Reviewer’s report

Title: 'Health space' visualization and identification of personalized molecular phenotype and treatment responses according to the underlying relevant biological processes

Version: 1 Date: 27 May 2011

Reviewer: Martin Kussmann

Reviewer's report:

General comment:
Display of multivariate data based on aggregation of variables according to prior knowledge is not as such new. The use of Ingenuity Pathways to do so may be new, but does not represent a revolutionary breakthrough in data visualization. However, it is a valid and useful method.

Minor essential revisions:
1) Improve explanation of the Demonstration dataset in the "Background" section. It is not stated that the study was set up as a crossover trial, which leads to confusion in the results section.

2) Section "Results" - Subtitle "Dietary compounds have a large effect on the response": The claim that "subject 1 [reacts primarily] on the oxidative and inflammatory axis" does not appear supported by the figure. Is this a typo, where a different subject was meant? Furthermore, while for subject 33 the claim clearly mentions a reaction along one axis and the claim for subject 25 explicitly mentions a combined response along two axes, the wording of the claim for subject 1 seems to imply one axis displaying both oxidative and inflammatory pathways (which of course does not exist in the figure). Be careful with plural and singular axis/axes.

3) Section "Results" - Subtitle "Constructing the model": The last paragraph states that the vector directions were also visualized and that they were similar across patients. This visualization is missing from the figure. The vectors really must be shown in figure 1.

4) Section "Results" - Subtitle "Dietary compounds have a large effect on the response": A claim is made that using a subset / expanded version of the original dataset proves that the method works on different datasets. Are these datasets sufficiently different to justify such a claim?

5) Table 1: The reference to this table does not state that it is in the supplementary data. Furthermore the data shown is a set of three lists, not a table. It would be useful to have further data in a real table. Specifically p-values and parameter class (chemistry, protein or metabolite)

6) Language: The English is generally good. It is consistently comprehensible and clear, but several systematic mistakes need to be corrected and a general
language correction would be helpful. Specifically: prior TO and after; correct usage of built vs. build, than vs. then, extent vs. extend; fragmentary or incomplete sentences.

Discretionary revisions:

a) The use of the treated group to define the healthy state is unconventional. While the discussion does justify this choice, and there is no problem with it from the point of view of validating the health space concept, it would be helpful to place a short sentence stating that the justification for this choice will follow in the discussion at the point where it is first mentioned.

b) Abbreviations: Abbreviations such as PBMC are not defined in the text. For statistical methods such as GSEA this is less of a problem.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

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