Response letter to: feedback from Su, Valerie and Solène

Authors: F.P. Meijvis, A.E. Maranus, W.B. Eiling

**Response letter**

The feedback we received on our report was less informative and critical than we had anticipated. In general, most parts of the feedback simply described the structure and content of our paper (e.g., “*The comparison strategy is outlined at the end of the method, specifying that models will be evaluated*…”). In our view, such feedback lacks depth and insight. The most critical comments pertained to the need for more biological interpretation, such as identifying gene importance for each principal component (PC). However, we believe this goes beyond the scope of our report, as our primary focus is the comparison of logistic regression and (sparse) PCA for predicting tumor versus healthy samples using gene expression data. To address this point, we have added a paragraph in the conclusion section to highlight the relevance of our analysis and the potential for biological interpretations. Additionally, we revised the introduction to clarify that our research question is centered solely on comparing the predictive performance of (sparse) PCA and stepwise logistic regression in this context, rather than exploring biological insights into tumor gene expression. This adjustment ensures that the objective of our study is clearly communicated. In conclusion, while we were unable to incorporate a substantial portion of the feedback into our final report due to its limited relevance, we made the following two key changes:

1. Improving readability of the introduction: We added a concluding sentence to the introduction to enhance the flow and clarity of this section.
2. Clarifying the research focus: We adjusted the introduction to explicitly emphasize that our study examines the comparison of logistic regression and (sparse) PCA, rather than identifying genes involved in prostate tumor development. Furthermore, we included a paragraph in the conclusion to underline the broader relevance of this analysis for gene expression data and its potential for biological applications.