Three quarters of the human genome is transcribed. High-throughput sequencing supports alternative splicing for >95% of human genes. Most genes produce ≥10 different transcripts and contain many sites of alternative splicing. These estimations were obtained by sequencing millions of transcript fragments, each usually much shorter than the originating molecule (i.e. short reads).

Investigation into potential intramolecular coordination between sites of alternative splicing separated by distances greater than short read lengths is often impossible. Thus, we developed a novel RNA-templated DNA-DNA ligation technique (SeqZip) that maintains distant intramolecular sequence connectivity. We used SeqZip to investigate coordination between sites of alternative splicing in mouse Fibronectin. We also used SeqZip to examine potential coordination among three clusters of 92 cassette exons capable of producing 18,112 different *Drosophila melanogaster* *Dscam1* isoforms. No coordinated splicing was observed in either gene.

Accurate transcript assembly and annotation from short reads requires orthogonal forms of sequencing data including those enriched for reads originating from the 5´ and 3´ ends of long transcripts. Mammalian piRNAs, required for male fertility, are generated from long precursor transcripts. Through the integration of orthogonal data sets, we generated a list of 467 precursor transcripts from 214 genomic loci that account for >95% of adult pachytene piRNAs. Accurate transcript definitions lead to the identification of A-MYB, a transcription factor that regulates pachytene piRNA-producing genes during mammalian spermatogenesis.

While a powerful genome-wide approach, transcript annotation and study via short read high-throughput sequencing often requires supporting experimental and analytical tools.