

APPENDIX

Supplemental Information

Figure S1: A typical *Drosophila* gene, *gp210*.

Figure S2: Fly and mouse nascent fractions exhibit differences in splicing efficiencies even though the mature pA fractions are well-spliced in both species.

Figure S3: Cotranscriptional splicing efficiency of introns found in single intron genes does not correlate with intron length in either species, or with adjacent exon length in mouse.

Figure S4: Negative correlation between 3'SS ratio and 5'exon length is significant even among different 3'exon lengths, gene lengths, or distances from the end of the intron to the end of the gene.

Figure S5: 5'exon length is only significantly correlated with 3'SS ratio for internal fly introns less than 1,000nt and less than 10% of the gene length.

Figure S6: In mouse, alternative introns have higher 3'SS ratios than constitutive introns at greater distances from the 3'end of the gene.

Figure S7: Introns within single intron genes do not exhibit position effects in mouse liver, but do in *Drosophila* S2 cells.

Figure S8: Gene length is negatively correlated with 3'SS ratio of introns in mouse and fly, regardless of intron position relative to the end of the gene.

Figure S9: Only introns within single intron genes do not have any correlation between 3'SS ratio and gene length in mouse.

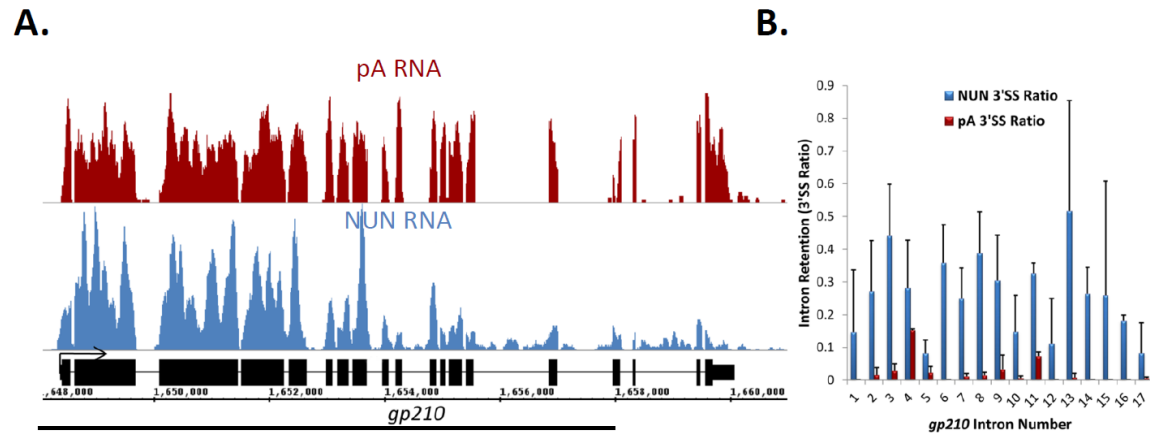


Figure S1: A typical *Drosophila* gene, *gp210*. **A.** An image of a typical fly gene, *gp210*, in the Integrated Genome Browser (IGB). pA RNA is in red, NUN RNA is in blue, and gene structure is in black. Note 5' to 3' abundance gradient in NUN fraction and low read signal within introns. Black bar represents 10,000bp. NUN average reads/bp ~30. **B.** Quantitation of intron retention for the introns of *gp210*. Intron Retention as 3'SS Ratio = Reads in last 25bp of Intron / Reads in first 25bp of 3' Exon.

