variants with an allele frequency of less than 0.005 in Asian populations based on gnomAD (https://gnomad.broadinstitute.org/). These variants were found to effect alternative splicing or amino acid sequence of *RB1*, according to the annotation [65]. Consequently, we consider these rare germline variants to be pathogenic (Table S14). Nearly half of the variants (48 %) were distributed in the *RB1* pocket domains that covered exons 12 to 18 (Fig. S11b).

Together, among the 16 patients with *RB1* alterations, four patients exhibited germline variants in their blood samples, while 12 patients had somatic variants specifically detected in their tumor tissues. Additionally, four patients presented somatic SVs involving *RB1* in their tumor tissues. Some patients had multiple alterations in *RB1*, such as two somatic SNVs in RB07, two germline SNVs in RB16, and two somatic SVs in RB05 (Figs. 3 and 4, Table 2, Fig. S12). Furthermore, we observed loss of heterozygosity (LOH) events affecting *RB1* in more than 56 % of the patients (Fig. S7, S13-14, Table S15).

To further verify somatic SVs affecting *RB1*, we did a PCR assay to validate a 1059-bp somatic DEL disrupting *RB1* in RB12 (Fig. S12b). The fragment length of the PCR production should theoretically be 1469 bp in normal. Our results showed the presence of two fragments with lengths of approximately 1500 bp and 380 bp in the RB12 tumor sample, confirming this somatic SV. For the remaining samples who lacked additional tissues or DNA for PCR validation and Sanger sequencing, we used BLASTN for pairwise alignment of the supporting LRS reads for their SVs with the reconstructed sequence (Methods). Dotplot shows that these large SVs in *RB1* are reliable (Fig. S15, Table S8).

In addition to genomic variants, aberrant DNA methylation (DNAme), such as focal hypermethylation at promoters of tumor suppressor genes (TSGs), may also lead to TSG inactivation and drive cancer initiation [66]. For instance, despite no pLoF variant being found in RB06 and RB09 at the genomic level, we found biallelic hypermethylation in the *RB1* promoter of the two tumor samples (Fig. 5, S16, Table S16), which could potentially silence the expression of *RB1*. This

Table 2
Predicted pathogenic variants in *RB1*.

	RB1 SNV/InDel		*RB1* pLoF SV		*RB1* promoter hypermethylation		*RB1* LOH	*RB1* biallelic inactivation
	Germline	Somatic	Germline	Somatic	Germline	Somatic		
RB01	c.607+1G > A	c.C1735T:p.R579*X*	–	–	NA	NA	×	yes
RB02	–	c.T1472C:p.L491P	–	DEL:Start-End	NA	NA	✓	yes
RB03	–	c.T2272C:p.S758P	–	–	NA	NA	×	unknown
RB04	–	c.C1399T:p.R467*X*	–	–	NA	NA	✓	yes
RB05	–	–	–	DEL:intron2-End TRA: intron17–14:47,925,869	NA	NA	×	yes
RB06	–	–	–	–	×	✓	×	yes
RB07	–	c.C1399T:p.R467*X*, c.C1981T:p.R661W	–	–	×	×	×	yes
RB08	–	c.1448_1449del:p.H483fs	–	–	×	×	✓	yes
RB09	–	–	–	–	×	✓	×	yes
RB10	–	c.C1909T:p.Q637*X*	–	–	×	×	✓	yes
RB11	–	c.C1654T:p.R552*X*	–	–	×	×	✓	yes
RB12	–	splice_region: TTTTTACTTTTAGTAAAAAA > –	–	DEL:intron16-exon17	×	×	×	yes
RB13	–	c.2325+1G > A	–	DEL:Start-End	×	×	✓	yes
RB14	c.2489+1G > T	–	–	–	×	×	✓	yes
RB15	c.C958T:p.R320*X*	c.C1027A:p.L343I	–	–	×	×	✓	yes
RB16	c.C1861A:p.R621S, c.2093delG:p.R698fs	–	–	–	×	×	✓	yes

Variants in *RB1*, including SNVs/InDels, SVs and promoter hypermethylation in germline or somatic forms.
Annotation of *RB1* is according to NM_000321.3; "Start" and "End" denote the start and end loci.
"-" denotes that the variants of this type are not detected in this sample.
"NA" denotes that no data is available for methylation analysis.

