
Figures and figure supplements

circFL-seq reveals full-length circular RNAs with rolling circular reverse transcription and nanopore sequencing

Zelin Liu *et al*

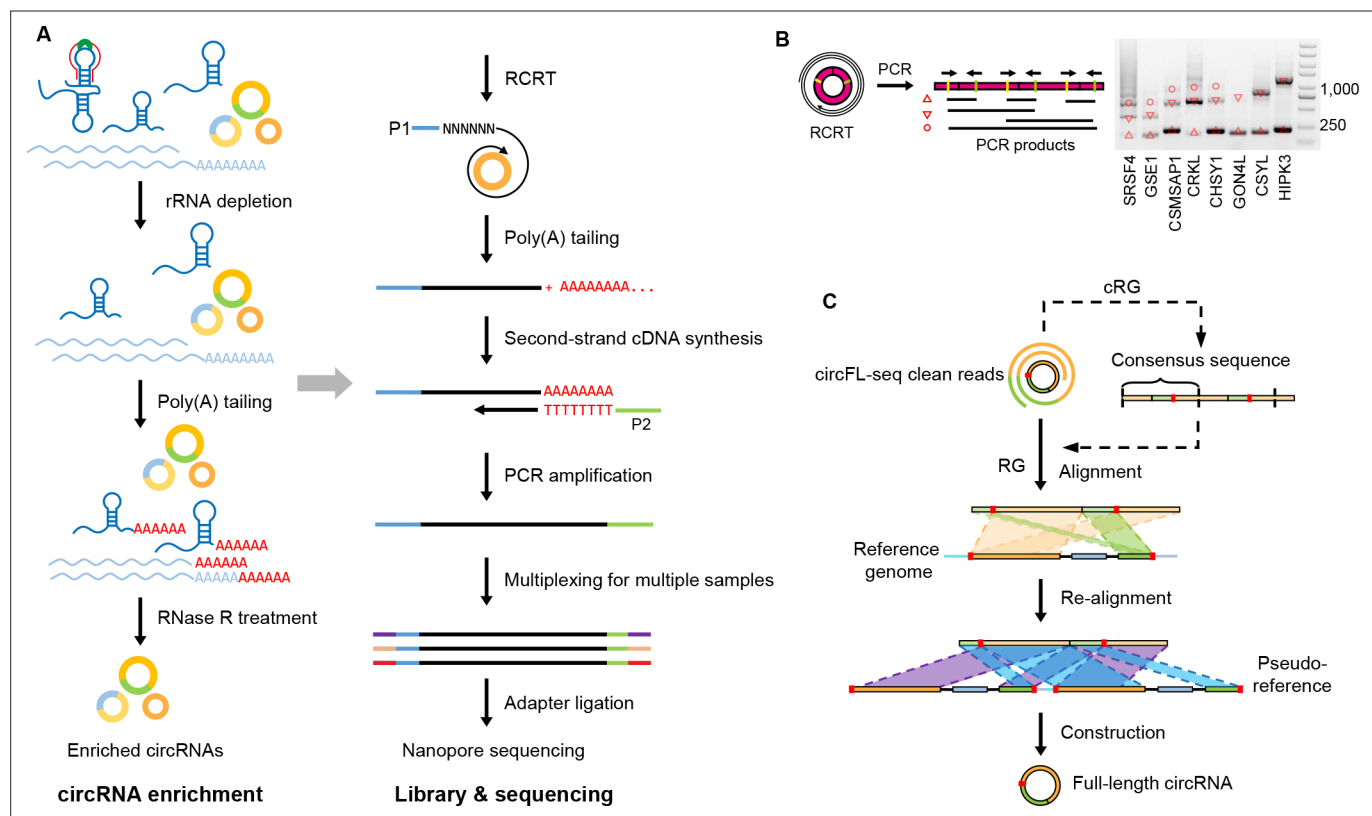


Figure 1. Diagram of circFL-seq workflow. **(A)** Experimental operation of circFL-seq consisted of circRNA enrichment, library construction, and nanopore sequencing. **(B)** PCR validation of rolling circle products from the circFL-seq cDNA library. The yellow and green lines indicate the positions of the PCR primers. The upward triangle, downward triangle, and circle symbols denote the 0-circle, 1-circle, and 2-circle cDNA products. **(C)** Computational pipeline of circFL-seq. circFL-seq clean reads were directly used in RG mode or were self-corrected for consensus sequences in cRG mode to reconstruct full-length circRNAs. circRNA, circular RNA.

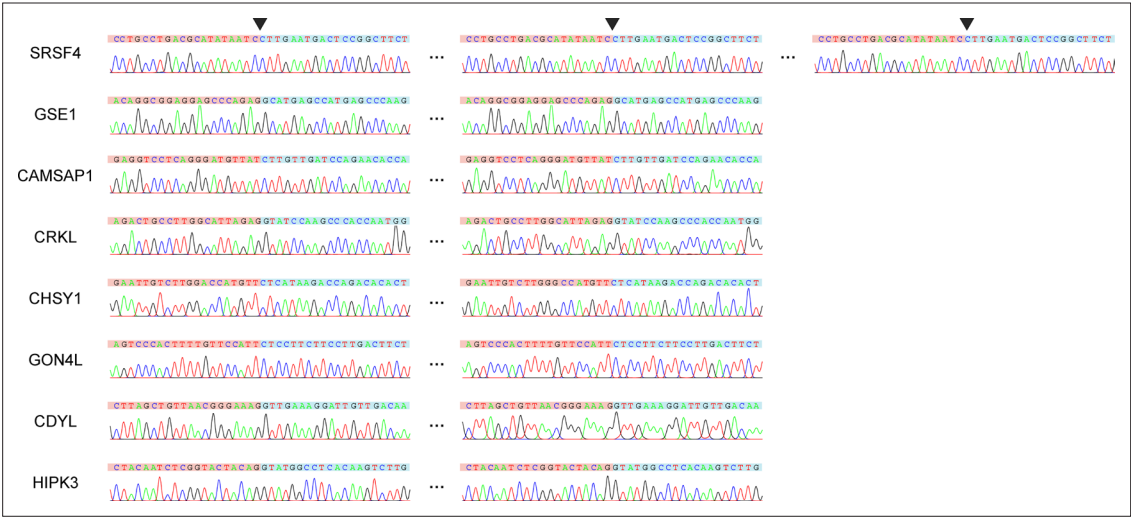


Figure 1—figure supplement 1. Sanger sequencing of rolling circular bands. PCR was performed with the circFL-seq library of HEK293T cells as a template.

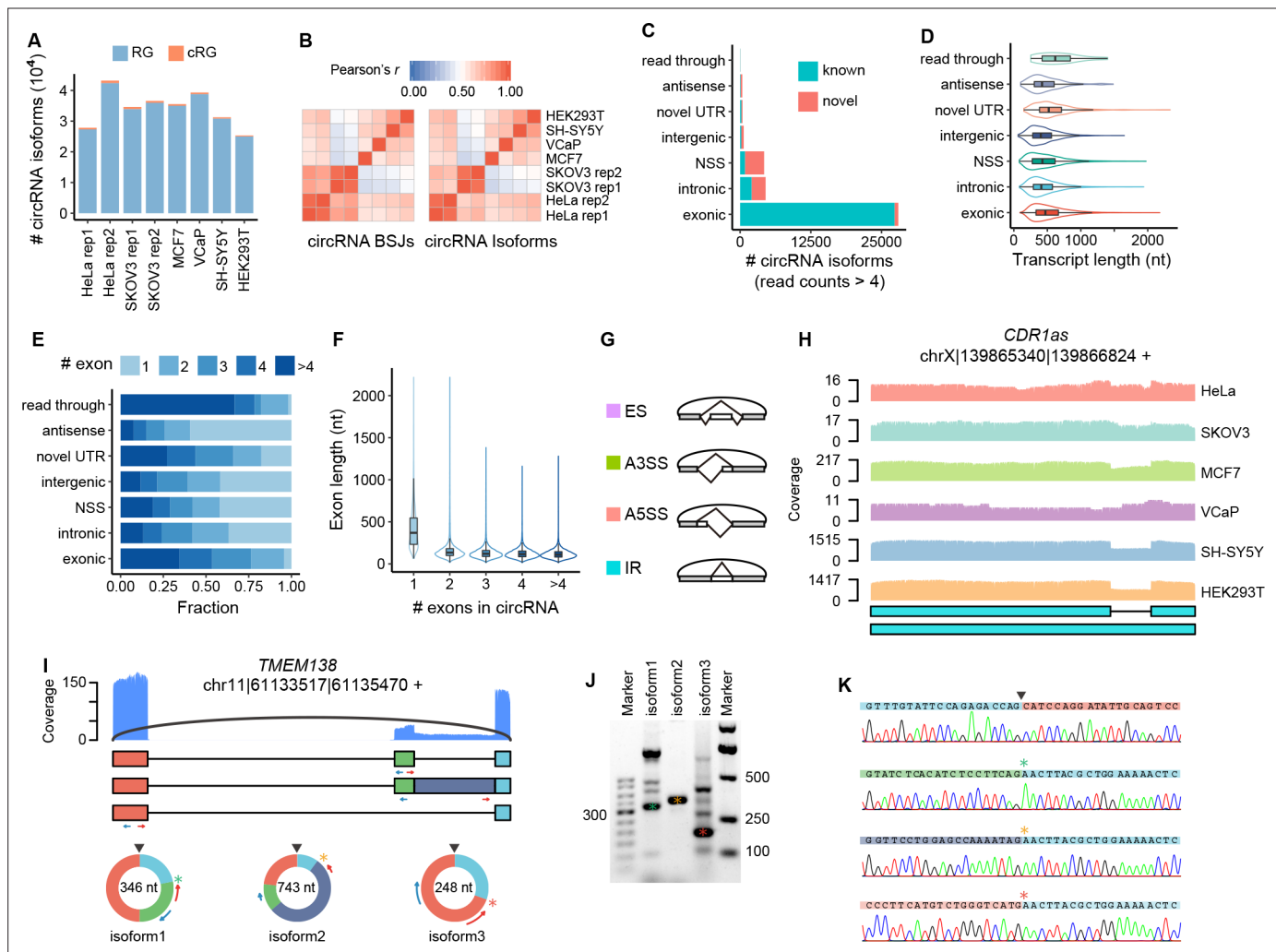


Figure 2. Analysis of full-length circRNA in eight samples. **(A)** Stacked bar plot represents the number of full-length circRNA isoforms detected by RG and cRG for six cell lines. **(B)** Expression correlation matrix for circRNA BSJs and isoforms of all samples. The color scale corresponds to Pearson's correlation coefficient. **(C)** Stacked bar plot represents the number of circRNA isoforms with read counts ≥ 5 from known or novel BSJs based on the circRNA database. **(D)** Boxplot showing the length distribution per isoform for circRNA isoforms with read counts ≥ 5 in all samples. Box lefts or rights are lower or upper quartiles, the bar is the median, and whiskers are the median ± 1.5 × interquartile range. **(E)** Stacked bar plot showing the fraction of exon numbers per isoform for circRNA isoforms with read counts ≥ 5 in all samples. **(F)** Boxplot showing the length distribution per exon for circRNA isoforms with read counts ≥ 5 in all samples. Box bottoms or tops are lower or upper quartiles, the bar is the median, and the whiskers are the median ± 1.5 × interquartile range. **(G)** Diagram of four types of alternative splicing (AS) events in circRNA: exon skipping (ES), alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), and intron retention (IR). **(H)** Plot showing the coverage of full-length circRNA reads in the position of *CDR1as* for six cell lines (replicate data were merged). Structures of the two isoforms of *CDR1as* are shown at the bottom. **(I–K)** AS events (one ES and one IR) of circ-*TMEM138* detected by circFL-seq **(I)**, agarose gel electrophoresis **(J)**, and Sanger sequencing **(K)**. Red/blue arcs are forward/reverse primers for validation of back-splicing junctions (BSJs) and forward splicing junctions (FSJs). Asterisks denote FSJs. Downward triangles denote BSJs. BSJ, back-splicing junction; circRNA, circular RNA; RG, reference guide.

