Supplementary Materials for circMeta: a unified computational framework for genomic feature annotation and differential expression analysis of circular RNAs

Derivation of variance of beta-binomial distribution using mean and dispersion parameterization in the proposed hierarchical model

$$\begin{aligned} var(x_{ijl}) &= E(var(x_{ijl}|\mu_{ij})) + var(E(x_{ijl}|\mu_{ij})) \\ &= E(n_{ijl}\mu_{ij}(1-\mu_{ij})) + var(n_{ijl}\mu_{ij}) \\ &= n_{ijl}E(\mu_{ij}(1-\mu_{ij}) + n_{ijl}var(\mu_{ij}) \\ &= n_{ijl}\mu_{ij}(1-\mu_{ij}) + n_{ijl}\rho_{ij}\mu_{ij}(1-\mu_{ij}) \\ &= n_{ijl}\mu_{ij}(1-\mu_{ij})(1+\rho_{ij}) \end{aligned}$$

Thus, $\hat{var}(x_{ijl}) = n_{ijl}\hat{\mu}_{ij}(1-\hat{\mu}_{ij})(1+\hat{\rho}_{ij})$. We plug in $\hat{\mu}_{ij}$ and shrunk estimate $\hat{\rho}_{ij}$ to obtain the estimate of $\hat{var}(x_{ijl})$.
The MOM estimate of μ_{ij} is μ_{ij} as $\hat{\mu}_{ij} = \frac{\sum_l x_{ijl}}{\sum_l n_{ijl}}$. Thus, we have

$$\begin{aligned} var(\hat{\mu}_{ij}) &= var(\frac{\sum_l x_{ijl}}{\sum_l n_{ijl}}) \\ &= (\frac{1}{\sum_l n_{ijl}})^2 var(\sum_l x_{ijl}) = (\frac{1}{\sum_l n_{ijl}})^2 \sum_l var(x_{ijl}) = (\frac{1}{\sum_l n_{ijl}})^2 \sum_l (n_{ijl}\mu_{ij}(1-\mu_{ij})(1+\rho_{ij})) \end{aligned}$$

Thus, we could obtain the estimate of $\hat{var}(\hat{\mu}_{ij})$ by plugging in estimated $\hat{var}(x_{ijl})$.

RNA preparing and RT-qPCR

RNA samples from human fetal cortex and cerebellum were QC'ed by Nanodrop and Qubit. 500 ng of RNA was reverse transcribed using random hexamers and SuperScript III (Invitrogen 18080-051) according to the manufacturer's protocol. RT-qPCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific 4367659). Relative transcript levels were calculated using the Delta-Delta-CT method. The primers for RT-qPCR are shown in Table S3.

Figure

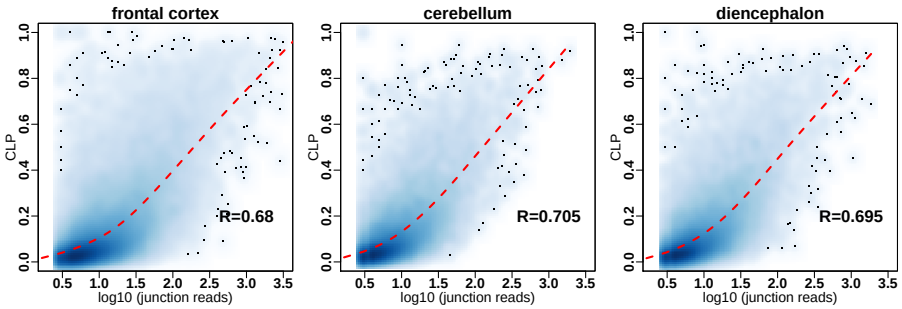
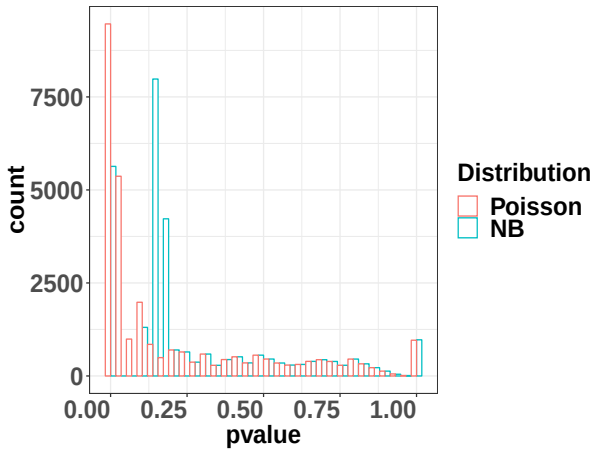
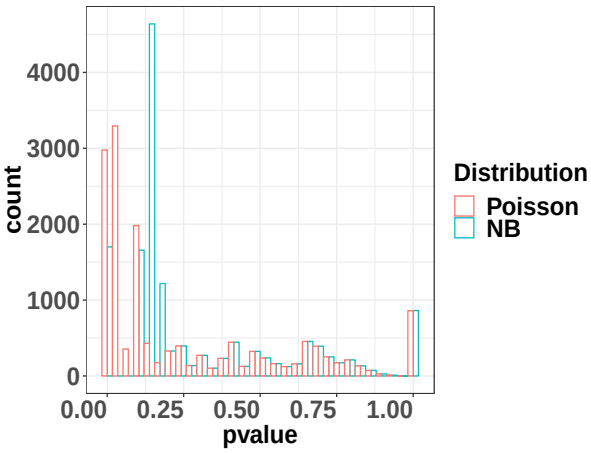


