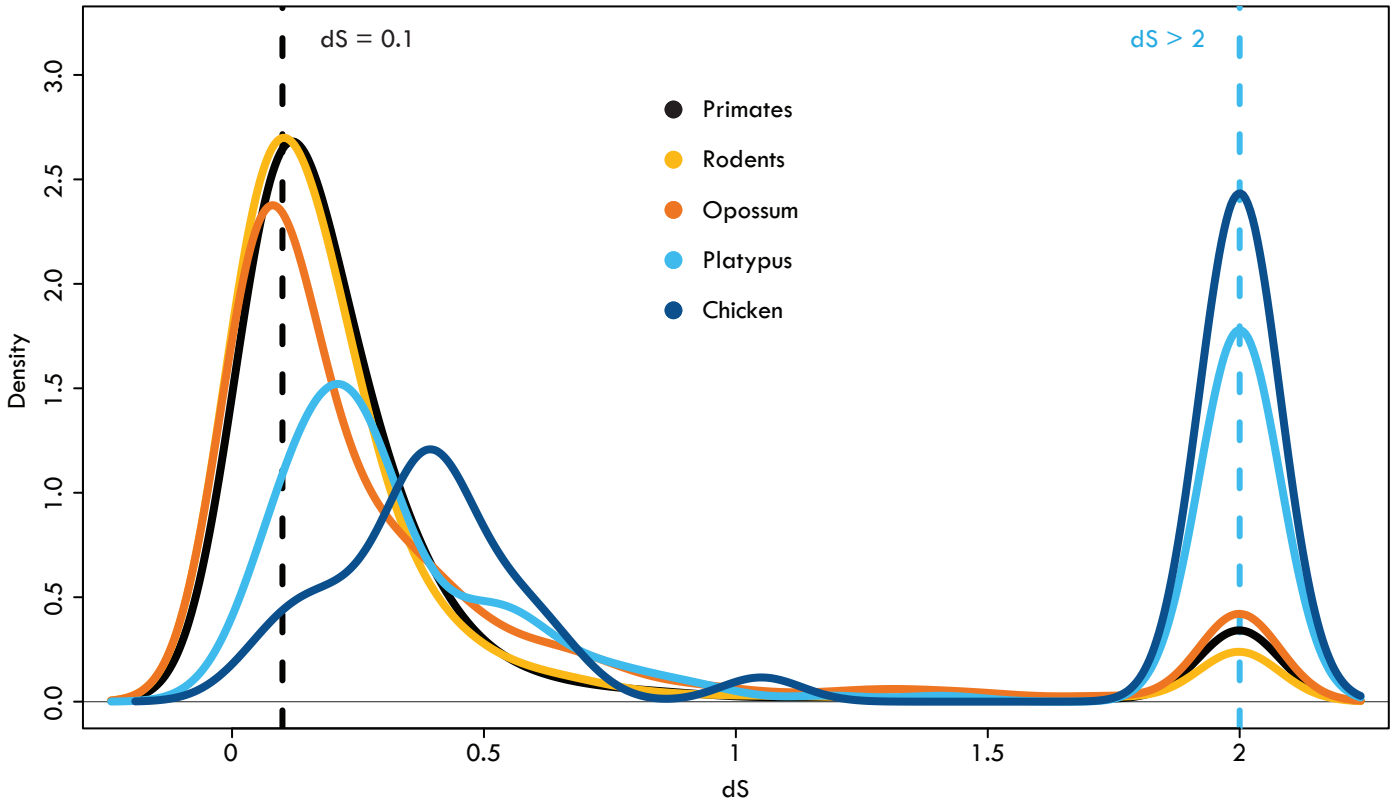
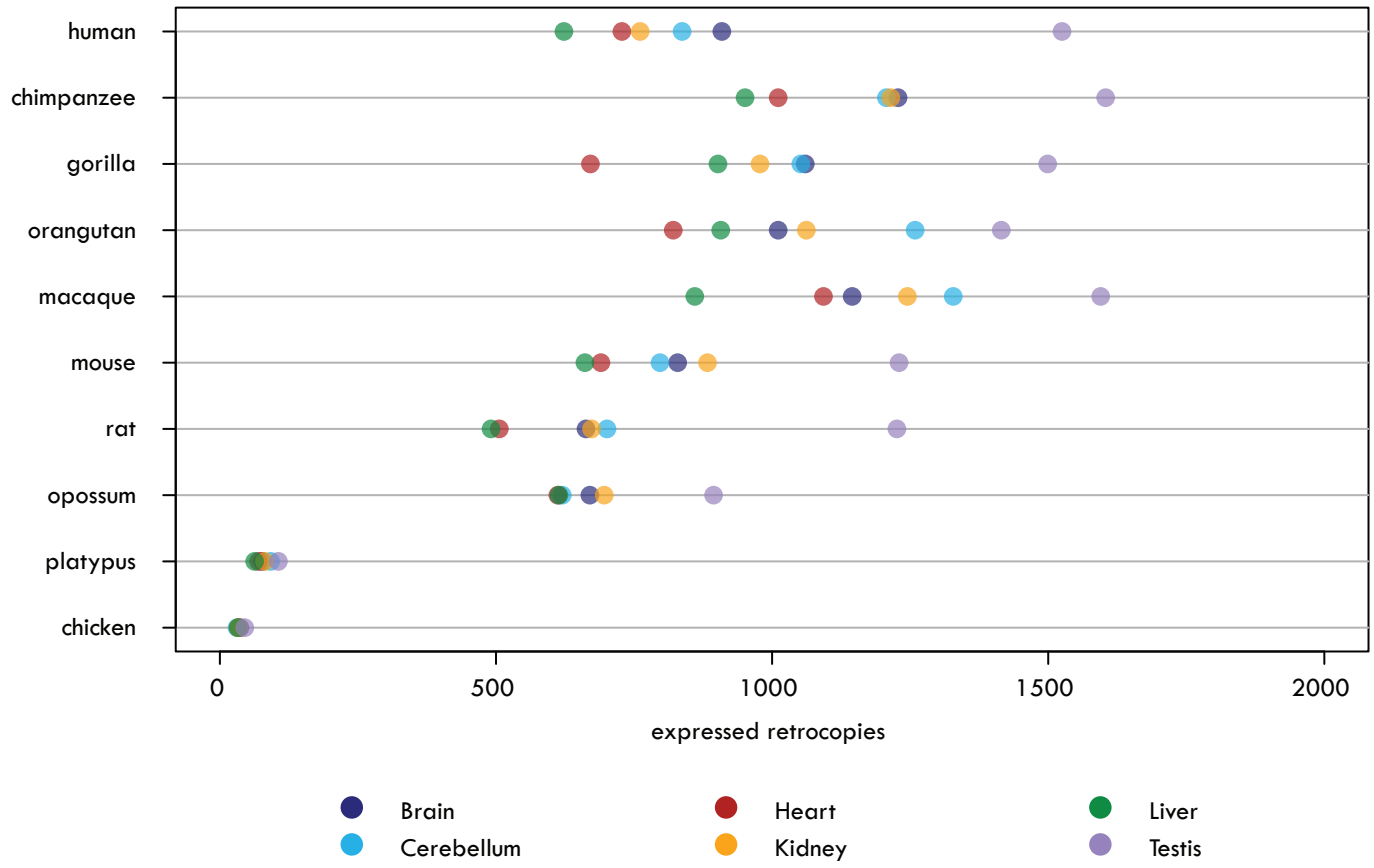
Supplemental Figure S1