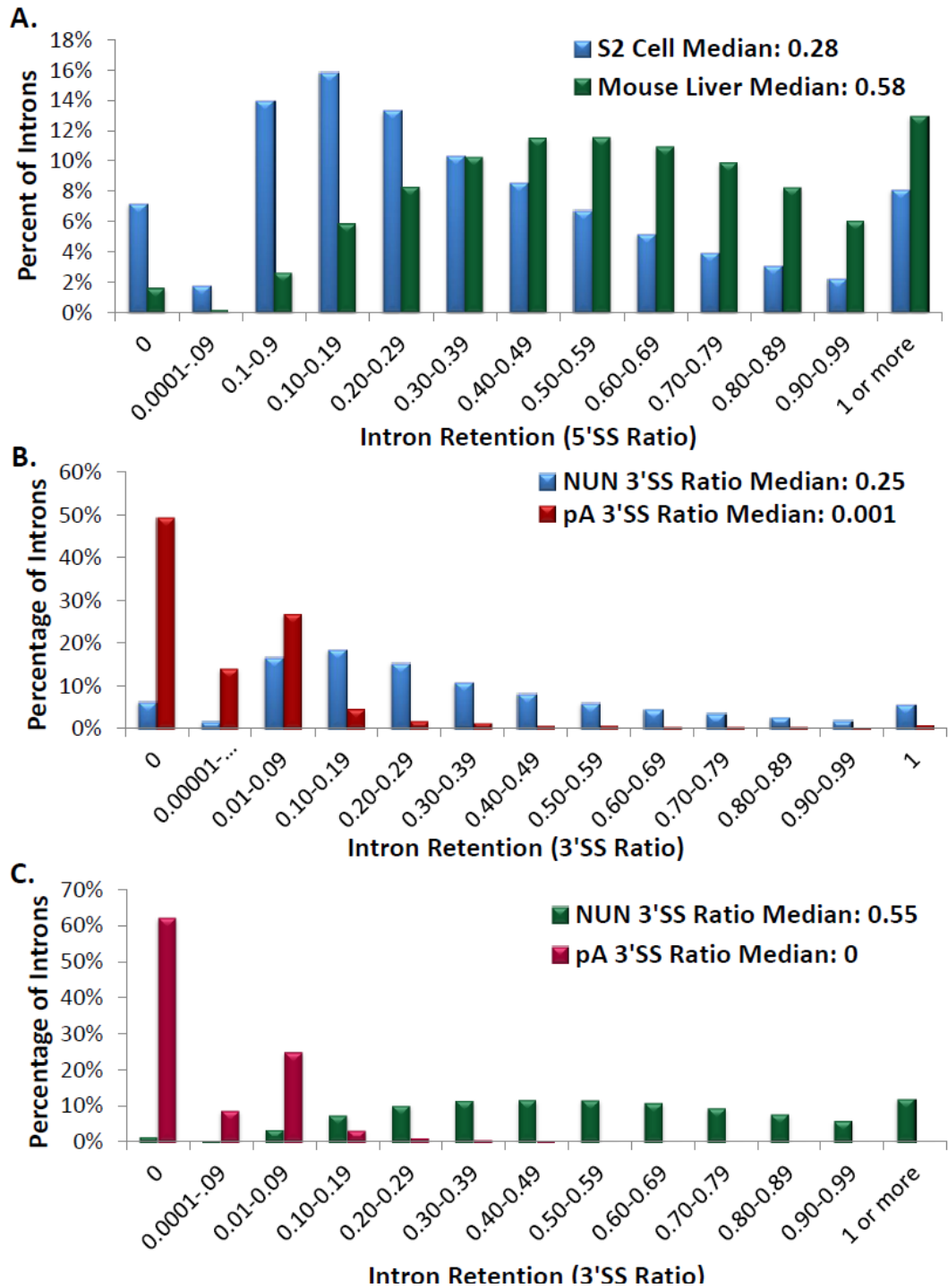


Figure S2: Fly and mouse nascent fractions exhibit variable splicing efficiencies even though mature pA fractions are well spliced in both species. **A.** A histogram of the percent of all introns of abundantly transcribed genes, grouped by 5'SS ratio in mouse and fly. Total sample size: Fly=19,607, Mouse=58,492. **B.** A histogram of the percent of all introns of abundantly transcribed genes, grouped by 3'SS ratio in fly. pA RNA is in red and NUN RNA is in blue. Total sample size: NUN=20,335, pA=23,795 introns. **C.** The same histogram for introns of abundantly transcribed mouse genes. pA RNA is in magenta and NUN RNA is in green. Total sample size: NUN=58,493, pA=58,491 introns.

