Supplementary Figure 1: Constructing and searching a database of exon-exon junctions. a) For each FlyBase annotated intron, 36 bases were taken from both neighboring exons and concatenated (gray bases represent the intron separating the two exons). Reads that could not be mapped to the reference genome were then searched against this database using MAQ. b) Examples of read mappings allowed are shown. Note that each mapped read must have at least 5 non-mismatch bases mapped to either side of the exon-exon junction. Mismatches are shown in red.

Supplementary Figure 2: Searching for imperfect intron deletions. a) For each FlyBase annotated intron having average read depth lower than 1.0, 36 bases on each side of the region with zero read depth were concatenated to form another database. Gray nucleotides represent the putatively deleted intron. Reads that could not be mapped to the reference genome were then searched against this database using BWA, which allows gaps. b) An example of read mappings confirming an imperfect intron deletion is shown. In this example read mappings show that the deletion does not span the entire region of zero depth. Instead, the deletion begins with the “CG” just upstream of the intron, and ends before the “CAAG” at the end of the intron.

Supplementary Figure 3: Using paired-ends to validate retroCNVs and intron deletions. Whether due to a missing intro or a retrotransposition event, paired-reads crossing an exon-exon junction not separated by an intron will appear farther apart then expected when mapped to a reference genome containing an intron.

Supplementary Figure 4: PCR validation of retroCNVs and intron deletions. a) The design of the PCR experiment is shown. Primers are shown as colored arrows, with the blue arrow representing the downstream primer and the red arrow representing the upstream primer. When a retroCNV is present, two amplified fragments are expected: one large fragment containing an intron (from the parental copy) and one small fragment missing an intron). When an intron is deleted, only the smaller fragment is expected. b) A gel image of PCR validation of retrogenes. Note that validated retroCNVs yield two fragments, one large and one small,

while both unconfirmed retroCNVs (likely false positives) and controls yield only a large fragment.