Response to comments by Reviewer #1

Thank you very much for your detailed reading of the paper and your insightful questions. You raise a number of good points and we have modified the manuscript in several places based upon your feedback. In our opinion, these revisions have nicely improved upon the original submission.

1. *This marginal prospective: the statistical definition and considerations are well presented. But the question is: practically, how useful is this? That is, in a typical real data analysis, how many covariates are expected to be fully independent of the response?*

   You raise a good point – certainly, the real world is messy and full of indirect connections, making complete independence unlikely to be literally true. However, we would argue that this is often true in hypothesis testing. Is a null hypothesis of precisely zero effect realistic? Usually not, but it is often informative to test the hypothesis anyway – if it cannot be rejected, any claim about a positive or negative effect should be viewed skeptically.

   This is very much how we see the value of the proposed mFDR approach: yes, its framework is surely an oversimplification, but it provides a straightforward and informative answer about how much trust you should place in the selected variables. For example, in the Shedden study, the mFDR suggests that up to 80% of the genes selected by the lasso model with cross validation could be false discoveries. It's possible that some of these genes have indirect relationships with survival, but the point is that for this model, you would select almost as many genes just by random chance even if none of them were related to survival. However, we *can* be confident about some of the genes selected: those which are present in the model where the mFDR is controlled at 10%.

   We now include a paragraph in the discussion ("A natural question is whether marginal independence is a realistic assumption. . . ") raising this question and discussing it much as we have in this response.

2. *The assumption of zero correlation may be too strong. How about weak correlations? Which may go to zero at a certain rate?*

   We're afraid we don't quite know what you mean here. The current assumption in the manuscript is that

   $$\frac{1}{n}\mathbf{x}_j \mathbf{W} \mathbf{X} \to \mathbf{0};$$

   In other words, what you seem to be describing is already the assumption in the paper: we don't assume that the correlations are exactly zero, but rather that they are going to zero asymptotically.

3. *The authors consider the logistic and Cox regressions. I do not see a very strong reason for sticking to the two models, that is, it seems that a large number of models can be accommodated. Is this true? If so, it definitely should be pointed out.*

   This is a good point – we concentrated on these models because of how widely used they are, but our intention was indeed to demonstrate that these ideas apply to a wide range of models. We have restructured Section 2 of the manuscript to make this point. Section 2.4 is now titled "Likelihood models", and begins with the following introduction:

   > The results derived in this manuscript are valid for any likelihood-based model for which classical asymptotic results hold, and are thus applicable to a wide variety of models. In this section, we provide additional detail for two such models, logistic regression and Cox regression, that are of particular interest due to their widespread use.

   The logistic regression and Cox regression portions of the paper are now subsections 2.4.1 and 2.4.2, respectively.

4. *The condition on dimensionality is not clear. Does the approach demand a finite number of covariates? If so, this is a limitation; and should be clarified. And discussions on possibly accommodating high dimensional data should be provided.*

   Our approach with this paper was to provide a proof in the fixed-p case and to investigate the high-dimensional ($p > n$) performance of the method via simulation. For example, in the simulation study of Section 3.3, $n = 400$ and $p = 1,000$. In practice and in simulations, the method seems to work well for high-dimensional data – better, in fact, than several competing approaches.

   However, your point is well-taken – it would certainly be of interest to attempt extending our proof of Theorem 1 to allow p to increase as a function of n as $n \to \infty$ and determine how this affects convergence. We now point out this limitation in the paper and mention that extending the proof in this manner is an important subject for future research.