12 - 14 June 2026 Tokyo , Japan

Validation of Novel RNA-seq–Derived Transcripts Identified by StringTie Using Isoform-Specific RT-PCR and Sanger Sequencing of the Validated Transcripts

Maiyon Park , Vartan Shahnazarian , Jason Ly , Daniel Lim , Sharon Asariah , San San Lwin , Joseph Dhahbi

California University of Science and Medicine (CUSM), the United States

Abstract

StringTie is a widely used transcriptome assembly and quantification tool for RNA-seq data. It reconstructs transcript structures and identifies known and potentially novel isoforms. A key advantage of StringTie is its ability to detect transcript models that are absent from existing annotations. It achieves this by reconstructing exon–intron architectures and splice junctions at each genomic locus, enabling discovery of previously unannotated transcripts. These novel transcript loci are labeled with “MSTRG” when they do not correspond to known gene identifiers. Such MSTRG transcripts may represent entirely new genes or alternative isoforms arising from known loci. Despite their potential importance, novel transcripts identified computationally are often not experimentally validated. To address this gap, we investigated whether StringTie-predicted novel transcripts could be detected in cells. Total RNA was isolated from HEK293T cells, and isoform-specific primers were designed to target 17 novel transcripts containing unique exon–exon junctions. Reverse transcription-PCR analysis revealed that 10 of the 17 transcripts produced products of the expected size, suggesting positive detection. These PCR products are currently being validated by Sanger sequencing at the UCR Genomics Core. To further characterize these transcripts, open reading frames were predicted using the NIH ORFfinder tool. Subsequent analysis with NCBI Protein BLAST indicated that the predicted proteins correspond to either variants of known proteins or previously uncharacterized proteins. These variants include N-terminal and C-terminal truncations, as well as sequences containing mutations. We will perform immunoprecipitation and protein sequencing to confirm expression and investigate their functional roles, with a particular focus on their association with Alzheimer’s disease.

Keywords: Aging; Alzheimer; MSTRG