Research article

Identification of driver genes with aberrantly alternative splicing events in pediatric patients with retinoblastoma

Zhenlei Yang, Jie Wang, Ruixi Zhu*

Department of Ophthalmology, Heilongjiang Province Hospital, Heilongjiang 150036, China

* Correspondence: E-mail: zhuohuitangd60@163.com.

Abstract: Retinoblastoma (RB) is one of the most common cancer in children. However, the specific mechanism about RB tumorigenesis has not been fully understood. In this study, to comprehensively characterize the splicing alterations in the tumorigenesis of RB, we analyzed the differential alternative splicing events in RB. Specifically, the isoforms of RB1 were downregulated in the RB samples, and a large proportion of differentially expressed genes had multiple differentially expressed transcripts (64%). We identified 1453 genes with differential alternative splicing, among which, SE accounted for the majority, followed by MXE, RI, A3SS, and A5SS. Furthermore, the biological function related to the normal function of eyes, and E2F family TFs were significantly enriched by the genes with differential alternative splicing. Among the genes associated with visual sense, ABCA4 was found to have two mutually exclusive exons, resulting in two isoforms with different functionalities. Notably, DAZAP1 was identified as one of the critical splicing factors in RB, which was potentially involved in E2F and RB pathways. Functionally, differential binding sites in DAZAP1 protein were significantly observed between RB and normal samples. Based on the comprehensive analysis of the differential alternative splicing events and splicing factors, we identified some driver genes with differential alternative splicing and critical splicing factors involved in RB, which would greatly improve our understanding of the alternative splicing process in the tumorigenesis of RB.

Key words: retinoblastoma; transcriptome profiling; alternative splicing; genes

1. Introduction

Retinoblastoma (RB) stands as the most common intraocular cancer in children, where white eye reflex and strabismus are the first signs [1,2]. RB can be categorized as two different types: congenital

retinoblastoma and sporadic retinoblastoma, which are determined by germline and somatic pathogenic mutations on RB1, the most important gene in the pathogenesis of RB [3]. Sporadic (non-heritable) retinoblastoma accounts for approximately half of RB cases, where there is often only one tumor developed in a single eye [4,5]. But for those with germline mutations on RB1, it would often progress to bilateral tumors [6,7]. As the functional retinoblastoma protein (pRB) is essential for chromosomal stability, the loss-of-function in pRB confers unlimited proliferation potential to RB [8,9]. In addition, mutations in RB1 are absent in some RB cases, and amplification of the MYCN oncogene might initiate retinoblastoma [10].

Disease manifestations in RB are often complex, associated with a series of factors, and the splicing pattern of RB1 mRNA could be one of them [11]. The alternative splicing process significantly promotes protein diversity by the spliceosome, composed of small RNAs and proteins. It is now well established that aberrant splicing affect tumor initiation and progression [12]. For example, the next generation sequencing of DNA and RNA in paired tumor and normal tissues have provided important insights into the consequences of somatic mutations on the splicing alterations in cancer [13,14]. Particularly, splicing mutations in NOTCH1 have been reported to result in an abnormal isoform of NOTCH1 in chronic lymphocytic leukemia (CLL) [15]. Moreover, abnormal expression of splicing factors has been frequently observed in several cancers. For instance, splicing factors such as SRSF1, SRSF3, and SRSF6 often act as oncogenic driver genes [16–18], while the tumor suppressing splicing factors include RBM5, RBM6, and RBM10 [19].

It is now well established that widespread alterations in RNA splicing have been cancer cells as compared to the normal cells; however, the splicing alterations in RB have not been systematically analyzed. To comprehensively characterize the splicing alterations in the tumorigenesis of RB, we conducted the differential splicing analysis by comparing the RB with normal tissues, anticipating to uncover some driver genes with aberrant splicing events and oncogenic splicing factors.

2. Materials and methods

2.1. Data acquisition

The fastq files from the discovery and validation datasets were obtained from Sequence Read Archive (SRA) database with accession SRP182935 and SRP133555 by SRA Toolkit, respectively. The reads cleaned by the earlier study [20] were used for the downstream analysis. The pathways and target genes of the transcription factors were obtained from MSigDB.

2.2. Read mapping and gene/transcript quantification

The cleaned reads were aligned to human reference genome hg19 with GENCODE annotation v19 by STAR aligner. The resulting SAM files were then processed by Samtools. Gene/transcript were quantified by R Subread package with featurecounts [21].

2.3. Differential alternative splicing analysis

The differential expression for each one of the alternative splicing events was conducted by rMATS [22] with the default parameters. Specifically, the SAM files were used as input of rMATS.

