Supplementary Information

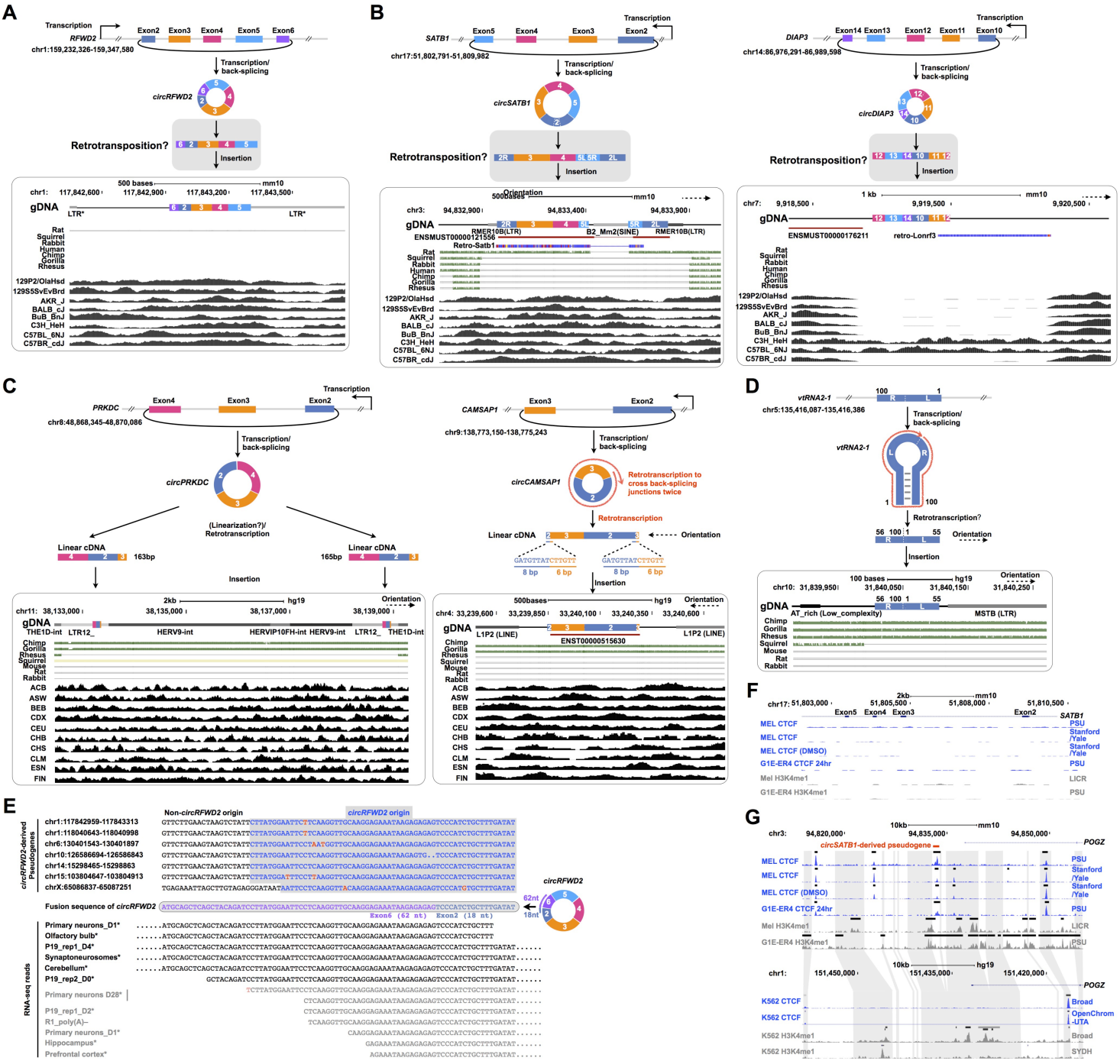

Figure S1. Characteriation of mouse and human circRNA-derived pseudogenes.

(A) The *circRFWD2*-derived pseudogenes could be identified in different mouse strains, but are not conserved in other species, such as rat and human. *, MMERVK10C-int LTR retrotransposon sequences.

