

Full-length circRNA sequencing method using low-input RNA and profiling of circRNAs in MPTP-PD mice on nanopore platform

Ying Wang,^{‡a} Xiaohan Li,^{‡ab} Wenxiang Lu,^a Fuyu Li,^a Lingsong Yao,^a Zhiyu Liu,^a Huajuan Shi,^a Weizhong Zhang,^{*c} and Yunfei Bai^{*a}

* Corresponding authors

a State Key Lab of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu, 210096, China.

E-mail: whitetcf@seu.edu.cn

b Hebei Key Laboratory of Animal Physiology, Biochemistry and Molecular Biology, College of Life Sciences, Hebei Normal University, Shijiazhuang, 050024, China.

c Department of Ophthalmology, First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, 210029, China.

E-mail: wei-zhongzhang@njmu.edu.cn

[‡] Y. W. and X. L. contributed equally to this paper.

Supporting information

1. Title page (page S-1)
2. Custom code for circRNA analysis (page S-2)
3. Figure S1 (page S-4)
4. Figure S2 (page S-5)
5. Figure S3 (page S-6)
6. Figure S4 (page S-7)
7. Figure S5 (page S-8)
8. Figure S6 (page S-9)
9. Figure S7 (page S-9)
10. Figure S8 (page S-10)
11. Table S1-S2 (page S-11)
12. Table S3 (page S-12)
13. Table S4 (page S-13)

Custom code for circRNA analysis

1. Nanopore data QC

ONT raw data were filtered by NanoFilt (v2.8.0) to obtain reads with QV ≥ 7 .

```
NanoFilt -q 7 -l 500 --headcrop 50 $sample.raw.fastq | gzip > $sample.trim.fastq
```

For libraries from different brain regions, each sample was separated according to the 24-nt barcode within PCR2 primers by Nanoplexer (v0.1.2). We ran porechop on all samples fastq file.

Note that we added the barcode sequences into the config file of porechop before running.

```
nanoplexer -b barcode.fa -p /output/ -l $sample.log $sample.raw.fastq
```

```
porechop-runner.py -i $sample.fastq -o $sample.trim.fastq
```

2. CircRNA detection and annotation for ONT data

We ran CIRI-long (v2.0) on all trimmed ONT sequencing data and then convert the info files into bed format.

```
CIRI-long call -l $sample.trim.fastq -o /path/$sample -p $sample -r mm9.fa -a mm9.refGene.gtf
```

```
python CIRI-long/misc/convert_bed.py /path/$sample/${sample}.info /path/$sample/${sample}.bed
```

```
cat *.bed | sort -u > all.bed
```

```
bedtools getfasta -fi mm9.fa -bed all.bed -name -split -fo all.circRNA.fa
```

3. AS events generating for ONT libraries

We converted all the bed file into gtf format and then generated the AS events in all samples.

```
suppa.py generateEvents -i AS/$sample/$sample.gtf -o AS/$sample}/${sample}.event -e SE SS RI -f ioe
```

4. NGS data QC

All raw data were assessed by Fastqc and then filtered by Trimmomatic (v 0.39).

```
fastqc --extract -o /out/$sample $sample_1.fq.gz $sample_2.fq.gz
```

```
java -jar Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33 $sample_1.fq.gz $sample_2.fq.gz
```

```
$sample_1.clean.fq.gz $sample_1.un.fq.gz $sample_2.clean.fq.gz $sample_2.un.fq.gz
```

```
ILLUMINACLIP:/path/Trimmomatic-0.39/adapters/NexteraPE-PE.fa:2:30:10 SLIDINGWINDOW:5:20 LEADING:5  
TRAILING:5 MINLEN:50
```

5. CircRNA detection and annotation for NGS data

We ran CIRI-full for all samples to detection circRNAs and AS within circRNAs.

```
java -jar CIRI_Full_v2.1.1.jar Pipeline -1 $sample_1.clean.fq.gz -2 $sample_2.clean.fq.gz -r
```

```
/Mus_musculus/mm9.fa /Mus_musculus/mm9.refGene.gtf -o $sample -d /out/prefix
```

6. miRNA analysis

Mouse miRNAs seeds was downloaded from miRTarBase and then miRNA binding sites were predicted using TargetScan.

```
perl targetscan_70.pl mmu_miRNA_seed.txt all.circRNA.fa.targetscan.sequence.txt
```

```
all.circRNA.fa.targetscan.out.txt
```