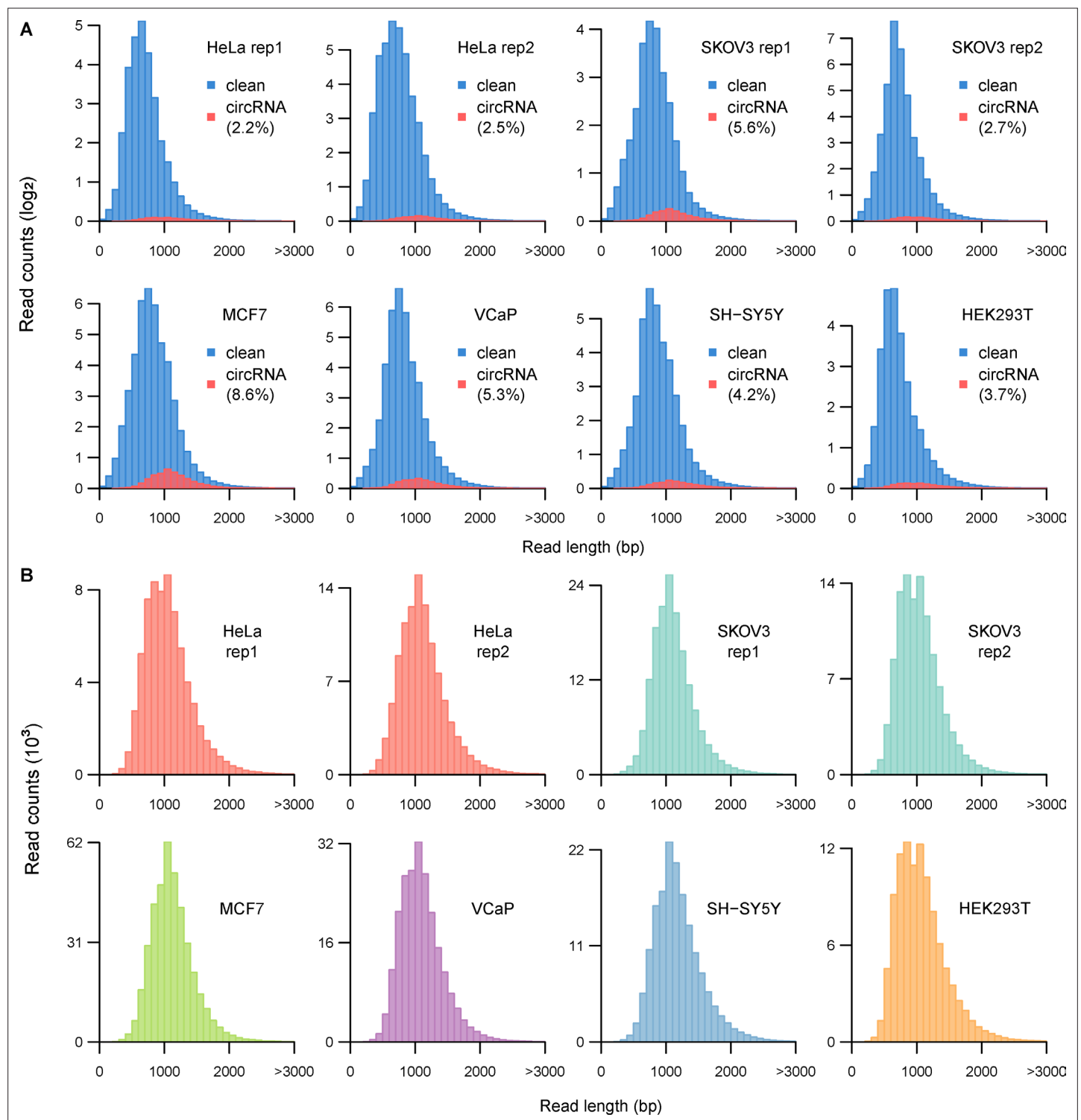


Figure 2—figure supplement 1. Clean read distribution of circFL-seq data of six cell lines. **(A)** Histograms showing the distribution of clean reads (blue) and full-length circRNA reads (red) for each sample. The percentages represent the circRNA read proportion in clean reads. **(B)** Histograms showing the distribution of full-length circRNA read amounts. circRNA, circular RNA.

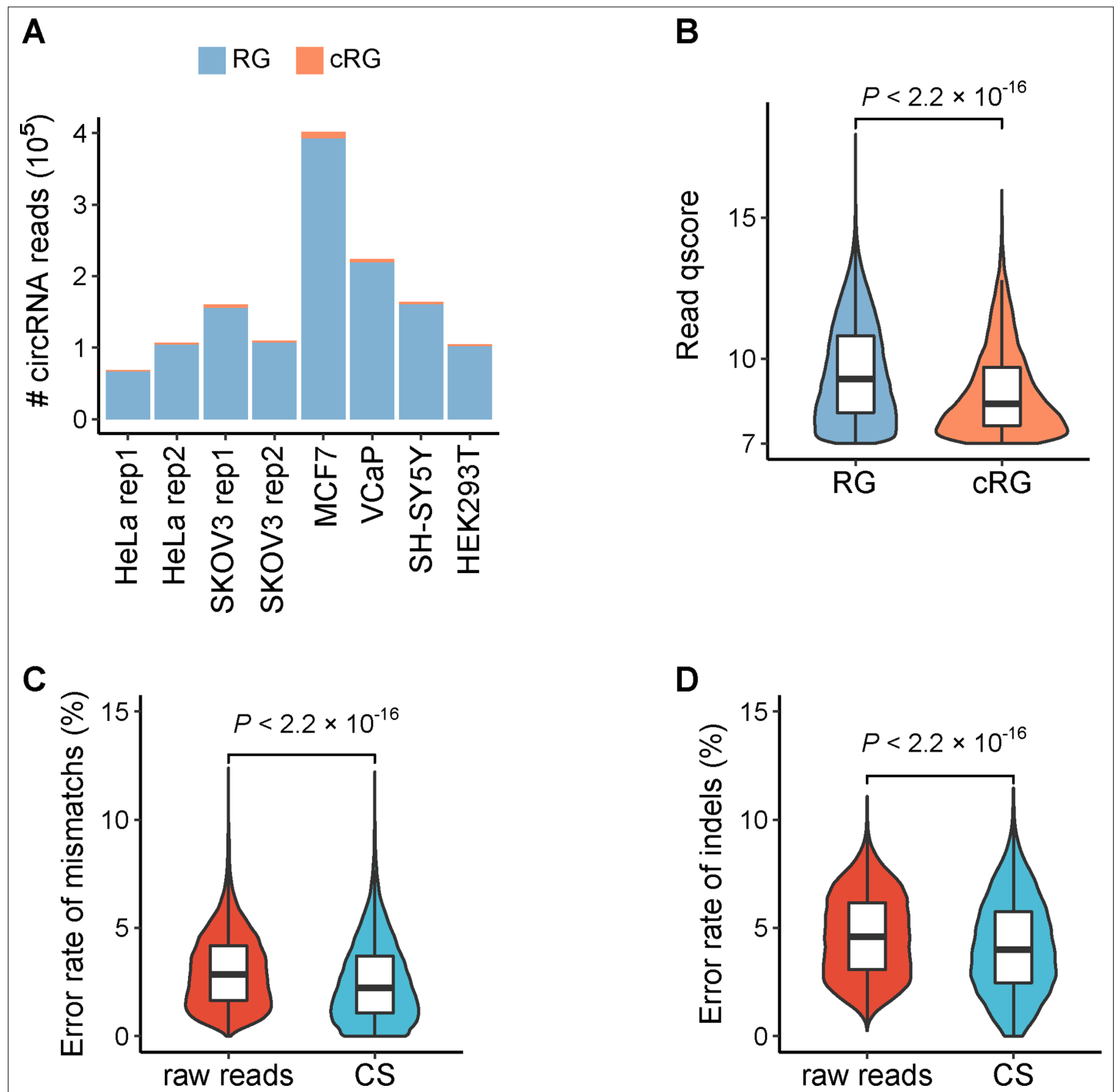


Figure 2—figure supplement 2. CircRNA reads identified from circFL-seq data of six cell lines. **(A)** Stacked bar plot representing the number of full-length circRNA isoforms detected by RG and cRG for eight samples. **(B)** Boxplot showing the read qscore distribution of circRNA isoforms of all samples. The qscore representing the read quality was extracted from the sequencing summary file. Boxplot showing the error rate of mismatches **(C)** and indels **(D)** for raw reads and consensus sequence (CS) of cRG-identified circRNA isoforms. circRNA, circular RNA; RG, reference guide.