The false discovery rate (FDR) of the differential analysis at 0.05 was used to determine the differential alternative splicing events.

2.4. Differential expression analysis

The raw count was used by R DESeq2 [23] package to conduct the differential expression analysis. The expression levels were normalized. The size factors and dispersion were estimated by DESeq function. The negative binomial general linear model (GLM) was fitted. Wald statistics and log2 fold change were used to evaluate the statistical significance (FDR < 0.05 and log2 fold change >1 or < −1) as previous study described [24,25].

2.5. Gene set enrichment analysis (GSEA)

Fisher's exact test and Kolmogorov-Smirnov test were used to implement the GSEA. The Fisher's exact test was to test the enrichment of a given gene set in the pathway or target genes of transcription factor (TF). The Kolmogorov-Smirnov test was to use normalized enrichment score (NES) by considering the statistical significance of genes in the pathway or target genes of TFs. These analyses were implemented in R clusterProfiler [26] package.

3. Results

3.1. The landscape of the alternative splicing events in retinoblastoma

To profile the alternative splicing events in retinoblastoma (RB), we analyzed the differentially alternative splicing levels by comparing 7 RB samples with 3 normal retinas using rMATS [22]. As RB1 gene is a well-recognized driver gene for RB oncogenesis, we found that RB1 also had several transcripts, and was downregulated in the tumor samples (Figure 1A). The differential expression analysis revealed that the genes with multiple isoforms accounted for 64% of the differentially expressed genes, suggesting that the alternative splicing events were frequently observed in RB (Figure 1B). With the thresholds of FDR at 0.05 and fold change at 2, we identified 5528 differentially expressed exons and 9925 transcripts with differentially alternative splicing (Figure 1C). Furthermore, the RB and normal retina tissues showed significantly different splicing levels based on Percent Spliced-In (PSI), indicating that the RB samples exhibited significantly different alternative splicing patterns from the normal tissues (Figure 1D).

3.2. The patterns of aberrantly alternative splicing events in RB

The aberrantly alternative splicing events could be classified into five patterns (Figure 2A), including exon skipping (SE), mutually exclusive exons (MXE), intron retention (RI), alternative 5' splicing site (A5SS), and alternative 3' splicing site (A3SS). Totally, we identified 1,453 genes with differential alternative splicing, among which, SE accounted for the majority, followed by MXE, RI, A3SS, and A5SS (Figure 2B). Notably, 104 genes were found to have both MXE and SE. Moreover, we conducted hierarchical clustering analysis of those genes with differential alternative splicing, and found that these genes could differentiate the RB samples from the normal samples (Figure 2C),

