

Digging in the RNA-seq garbage: evaluating the characteristics of unmapped RNA-seq reads in normal tissues

M. Carrara[✉], R. Calogero

Molecular Biotechnology Center, University of Torino, Torino, Italy

Motivations

RNA-seq has the potential to discover genes created by complex chromosomal rearrangements. 'Fusion' genes formed by the breakage and re-joining of two different chromosomes have repeatedly been implicated in the development of cancer. However, due to the heterogeneous nature of solid tumors, pathological fusions can be confounded with read-through fusions across adjacent genes in the genome, or transcription-induced chimeras (TICs) present in normal tissues.

Methods

The pipe-line used to detect TICs is the following:

- Removing PE reads associated to transcripts (Bowtie, followed by SHRIMP on everything not mapped by Bowtie);
- removing reads mapping to spiked-in PhiX control;

- trimming linkers;
- detecting TICs with single end mapping (SHRIMP).

Results

We detected TICs events in 16 normal tissue samples. The information derived by our analysis indicates the presence of significant number of TICs in normal tissues. Some of these TICs can be erroneously associated to cancer development. In some cases TICs expression is corroborated by large number of reads and their expression is significantly spread over different tissues. Our data highlight that TICs could be erroneously associated to cancer aberrations, if RNA-seq analysis is only evaluated in tumor samples, not considering the parallel analysis of normal tissue samples associated to the tumor.