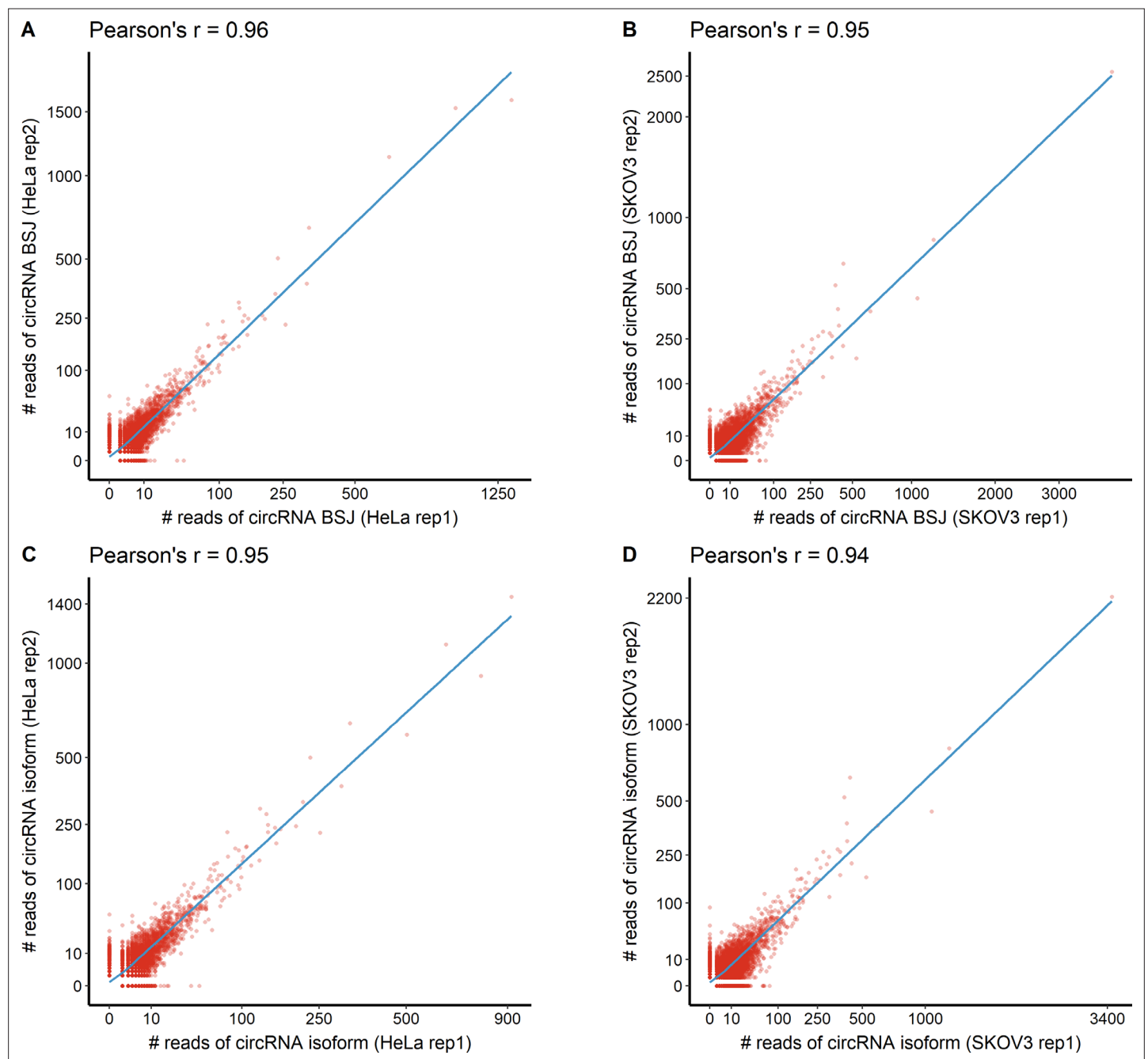


Figure 2—figure supplement 3. Scatter plot showing the correlation of circRNA at the BSJ level (A, B) and isoform level (C, D) between circFL-seq replicates. CircRNA BSJs/isoforms with read counts > 0 in at least one replicate were included. BSJ, back-splicing junction; circRNA, circular RNA.

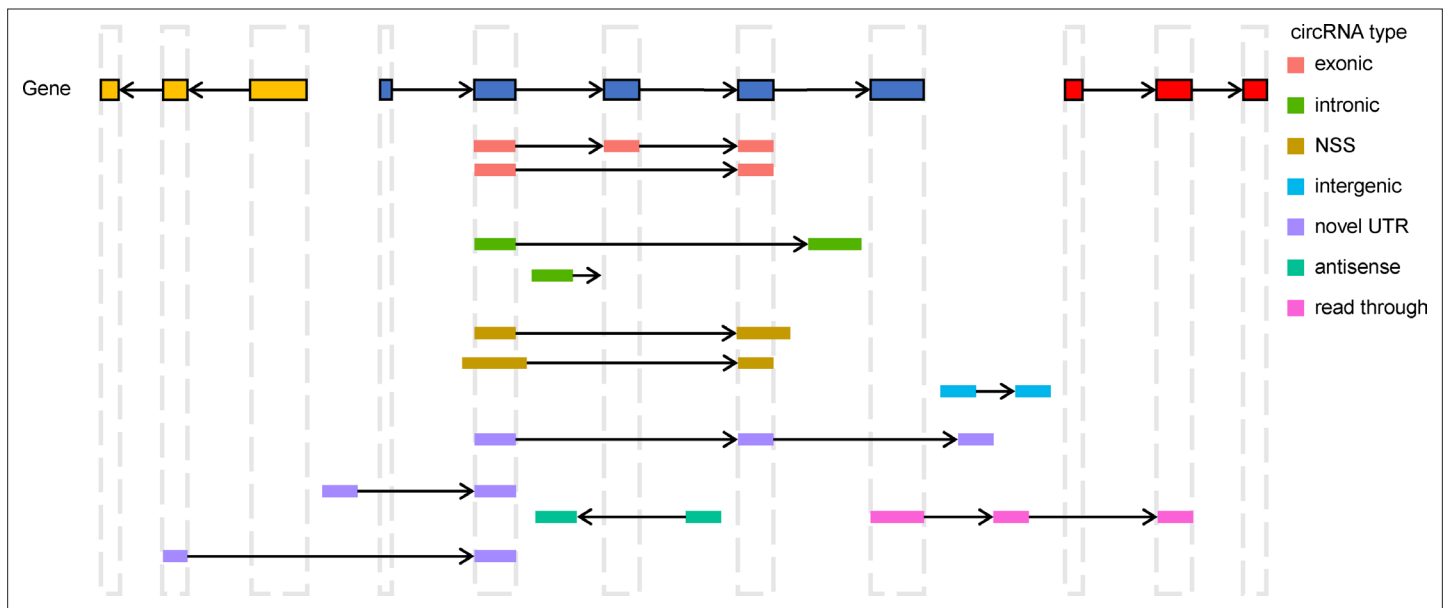


Figure 2—figure supplement 4. Diagram of circRNA types. The classification is based on the positions of BSJs and boundary exons following the principles below. Exonic: the circRNA body is totally located inside of one gene from the same strand. Both of the BSJs are identical to annotated junctions. Intronic: the circRNA body is also totally located inside one gene from the same strand. However, at least one of the boundary exons does not overlap with any annotated exon. Novel splicing site (NSS): the circRNA body is also totally located inside one gene from the same strand. However, both boundary exons overlap with annotated exons, with at least one BSJ different from the annotated linear junction. Intergenic: the whole body of circRNA is located in an intergenic region. Novel UTR: the circRNA body partially overlaps with only one gene from the same strand, and at least one BSJ is located in an intergenic region. Antisense: there is no overlap between circRNA and any gene from the same strand. However, the circRNA overlaps gene(s) from the antisense strand. Read-through: BSJs are located in different genes with the same strand. BSJ, back-splicing junction; circRNA, circular RNA.

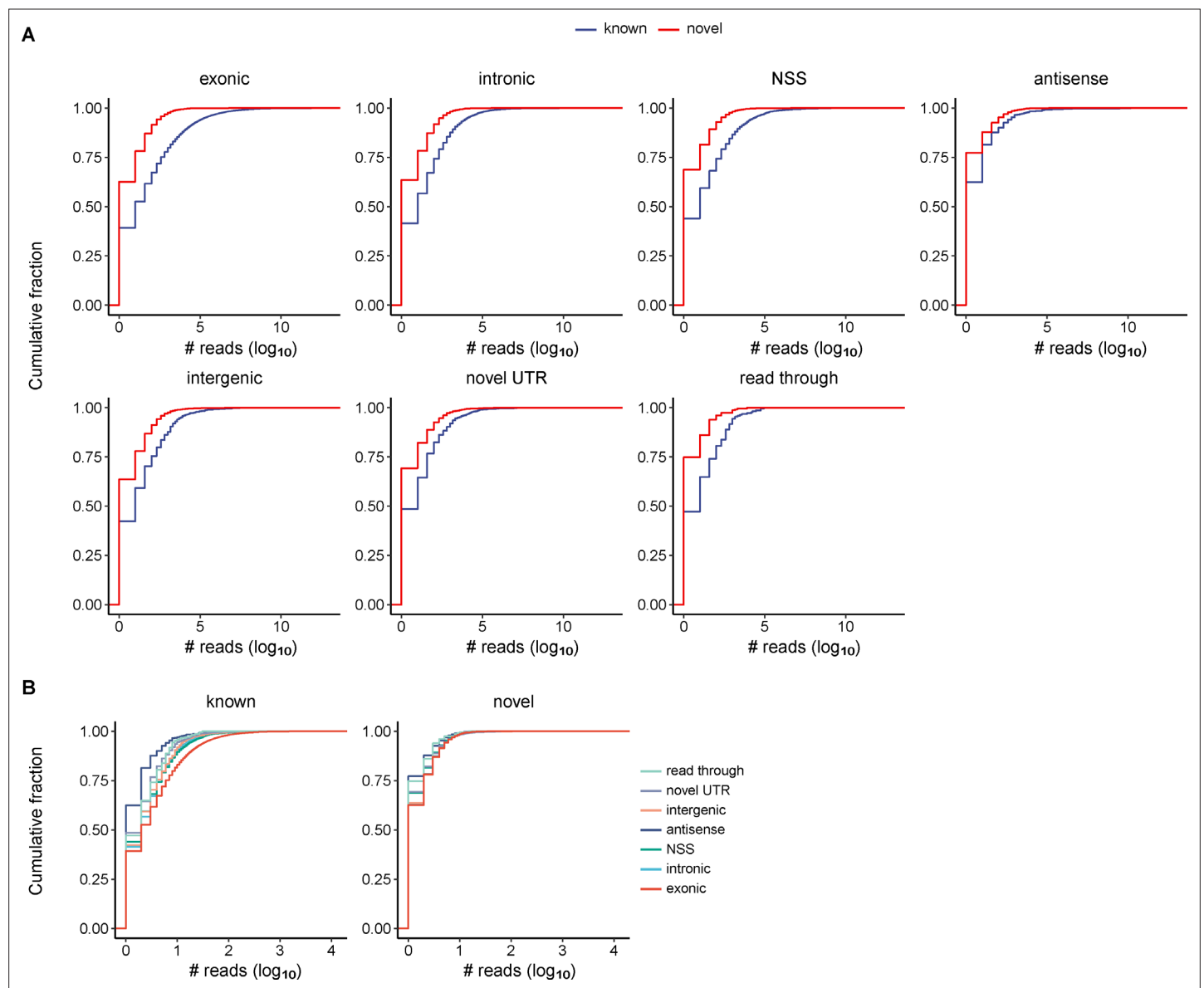


Figure 2—figure supplement 5. Cumulative distribution of read counts for circRNA isoforms identified by circFL-seq from six cell lines. CircRNA isoforms were classified based on database status and annotation types. **(A)** For each annotation type (exonic, intronic, NSS, intergenic, novel UTR, antisense, and read-through), the cumulative distribution of read counts is classified to known and novel status based on whether their BSJs are annotated in the circRNA database. **(B)** For known and novel status, the cumulative distribution of read counts is classified into seven annotation types. BSJ, back-splicing junction; circRNA, circular RNA.