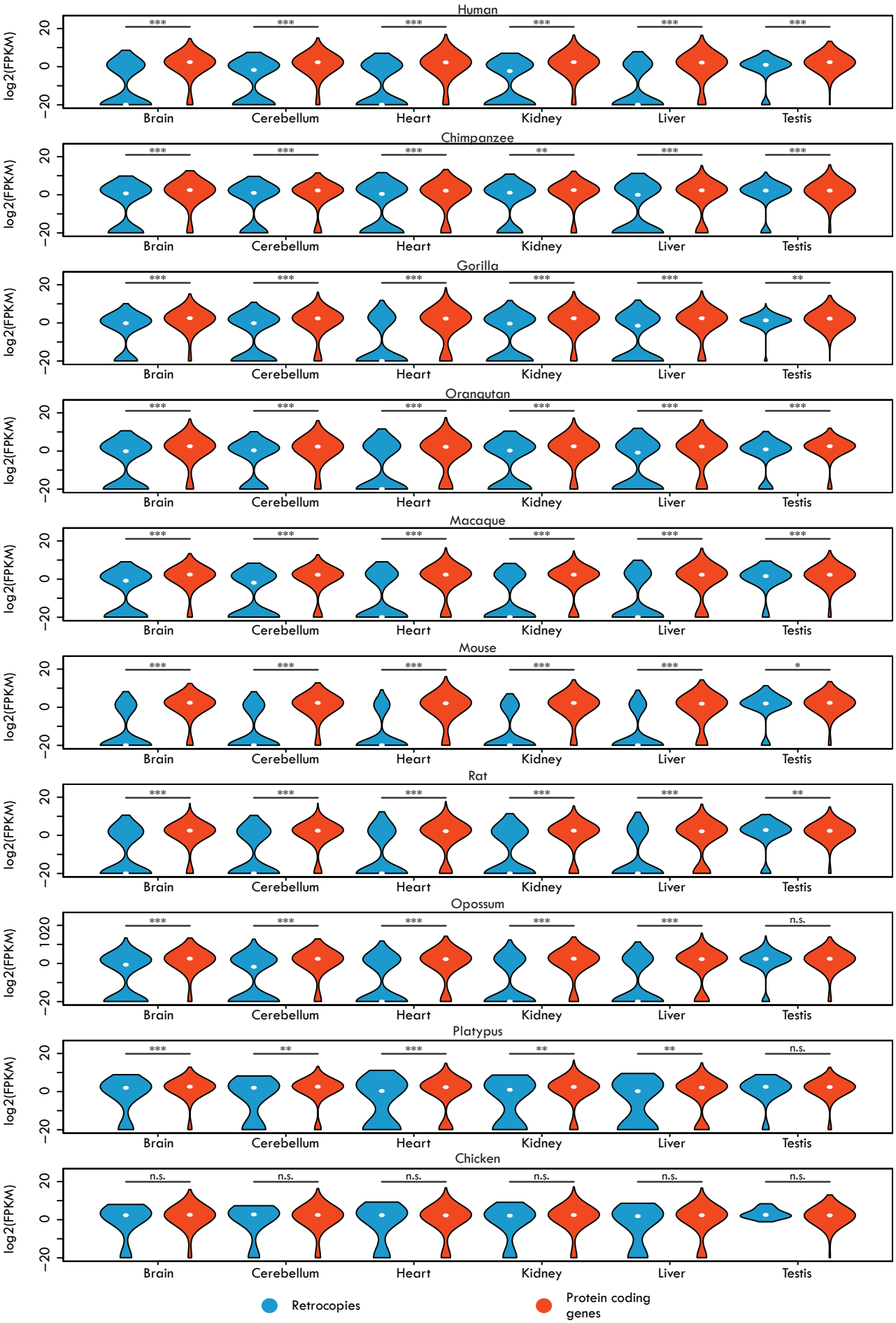
Supplemental Figure S1. d_S dating of retrocopies. Density plot showing the distribution of d_S values for all retrocopies (d_S values from all primates and rodents have been collapsed due to their similarity). d_S values higher than 2 have been collapsed.

Supplemental Figure S2



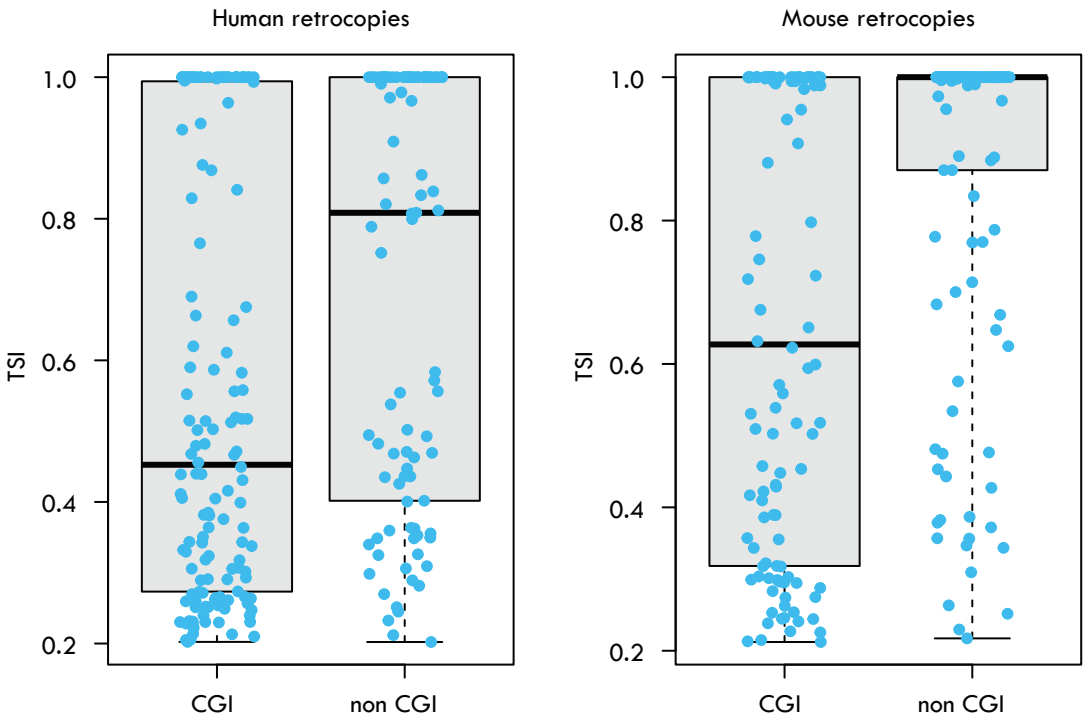
Supplemental Figure S2. Number of expressed retrocopies per tissue based on read subsampling. The number of expressed retrocopies (≥ 1 unique read mapping on their locus) on each organ was calculated after subsampling 5 million uniquely mapped reads from all samples derived from the same organ.