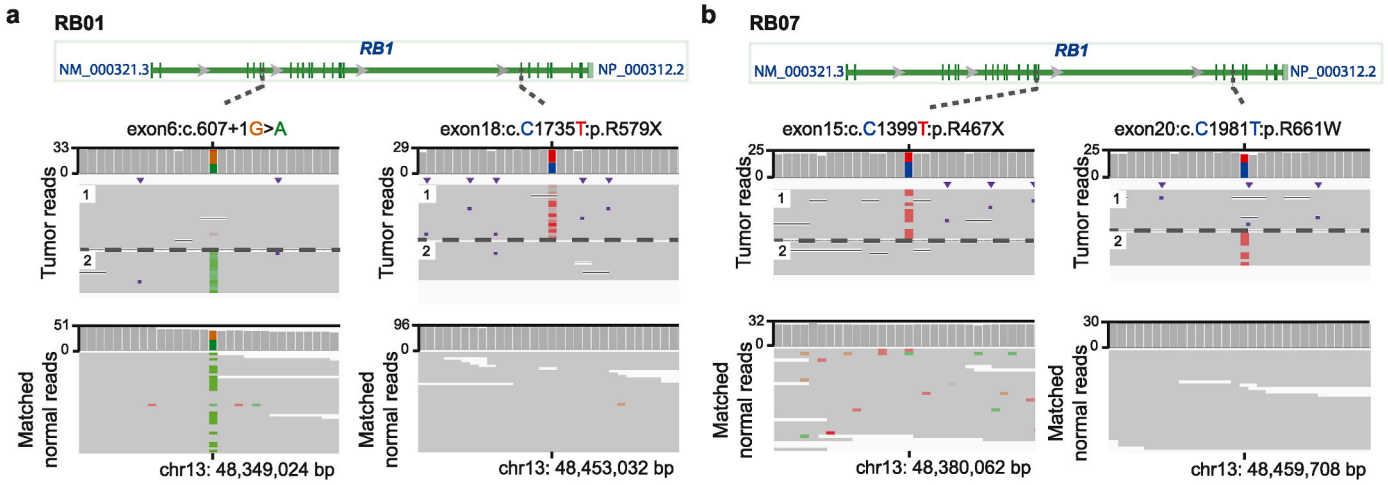

Fig. 3. *RB1* biallelic inactivation by small variants. **a,** One somatic stopgain SNV (exon18:c.C1735T:p.R579*X*) and one germline splicing SNV (exon6:c.607+1G > A) are located on separated haplotypes, forming a 'two-hit' model in RB01 sample. **b,** Two somatic SNV (exon15:c.C1399T:p.R467*X* and exon20:c.C1981T:p.R661W) locate on separated haplotypes, forming a compound heterozygous model in RB07 sample.

highlights the advantage of simultaneous detection of SV and methylation by ONT data in exploring the pathogenesis of RB in case of no detectable mutation before.

2.4. Two-hit of *RB1* in RB

By taking advantage of long-read data, we detected the phasing information of variants in *RB1* (**Methods**), which is missed by the short-read-based method in phasing of cancerous variants and linking together adjacent variants [46]. Our analysis showed various forms of *RB1* inactivation in tumor tissues. For instance, in RB01, we detected one germline splicing SNV (NM_000321, exon6:c.607+1G > A) and one somatic stopgain SNV (exon18:c.C1735T:p.R579*X*) in different exons of *RB1*, belonging to two distinct haplotypes (Fig. 3a). In RB07, we detected no germline variant but two somatic mutations (exon15:c.C1399T:p.R467*X* and exon20:c.C1981T:p.R661W), located on different haplotypes (Fig. 3b). Our study also revealed the significant impact of pLoF SVs on *RB1* inactivation, which has been comparatively underreported until now. In RB12, we detected a 20-bp somatic InDel flanking the splicing donor site, which was further validated to result in the skipping of exon 15 by transcript analysis of *RB1*, and one 1059-bp somatic DEL located in different haplotypes (Fig. 4a–S12a-b). In RB05, we found that a 17 Mb somatic DEL, covering the intron 2 to transcription termination site (TTS) of *RB1*, and a TRA was simultaneously detected inside this region, which was validated by the supporting LRS reads (Fig. 4b). These two SVs presented in the same clonal status when calculated by PyClone-VI [58], indicating the biallelic inactivation of *RB1* by complex somatic SVs in RB05. In RB16, we found two heterozygous germline variants (exon19:c.C1861A:p.R621S and exon20:c.2093delG:p.R698fs) in the same haplotype. Interestingly, we observed that the tumor tissue had a germline variant (exon19:c.C1861A) with allele frequency (AF) of 1.0 due to losing normal allele by LOH, resulting