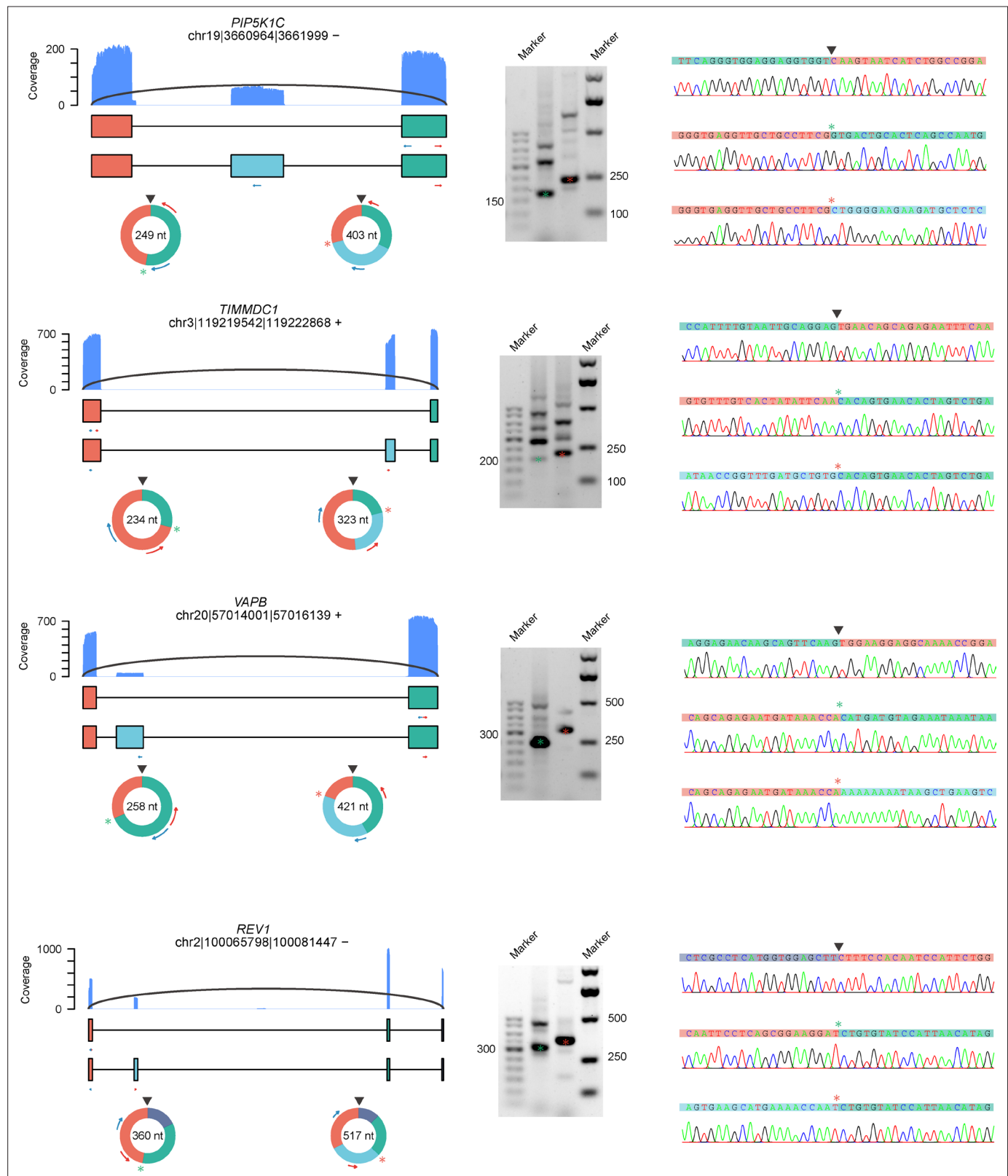


Figure 2—figure supplement 6. CircRNAs with exon skipping validated by RT-PCR and Sanger sequencing in HeLa cells. RT-PCR was performed with RNase R-treated RNA. The coverage of full-length circRNA reads mapped to the reference genome is shown. circRNA, circular RNA.

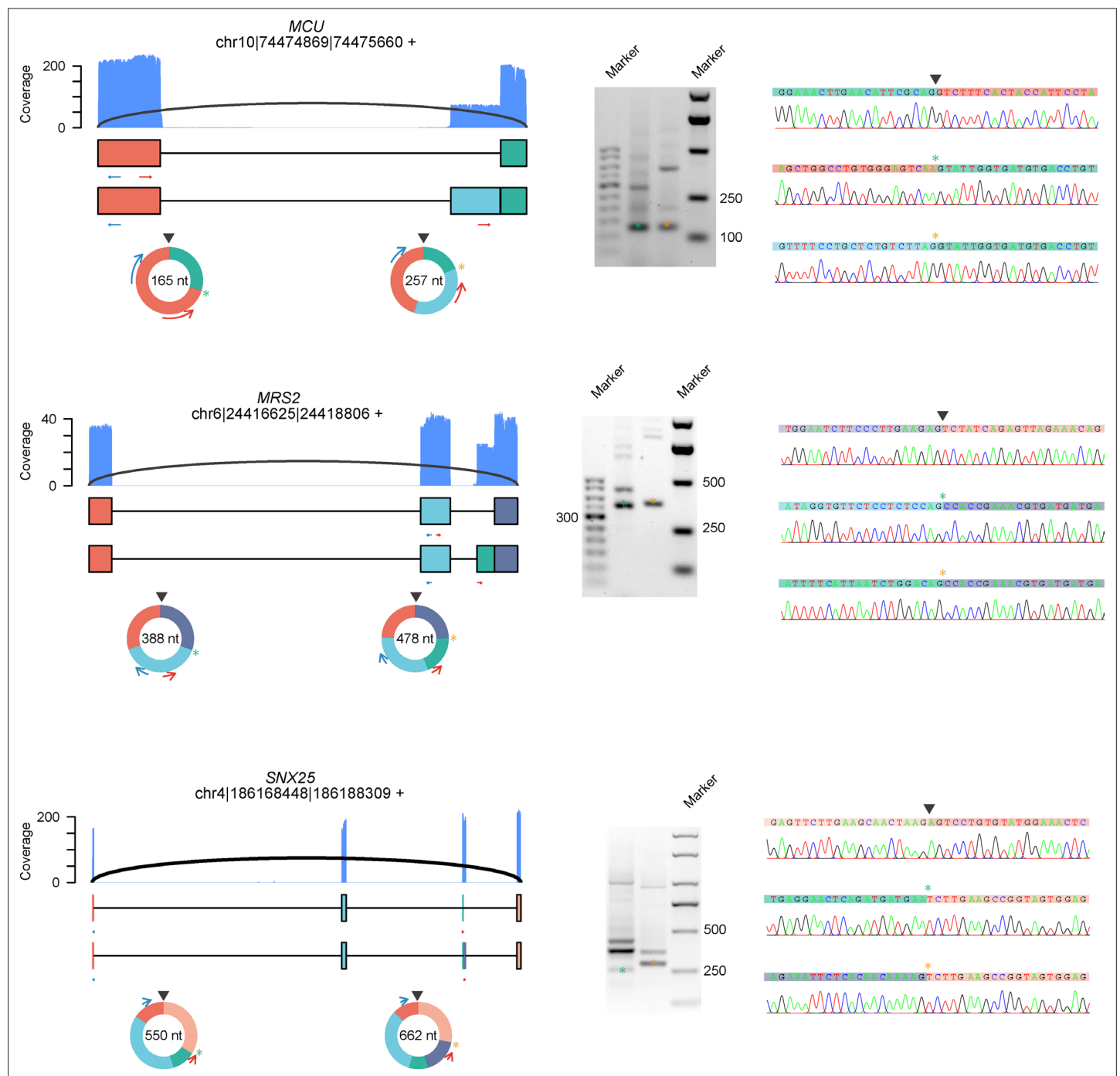


Figure 2—figure supplement 7. CircRNAs with alternative 3'/5' splicing sites (A3SS for circRNAs from *MCU* and *MRS2*, A5SS for circRNA from *SNX25*) were validated by RT-PCR and Sanger sequencing in HeLa cells. RT-PCR was performed with RNase R-treated RNA. The coverage of full-length circRNA reads mapped to the reference genome is shown. A3SS, alternative 3' splice site; alternative 5' splice site; circRNA, circular RNA.

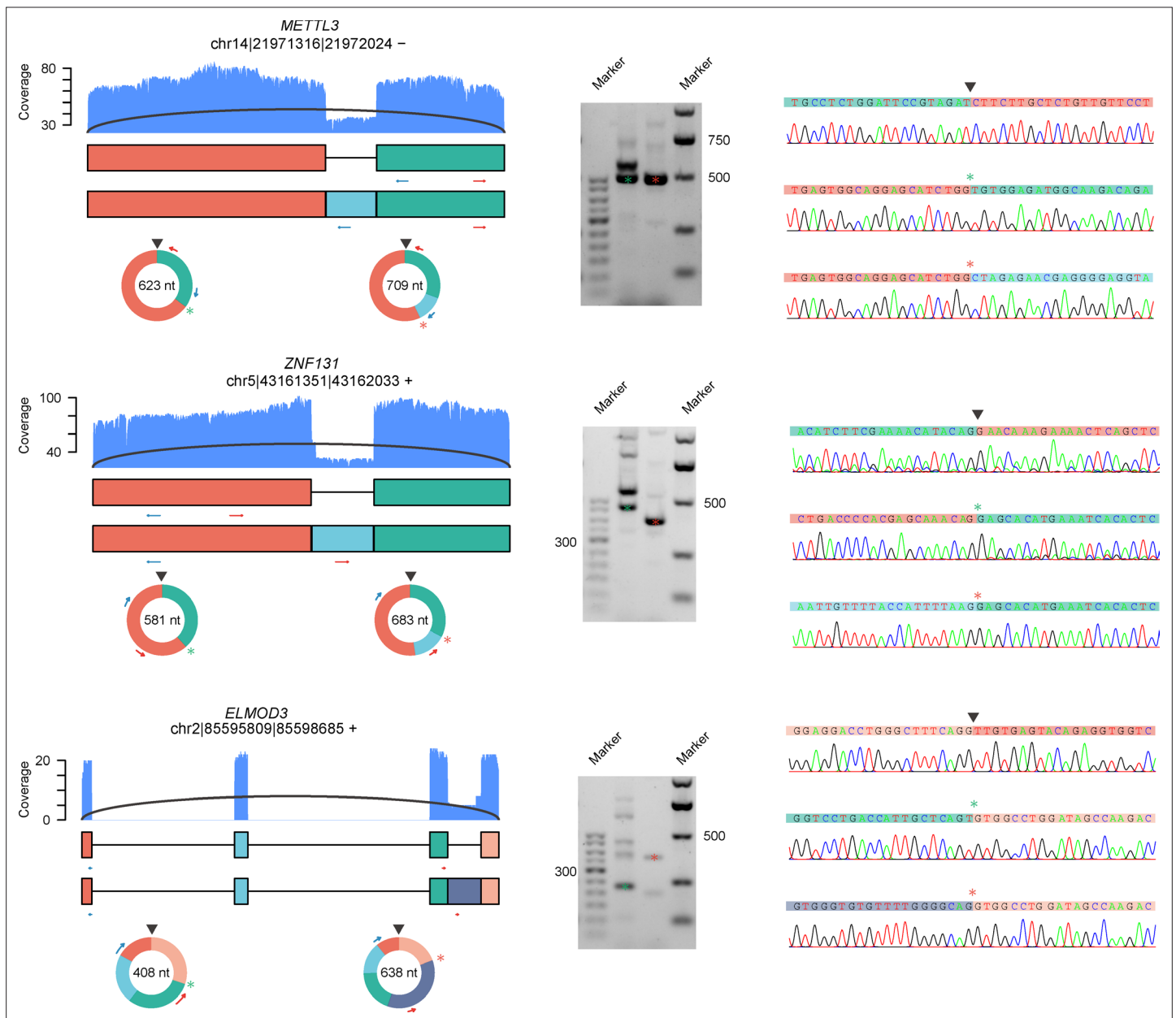


Figure 2—figure supplement 8. CircRNAs with intron retention validated by RT-PCR and Sanger sequencing in HeLa cells. RT-PCR was performed with RNase R-treated RNA. The coverage of full-length circRNA reads mapped to the reference genome is shown. circRNA, circular RNA.

