Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort

Jaya Singh,[1] Avshesh Mishra,[1] Arunachalam Jayamuruga Pandian,[2] Ashwin C. Mallipatna,[3] Vikas Khetan,[4] S. Sripriya,[2] Suman Kapoor,[1] Smita Agarwal,[1] Satish Sankaran,[1] Shanmukh Katragadda,[1] Vamsi Veeramachaneni,[1] Ramesh Hariharan,[1,5] Kalyanasundaram Subramanian,[1] Ashraf U. Mannan[1]

(The first two authors contributed equally to this work.)

[1]*Strand Center for Genomics and Personalized Medicine, Strand Life Sciences, Bangalore, India;* [2]*Sankara Nethralaya ONGC Department of Genetics and Molecular Biology, Vision Research Foundation, Chennai, India;* [3]*Department of Pediatric Ophthalmology and Strabismology, Narayana Nethralaya, Bangalore, India;* [4]*Department of Vitreo Retina, Sankara Nethralaya, Chennai, India;* [5]*Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India*

Purpose: Retinoblastoma (Rb) is the most common primary intraocular cancer of childhood and one of the major causes of blindness in children. India has the highest number of patients with Rb in the world. Mutations in the *RB1* gene are the primary cause of Rb, and heterogeneous mutations are distributed throughout the entire length of the gene. Therefore, genetic testing requires screening of the entire gene, which by conventional sequencing is time consuming and expensive.
Methods: In this study, we screened the *RB1* gene in the DNA isolated from blood or saliva samples of 50 unrelated patients with Rb using the TruSight Cancer panel. Next-generation sequencing (NGS) was done on the Illumina MiSeq platform. Genetic variations were identified using the Strand NGS software and interpreted using the StrandOmics platform.
Results: We were able to detect germline pathogenic mutations in 66% (33/50) of the cases, 12 of which were novel. We were able to detect all types of mutations, including missense, nonsense, splice site, indel, and structural variants. When we considered bilateral Rb cases only, the mutation detection rate increased to 100% (22/22). In unilateral Rb cases, the mutation detection rate was 30% (6/20).
Conclusions: Our study suggests that NGS-based approaches increase the sensitivity of mutation detection in the *RB1* gene, making it fast and cost-effective compared to the conventional tests performed in a reflex-testing mode.

Retinoblastoma (Rb) is a malignant tumor of the developing retina that occurs in children, usually before the age of five years, and it causes childhood blindness [1]. According to the World Health Organization (WHO), the average age-adjusted incidence rate of Rb in the United States and Europe is 2–5 cases per million children (approximately 1 in 14,000–18,000 live births) [2]. As per the latest National Cancer Registry Program (NCRP) report, in India, the age-adjusted rates of Rb incidence are estimated to be 1.9–12.3 and 1.3–6.7 per million in boys and girls, respectively [3]. Due to its early age of occurrence and the risk of second cancers (soft tissue sarcomas, osteosarcomas, and melanomas) at later stages of life, early molecular diagnosis and treatment options must be considered for better management of the disease [4,5].

India has the highest number of Rb cases, where almost 20% of the world's Rb patients reside in India [4]. In developed countries, children with Rb have a disease-free survival rate greater than 90%, compared to developing nations, where it is substantially lower, at 10–30% [6,7]. As with other developing countries, late diagnosis, lack of awareness, and the inaccessibility of specialized care are the major reasons for tumor metastasis in India [4]. The burden of Rb on the Indian health care system has been steadily increasing, thus stressing the need for cost-effective methods for early detection, surveillance, and disease management.

Rb is a tumor that occurs in both heritable (25–30%) and non-heritable (70–75%) forms. A heritable disease is defined by the presence of a germline mutation in the *RB1* gene (Gene ID: 5925, OMIM 614041), which is followed by a somatic mutation in the developing retina. It can result in tumors affecting either one (unilateral) or both (bilateral) eyes. In the non-heritable form of Rb, both mutations occur in the somatic cells, leading only to unilateral tumors [8]. Usually, a familial, bilateral, or multifocal disease is suggestive of a heritable disease, whereas older children with a unilateral tumor are more likely to have the non-heritable form of the disease [9].

Correspondence to: Ashraf U. Mannan, Strand Life Sciences, Kirloskar Business Park, Bellary Road, Hebbal, Bangalore-560024, India; Phone: +91-80-40787353; FAX: +91-80-40787299; email: ashraf@strandls.com

In India, the few studies that have been conducted to determine the prevalence of *RB1* mutations in various Indian cohorts reported mutation detection rates ranging from 33% to 85% for both unilateral and bilateral cases [10-14]. In one of the first studies in India, Ata-ur-Rasheed et al. screened 21 patients with Rb using the Sanger sequencing method and identified *RB1* mutations in seven patients, and the mutation detection rate was 33.3% [11]. In another study, Kiran et al. screened 47 patients by single-strand conformation polymorphism (SSCP) followed by sequencing and reported a mutation detection rate of 46% [13]. The screening of a relatively large cohort of 74 patients using a combinatorial approach including fluorescent quantitative multiplex PCR, fluorescent genotyping, restriction fragment length polymorphism (RFLP), and sequencing, Ali et al. reported a detection rate of 66% [10]. In a recent study, Deverajan et al. screened 33 patients from Southern India by targeted next-generation sequencing (NGS) and reported a mutation detection rate of 85% [12]. Collectively, from these studies, it is evident that there is a high variability in the reported detection rates of *RB1* mutations in various Indian cohorts.

The *RB1* gene shows a wide spectrum of mutations, including single nucleotide variations (SNVs), small insertions/deletions (indels), and large deletions/duplications. These mutations are distributed throughout the entire length of the gene, spanning 27 exons, and no hotspots have been reported. Conventional genetic testing of the *RB1* gene involves screening of all 27 exons and the flanking intronic regions by Sanger sequencing, followed by a deletion/duplication analysis by multiplex ligation-dependent probe amplification (MLPA). This sequential testing strategy performed in a reflex-testing mode is time consuming and expensive. New advances in genomic technologies, such as NGS, allow us to detect all types of variants, such as SNVs, indels, and structural variants, including large deletions/duplications, at a significantly lower cost than traditional methods. In the current study, we used an improved NGS-based method to screen the *RB1* gene in the DNA isolated from blood or saliva samples from an Indian Rb cohort (50 cases) and detected all types of germline mutations, including large deletions ranging from a single exon to a whole gene (>178 kb) deletion. Moreover, we report a mutation detection rate of 100% in bilateral Rb (22) cases.