indicating that aberrantly alternative splicing events were widespread in the tumorigenesis of RB.
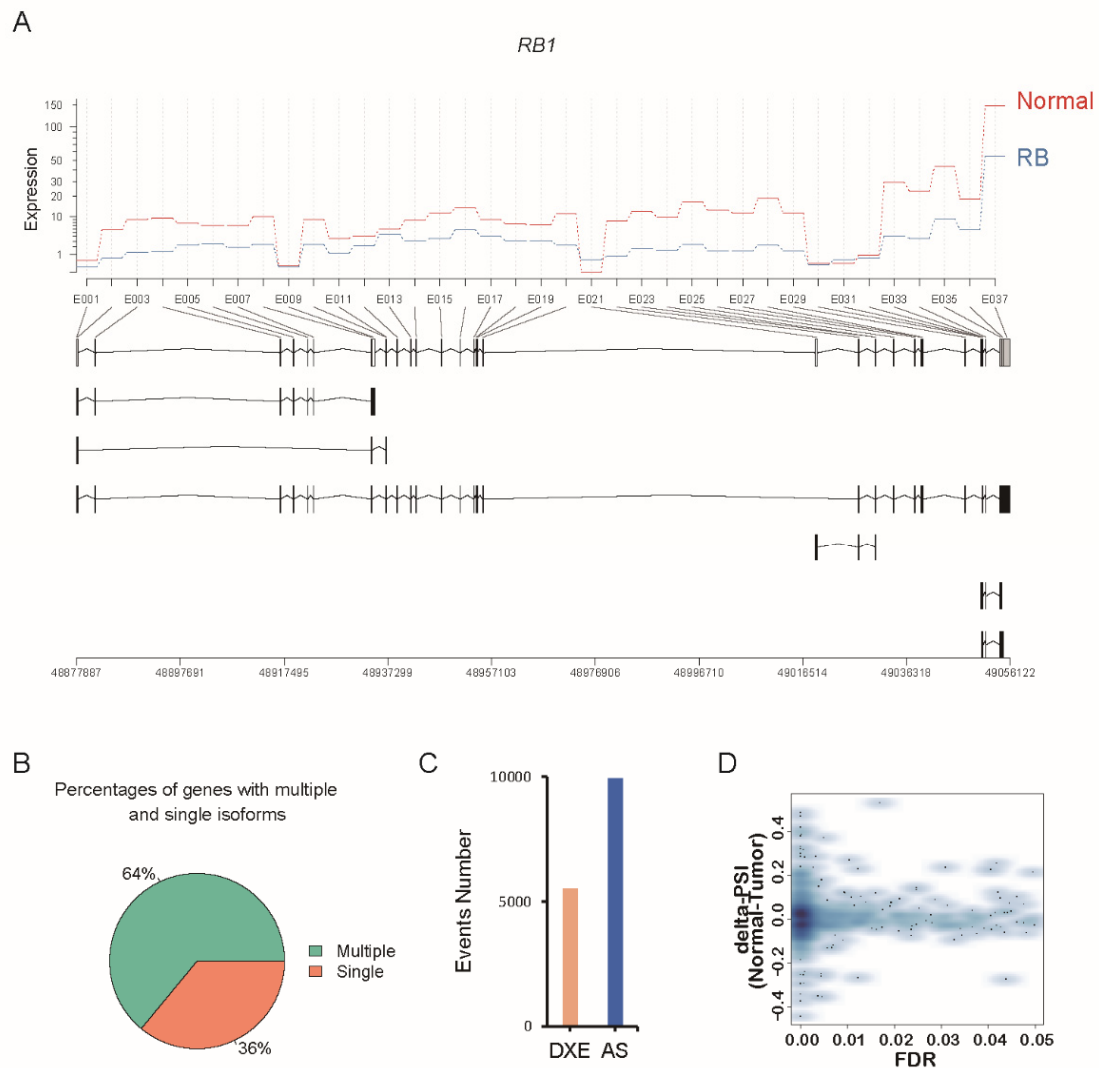

Figure 1. The overview of alternative splicing events in RB. (A) The exon expression levels of RB1 gene in RB and normal tissues. (B) Percentages of genes with multiple and single isoforms. (C) The number of differentially expressed exons (DXE) and alternative splicing (AS) events.

3.3. Functional annotation of the genes with differential alternative splicing in RB

To annotate the genes with differential alternative splicing in RB, we conducted gene set enrichment analysis (GSEA) on those genes against gene sets of gene ontology (GO) and transcription factor (TF) target genes. Specifically, the biological function related to the normal function of eyes, such as detection of visible light, phototransduction of visible light, and detection of light stimulus, were significantly enriched by those genes (Figure 3A). The GSEA against the TF target genes revealed that E2F family transcription factors were significantly enriched (Figure 3B). Moreover, ABCA4, CNGA1, PDE6B, PDE6G, SAG, TULP1, NRL, RLBP1, and NPTN were closely

associated with biological function related to visual sense (Figure 3C). Particularly, MXE was observed in ABCA4, suggesting that the loss-of-function in ABCA4 due to MEX might cause disorder in retinas (Figure 3D).
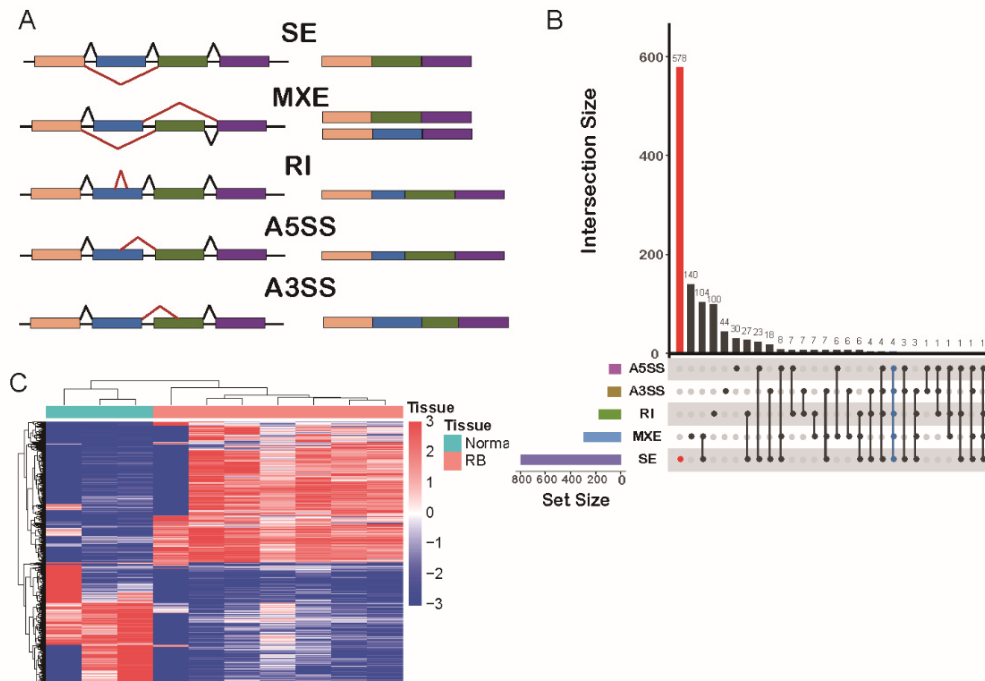