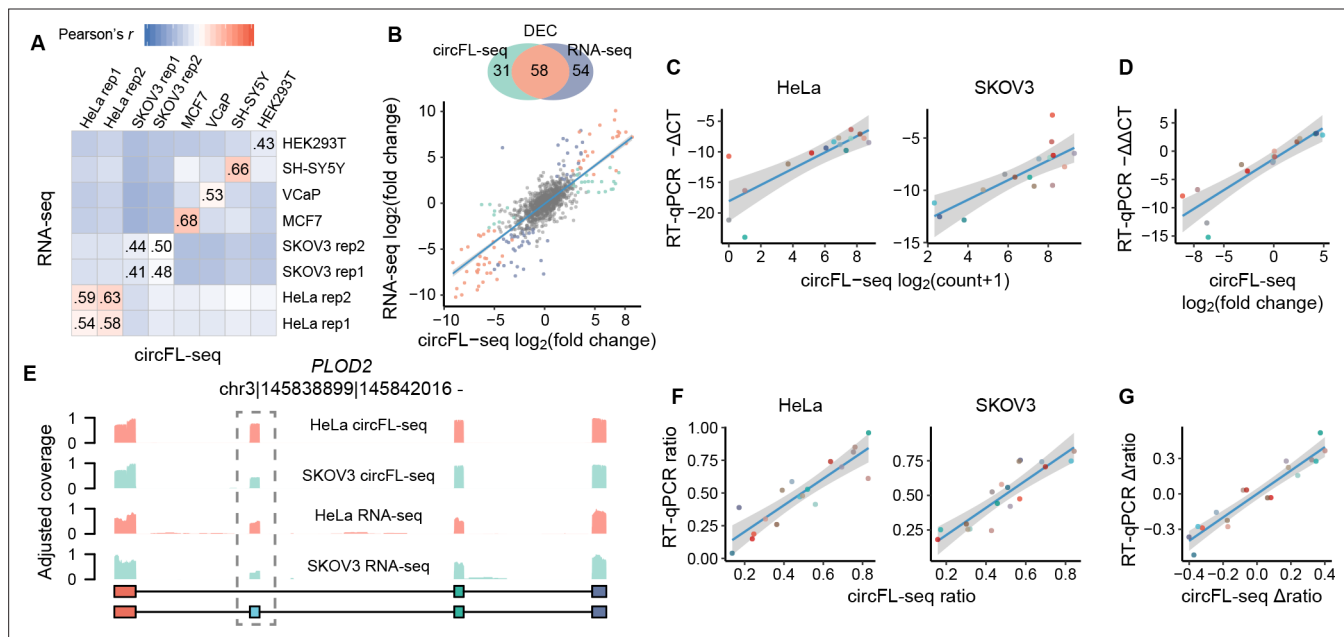


Figure 3. Quantification of circRNA at the BSJ and isoform levels. **(A)** Expression correlation matrix of circRNA BSJ quantified by circFL-seq and RNA-seq for six cell lines. The numbers in the matrix represent Pearson's correlation coefficients. **(B)** Comparison of differentially expressed circRNA (DEC) detection between circFL-seq and RNA-seq. Top panel: Venn diagram showing the number of DECs detected by circFL-seq (green), RNA-seq (purple), and both methods (orange). Bottom panel: scatter plot showing the correlation of fold change (log base 2) for HeLa and SKOV3 cells between circFL-seq and RNA-seq. **(C)** Scatter plot showing the correlation of the expression levels of 16 circRNA BSJs for HeLa (left) and SKOV3 (right) cells between circFL-seq and RT-qPCR. **(D)** Scatter plot showing the correlation of fold changes (log base 2) of the 16 BSJs for HeLa and SKOV3 cells between circFL-seq and RT-qPCR. **(E)** Plot showing the adjusted coverage of full-length circRNA reads and RNA-seq reads in the position of circRNA from *PLOD2*. The circular structures of the two circRNA isoforms are shown in the lower panel. **(F)** Scatter plot showing the correlation of the transcript ratio of 18 circRNA isoforms from nine circRNA BSJs (each BSJ has two isoforms) for HeLa (left) and SKOV3 (right) cells between circFL-seq and RT-qPCR. The relative expression of target BSJs/isoforms quantified by RT-qPCR was determined with RNase R-treated samples and *GAPDH* from total RNA without RNase R treatment as a reference. **(G)** Scatter plot showing the correlation of the differential ratio (Δratio) of the 18 isoforms for HeLa and SKOV3 cells between circFL-seq and RT-qPCR. The shaded areas denote 95 % confidence intervals. BSJ, back-splicing junction; circRNA, circular RNA.

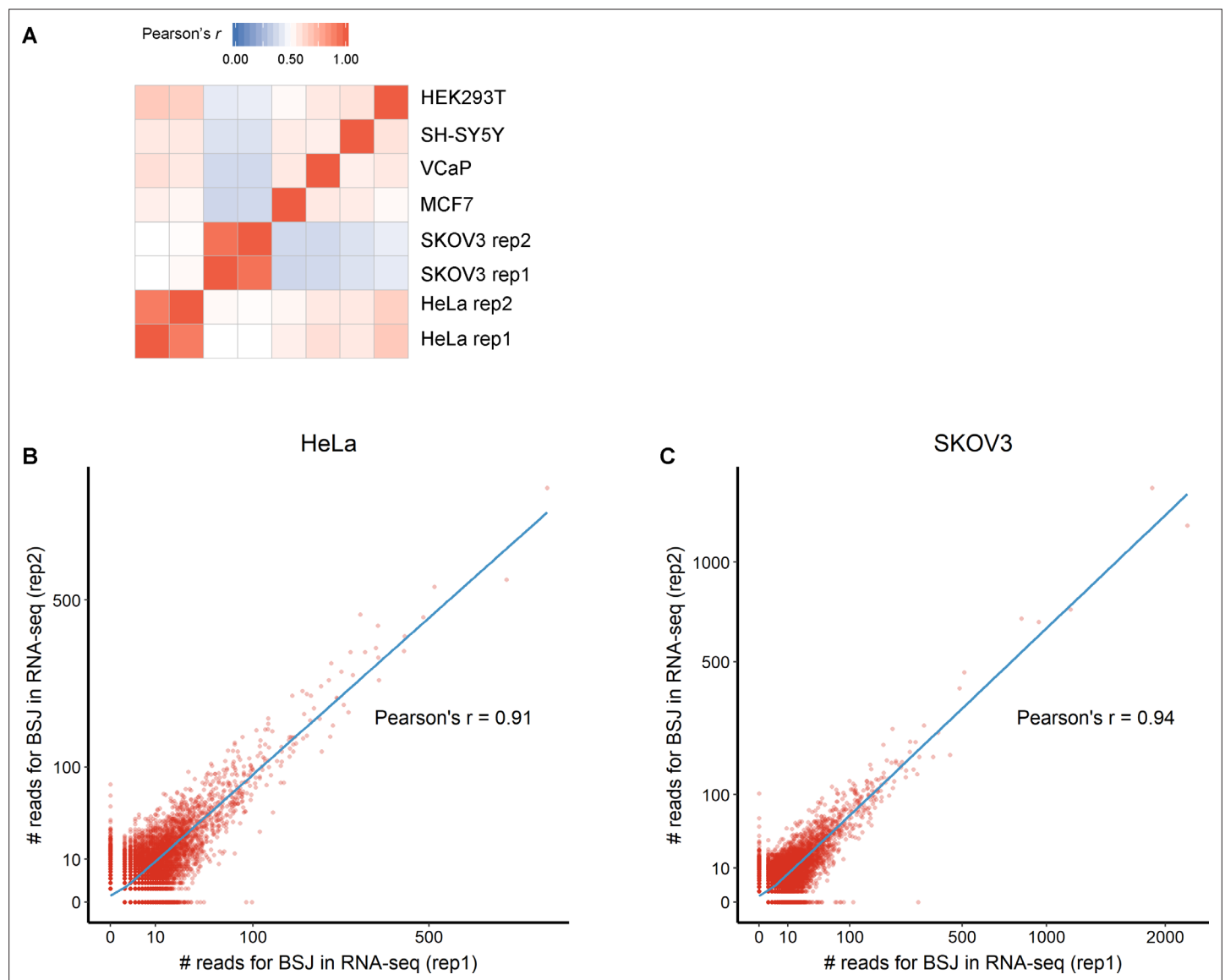


Figure 3—figure supplement 1. Correlations of circRNA BSJs among RNA-seq samples from six cell lines. **(A)** Expression correlation matrix for circRNA BSJs among six cell lines. The color scale corresponds to Pearson's correlation coefficients. Scatter plot showing the correlations of BSJs between HeLa **(B)** or SKOV3 **(C)** replicates. CircRNA BSJs with read counts > 0 in at least one replicate were included.