Supplemental Figure S3



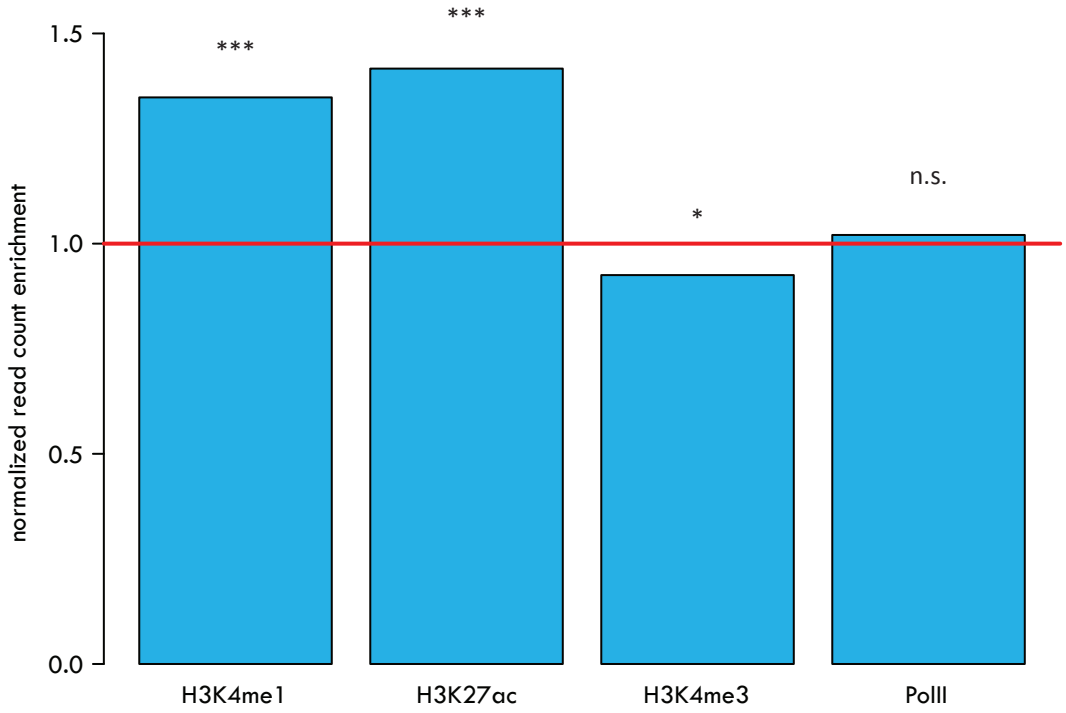
Supplemental Figure S3. Expression level of retrocopies and annotated protein coding genes across species and organs. Significant differences (Mann-Whitney U-test with Benjamini-Hochberg correction): *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; n.s.: $P > 0.05$.

Supplemental Figure S4



Supplemental Figure S4. Tissue specificity of human and mouse retrocopies with CGI and non CGI promoters. Blue dots indicate the TSI of each retrocopy. Whiskers extend up to the maximum and minimum values.

Supplemental Figure S5

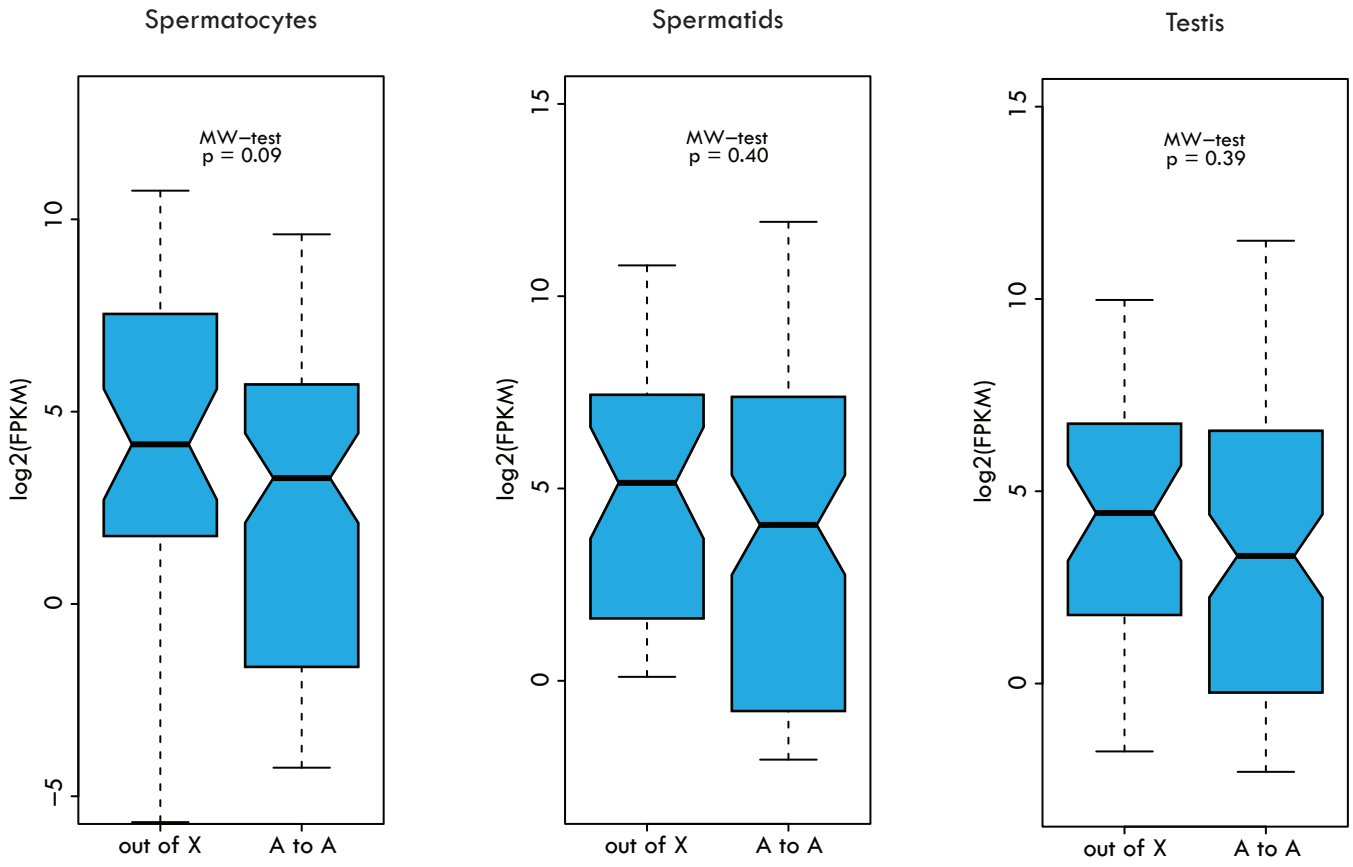


Supplemental Figure S5. Enrichment of enhancer-associated marks at mouse proto-promoters. The height of the bars represent the enrichment in the number of ChIP-seq reads mapping at mouse proto-promoter regions with respect to the reads mapping on the mouse regions orthologous to the loci of rat-specific non expressed retrocopies The read number was normalized by the number of input reads mapping on the same regions with the following formula:

$$\frac{\text{n° ChIP seq reads (proto promoter)} \,/\, \text{n° ChIP seq reads (orthologous sites non exp retrocopies)}}{\text{n° input reads (proto promoter)} \,/\, \text{n° input reads (orthologous sites non exp retrocopies)}}$$

