



# Long-range alternative splicing contributes to neoantigen specificity in glioblastoma

Mingjun Ji<sup>1</sup>, Qing Yu<sup>1</sup>, Xin-Zhuang Yang<sup>2</sup>, Xianhong Yu<sup>3</sup>, Jiabin Wang<sup>1</sup>, Chunfu Xiao<sup>1</sup>, Ni A. An<sup>4</sup>, Chuanhui Han<sup>5</sup>,  
Chuan-Yun Li <sup>1,6,7,\*</sup>, Wanqiu Ding <sup>1,8,\*</sup>

<sup>1</sup>State Key Laboratory of Protein and Plant Gene Research, Laboratory of Bioinformatics and Genomic Medicine, Institute of Molecular Medicine, College of Future Technology, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China

<sup>2</sup>Center for Bioinformatics, National Infrastructures for Translational Medicine, Institute of Clinical Medicine and Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 9 Dongdan Santiao, Dongcheng District, Beijing 100730, China

<sup>3</sup>Academic Department, Shanghai MobiDrop Co., Ltd., Room 308, Building 1, No. 351 Guoshoujing Road, Shanghai Free Trade Pilot Zone, Shanghai 200000, China

<sup>4</sup>Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, China

<sup>5</sup>School of Basic Medical Sciences, Peking University, No. 38 Xueyuan Road, Haidian District, Beijing 100191, China

<sup>6</sup>Chinese Institute for Brain Research, No. 26 Science Park Road, Changping District, Beijing 102206, China

<sup>7</sup>Southwest United Graduate School, 121 Dajie, Wuhua District, Kunming 650092, China

<sup>8</sup>Bioinformatics Core Facility, Institute of Molecular Medicine, College of Future Technology, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China

\*Corresponding authors. Chuan-Yun Li, College of Future Technology, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China.

E-mail: [chuanyunli@pku.edu.cn](mailto:chuanyunli@pku.edu.cn); [wanqiu.ding@pku.edu.cn](mailto:wanqiu.ding@pku.edu.cn)

E-mail: [dingwq@pku.edu.cn](mailto:dingwq@pku.edu.cn)

## Abstract

Recent advances in neoantigen research have accelerated the development of immunotherapies for cancers, such as glioblastoma (GBM). Neoantigens resulting from genomic mutations and dysregulated alternative splicing have been studied in GBM. However, these studies have primarily focused on annotated alternatively-spliced transcripts, leaving non-annotated transcripts largely unexplored. Circular ribonucleic acids (circRNAs), abnormally regulated in tumors, are correlated with the presence of non-annotated linear transcripts with exon skipping events. But the extent to which these linear transcripts truly exist and their functions in cancer immunotherapies remain unknown. Here, we found the ubiquitous co-occurrence of circRNA biogenesis and alternative splicing across various tumor types, resulting in large amounts of long-range alternatively-spliced transcripts (LRs). By comparing tumor and healthy tissues, we identified tumor-specific LRs more abundant in GBM than in normal tissues and other tumor types. This may be attributable to the upregulation of the protein quaking in GBM, which is reported to promote circRNA biogenesis. In total, we identified 1057 specific and recurrent LRs in GBM. Through in silico translation prediction and MS-based immunopeptidome analysis, 16 major histocompatibility complex class I-associated peptides were identified as potential immunotherapy targets in GBM. This study revealed long-range alternatively-spliced transcripts specifically upregulated in GBM may serve as recurrent, immunogenic tumor-specific antigens.

**Keywords:** circRNA; alternative splicing; neoantigen

## Introduction

Immune checkpoint inhibitors (ICIs) improve survival in many cancer types, such as melanoma and hepatocellular carcinoma, but are not equally effective in all types of solid tumors [1]. In particular, cancers with a high tumor mutational burden (TMB) benefit more from ICIs than other cancer types. Therefore, combinations of different treatment strategies are needed for cancer types with low TMB, such as glioblastoma (GBM) [2]. Cancer immunotherapy based on neoantigens that are recognized by tumor-reactive T cells has attracted increasing attention in recent years [3, 4]. Tumor-specific neoantigens are exclusively present in the major histocompatibility complex (MHC) molecules on the surfaces of cancer cells and are absent in normal tissues; thus, they are considered as safe and potent targets for T cell-based immunotherapies [5].

The majority of well-studied neoantigens are derived from genomic mutations, particularly single-nucleotide variants (SNVs) in coding regions [5–7]. Recent studies have suggested that tumor-specific fusion genes, transposable elements, RNA editing, and alternatively-spliced isoforms could also generate neoepitopes that elicit a T-cell response [8–11]. Tumors consistently exhibit more alternative splicing events than healthy tissues, and more neoepitopes in tumors are derived from alternative splicing than from other somatic SNV events, indicating that alternative splicing events offer an increased number of targets compared with SNV events [11, 12]. Additionally, many neoepitopes are derived from noncoding regions and non-canonical open reading frames [13–15]. However, the role of non-annotated, alternatively-spliced transcripts in neoantigen-based cancer immunotherapies remains to be elucidated.

Received: May 27, 2024. Revised: August 14, 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Circular ribonucleic acid (circRNAs) are a class of covalently closed, single-stranded RNAs, typically generated via back-splicing reactions [16]. Back-splicing occurs when the spliceosome joins a downstream 5'-splice site to an upstream 3'-splice site and can be viewed as a type of alternative splicing [17]. In recent years, diverse functions for circRNAs have been reported, including sponging microRNAs (miRNAs), interacting with RNA-binding proteins (RBPs), producing small peptides and competing with linear splicing [18]. Recently, immunopeptidomics analyses revealed that circRNA-derived peptides can be presented by MHC [19]. It has been reported that circRNAs are correlated with the presence of linear transcripts with exon skipping events, but little attention has been given to these alternatively-spliced linear transcripts [20, 21]. Therefore, whether these transcripts coexist with circRNAs and to what extent this phenomenon occurs remain largely unknown. In addition, these circRNA-dependent alternatively-spliced linear transcripts may participate in neoantigen generation. Whether such transcripts contribute to proteome diversity in tumors and thus could become potent sources of neoantigens in cancer immunotherapies remains to be further explored.

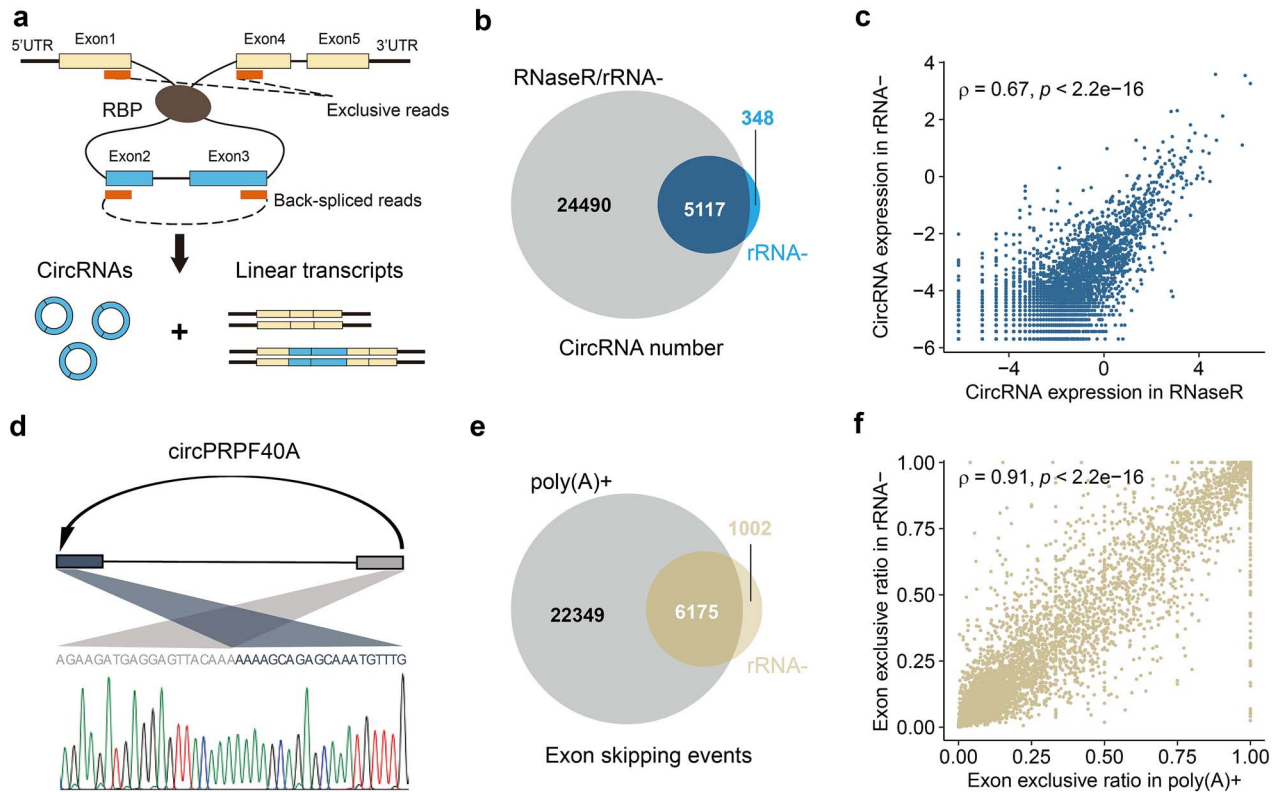
Here, we focused on circRNAs and alternatively-spliced linear transcripts with exon skipping events, and investigated their potential applications in cancer immunotherapy. We revealed that circRNAs are correlated with corresponding long-range alternatively-spliced linear transcripts (LRs) in various tumor types. We further found that there were more LRs in GBM than in normal tissues and other tumor types and identified >1000 transcripts with neoantigen-encoding potential. Additionally, an analysis of mass spectrometry (MS)-based immunopeptidomics data identified 16 MHC class I-associated peptides derived from GBM-specific LRs, indicating that LRs provide a potential source of recurrent, immunogenic tumor-specific antigens in GBM.