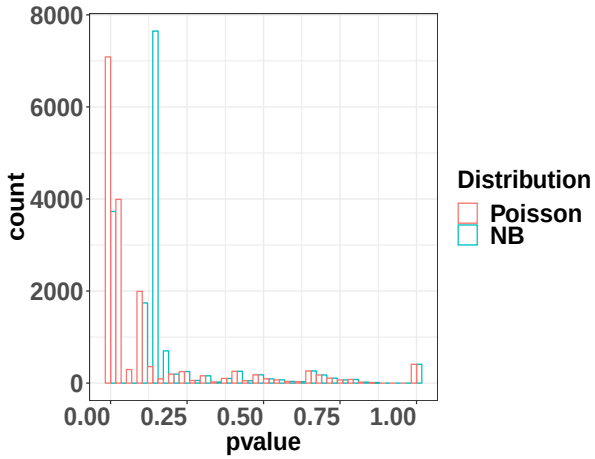
Figure S1: Correlation of logarithmic circular read counts and Circular-to-Linear Proportion (CLP) for common circRNAs identified by findcirc, CIRCexplorer and CIRI in frontal cortex, cerebellum and diencephalon respectively.



(a) p-values of GOF (frontal cortex)



(b) p-values of GOF (cerebellum)



(c) p-values of GOF (diencephalon)

Figure S2: p-values of Goodness Of Fit (GOF) for Poisson distribution and Negative binomial distribution for junction reads of common circRNAs identified by findcirc, CIRCexplorer and CIRI in frontal cortex, cerebellum and diencephalon respectively. The significant level 0.05 is indicated by a dash vertical line.