Figure 2. Differential alternative splicing patterns in RB. (A) The schematic diagrams for the five differential alternative splicing patterns. (B) The number of genes across the five alternative splicing patterns. (C) The expression profiles of the genes with differential alternative splicing in RB.

3.4. Identification of critical splicing factor involved in RB

We collected 10 major splicing factors including DAZAP1, ESRP1, ESRP2, MBNL1, QKI, RBM25, RBM4, SF1, TIAL1, and ZRANB2 to identify the upstream regulators of the aberrantly alternative splicing events. Moreover, we also conducted weighted gene co-expression network analysis (WGCNA) on the transcripts of genes with differential alternative splicing. Consequently, nine functional modules were identified, and linked to the 10 splicing factors and the tissues (Figure 4A). Notably, the black module was associated with both tumor tissues and DAZAP1 at the highest statistical significance (Figure 4A), suggesting that genes in the black module might be regulated by DAZAP1 at RNA splicing, and might be implicated in the tumorigenesis of RB. Further analysis of the binding motif revealed that the RB samples had significantly different binding sites from the normal tissues in DAZAP1 (Figure 4B). Consistently, this finding was also validated in an independent RNA-seq dataset (Figure 4C). These results indicated that DAZAP1 might be the critical splicing factor involved in RB.

To further shed light on the function of the black module, we first visualized the network of the

black module, and found that SLC38A5, SNRPB, TXNL4A, UBE2K, and RUVBL1 had relatively high connectivity (Figure 5A). More importantly, the genes within the black module were significantly enriched in RB pathway and E2F pathway by GSEA (adjusted P-value < 0.05, Figure 4B), suggesting that RB and E2F pathway played key roles in the tumorigenesis of RB.
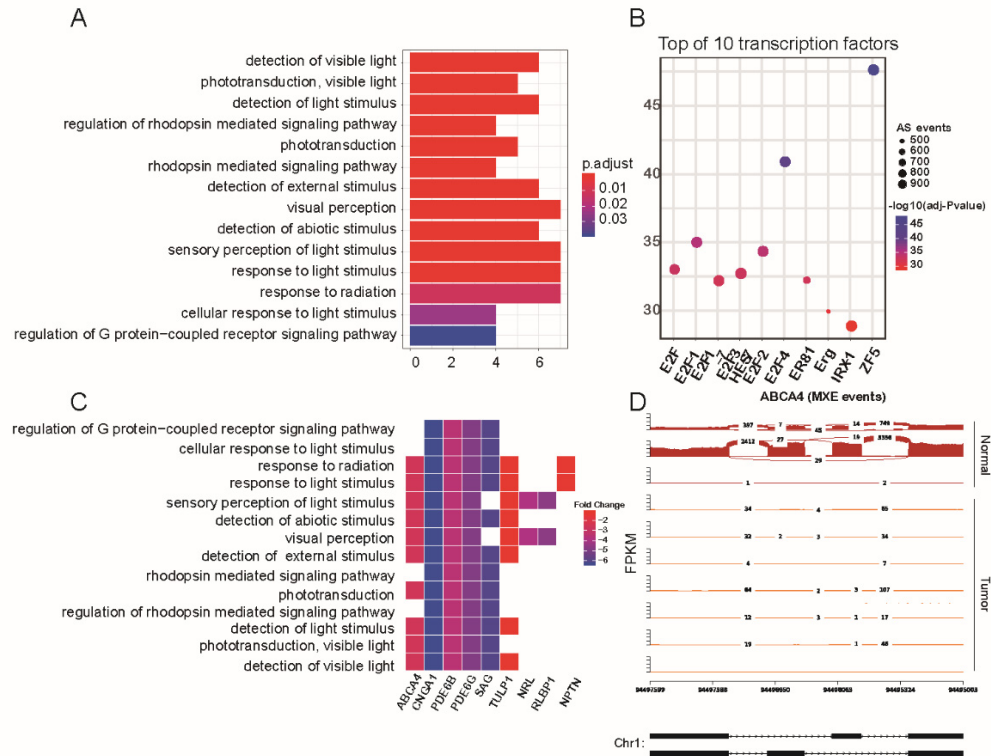