## Results

### Accurate and quantitative profiling of both circular and linear transcripts

The biogenesis of circRNAs can be regulated by *cis*-elements, such as intronic complementary elements (ICSs), and *trans*-factors, such as RBPs. In the process of circRNA formation, complementary paired RNA duplexes generated by ICS may bring distant splice sites closer together, potentially facilitating the formation of both circRNAs and linear isoforms with long-range splices (Fig. 1a). Previous case studies have suggested that linear RNAs with exon skipping events may be involved in circRNA biogenesis. However, it remains unknown whether this mechanism is ubiquitous. To explore the correlation between circRNAs and alternative splicing on a genome-wide scale, we first developed a computational pipeline to identify and quantify circRNAs and corresponding alternatively-spliced linear transcripts using paired-end, ribosomal RNA (rRNA)-depleted RNA sequencing (RNA-seq) data (Supplementary Fig. S1 and Supplementary Methods). CircRNAs are identified by considering the mapping results of both ends for paired reads. Not only one end supports the back-spliced junction, but also the other end is required to be mapped to the same circularized exons or identified as a consistent candidate back-spliced junction. Additionally, considering the challenges of accurately mapping long-range spliced junction reads, a junction database of all possible combinations of annotated exons was constructed as the junction reference. While RNase R-treated RNA-seq and poly(A)-positive RNA-seq are used to identify circRNAs and mature linear transcripts, respectively, rRNA-depleted

RNA-seq represents a unique means to simultaneously profile the circular and linear forms of transcripts expressed in the same sample. We first applied the pipeline to in-house rRNA-depleted RNA-seq data from macaque cerebellar tissue [22], identified and quantified both circular and linear events. To evaluate the efficiency and specificity of rRNA-depleted RNA-seq in profiling both transcript species, we then performed RNase R-treated RNA-seq and poly(A)-positive RNA-seq for the same cerebellar sample. The majority (93.63%) of high-confidence circRNAs (back-spliced reads  $\geq 5$ ) identified in rRNA-depleted samples were also detectable in the RNase R-treated samples (Fig. 1b), and the expression of circRNAs (quantified as RPM, back-spliced reads per million mapped reads) in these two systems showed a significantly positive correlation (Fig. 1c). Furthermore, 15 circRNA candidates were randomly selected for validation by Sanger sequencing, and the expression of 14 of these candidates was successfully verified (Fig. 1d and Supplementary Fig. S2). Regarding alternative splicing, the exon-skipping events identified by poly(A)-positive RNA-seq of the same tissue were highly consistent with those identified by rRNA-depleted RNA-seq (86.04% of the events shared, Spearman correlation coefficient of the exon exclusive ratio is 0.91, Fig. 1e and (f)). Additionally, we compared the performance of our computational pipeline with existing methods on both simulated and real data (Supplementary Methods). We produced two simulated datasets based on CIRI-simulator tool [23]. One is based on circRNA list only, and the other is the mixed data (containing linear transcripts as background) to simulate complex transcriptomic environment. Five commonly used, well-performing tools were selected for comparisons: CIRI2, CIRIquant, find\_circ, CIRCexplorer3, and CircSplice [24–28]. From the results, we found that our pipeline shows the highest precision (99.84% for circRNA only datasets and 99.71% for the mixed datasets) and the lowest number of false positives (Supplementary Fig. S3). The recall, and F1 score of our pipeline, also indicates good performance, second only to CIRI tool suite. For real data, the recent benchmark research compared 16 tools in deep sequencing data of three cell lines (HLF, SW480, and NCI-H23), and reported the validation results of 1516 circRNAs selected from the 16 tools [29]. For comparisons, we first identified circRNAs using our pipeline based on the deep sequencing data of the three cell lines, and further assessed each method's performance from three aspects: (i) the sensitivity calculated based on 957 true-positive circRNAs, validated by all three validation methods (quantitative PCR [qPCR], RNaseR treatment, and amplicon sequencing). (ii) The precision calculated based on the circRNAs overlapped with 1516 validated circRNAs. (iii) The proportion of circRNA candidates that were also detected by other tools. From the results, we found that our pipeline has comparable sensitivity to several well-performing tools (Supplementary Fig. S4a). Of note, this sensitivity is based on a biased set of circRNAs selected from the 16 tools, excluding the tool presented in this paper. The overlap of these circRNAs is not a representative random sample of all circRNAs, which may lead to an underestimation of our pipeline's performance. When we used circRNAs that passed at least one validation method as the positive list to calculate precision, our pipeline ranked in the top 5 for performance (Supplementary Fig. S4b). Besides, nearly 99% circRNAs identified by our pipeline were also supported by at least two additional tools, which further indicate the reliability of the circRNA list (Supplementary Fig. S4c). Taken together, these observations indicate that the identification results by our pipeline provides an appropriate basis for studying the quantitative relationship between circular and linear transcripts encoded by the same gene.



**Figure 1.** Accurate quantification of circRNAs and exon-skipping events in rRNA-depleted RNA-seq. (a) A schematic model of RNA back-splicing, in which a circRNA arises along with an alternatively-spliced linear transcript that skips the circularized exons, and the presumed regulation of these two events by RBP. (b) Comparison of circRNAs identified via rRNA-depleted RNA-seq and those detected via rRNA-/RNase R-treated RNA-seq in the same sample. (c) The expression levels of circRNAs estimated from rRNA-depleted and rRNA-/RNase R-treated RNA-seq (in logRPM) were compared in a scatter plot. Spearman correlation coefficient was calculated. (d) Sanger sequencing validation for circRNA (circPRPF40A) is shown as an example. (e) Comparison of exon-skipping events identified by rRNA-depleted RNA-seq and those detected in poly(A)-positive RNA-seq assays using the same sample. (f) The exon exclusive ratios of exon-skipping events estimated from rRNA-depleted RNA-seq and poly(A)-positive RNA-seq were compared in a scatter plot.

## The co-occurrence of circular ribonucleic acids and long-range alternatively-spliced transcripts

To further facilitate the quantitative comparison between circRNA biogenesis and long-range exon-skipping events, we defined the ‘circularized ratio’, as the proportion of back-spliced circRNA reads relative to the sum of local splicing reads, to measure the capacity of an mRNA to generate circRNAs; further, we defined the ‘exclusive ratio’ as the proportion of exclusive reads relative to the sum of local splicing reads to measure the frequency of long-range alternatively-spliced transcript generated by a particular mRNA (Fig. 1a, Supplementary Methods). Based on this, we collected rRNA-depleted RNA-seq data from five tumor cell lines, including A549, IMR-32, MDA-MB-468, RD, and RH4 [30–33], and used the established pipeline to investigate the co-occurrence of circRNAs and corresponding alternatively-spliced linear transcripts in tumors. As expected, the abundance of exons capable of forming circRNAs in linear transcripts was significantly negatively correlated with the circularized ratio of these exons across various tumor types (Fig. 2a). Considering that the expression levels of exons in linear transcripts may be influenced by the overall transcription activity of the host genes, we further calculated the exclusive ratio to determine the proportion of linear transcripts with exon skipping events resulting from exon circularization. Correspondingly, the proportion of circularization and the exclusive ratio exhibited a significantly positive correlation across different tumors (Fig. 2a). These results indicate ubiquitous co-occurrence of circRNA biogenesis and alternative splicing events in various tumor types, suggesting that it may represent a

prevalent form of splicing in different physiological processes. Consequently, we obtained a total of 47,203 circRNAs in tumor cells with at least one back-spliced read. In 4400 (or 9.32%) of these events, the alternative splicing events with skipping of the circularized exons were detected. Notably, the rate of co-occurrence increased with increasing host gene expression or supported back-spliced reads (Fig. 2b and Supplementary Fig. S5). For the 6065 circRNAs supported by at least five back-spliced reads, a greater proportion (1020/6065, or 16.82%) of alternative splicing events could be detected in the five tumor cell lines. Moreover, considering the challenges of accurately mapping long-range spliced junction reads, as well as the fact that circRNAs are more stable than their corresponding linear transcripts [34], the levels of circRNA-dependent alternatively-spliced linear transcripts that we detected might be substantially underestimated.

Next, we explored the characteristics of these circRNA-dependent alternatively-spliced transcripts, and found that a large proportion of circRNA-dependent alternative splicing events were absent from the annotated transcripts, with 90.62% in A549 (Fig. 2c) and 88.66% in all five tumors being non-annotated. Additionally, these alternatively-spliced transcripts tended to span longer genomic regions (Wilcoxon rank-sum test,  $P$  value  $< 2.2 \times 10^{-16}$ ) and skip more exons (Fisher’s exact test,  $P$  value  $< 2.2 \times 10^{-16}$ ) than did the annotated transcripts (Fig. 2d and e and Supplementary Fig. S6), indicating a special splicing process. We subsequently denoted these circRNA-dependent alternatively-spliced transcripts as LRs.