7. Translation prediction

Noted that we first repeat each sequence 4 times before ORF prediction. IRESfinder was used to identified the IRES in circRNAs. To identify potential m6A sites in circRNAs, we extracted m6A modification peaks detected by three different peak calling tools (exomePeak, MetPeak, and MACS2) from the REPIC database, followed by aligning them to circRNA sequences using blat and the presence of m6A consensus motif "RAC"

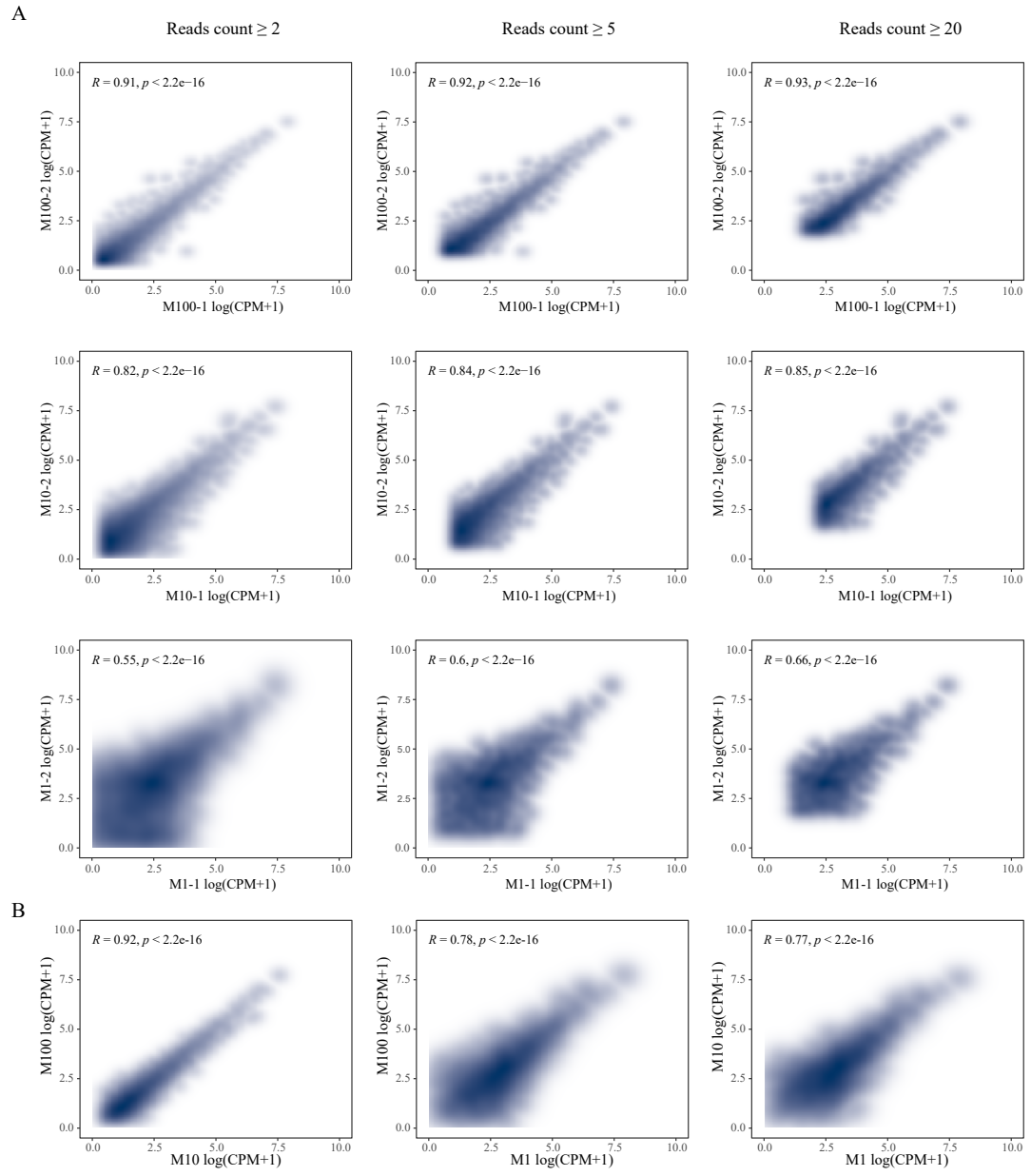


Figure S2. Expression correlation for circRNA BSJs of technical replicates (A) and different input level (B).