Figure 3. The gene ontology (GO) terms and transcription factors (TFs) potentially involved in RB. (A) The GO terms enriched by the genes with differential alternative splicing. (B) The TFs enriched by their target genes with differential alternative splicing. (C) The genes associated with the enriched GO terms. (D) The differential MXE splicing events in ABCA4 between RB and normal tissues.

4. Discussions

Retinoblastoma is one of the most common cancer in children. Despite of advances in the pediatric medicine, the mortality of retinoblastoma still remains relatively high in developing countries, partly due to the likely delayed diagnosis. In this study, to comprehensively characterize the splicing alterations in the tumorigenesis of RB, we analyzed the differential alternative splicing events in RB. Specifically, the isoforms of RB1 were downregulated in the RB samples, and a large proportion of differentially expressed genes had multiple differentially expressed transcripts (64%). Both germline and somatic splicing variants in RB1 were recurrently observed in RB samples [27,28]. Moreover, we identified 1453 genes with differential alternative splicing, among which, SE accounted for the majority, followed by MXE, RI, A3SS, and A5SS. Consistently, SE has been found in RB1 resulted from splicing mutations [29]. The A5SS and A3SS in RB1 were also reported in RB [30,31]. However, for other genes involved in RB tumorigenesis have not been reported.
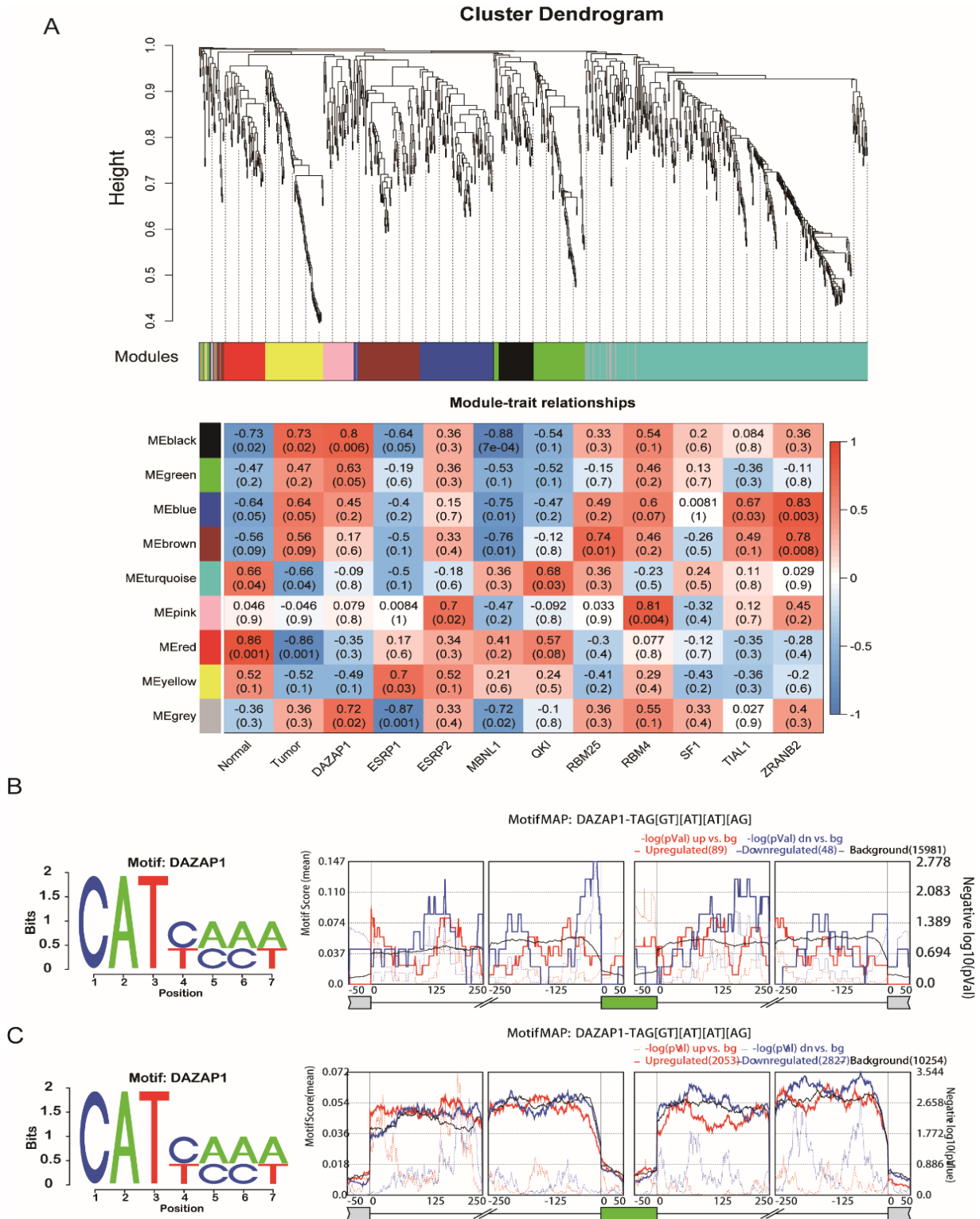
Figure 4. The association between the splicing factors with the functional modules. (A) The functional modules identified by WGCNA algorithm based on the genes with differential alternative splicing and their correlation with the splicing factors. The binding motif within DAZAP1 and its differential binding peaks between RB and normal tissues using the discovery (B) and validation dataset (C).
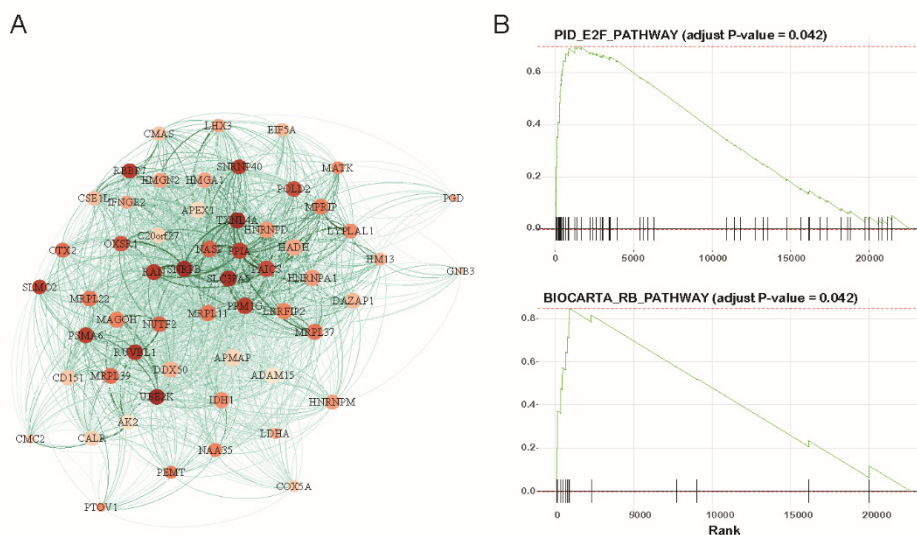
Figure 5. The functional characterization of black module. (A) The network visualization of the genes within the black module. (B) The candidate pathways that the genes within the black module might participate in.