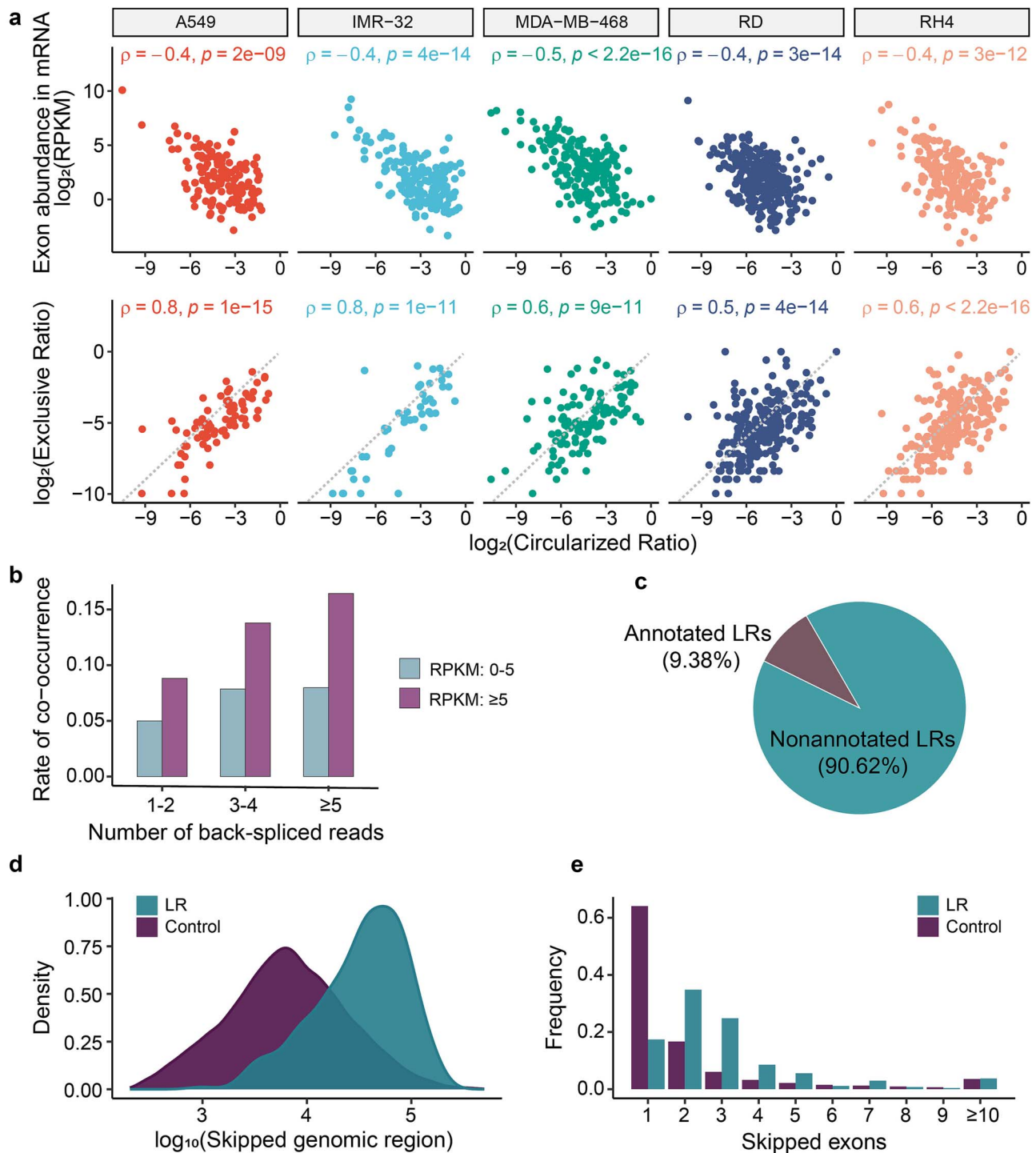
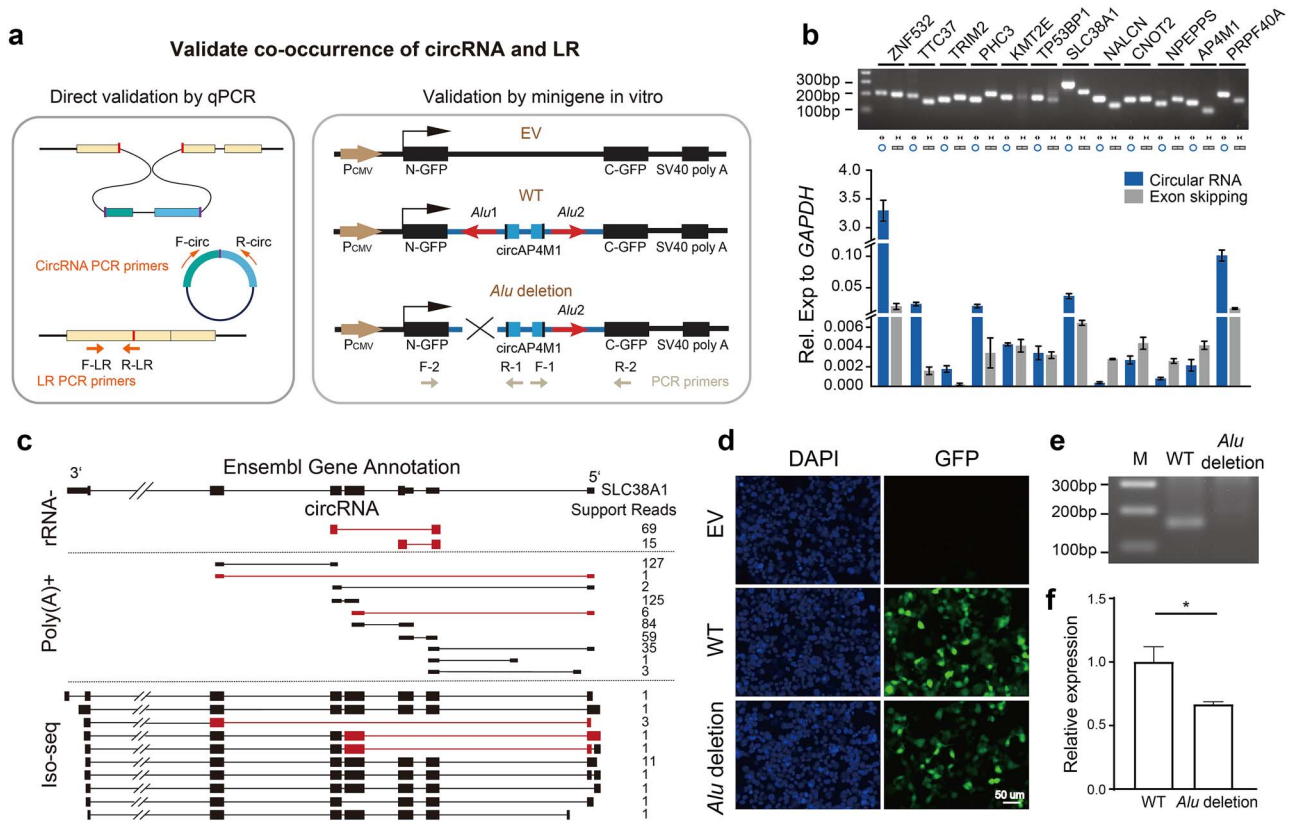


Figure 2. LR are correlated with circRNA biogenesis in tumors. (a) The correlation of the circularized ratio with exon abundance (upper panel) and the exclusive ratio of circularized exons in linear transcripts (lower panel). Spearman correlation coefficients were calculated. (b) For circRNA-encoding genes, the proportions of circRNA regions detected to concurrently encode alternatively-spliced linear transcripts without circularized exons were shown for the A549 cell line. These genes were categorized on the basis of the expression of the circRNAs (estimated using the number of back-spliced supporting reads) and of the host genes (estimated in RPKM). (c) Pie plot showing the proportion of annotated or non-annotated LR in A549 cells. (d) The length distribution of  $\log_{10}$ -transformed skipped genomic regions (base pairs) of LR and annotated alternatively-spliced transcripts (control) in the A549 cell line. (e) Distribution of the number of skipped exons in identified LR and annotated alternatively-spliced transcripts (control) in A549 cells.