METHODS

Clinical diagnosis and patients: Saliva or peripheral blood samples were obtained from 50 unrelated patients with an indication of Rb referred to our laboratory between March 2014 and January 2016. Informed consent was obtained from all subjects and sequencing of the patients' samples for this study was approved by the Institutional Ethics Committee of Strand Life Sciences. A clinical diagnosis of Rb was confirmed through a clinical examination conducted by the referring ophthalmologist. There were 20 patients with unilateral Rb, 22 patients with bilateral Rb, and 8 patients with unavailable information on laterality.

DNA was extracted from saliva samples using the PrepIT-L2P kit (DNA Genotek, Canada), as per the manufacturer's instructions. For blood samples, either the QIAamp DNA Mini Kit (Qiagen, Germany) or the Nucleospin kit (Macherey-Nagel, Germany) was used for DNA isolation, as per the manufacturer's instructions. The concentration of DNA was determined using the Qubit fluorimeter (Life Technologies).

Library preparation and targeted NGS: Targeted NGS was performed on patient genomic DNA using the Trusight Cancer sequencing panel (Illumina) that contains 1,736 genomic regions from 94 genes suspected of having a role in cancer predisposition, including the *RB1* gene. An analytical validation of our panel has shown a sensitivity of 98.2%, specificity of 100%, and reproducibility of 99.5%. The gene coverage analysis on this panel revealed that exonic and flanking intronic regions of the *RB1* gene (NM_000321) showed coverage of >99% (≥ 20 reads) with a mean read depth of 405X. The Nextera DNA library preparation protocol (Illumina) to convert input genomic DNA (gDNA) into adaptor-tagged indexed libraries was essentially performed as previously described [15]. The tagged and amplified sample libraries were checked for quality and they were quantified using the BioAnalyzer (Agilent). Up to 6–10 pM of the pooled library was loaded and sequenced on the MiSeq platform (Illumina), according to the manufacturer's instructions.

NGS – data analysis and interpretation: The trimmed FASTQ files were generated using MiSeq Reporter (Illumina). The reads were aligned against the whole genome build: hg19 using Strand NGS v2.5. Data analysis and interpretation were performed using Strand NGS v2.5 and StrandOmics v3.0 (a proprietary clinical genomics interpretation and reporting platform from Strand Life Sciences), as previously described [15]. In brief, StrandOmics is a clinical interpretation and reporting platform that combines knowledge from internal curated literature content (approximately 40,000 extra curated variant records), along with various publically available data sources such as Uniprot, OMIM, HGMD, ClinVar, ARUP, dbSNP, 1000 Genomes, Exome Variant Server, and Exome Aggregation Consortium (ExAC). In addition to databases, bioinformatics prediction tools, such as SIFT, PolyPhen HVAR/HDIV, Mutation Taster, Mutation Assessor,

FATHMM, LRT for missense variants, and NNSPLICE; and ASSP tools for variants in essential splice sites and exon–intron boundaries, have also been integrated to assess the pathogenicity of the variants. This integrated knowledge is then used to prioritize automatically a list of variants based on American College of Medical Genetics and Genomics (ACMG) guidelines [16], the inheritance model, disease phenotype, sequence conservation across various species, and allelic frequency in our laboratory's internal patient pooled database (PPDB). A variant was labeled 'novel' when it had not been previously reported in the literature or in any public database (as mentioned above).

Variant calling and classification: Reads with average base quality <Q20 were excluded from the variant calling process, and the Bayesian approach was used to identify the consensus genotype at the variant locus. Each called variant was assigned a Phred equivalent score that represents base-calling error probabilities. The identified variants in this study were called with a read quality >Q30 and a confidence score >50.

The identified variants were labeled according to the ACMG recommended standards for the interpretation and reporting of sequence variations [16]. The variants were classified into five categories: 1) pathogenic, 2) likely pathogenic, 3) variant of uncertain significance (VUS), 4) likely benign, and 5) benign.

Copy number variation analysis for large deletion/duplication: In addition to SNVs and small indels, a copy number analysis was performed to identify large deletions or insertions ranging from a single exon to whole gene deletion. This was done by taking each non-overlapping target region in turn, of which there are 1,736, and comparing normalized read coverage across 8–11 other samples from the same run. Normalized coverage-based copy number values (CNVs) and Z-scores [17] for each panel region were computed using StrandNGS v2.5. For each sample, potential copy number changes in the *RB1* gene were identified by manual interpretation based on the following cut-offs: CNV >3, Z-score >2 for duplications and CNV <1.2, Z-score <-2 for deletions.

Split read analysis for the identification of break points: Reads that did not align with an alignment score >95% were subjected to split read alignment [18]. Here, the input reads were split into two segments and each segment was mapped independently to the reference genome. The minimum size of the major segment was 35 bp and that of the minor segment was 15 bp. The split segments were required to align uniquely, with an alignment score of at least 97%. Based on these split read alignment scores, a structural variant (SV) caller was used to call out large deletions, insertions, inversions, and translocation events. These split read alignment and

SV calling algorithms are integrated into StrandNGS v2.5, which was used to perform this analysis. A threshold of five split reads supporting the SV event was used for calling them out. Further confirmation of the SV event was performed by looking at the event in the StrandNGS elastic genome browser and verifying that the break points across all split reads are unique and that the other partially aligned reads support the same event. For deletion events spanning one or more exons, the CNV analysis would also show significantly lower normalized coverages at these locations, thus providing further evidence of the event.

Confirmation of the detected variants by Sanger sequencing or MLPA: All the pathogenic variants detected in the patient samples were confirmed by Sanger or MLPA. In case of SNVs and indels, primers flanking each variant were designed, and the genomic region encompassing the variant was amplified by PCR. Details of primer sequences and PCR conditions are provided in Appendix 1 (Appendix 1, Appendix 2, Appendix 3, Appendix 4, and Appendix 5 are available as online supplementary information). The PCR products were purified using the Gene Jet PCR Purification Kit (Thermo Fisher), according to the manufacturer's instructions. The purified PCR products were sequenced using both forward and reverse primers (which were used for the PCR amplification) using the BigDye® Terminator v3.1 kit (Life Technologies). The sequencing PCR products were purified and subsequently analyzed by the 3500DX Genetic Analyzer (Life Technologies), as described previously [15]. MLPA was performed with 50 ng of gDNA, according to manufacturer's instructions, using the SALSA MLPA P047-*RB1* kit (MRC-Holland, The Netherlands). Probe amplification products were run on the Genetic Analyzer 3500DX (Life Technologies). MLPA peak plots were visualized and normalized, and the dosage ratios were calculated using the Coffalyser.Net software (MRC-Holland, The Netherlands). A threshold ratio of >1.3 denotes duplication and a ratio of <0.7 denotes deletion.