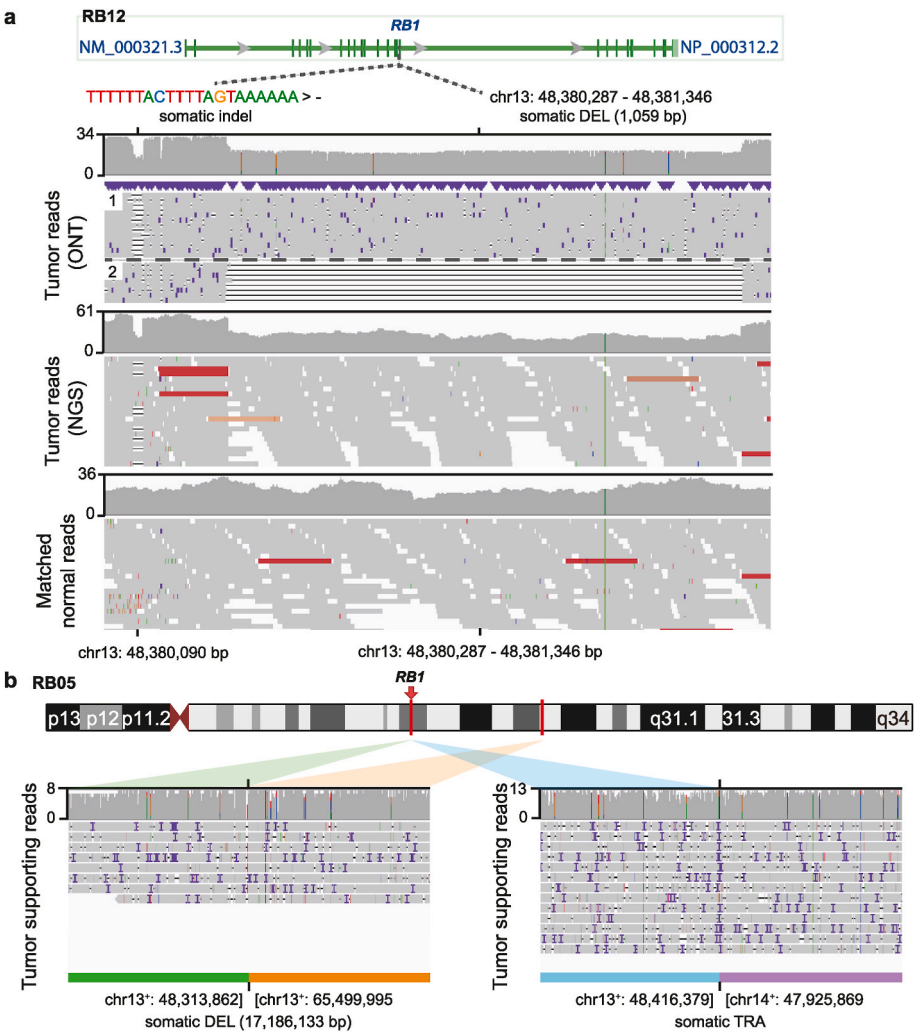

Fig. 4. *RB1* biallelic inactivation by small variants and SVs. **a,** One somatic InDel flanking the splicing donor site and one 1059-bp somatic DEL located in different haplotypes in RB12. **b,** Two somatic SVs (DEL and TRA) form complex heterozygous somatic SVs in RB05.

in complete inactivation of *RB1* in tumor tissue (Fig. S12c). To sum, no sample had two germline variants on different haplotypes, possibly due to the homozygous embryo lethal effect of complete germline inactivation of *RB1* that had been observed in mice [67,68]. Totally, our findings showed that except for RB03 who lacked the methylation data for phasing, all patients had biallelic inactivation of *RB1* in tumor tissues that fitted the two-hit model with multiple forms, such as a germline variant plus a somatic mutation, entire somatic mutations, and germline variants plus LOH, but not germline homozygous mutation.

2.5. Inferring relative timing of somatic events in RB

Somatic evolution plays a key role in the development of various types of cancers, which has been the subject of extensive investigation [69,70]. For RB, the genomic mutation process during tumor development is an as-yet unidentified event, presumably due to the limited variants in RB [71,72]. On this line of reasoning, we tried to combine the information of SNVs, InDels, SVs, as well as CNAs to access the relative timing of genetic progression in RB [57].

We retained only events that occurred in at least three samples across the cohort. The results showed that nearly all the estimated genetic progression originated from *RB1* variants. Examples of these genomic alterations are described here along with underlying structural rearrangements (Fig. 6, Fig. S17, Table S17-18). In RB16, we observed that copy neutral LOH (*CN*-LOH) of Chr13 in early clonal stage, which resulted in the inactivation of *RB1*, followed by amplifications of Chr1q and Chr7q. Later events included diverse alterations, such as copy loss LOH (CL-LOH) of Chr16 (Fig. 6a). In RB01, the initial variants included mutations in *RB1* and amplifications of Chr1q, Chr6p, and Chr13q, followed by deletions of Chr8p, Chr11q, Chr12p13, and Chr16. (Fig. 6b). In RB13, *RB1* variants tended to occur early, followed by amplification of Chr13q and CL-LOH of Chr5q (Fig. 6c). In RB05, who had co-occurrence of *RB1* somatic SV and *MYCN* amplification, variants of *RB1*, amplifications of Chr1q, Chr2p24, and deletions of Chr13q14.2, occurred before the deletions of Chr8p, Chr11p, Chr12p13 and Chr16q (Fig. 6d).

