Author's response to reviews

Title: An online tool for evaluating diagnostic and prognostic gene expression biomarkers in bladder cancer

Authors:

Garrett M Dancik (dancikg@easternct.edu)

Version: 3 Date: 19 May 2015

Author's response to reviews: see over
Dear Dr. Bobby Shayegan

I am pleased to submit a REVISED manuscript, titled “An online tool for evaluating diagnostic and prognostic gene expression biomarkers in bladder cancer”, for consideration for publication. As requested, changes to the manuscript are marked up using the yellow highlight feature of Microsoft Word. We thank you and the reviewers for useful comments regarding the manuscript. Three major changes were made in response to these comments. First, a new table has been added more thoroughly describing the patient cohorts analyzed, patient exclusion, and processing. Second, patients profiled in multiple cohorts are excluded from multiple analyses so as not to bias the results. Third, the user now has the option of removing treated patients from the survival analyses. In addition, the patient selection criteria has been clarified and the format of the results modified to be more readable. The reviewer comments are provided below, along with the responses which I hope will make this manuscript publishable in *BMC Urology*.

**Reviewer #1:**

**Comment:** The author comments that there will be regular updates – yet a large MIBC dataset from Oct 2014 (Mitra AP et al JNCI 2014; PMID 25344601) is not included at this early stage. Would it be possible to add this now?

**Author Reply:** This dataset, with expression profiles available from Gene Expression Omnibus at Accession #GSE57933, was not included because the clinical outcome information is not publicly available. I reached out to one of the coauthors on the manuscript, however the authors are not in favour of making this data available at the present time.

**Comment:** line 157/158: only 2 were NMI, and these were removed?

**Author Reply:** This is correct. There was a typo in the previous version that had made this statement confusing. This line has now been corrected (now at line 164-165).

**Comment:** line 164/165: add references for two different grading systems (WHO 1973 and WHO 2004); best also to cite latest TNM staging for staging breakdown

**Author Reply:** Done (now at line 175-176).

**Comment:** line 183-188: why not progression free survival as an endpoint for NMIBC?

**Author Reply:** Unfortunately, none of the datasets included have both progression status and time to progression available.

**Comment:** Table 1 – not sure what the numbers mean under OS/DSS/RFS – number of patients with data available for that endpoint?

**Author Reply:** This is correct, The numbers under OS/DSS/RFS correspond to the number of patients with data available for each endpoint. To clarify this, I now explicitly state this in the Table caption.
**Comment:** Table 1 – the authors say that they have eliminated 2 patients from MDA-1 because only 2 were not MI (see above) – but they do not include MI vs NMI in Table 1.

**Author Reply:** Table 1 includes only the patients included in the database that are analyzed, and this is now explicitly stated in the Table caption. The notation of '-' is also now explicitly defined as indicating insufficient sample size for analysis. In the case of MDA-1, because the 2 patients with NMI tumors were removed, comparison of MI and NMI tumors is not done, which is why the corresponding cell in the table contains a '-'.

**Comment:** Figure 2B – the pie charts are not intuitive for the section on survival. Perhaps because poor vs. good prognosis is not in the title above the pie chart like the comparators are for panel A.

**Author Reply:** I have changed the comparators to include "poor vs. good prognosis".

**Comment:** Figure 2C – it would make more sense if the light blue was between the pink and the dark blue (dark blue on bottom and dark red on top – these are the extremes).

**Author Reply:** Agreed. This has now been changed.

**Comment:** In the supplementary Excel file it would be easier to have an explanation of the numbers in the respective table and not all together on the first worksheet. Could also make it clearer - for example, for grade make the first row the # of LG and the second row the # of HG, then third row is the HR and forth row is p-value.

**Author Reply:** I modified the Excel file so that an explanation is included with each table and so that rows are used to present the results for each cohort.

**Comment:** The heterogeneity in therapy between and within cohorts (eg. BCG and adjuvant chemo in CNUH cohort) detracts from generalizability.

**Author Reply:** Agreed. This limitation is now mentioned in the Conclusion. In addition, the user now has the option of excluding patients treated with BCG intravesical therapy and adjuvant chemotherapy.

**Reviewer #2:**

**Comment:** Patient selection methodology is somewhat confusing. Does the author state that all patients who received neoadjuvant or adjuvant chemotherapy were excluded?

**Author Reply:** Patients receiving neoadjuvant chemotherapy are excluded from the prognostic biomarker evaluation (i.e., survival analysis). Initially, I did not exclude patients receiving adjuvant chemotherapy (or intravesicaul therapy) because treatment was either not associated with better outcomes (Supporting Fig. 1) and/or confounded with TNM staging. Specifically in the case of DFCI, the majority of chemotherapy-treated patients (27/32) had nodal involvement
or distant metastases. However, I now give the user the option of excluding patients receiving adjuvant chemotherapy (or intravesical therapy) from the prognostic biomarker evaluation analysis. This is now explicitly stated in the Prognostic biomarker evaluation and survival analysis section.

Comment: Table 1 reports that significant amount of survival data is missing from the various included datasets. This is an important limitation and should be discussed in the manuscript. Only one dataset has reported RFS.

Author Reply: I agree that the lack of survival data is a limitation, and this is now stated in the Conclusion. In addition, I have added Supporting Table 2 which shows that stage (NMI vs. MI) and grade (LG vs. HG) are consistently associated with outcome when the best available endpoint is used.

Comment: Where there any other important data point such as tumor stage, or grade also missing for any of the patients? A table describing the patient cohort would be helpful.

Author Reply: In some cases, clinical data such as tumor stage are not available for all patients. This can also be determined by looking at Table 1, which lists the total number of patients included in BC-BET for each cohort, as well as the number of patients with stage and grade analyzed. Additional information about each cohort, including those patients not meeting the selection criteria, are now listed in Supporting Table 1 to provide more details about each cohort.

Comment: Is there any possibility of overlap of patients between any of these datasets? i.e. same patients being presented in different report from the same institution.

Author Reply: There is overlap of patients between and within some of the datasets, which the initial version using the available processed data did not account for. In order to prevent this overlapping of patients from biasing the results, duplicate patients within a cohort are combined and duplicate patients across cohorts are excluded so that no patient is included more than once in the same analysis. Details are provided in Patient cohorts and gene expression datasets under Construction and Content and in Supporting Table 1.

Comment: The BC-BET website states that 14 patient datasets are utilized whereas the manuscript states 13.

Author Reply: BC-BET contains 13 patient datasets and this has been corrected on the website.

Thank you for your consideration,

Garrett Dancik, PhD
Assistant Professor
Department of Mathematics and Computer Science
Eastern Connecticut State University