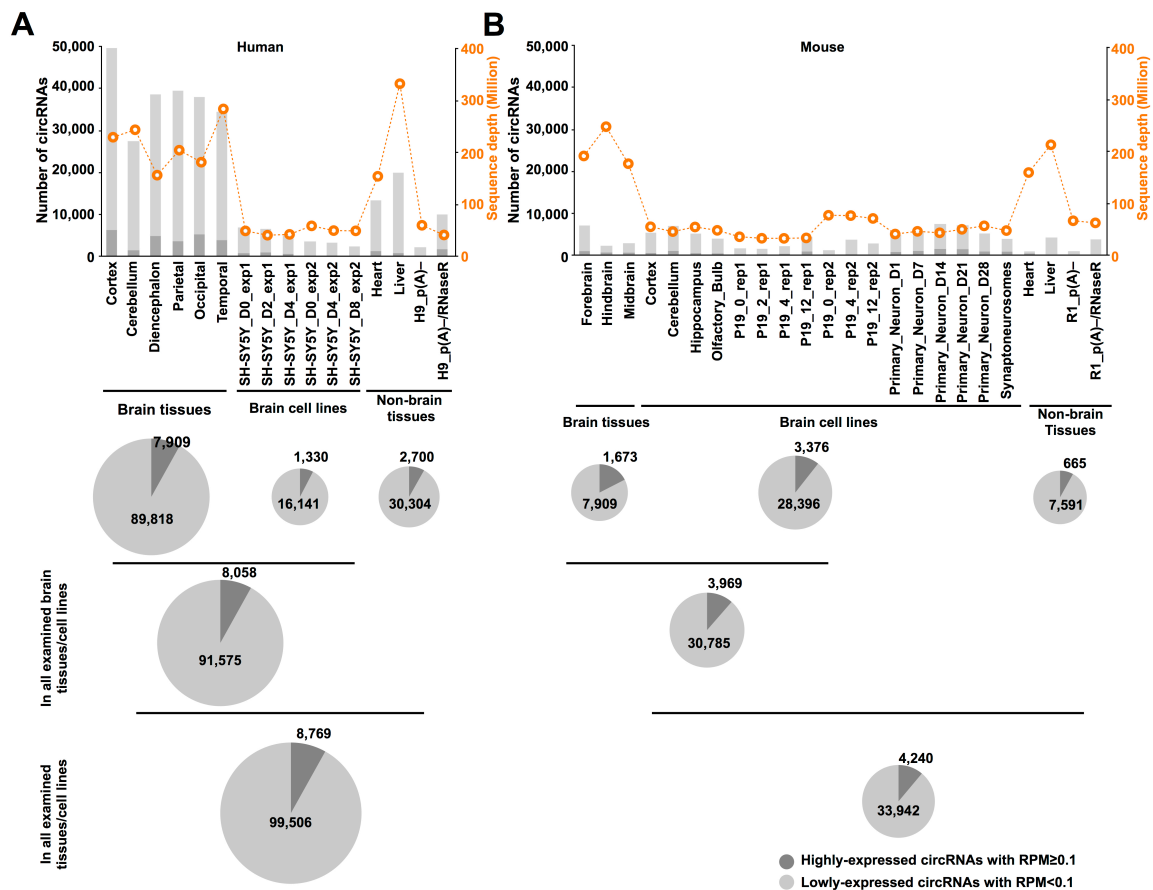


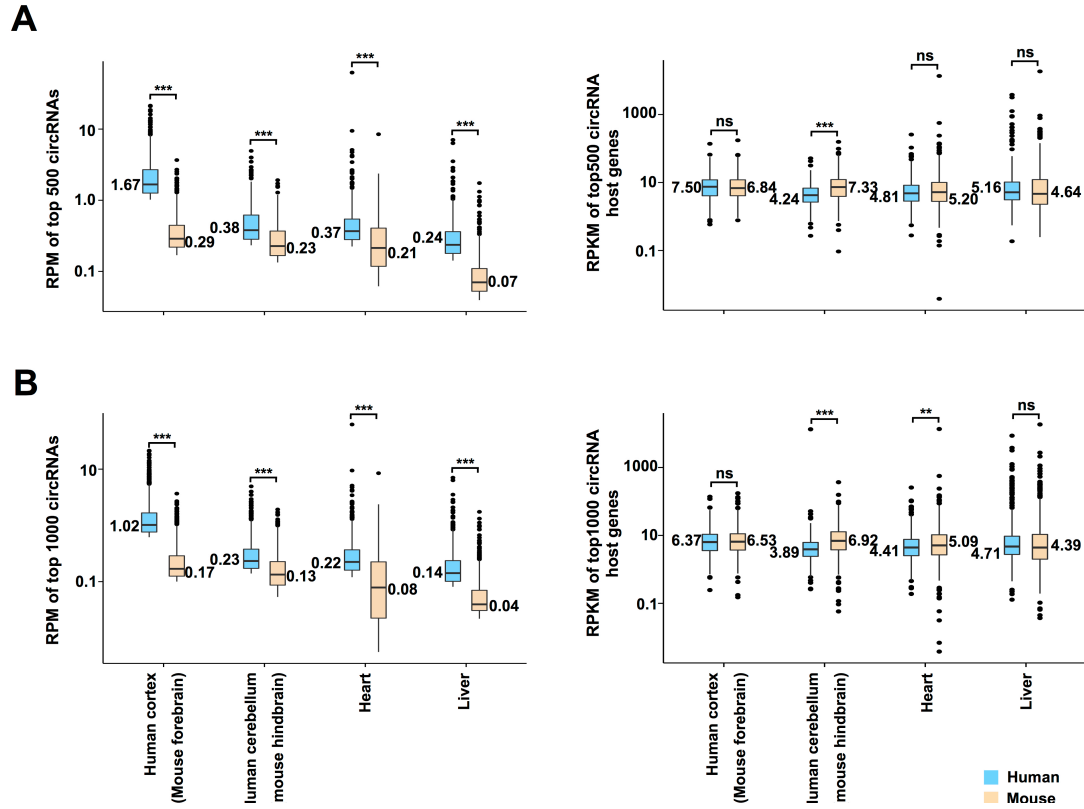
## Supplemental figure legends



**Figure S1. Identification of circRNAs from different tissues and cell lines in human and mouse.**

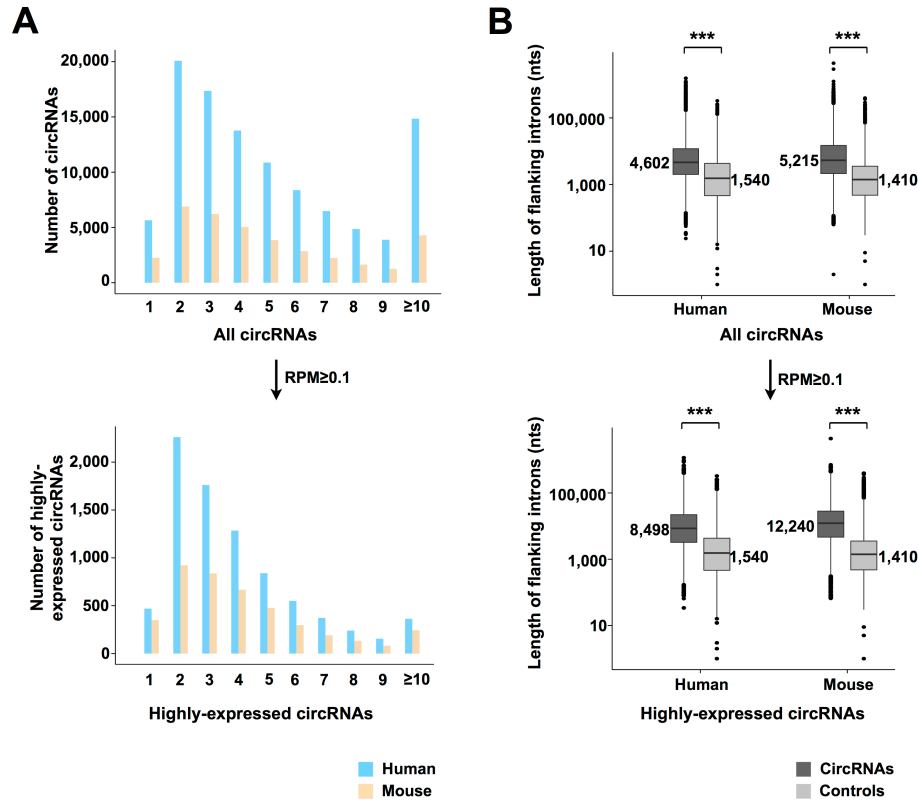
(A) Histogram of identified circRNAs in various human samples. Dark gray columns indicate the highly-expressed circRNAs with RPM  $\geq 0.2$  in RNase R-treated samples or RPM  $\geq 0.1$  in all other samples. Sequence depths for all examined human samples are indicated in orange. Total numbers of circRNAs in different tissues/cell lines were indicated in pie plots. (B) Histogram of identified circRNAs in various mouse samples. Dark gray columns indicate the highly-