To further evaluate the conservation of this splicing mechanism, we conducted an analysis of circRNAs and LR in cerebellar tissue derived from rhesus macaques. Macaques, being phylogenetically close to humans, share key cis-regulatory elements, such as *Alu* sequences, that govern the biogenesis of circRNAs. The accessibility of their tissue samples, combined with these genetic similarities, renders them an optimal model for experimental

validation. We found that circRNAs are similarly associated with LR in macaques (Supplementary Fig. S7a and b). For 5412 circRNAs supported by at least five back-spliced reads, 849 (15.69%) corresponding LR could also be detected (Supplementary Fig. S7c), a rate comparable to previous results (Fig. 2b). qPCR further verified the co-occurrence of circRNAs and the corresponding LR in 12 of 15 randomly-selected genes encoding both isoform



**Figure 3.** Experimental verification of the association between circRNA biogenesis and LRs. (a) Diagram of the experimental design to validate the co-occurrence of circRNA and LR. For qPCR experiment, primers for circRNA back-spliced junction and LR junction are designed to quantify the expressions. For *in vitro* minigene experiment, the structures of the pZW1 vector (EV), as well as two constructs with the inserted sequences of pZW1 circular RNA (WT and Alu deletion) are shown, with the two divergent *Alu* elements in flanking introns (*Alu1* and *Alu2*). The positions of two pairs of PCR primers used to quantify the expression of circular (F-1, R-1) and linear (F-2, R-2) transcripts are indicated by arrows. (b) Gel electrophoresis of the qPCR products for 12 successfully-validated genes encoding both a circRNA and an alternative linear isoform without the circularized exons, along with qPCR quantification results. (c) An example gene exhibiting the co-occurrence of circRNA and a circularized exon skipping event supported by rRNA-depleted RNA-seq, poly(A)-positive RNA-seq data and Iso-seq data. Spliced junctions supporting the alternatively-spliced linear transcripts with circularized exon skipping were highlighted. (d) Representative fluorescence images of HEK293T cells transfected with the empty vector pZW1 (EV), AP4M1-pZW1 construct (WT), or a mutant construct in which the *Alu* sequence in the flanking intron was deleted (*Alu* deletion). DAPI staining was performed to identify the nuclei, and long-range spliced transcript was indicated by GFP signals (scale bar represents 50  $\mu\text{m}$ ). (e) The expression of circAP4M1 in cells transfected with the WT or mutant construct (*Alu* deletion) was confirmed by agarose gel electrophoresis. (f) The long-range spliced transcript was quantified by qPCR in WT or mutant construct (*Alu* deletion).

species, as supported by the high-throughput data (Fig. 3a and b). To further investigate whether these LRs are indeed mature transcripts rather than premature or splicing byproduct, we analyzed PacBio Iso-seq data on poly(A)-positive RNAs derived from a matched tissue sample [35]. A total of 105 out of 343 full-length isoforms with these alternatively-spliced junctions supported by at least three reads in short-read data were directly detected in the Iso-seq data (Fig. 3c), revealing that this type of alternatively-spliced linear transcript coexisting with circRNAs indeed represents mature isoforms.

To explore whether LR formation is modulated by circRNA biogenesis, we cloned the circularized exons of the *AP4M1* gene, along with its full-length flanking introns, into the middle position of the *egfp* gene based on a previously reported design [36] and evaluated the influence of circRNA biogenesis on LR formation (Fig. 3a and Supplementary Methods). In this context, the formation of circRNA would excise the inserted fragments and enable the expression of the *GFP* gene. As the *Alu* sequence is considered an important cis-element that influences circRNA biogenesis [36, 37], a mutant construct with a deletion of the *Alu* sequence in the flanking introns was also constructed to prevent circRNA generation. Both constructs were then transfected

into HEK293T cells to quantify the expression of the circRNAs and alternatively-spliced linear transcripts (Fig. 3a). Subsequent fluorescence microscopy and qPCR revealed that deletion of the intronic *Alu* sequences disrupted circRNA biogenesis and significantly decreased the expression of the corresponding circularized exon-skipped linear isoform (Fig. 3d and f). Taken together, our observations in tumors, cell lines and macaque tissues suggest that circRNA biogenesis is accompanied by alternative splicing of linear transcripts to generate novel isoforms with the circularized exons skipped.

### Long-range alternative splicing is upregulated in glioblastoma

CircRNAs are known to be aberrantly expressed in various physiological and disease conditions [38], suggesting that LRs may be differentially expressed in tumors and thus serve as a previously neglected source of neoantigens for cancer immunotherapy. To investigate this possibility, we utilized all circRNAs from both normal and tumor tissues in Cancer-Specific CircRNA Database (CSCD) [39] to identify the LRs in six types of tumors from TCGA and six matched normal tissues from GTEx (Fig. 4a). Overall, we identified 50 085 and 36 177 LRs in tumor tissues and their

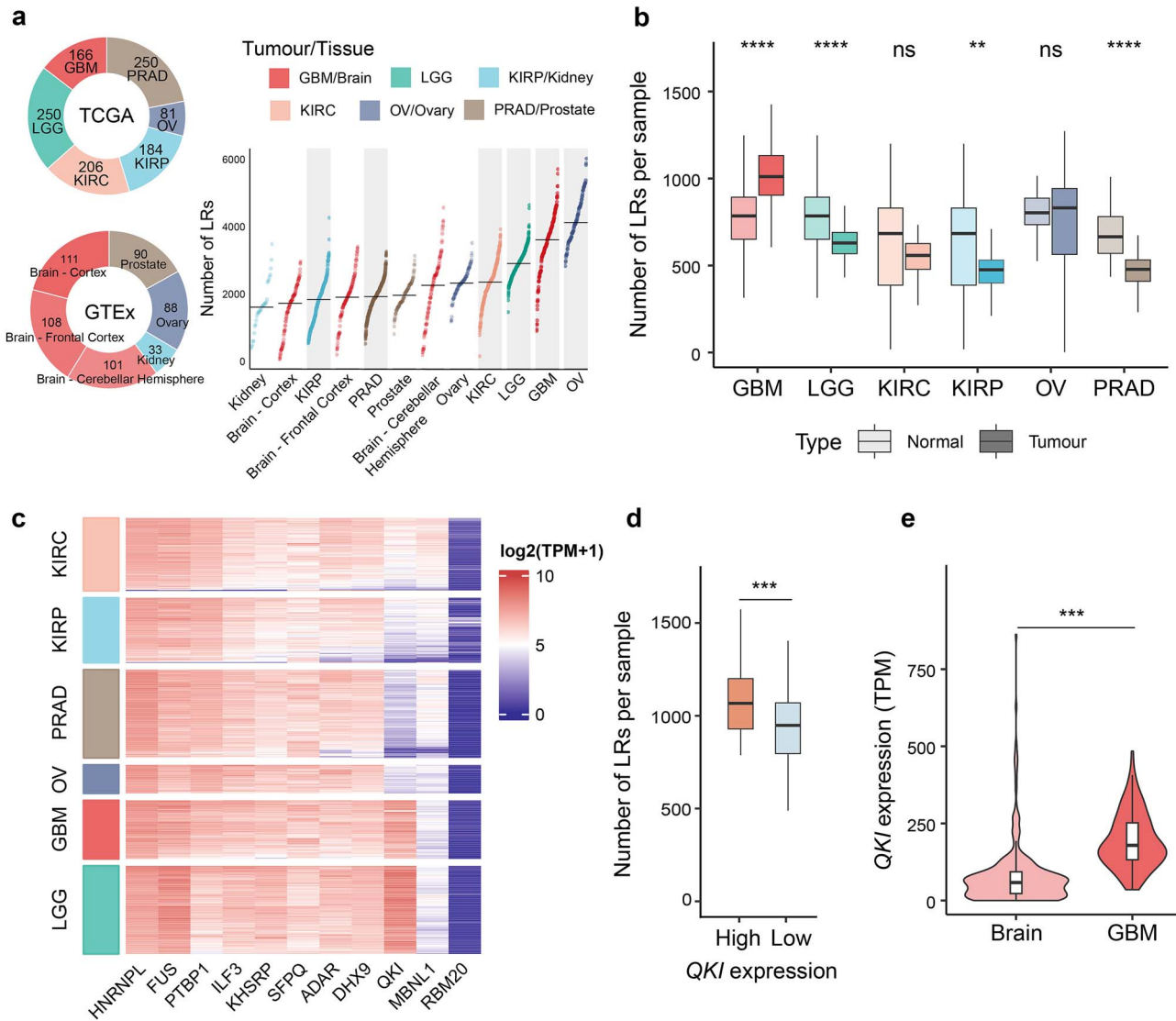


Figure 4. Long-range alternative splicing is specifically upregulated in GBM. (a) Donut charts showing the number of tumor and normal tissue samples from the TCGA and GTEx databases. The number of LR per sample was compared in tumor and normal tissue after downsampling to 20 million reads. Wilcoxon rank-sum test, ns: not significant; \*: P value <.05; \*\*: P value <.01; \*\*\*: P value <.001. (c) The expression of circRNA regulators across tumor samples. (d) The number of LR per sample in GBM was compared between the two groups classified by QKI expression. Wilcoxon rank-sum test, ns: not significant; \*: P value <.05; \*\*: P value <.01; \*\*\*: P value <.001. (e) QKI expression was compared in normal brain tissue and GBM samples. Wilcoxon rank-sum test, ns: not significant; \*: P value <.05; \*\*: P value <.01; \*\*\*: P value <.001.

corresponding normal tissues, respectively. For each tissue type, the number of LR in most tumor types was obviously higher than that in normal tissues (Fig. 4a). However, these disparities may be attributable to differences in sequencing depth between samples. To address this, all the samples were downsampled to equal sequencing depths, and the number of LR per sample was compared between the tumor and normal tissues. We found that LR are differentially regulated across tumor types. Notably, the number of LR in GBM samples was significantly higher than in normal brain tissue and other tumor types (Fig. 4b and Materials and methods).