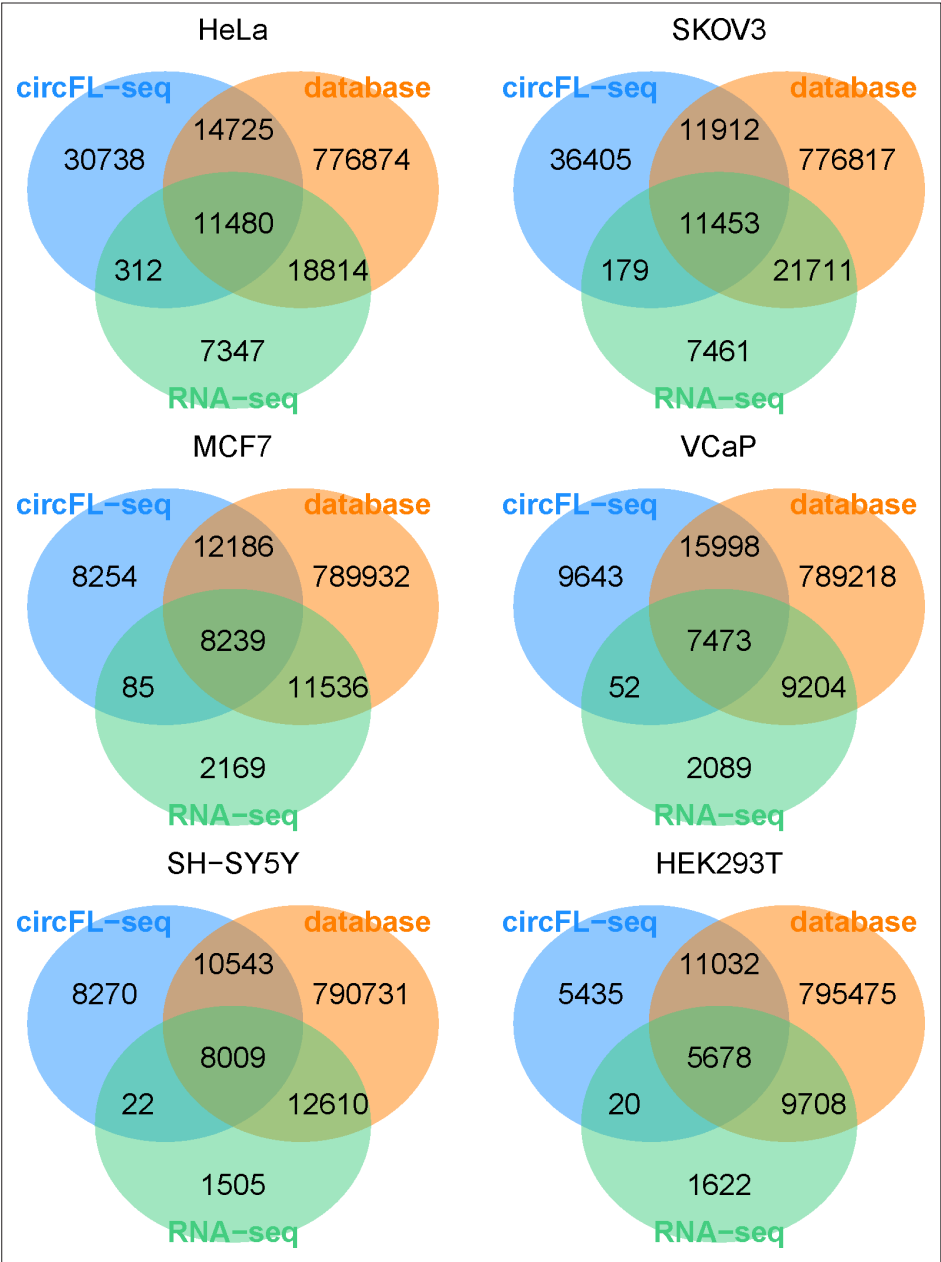
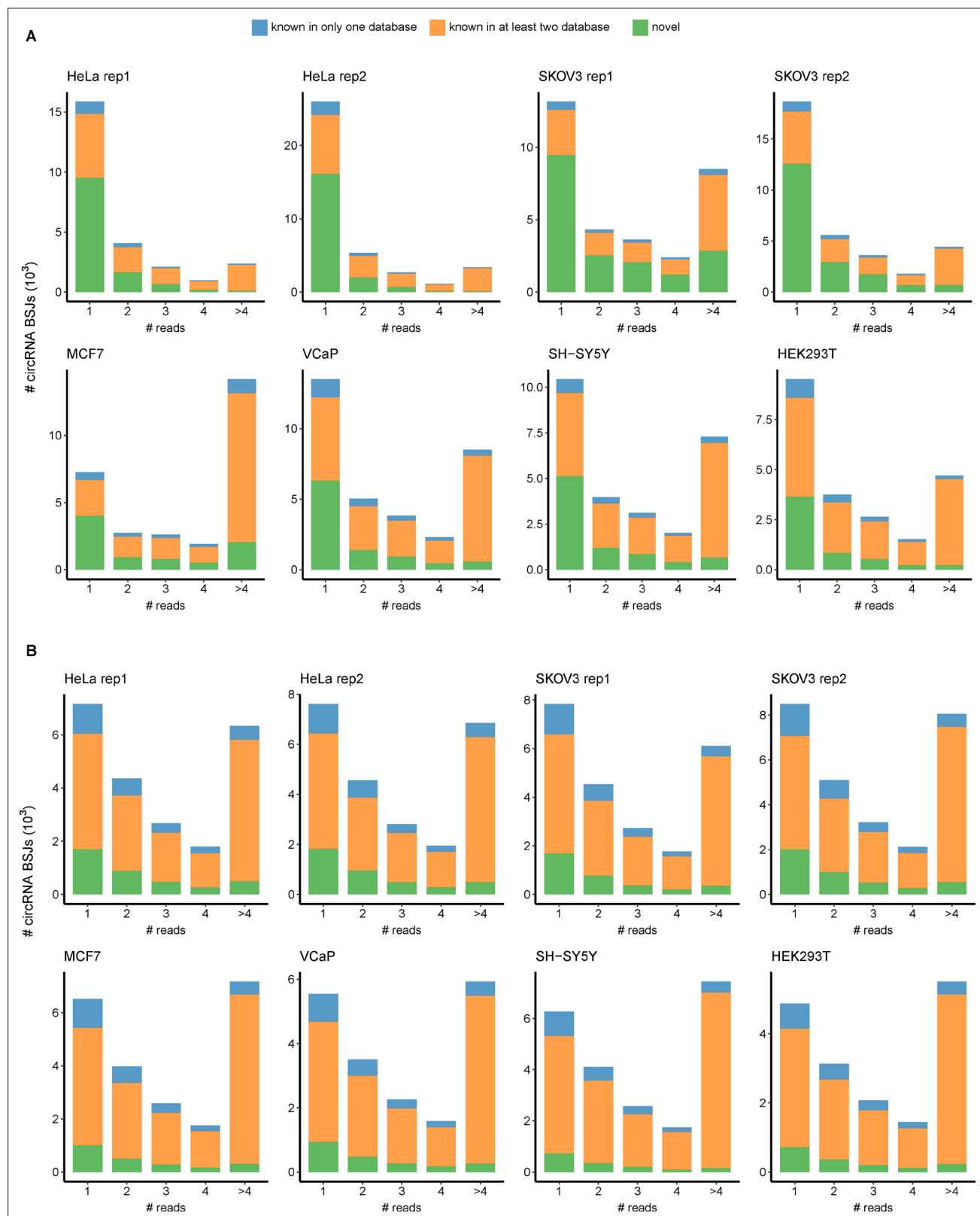


Figure 3—figure supplement 2. Venn diagram of BSJs detected by circFL-seq, RNA-seq, and database. BSJ, back-splicing junction.



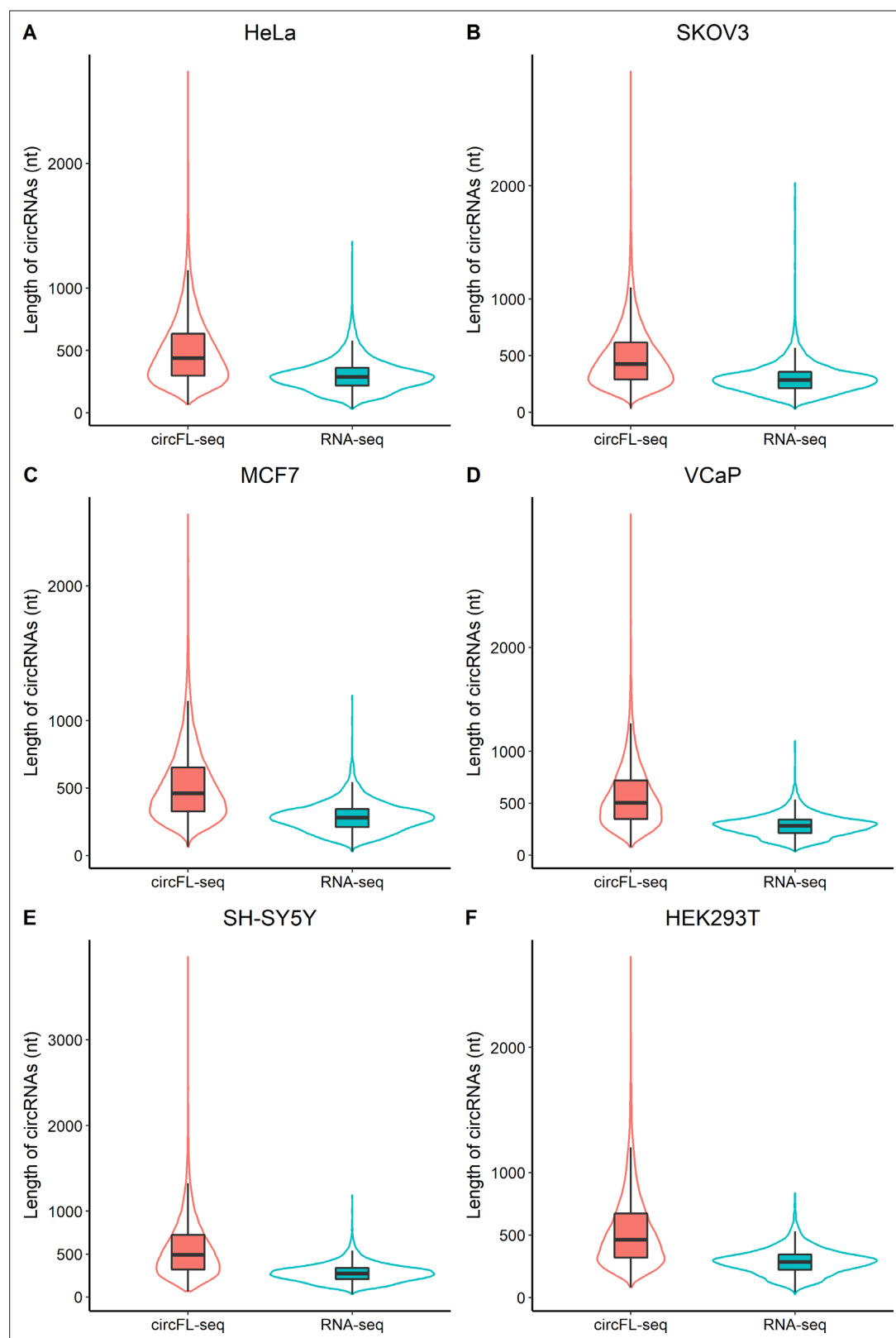


Figure 3—figure supplement 4. Comparison of circFL-seq and RNA-seq for length of full-length circRNA of six cell lines. For RNA-seq, full-length circRNAs were reconstructed by CIRI-full. circRNA, circular RNA.

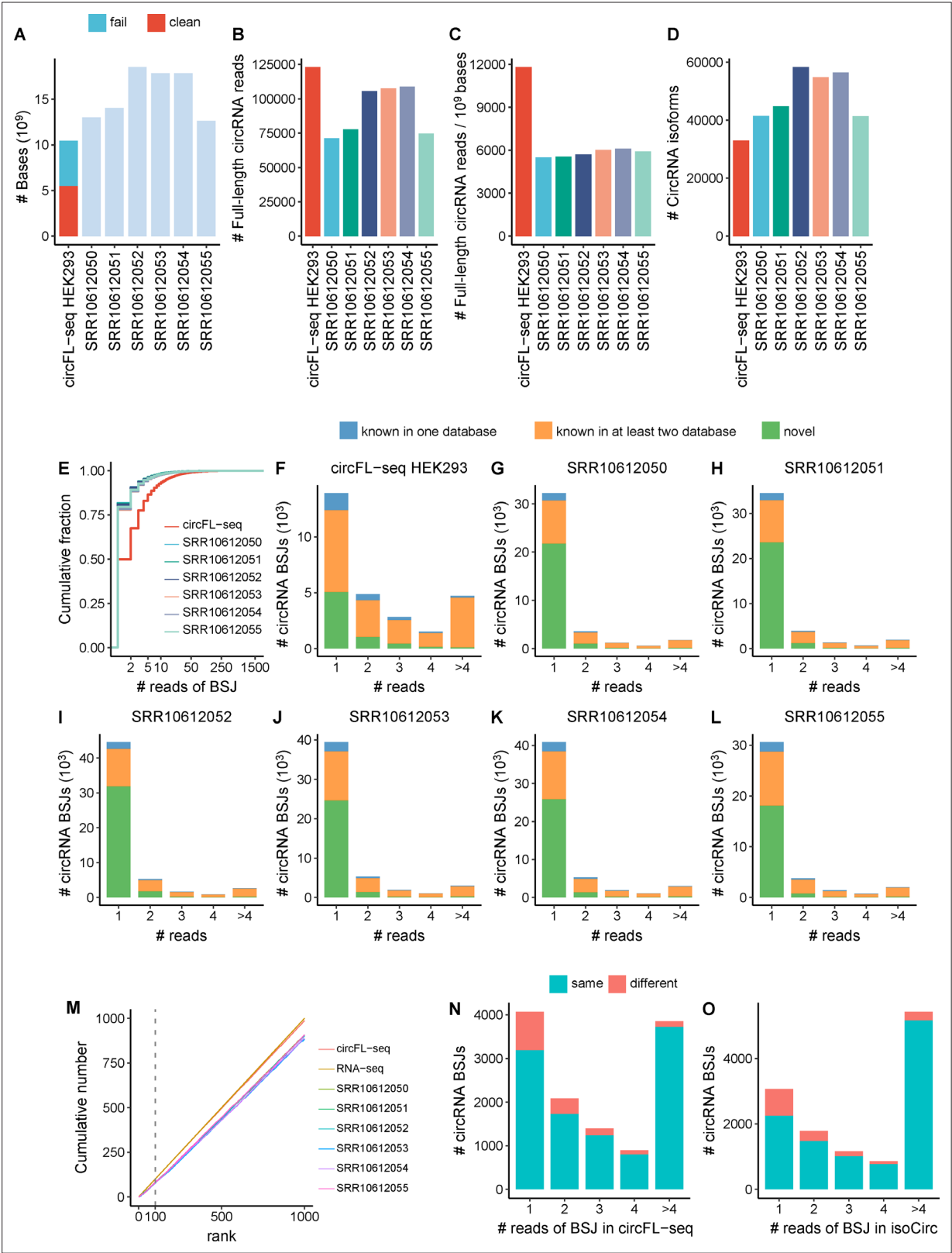


Figure 3—figure supplement 5. Comparison of circFL-seq and isoCirc for full-length circRNA detection in the HEK293 cell line. **(A)** Stacked bar plot showing the number of sequenced bases. circFL-seq includes fail and clean bases. Fail bases are from low-quality reads with qscore <7 and trimmed adapters. **(B)** Bar plot showing the number of full-length circRNA reads. **(C)** Bar plot showing full-length circRNA reads per 10⁹ raw sequenced bases. **(D)** Bar plot showing the number of full-length circRNA isoforms. **(E)** Cumulative distribution of read counts of BSJs. **(F–L)** Stacked bar plot showing the