expressed circRNAs with  $\text{RPM} \geq 0.2$  in RNase R-treated samples or  $\text{RPM} \geq 0.1$  in all other samples. Sequence depths for all examined mouse samples are indicated in orange. Total numbers of circRNAs in different tissues/cell lines were indicated in pie plots.



**Figure S2. Expression of top 500 or top 1000 circRNAs and their cognate mRNAs.**

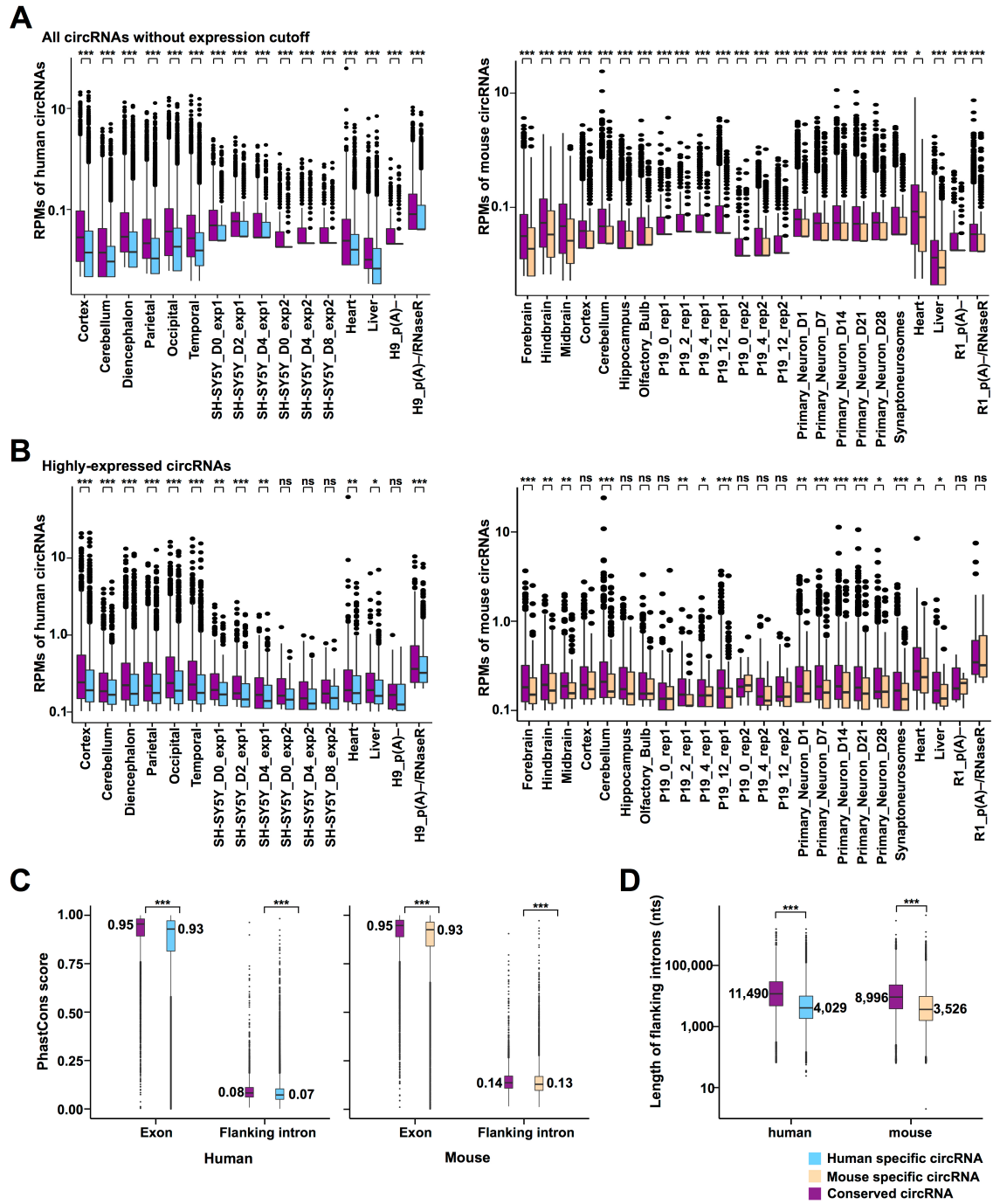
The expression level of top 500 (A, left panel) or top 1000 (B, left panel) circRNAs in human (blue) is higher than those in mouse (yellow). In contrast, the expression level of top 500 (A, right panel) or top 1000 (B, right panel) circRNA cognate mRNAs is similar between human (blue) and mouse (yellow). \*\*  $p$  value < 0.01, \*\*\*  $p$  value < 0.001, Wilcoxon rank-sum test.



