Author's response to reviews

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

Authors:

Serban Nacu (serban.nacu@gmail.com)
Wenlin Yuan (wenliny@gene.com)
Zhengyan Kan (kanzhengyan@yahoo.com)
Deepali Bhatt (bhattd@gene.com)
Celina S Rivers (rivers@gene.com)
Jeremy Stinson (stinson@gene.com)
Brock A Peters (brock.peters@gmail.com)
Zora Modrusan (modrusan@gene.com)
Kenneth Jung (kjung@gene.com)
Somasekar Seshagiri (sekar@gene.com)
Thomas D Wu (twu@gene.com)

Version: 3 Date: 28 September 2010

Author's response to reviews:

Qihong Huang, M.D., Ph.D.
Associate Editor, BMC Medical Genomics

Dear Dr. Huang:

Thank you for the reviews on our manuscript on "Deep sequencing analysis of transcription-induced chimeras in human prostate adenocarcinoma and reference samples" for publication in BMC Medical Genomics.

We have revised our manuscript accordingly. In our revision, we have addressed the reviewers' concerns in a substantive and comprehensive manner. Specifically, we have added a detailed analysis of domains and tissue specificity for all TICs, as requested by reviewer 1. As a result, the revised manuscript now contains an additional figure, as well as four new subpanels to previous figures. We have also enhanced our abstract and addressed the concerns of reviewers 2 and 3. We have submitted our microarray and sequencing data to NCBI Gene Expression Omnibus, so that it is now publicly available. Our point-by-point response to the reviewer's comments is appended to the bottom of this letter.

We have also made the editorial changes that you requested, including structuring the abstract and adding sections for authors' contributions and competing interests. Note that we have separated our supplemental material into 7 separate additional files. They are essentially unchanged from the original submission.
Thank you for reviewing our changes. We look forward to hearing from you.

Sincerely,

Thomas Wu, M.D., Ph.D.
Senior Scientist
Genentech, Inc.

Reviewer 1, Arul Chinnaiyan

1. The reviewer suggests that we plot the total distribution of possible exon-exon junctions.

Response: We have added this in a new subpanel E of Figure 2, which shows the distribution of splice distances in our database, and confirms that the underlying distribution is different from that in Fig. 2D.

2. The reviewer asks if we observe any global tissue-specific expression of the 5' partners of TICs.

Response: We have performed a comprehensive analysis of tissue specificity of the 5' genes, and show the results in a new Figure 6, and we describe this heatmap in the text, stating that 5' genes have various levels of tissue specificity.

3. The reviewer asks if we can analyze the protein domain structure of our major reported events.

Response: We have computed the protein domain structure for all TIC events and their component genes, and now report this information in Supplemental Table S2, with a summary provided in a new subpanel F of Figure 2. We have also added sections in the Results and Methods sections discussing our domain analysis for these events, as well as for the TICIE events and the distant fusion events that we confirmed.

4. The reviewer suggests that it would be beneficial to add a subpanel to Figure 3, to show the gene expression of an average gene.

Response: We would like to cooperate with the reviewer, but we do not know how to select an "average" gene. In addition, the intent of this figure is to show the relationship between expression of the TIC and that of the upstream and downstream genes, and we have TIC expression only from RT-PCR experiments on a selected number of TIC fusions. We have added new subpanels G and H to Figure 3 to show that TIC splicing efficiency differs among different TIC events.

Reviewer 2, Rolf Skotheim

1. The reviewer asks us to clarify in the abstract that we did perform experimental sequencing work on samples and to describe the samples.
Response: We agree with the reviewer's comment, and have revised our abstract accordingly.

2. The reviewer asks us to augment the information on experimental procedures in the Methods and Materials section.

Response: We have augmented the experimental information regarding the CGH microarrays. The text we have used is identical to that in a recent Nature paper, which also involves an analysis of these prostate tumor CGH microarrays, but not the normal microarrays. We have also augmented the description of our qRT-PCR procedures.

3. The reviewer requests that the data be deposited into a public repository.

Response: The sequencing and microarray data are now available publicly at NCBI GEO, under Series accession number GSE24284, and we have added text to the Methods section to declare this.

4. The reviewer asks for clarification on (a) whether the fusion transcripts are trans-splicing or genomic in origin, and (b) whether TICs are cis- or trans-splicing events.