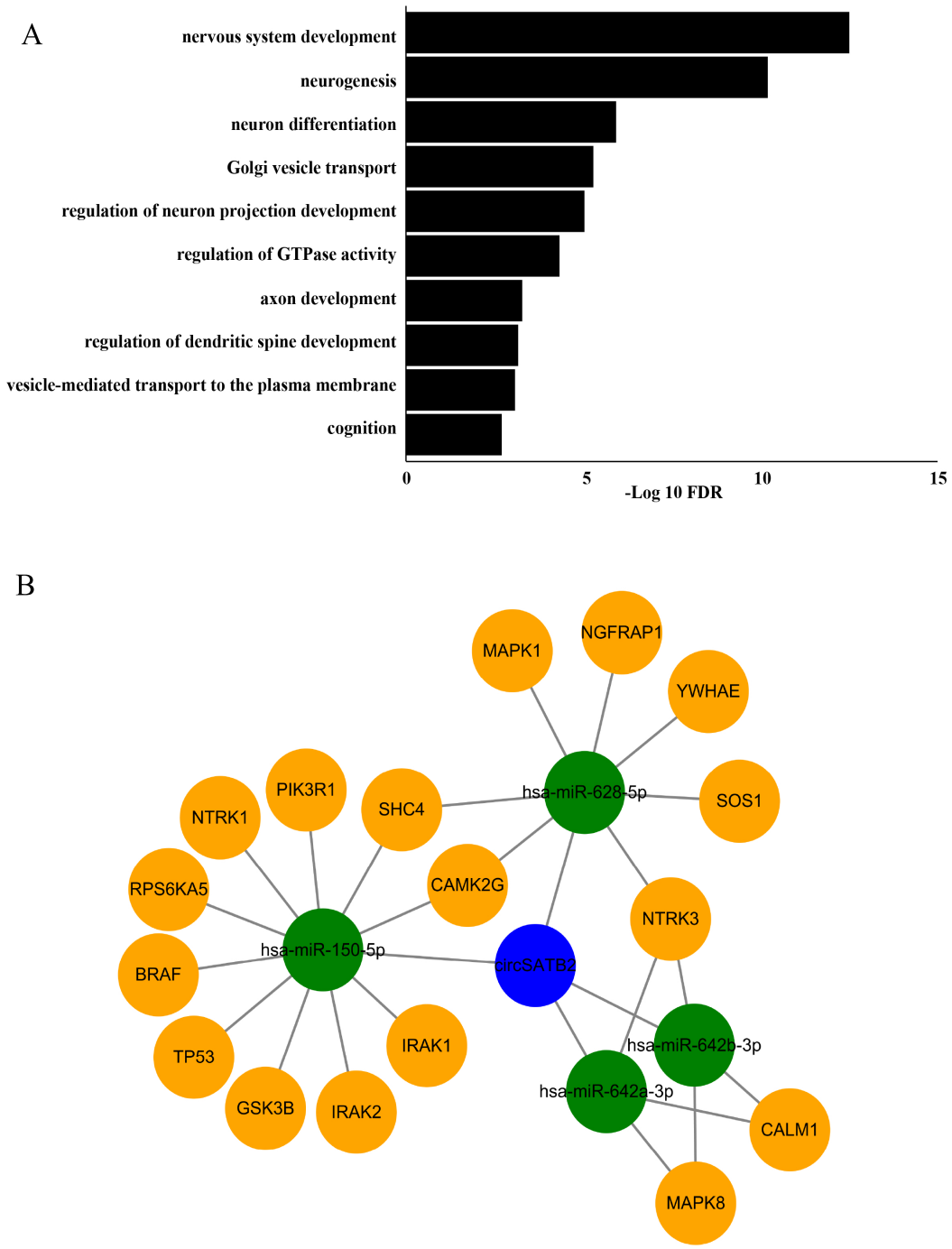


Figure S3: (A) Gene ontology (GO) analysis using host genes harboring circRNAs expressed higher in frontal cortex than cerebellum. Bar charts showing selected biological process ranked by false discovery rate (FDR) (FDR<0.05). (C) The putative miRNAs that could be sponged by circSATB2 are predicted by starBase database and indicated in dark green. The mRNAs regulated by miRNA changes are shown in yellow. miRNAs that are not expressed in corresponding tissues are marked in grey. The circRNA-miRNA-mRNA network is related to ErbB signaling pathway.

Table S1: Expression level (read counts) of miRNAs in cerebellum and frontal cortex associated with circRELL1 and circSATB2

	miRNA	cerebellum	frontal cortex	circRNA
1	hsa-miR-9	0	0	circRELL1
2	hsa-miR-186	34	26	circRELL1
3	hsa-miR-526b	3	2	circRELL1
4	hsa-miR-17	18	3	circRELL1
5	hsa-miR-20a	7	2	circRELL1
6	hsa-miR-106a	13	3	circRELL1
7	hsa-miR-106b	29	21	circRELL1
8	hsa-miR-20b	20	6	circRELL1
9	hsa-miR-93	41	39	circRELL1
10	hsa-miR-302e	2	2	circRELL1
11	hsa-miR-302a	2	2	circRELL1
12	hsa-miR-520e	3	4	circRELL1
13	hsa-miR-520b	2	2	circRELL1
14	hsa-miR-373	17	7	circRELL1
15	hsa-miR-337	11	15	circRELL1
16	hsa-miR-150	9	7	circSATB2
17	hsa-miR-3614	57	43	circSATB2
18	hsa-miR-628	12	20	circSATB2
19	hsa-miR-642b	3	2	circSATB2
20	hsa-miR-642a	12	9	circSATB2