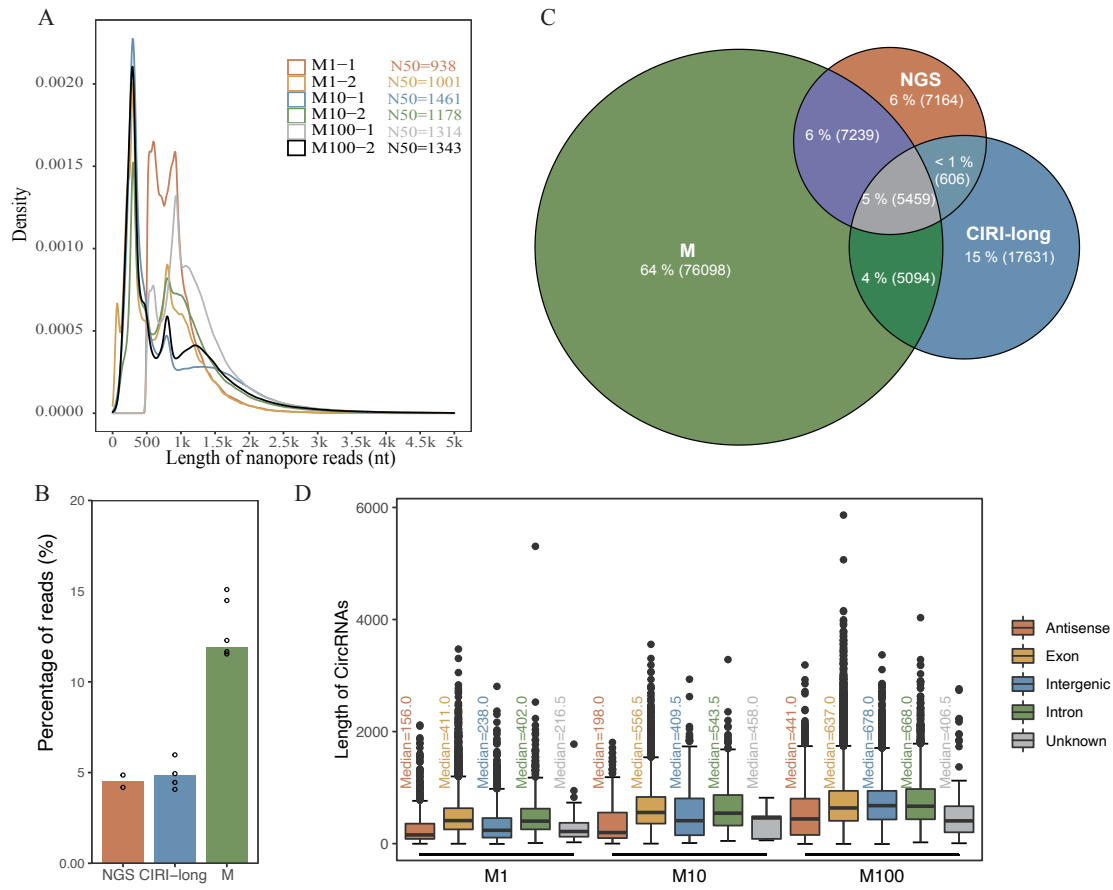


Figure S3. CircRNAs detected by CIRC-long datasets, our low-input RNA FL circRNA sequencing datasets and NGS datasets. (A) Length distribution of nanopore sequencing reads. (B) Fraction of supporting reads in three datasets. (C) Venn plot of circRNAs at the BSJ level among three datasets. (D) Length distribution of different type of FL circRNAs detected in our low-input RNA FL circRNA sequencing datasets.