To investigate the mechanism of the upregulated LR in GBM samples, we examined the expression of RBPs, known to regulate circRNA biogenesis, across tumor types [17, 40]. We found that the expression of quaking (QKI), a protein reported to promote circularization [41], is significantly upregulated in GBM (Fig. 4c). Further analysis revealed a positive correlation between

QKI expression and the number of LR, indicating that higher QKI expression is associated with increased LR in GBM (Fig. 4d). In addition, consistent with the finding of more LR in GBM, QKI expression in GBM was significantly higher than in normal brain tissues (Fig. 4e). Taken together, these findings suggest that QKI may promote the expression of LR in GBMs.

### Glioblastoma-specific long-range alternatively-spliced transcripts are potential recurrent neoantigens

As LR are specifically upregulated in GBM compared to brain tissue (Fig. 4b), we then investigated whether LR may serve as a source of neoantigens for cancer immunotherapies. We identified a total of 11 639 LR specifically expressed in GBM tissues but not in normal brain tissues, with an average of >200 specific LR per individual. As a note, GBM is a cancer type with a low TMB that harbors an average of only 30–50 nonsynonymous mutations per

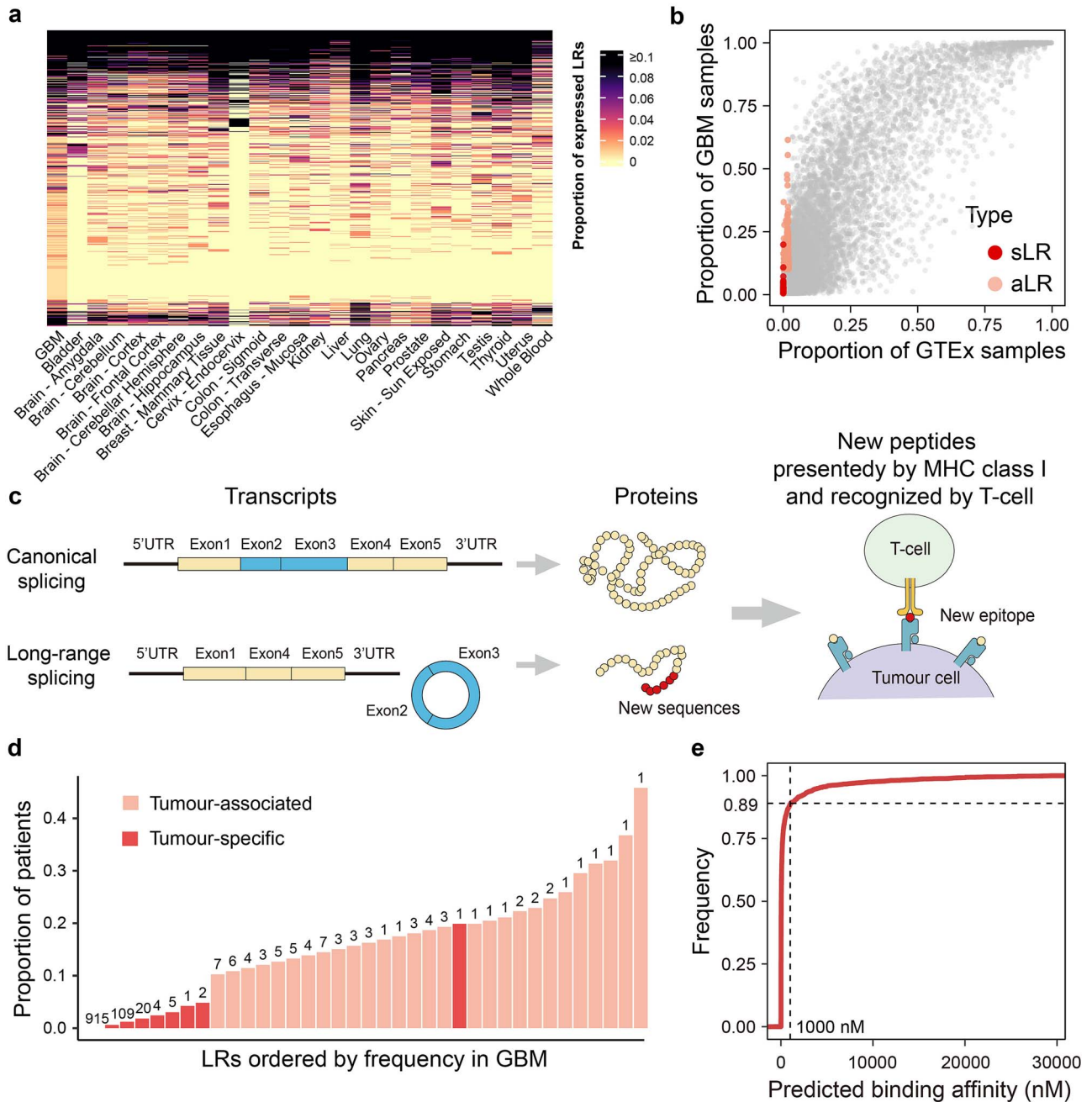


Figure 5. GBM-specific LR are potential recurrent neoantigens. (a) The proportion of samples expressing a given LR (rows) in each tissue type (columns). (b) Frequency of LR (dots) in GBM and all normal samples. GTEx samples were merged to calculate the overall LR expression proportion. (c) Diagram illustrating the potential of LR to produce neoantigens. (d) Proportion of samples expressing an individual LR (number of LR indicated above each bar) ordered by sample frequency in the TCGA-GBM cohort. (e) Cumulative curve of the predicted BA between MHC class I and LR. The top 10 most prevalent HLA alleles were used, and the cut-off was set at 1000 nM.

individual [42]. This suggests that the abundant specific LR in GBM could expand the pool of neoantigens contributing to GBM immunotherapy.

To further investigate the specificity and recurrence of LR in GBM, we expanded our analysis to RNA-seq data from a total of 24 human normal tissues in 1434 individuals (Supplementary Fig. S8) and estimated the proportion of samples expressing each LR in all tissues. Among the 24 937 LR expressed in GBM, we found that some transcripts were exclusively expressed in GBM but absent in any normal tissue (Fig. 5a), indicating the high specificity of LR in GBM. We further classified LR in GBM based on the degree

of specificity. LR expressed exclusively in tumors and absent in any normal tissue were defined as tumor-specific LR (sLR), while those expressed in >10% of tumor samples but <2% of normal tissue samples were categorized as tumor-associated LR (aLR). We identified a total of 3082 sLR and 279 aLR in GBM (Fig. 5b). Most sLR were expressed in only one sample (2685 out of 3082), exhibiting individual-specific expression. However, 12.88% of sLR (397 out of 3082) were present in >1% of patients, with the most broadly prevalent sLR observed in 19.88% of patients, indicating that GBM-specific LR are recurrent to some extent (Fig. 5b).

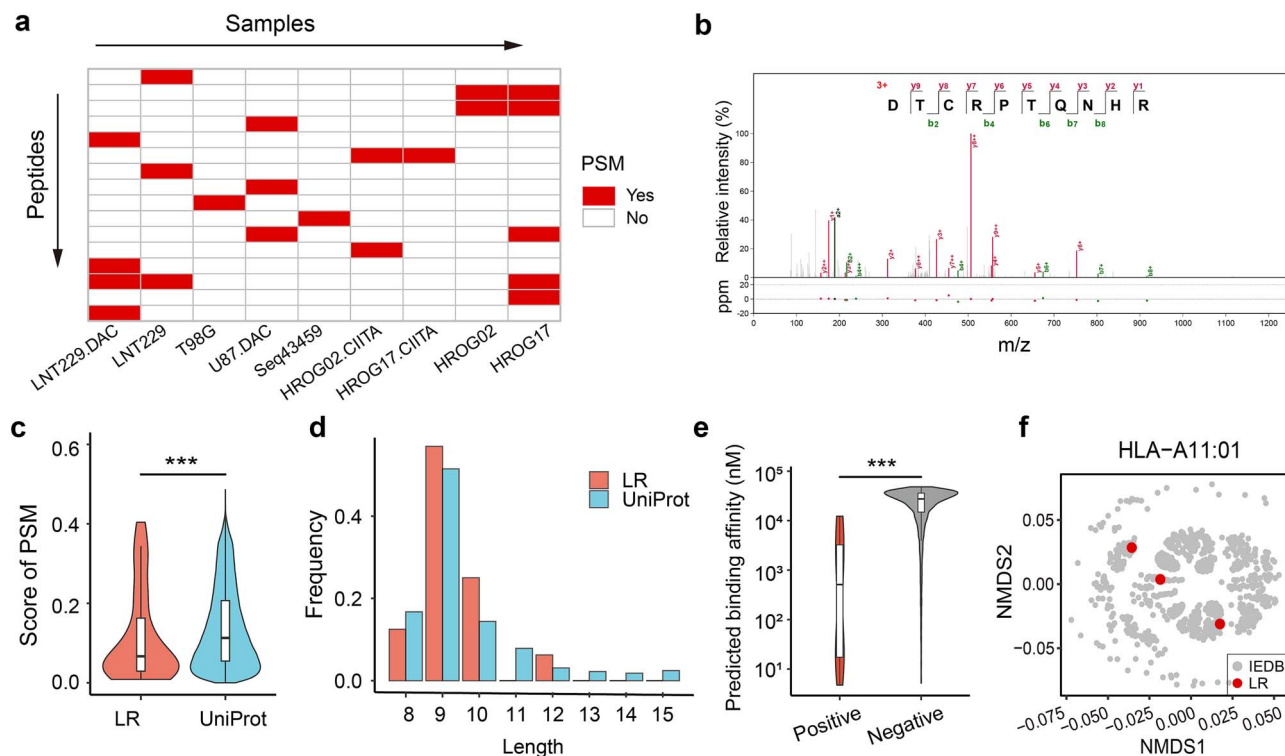


Figure 6. LR-derived peptides in GBM were detected in MS-based immunopeptidomics datasets. (a) Heatmap summarizing the identified peptides (rows) found in each sample (columns) in the MS-based immunopeptidomics dataset derived from GBM cell lines or tissues. (b) Spectra of one identified peptide. (c and d) The PSM score identified in pFind3 (c) and peptide length distribution (d) for peptides identified from LR and UniProt-annotated transcripts. (e) The predicted BA of the corresponding HLA allele in that sample with peptides from MS-supported LR regions (positive) and other regions in the same LR (negative). (f) NMDS representation showing motif compliance for HLA-A11:01-binding peptides comparing identified LR-derived peptides and immunopeptide from the IEDB.