RESULTS

The mutation spectrum in the patients with Rb: In total, we screened 50 DNA samples of unrelated patients with Rb for mutations in the *RB1* gene using NGS. The demographic profile and clinical characteristics of all the subjects are provided in Appendix 2. In 33 patients, pathogenic or likely pathogenic variants (hereby referred to as mutations) were identified (Table 1 and Figure 1), accounting for 66% (33/50) of all cases (Figure 2A). The spectrum of identified mutations includes 19 SNVs (11 nonsense, three missense, and five splice site variants), eight indels (six deletions, one indel, and one duplication), and six large deletions (single

exon deletion to whole gene deletions; Figure 3). All the SNVs and indels identified by NGS were confirmed by Sanger sequencing, and large deletions were confirmed by MLPA analysis, which implies 100% concordance between the NGS findings and Sanger/MLPA data. We detected 29 unique mutations, of which 12 were novel (Table 1). None of the 12 identified novel mutations in our study were found in the 1,200 control chromosomes. Interestingly, among the 11 nonsense mutations identified in our study, the majority (91%) were substitutions of arginine residue to stop codon due to a C to T transition (Table 1). We also detected two recurrent nonsense mutations: p.Arg455Ter (3X) and p.Arg579Ter (2X; Table 1). We detected three missense mutations (p.Gln702Lys, p.Cys712Arg and p.Trp563Cys), all of which lie in the A/B "pocket" domain of the protein [19,20].

Correlation between laterality and mutation detection rate: To determine whether the mutation detection rate in our screen was correlated with the laterality of the Rb patients, we stratified the patients into three categories, namely bilateral, unilateral, and unknown laterality. Of the 50 patients, 22 were diagnosed with bilateral Rb (BRb), while 20 patients showed a unilateral form of Rb (URb). For eight patients, laterality information was unavailable. In BRb patients, the mutation detection rate was 100% (22/22; Figure 2B). In URb cases, the mutation detection rate was 30% (6/20; Figure 2C) and in unknown cases, mutations were detected in 62.5% (5/8) of patients (Figure 2D). Overall, the mutation detection frequency was 66% (33/50 cases; Figure 2A).

Detection of large deletions in the RB1 gene: Using the CNV analysis, we detected six large deletions in our cohort. The spectrum of deletions ranged from a single exon deletion (one case) to multi-exon (three cases) to whole gene deletions (two cases; Table 1). The deletions identified by NGS in the patient samples (RB6, RB30, RB31, RB32, and RB33) were confirmed by MLPA (Appendix 3). In two of these samples (RB6 and RB31), we were able to detect the exact break point of the identified deletion in the genomic sequence by a split-read alignment analysis (Appendix 4). In patient RB6 with URb, the deletion of exons 8–11 was detected by CNV analysis. Using the split-read alignment of the sequence reads, the 5′ break point could be identified at 2,574 bp upstream (chr13:4893377) of exon 8 and the 3′ break point was mapped 678 bp downstream (chr13:48943418) of exon 11 of the *RB1* gene (c.719–2574_1127+678delinsC; Appendix 4). In patient RB31 with BRb, a partial deletion of 21 bases (chr13:49050959) at the 3′ end of exon 25 and a complete deletion of exons 26 and 27 were detected by CNV analysis. Using a split-read alignment of the sequence reads, the 3′ break point could be identified at 3,849 bp (chr13:49059971)

downstream of 3′ UTR in the *RB1* gene [c.2643_(*1915+3849) del] (Appendix 4). The exact break points of the identified deletions were confirmed by Sanger sequencing (Appendix 4).

Identification of genetic mosaicism in URb cases: Individuals who have URb without an identified heterozygous germline *RB1* mutation are at risk for low-level mosaicism [1]. In our screen, two patients (RB12 and RB15) were found to carry nonsense variants: p.Arg445Ter (c.1333C>T) and p.Arg455Ter (c.1363C>T), respectively. In the RB12 case, the c.1333C>T variant had 21.7% supporting reads (out of 461 reads; Appendix 5) and in RB15, the c.1363C>T variant had 17% supporting reads (out of 909 reads; Appendix 5). When Sanger sequencing was performed, in the electropherogram, the relative peak intensity of the 'T' allele was much weaker than the reference 'C' allele in the specimen DNA samples (Appendix 5). Thus, in these individuals, there could be a possibility of genetic mosaicism in relation to the identified *RB1* mutation.

DISCUSSION

Germline mutations have been reported throughout the *RB1* gene in Rb patients, and only a few of these reported mutations are recurrent. Previously, several Indian studies conducted screening of the *RB1* gene in Rb patients and reported mutation detection rates in the range of 33% to 85% [10-14]. These studies highlight the limitations of the techniques used in these studies because, in principle, 100% of bilateral Rb patients carry germline mutations in the *RB1* gene. To confirm the molecular diagnosis of Rb, several different genetic testing methods have been used traditionally, such as Sanger sequencing, quantitative multiplex PCR, cytogenetic testing, MLPA, and array-Comparative Genomic Hybridization (aCGH) [14,21-23]. Sanger sequencing is used to detect point mutations and indels; when negative, another method (as mentioned above) is used to detect large deletions/duplications/insertions. This sequential mode (reflex) of testing is time consuming and expensive.

Compared to the reflex-testing mode, our current study shows that a NGS-based method is able to screen the complete *RB1* gene and can detect all types of mutations, including large deletions. In our study, among patients affected with BRb (22 cases), the mutation detection rate was 100%. Recently, a NGS-based test was used by Li et al. to screen the entire *RB1* gene to detect all types of *RB1* mutations, such as point mutations, small indels, and large deletions or duplications on a single test platform [24]. Our strategy had notable similarities with that reported by Li et al., including 100% concordance between the NGS output and

TABLE 1. LIST OF MUTATIONS IN THE *RB1* GENE, IDENTIFIED IN PATIENTS WITH RETINOBLASTOMA (RB).

