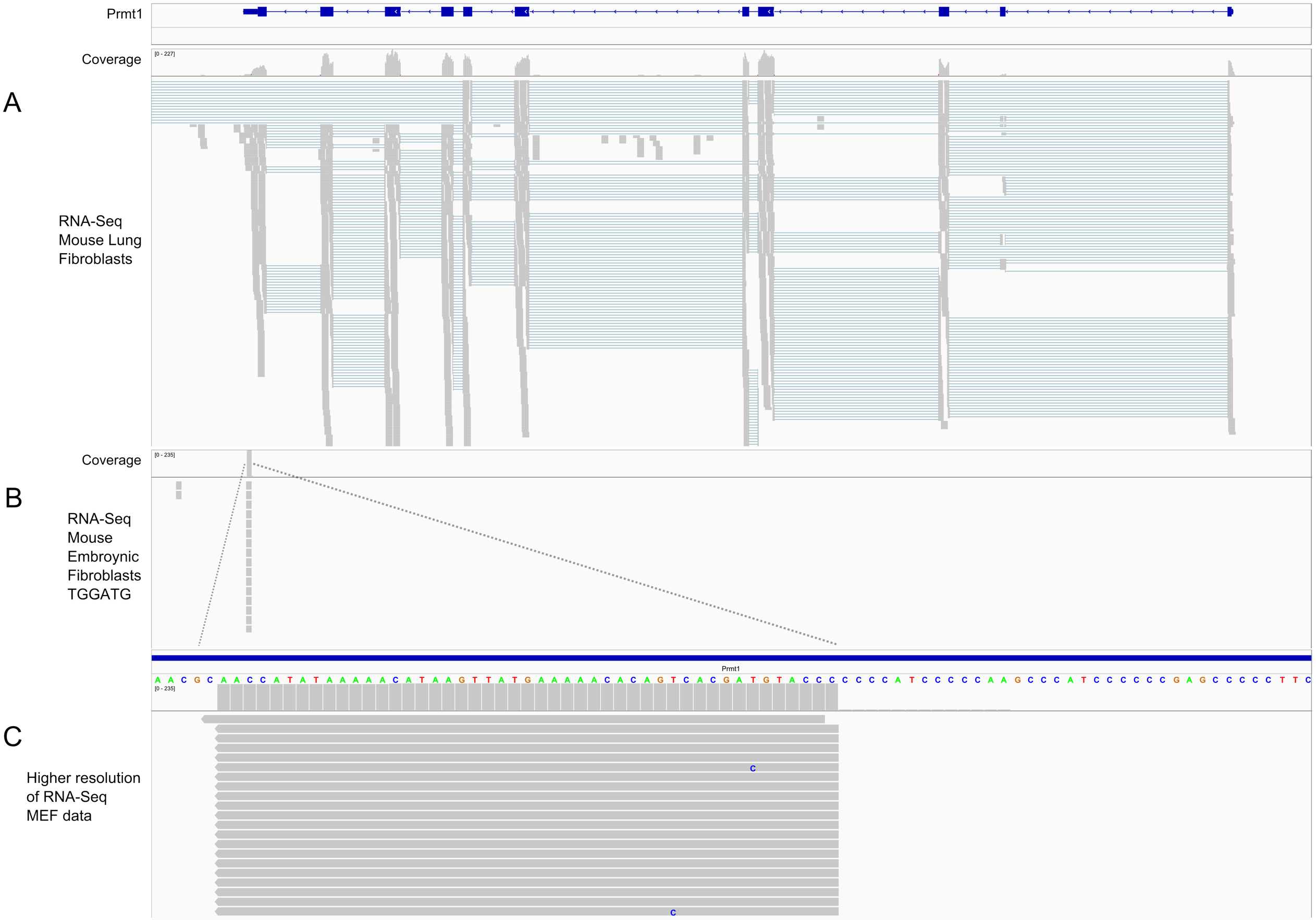

Supplementary figure 1. RNA-Seq data for the gene Prmt1 (top panel) from mouse fibroblasts from (A) Guttman et al. and (B) Islam et al. visualized using the Integrative Genomics Viewer (IGV). The coverage track displays the depth of the reads displayed across the Prmt1 gene as a grey bar chart. The Guttman dataset was mapped using Tophat and exon-exon boundaries are shown as lines connecting different exons. RNA-Seq reads from Guttman et al. are mapped all across the exons. In comparison to the Islam et al. dataset, the majority of the RNA-Seq reads are mapping to the 3′ UTR. The single end reads were mapped using Bwa. The bottom panel (C) zooms into the 3′ UTR region and the nucleotide information is shown for the positive strand of the genome; the Prmt1 is on the negative strand. Upstream of the reads in panel C are stretches of guanosines (on the negative strand), and during first strand cDNA synthesis they become sites for strand invasion during template switching. Since the Islam et al. protocol uses oligo-dT primers, almost all first strand cDNA synthesis is interrupted and all reads stop at the 3′ UTR.

Supplementary figure 2. RNA-Seq data for the gene Tuba1a (top panel) from mouse fibroblasts from (A) Guttman et al. and (B, C) Islam et al. visualized using the Integrative Genomics Viewer (IGV). The coverage track displays the depth of the reads displayed across the Prmt1 gene as a grey bar chart. The Guttman dataset was mapped using Tophat and exon-exon boundaries are shown as lines connecting different exons. RNA-Seq reads from Guttman et al. are mapped all across the exons. We show RNA-Seq data from two mouse fibroblast cells in the Islam et al. dataset. In panel (B), we observe a large peak where the majority of reads map to. These reads were detected as artifacts from our filtering method due to the high sequence complementarity of the upstream sequence to the barcode. In panel (C), the large peak in panel (B) is missing, and reads are more distributed along the last exon. While the RNA-Seq data from Islam et al., is made from different heterogenous single cells it seems unlikely that these are true differences given that we do not observe this peak in other mouse fibroblast data and our filtering method detects this as a putative region for strand invasion.