We therefore speculated that some of these LR may undergo frameshift mutations during translation, thereby generating novel protein sequences that are presented by the MHC of tumor cells and recognized by T cells, serving as potential tumor neoantigens (Fig. 5c). We then evaluated the possibility of ORF frameshifting and obtained a total of 1057 sLRs and 76 aLRs predicted to generate novel protein sequences (Supplementary Fig. S9). Most sLRs with ORF frameshifts were present in only one sample (915 of 1057 GBM samples), whereas 142 (13.43%) were expressed in >1% of the GBM samples (Fig. 5d). In contrast, aLRs were more recurrent, with the most frequent aLR detected in 45.78% of GBM samples (Fig. 5d). However, tumor-specific antigens are preferred over tumor-associated antigens in immunotherapy because they can prevent central immune tolerance and minimize adverse effects [43]; thus, sLRs are particularly valuable sources of tumor neoantigens. To further investigate the neoantigen potential of sLRs, we utilized the NetMHCpan tool to predict the binding affinity (BA) between the translated sequences of LR and the ten most prevalent human leukocyte antigen (HLA) alleles within populations [44] (Materials and methods). Remarkably, >89% of the analyzed sLRs were predicted to generate at least one short peptide harboring novel sequences capable of binding to one or more HLA alleles (Fig. 5e). Notably, certain HLA alleles, such as HLA-A11:01, demonstrated the potential to present ~50% of the analyzed sLRs (Supplementary Fig. S10), highlighting sLRs as a reservoir of novel antigens.

To further investigate the neoantigen presentation capability of these GBM-specific LR, high-resolution immunopeptidomics datasets from both GBM cell lines and tissue samples were

collected and analyzed [45–47] (Materials and methods). Remarkably, a total of 16 peptides were detected in at least one sample (Fig. 6a), and the spectra showed high match quality (Fig. 6b). We then comprehensively evaluated the authenticity of the identified peptides. Compared with the UniProt-annotated peptides, the sLR-derived peptides demonstrated even lower peptide spectrum match (PSM) scores, indicating higher matching quality of the peptide spectra (Fig. 6c). In addition, the length distribution of sLR-derived peptides was similar to that of UniProt peptides, and the sLR-derived peptides exhibited a higher affinity for MHC molecules compared to unidentified peptides from the same LR transcript (Fig. 6d and e). Additionally, nonmetric multidimensional scaling (NMDS) analysis of HLA-A11:01 revealed that the identified peptides contained the expected HLA-I-binding motifs (Fig. 6f). These results thus demonstrate the high confidence of the identified peptides, indicating that sLR is a reliable source of neoantigens. Considering that the relatively lower sensitivity of MS [48] and the constraints of peptide identification due to sample availability and HLA alleles, the number of sLR-derived MHC-associated peptides is likely substantially underestimated.

## Discussion

Previous case studies have reported linear isoforms of certain mRNAs excluding circularized exons [20, 21, 49]. However, the extent to which these phenomena are ubiquitous, whether the corresponding linear transcripts truly exist, and whether these transcripts could be useful in cancer immunotherapies remain unknown. Here, we developed a bioinformatics pipeline to simultaneously identify circRNAs and corresponding exon



skipping events in rRNA-depleted RNA-seq data. Overall, an average of 15% of circRNAs were found to coexist with corresponding circRNA-dependent LRs. Further experiments validated the co-occurrence of circRNAs and long-range alternatively-spliced transcripts. Additionally, we found that the proportion of circularization and the exclusive ratio exhibited a significantly positive correlation across different tumors (Fig. 2a). Since circRNAs are generally more stable than linear RNAs, and LR linear transcripts are more prone to degradation, this can lead to an underestimation of authentic correlations. Subsequent experiments revealed that deletion of the intronic Alu sequences disrupted circRNA biogenesis and significantly decreased the expression of the corresponding circularized exon-skipped linear isoform, indicating that LR formation is affected while not entirely regulated by the circRNA biogenesis. Other factors, such as the canonical alternative splicing and the structure of pre-mRNA, may also influence the generation of long-range transcripts, which requires further investigation.

Exon circularization is facilitated by complementary pairing sequences in the flanking introns of circularized exons [17]. Despite sharing common cis-elements, circRNA expression exhibits cell type-specific or tissue-specific patterns [50–52], indicating the influence of *trans*-factors on circRNA biogenesis. Multiple splicing factors and RBPs, such as RNA-binding motif protein 20 (RBM20), muscleblind (MBNL1), and QKI, are involved in the regulation of exon circularization [41, 53–59]. This suggests that circRNA regulation likely results from the concerted action of multiple RBPs, but the crosstalk between different RBPs and their regulatory roles in circRNA biogenesis remain undefined. In this study, we found an upregulation of LRs in GBM while a downregulation in other tumors/tissues. It has been reported that in several tumors, including renal cell carcinoma, ovarian cancer, lung adenocarcinoma, and prostate cancer, circRNA abundance is lower than that in normal tissues, possibly due to dilution effects from rapid cell division [17, 60–62], which is consistent with our finding of downregulated LRs in these tumor tissues. Moreover, we observed upregulated expression of QKI in GBM, which may explain the specificity of LR generation in GBM samples. Notably, although QKI also showed specific high expression in LGG (Fig. 4c), the expression level of QKI in LGG samples did not show a positive correlation with LR number, suggesting that other regulatory proteins may be involved in LGG (Supplementary Fig. S11). Further investigations are thus needed to clarify the regulation of RBPs affecting circRNA and LR generation in different tumors, as well as the interplay between these RBPs.

Research on neoantigens in GBM has primarily focused on SNVs [8, 63], alternative splicing [11, 64], and transposons [65]. Therefore, the neoantigens that arise from non-annotated transcripts in GBM are of great interest for further investigation. Here, we demonstrated that LRs, which are largely non-annotated transcripts, are specifically upregulated in GBM and have the potential to generate multiple tumor-specific MHC class I-associated peptides. Analysis of transcriptomic data from 1434 samples across 24 types of normal tissues revealed no expression of the LRs that generate these peptides in healthy individuals. Furthermore, searches in peptide databases such as PeptideAtlas [66] and the Human Proteome Map [67] found no supporting evidence of these peptides in healthy peptidomic data, which further confirms their lack of expressions in healthy tissue, indicating they possibly prevent central immune tolerance and are prone to elicit T cell response. Additionally, these peptides were identified from MHC-eluted peptide data obtained via MS, indicating they can be processed and presented by MHC in tumor cells, which is an

important prerequisite for eliciting T-cell immune responses. Statistical data from previous reports show that ~30% of predicted tumor neoepitopes in melanoma patients that bind MHC will be recognized by a T cell [68, 69], indicating the immunogenicity of predicted epitopes to some extent. Our findings thus expand the repertoire of tumor-specific non-annotated neoantigens in GBM, providing additional targets for future antigen screening and personalized immunotherapy. However, further experimental validation is needed to determine whether these peptides can be recognized by T cells.

## Materials and methods

### Identification of long-range alternatively-spliced transcripts in the Genotype-Tissue Expression and the Cancer Genome Atlas databases

All lists of circRNAs were downloaded from the CSCD database [39]. Bam files of RNA-seq data were downloaded from the TCGA and GTEx databases, and unique alignments were then extracted with SAMtools (version: 1.16.1) for downstream analysis. Supporting reads and the exclusive ratio of circRNA-dependent LRs were calculated as described above. To avoid biases introduced by sequencing depth and transcriptional noise, authentic LRs were obtained with the filtering criteria of  $\geq 2$  exclusive junction reads and  $\geq 5$  total junction reads. Otherwise, the exclusive ratio at that position was defined as not available (NA). To identify tumor-specific LRs, we merged all the samples in the GTEx database to calculate the overall proportions of each LR. To exclude sequencing bias, we defined the 'NA ratio' to identify LR gene loci that are not expressed in patients. LRs with an 'NA ratio'  $> 50\%$ , indicating that the LR cannot be detected in  $> 50\%$  of individuals in the GTEx or TCGA project, were discarded from downstream analysis.