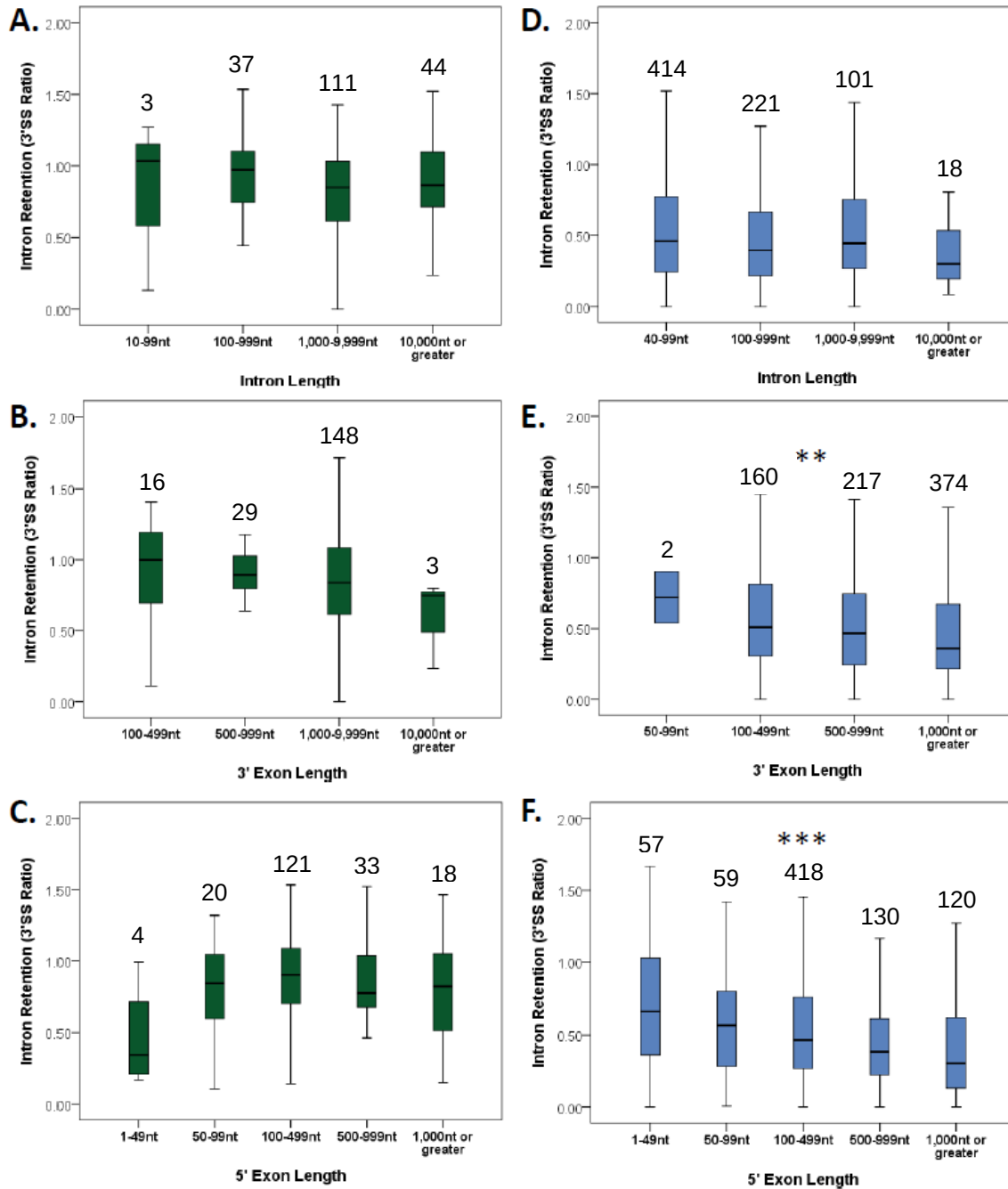


Figure S3: Cotranscriptional splicing efficiency of introns found in single-intron genes does not correlate with intron length in either species, or with adjacent exon length in the mouse. A, D. Intron length does not correlate with 3'SS ratio in mouse (A) ($P=0.414$, Kruskal-Wallis Test) or fly (D) ($P=0.434$, Kruskal-Wallis Test). **B, E.** 3' exon length does not correlate with 3'SS ratio in mouse (B) ($P=0.143$, Kruskal-Wallis Test), but does in fly (E) ($P=0.003$, Kruskal-Wallis Test). **C, F.** 5' exon length does not correlate with 3'SS ratio in mouse (C) ($P=0.273$, Kruskal-Wallis Test), but does in fly (F) ($P<0.001$, Kruskal-Wallis Test).