**Figure S3. Genomic features of circRNAs in human and mouse.**

(A) Numbers of circularized exons in all (top panel) or highly-expressed (bottom panel) circRNAs in human (blue) or mouse (yellow).

(B) Length of circRNA-flanking introns in all (top panel) or highly-expressed (bottom panel) circRNAs in both human and mouse. \*\*\*  $p$  value < 0.001, Wilcoxon rank-sum test.



**Figure S4. Characterization of conserved circRNAs.**

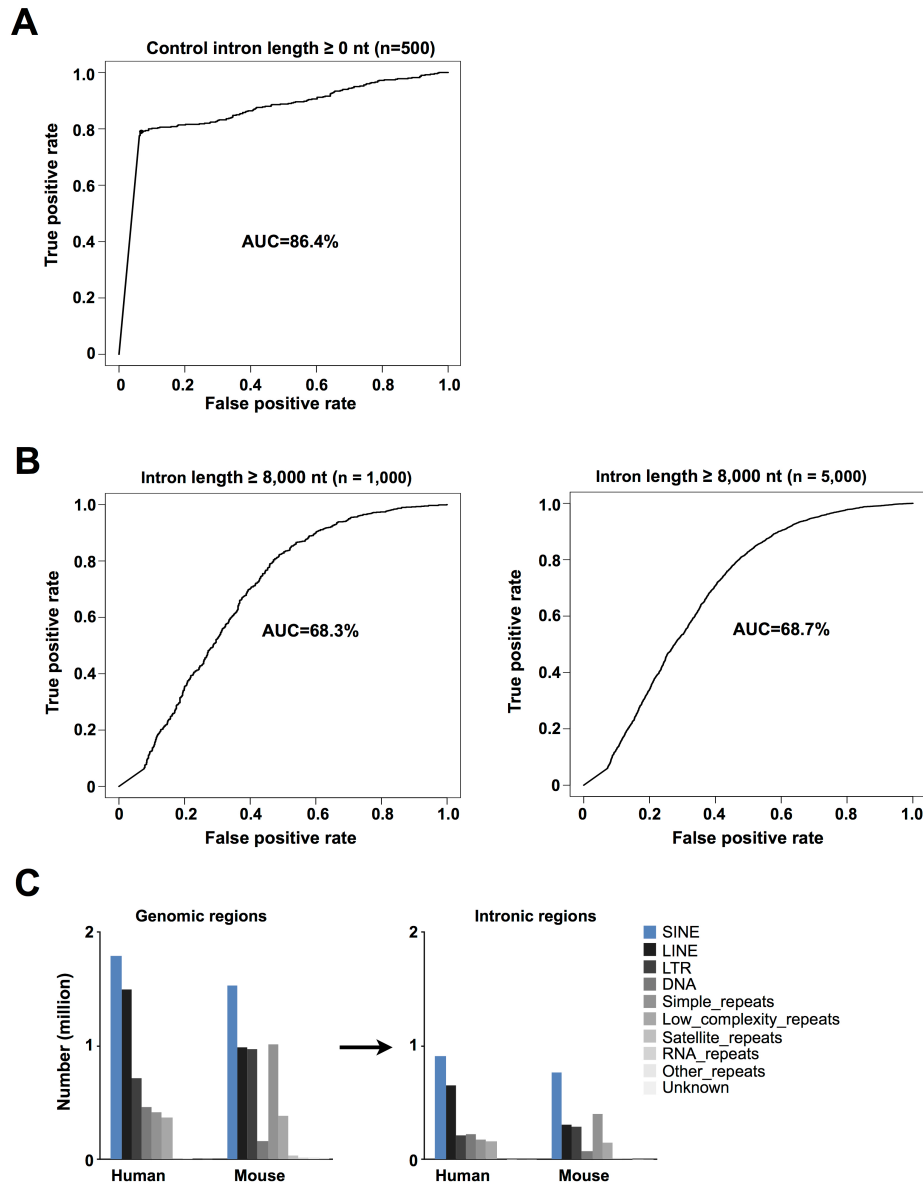
(A) The expression level of conserved (purple) circRNAs is higher than species-specific circRNAs in human (blue) and mouse (yellow) in all examined samples. \*

$p$  value < 0.05, \*\*\*  $p$  value < 0.001, Wilcoxon rank-sum test.