Patient No.	Exon/ Intron	Type of mutation	c.DNA	Amino acid change	Rb presentation	Reference/ Novel
RB1	Ex 1	Indel	c.46_68delGCCGCCGCCGCCGGAACCCCCGGCACC	-	Unilateral	Novel
RB2	Ex 2	Indel	c.178_182delTTATG	p.Leu60SerfsTer48	Bilateral	8651278[b]
RB3	Ex 4	Indel	c.400_407delTTACTAAA	p.Leu134ArgfsTer3	Bilateral	Novel
RB4	Int 5	SS	c.539+1G>A	-	Unilateral	10991691[b]
RB5	Ex 6	Indel	c.540-6_545delTAATAGGATATC	-	Bilateral	Novel
RB6	Ex 8-11	LD	c.719-2574_1127+678delinsC	p.Lys240ArgfsTer3	Unilateral	Novel
RB7	Ex 8	NS	c.751C>T	p.Arg251Ter	Bilateral	20447117[b]
RB8	Int 9	SS	c.940-1G>A	-	Bilateral	Novel
RB9	Ex 11	NS	c.1072C>T	p.Arg358Ter	Bilateral	rs121913301[a]
RB10	Int 12	SS	c.1215+1G>A	-	Bilateral	rs587776783[a]
RB11	Ex 13	Indel	c.1281delA	p.Glu428ArgfsTer29	Unknown	Novel
RB12	Ex 14	NS	c.1333C>T	p.Arg445Ter	Unilateral	rs121913302[a]
RB13	Ex 14	NS	c.1363C>T	p.Arg455Ter	Unilateral	rs121913302[a]
RB14	Ex 14	NS	c.1363C>T	p.Arg455Ter	Unknown	rs121913302[a]
RB15	Ex 14	NS	c.1363C>T	p.Arg455Ter	Bilateral	rs121913302[a]
RB16	Int 15	Indel	c.1421+5_1421+17delTTTTTTTACTTTT	-	Bilateral	Novel
RB17	Ex 17	NS	c.1654C>T	p.Arg552Ter	Bilateral	25742471[b]
RB18	Ex 17	MS	c.1689G>C	p.Trp563Cys	Unknown	15605413[b]
RB19	Int 17	SS	c.1696-3T>G	p.Asp566ArgfsTer45	Bilateral	Novel
RB20	Ex 18	NS	c.1723C>T	p.Gln575Ter	Bilateral	rs587778864[a]
RB21	Ex 18	NS	c.1735C>T	p.Arg579Ter	Bilateral	rs121913305[a]
RB22	Ex 18	NS	c.1735C>T	p.Arg579Ter	Bilateral	rs121913305[a]
RB23	Ex 19	Indel	c.1953T > TA	p.K653fs	Bilateral	Novel
RB24	Ex 19	Indel	c.1959_1960dupA	p.Val654SerFsTer14	Bilateral	15605413[b]
RB25	Ex 20	MS	c.2104C>A	p.Gln702Lys	Bilateral	12541220[b]
RB26	Ex 21	MS	c.2134T>C	p.Cys712Arg	Unilateral	9671401[b], 10486322[b]
RB27	Int 22	SS	c.2325+1G>A	-	Unknown	15884040[b]
RB28	Ex 23	NS	c.2359C>T	p.Arg787Ter	Bilateral	rs137853293[a]
RB29#	Ex 24-27	LD	c.(2479+1_2490-1)_(2787+1_2788-1)del	-	Bilateral	Novel
RB30	Ex 25	LD	c.(2520+1_2521-1)_(2663+1_2664_1)del	-	Bilateral	Novel
RB31	Ex 25_27*	LD	c.2643_(*1915+3849)del	-	Bilateral	Novel

Patient No.	Exon/ Intron	Type of mutation	c.DNA	Amino acid change	Rb presentation	Reference/ Novel
RB32	-	WGD	c.(?_-166)_(*1819_?)del	-	Unknown	cUnknown
RB33	-	WGD	c.(?_-166)_(*1819_?)del	-	Bilateral	cUnknwon

Abbreviations: Indel: small insertion or deletion, LD: Large deletion, MS: Missense, NS: Nonsense, SS: Splice site, WGD: Whole gene deletion, adbSNP database, bPMID, *Partial deletion of exon 25 with complete deletion of exons 26 and 27, c Breakpoints not known, #All LD and WGD except in the RB29 sample were confirmed by MLPA (Multiplex Ligation-dependent Probe Amplification).
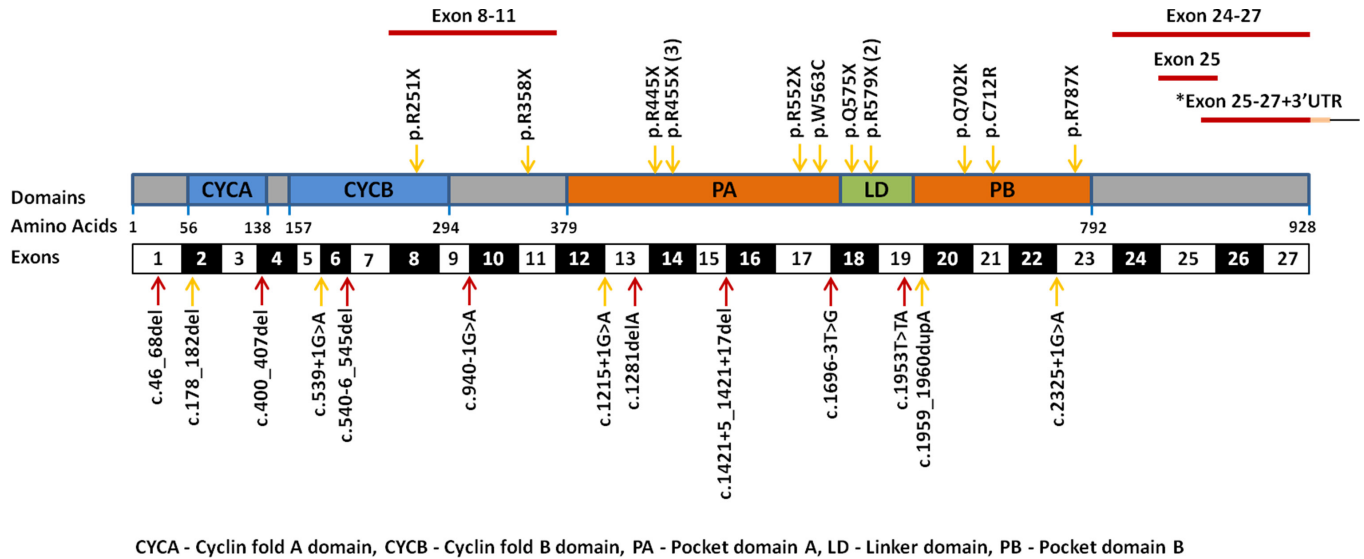
Figure 1. Schematic diagram representing the structural domains of the RB1 protein, along with 27 exons, and representing the localization of the identified mutations in the retinoblastoma (Rb) cohort. The novel mutations identified in our study are indicated by red arrows and known mutations are marked by yellow arrows. Among the 29 unique mutations identified in 33 patients in our Rb cohort, 16 mutations are located in the region encoding for the A/B pocket domain and five mutations are located in the cyclin domain of the RB1 protein. We identified six large deletions in our cohort, but two whole gene deletions are not shown in the representation; the other four deletions are indicated by red bars. Note: *partial deletion of exon 25 with complete deletion of exons 26–27 and 3′UTR (untranslated region).