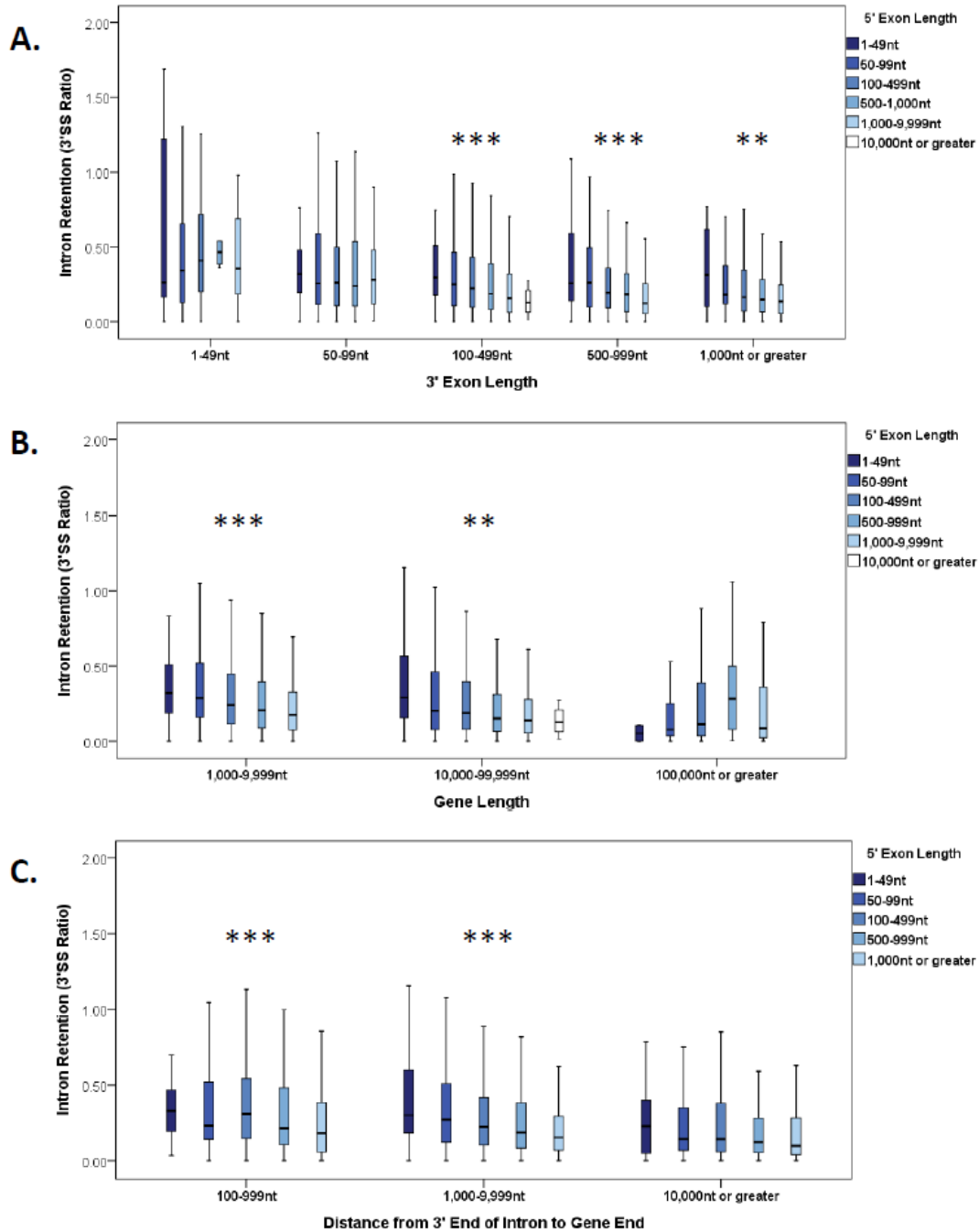


Figure S4: Negative correlation between 3'SS ratio and 5' exon length is significant even among different 3' exon lengths, gene lengths, or distances from the end of the intron to the end of the gene. **A.** 3'SS ratio is negatively correlated with 3' exon length for exons which are 100nt or greater, but not for the minority of exons that are less than 100nt long. (100-499nt $P < 0.001$; 500-999nt $P < 0.001$; 1,000nt+ $P = 0.003$, Kruskal-Wallis Test) **B.** 3'SS ratio is negatively correlated with gene length for genes less than 100,000nt long, but not the minority of genes which are longer than 100,000nt. (1,000-9,999nt $P < 0.001$; 10,000-99,999nt $P = 0.006$, Kruskal-Wallis Test) **C.** 3'SS ratio is negatively correlated with the distance of the intron from the end of the gene for introns less than 10,000nt away, but not further (100-999nt $P > 0.001$; 1,000-9,999nt $P < 0.001$, Kruskal-Wallis Test).