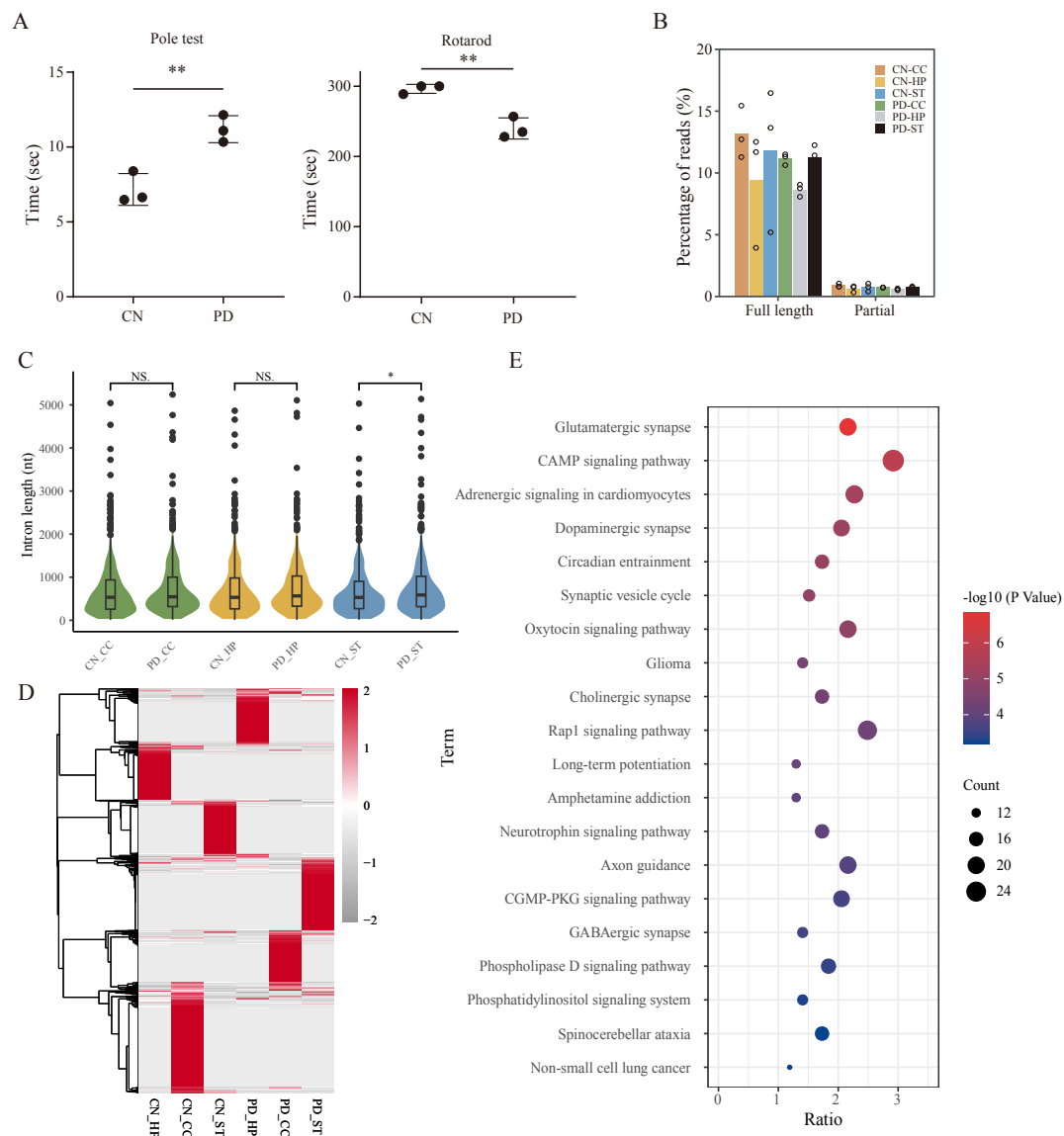


Figure S4. (A) Exercise capacity test of PD mice induced by MPTP. The difference in climbing time and Fatigue rod time difference before and after molding. (B) Fraction of FL and partial supporting reads in different brain region samples. ** indicates p-value < 0.01. (C) Violent plot of RI length distribution in different tissues. * indicates p-value < 0.02. (D) Heatmap of the circRNAs contained RI events in different tissues. (E) Bubble plot of KEGG pathways of host genes of the circRNAs contained RI events in different tissues.