(B) When detected with highly-expressed circRNAs, the expression level of conserved (purple) circRNAs is higher than species-specific circRNAs in most examined human (blue) and mouse (yellow) samples. Partially due to the small number of highly-expressed circRNAs, the comparison might be not significant (ns) in some samples. \*  $p$  value < 0.05, \*\*\*  $p$  value < 0.001, ns, not significant, Wilcoxon rank-sum test.

(C) The PhastCons scores of conserved (purple) and species-specific (blue, human; yellow, mouse) circRNAs in exon, flanking intron and predicted complementary sequence. \*\*\*  $p$  value < 0.001, Wilcoxon rank-sum test.

(D) Length of circRNA-flanking introns of conserved (purple) or species-specific circRNAs (blue for human or yellow for mouse). \*\*\*  $p$  value < 0.001, Wilcoxon rank-sum test.



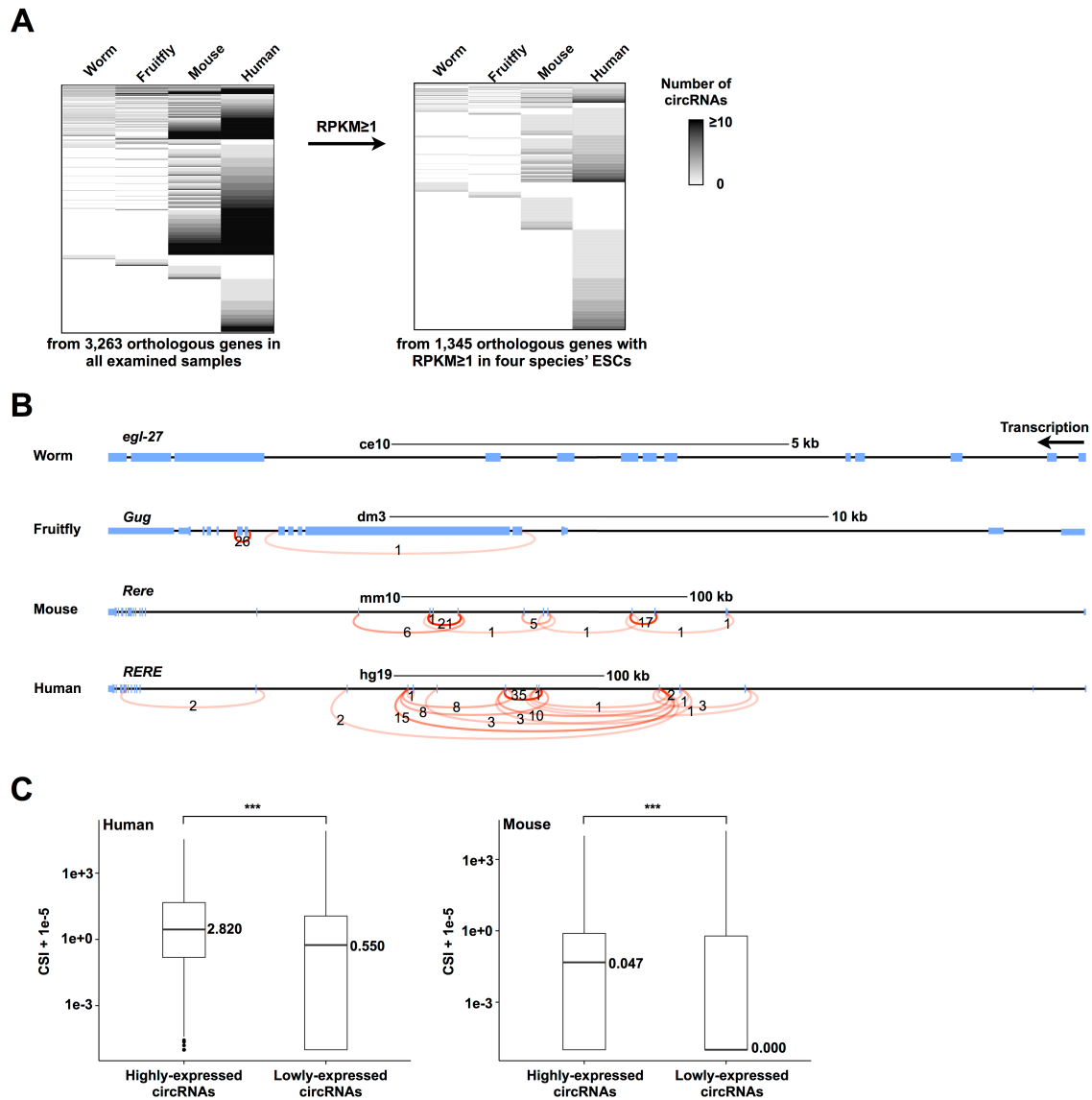
**Figure S5. ROC curve of CSI and distribution of different types of repetitive elements in human and mouse.**