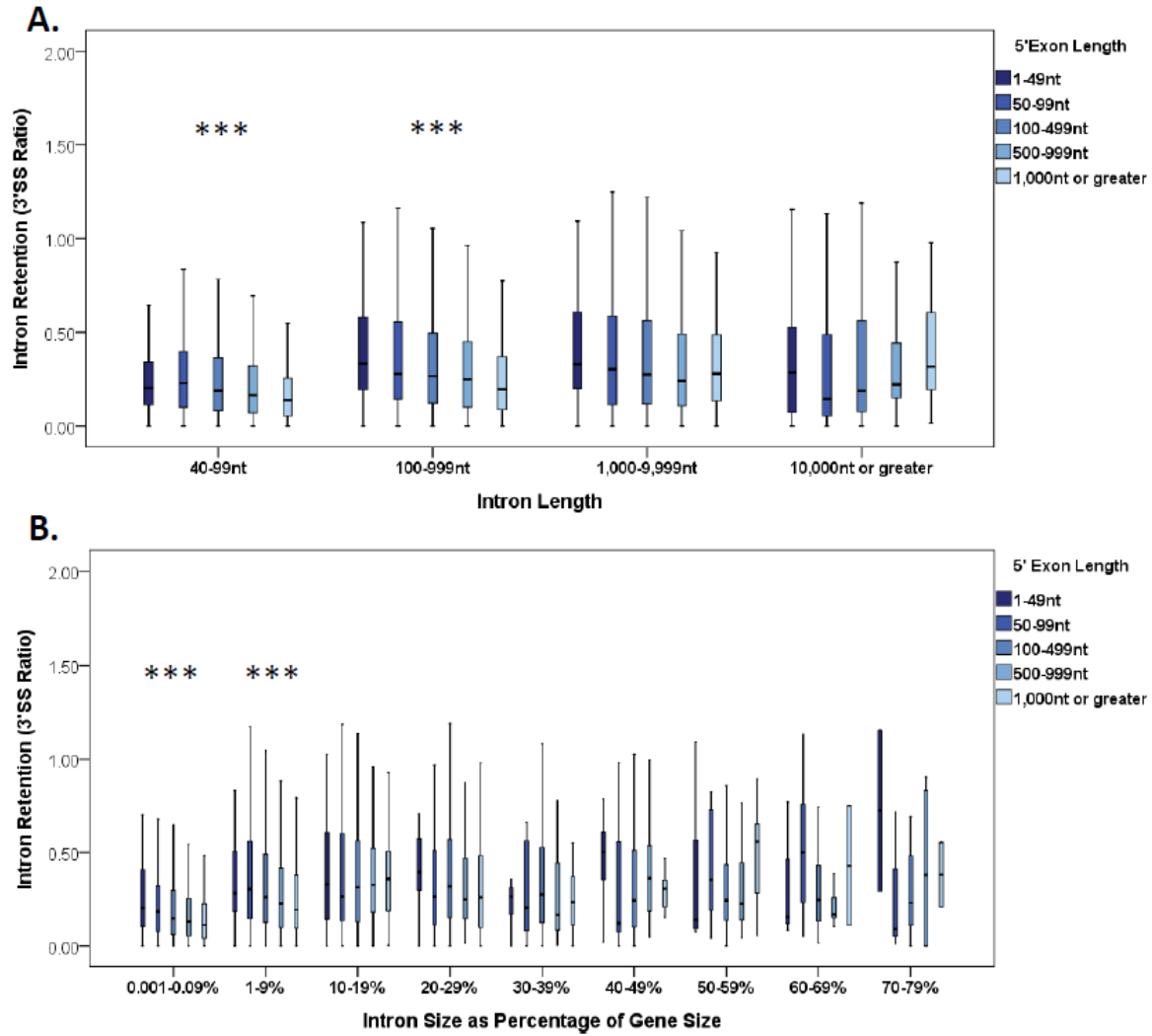


Figure S5: 5' exon length is only significantly correlated with 3'SS ratio for internal fly introns less than 1,000nt and less than 10% of gene length. A. Fly internal introns less than 1,000nt long show a statistically significant correlation between 5' exon length and 3'SS ratio (40-999nt $P < 0.001$, Kruskal-Wallis Test). **B.** Fly internal introns that are less than 10% of the size of the gene show a statistically significant correlation between 5' exon length and 3'SS ratio ($>10\%$ $P < 0.001$, Kruskal-Wallis Test).