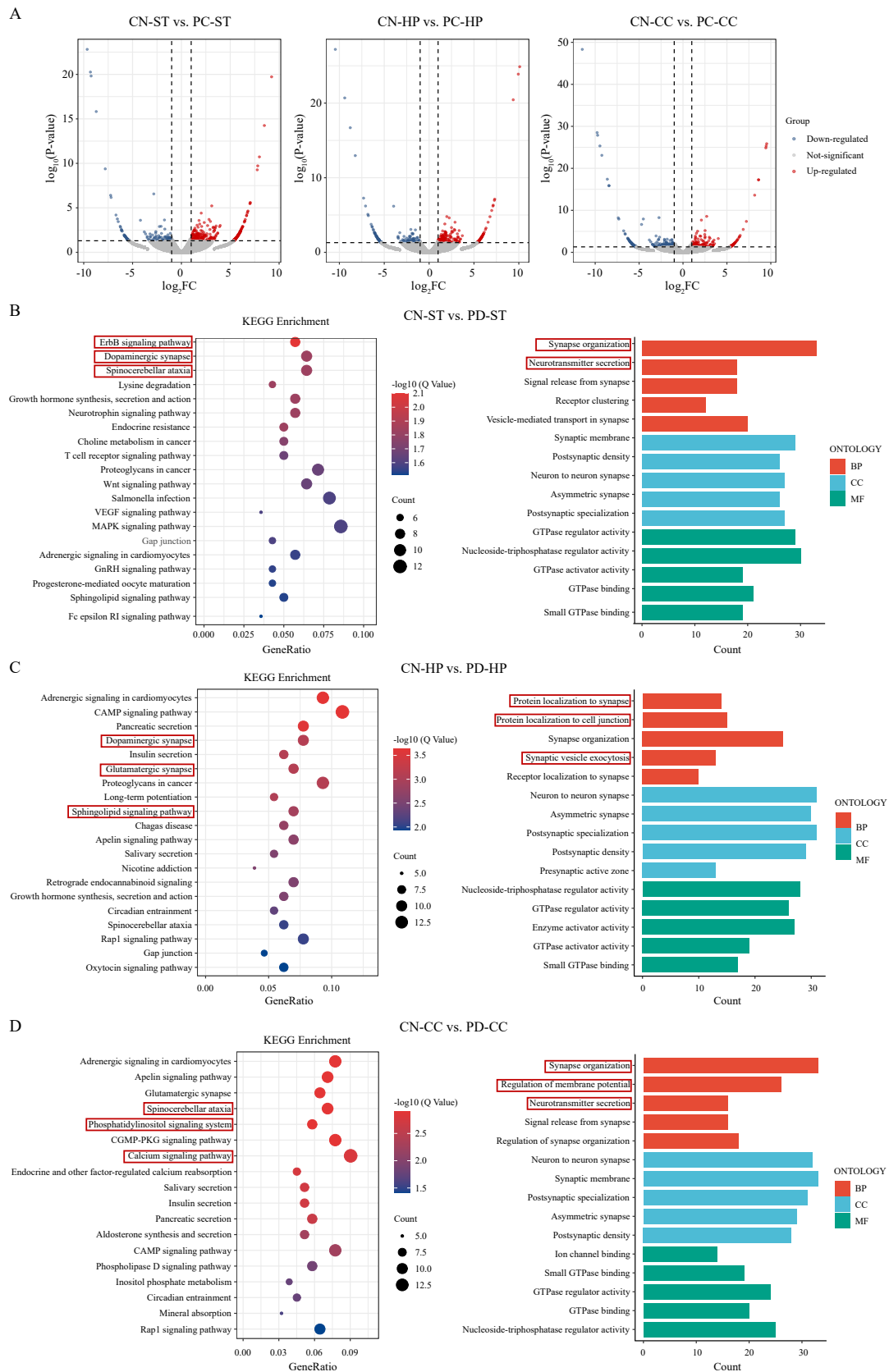


Figure S5. Different expression and function analysis of circRNAs in different groups. (A) Volcano plots of DE circRNAs in three brain regions between PD and CN. (B-C) KEGG pathway and GO term function analysis of host genes of the DE circRNAs in three groups.