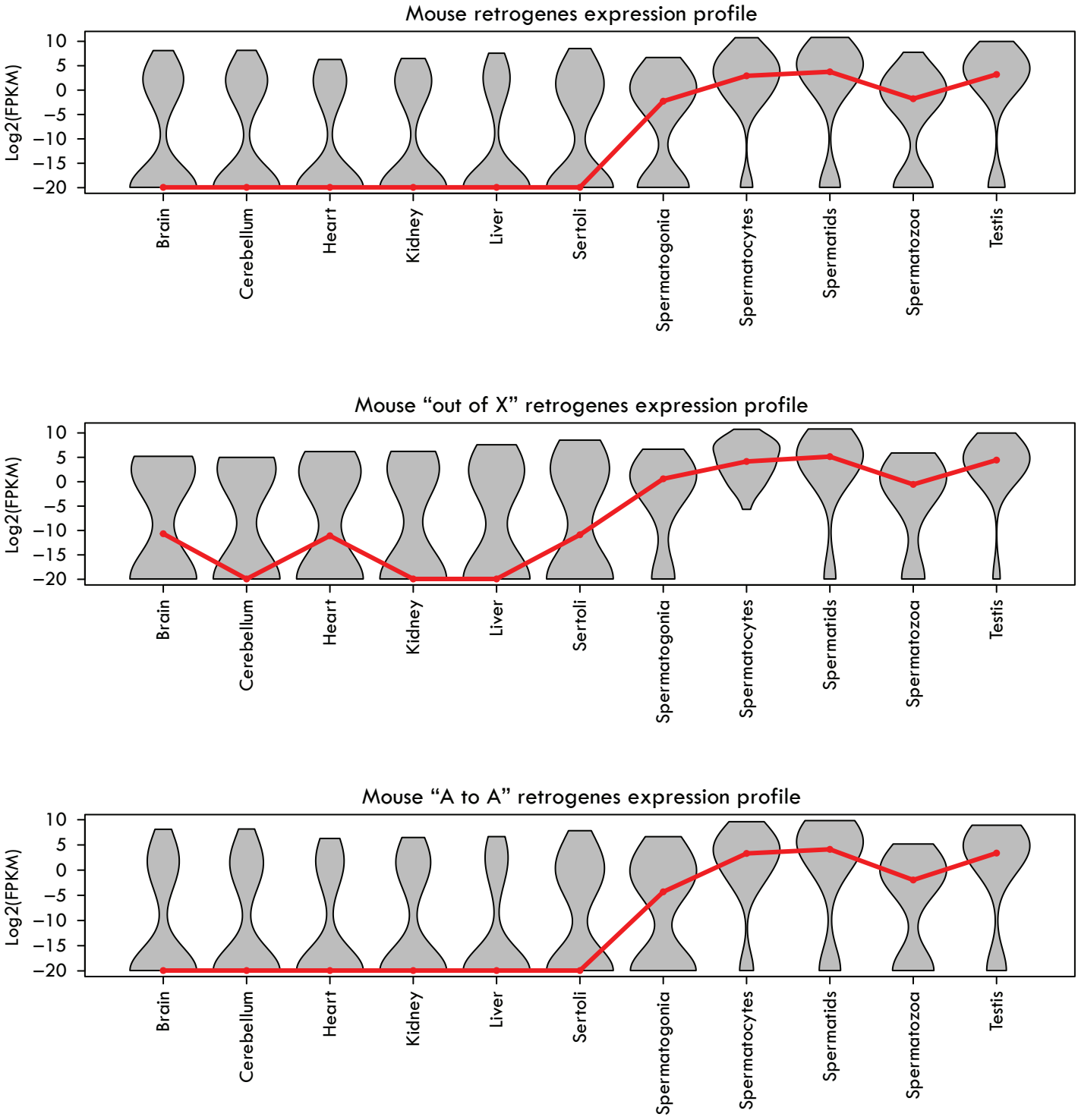
The red line indicates equal number of reads mapping on the two classes of integrations sites (ratio = 1); values higher than 1 indicate enrichment of ChIP-seq reads at proto-promoter sites, while values below 1 indicate depletion. Significant enrichment/depletion was tested with a $\chi 2$ test. Significant differences ($\chi 2$ test with Benjamini-Hochberg correction): *** $P < 0.001$; * $P < 0.05$; n.s.: $P > 0.5$.

Supplemental Figure S6



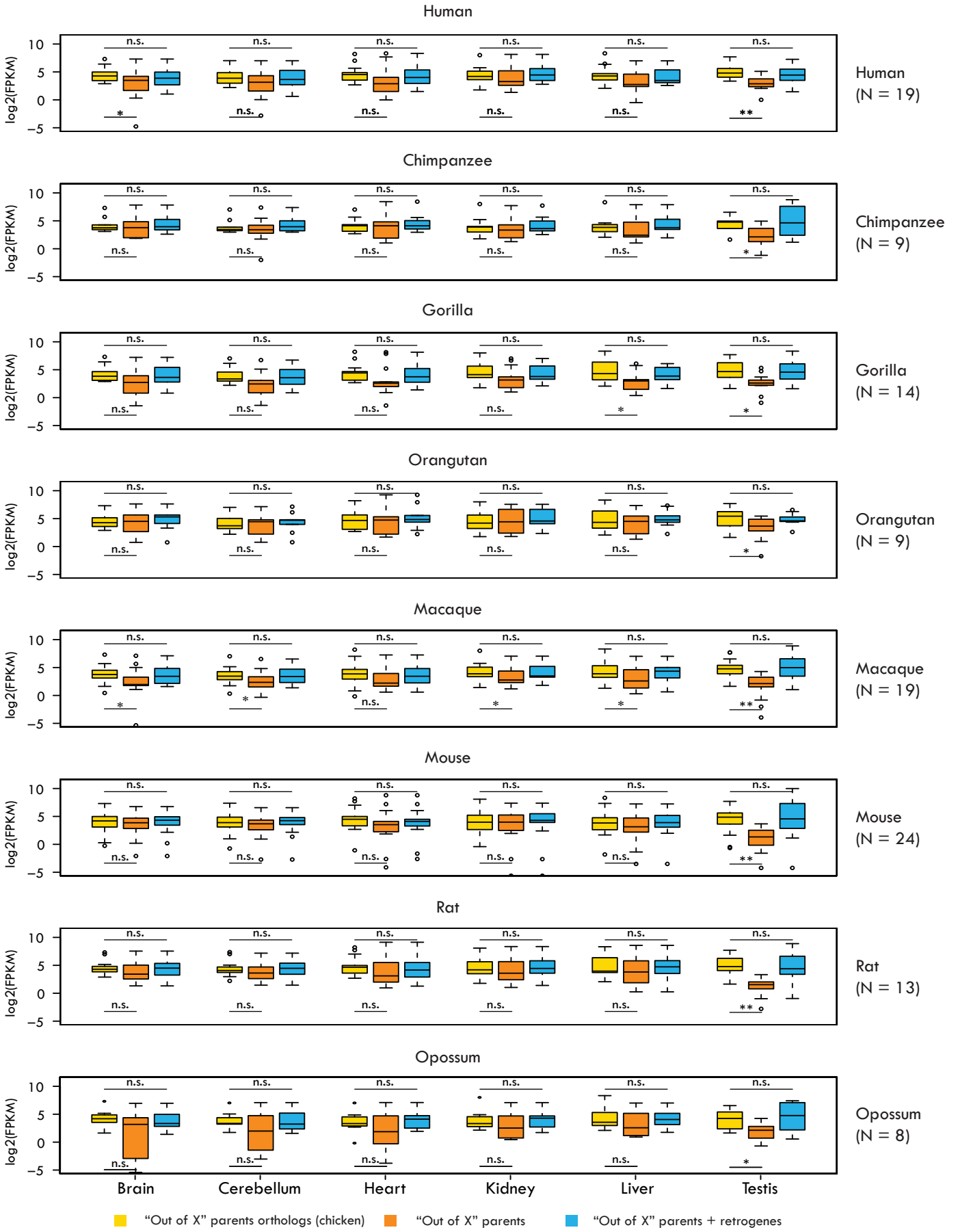
Supplemental Figure S6. Expression level of mouse retrogenes in spermatocytes, spermatids and testis. Differences in expression between "out of X" and "autosome to autosome" retrogenes were tested using a Mann-Whitney U-test and the *P* values corrected for multiple tests (Benjamini-Hochberg correction). Whiskers extend up to 1.5 times the interquartile range; outliers were removed for graphical purposes. "A to A": "autosome to autosome".