Sanger confirmation and the detection of low-level mosaic *RB1* mutations using the NGS test [24]. In Indian Rb cohorts, conventional testing was able to detect mutations in the range of 36% to 83% in BRb cases [4]. In a recent study, Devarajan et al. used a NGS-based approach to screen the *RB1* gene in an Indian cohort and reported a detection rate of 85.7% (18/21) in the BRb cases [12]. Interestingly, in another recent study, Grotta et al. used a combined approach of NGS and aCGH and still could detect mutations in only 96.5% (28/29) of the BRb cases through this reflex mode of testing [22]. Overall, it appears that our NGS-based testing has a higher sensitivity than previous studies using both conventional tests and other NGS-based tests [10-14].

In our study, among the URb cases, the mutation detection rate was 30% (6/20). In previous studies with Indian cohorts with a significant number of URb cases, the mutation detection rate was reported in the range of 18% to 23.8% [21,23].

Through a CNV analysis, embedded in our NGS-based approach, we could detect six large deletions in our cohort ranging from a single exon to whole gene deletion. Among the six deletions, four were detected in BRb cases, one in an URb case, and one in a case where laterality was unknown. The overall detection rate of large deletions in our study was 12% (6/50) and in BRb cases, it was 18.2% (4/22), which is similar to the findings previously reported (9.5% to 20.5%)

in other Indian cohorts [14,21,23,25]. Moreover, using the split read alignment of the sequence reads [18,26], we could identify the precise break points in the *RB1* gene in two of six deletions. We could confirm the break points of these two deletions using PCR amplification of the break point regions and Sanger sequencing. The identification of break points in cases with a large deletion by split read alignment allows us to establish a precise Sanger sequencing-based assay that is fast and economical for screening other at-risk family members.

In our study, we identified 11 nonsense mutations. Interestingly, ten of these 11 variants involved a substitution of arginine residue with a stop codon. At the nucleotide level, all mutations were C to T transitions. Previously, it has been reported that in the *RB1* gene, the majority of nonsense mutations occur due to C to T transitions at CpG dinucleotides (CpGs) as a result of the deamination of 5-methylcytosine to thymidine within these CpGs [27]. The occurrence of nonsense mutations at CpGs in the *RB1* gene appears to be determined by several factors, such as the constitutive presence of methylation at cytosines within CpGs, the specific codon within which the cytosine is methylated, and the region of the gene within which that codon resides [27]. In four of the mutated CGA codons (p.Arg251 in exon 8, p.Arg445 and p.Arg455 in exon 14, and p.Arg579 in exon18) of the *RB1* gene, a high frequency of constitutive methylation has been reported [27]. We detected the p.Arg455Ter mutation 3X

and the p.Arg579Ter mutation 2X. These two variants have been previously reported as recurrent mutations in patients affected with Rb [28].

We detected three missense mutations in our cohort and all of these mutations were located in the A/B pocket domain (379–792 residues) of the protein. The A/B pocket domain is essential for the interaction of the RB1 protein with the E2F transcription factor [29]. Previously, Richter et al. reported that 13 of 15 missense mutations identified in their study were located in the A/B pocket domain, thus suggesting that missense mutations occur frequently in this domain of the RB1 protein and highlighting the functional importance of this domain in the protein function [28].

In two cases (RB12 and RB15), the supporting read fractions for the identified variants were much lower (approximately 20%) than the expected ratio of 50%, suggesting the

possibility of mosaicism. The incidence of mosaicism was estimated to be 30% and 6% in sporadic BRb and URb cases, respectively [30]. The use of deep sequencing technology, such as NGS, which has an increased sensitivity, enables us to detect low-level mosaicism in the *RB1* gene. The identification of a mosaic mutation in Rb cases has important clinical implications, as it confirms a genetic diagnosis and alters genetic counseling, surveillance, and disease management measures.