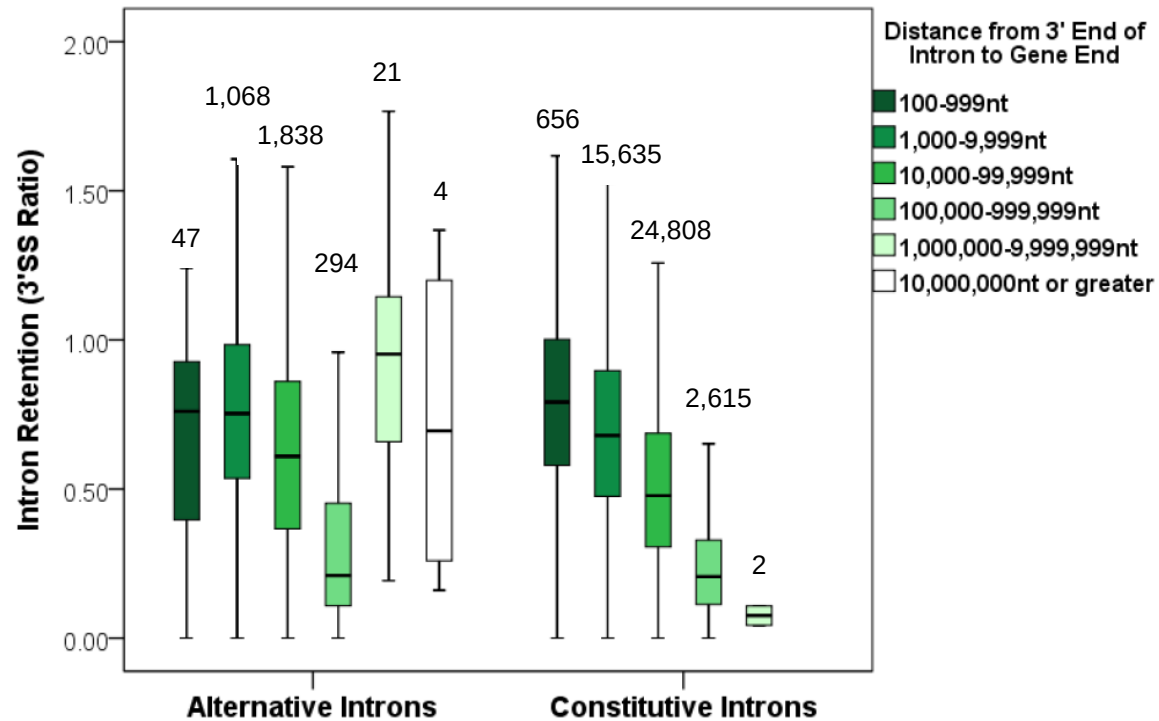


Figure S6: In mouse, alternative introns have higher 3'SS ratios than constitutive introns at greater distances from the 3' end of the gene. Internal, alternatively-annotated introns have significantly different 3'SS ratios than constitutively-annotated introns at every distance from the 3' end of the gene (100-999nt: $P=0.04$, Mann-Whitney U-Test; 1,000-99,999nt: $P<0.001$, Mann-Whitney U-Test; 100,000-999,999nt: $P=0.038$, Mann-Whitney U-Test; 1,000,000-99,999,999nt: $P=0.008$, Mann-Whitney U-Test). Additionally, only alternatively-annotated introns occur 10,000,000nt or greater away from the 3' end of the gene, and these are significantly retained.

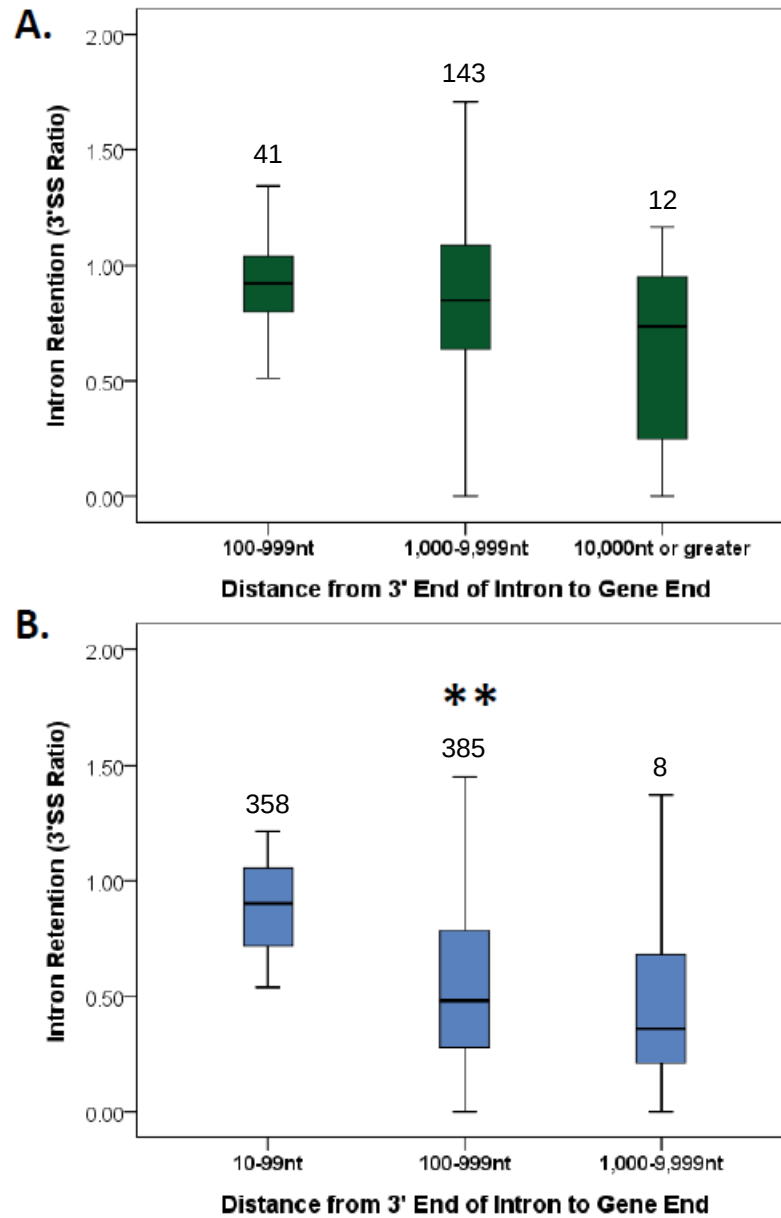


Figure S7: Introns within single intron genes do not exhibit position effects in mouse liver, but do in *Drosophila* S2 cells. A. There is no statistically significant difference in the distributions of mouse introns within single-intron genes based on their distance to the 3' end of the gene ($P=0.052$, Kruskal-Wallis Test). **B.** In fly, the closer an intron within a single-intron gene is to the end of that gene, the higher the 3'SS ratio ($P=0.004$, Kruskal-Wallis Test).