Figure 3—figure supplement 5 continued on next page

Figure 3—figure supplement 5 continued

distribution of known or novel circRNA BSJs for different read counts. **(M)** Plot showing the cumulative number of top expressed circRNAs of circFL-seq, isoCirc, and RNA-seq detected in the database. Stacked bar plot showing the distribution of read counts of common circRNA BSJs detected in circFL-seq **(N)** and isoCirc **(O)**. 'Same' and 'different' represent BSJs w/wo the same isoforms between circFL-seq and isoCirc. The isoCirc analysis in **(N, O)** combines circRNA results from all six HEK293 isoCirc libraries. BSJ, back-splicing junction; circRNA, circular RNA.

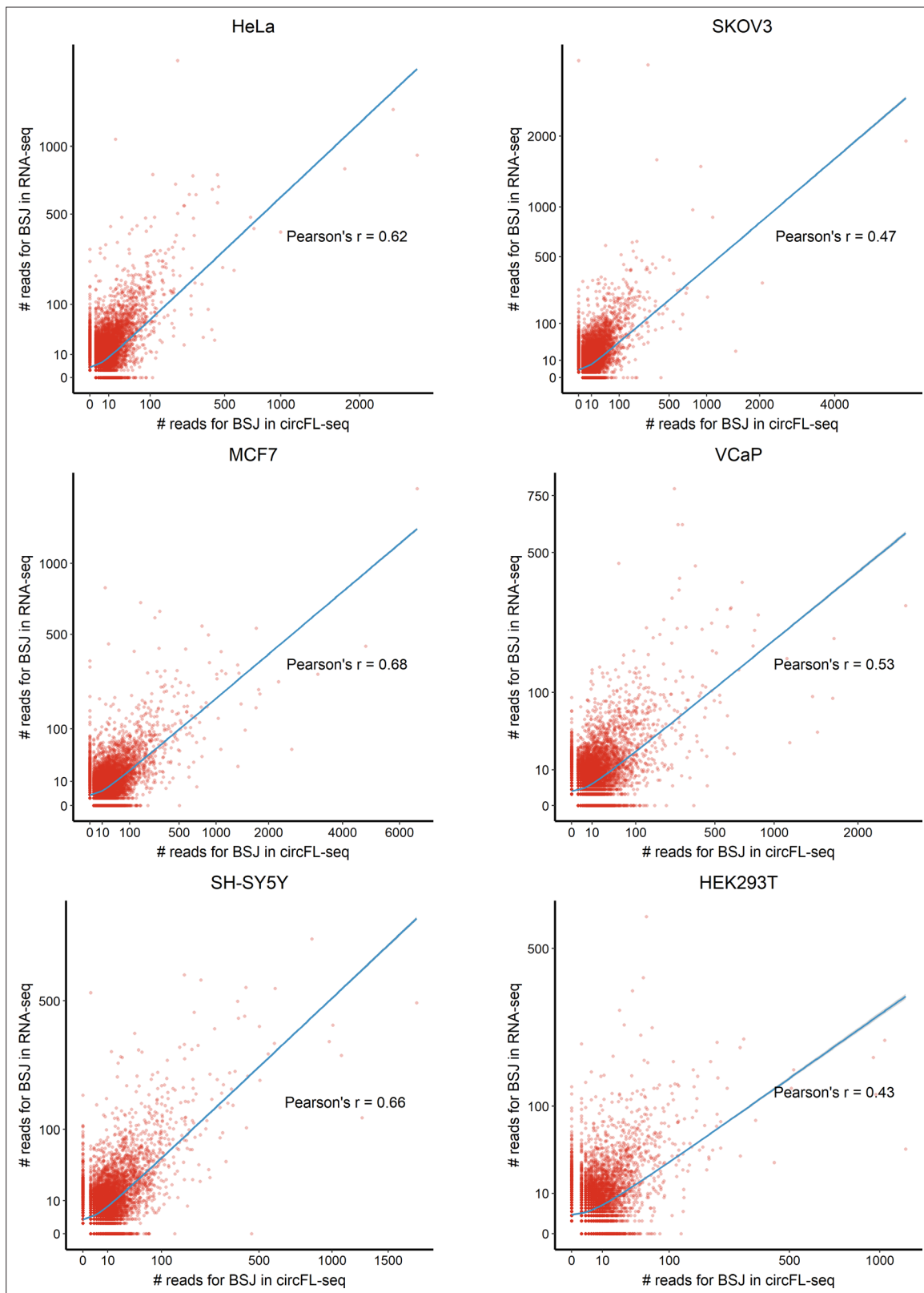


Figure 3—figure supplement 6. Scatter plot showing the correlation of circRNA BSJs between circFL-seq and RNA-seq samples of six cell lines. BSJ, back-splicing junction; circRNA, circular RNA.

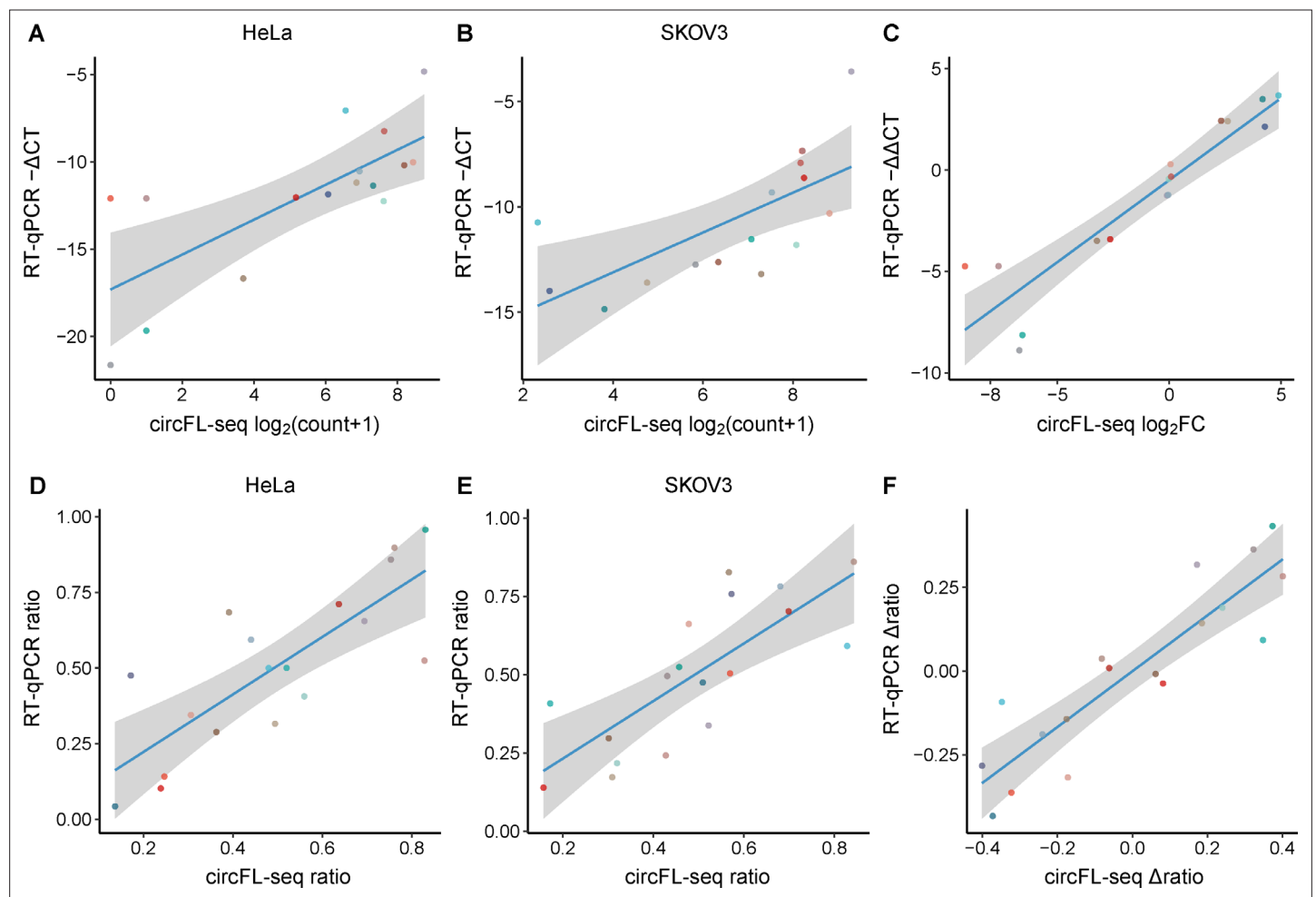


Figure 3—figure supplement 7. Evaluation of circRNA quantification between circFL-seq and RT-qPCR. The relative expression of target BSJs and isoforms quantified by RT-qPCR was performed with samples without RNase R treatment. *GAPDH* from total RNA without RNase R treatment was used as a reference. (A, B) Scatter plot showing the correlation of expression levels of 16 circRNA BSJs for HeLa (A) and SKOV3 (B) cells between circFL-seq and RT-qPCR. (C) Scatter plot showing the correlation of fold change (log base 2) of the 16 BSJs for HeLa and SKOV3 cells between circFL-seq and RT-qPCR. (D, E) Scatter plot showing the correlation of the transcript ratio of 18 circRNA isoforms from 9 circRNA BSJs (each BSJ has two isoforms) for HeLa (D) and SKOV3 (E) cells between circFL-seq and RT-qPCR. (F) Scatter plot showing the correlation of the differential ratio (Δ ratio) of the 18 isoforms for HeLa and SKOV3 cells between circFL-seq and RT-qPCR. BSJ, back-splicing junction; circRNA, circular RNA.