Furthermore, the gene set enrichment analysis (GSEA) efficiently identified key GO terms and TFs enriched by those genes with differential alternative splicing. Specifically, the biological function related to the normal function of eyes, such as detection of visible light, phototransduction of visible light, and detection of light stimulus, were identified. These biological function were highly associated with the loss of normal function in retina and vision disorders [32]. Moreover, the E2F family TFs were identified by GSEA. As pRB could restrain cell proliferation by binding E2F TFs, loss-of-function in RB1 might result in hyperactive in E2F family TFs [33]. Among the genes associated with visual sense, two isoforms of ABCA4 had two mutually exclusive exons. Particularly, ABCA4 has been reported to be implicated in eye diseases [34,35]. In addition, we also investigated the potential splicing factors involved in RB and associated them with functional modules. Notably, the black module, which was potentially involved in E2F and RB pathways, might be regulated by DAZAP1 at the highest statistical significance, and DAZAP1 was found to participate in clear cell renal cell carcinoma (CCRCC), pancreatic cancer, and acute lymphoblastic leukemia (ALL) [36–38]. Functionally, DAZAP1 has been reported to integrate splicing regulation into MEK/Erk-induced cell proliferation and migration [39]. In accordance with this, we also found that RB samples had significantly different binding sites from the normal tissues in DAZAP1 in both the discovery and validation datasets.

Based on the comprehensive analysis of the differential alternative splicing events and splicing factors, we identified some driver genes with differential alternative splicing and critical splicing factors involved in RB, which would greatly improve our understanding of the alternative splicing process in the tumorigenesis of RB.