### In silico translation of long-range alternatively-spliced transcripts and major histocompatibility complex class I-binding affinity prediction

LRs harboring a start codon upstream of the circularized exons were selected for *in silico* translation. We then removed LRs for which the lengths of the skipped regions were multiples of three. The nucleotide sequences of the LRs were obtained using bedtools (version: 2.26.0) [70], and a customized Perl script was used to translate the nucleotide sequences into protein sequences according to the original reading frame until the first stop codon was identified. The new protein sequences were further filtered by performing a BLASTP (version: 2.5.0) search against known proteins from the human UniProt/SwissProt database (updated on 08/12/2023). To analyze the neoantigen-encoding potentials of these proteins, NetMHCpan-4.1 [71] was used to predict the BA of peptides from LRs and MHC class I molecules. Peptides with BA  $< 1000$  nM were considered potential MHC class I-associated peptides. LRs with at least one potential MHC class I-associated peptide were considered potential sources of neoantigens. The ten most prevalent types of HLA alleles were used in this analysis [44].

### Mass spectrometry-based immunopeptidomics data analysis

MS-based immunopeptidomics datasets were obtained from PXD020079, PXD008127 and PXD003790 [45–47]. The data were analyzed by the pFind3 tool [72, 73] in open search mode using the following parameters: no enzyme, precursor mass tolerance of 20 ppm and fragment mass tolerance of 20 ppm. A false discovery rate (FDR) of 5% was applied at the peptide level, and

no FDR filter was used at the protein level. A custom protein database established by integrating the predicted sLR-derived protein sequences with the existing human Swiss-Prot database (updated on 08/12/2023) was used. Only LR identified in the MS data were considered translated and retained for downstream analysis. The positive group included peptides directly identified by immunopeptidomics, while the negative group included peptides derived from the same LRs but not identified in the immunopeptidomics data. NMDS was analyzed according to a previously published method [74].

### Key Points

Our contributions are summarized as follows:

- We developed an in-house pipeline to identify and quantify both circRNAs and alternative splicing events in the same sample and found that long-range alternative splicing is pervasively correlated with circRNA biogenesis across species and various tumor types.
- Through comparative transcriptome analysis, we discovered that long-range alternative splicing is specifically upregulated in GBM, accompanied by increased expression of QKI, an RBP reported to promote circularization.
- Combining proteogenomic analysis, we revealed that GBM-specific long-range spliced transcripts generate MHC class I-associated peptides, thereby expanding the pool of neoantigens for cancer immunotherapy in GBM.

## Acknowledgements

We acknowledge Dr IC Bruce for critical reading of the manuscript. We thank Dr Shaokun Shu at School of Basic Medical Sciences and Dr Aibin He at College of Future Technology from Peking University, and Dr Meng-Qiu Dong and Dr Yong Cao at the National Institute of Biological Sciences for insightful suggestions. The results presented here are based on data generated by TCGA Research Network (<https://www.cancer.gov/tcga>) and The Genotype-Tissue Expression Project.

## Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: The authors declare that they have no competing interests.

## Funding

This work was supported by grants from the Ministry of Science and Technology of China (National Key Research and Development Program of China, 2018YFA0801405 and 2019YFA0801801), the National Natural Science Foundation of China (32400535), and the Chinese Institute for Brain Research (2020-NKX-XM-11).

## Data availability statement

The rRNA-depleted RNA-seq data of tumor cells for circRNA and LR identification were acquired from SRR19977092, SRR19977102, SRR20966721, SRR2962437, and SRR3083993 [30–33]. MS-based immunopeptidomics datasets were obtained from PXD020079, PXD008127, and PXD003790 [45–47]. The macaque cerebellum rRNA-depleted RNA-seq data and poly(A)-positive RNA-seq data were acquired at accession number PRJNA397934 [22].

The macaque cerebellum RNase R-treated RNA-seq data were acquired at accession number SRR6456115. The Iso-seq data were acquired at accession numbers SRR5038768 and SRR5038792 [35]. The code for the LR identification pipeline is available on GitHub (<https://github.com/fulorid/Long-range-alternative-splicing-identification>).

## Author contributions

Mingjun Ji: Formal Analysis, Visualization, Data Curation, Writing—Original Draft, Writing—Review & Editing; Qing Yu: Investigation, Data Curation; Xin-Zhuang Yang: Methodology, Formal Analysis; Xianhong Yu: Validation; Jiaxin Wang, Chunfu Xiao, Ni A. An, Chuanhui Han: Investigation, Data Curation; Chuan-Yun Li, Wanqiu Ding: Conceptualization, Supervision, Methodology, Project Administration, Funding Acquisition, Writing—Original Draft, Writing—Review & Editing.

## References

1. Haslam A, Prasad V. Estimation of the percentage of US patients with cancer who are eligible for and respond to checkpoint inhibitor immunotherapy drugs. *JAMA Netw Open* 2019;**2**:2. <https://doi.org/10.1001/jamanetworkopen.2019.2535>.
2. Medikonda R, Dunn G, Rahman M. et al. A review of glioblastoma immunotherapy. *J Neurooncol* 2021;**151**:41–53. <https://doi.org/10.1007/s11060-020-03448-1>.
3. Fang X, Guo Z, Liang J. et al. Neoantigens and their potential applications in tumor immunotherapy. *Oncol Lett* 2022;**23**:88. <https://doi.org/10.3892/ol.2022.13208>.
4. Pearlman AH, Hwang MS, Konig MF. et al. Targeting public neoantigens for cancer immunotherapy. *Nat Cancer* 2021;**2**:487–97. <https://doi.org/10.1038/s43018-021-00210-y>.
5. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;**348**:69–74. <https://doi.org/10.1126/science.aaa4971>.
6. Robbins PF, Lu Y-C, El-Gamil M. et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* 2013;**19**:747–52. <https://doi.org/10.1038/nm.3161>.
7. Lang F, Schrörs B, Löwer M. et al. Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat Rev Drug Discov* 2022;**21**:261–82. <https://doi.org/10.1038/s41573-021-00387-y>.
8. Yang W, Lee K-W, Srivastava RM. et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med* 2019;**25**:767–75. <https://doi.org/10.1038/s41591-019-0434-2>.
9. Zhang M, Fritsche J, Roszik J. et al. RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat Commun* 2018;**9**:3919. <https://doi.org/10.1038/s41467-018-06405-9>.
10. Shah NM, Jang HJ, Liang Y. et al. Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat Genet* 2023;**55**:631–9. <https://doi.org/10.1038/s41588-023-01349-3>.
11. Kahles A, Lehmann K-V, Toussaint NC. et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 2018;**34**:211–224.e6. <https://doi.org/10.1016/j.ccell.2018.07.001>.
12. Frankiw L, Baltimore D, Li G. Alternative mRNA splicing in cancer immunotherapy. *Nat Rev Immunol* 2019;**19**:675–87. <https://doi.org/10.1038/s41577-019-0195-7>.
13. Laumont CM, Vincent K, Hesnard L. et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl*