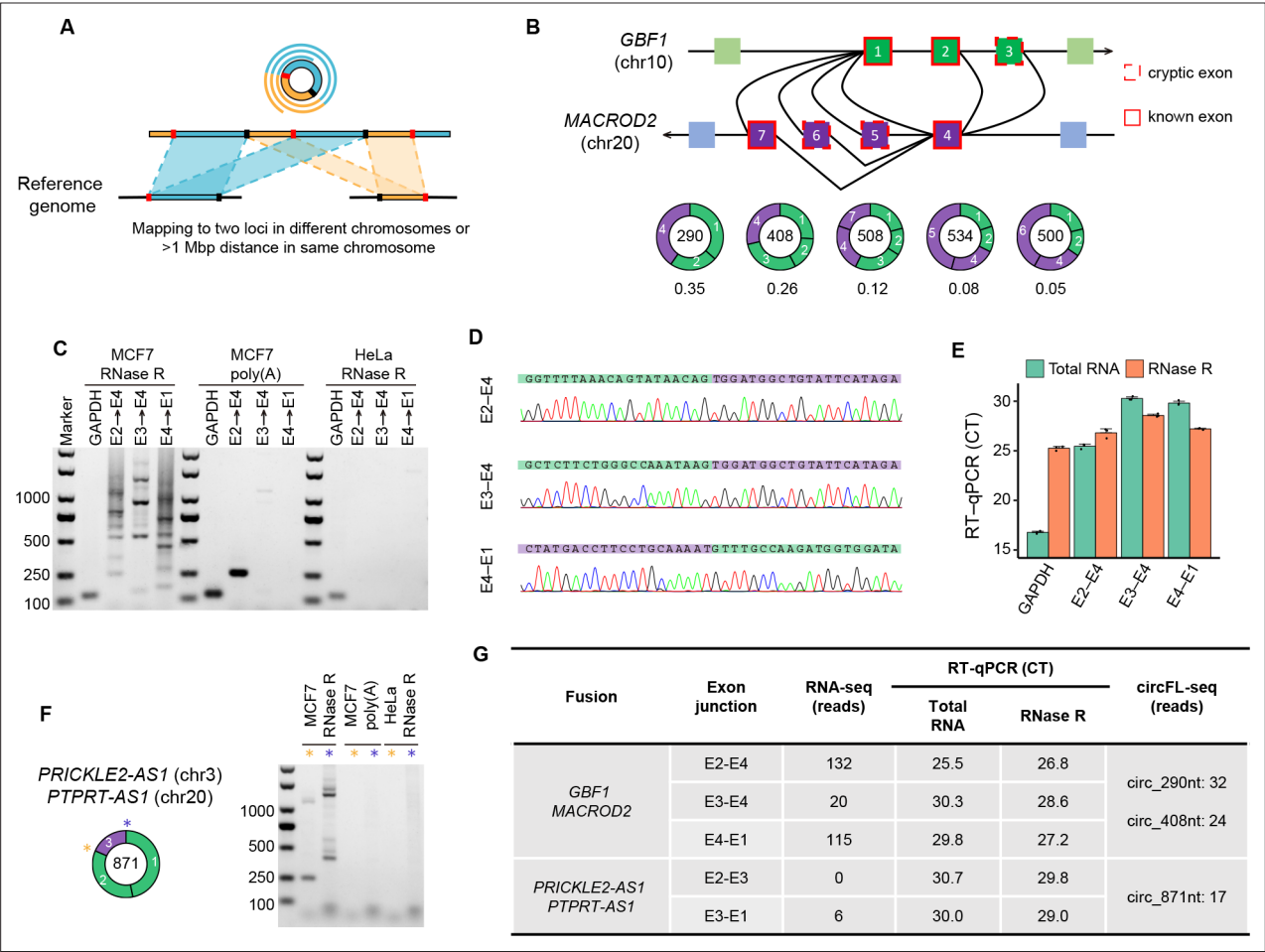
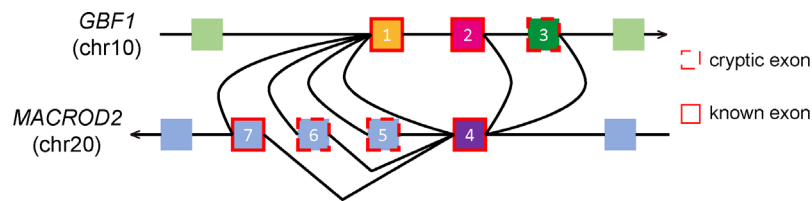


Figure 4. Detection and validation of fusion circRNA (f-circRNA) in the MCF7 cell line. **(A)** Diagram of identification of f-circRNA with circFL-seq data. **(B)** Diagram of five high-quality f-circRNA isoforms (read counts \geq 5) fused by *GBF1* and *MACROD2*. The transcript ratio represents the fractions of the isoforms. **(C–E)** Validation of f-circRNA junctions from *GBF1/MACROD2* by agarose gel electrophoresis **(C)**, Sanger sequencing **(D)**, and RT-qPCR **(E)**. **(C)** Agarose gel electrophoresis showing the RT-PCR products of f-circRNA junctions with RNase R-treated MCF7 and HeLa RNA and poly(A) selected MCF7 RNA as a template. **(F)** Agarose gel electrophoresis showing the RT-PCR products of f-circRNA junctions from *PRICKLE2-AS1/PTPRT-AS1*. **(G)** Information on five f-circRNA junctions detected by circFL-seq, RNA-seq, and RT-qPCR.



E1-E2-E4, circRNA length: 290 nt

The Sanger sequencing validation sequences:

GGGGGGGGGTTGCCAAGATGGTGGATAAGAAATATTACATCATTCAAGGGG
 AGATTAAACATTGTGGTTGGGGCCATCAAACGAAATGCCCGATGGAGCACCC
 ATACACCACTGGATGAAGAACGGGATCCTCTGCTGCATAGTTTCGGTCATCT
 AAAGGAGGTTTTAAACAGTATAACAGTGGATGGCTGTATTCATAGAGCAGC
 CGGCCCCCTGTTTGCTAGCTGAATGTCGTAACCTGAATGGCTGTGATACTGG
 ACATGCAAAAATCACATGTGGCTATGACCTTCCTGCAAAAATGTTTGCCAAG
 ATGGTGGATAAGAAATATTACATCATTCAAGGGGAGATTAAACATTGTGGTTG
 GGGCCATCAAACGAAATGCCCGATGGAGCACCCATACACCACTGGATGAA
 GAACGGGATCCTCTGCTGCATAGTTTCGGTCATCTAAAGGAGGTTTTAAAC
 AGTATAACAGTGGATGGCTGTATTCATAGAGCAGCCGGCCCCCTGTTTGCTAG
 CTGAATGTCGTAACCTGAATGGCTGCCCCCCCC

E1-E2-E3-E4, circRNA length: 408 nt

The Sanger sequencing validation sequences:

GGGGGGGGGTGGATCTTCTGACATGCTTCTCTCCTGAAATAACTGGATGGCT
 CTTCTGGGCCAAATAAGTGGATGGCTGTATTCATAGAGCAGCCGGCCCCCTG
 TTTGCTAGCTGAATGTCGTAACCTGAATGGCTGTGATACTGGACATGCAAA
 AATCACATGTGGCTATGACCTTCCTGCAAAAATGTTTGCCAAGATGGTGGATA
 AGAATATTACATCATTCAAGGGGAGATTAAACATTGTGGTTGGGGCCATCAA
 ACGAAATGCCCGATGGAGCACCCATACACCACTGGATGAAGAACGGGATC
 CTCTGCTGCATAGTTTCGGTCATCTAAAGGAGGTTTTAAACAGTATAACAGC
 TTCAAAGCTTTTGAGGCTGGATAGGACTTTGATGATAAAGCTCCTTTCAG
 GTCTTATGGATCTTCTGACATGCTTCTCTCCTGAAATAACTGGATGGCTCTT
 CTGGGCCAAATAAGTGGATGGCTGTATTCATAGAGCAGCCGGCCCCCTGTTT
 GCTAGCTGAATGTCGTAACCTGAATGGCTGCCCCCCCC

Figure 4—figure supplement 1. Sanger validation of sequences of f-circRNA from GBF1/MACROD2. The forward and reverse primers are highlighted. f-circRNA, fusion circRNA.

PRICKLE2-AS1 (chr3):
PTPRT-AS1 (chr20)



E1-E2-E3, circRNA length: 871 nt

The Sanger sequencing validation sequences:

GGGGGGGAACTAGAACGGTCGTGTCCCTAATTAAAGGCCACGGGCCCAAT
 TCTAGCTGGTTCTACTTGTGACTCATTGAGTGATCTTGGGAAAACCTCTCTGC
 AACTCAAATGTTTCATTTTAAAGATGAGAAGTTTGGATCAATAACATCTAAGG
 TGTGGTGTGTTTGTCTGCCTCCTGGGAAGCAGCTTTCAAACTCATTGTTT
 AAGTGTCTTGGGAATGAAAAAGACAAACCATCCCGTCAAGGGAACAATAG
 TTTCTACTTGGGGGGAAAGTACATCTTGCCAATTTTCAGTTTGCATTGTCCCA
 AACATGTTTGTAAACTCTACCCTGAATTATCTGTGATGAATACAGTATAAA
 GTTAACCCCCCACATGGAATGACTAACAAAAGCAGATGCGAAACAGCTAA
 CAAAGAAAGCAAACACAAAGATAAAACAGCATGCGTCAAATCCTATGTCA
 ATCATAAGAAAATGTTTTTCACATCTAATCATCGAAGATTAAGGGCTGGATA
 TTTAAGGAATGTGAAAGAGGATGTCCTTCAGCAAACAATTTAAATAAGGCA
 AGAAATGAACGTGATATTTTCTCCGCCTCCTCTTCTGACTGAGAAGATGA
 TTCCTGGAGATAATCCACTTGTTTATCCGTGGATGTGAACATAATTTGGAGG
 CAGCAGTCACTCCAGATGGCCCGCTGAAGCTGGGAGTCCTGAGTTAATTC
 AAGCCAAATTTCTCACTCCCTGGAGGAGCAGAGTGGAGGGTGTGTGTGCA
 TGGAGAAGTCCAAGATTTTCATATCTGGAAAAGAAGACTGGGAGAGGCCAG
 CATGAATGGCCACTGTCCTCGCCAAATCTGGATGGTATGTCTTAGGGTACCG
 TGGGACAAAACCTAGAACGGTCGTGTCCCTAATTAAAGGCCACGGGCCCAA
 TTCTAGCTGGTTCTACTTGTGACTCATTGAGTGATCTTGGGAAAACCTCTCTG
 CAACTCAAATGTTTCATTTTAAAGATGAGAAGTTTGGATCAATAACATCTAAG
 GTGTGGTGTGTTTGTCTGCCTCCTGGGAAGCAGCTTTCAAACTCATTGTT
 CAAGTGTCTTGGGAATGAAAAAGACAAACCATCCCGTCAAGGGAACAATA
 GTTTCTACTTAGGGGGAAAGTACATCTTGCCAATTCAGTTTGCATTGTCCC
 AAACATGTTTGTAAACTCTACCCTGAATTATCTGTGATGAATACAGTATAA
 AGTTAACCCCCCACATGGAATGACTAACAAAAGCAGATGCGAAACAGCTA
 ACAACCCCCCCCCC

Figure 4—figure supplement 2. Sanger validation of the sequence of f-circRNA from *PRICKLE2-AS1/PTPRT-AS1*. The forward and reverse primers are highlighted. f-circRNA, fusion circRNA.