Supplemental Figure S7



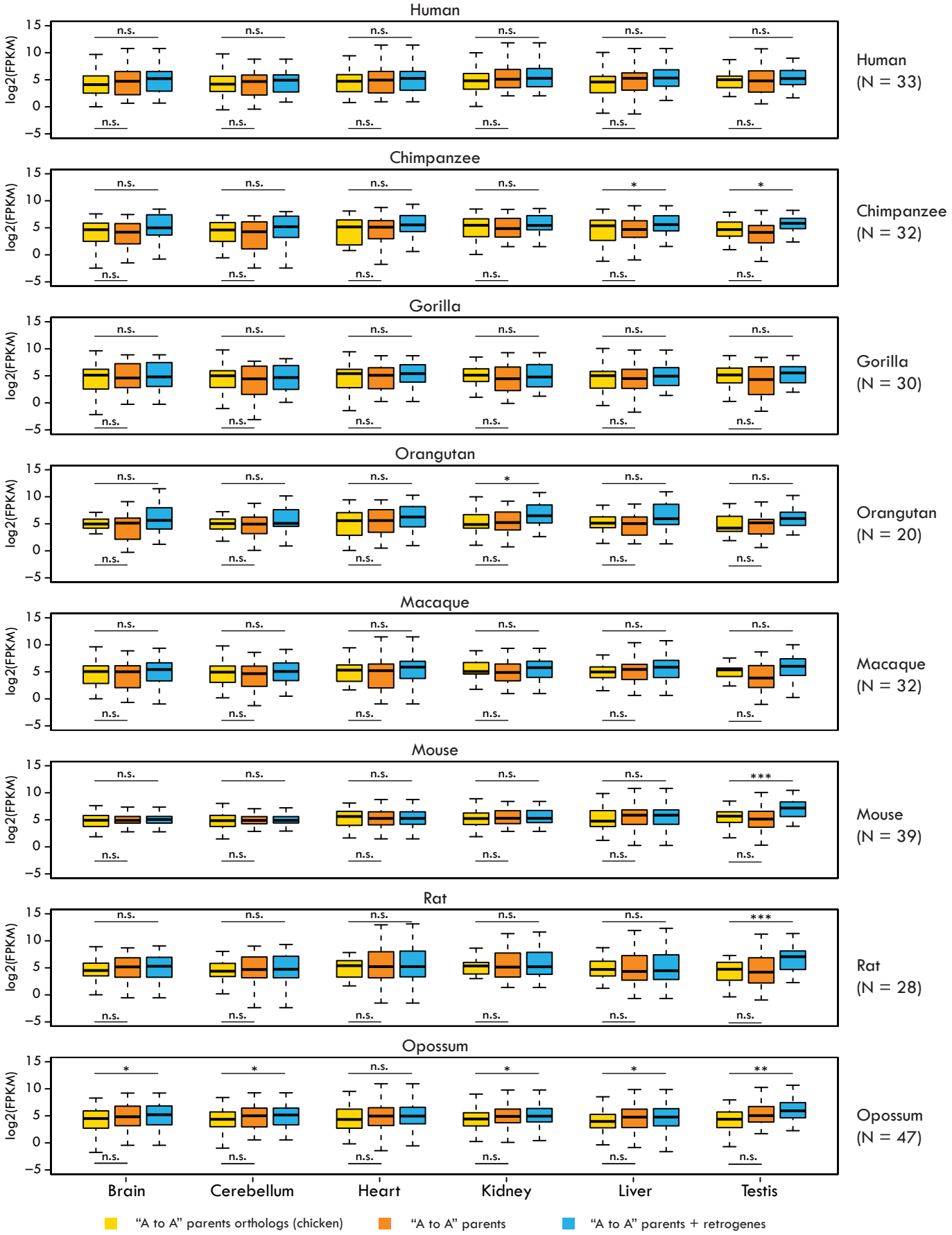
Supplemental Figure S7. Expression level of mouse retrogenes across tissues. Plots show the expression profiles of all retrogenes, of the “out of X retrogenes”, and of the “autosome to autosome” retrogenes in different mouse tissues/organs. Red lines connect the median values from each tissue/organ. “A to A”: “autosome to autosome”.