India has the highest number of patients with Rb, accounting for approximately 20% of the global Rb population [4]. The number of new cases is increasing each year, as the population of India is on the rise. As a result, treatment and disease management measures for patients with Rb are causing an increased financial burden on the Indian health care system. In the *RB1* gene, heterogeneous mutations
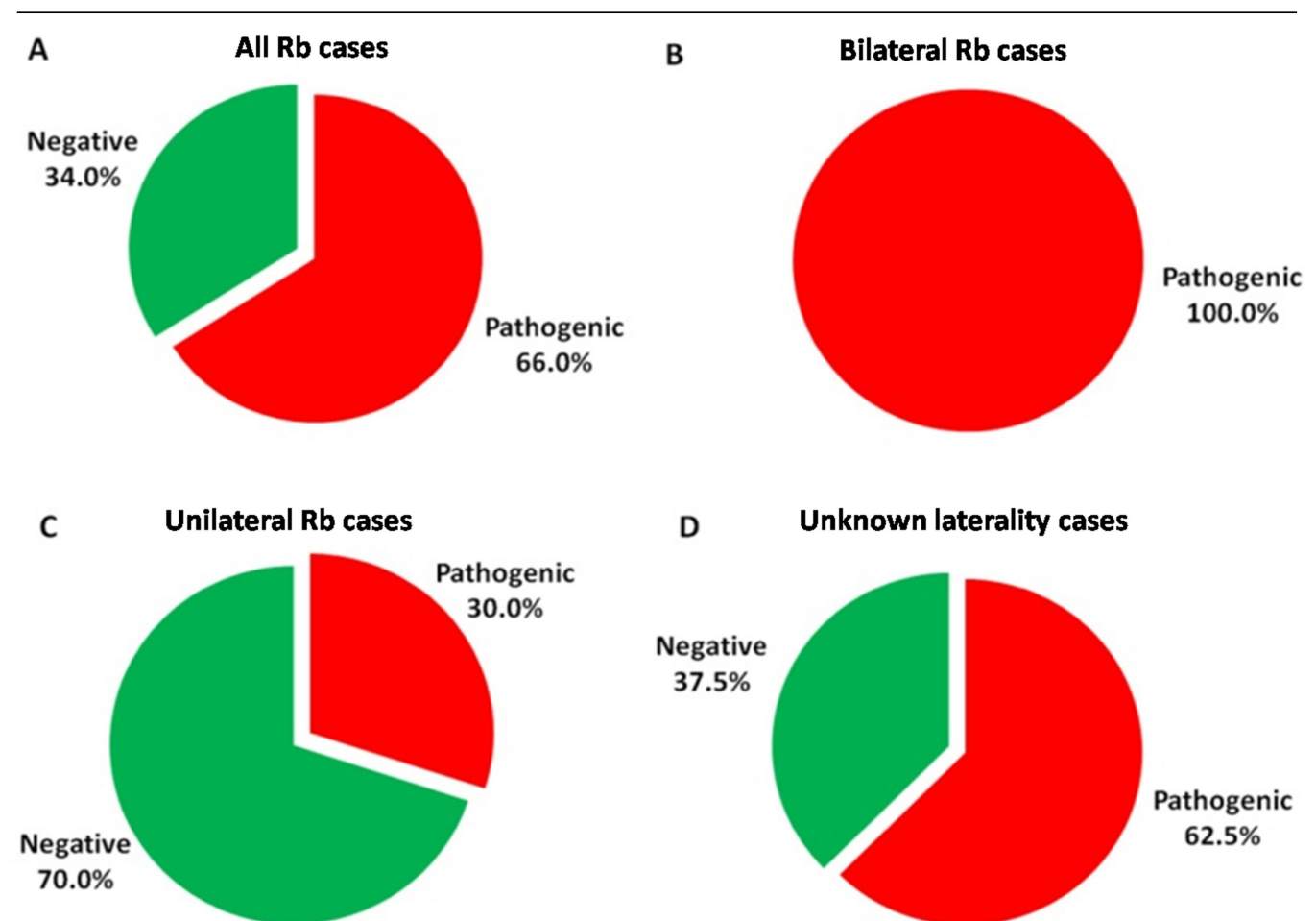

Figure 2. Mutation detection rate in the retinoblastoma (Rb) cohort. We stratified our cohort based on the Rb presentation into three groups, namely bilateral Rb (BRb), unilateral Rb (URb), and unknown literality, to determine whether the mutation detection rate was correlated with Rb presentation. **A**: A pie chart depicting an overall mutation detection rate of 66% (33/50 cases) in all Rb cases screened in our cohort. **B**: In BRb cases, the mutation detection rate was 100% (22/22). **C**: The mutation detection rate was 30% (6/20) in URb cases. **D**: In unknown laterality cases, the mutation detection rate was 62.5% (5/8).

are distributed throughout the entire length of the gene, suggesting that in terms of conventional tests, no single technology will be fully sensitive and efficient; a combination of tests will be necessary for confirmation of a genetic diagnosis, which is time consuming and costly. Our study indicates that NGS-based comprehensive testing of Rb patients will be at least six times more economical than reflex mode testing by Sanger, followed by MLPA for negative cases. In India, there is a pressing need for a cost-effective and comprehensive genetic testing method for the diagnosis and early detection of Rb. In the current study, we report a 100% mutation detection rate in patients with BRb. Our study suggests that a NGS-based approach increases the sensitivity of mutation detection in the *RB1* gene and helps in the confirmation of a genetic diagnosis in patients and at-risk family members compared to conventional tests performed in reflex testing mode. Our finding strongly supports the incorporation of a NGS-based approach for the routine genetic testing of Rb in India, as it is highly sensitive, accurate, fast, and economically feasible.

APPENDIX 1. PCR CONDITIONS AND PRIMER SEQUENCES FOR MUTATIONS IDENTIFIED IN THE RB1 GENE

To access the data, click or select the words "Appendix 1."

APPENDIX 2. THE DEMOGRAPHIC PROFILE AND CLINICAL CHARACTERISTICS OF PATIENTS WITH RETINOBLASTOMA (RB).

To access the data, click or select the words "Appendix 2."

APPENDIX 3. DETECTION OF LARGE DELETION IN THE *RB1* GENE IN THE PATIENT SAMPLES.

To access the data, click or select the words "Appendix 3." Detection of copy number variation (CNV) and confirmation by MLPA (multiplex ligation-dependent probe amplification). Figures on the left panel (A, C, E, G, I and K) represent CNV analysis and on the right panel (B, D, F, H and J) represent MLPA analysis. In the MLPA plot, x-axis represents genomic regions and y-axis represents dosage quotient (DQ). DQ distribution of 0.8–1.2 represent normal copy, 0.4–0.65 represents heterozygous deletion and 1.3–1.65 represents heterozygous duplication. In the sample RB6, the CNV analysis showed a heterozygous deletion of exon 8–11 (**A**), which was confirmed by MLPA analysis (**B**). Similarly, the CNV analysis in the sample RB30, revealed a heterozygous deletion of exon 25 (**C**), which was confirmed by MLPA analysis (**D**). The CNV analysis in the sample RB31, revealed a heterozygous deletion of exon 26–27 and a partial deletion of exon 25 (depicted in dotted circle; **E**), which was confirmed by MLPA analysis (**F**). The CNV analysis
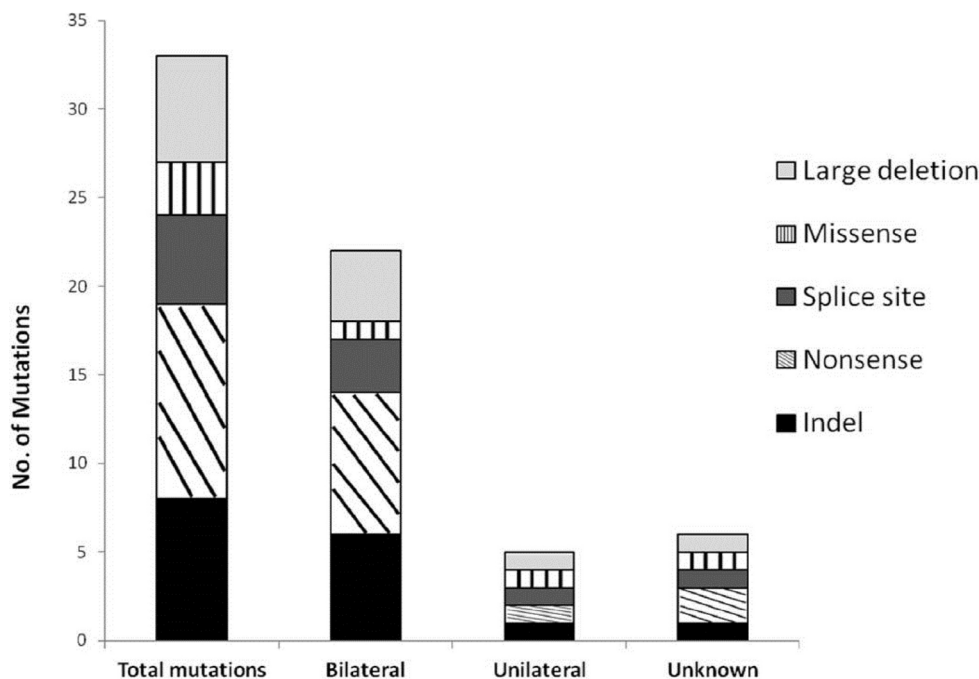