Overall, across all samples in our study, we found *RB1* was the only driver gene that had mutations during RB genetic progression. Based on the PhylogicNDT LeagueModel simulation results, these rankings across samples counted the probability order of occurrence of lesions, summarizing whether each mutation occurred earlier or later in the development of the tumor. We inferred that *RB1* mutated among the earliest events, then the other chromosomes presented instability (Fig. 6e–f). Our analysis is similar to the tumor progression model proposed by previous study [73]. The relative timing of somatic events of individual tumor specimens was highly variable, which was also observed in different types of cancers [57]. From the mutation time and prevalence, we found that common drivers typically occur before rare drivers [57]. Therefore, our study inferred the relative timing of somatic events during RB tumorigenesis, showing that the genetic progression of RB

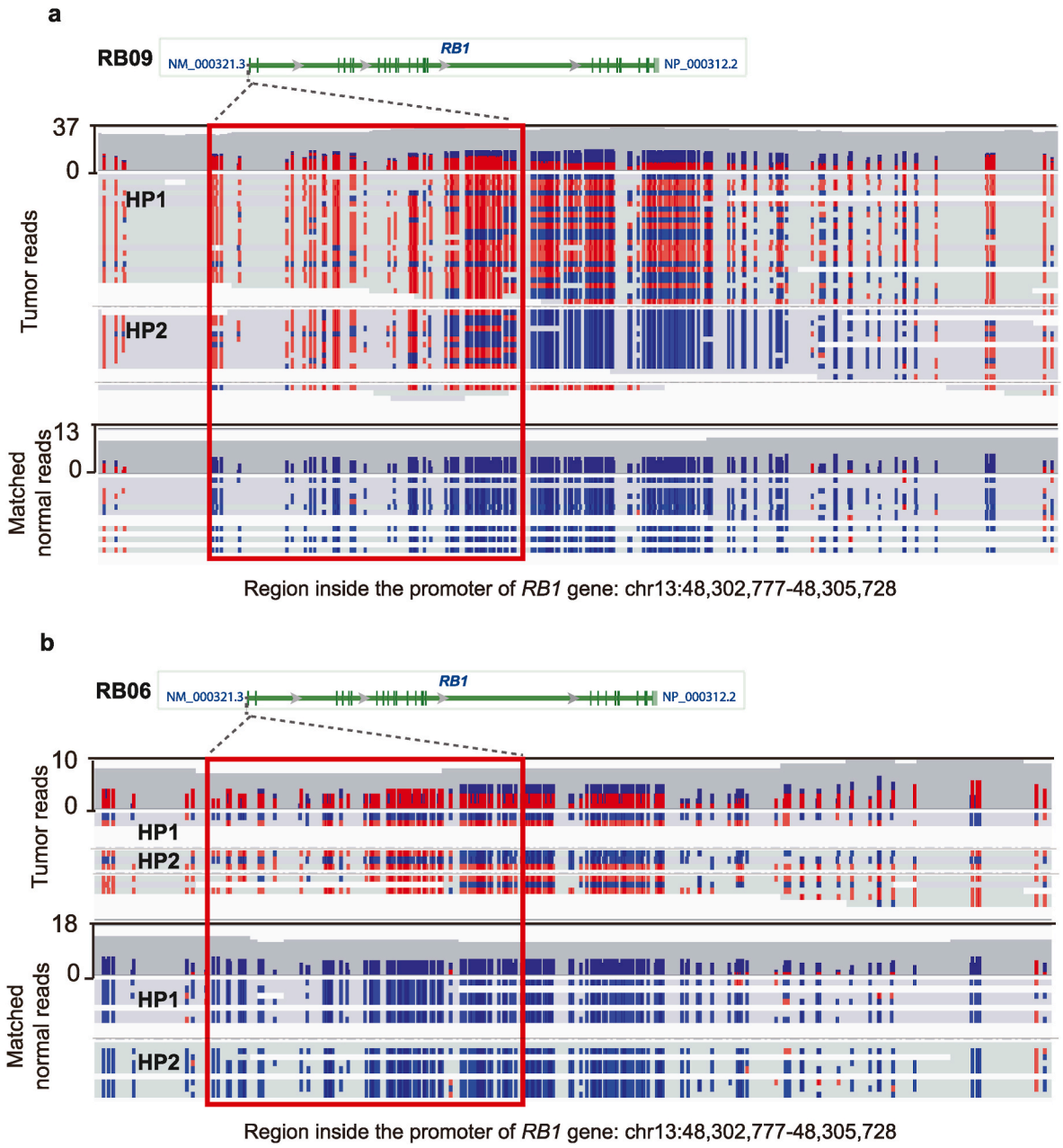

Fig. 5. *RB1* biallelic inactivation by methylation. Tumor sample is showed in the top track and blood (normal) in the bottom track. Red indicates hypermethylation and blue indicates hypomethylation. HP1 and HP2 are the different haplotypes inferred from PEPPER-Margin-DeepVariant pipeline. The range in the red box is the biallelic hypermethylation in *RB1* promoter region. **a,** Somatic hypermethylation of *RB1* promoter region in RB09. **b,** Somatic hypermethylation of *RB1* promoter region in RB06.

follows trends although it is far from deterministic.