Supplemental Figure S8



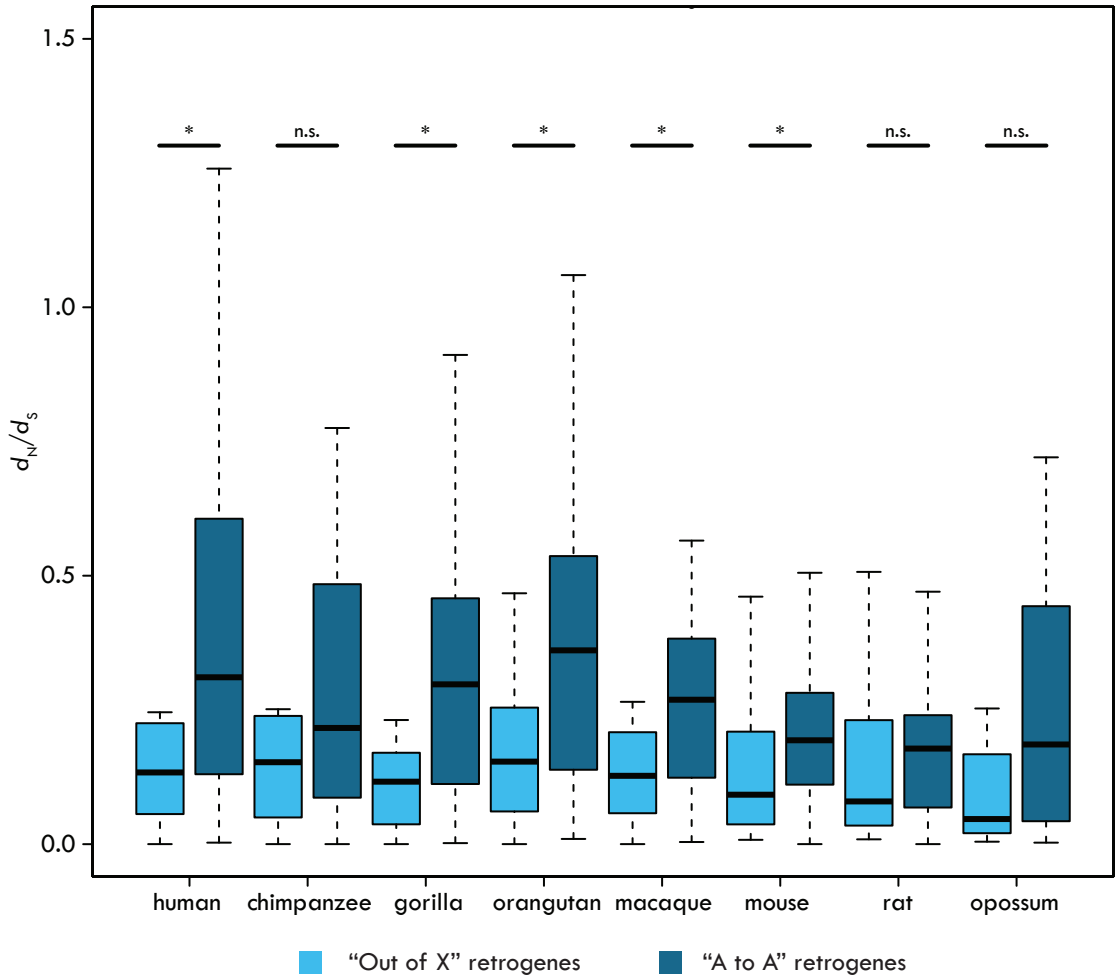
Supplemental Figure S8. Dosage compensation of "out of X" retrogenes. The tissue-specific expression of the orthologous parental genes of "out of X" retrogenes in chicken (yellow), of the "out of X" parental genes (orange) and the summed expression of "out of X" retrogenes and their parental genes (blue) shows that retrogenes compensate the significant decrease in expression of their parents in testis. The numbers of gene pairs tested in all species are reported at the left of each plot. Significant differences (paired Mann-Whitney U-test with Benjamini-Hochberg correction): **: $P < 0.01$; *: $P < 0.05$; n.s.: $P > 0.05$. Whiskers extend up to 1.5 times the interquartile range; a part of the outliers were removed for graphical purposes.

Supplemental Figure S9



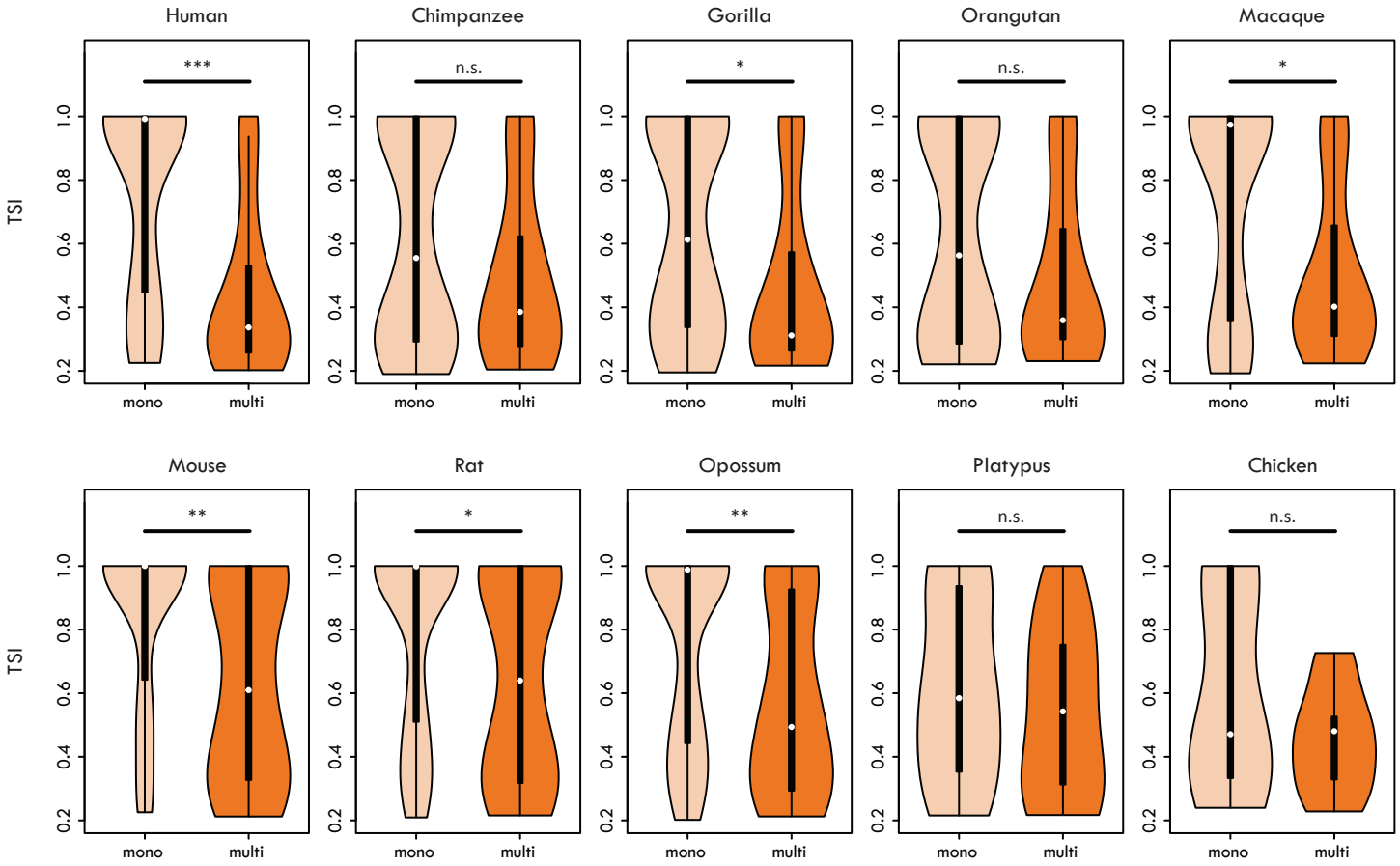
Supplemental Figure S9. Dosage compensation of "A to A" retrogenes. The tissue-specific expression of the orthologous parental genes of "A to A" retrogenes in chicken (yellow), of the "A to A" parental genes (orange) and the summed expression of "A to A" retrogenes and their parental genes (blue) shows that autosomal parental genes do not differ from their chicken orthologs, and that retrogenes tend to overcompensate the expression of their parents in testis. The numbers of gene pairs tested in all species are reported at the left of each plot. Significant differences (paired Mann-Whitney U-test with Benjamini-Hochberg correction): ***: $P < 0.001$; **: $P < 0.01$; *: $P < 0.05$; n.s.: $P > 0.05$. Whiskers extend up to 1.5 times the interquartile range; outliers were removed for graphical purposes.