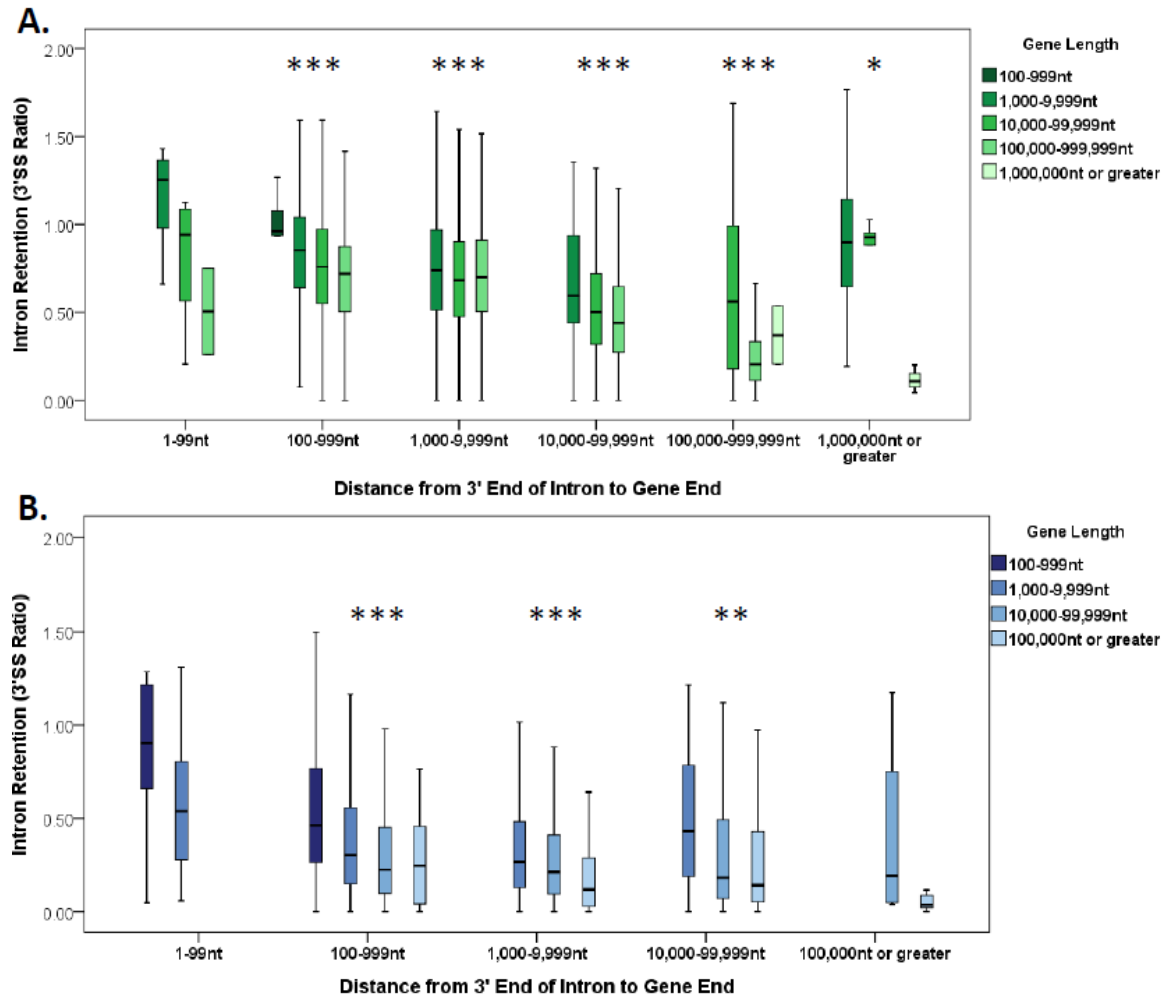


Figure S8: Gene length is negatively correlated with 3'SS ratio of introns in mouse and fly, regardless of intron position relative to the end of the gene. A. In mouse liver, all introns further than 100nt away from the end of the gene show statistically significant differences in 3'SS ratio based on gene length (100-999,999nt away from 3' end of gene $P < 0.001$; 1,000,000nt + $P = 0.026$, Kruskal-Wallis Test). **B.** In fly tissue, all introns between 100-100,000nt away from the 3' end of the gene show statistically significant differences in 3'SS ratio based on gene length (100-9,999nt $P < 0.001$; 10,000-99,999nt $P = 0.006$, Kruskal-Wallis Test)