(B) Identification of other high-confidence circRNA-derived pseudogenes in mouse reference genome. Left, mouse *circSATB1*-derived pseudogene. A *circSATB1* that contains exons 2, 3, 4 and 5 with the back-splicing exon5-exon2 junction sequence was produced from the mouse chr17:51,802,791-51,809,982 locus (top), and could be retrotransposed (middle, grey box) to generate a *circSATB1*-derived pseudogene on mouse chromosome 3 (bottom). Note that the *circSATB1*-derived pseudogene could be identified in all examined mouse strains and is also conserved in rat. Right, mouse circ*DIAP3*-derived pseudogene. A circ*DIAP3* that contains exons 10, 11, 12, 13 and 14 with the back-splicing exon14-exon10 junction sequence was produced from the mouse chr14:86,976,291-86,989,598 locus (top) and could be retrotransposed (middle, grey box) to generate a *circDIAP3*-derived pseudogene on mouse chromosome 7 (bottom). Note that the *circDIAP3*-derived pseudogene could be detected in some but not all the examined mouse strains.

(C) Identification of high-confidence circRNA-derived pseudogenes in human reference genome. Left, human *circPRKDC*-derived pseudogenes. A *circPRKDC* that contains exons 2, 3 and 4 with the back-splicing exon4-exon2 junction sequence was produced from the human chr8:48,868,345-48,870,086 locus (top), and could be retrotransposed (middle, grey box) to generate two *circPRKDC*-derived pseudogenes on human chromosome 11 (bottom). Note that these two *circPRKDC*-derived

pseudogenes could be identified in all examined human populations and is also conserved in Chimp and Gorilla, but not in Rhesus. Right, human *circCAMSAP1*-derived pseudogene. A *circCAMSAP1* that contains exons 2 and 3 with the back-splicing exon3-exon2 junction sequences was produced from the human chr9:138,773,150-138,775,243 locus (top) as previously annotated [*1*], and could be retrotransposed (middle, grey box) to generate a *circCAMSAP1*-derived pseudogene on human chromosome 4 (bottom). Of note, the *circCAMSAP1*-derived pseudogene contains eight nucleotides of exon3 upstream to the full length of exon2 sequences on one end, and six nucleotides of exon2 downstream to the full length of exon3 sequences on the other end, suggesting that the retro-transcription of *circCAMSAP1* could presumably pass its exon3-exon2 back-splicing junction twice (middle). The *circCAMSAP1*-derived pseudogene could be identified in all examined human populations and is also conserved in Chimp and Gorilla, but not in Rhesus.

(D) A pseudogene with non-colinear exon-exon junctions of human vault RNA (*vtRNA2-1*) origin. The predicted secondary structure of *vtRNA* [*2*, *3*], and its closely located 5' and 3' ends could be read through by reverse transcription (middle, red curved line around the structured vtRNA), theoretically, resulting in the *vtRNA*-derived pseudogene with a non-colinear exon-exon junction, even though the parental vtRNA itself was not covalently closed. Note that the *vtRNA*-derived pseudogene is also conserved in primates.

(E) Mapping of RNA-seq reads to the back-splicing junction of *circRFWD2* without a mismatch. Possibly due to the high similarity to its related pseudogenes (top), RNA-seq reads failed to be mapped to the *circRFWD2* exon6-exon2 back-splicing junction

sites with most currently available methods [*1, 4*], but identified in this study by aligning to the 190 bp back-splicing junction references requiring no mismatches (Materials and Methods). Top, the sequences in *circRFWD2*-derived pseudogenes. Middle, the back-splicing junction sequences of *circRFWD2*. Bottom, real RNA-seq reads that perfectly mapped to the back-splicing junction of *circRFWD2*, but not to the *circRFWD2*-derived pseudogenes without a mismatch.

(F) No clear CTCF binding site was examined in the mouse *SATB1* region. Correspondingly, this area is also suggested to be silent as enhancer. Blue, CTCF binding signals. Grey, H3K4me1 binding signals.

(G) Neither the *circSATB1*-derived pseudogene homologue nor the CTCF binding site could be found in the examined human genomes and with available CTCF datasets. Blue peaks, CTCF binding signals. Grey peaks, H3K4me1 binding signals. Black bars over the binding signals, predicted CTCF/H3K4me1 binding regions. Mouse and human conserved regions were highlighted with grey shadow.

References

1. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. *Cell* 2014; **159**:134-147.
2. Nandy C, Mrázek J, Stoiber H, Grässer FA, Hüttenhofer A, Polacek N. *J Mol Biol* 2009; **388**(4):776-84.
3. Stadler PF, Chen JJ, Hackermüller J, *et al*. *Mol Biol Evol* 2009; **26**(9):1975-91.
4. Memczak S, Jens M, Elefsinioti A, *et al. Nature* 2013; **495**;333-338.