Supplemental Figure S10



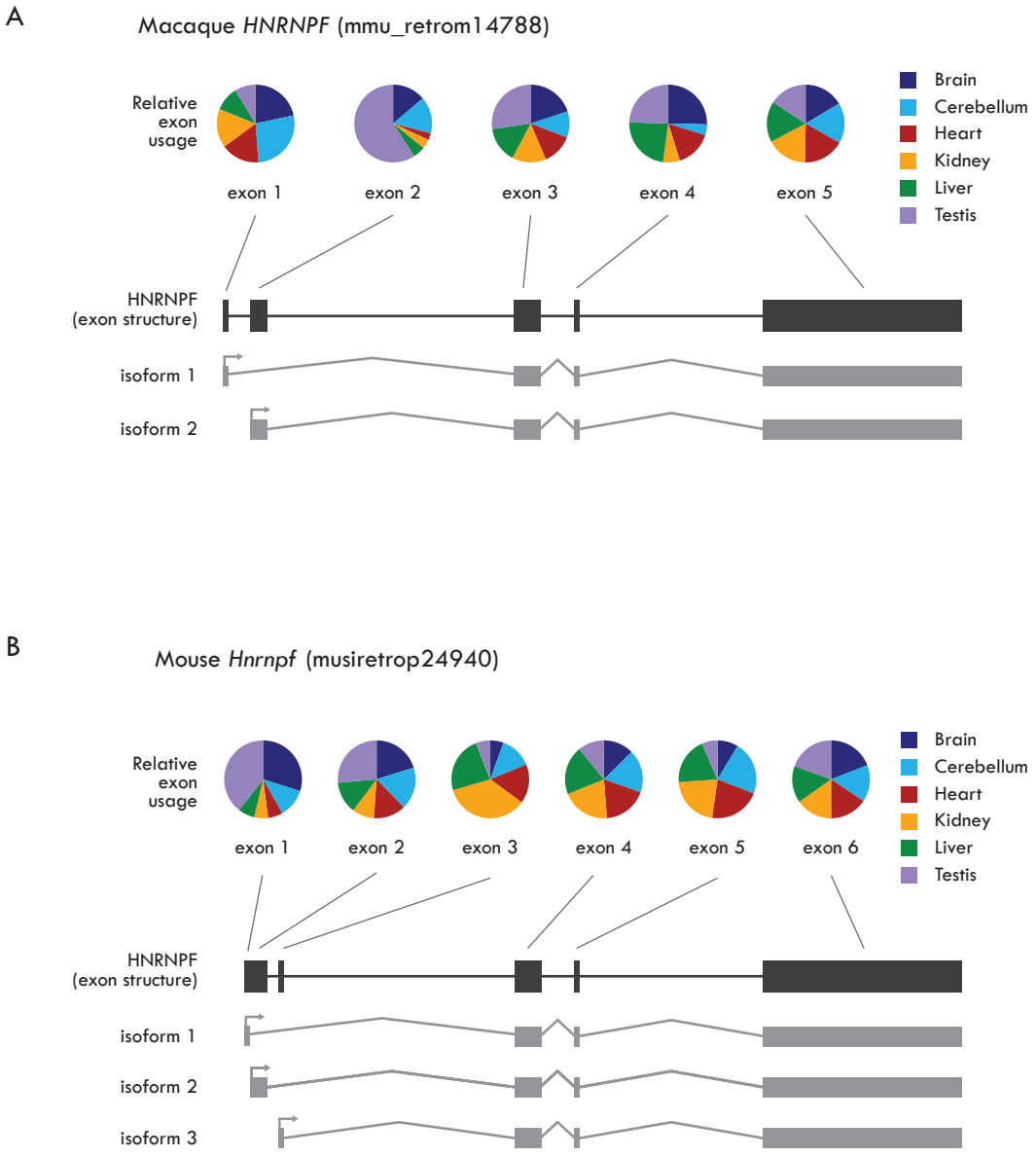
Supplemental Figure S10. Differences in d_N/d_S ratios between "out of X" and "autosome to autosome" retrogenes. d_N/d_S ratios were calculated as described in the Methods ("out of X analysis"). Significant differences (Mann-Whitney U-test with Benjamini-Hochberg correction): *: $P < 0.05$; n.s.: $P > 0.05$. Whiskers extend up to 1.5 times the interquartile range; outliers were removed for graphical purposes. "A to A": autosome to autosome.