Response: Our experimental design was not designed to distinguish between trans-splicing events where genes on separate transcripts are fused together, and cis-splicing, where the genes are located on a single transcript. However, we believe that trans-splicing is a rare event. Finta and Zaphiropoulos, J. Biol Chem 277, 2002, 5882-90, believe the occurrence is 0.15%. Accordingly, we believe that most evidence suggests that TIC events are due to cis- mechanism, and we state this in the Discussion section. As for gene fusion events, we cannot distinguish between genomic events from trans-splicing events either. However, our CGH microarray evidence does support a genomic event for TMPRSS2-ERG in one sample, and we believe that most evidence suggests that gene fusion events have a genomic origin, rather than trans-splicing. To help clarify this distinction, we have made our wording more uniform throughout the manuscript, by removing occurrences of the phrase "TIC fusion" and using just "TIC".

5. The reviewer asks us to provide additional information in the abstract and to clarify what was done.

Response: We have greatly revised our abstract to provide additional information and clarify our work.

6. Page 14, line 1. The reviewer asks how the reading frame of an upstream gene can be affected by a TIC or TIC event.

Response: If the fusion breakpoint occurs after the transcription start site (TSS) of the 5' gene, then the reviewer is correct in that the TSS and reading frame of the 5' gene should be preserved. However, if the fusion breakpoint occurs before
the original TSS of the 5' gene, then the fusion gene must have a new TSS, either in the 5' or 3' gene. In these cases, we predict the TSS and coding region based on the longest ORF, and this may imply translation of the 5' gene in a different reading frame. Nevertheless, the reviewer's comment has caused us to recompute our results again more carefully, and we have revised Fig. 2C and our associated text accordingly, with a clearer explanation of our analysis.

7. Page 15, line 20. In accordance with the reviewer's comment, we have changed our description of the UHR sample to include just the 10 cell lines.

8. Page 27, lines 2-3. Increased expression of the downstream fusion partner should not be required.

Response: We agree with the reviewer that fusion genes can have activity by altering the protein, rather than increasing the downstream gene expression. However, we believe that it is still important to point out the lack of enhanced downstream gene expression in these cases, since the readers will want to know. We thank the reviewer for the citation to his work, and we cite his paper in our Discussion section.

9. page 21, 1.7. The word "million" is included unnecessarily.

Response: We have removed the two occurrences of the word "million".

10. Table 2. The authors should discuss the possibility of contamination of TMPRSS2-ERG in N1.

Response: We now mention this possibility, which is a likely interpretation.

11. Figure 3. The authors should include sample labels and a y-axis scale for RPKM values.

Response: We have revised the figure as suggested.

12. Figure 4A. The relationship between SLC45A3-ELK4 and ERG is more complex than discussed in the text.

Response: We agree that SLC45A3-ELK4 is also low in many ERG-negative tumors, but its high expression appears to be restricted to ERG-negative tumors. We have expanded our discussion of Fig. 4A and clarified our point, which is that low expression of ERG appears to be a prerequisite, but not a sufficient condition, for high levels of the TIC event.

13. Figure 4B,C and Figure 5A-D. The authors should add gene symbols to the figure.

Response: We have revised the figures as suggested.

Reviewer 3, Gilbert Omenn
1. Proteomics can be a useful adjunct to RNA-Seq studies.
Response: We thank the reviewer for reminding us of this, and we cite one of his papers in our discussion section.

2. The ConSig score for gene fusions may be useful.
Response: We thank the reviewer for the reference. However, if we understand correctly, it appears that the score may be more applicable to gene fusions than to the TIC events that are the focus of this paper, and we do not have the resources at this time to apply this score to our findings. Nevertheless, we are citing the paper as an example of the apparent rarity of some gene fusions, such as R3HDM2-NFE2.

3. The filters to remove false positives should receive more emphasis and detailed results.
Response: Our Results section does contain four reasonably detailed paragraphs that describe our step for removing false positives and the numbers of reads that were filtered by each step. We further compared our findings with AceView, EST evidence, and prior studies of TICs and the MAQC samples, and found a substantial amount of support for our findings.

4. Clinical correlates would be desirable if more patients were analyzed.
Response: We agree with the reviewer that clinical correlations would require larger sets of patients analyzed to be useful.