3. Discussion

In this study, we use LRS and SRS to interrogate comprehensive pathogenic variants of 16 RB tumors, which is the largest cohort that employed LRS to investigate somatic SVs in RB to date. The addition of LRS enables us to detect more SVs across the genome due to their longer reads allowing for covering low-complex regions where SVs tend to occur [13,15]. For instance, when we compared the confident results from 11 samples' whole genome LRS and SRS data, note that these SVs were manually checked using IGV and could be considered reliable, we found that more SVs were identified from LRS than SRS (140 vs. 122). Notably, all but one of the 18 INSs were uniquely detected by LRS, indicating that LRS exhibits increased sensitivity in detecting SVs in RB, particularly the mutational type of INSs (Fig. S5).

As for the variants in *RB1*, whose inactivation is generally thought to be the cause of RB initiation, although we thoroughly analyzed genomic data, there are still no detectable causal variants in two samples (RB06 and RB09). To address this issue, we performed methylation analysis of the 11 samples based on ONT data, the results showed that there are somatic DNA hypermethylation of *RB1*'s promoters in the tumor tissues of RB06 and RB09, lending weight to the notion that combinatory use of LRS and SRS could better explain the genomic and epigenomic features of cancer genomes to pinpoint the causal variants of RB.

Despite the comprehensive characterization of pathogenic variants in RB, we are aware that the resolution of causal variants at a haplotype level are required to clarify biallelic loss of *RB1*. We performed

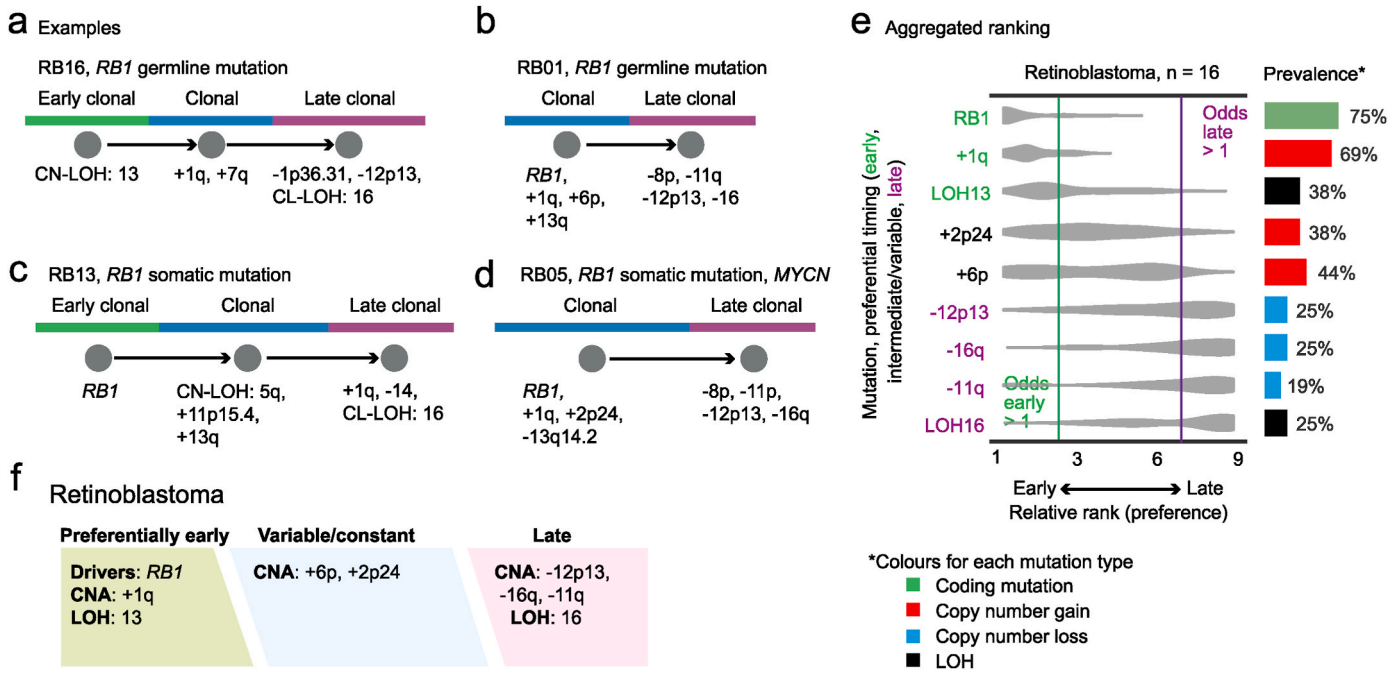

Fig. 6. Relative timing of somatic events in RB. Examples shows the relative timing of genetic progression in individual patients, the constituent data for the ordering model process. a, Timing of somatic events of RB16 who had two germline SNVs in RB1. b, Timing of somatic events of RB01 who had germline and somatic SNVs in RB1. **c,** Timing of somatic events of RB13 who had somatic LOH in *RB1*. **d,** Timing of somatic events of RB05 who had somatic SVs in *RB1* and *MYCN* amplification. **e,** Preferential ordering diagrams for RB across 16 samples. **f,** Relative timing of somatic events during RB development. Direction from left to right represents timelines between early and late stage of tumor.