- Med* 2018;**10**:eaa5516. <https://doi.org/10.1126/scitranslmed.aau5516>.
14. Laumont CM, Daouda T, Laverdure J-P. et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat Commun* 2016;**7**:10238. <https://doi.org/10.1038/ncomms10238>.
  15. Ouspenskaia T, Law T, Clauser KR. et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol* 2022;**40**:209–17. <https://doi.org/10.1038/s41587-021-01021-3>.
  16. Zhou W-Y, Cai Z-R, Liu J. et al. Circular RNA: metabolism, functions and interactions with proteins. *Mol Cancer* 2020;**19**:172. <https://doi.org/10.1186/s12943-020-01286-3>.
  17. Yang L, Wilusz JE, Chen L-L. Biogenesis and regulatory roles of circular RNAs. *Annu Rev Cell Dev Biol* 2022;**38**:263–89. <https://doi.org/10.1146/annurev-cellbio-120420-125117>.
  18. Kristensen LS, Andersen MS, Stagsted LVW. et al. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019;**20**:675–91. <https://doi.org/10.1038/s41576-019-0158-7>.
  19. Ferreira HJ, Stevenson BJ, Pak H. et al. Immunopeptidomics-based identification of naturally presented non-canonical circRNA-derived peptides. *Nat Commun* 2024;**15**:2357. <https://doi.org/10.1038/s41467-024-46408-3>.
  20. Zaphiropoulos PG. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc Natl Acad Sci U S A* 1996;**93**:6536–41. <https://doi.org/10.1073/pnas.93.13.6536>.
  21. Kelly S, Greenman C, Cook PR. et al. Exon skipping is correlated with exon circularization. *J Mol Biol* 2015;**427**:2414–7. <https://doi.org/10.1016/j.jmb.2015.02.018>.
  22. An NA, Ding W, Yang X-Z. et al. Evolutionarily significant A-to-I RNA editing events originated through G-to-A mutations in primates. *Genome Biol* 2019;**20**:24. <https://doi.org/10.1186/s13059-019-1638-y>.
  23. Zeng X, Lin W, Guo M. et al. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol* 2017;**13**:e1005420. <https://doi.org/10.1371/journal.pcbi.1005420>.
  24. Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief Bioinform* 2018;**19**:803–10. <https://doi.org/10.1093/bib/bbx014>.
  25. Zhang J, Chen S, Yang J. et al. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun* 2020;**11**:90. <https://doi.org/10.1038/s41467-019-13840-9>.
  26. Memczak S, Jens M, Elefsinioti A. et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**:333–8. <https://doi.org/10.1038/nature11928>.
  27. Ma X-K, Wang M-R, Liu C-X. et al. CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics Proteomics Bioinformatics* 2019;**17**:511–21. <https://doi.org/10.1016/j.gpb.2019.11.004>.
  28. Feng J, Chen K, Dong X. et al. Genome-wide identification of cancer-specific alternative splicing in circRNA. *Mol Cancer* 2019;**18**:35. <https://doi.org/10.1186/s12943-019-0996-0>.
  29. Vromman M, Anckaert J, Bortoluzzi S. et al. Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nat Methods* 2023;**20**:1159–69. <https://doi.org/10.1038/s41592-023-01944-6>.
  30. Ashouri A, Sayin VI, Van den Eynden J. et al. Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events. *Nat Commun* 2016;**7**:13197. <https://doi.org/10.1038/ncomms13197>.
  31. Scheckel C, Drapeau E, Frias MA. et al. Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. *Elife* 5:e10421.
  32. Dattilo D, Di Timoteo G, Setti A. et al. The m6A reader YTHDC1 and the RNA helicase DDX5 control the production of rhabdomyosarcoma-enriched circRNAs. *Nat Commun* 2023;**14**:1898. <https://doi.org/10.1038/s41467-023-37578-7>.
  33. Ross MO, Xie Y, Owyang RC. et al. PTPN2 copper-sensing rapidly relays copper level fluctuations into EGFR/CREB activation and associated CTR1 transcriptional repression. *Nat Commun* 2024;**15**:6947.
  34. Li Y, Zheng Q, Bao C. et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res* 2015;**25**:981–4. <https://doi.org/10.1038/cr.2015.82>.
  35. Zhang S-J, Wang C, Yan S. et al. Isoform evolution in primates through independent combination of alternative RNA processing events. *Mol Biol Evol* 2017;**34**:2453–68. <https://doi.org/10.1093/molbev/msx212>.
  36. Zhang X-O, Wang H-B, Zhang Y. et al. Complementary sequence-mediated exon circularization. *Cell* 2014;**159**:134–47. <https://doi.org/10.1016/j.cell.2014.09.001>.
  37. Jeck WR, Sorrentino JA, Wang K. et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;**19**:141–57. <https://doi.org/10.1261/rna.035667.112>.
  38. Lee ECS, Elhassan SAM, Lim GPL. et al. The roles of circular RNAs in human development and diseases. *Biomed Pharmacother* 2019;**111**:198–208. <https://doi.org/10.1016/j.biopha.2018.12.052>.
  39. Xia S, Feng J, Chen K. et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* 2018;**46**:D925–9. <https://doi.org/10.1093/nar/gkx863>.
  40. Shen H, An O, Ren X. et al. ADARs act as potent regulators of circular transcriptome in cancer. *Nat Commun* 2022;**13**:1508. <https://doi.org/10.1038/s41467-022-29138-2>.
  41. Conn SJ, Pillman KA, Toubia J. et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015;**160**:1125–34. <https://doi.org/10.1016/j.cell.2015.02.014>.
  42. Alexandrov LB, Nik-Zainal S, Wedge DC. et al. Signatures of mutational processes in human cancer. *Nature* 2013;**500**:415–21. <https://doi.org/10.1038/nature12477>.
  43. Wang C, Yu M, Zhang W. Neoantigen discovery and applications in glioblastoma: an immunotherapy perspective. *Cancer Lett* 2022;**550**:215945. <https://doi.org/10.1016/j.canlet.2022.215945>.
  44. He Y, Li J, Mao W. et al. HLA common and well-documented alleles in China. *HLA* 2018;**92**:199–205. <https://doi.org/10.1111/tan.13358>.
  45. Forlani G, Michaux J, Pak H. et al. CIITA-transduced glioblastoma cells uncover a rich repertoire of clinically relevant tumor-associated HLA-II antigens. *Mol Cell Proteomics* 2021;**20**:100032. <https://doi.org/10.1074/mcp.RA120.002201>.
  46. Shraibman B, Barnea E, Kadosh DM. et al. Identification of tumor antigens among the HLA peptidomes of glioblastoma tumors and plasma\*. *Mol Cell Proteomics* 2018;**17**:2132–45. <https://doi.org/10.1074/mcp.RA118.000792>.
  47. Human Leukocyte Antigen (HLA). Peptides derived from tumor antigens induced by inhibition of DNA methylation for development of drug-facilitated immunotherapy. *Mol Cell Proteomics* 2016;**15**:3058–70. <https://doi.org/10.1074/mcp.M116.060350>.
  48. Becker JP, Riemer AB. The importance of being presented: target validation by immunopeptidomics for epitope-specific immunotherapies. *Front Immunol* 2022;**13**:883989.
  49. Zaphiropoulos PG. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in

- epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol* 1997;**17**:2985–93. <https://doi.org/10.1128/MCB.17.6.2985>.
50. Salzman J, Chen RE, Olsen MN. et al. Cell-type specific features of circular RNA expression. *PLoS Genet* 2013;**9**:e1003777. <https://doi.org/10.1371/journal.pgen.1003777>.
51. Zhang X-O, Dong R, Zhang Y. et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 2016;**26**:1277–87. <https://doi.org/10.1101/gr.202895.115>.
52. Dong R, Ma X-K, Chen L-L. et al. Increased complexity of circRNA expression during species evolution. *RNA Biol* 2017;**14**:1064–74. <https://doi.org/10.1080/15476286.2016.1269999>.
53. Ashwal-Fluss R, Meyer M, Pamudurti NR. et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014;**56**:55–66. <https://doi.org/10.1016/j.molcel.2014.08.019>.
54. Ivanov A, Memczak S, Wyler E. et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 2015;**10**:170–7. <https://doi.org/10.1016/j.celrep.2014.12.019>.
55. Kramer MC, Liang D, Tatomer DC. et al. Combinatorial control of drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev* 2015;**29**:2168–82. <https://doi.org/10.1101/gad.270421.115>.
56. Rybak-Wolf A, Stottmeister C, Glažar P. et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015;**58**:870–85. <https://doi.org/10.1016/j.molcel.2015.03.027>.
57. Khan MAF, Reckman YJ, Aufiero S. et al. RBM20 regulates circular RNA production from the titin gene. *Circ Res* 2016;**119**:996–1003. <https://doi.org/10.1161/CIRCRESAHA.116.309568>.
58. Errichelli L, Dini Modigliani S, Laneve P. et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat Commun* 2017;**8**:14741. <https://doi.org/10.1038/ncomms14741>.
59. Li X, Liu C-X, Xue W. et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* 2017;**67**:214–227.e7. <https://doi.org/10.1016/j.molcel.2017.05.023>.
60. Bachmayr-Heyda A, Reiner AT, Auer K. et al. Correlation of circular RNA abundance with proliferation—exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep* 2015;**5**:8057. <https://doi.org/10.1038/srep08057>.
61. Chen S, Huang V, Xu X. et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell* 2019;**176**:831–843.e22. <https://doi.org/10.1016/j.cell.2019.01.025>.
62. Vo JN, Cieslik M, Zhang Y. et al. The landscape of circular RNA in cancer. *Cell* 2019;**176**:869–881.e13. <https://doi.org/10.1016/j.cell.2018.12.021>.
63. Turajlic S, Litchfield K, Xu H. et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* 2017;**18**:1009–21. [https://doi.org/10.1016/S1470-2045\(17\)30516-8](https://doi.org/10.1016/S1470-2045(17)30516-8).
64. Wang L, Shamardani K, Babikir H. et al. The evolution of alternative splicing in glioblastoma under therapy. *Genome Biol* 2021;**22**:48. <https://doi.org/10.1186/s13059-021-02259-5>.
65. Bonté P-E, Arribas YA, Merlotti A. et al. Single-cell RNA-seq-based proteogenomics identifies glioblastoma-specific transposable elements encoding HLA-I-presented peptides. *Cell Rep* 2022;**39**:110916. <https://doi.org/10.1016/j.celrep.2022.110916>.
66. Desiere F, Deutsch EW, King NL. et al. The PeptideAtlas project. *Nucleic Acids Res* 2006;**34**:D655–8. <https://doi.org/10.1093/nar/gkj040>.
67. Kim M-S, Pinto SM, Getnet D. et al. A draft map of the human proteome. *Nature* 2014;**509**:575–81. <https://doi.org/10.1038/nature13302>.
68. Vitiello A, Zanetti M. Neoantigen prediction and the need for validation. *Nat Biotechnol* 2017;**35**:815–7. <https://doi.org/10.1038/nbt.3932>.
69. Carreno BM, Magrini V, Becker-Hapak M. et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 2015;**348**:803–8. <https://doi.org/10.1126/science.aaa3828>.
70. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
71. Reynisson B, Alvarez B, Paul S. et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020;**48**:W449–54. <https://doi.org/10.1093/nar/gkaa379>.
72. Chi H, Liu C, Yang H. et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol* 2018;**36**:1059–61. <https://doi.org/10.1038/nbt.4236>.
73. Shao G, Cao Y, Chen Z. et al. How to use open-pFind in deep proteomics data analysis? A protocol for rigorous identification and quantitation of peptides and proteins from mass spectrometry data. *Biophys Rep* 2021;**7**:207–26. <https://doi.org/10.52601/bpr.2021.210004>.
74. Sarkizova S, Klaeger S, Le PM. et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat Biotechnol* 2020;**38**:199–209. <https://doi.org/10.1038/s41587-019-0322-9>.