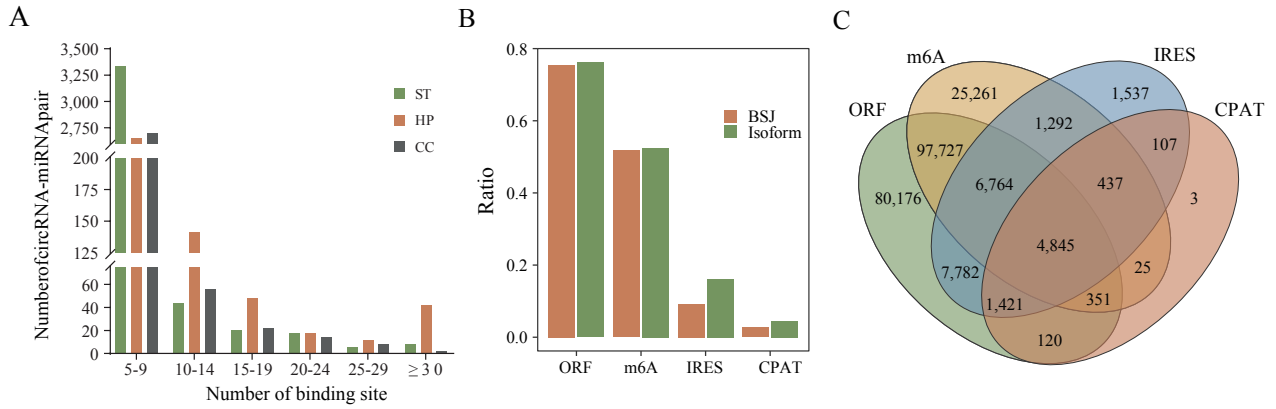


Figure S6. (A) Barplot of miRNA binding sites of all circRNAs. (B) Ratio of the circRNAs possessed ORF sequence/m6A sites/IRES sequence/coding potential assessment translatability evidences in BSJ and isoform level, respectively. (C) Venn plot shows the overlap number of circRNAs possessed four translatability evidences.

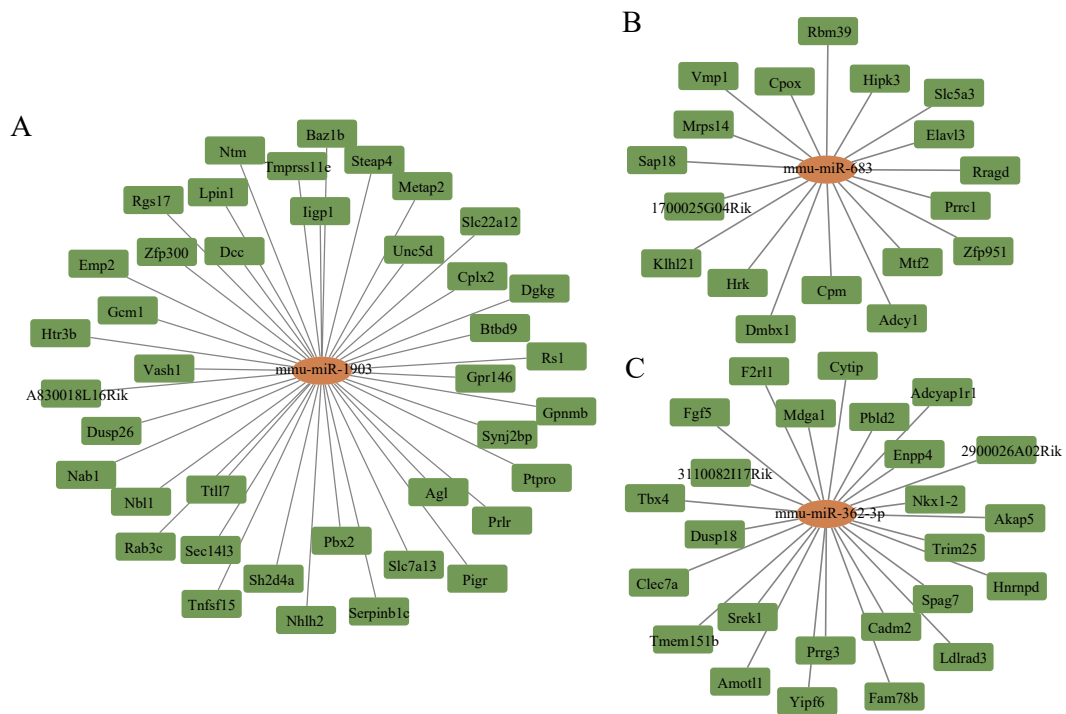


Figure S7. Network of miRNA/mRNA based on the predicted relationship according to miRTarBase. Orange oval represented the miRNAs and green rectangle represented the mRNAs.

