Reviewer's report

**Title:** Deep sequencing analysis of transcription-induced chimeras in human prostate adenocarcinoma and reference samples

**Version:** 3  **Date:** 25 November 2010

**Reviewer:** Gilbert S Omenn

**Reviewer's report:**

The responses to reviews seem quite adequate. The paper has been clarified and improved.

Editorial: In the Abstract and elsewhere the style "2 out of 5" is used; the word "out" is unnecessary.

Discretionary Revisions

I share here additional comments from my colleague, Dr. Christopher Maher, an expert on paired-end and short read next-generation sequencing and inference of chimeras:

They have addressed most your and Arul's original comments. I still think it is very descriptive and lacks the clinical components desirable for BMC Medical Genomics.

- The authors state that “Prostate and reference samples exhibit a wide range of TIC events, involving more genes than previously estimated using ESTs” in their abstract. One discussion point to consider is whether existing publications focused on reads passing filters that may have been biased against TICs.

- Are the authors restricting themselves by merely requiring adjacent same strand transcripts? Would this include the relevant category of transcripts that are separated by a transcript on the opposite orientation? Or genes separated by 1 gene in the same orientation? Arbitrary distance metrics would miss many of these cases.

- The authors should indicate the percentage of positive controls in both UHR (5/7) and HBR for long-range and TICs. Additionally, it would be beneficial to have an overall summary statistic of their false positive rates.

- With regards to tissue-specificity, the authors could generate a simple metric for tissue-specificity such as whether the index cases for the TICs are in a tissue that has a higher median expression than remaining cohorts. For example, a gene such as SLC45A3 should certainly show a markedly higher expression level in prostate than the remaining tissues sampled.

- In the conclusion the authors write “We believe that the vast majority of our TIC events represent true biological events, as supported by our experimental
verification of several events by qRT-PCR, and supporting evidence from EST sequences and previous surveys of TIC events”. A percentage would be more informative than ‘vast majority’.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.