haplotype phasing analysis of variants in *RB1* using ONT long reads. Since the samples (RB01-RB05) lack raw fast5 data for methylation detection, we assigned the genomic variants of 16 samples and epigenomic variants of 11 samples into two haplotypes (**Methods**). We found that except for RB03 who lacked the methylation data for phasing, all sample presented biallelic inactivation of *RB1* in various forms. It is worth noting that, in some cases, fully inactivation of *RB1* could only be explained by LRS data, such as the complex heterozygous mutations by a small InDel and a large SV in RB12, and the biallelic hypermethylated promoter of *RB1* in RB06 and RB09. One important caveat of our analysis is the comprehensive identification and phasing of SNVs, InDels, SVs and methylation by combinatorial analysis of LRS and SRS lay the foundation for interpretation of biallelic loss of *RB1*.

There are some limits in this study. Firstly, with the respect to genetic progression in RB, although we inferred the relative timing and prevalence using all genetic variants of 16 samples, more samples with different stages of disease are required to investigate the details of evolutionary trajectories [57,59]. Secondly, considering that LRS could cover the regions not easily accessible to SRS or arrays technologies [74], we analyzed CNAs and ploidy features using LRS, but the ability to identify LOH is inadequate because of error-prone detection of small variants by ONT. Once the accuracy of ONT data and the detection method improved, we will obtain a more thorough overview of genomic feature based on ONT data alone. Thirdly, we were unable to perform comprehensive validation experiments for the somatic SVs due to the limited size of the RB tumor tissue samples. The DNA extracted from these samples was only sufficient for constructing sequencing libraries for NGS and ONT platforms. While we were able to orthogonally validate the SVs in samples RB08 and RB12 using PacBio HiFi sequencing and PCR, respectively, the lack of additional tissues or DNA prevented further validation in other samples. Future studies with larger tumor samples that provide sufficient DNA for both sequencing and validation experiments will help independently confirm the accuracy of SVs identified by LRS in RB genomes, particularly those uniquely detected by ONT and those impacting key RB driver genes like *RB1*.

Overall, our study characterizes novel somatic SVs and delivers a thorough overview of the genetic landscape of the RB genome, which expands our understanding of pathogenic variants in RB. By leveraging the capabilities of LRS data, we are able to uncover the phasing information that effectively explains the biallelic inactivation of *RB1* as well as the relative timing of genetic progression in RB. This information advances our understanding of the occurrence and progression of RB, filling in gaps left by previous studies.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Zhongshan Ophthalmic Center, Sun Yat-sen University. The informed consent was signed by all the patients. According to ethical and legal standards, every specimen was made and handled anonymously.

Funding

This research was supported by grants from the National Key R&D Program of China (2019YFA0904400, Z.X.) and the GuangZhou city, Schools and Hospitals (202201020336, Z.X.).

Data sharing statement

Whole genome sequencing data has been submitted to GSA-Human BioProject (https://ngdc.cncb.ac.cn/gsa-human/) under accession number PRJCA008987, including Nanopore data of RB01-RB16, PacBio HiFi data of RB08 and Illumina WGS data of RB06-RB16. The Illumina WES data of RB01-RB05 was obtained from our previous study [23]. The IGV images and dotplot figures from BLASTN have been uploaded to the Zenodo (https://zenodo.org/records/11504094).

Code availability

The codes of data analysis are publicly available at https://github.

com/xie-lab/EC-SSV.

CRediT authorship contribution statement

Jingjing Zheng: Writing – review & editing, Writing – original draft, Visualization, Investigation, Data curation. **Tong Li:** Writing – review & editing, Visualization, Data curation. **Huijing Ye:** Data curation, Writing – review & editing. **Zehang Jiang:** Data curation. **Wenbing Jiang:** Validation. **Huasheng Yang:** Conceptualization. **Zhikun Wu:** Writing – review & editing, Data curation. **Zhi Xie:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We extend our sincere gratitude to all the patients with RB who participated in our study, and we wish them good health. We would like to acknowledge the Center for Precision Medicine at Sun Yat-sen University for the long-term support. We would also like to thank Cheng Wu, Shuai Ouyang, and Yun Zhang for their assistance with sample collection and bioinformatic analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.canlet.2024.217121.