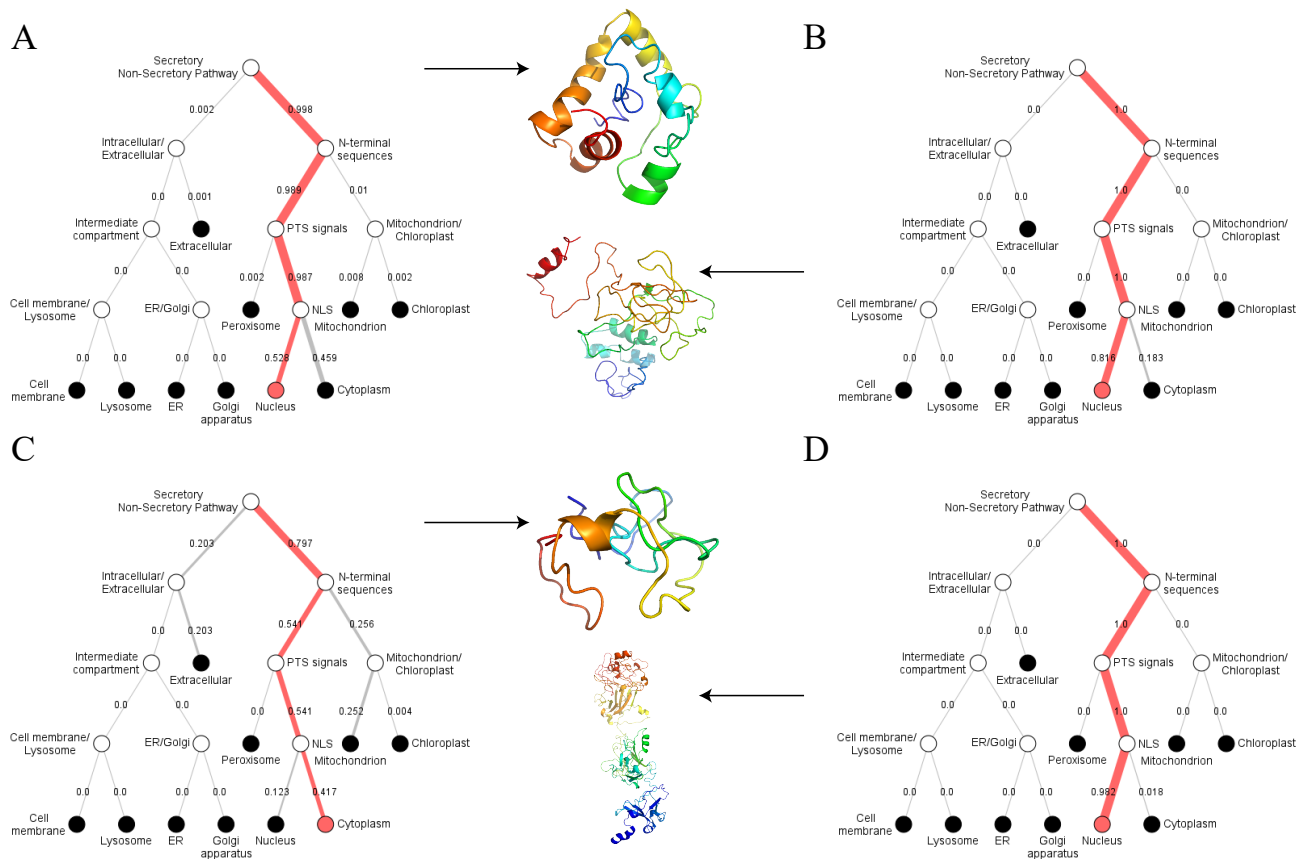


Figure S8. The subcellular location and three-dimensional structure of the predicted peptides from candidate circRNAs. Peptides encoded by mmu_circCDYL_001 consisted of 103(A) or 230 (B) amino acids, and mmu_circBNC2_001 consisted of 74 (C) or 681 (D) amino acids.

Table S1. 24-nt barcoded pprimers

Name	Sequence (5'>3')
B1P1	TATAGCCTATAGAGGCCCTATCCTAATGATACGGCGACCACCGA
B1P2	TATAGCCTATAGAGGCCCTATCCTCAAGCAGAAGACGGCATAACGA
B2P1	GGCTCTGAAGGCGAAGTAATCTTAATGATACGGCGACCACCGA
B2P2	GGCTCTGAAGGCGAAGTAATCTTACAAGCAGAAGACGGCATAACGA
B3P1	AGGCTATAGCCTCTATAGGATAGGAATGATACGGCGACCACCGA
B3P2	AGGCTATAGCCTCTATAGGATAGGCAAGCAGAAGACGGCATAACGA
B4P1	TCAGAGCCCTTCGCCTTAAGATTAATGATACGGCGACCACCGA
B4P2	TCAGAGCCCTTCGCCTTAAGATTAAGCAGAAGACGGCATAACGA
B5P1	ATTACTCGTCCGGAGACGCTCATTAAATGATACGGCGACCACCGA
B5P2	ATTACTCGTCCGGAGACGCTCATTCAAGCAGAAGACGGCATAACGA
B6P1	GAGATTCCATTCAGAAGAATTCGTAATGATACGGCGACCACCGA
B6P2	GAGATTCCATTCAGAAGAATTCGTCAGCAGAAGACGGCATAACGA
Chr15-F	TCTCACGGAAAAGTCGCAGT
Chr15-R	TCCACTGATGCTGGTGCTTT
RI-17-F	TGAGCTGACTCTTGCTACCTG
RI-17-R	GTAAAGAGCCACCTCCACAGAA