Supplemental Figure S11



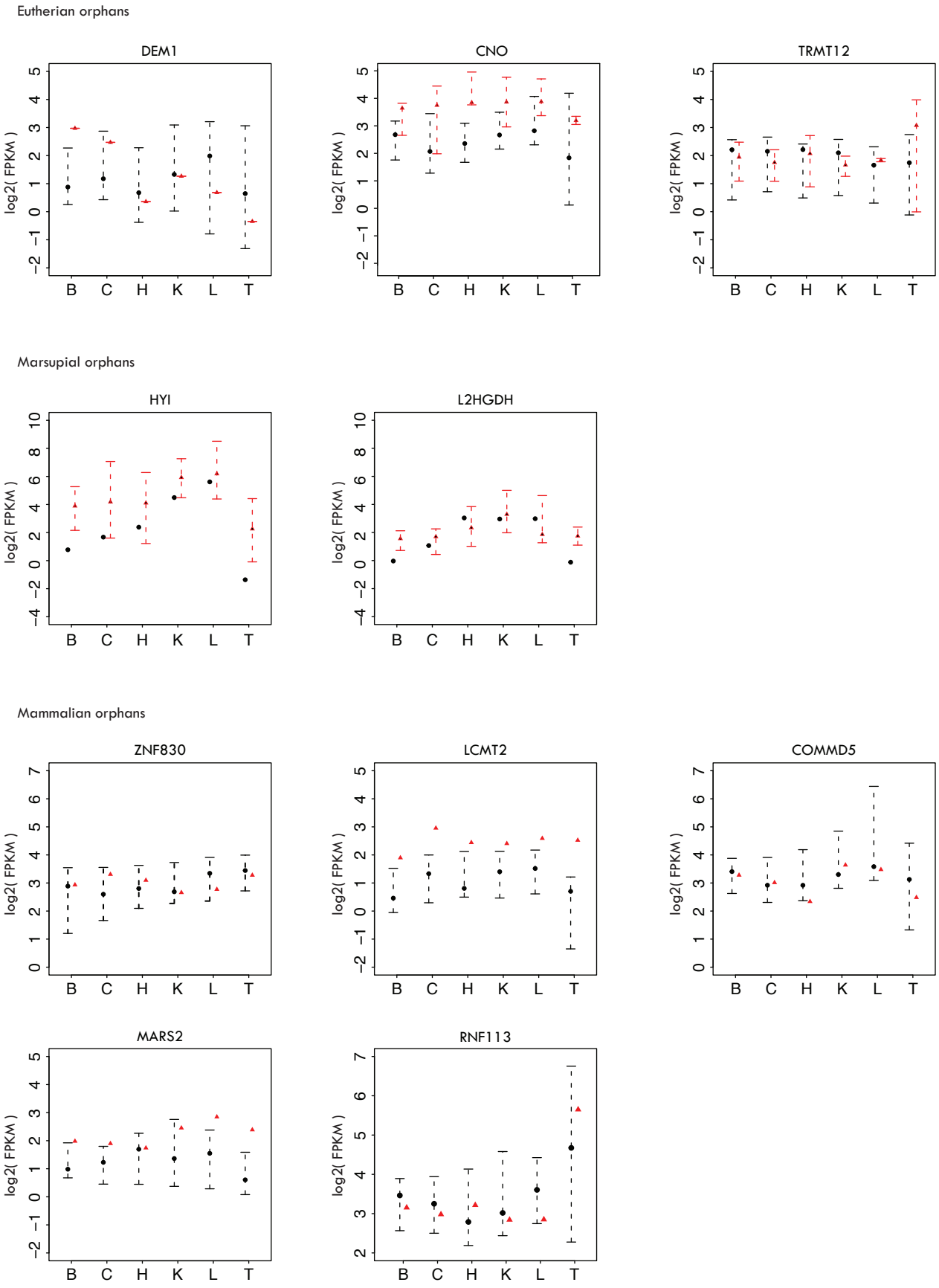
Supplemental Figure S11. Tissue specificity of monoexonic and multiexonic (only new 5' exons considered) retrogenes. The violins indicate retrogene TSI distribution. Significant differences (Mann-Whitney U-test with Benjamini-Hochberg correction): ***: $P < 0.001$; **: $P < 0.01$; *: $P < 0.05$; n.s.: $P > 0.05$. Mono: monoexonic retrogenes; multi: multiexonic (only those with new 5' exons) retrogenes.

Supplemental Figure S12



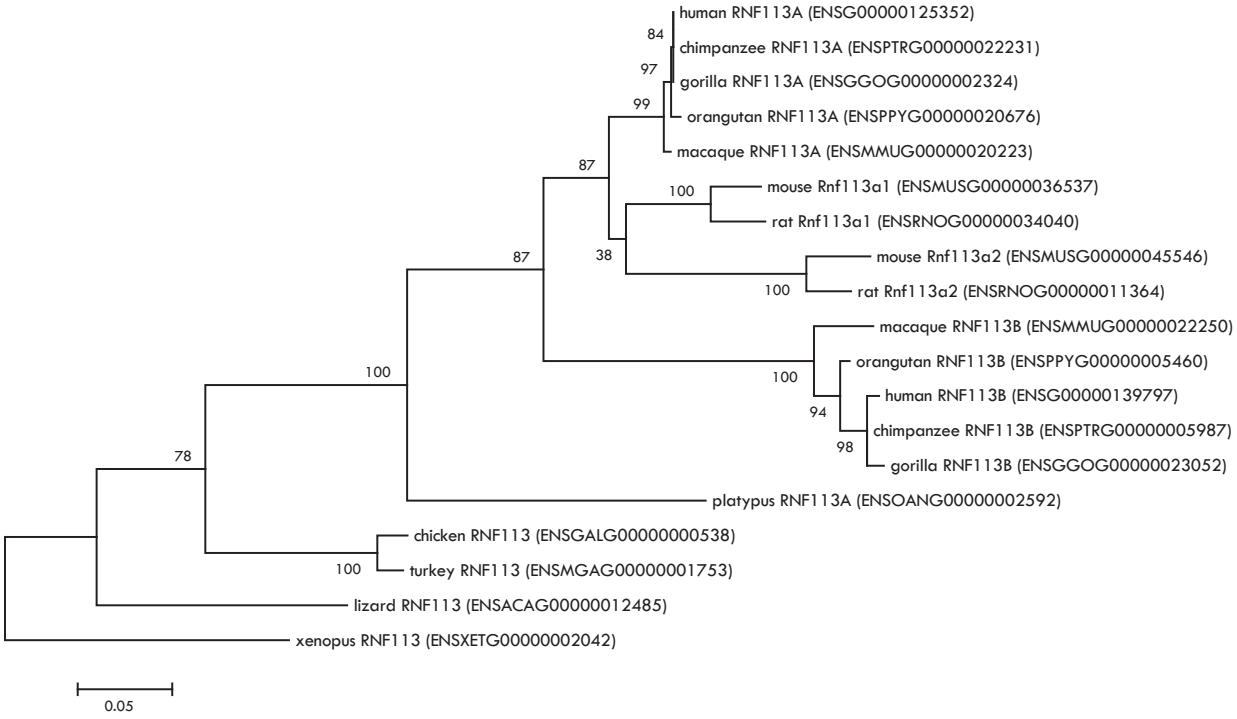
Supplemental Figure S12. Differential exon usage of the *HNRNPF* retrogene in macaque (A) and mouse (B). Top: Relative contribution of each organ in the amount of unique reads mapped on each exon. Relative exon usage is obtained by normalizing the read count of each exon by the total number of reads mapped on the whole gene for each organ. Exon 2 in macaque and exon 1 in mouse are significantly more highly transcribed in testis (DEXSeq analysis, Benjamini-Hochberg corrected $P < 0.01$); exon 3 in mouse is significantly more highly expressed in kidney (DEXSeq analysis, Benjamini-Hochberg corrected $P < 0.01$); other exons are equally expressed in all organs. Bottom: exon structure (black) and alternative transcripts (grey) of the *HNRNPF* gene.

Supplemental Figure S13



Supplemental Figure S13. Expression profile of orphan retrogenes. The expression levels of orphan retrogenes (in black) in six organs are compared to those of their parental genes (in red) in outgroup species. In each plot, black dots represent the expression level of the orphan retrogene in each tissue. When the orphan retrogene is present in more species, the black dots represent the median expression across all orthologues, while the error bars represent the range of expression. Expression levels of outgroup parental genes are represented in the same way (with red triangles/error bars). B: brain; C: cerebellum; H: heart; K: kidney; L: liver; T: testis.

Supplemental Figure S14



Supplemental Figure S14. Phylogenetic tree of the RNF113 gene. The tree was reconstructed based on the cDNA sequences of all genes (downloaded from Ensembl) using MEGA5 (Tamura et al. 2011). Multiple sequence alignment was performed with MUSCLE (Edgar 2004). The evolutionary history of the RNF113 gene was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei 1993). The tree with the highest log likelihood (-4753.7240) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.