Reviewer's report

Title: Immunological network signatures of cancer progression and survival

Version: 1 Date: 10 December 2010

Reviewer: Peter Lee

Reviewer's report:

The authors have developed an information theory based method to assign immune and cancer relevance scores to all human genes. In the manuscript, they describe the method and benchmark it to known immunological genes. In addition, immune gene and cancer relevance scores are depicted in the context of the interactome. The authors utilize their method to compare immune signatures amongst normal tissues and in immune network signatures and clinical outcomes amongst skin-related tissues from melanoma patients. The method developed by the authors enables the identification of immune components in mixed tissue patient gene expression profiles and provides some understanding of the immune related mechanisms in the cancer setting.

This paper introduces a new angle on microarray gene expression interpretation, enabling detection of the immunological components in each sample and identifies the associated protein interactions based on immune relevance, thus revealing the genes, mechanisms and immune cell types that take part in melanoma progression. However, the wording, figure presentation and organization in the results and methods sections are not clear and require editing and re-wording.

Major points

(1) In the methods section (page 19, under “defining the dictionary of terms for immune and neoplasm context”) it is stated that “strict scrutiny” was applied to accept terms into the final list. Further explanations are needed regarding this process and the basis for accepting/rejecting terms.

(2) In the methods section (page 21, under “an immunological and cancer relevance score for all human genes using information theory and text mining”), Shannon’s entropy formula is given. However, it is not clear how the entropy value that is calculated using this formula (Hc) is used, and what is its relation to the KL formula. Moreover, in the results part the authors state that the KL formula is used as the immune relevance score for the genes.

(3) In the methods section (page 18, under “constructing a validated human interactome & network analysis), the authors do not indicate how the three interaction databases were integrated.

(4) In supplementary file #5 the authors do not explain about the composit immunological relevance and connectivity y score (last column), or how it was calculated.
(5) In the results section, the section “Immunological comparisons of normal tissues and robustness of tissue specific immune interactions” is very hard to understand and needs wording revision to clarify. For example –the authors refer to “differentially expressed genes”, but it is not clear from the text what was under comparison to yield these differentially expressed genes, what test was used. In addition, the reference referring to the expression data source is missing. The text which contains these explanations exists in the methods section, under “microarray gene expression analysis and a composite expression and immunological relevance score”, but is only referred to in the text in later sections. In addition, it is not understood where to look for the CD4 and CD8 T cells that the authors refer to, in Figure 4.

(6) K-means test was applied with the five centrality measures, and the eccentricity measure was the only one that clustered the tissues in a biologically meaningful manner. What do the authors mean by “biologically meaningful” in this context? In addition, the authors write that the immune cells had the lowest average values. Does this refer to centrality values? This should be added to the text for clarity.

(7) The authors often refer to supplementary file 6 (Figure S1) in the main text, and also draw conclusion for this section partly based on it. Given this, the authors should consider putting this figure in the main text.

Minor points

(1) General comment regarding the methods section: it is often not clear in the text which method to refer to. Thus, instead of referring the reader to the methods section, it might be more clear to specify the title in the methods section, and also order the methods in the methods section according to the order they are referred to in the text.

(2) The methods section (page 19, under “defining the dictionary of terms for immune and neoplasm context”) states that 1996 immune and 562 neoplasm terms were identified, but in the results section its written that 1921 immune terms (the neoplasm terms are not mentioned). According to supp1 there are 1921 and 562 immune and neoplasm terms, respectively.

(3) Page 6 – a reference is missing for “using established text mining procedures”.

(4) The threshold used to filter out genes that do not have a sufficient immune relevance score is not clear, and should be mentioned in the methods section. In the results section it is mentioned at one point that the threshold is >1. However, it is not clear whether this threshold is used throughout the study, and why was this threshold used.

(5) Figure 1: “A” and “B” captions are missing for figures 1A and 1B.

(6) Figure 2 legend: spelling mistake – authors wrote “al” instead of “all”.

(7) Figure 3: there is no caption or explanation for the heatmap scale (colors).

(8) Figure 4: “A” and “B” captions are missing for figures 4A and 4B.
Figure 4 legend: spelling mistake – line 3: authors wrote “that” instead of “than”; line 6: authors wrote “there” instead of “their”. A general comment regarding figure 4 – the text is not clear regarding what figure 4 is trying to show. In addition, there are no captions for the matrix’s x and y axis or the heatmap scale (colors).

Figure 4B: explanations regarding this figure, including explanations in the figure legend, are not clear and need rewording. In the figure legend, the authors refer to panel A, but it is not clear what it is. The figure itself lacks captions for x-y axis.

Figure 5 legend: abbreviations for tissue names are missing, as the main text only refers to three of these (normal skin, primary skin tumor and metastatic melanoma). In addition, it is not clear what tissues are considered “normal” and “cancer”. For clarity, the normal and cancer tissues should be grouped together in the table. In the figure itself, a scale that explains the colors is missing.

Table 1: the authors should add a column stating the biological conclusion drawn from the up/down regulated highest graded genes for each comparison (such as the examples given in the text). Also, it is not stated what threshold was used in picking the highest graded immune genes.

The authors refer to the differences between in situ melanoma and squamous cell carcinoma as being the highest in the text (page 11), give examples to some genes, but the network for these genes is missing.

In the methods section (page 22, under “microarray gene expression analysis and a composite expression and immunological relevance score”), some details are missing: (a) the authors refer to a “p-val<0.01”, but it is not stated what test was used and what was compared/tested? (2) the authors used only probes with p.val<0.01 in more than 50 of the patients – how many patients were there overall in this study?

Spelling mistake on page 13, line 6: instead of “the same analysis applied” write “the same analysis was applied”.

Level of interest: An article of importance in its field

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.