Figure 3. Types of mutations and spectrum in the retinoblastoma (Rb) cohort. The spectrum of mutation types detected in our cohort was missense, nonsense, splices site, indel, and large deletion types. In overall cases, we detected 11 nonsense, three missense, five splice site variants; eight indels; and six large deletions. In bilateral Rb (BRb) cases, eight nonsense, one missense, three splice site variants; six indels; and four large deletions were detected. In unilateral Rb (URb) cases, one nonsense, one missense, and one splice site variants; one indel; and one large deletion were detected, and in unknown literality cases, two nonsense, one missense, and one splice site variant; one indel, and one large deletion were detected.

in the sample RB32 showed heterozygous whole *RB1* gene deletion (**G**), which was confirmed by MLPA analysis (**H**). The CNV analysis in the sample RB33 showed heterozygous whole *RB1* gene deletion (**I**); although the CNV data showed higher heterogeneity, however, this deletion was confirmed by MLPA analysis (**J**). In both cases of whole gene deletion (RB32 and RB33), MLPA analysis revealed that in addition to the whole *RB1* gene deletion, the upstream and downstream genomics regions flanking the *RB1* gene were also deleted. The CNV analysis in the sample RB29, revealed a heterozygous deletion of exon 24–27, this sample also showed higher heterogeneity in the CNV data compared to other samples; however, considering deletion of multiple continuous exons, it is unlikely that it is false positive (**K**), as, additional DNA was unavailable for the sample RB29 therefore MLPA confirmation could not be performed.

APPENDIX 4. SPLIT-READ ALIGNMENT ANALYSIS.

To access the data, click or select the words "Appendix 4." By split read alignment of the sequence reads, we could detect the precise breakpoints in the *RB1* gene in 2 out of 6 deletions. In the sample RB6, for the identified deletion (c.719–2574_1127+678delinsC), the break points were: 2574 bp upstream (chr13:4893377) of exon 8 (5′ break-point) and the 3′ break-point was mapped to 678 bp downstream (chr13:48943418) of exon11 of the *RB1* gene (**A**). The identified deletion in the sample RB6, was confirmed by Sanger sequencing (**B**). In the sample RB31, for the identified deletion [c.2643_(*1915+3849)del], the break points were: 21 bases (chr13:49050959) at 3′ end of exon 25 and complete deletion of exon 26 and exon 27 and the 3′ break-point was mapped to 3849 bp (chr13:49059971) downstream of 3′UTR of the *RB1* gene (**C**). The identified deletion in the sample RB31, was confirmed by Sanger sequencing (**D**).

APPENDIX 5. INDICATION OF GENETIC MOSAICISM IN THE UNILATERAL RB (URB) CASES.

To access the data, click or select the words "Appendix 5." In the sample RB12, NGS data showed that the variant, c.1333C>T, had 21.7% supporting reads (out of 461 reads; **A**). In the Sanger electropherogram, the identified variant, c.1333C>T, was detectable and the relative peak intensity of the variant nucleotide 'T' was weaker as compared to reference nucleotide 'C' (**B**). In sample RB15, NGS data showed that the variant, c.1363C>T, had 17% supporting reads (out of 909 reads; **C**). In the Sanger electropherogram, the identified variant, c.1363C>T, was detectable and the relative

peak intensity of the variant nucleotide 'T' was weaker as compared to reference nucleotide 'C' (**D**).

REFERENCES

1. Lohmann DR, Gallie BL. Retinoblastoma. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Fong CT, Mefford HC, Smith RJH, Stephens K, editors. GeneReviews® [Internet] Seattle (WA): University of Washington; 1993.

2. Retinoblastoma. nion for International Cancer ControlWHO. 2014.

3. Satyanarayana L, Asthana S, Labani SP. Childhood cancer incidence in India: a review of population-based cancer registries. Indian Pediatr 2014; 51:218-20. [PMID: 24736911].

4. Dimaras H. Retinoblastoma genetics in India: From research to implementation. Indian J Ophthalmol 2015; 63:219-26. [PMID: 25971166].

5. Guerin S, Hawkins M, Shamsaldin A, Guibout C, Diallo I, Oberlin O, Brugieres L, de Vathaire F. Treatment-adjusted predisposition to second malignant neoplasms after a solid cancer in childhood: a case-control study. J Clin Oncol 2007; 25:2833-9. [PMID: 17602088].

6. Canturk S, Qaddoumi I, Khetan V, Ma Z, Furmanchuk A, Antoneli CB, Sultan I, Kebudi R, Sharma T, Rodriguez-Galindo C, Abramson DH, Chantada GL. Survival of retinoblastoma in less-developed countries impact of socioeconomic and health-related indicators. Br J Ophthalmol 2010; 94:1432-6. [PMID: 20733021].

7. MacCarthy A, Draper GJ, Steliarova-Foucher E, Kingston JE. Retinoblastoma incidence and survival in European children (1978–1997). Report from the Automated Childhood Cancer Information System project. Eur J Cancer 2006; 42:2092-102. [PMID: 16919773].