(A) ROC (receiver operating characteristic) curve was plotted to evaluate CSI performance and the area under the curve (AUC) achieves 86.4% when using 500 control intron pairs without length limitation.

(B) ROC curve was plotted to evaluate CSI performance and the AUC of the ROC curve was used to evaluate the specificity and sensitivity of CSI from 1,000 or 5,000 flanking introns of highly-expressed circRNAs and 1,000 or 5,000 control intron pairs with lengths  $\geq 8,000$  nts.

(C) Distribution of different types of repetitive elements in genomic (left panel) and intronic (right panel) regions in human and mouse.





**Figure S6. Increased alternative circularization during species evolution.**

(A) Increased alternative circularization during species evolution in 3,263 orthologous genes from all examined samples (left panel) and in 1,345 orthologous genes with their linear mRNAs expressed at RPKM  $\geq 1$  in ESCs of worm, fruitfly, mouse and human (right panel).

(B) An example of increased alternative circularization at *RERE* locus in ESCs of four species during species evolution. Identified circRNAs at *RERE* locus with their back-spliced junction reads (numbers) from these ESC lines are indicated by red arc lines.

(C) Boxplots of CSIs. CSIs of highly-expressed circRNAs are much higher than those of lowly-expressed ones in both human (left panel) and mouse (right panel).

## **Supplemental Table Legends**

### **Table S1. List of RNA-seq datasets used in this study.**

Total of 16 human, 24 mouse, 30 fruitfly and 12 worm samples were used for circRNA identification in this study, listed with species, tissue/cell line information, GEO number, raw read number, mapped read number with TopHat2/TopHat-Fusion and mapping ratio.

### **Table S2. Lists of circRNAs identified in human samples, related to Fig. 1A and S1A.**

In total, 108,275 circRNAs were identified in examined human samples, with 15,551 conserved ones highlighted in red. The genomic locations, strand information, exon numbers/sizes/offsets, host gene symbols and flanking intron locations are listed.

### **Table S3. Lists of circRNAs identified in mouse samples, related to Fig. 1A and S1B.**

In total, 38,182 circRNAs were identified in examined mouse samples, with 15,517 conserved ones highlighted in red. The genomic locations, strand information, exon numbers/sizes/offsets, host gene symbols and flanking intron locations are listed.

### **Table S4. Lists of circRNAs identified in fruitfly samples.**

In total, 7,378 circRNAs were identified in examined fruitfly samples. The genomic locations, strand information, exon numbers/sizes/offsets, host gene symbols and flanking intron locations are listed.

**Table S5. Lists of circRNAs identified in worm samples.**

In total, 3,840 circRNAs were identified in examined worm samples. The genomic locations, strand information, exon numbers/sizes/offsets, host gene symbols and flanking intron locations are listed.

**Table S6. Expression level of orthologous genes in ESCs and number of circRNAs identified in orthologous genes, related to Fig. 5C.**

In total, 3,263 orthologous genes in four species were retrieved from OrthoDB<sup>25</sup>.

Among them, 1,345 orthologous genes with their linear mRNA expression at  $\text{RPKM} \geq 1$  in ESC samples from all four species were marked by red. The numbers of circRNAs generated from 3,263 (highlighted in yellow) or 1,345 (highlighted in blue) orthologous genes were retrieved from all samples or related ESC samples, respectively.