References

[1] H. Dimaras, et al., Retinoblastoma, Nat. Rev. Dis. Prim. 1 (2015) 15021.
[2] H. Dimaras, et al., Retinoblastoma, Lancet 379 (2012) 1436–1446.
[3] J.H. Francis, et al., Molecular changes in retinoblastoma beyond RB1: findings from next-generation sequencing, Cancers 13 (2021).
[4] J. Zhang, et al., A novel retinoblastoma therapy from genomic and epigenetic analyses, Nature 481 (2012) 329–334.
[5] J. McEvoy, et al., RB1 gene inactivation by chromothripsis in human retinoblastoma, Oncotarget 5 (2014) 438–450.
[6] S. Richter, et al., Sensitive and efficient detection of RB1 gene mutations enhances care for families with retinoblastoma, Am. J. Hum. Genet. 72 (2003) 253–269.
[7] D. Parma, M. Ferrer, L. Luce, F. Giliberto, I. Szijan, RB1 gene mutations in Argentine retinoblastoma patients. Implications for genetic counseling, PLoS One 12 (2017) e0189736.
[8] H.R. Davies, et al., Whole-genome sequencing of retinoblastoma reveals the diversity of rearrangements disrupting RB1 and uncovers a treatment-related mutational signature, Cancers 13 (2021).
[9] C.J. Dommering, et al., RB1mutation spectrum in a comprehensive nationwide cohort of retinoblastoma patients, J. Med. Genet. 51 (2014) 366–374.
[10] E.A. Price, et al., Spectrum of RB1 mutations identified in 403 retinoblastoma patients, J. Med. Genet. 51 (2014) 208–214.
[11] T. Cai, et al., Mutation spectrum of RB1 mutations in retinoblastoma cases from Singapore with implications for genetic management and counselling, PLoS One 12 (2017).
[12] F. Dubois, N. Sidiropoulos, J. Weischenfeldt, R. Beroukhim, Structural variations in cancer and the 3D genome, Nat. Rev. Cancer 22 (2022) 533–546.
[13] G.A. Logsdon, M.R. Vollger, E.E. Eichler, Long-read human genome sequencing and its applications, Nat. Rev. Genet. 21 (2020) 597–614.
[14] Y. Liu, et al., DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation, Genome Biol. 22 (2021) 295.
[15] M.J.P. Chaisson, et al., Multi-platform discovery of haplotype-resolved structural variation in human genomes, Nat. Commun. 10 (2019) 1784.
[16] M. Cretu Stancu, et al., Mapping and phasing of structural variation in patient genomes using nanopore sequencing, Nat. Commun. 8 (2017) 1326.
[17] M. Nattestad, et al., Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line, Genome Res. 28 (2018) 1126–1135.
[18] J. Huddleston, et al., Discovery and genotyping of structural variation from long-read haploid genome sequence data, Genome Res. 27 (2017) 677–685.
[19] W. De Coster, et al., Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome, Genome Res. 29 (2019) 1178–1187.
[20] Y. Sakamoto, et al., Phasing analysis of lung cancer genomes using a long read sequencer, Nat. Commun. 13 (2022) 3464.
[21] K.M. Schieffer, et al., Molecular classification of a complex structural rearrangement of the RB1 locus in an infant with sporadic, isolated, intracranial, sellar region retinoblastoma, Acta Neuropathol Commun 9 (2021) 61.
[22] K. Nakamichi, A. Stacey, D. Mustafi, Targeted long-read sequencing allows for rapid identification of pathogenic disease-causing variants in retinoblastoma, Ophthalmic Genet. 43 (2022) 762–770.
[23] C. Wu, et al., Single-cell characterization of malignant phenotypes and microenvironment alteration in retinoblastoma, Cell Death Dis. 13 (2022) 438.
[24] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (2018) i884–i890.
[25] Z. Wu, et al., Structural variants in the Chinese population and their impact on phenotypes, diseases and population adaptation, Nat. Commun. 12 (2021).
[26] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26 (2010) 589–595.
[27] K. Cibulskis, et al., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nat. Biotechnol. 31 (2013) 213–219.
[28] A. McKenna, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303.
[29] S. Kim, et al., Strelka2: fast and accurate calling of germline and somatic variants, Nat. Methods 15 (2018) 591–594.
[30] S.M.E. Sahraeian, et al., Deep convolutional neural networks for accurate somatic mutation detection, Nat. Commun. 10 (2019) 1041.
[31] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (2010) e164.
[32] H. Li, Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics 34 (2018) 3094–3100.
[33] F.J. Sedlazeck, et al., Accurate detection of complex structural variations using single-molecule sequencing, Nat. Methods 15 (2018) 461–468.
[34] T. Jiang, et al., Long-read-based human genomic structural variation detection with cuteSV, Genome Biol. 21 (2020) 189.
[35] Y. Shiraishi, et al., Precise characterization of somatic complex structural variations from tumor/control paired long-read sequencing data with nanomonsv, Nucleic Acids Res. 51 (2023) e74, e74.
[36] J.R. Wang, J. Holt, L. McMillan, C.D. Jones, FMLRC: hybrid long read error correction using an FM-index, BMC Bioinf. 19 (2018) 50.
[37] M. Kirsche, et al., Jasmine and Iris: population-scale structural variant comparison and analysis, Nat. Methods 20 (2023) 408–417.
[38] M. Kirsche, et al., Jasmine: population-scale structural variant comparison and analysis, bioRxiv (2021) 445886, 2021.05.27.
[39] J.T. Robinson, H. Thorvaldsdottir, A.M. Wenger, A. Zehir, J.P. Mesirov, Variant review with the integrative genomics viewer, Cancer Res. 77 (2017) e31–e34.
[40] V. Geoffroy, et al., AnnotSV: an integrated tool for structural variations annotation, Bioinformatics 34 (2018) 3572–3574.
[41] J. Espejo Valle-Inclan, et al., A multi-platform reference for somatic structural variation detection, Cell Genom 2 (2022) 100139.
[42] X. Chen, et al., Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications, Bioinformatics 32 (2016) 1220–1222.
[43] T. Rausch, et al., DELLY: structural variant discovery by integrated paired-end and split-read analysis, Bioinformatics 28 (2012) i333–i339.
[44] J.A. Wala, et al., SvABA: genome-wide detection of structural variants and indels by local assembly, Genome Res. 28 (2018) 581–591.
[45] N.J. Loman, J. Quick, J.T. Simpson, A complete bacterial genome assembled de novo using only nanopore sequencing data, Nat. Methods 12 (2015) 733–735.
[46] K. Shafin, et al., Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads, Nat. Methods 18 (2021) 1322–1332.
[47] I. Lee, et al., Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing, Nat. Methods 17 (2020) 1191–1199.
[48] I. Martincorena, et al., Universal patterns of selection in cancer and somatic tissues, Cell 171 (2017) 1029–1041 e21.
[49] L. Jiang, et al., WITER: a powerful method for estimation of cancer-driver genes using a weighted iterative regression modelling background mutation counts, Nucleic Acids Res. 47 (2019) e96.
[50] B.S. Pedersen, A.R. Quinlan, J. Hancock, Mosdepth: quick coverage calculation for genomes and exomes, Bioinformatics 34 (2018) 867–868.
[51] A. Magi, et al., Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data, Bioinformatics 35 (2019) 4213–4221.
[52] S. Zaccaria, B.J. Raphael, Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data, Nat. Commun. 11 (2020) 4301.
[53] M.U. Ahsan, Q. Liu, L. Fang, K. Wang, NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks, Genome Biol. 22 (2021) 261.
[54] C.A. Maher, R.K. Wilson, Chromothripsis and human disease: piecing together the shattering process, Cell 148 (2012) 29–32.
[55] I. Cortes-Ciriano, et al., Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing, Nat. Genet. 52 (2020) 331–341.
[56] H.E. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots, BMC Bioinf. 14 (2013).
[57] M. Gerstung, et al., The evolutionary history of 2,658 cancers, Nature 578 (2020) 122–128.
[58] S. Gillis, A. Roth, PyClone-VI: scalable inference of clonal population structures using whole genome data, BMC Bioinf. 21 (2020) 571.