*B1-B6 represented the 24-nt barcodes

Table S2. Statics of sequencing data from different input level RNA

Sample	Clean reads	Clean base (G)	GC (%)	CircRNA number	Full length (%)	BSJ (%)
M1-1	7,724,328	7.09	54.4	8,054	6.52	/
M1-2	3,621,652	2.54	53.4	9,097	15.1	/
M10-1	2,298,833	2.02	49.2	22,147	11.7	/
M10-2	4,167,743	3.89	49.0	37,205	14.5	/
M100-1	4,511,627	5.57	49.9	36,806	12.3	/
M100-2	3,111,603	2.74	48.6	29,342	11.5	/
M10-NGS-1	35,163,580	5.09	56.3	14,598	/	4.86
M10-NGS-2	29,094,750	4.22	57.2	12,591	/	4.18

Table S3. Statics of sequencing data from different brain regions in healthy and PD mice

	Sample	Clean reads	Clean base (G)	GC (%)	CircRNA number	Full length (%)	CircBase	CircAtlas
CN	HP-rep1	3,185,248	3.77	45.0	13,771	4.26	4,982	10,085
	HP-rep2	3,220,365	3.79	49.8	34,751	12.5	6,200	16,456
	HP-rep3	2,586,338	3.2	50.0	32,882	13.3	5,974	15,537
	CC-rep1	1,380,796	1.65	50.0	24,521	16.5	5,665	13,407
	CC-rep2	3,142,683	3.65	48.6	39,074	12.1	6,162	16,387
	CC-rep3	2,758,445	2.89	50.3	36,714	13.5	6,364	17,231
	ST-rep1	2,590,539	3.39	44.5	14,830	5.59	4,984	10,596
	ST-rep2	2,559,207	2.63	50.2	42,831	17.5	6,300	16,769
	ST-rep3	2,428,175	2.93	49.7	32,141	14.4	6,021	15,758
PD	HP-rep1	3,335,877	4.65	46.2	25,835	8.63	5,979	15,422
	HP-rep2	2,307,448	2.97	47.7	20,645	9.4	5,550	13,352
	HP-rep3	1,718,584	2.34	47.3	16,981	9.55	5,112	11,452
	CC-rep1	2,869,815	3.98	46.2	30,174	12.2	6,038	16,404
	CC-rep2	1,824,360	2.43	47.2	21,437	12.1	5,469	13,251
	CC-rep3	2,685,527	3.47	46.8	26,840	11.3	5,969	15,546
	ST-rep1	3,786,063	5.02	47.6	39,436	13	6,606	19,132
	ST-rep2	1,742,111	2.37	46.7	18,774	10.7	5,171	12,014
	ST-rep3	2,990,368	3.53	48.4	31,479	12.2	6,367	17,283

Table S4. Potential translated circRNA

Brain region	Genomic Location	Host gene	Length (nt)	riboCIRC_ID	Evidence
ST	chr5:104314926-104315309	<i>Klhl8</i>	383	mmu_circKLHL8_001	RPF/cORF/m6A
	chr5:107063850-107067654	<i>Zfp644</i>	3,536	mmu_circZFP644_001	RPF/cORF/m6A
HP	chr11:117151712-117152350	<i>Septin9</i>	638	mmu_circSeptin9_001	RPF/cORF/m6A
	chr14:35504868-35508291	<i>Wapl</i>	1,152	mmu_circWAPL_001	RPF/cORF/m6A
	chr15:82681943-82687714	<i>Tcf20</i>	5,771	mmu_circTCF20_001	RPF/cORF/m6A
	chr19:28740243-28740726	<i>Glis3</i>	483	mmu_circGLIS3_002	RPF/cORF/MS/m6A
	chr4:108390394-108400150	<i>Zfyve9</i>	2,217	mmu_circZFYVE9_001	RPF/cORF/m6A
CC	chr13:35907778-35908444	<i>Cdyl</i>	666	mmu_circCDYL_001	RPF/cORF/MS/m6A
	chr4:83937597-83939566	<i>Bnc2</i>	1,969	mmu_circBNC2_001	RPF/cORF/MS/m6A