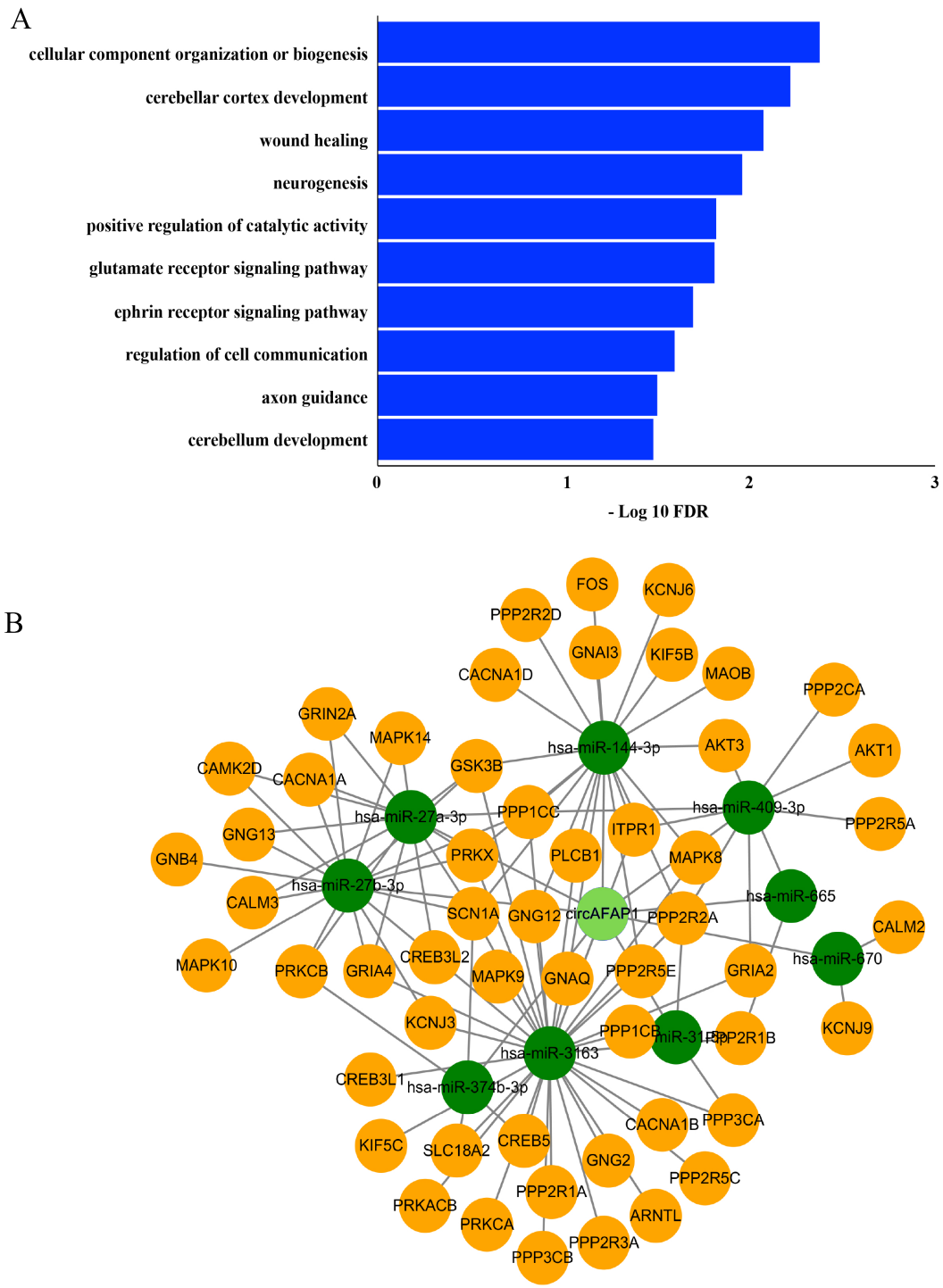


Figure S4: (A) Gene ontology (GO) analysis using host genes producing higher level of circRNAs in cerebellum than diencephalon. Bar charts showing selected biological process ranked by false discovery rate (FDR) (FDR<0.05). (B) The putative miRNAs that could be sponged by circAFAP1 are predicted by starBase database and indicated in dark green. The mRNAs networks that could be affected by miRNA changes is shown in yellow. miRNAs that are not expressed in corresponding tissues are marked in grey. The circRNA-miRNA-mRNA axis is related to dopaminergic synapse.