Conflict of interests

The authors declare there is no conflict of interest.

Data availability statement

The fastq files were obtained from Sequence Read Archive (SRA) database with accession SRP182935 and SRP133555 by SRA Toolkit.

References

1. J. Reinhard, N. Wagner, M. M. Kramer, M. Jarocki, S. C. Joachim, H. B. Dick, et al., Expression Changes and Impact of the Extracellular Matrix on Etoposide Resistant Human Retinoblastoma Cell Lines, *Int. J. Mol. Sci.*, **21** (2020), 4322.

2. I. Aerts, L. Lumbroso-Le Rouic, M. Gauthier-Villars, H. Brisse, F. Doz, [Retinoblastoma update], *Arch Pediatr.*, **23** (2015), 112–116.

3. F. Salviat, M. Gauthier-Villars, M. Carton, N. Cassoux, L. Lumbroso-Le Rouic, C. Dehainault, et al., Association Between Genotype and Phenotype in Consecutive Unrelated Individuals With Retinoblastoma, *JAMA Ophthalmol.*, **138** (2020), 843–850.

4. P. R. Mendoza, H. E. Grossniklaus, The Biology of Retinoblastoma, in *Progress in Molecular Biology and Translational Science,* 2015, 503–516.

5. D. Lohmann, Retinoblastoma, *Adv. Exp. Med. Biol.*, **685** (2010), 220–227.

6. M. Mehyar, M. Mosallam, A. Tbakhi, A. Saab, I. Sultan, R. Deebajah, et al., Impact of RB1 gene mutation type in retinoblastoma patients on clinical presentation and management outcome, *Hematol. Oncol. Stem Cell Ther.*, **13** (2020), 152–159.

7. T. T. Kivela, T. Hadjistilianou, Neonatal Retinoblastoma, *Asia. Pac. J. Oncol. Nurs.*, **4** (2017), 197–204.

8. P. Indovina, F. Pentimalli, N. Casini, I. Vocca, A. Giordano, RB1 dual role in proliferation and apoptosis: cell fate control and implications for cancer therapy, *Oncotarget,* **6** (2015), 17873–17890.

9. L. Zheng, W. H. Lee, Retinoblastoma tumor suppressor and genome stability, *Adv. Cancer Res.*, **85** (2002), 13–50.

10. D. E. Rushlow, B. M. Mol, J. Y. Kennett, S. Yee, S. Pajovic, B. L. Theriault, et al., Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies, *Lancet Oncol.*, **14** (2013), 327–334.

11. C. Rodriguez-Martin, F. Cidre, A. Fernandez-Teijeiro, G. Gomez-Mariano, L. de la Vega, P. Ramos, et al., Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene, *J. Hum.Genet.*, 61 (2016), 463–466.

12. S. C. W. Lee, O. Abdel-Wahab, Therapeutic targeting of splicing in cancer, *Nat. Med.*, **22** (2016), 976–986.

13. F. Supek, B. Minana, J. Valcarcel, T. Gabaldon, B. Lehner, Synonymous mutations frequently act as driver mutations in human cancers, *Cell*, **156** (2014), 1324–1335.

14. H. Jung, D. Lee, J. Lee, D. Park, Y. J. Kim, W. Y. Park, et al., Intron retention is a widespread mechanism of tumor-suppressor inactivation, *Nat. Genet.*, **47** (2015), 1242–1248.

15. V. Spina, D. Rossi, Overview of non-coding mutations in chronic lymphocytic leukemia, *Mol. Oncol.*, **13** (2019), 99–106.

16. R. Karni, E. de Stanchina, S. W. Lowe, R. Sinha, D. Mu, A. R. Krainer, The gene encoding the splicing factor SF2/ASF is a proto-oncogene, *Nat. Struct. Mol. Biol.*, **14** (2007), 185–193.

17. M. Cohen-Eliav, R. Golan-Gerstl, Z. Siegfried, C. L. Andersen, K. Thorsen, T. F. Orntoft, et al., The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers, *J. Pathol.*, **229** (2013), 630–639.

18. R. Jia, C. Li, J. P. McCoy, C. X. Deng, Z. M. Zheng, SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance, *Int. J. Biol. Sci.*, **6** (2010), 806–826.

19. E. G. Bechara, E. Sebestyen, I. Bernardis, E. Eyras, J. Valcarcel, RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation, *Mol. Cell*, **52** (2013), 720–733.