8. Retinoblastoma Treatment. (PDQ(R)): Health Professional Version. PDQ Pediatric Treatment Editorial Board: PDQ Cancer Information Summaries [Internet] Bethesda (MD): National Cancer Institute (US); 2002

9. Zajaczek S, Jakubowska A, Kurzawski G, Krzystolik Z, Lubinski J. Age at diagnosis to discriminate those patients for whom constitutional DNA sequencing is appropriate

in sporadic unilateral retinoblastoma. Eur J Cancer 1998; 34:1919-21. [PMID: 10023315].

10. Ali MJ, Parsam VL, Honavar SG, Kannabiran C, Vemuganti GK, Reddy VA. RB1 gene mutations in retinoblastoma and its clinical correlation. Saudi J Ophthalmol 2010; 24:119-23. [PMID: 23960888].

11. . Ata-ur-Rasheed M. Vemuganti G, Honavar S, Ahmed N, Hasnain S, Kannabiran C. Mutational analysis of the RB1 gene in Indian patients with retinoblastoma. Ophthalmic Genet 2002; 23:121-8. [PMID: 12187430].

12. Devarajan B, Prakash L, Kannan TR, Abraham AA, Kim U, Muthukkaruppan V, Vanniarajan A. Targeted next generation sequencing of RB1 gene for the molecular diagnosis of Retinoblastoma. BMC Cancer 2015; 15:320-[PMID: 25928201].

13. Kiran VS, Kannabiran C, Chakravarthi K, Vemuganti GK, Honavar SG. Mutational screening of the RB1 gene in Indian patients with retinoblastoma reveals eight novel and several recurrent mutations. Hum Mutat 2003; 22:339-[PMID: 12955724].

14. Thirumalairaj K, Abraham A, Devarajan B, Gaikwad N, Kim U, Muthukkaruppan V, Vanniarajan A. A stepwise strategy for rapid and cost-effective RB1 screening in Indian retinoblastoma patients. J Hum Genet 2015; 60:547-52. [PMID: 26084579].

15. Mannan AU, Singh J, Lakshmikeshava R, Thota N, Singh S, Sowmya TS, Mishra A, Sinha A, Deshwal S, Soni MR, Chandrasekar A, Ramesh B, Ramamurthy B, Padhi S, Manek P, Ramalingam R, Kapoor S, Ghosh M, Sankaran S, Ghosh A, Veeramachaneni V, Ramamoorthy P, Hariharan R, Subramanian K. Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: implications of embracing a multi-gene panel in molecular diagnosis in India. J Hum Genet 2016; [PMID: 26911350].

16. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med 2008; 10:294-300. [PMID: 18414213].

17. Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, Chan SL, Chan WC, Hui DS, Ng SS, Chan HL, Wong CS, Ma BB, Chan AT, Lai PB, Sun H, Chiu RW, Lo YM. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci USA 2013; 110:18761-8. [PMID: 24191000].

18. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet 2011; 12:363-76. [PMID: 21358748].

19. Chau BN, Pan CW, Wang JY. Separation of anti-proliferation and anti-apoptotic functions of retinoblastoma protein through targeted mutations of its A/B domain. PLoS ONE 2006; 1:e82-[PMID: 17183714].

20. Lee C, Chang JH, Lee HS, Cho Y. Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. Genes Dev 2002; 16:3199-212. [PMID: 12502741].

21. Bamne MN, Ghule PN, Jose J, Banavali SD, Kurkure PA, Amare Kadam PS. Constitutional and somatic RB1 mutation spectrum in nonfamilial unilateral and bilateral retinoblastoma in India. Genet Test 2005; 9:200-11. [PMID: 16225399].

22. Grotta S, D'Elia G, Scavelli R, Genovese S, Surace C, Sirleto P, Cozza R, Romanzo A, De Ioris MA, Valente P, Tomaiuolo AC, Lepri FR, Franchin T, Ciocca L, Russo S, Locatelli F, Angioni A. Advantages of a next generation sequencing targeted approach for the molecular diagnosis of retinoblastoma. BMC Cancer 2015; 15:841-[PMID: 26530098].

23. Parsam VL, Kannabiran C, Honavar S, Vemuganti GK, Ali MJ. A comprehensive, sensitive and economical approach for the detection of mutations in the RB1 gene in retinoblastoma. J Genet 2009; 88:517-27. [PMID: 20090211].

24. Li WL, Buckley J, Sanchez-Lara PA, Maglinte DT, Viduetsky L, Tatarinova TV, Aparicio JG, Kim JW, Au M, Ostrow D, Lee TC, O'Gorman M, Judkins A, Cobrinik D, Triche TJ. A Rapid and Sensitive Next-Generation Sequencing Method to Detect RB1 Mutations Improves Care for Retinoblastoma Patients and Their Families. J Mol Diagn 2016; [PMID: 27155049].

25. Joseph B, Raman R, Uthra S, Jagadeesan M, Ganesh A, Paul PG, Sharma T, Kumaramanickavel G. Genotype-phenotype correlation analysis in retinoblastoma patients from India. Asian Pac J Cancer Prev 2006; 7:619-22. [PMID: 17250439].

26. Zhang ZD, Du J, Lam H, Abyzov A, Urban AE, Snyder M, Gerstein M. Identification of genomic indels and structural variations using split reads. BMC Genomics 2011; 12:375-[PMID: 21787423].

27. Mancini D, Singh S, Ainsworth P, Rodenhiser D. Constitutively methylated CpG dinucleotides as mutation hot spots in the retinoblastoma gene (RB1). Am J Hum Genet 1997; 61:80-7. [PMID: 9245987].

28. Richter S, Vandezande K, Chen N, Zhang K, Sutherland J, Anderson J, Han L, Panton R, Branco P, Gallie B. Sensitive and efficient detection of RB1 gene mutations enhances care for families with retinoblastoma. Am J Hum Genet 2003; 72:253-69. [PMID: 12541220].

29. DiCiommo D, Gallie BL, Bremner R. Retinoblastoma: the disease, gene and protein provide critical leads to understand cancer. Semin Cancer Biol 2000; 10:255-69. [PMID: 10966849].

30. Chen Z, Moran K, Richards-Yutz J, Toorens E, Gerhart D, Ganguly T, Shields CL, Ganguly A. Enhanced sensitivity for detection of low-level germline mosaic RB1 mutations in sporadic retinoblastoma cases using deep semiconductor sequencing. Hum Mutat 2014; 35:384-91. [PMID: 24282159].