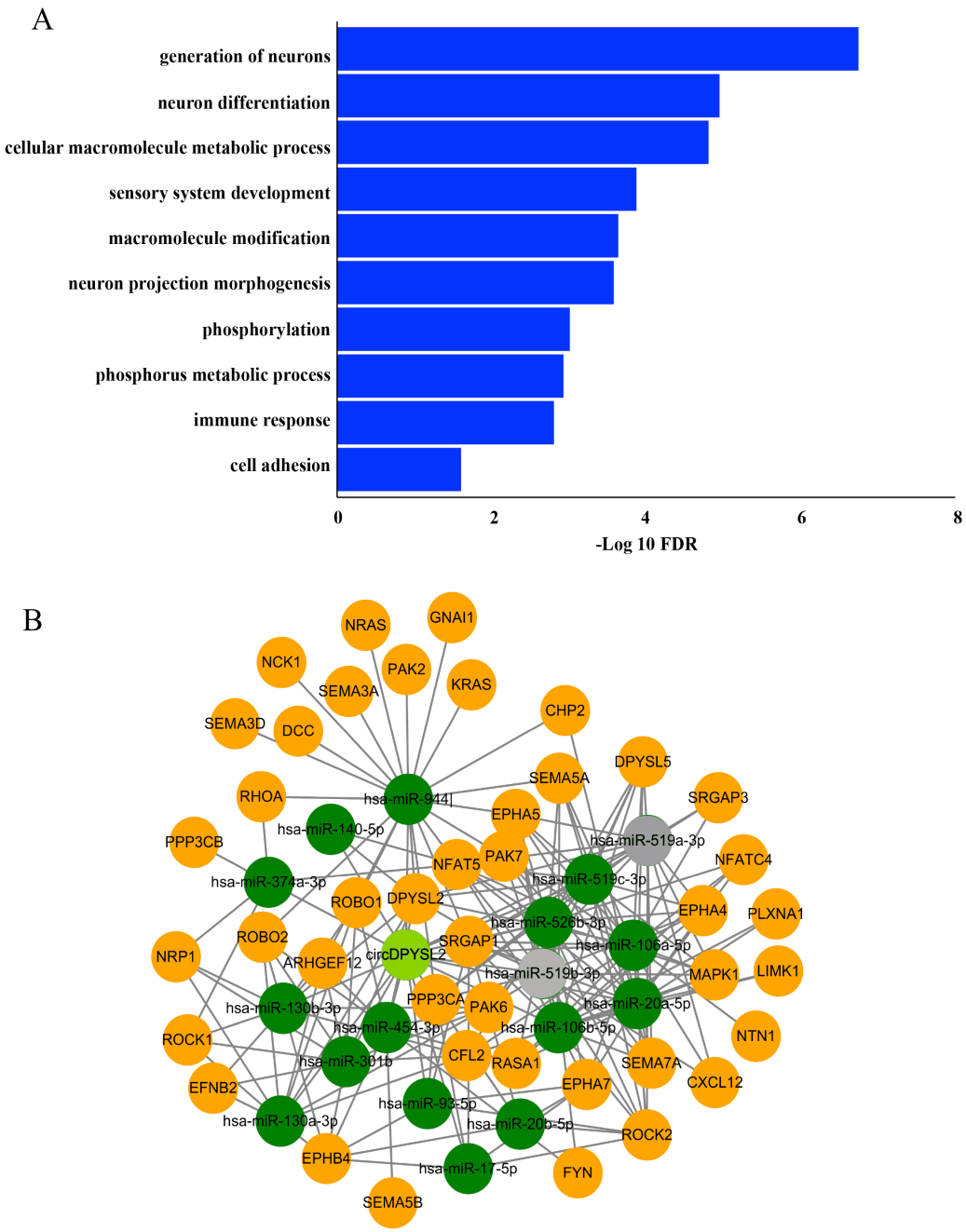


Figure S5: (A) Gene ontology (GO) analysis using host genes producing higher level of circRNAs in diencephalon than cerebellum. Bar charts showing selected biological process ranked by false discovery rate (FDR) (FDR<0.05). (B) The putative miRNAs that could be sponged by circDPYSL2 are predicted by starBase database and indicated in dark green. The mRNAs networks that could be affected by miRNA changes is shown in yellow. miRNAs that are not expressed in corresponding tissues are marked in grey. The circRNA-miRNA-mRNA axis is related to axon guidance.