20. S. Rajasekaran, L. D. Nagarajha Selvan, K. Dotts, R. Kumar, P. Rishi, V. Khetan, et al., Non-coding and Coding Transcriptional Profiles Are Significantly Altered in Pediatric Retinoblastoma Tumors, *Front. Oncol.*, **9** (2019), 221.

21. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, **30** (2014), 923–930.

22. S. Shen, J. W. Park, Z. X. Lu, L. Lin, M. D. Henry, Y. N. Wu, et al., rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data, *Proc. Nat. Acad. Sci. U. S. A.*, **111** (2014), E5593–5601.

23. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.*, **15** (2014), 550.

24. C. Gu, X. Shi, Z. Huang, J. Chen, J. Yang, J. Shi, et al., A comprehensive study of construction and analysis of competitive endogenous RNA networks in lung adenocarcinoma, *Biochim. Biophys. Acta Proteins Proteomics,* **1868** (2020), 140444.

25. X. Shi, T. Huang, J. Wang, Y. Liang, C. Gu, Y. Xu, et al., Next-generation sequencing identifies novel genes with rare variants in total anomalous pulmonary venous connection, *EBioMedicine,* **38** (2018), 217–227.

26. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS,* **16** (2012), 284–287.

27. V. Imperatore, A. M. Pinto, E. Gelli, E. Trevisson, V. Morbidoni, E. Frullanti, et al., Parent-of-origin effect of hypomorphic pathogenic variants and somatic mosaicism impact on phenotypic expression of retinoblastoma, *Eur. J. Hum. Genet.*, **26** (2018), 1026–1037.

28. C. J. Dommering, B. M. Mol, A. C. Moll, M. Burton, J. Cloos, J. C. Dorsman, et al., RB1 mutation spectrum in a comprehensive nationwide cohort of retinoblastoma patients, *J. Med. Genet.*, **51** (2014), 366–374.

29. J. Alonso, P. Garcia-Miguel, J. Abelairas, M. Mendiola, E. Sarret, M. T. Vendrell, et al., Spectrum of germline RB1 gene mutations in Spanish retinoblastoma patients: Phenotypic and molecular epidemiological implications, *Hum. Mutat.,* **17** (2001), 412–422.

30. R. Condorelli, L. Spring, J. O'Shaughnessy, L. Lacroix, C. Bailleux, V. Scott, et al., Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer, *Ann. Oncol.,* **29** (2018), 640–645.

31. F. Sanchez-Sanchez, M. Kruetzfeldt, C. Najera, S. Mittnacht, A novel constitutional mutation affecting splicing of retinoblastoma tumor suppressor gene intron 23 causes partial loss of pRB activity, *Hum. Mutat.,* **25** (2005), 223.

32. H. Dimaras, T. W. Corson, Retinoblastoma, the visible CNS tumor: A review, *J. Neurosci. Res.,* **97** (2019), 29–44.

33. H. Sun, Y. Wang, M. Chinnam, X. Zhang, S. W. Hayward, B. A. Foster, et al., E2f binding-deficient Rb1 protein suppresses prostate tumor progression in vivo, *Proc. Nat. Acad. Sci. U. S. A.,* **108** (2011), 704–709.

34. F. P. M. Cremers, W. Lee, R. W. J. Collin, R. Allikmets, Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations, *Prog. Retinal Eye Res.*, 2020, 100861, forthcoming.

35. C. P. Pang, D. S. Lam, Differential occurrence of mutations causative of eye diseases in the Chinese population, *Hum. Mutat.*, **19** (2002), 189–208.

36. T. Chen, W. Zheng, J. Chen, S. Lin, Z. Zou, X. Li, Z. Tan, Systematic analysis of survival-associated alternative splicing signatures in clear cell renal cell carcinoma, *J. Cell. Biochem.*, **121** (2020), 4074–4084.

37. M. Yu, W. Hong, S. Ruan, R. Guan, L. Tu, B. Huang, et al., Genome-Wide Profiling of Prognostic Alternative Splicing Pattern in Pancreatic Cancer, *Front. Oncol.*, **9** (2019), 773.

38. V. Prima, S. P. Hunger, Cooperative transformation by MEF2D/DAZAP1 and DAZAP1/MEF2D fusion proteins generated by the variant t(1;19) in acute lymphoblastic leukemia, *Leukemia*, **21** (2007), 2470–2475.

39. R. Choudhury, S. G. Roy, Y. S. Tsai, A. Tripathy, L. M. Graves, Z. Wang, The splicing activator DAZAP1 integrates splicing control into MEK/Erk-regulated cell proliferation and migration, *Nat. Commun.*, **5** (2014), 3078.