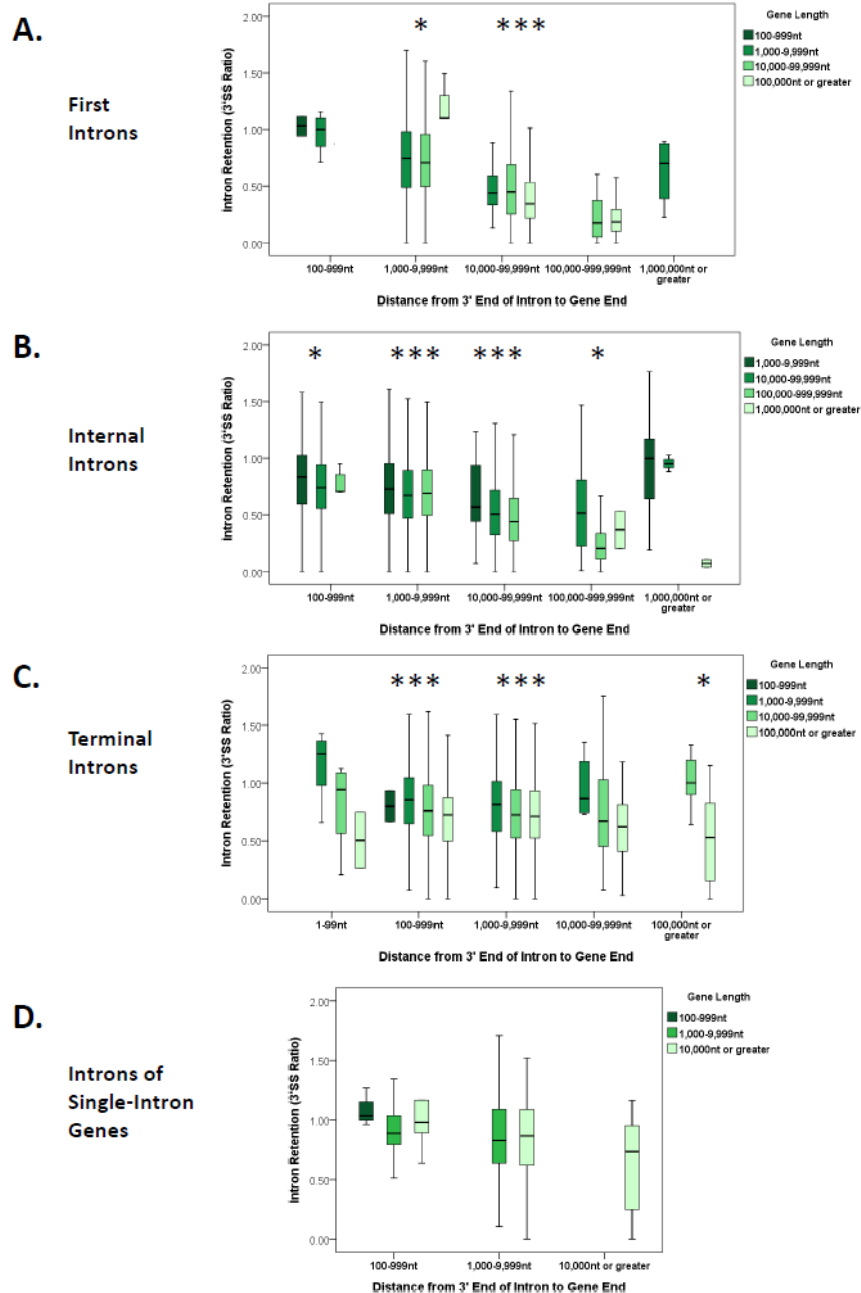


Figure S9: Only introns within single intron genes do not have any correlation between 3'SS ratio and gene length in mouse. A. First introns 1,000-99,999nt from the 3' end of a gene show statistically significant differences in 3'SS ratio by gene length (1,000-9,999nt $P=0.017$; 10,000-99,999nt $P<0.001$, Kruskal-Wallis Test). **B.** All internal introns except those 1,000,000nt or more away from the 3' end show a statistically significant correlation between 3'SS ratio and gene length (100-999nt $P=0.012$; 1,000-99,999nt $P<0.001$; 100,000-999,999nt $P=0.03$, Kruskal-Wallis Test). **C.** Most final introns show a statistically significant correlation between 3'SS ratio and gene length (100-9,999nt $P<0.001$; 100,000nt+ $P=0.028$, Kruskal-Wallis Test) **D.** There are no statistically significant differences in the 3'SS ratios of introns within single-intron genes when evaluated across gene lengths.