Table S2: Expression level (read counts) of miRNAs in cerebellum and diencephalon associated with circAFAP1 and circDPYSL2

	miRNA	cerebellum	diencephalon	circRNA
1	hsa-mir-3163	10	4	circAFAP1
2	hsa-mir-31	4	5	circAFAP1
3	hsa-mir-27a	7	7	circAFAP1
4	hsa-mir-27b	21	26	circAFAP1
5	hsa-mir-144	11	10	circAFAP1
6	hsa-mir-670	10	27	circAFAP1
7	hsa-mir-665	16	10	circAFAP1
8	hsa-mir-374b	18	16	circAFAP1
9	hsa-mir-409	18	27	circAFAP1
10	hsa-mir-130a	19	18	circDPYSL2
11	hsa-mir-519c	5	6	circDPYSL2
12	hsa-mir-519a	0	0	circDPYSL2
13	hsa-mir-519b	0	3	circDPYSL2
14	hsa-mir-17	18	4	circDPYSL2
15	hsa-mir-20a	7	7	circDPYSL2
16	hsa-mir-106a	13	9	circDPYSL2
17	hsa-mir-526b	3	3	circDPYSL2
18	hsa-mir-130b	45	47	circDPYSL2
19	hsa-mir-454	41	34	circDPYSL2
20	hsa-mir-301b	8	10	circDPYSL2
21	hsa-mir-93	41	42	circDPYSL2
22	hsa-mir-20b	20	11	circDPYSL2
23	hsa-mir-106b	29	22	circDPYSL2
24	hsa-mir-944	18	3	circDPYSL2
25	hsa-mir-374a	23	13	circDPYSL2
26	hsa-mir-140	22	12	circDPYSL2

Table S3: Primers for RT-qPCR

Primer.Name	species	Sequence
circRELL1_F_qPCR	human	GTCAAACCAGAAGGAACGGA
circRELL1_R_qPCR	human	ACATCCCTCTCGACAACACC
circZFAND6_F_qPCR	human	CCAGTTCAATGCACAGATGG
circZFAND6_R_qPCR	human	AAGCATAGGCACTTGGCTGT
circKCNN2_F_qPCR	human	TCTGATTGCCAGAGTCATGC
circKCNN2_R_qPCR	human	TCCAGTCATCTGCTCCATTG
circSATB2_F_detect	human	AAGATGTCTATCATGTTGTG
circSATB2_R_detect	human	GGCCACTGTCGCGTCGGGTG
circRAPGEF5_F	human	ACTGCAACTTGTGCCTCTCA
circRAPGEF5_R	human	GCACTGGACGAAAGGACAGT
circATRNL1_F1_detect/qPCR	human	ATTGCTCTGGTCATGGGAAG
circATRNL1_R1_detect/qPCR	human	TCATTGCCCCTTATTTCAGG
hGAPDH-R	human	GGCTGTTGTCATACTTCTCATGG
hGAPDH-F	human	GGAGCGAGATCCCTCCAAAAT

