CONTROL ID: 3386581
CONTACT (INSTITUTION ONLY): UTHSC
CONTACT (NAME ONLY): Hyo Young Choi
TITLE: SCISSOR: a novel framework for identifying structural changes in RNA transcripts
ABSTRACT BODY:
Abstract Body: We propose a statistical method, SCISSOR, for unsupervised screening of a range of structural alterations in RNA-seq data. Compared to other existing methods relying on a limited subset of RNA-seq data available, e.g. exon/gene level expression or junction split reads, we consider a novel shape property of aligned short read data through a base-level pileup file. This intact and uncompressed view of RNA-seq profile enables the unbiased discovery of structural alterations by looking for anomalous shapes in expression.

Shape changes in selecting sample outliers in RNA-seq, SCISSOR, is a series of procedures for transforming and normalizing base level RNA sequencing coverage data in a transcript independent manner, followed by a statistical framework for its analysis. The resulting high dimensional object is amenable to unbiased identification of structural alterations across RNA-seq cohorts with nearly no assumption on the mutational mechanisms underlying abnormalities. This enables SCISSOR to independently recapture known variants such as splice site mutations in tumor suppressor genes as well as novel variants that are previously unrecognized or difficult to identify by any existing methods including recurrent alternate transcription start sites and recurrent complex deletions in 3' UTRs.

In a cohort of 522 TCGA head and neck squamous cell carcinoma, SCISSOR identifies known as well as novel aberrations including abnormal splicing, intra-/intergenic deletions, small indels, and alternative transcription start/termination. In addition, its genome-wide analysis uncovers a novel type of variant gene transcription near intragenic CpG islands. Finally, we find that our approach through shapes can be also useful for picking up rare transcripts from individual samples such as leukocyte transcripts and stromal transcripts. We believe that this new approach holds promise for identifying otherwise obscured genetic aberrations.

Taken together, these results suggest that SCISSOR has great potential for broad applications including discovery of novel driver genes and mechanisms of genetic abnormalities, detection of non-coding RNA, and studies of single cell RNA-seq.
AUTHORS/INSTITUTIONS: H. Choi, D.N. Hayes, UTHSC, Memphis, Tennessee, UNITED STATES|H. Choi, D.N. Hayes, Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, UNITED STATES|J.S. Marron, Statistics and Operations Research, UNC, Chapel Hill, North Carolina, UNITED STATES|