[59] I. Leshchiner, et al., Inferring early genetic progression in cancers with unobtainable premalignant disease, Nat. Can. (Ott.) 4 (2023) 550–+.
[60] J. Ye, et al., Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.
[61] A. Fujimoto, et al., Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer, Genome Med. 13 (2021) 65.
[62] J. Espejo Valle-Inclan, et al., A multi-platform reference for somatic structural variation detection, Cell Genomics 2 (2022).
[63] M.R. Stratton, P.J. Campbell, P.A. Futreal, The cancer genome, Nature 458 (2009) 719–724.
[64] F. Martinez-Jimenez, et al., A compendium of mutational cancer driver genes, Nat. Rev. Cancer 20 (2020) 555–572.
[65] S. Richards, 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. 17 (2015) 405–424.
[66] H. Pan, et al., Discovery of candidate DNA methylation cancer driver genes, Cancer Discov. 11 (2021) 2266–2281.
[67] T. Jacks, et al., Effects of an Rb mutation in the mouse, Nature 359 (1992) 295–300.
[68] E.Y. Lee, et al., Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis, Nature 359 (1992) 288–294.
[69] J. Cairns, Mutation selection and the natural history of cancer, Nature 255 (1975) 197–200.
[70] I. Martincorena, P.J. Campbell, Somatic mutation in cancer and normal cells, Science 349 (2015) 1483–1489.
[71] C.A. Benavente, M.A. Dyer, Genetics and epigenetics of human retinoblastoma, Annu. Rev. Pathol. 10 (2015) 547–562.
[72] I.E. Kooi, et al., Somatic genomic alterations in retinoblastoma beyond RB1 are rare and limited to copy number changes, Sci. Rep. 6 (2016) 25264.
[73] K. Sampieri, et al., Genomic differences between retinoma and retinoblastoma, Acta Oncol. 47 (2009) 1483–1492.
[74] K. Lavrichenko, S. Johansson, I. Jonassen, Comprehensive characterization of copy number variation (CNV) called from array, long- and short-read data, BMC Genom. 22 (2021) 826.