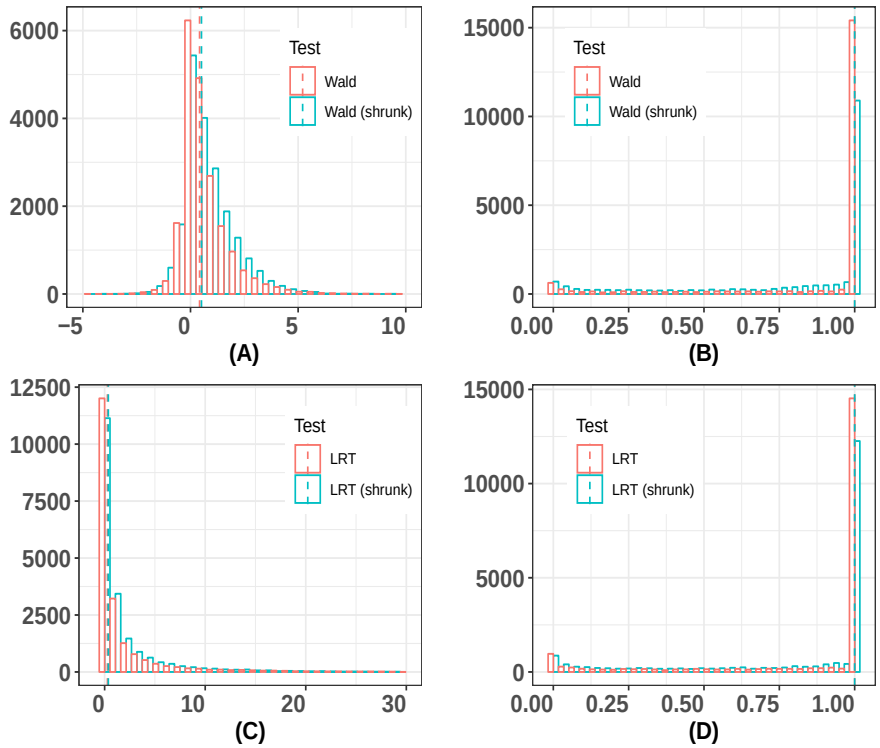


Figure S6: Differentialyl expressed circRNA analysis between frontal cortex and diencephalon. (A,B) Comparison of statistics distribution and FDR distribution between Wald test vs Wald test (shrunk); (C,D) Comparison of statistics distribution and FDR distribution between LRT vs LRT (shrunk). The median of test statistics or FDR is indicated by a dash vertical line.

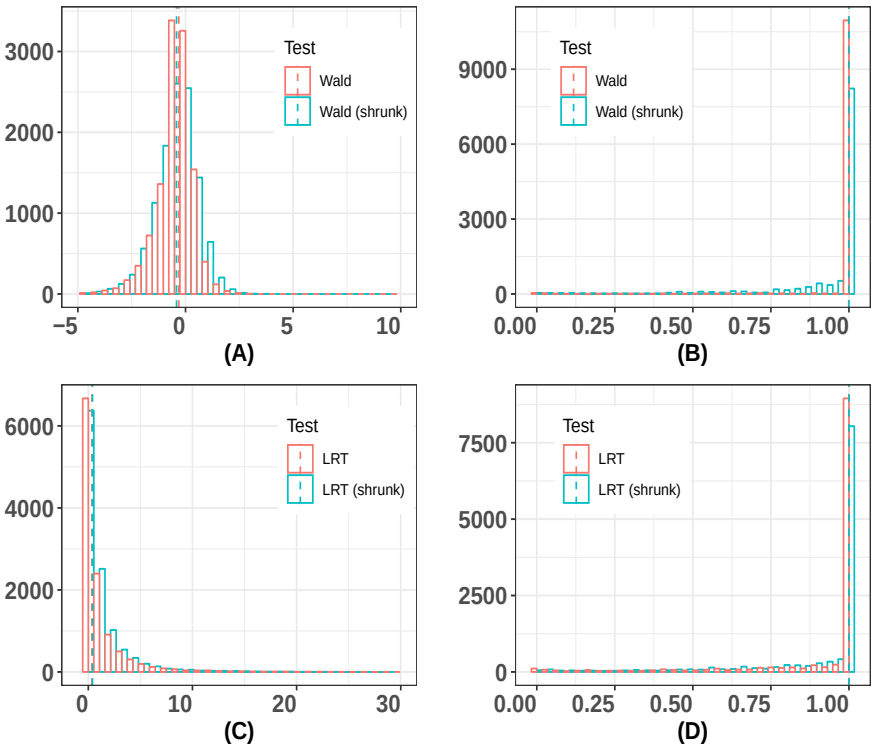


Figure S7: Differentially expressed circRNA analysis between cerebellum and diencephalon. (A,B) Comparison of statistics distribution and FDR distribution between Wald test vs Wald test (shrunk); (C,D) Comparison of statistics distribution and FDR distribution between LRT vs LRT (shrunk). The median of test statistics or FDR is indicated by a dash vertical line.