We thank the reviewers for their comments.

Review #1: Regarding the significance and impact of the work, isotonic regression has been used in a multitude of applications, a few of which are given in the introduction. One of the most natural application areas for isotonic regression is biology. Biologists have established that "genetic effects on phenotypes such as height, fitness or disease are monotone"' ([1]). See [1] for references to the biology literature, and [2] for a discussion on monotonic genetic effects on disease.

In the case of disease, the presence or absence of a disease follows a monotone relationship with respect to gene expression. Classifying between lung and skin cancer amounts to applying this principle to a subpopulation of individuals who have lung or skin cancer. We will certainly include this reasoning in the revision of our paper. The motivation for a *sparse* model is that certain genes should be more responsible for disease than others. Sparsity can be viewed as a kind of regularization; to prevent overfitting, we allow the regression to explain the results using only a small number of genes. By identifying the most relevant genes, sparse isotonic regression helps elucidate the mechanism of disease. We have discussed our work with a biostatistician who works on cancer detection, and are working on using applying our algorithms to histology data from his lab. We hope this will improve the detection accuracy of his method.

Review #2: We agree with your idea to move the presentation of the algorithms to the main text while moving the results in Section 4 to the Appendix. Regarding tractability, the sparse quadratic minimization problem solved by the IPIR algorithm is NP-hard; we will include a reference. Both LPSR and S-LPSR are linear programs, which can be solved in polynomial time. The second step of TSIR in the Noisy Output Model is a linearly constrained quadratic program that can be solved in polynomial time. TSIR in the Noisy Input Model can also be solved in polynomial time. A note about Lemma 1: we should have stated that for the Noisy Input Model, there is a polynomial-time procedure to obtain an optimal solution to Problem (21)-(23), by forming the associated linear program and then finding an integer optimal solution. The procedure to find an integer optimal solution is part of the proof of Lemma 1. We agree that it is important to discuss tractability of algorithms, so all of this will be clarified in the revision.

Regarding line 343 in the Appendix: Constraint (9) requires each F_i to be either 0 or 1. Therefore, the substitution that we gave for the objective function gives an equivalent integer linear program.

Algorithm 1 is indeed implemented with integer programming in Gurobi, as written. The algorithm can be quite slow, which motivates the need for the two-stage approach. Algorithm 2/3+4 does not solve the same thing as Algorithm 1, but rather is a heuristic. The idea is that a problem becomes more tractable when decisions are made in two steps instead of simultaneously. To clarify, the objective in Eq. 10 is tailored to the goal of support recovery alone, and should be viewed separately from Eq. 5.

Regarding experiments, we have now implemented k-NN with dimension reduction at your suggestion. Surprisingly, the performance is worse than k-NN, achieving close to the baseline performance of about 68%. Even without comparison to other approaches, the excellent performance of our algorithm shows that it is a promising approach.

Review #3: Regarding the optimality of the convergence rate, it is true that the error bounds are probably quite loose, due to large constants appearing in the bounds. Our main goal was to show statistical consistency of our algorithms, and we did not optimize the constants. Tightening the bounds would be an interesting direction for future research. We note that error bounds are given for a noisy input setting, which is not typically seen in the literature but is often encountered in practice, such as the cancer application we studied.

Regarding the synthetic experiments, we have now run 50 trials. For the final version of the paper, many more trials will be conducted. In addition to the original metric measuring the frequency of the correct support recovery by our computations, per the reviewer's suggestion we will also report our results on function estimation accuracy. We found those to be extremely encouraging. For example, with 250 samples and d = 10, IPIR had a function estimation accuracy of 92.2%, LPSR had an accuracy of 89.2%, and S-LPSR had an accuracy of 90.1%. The accuracy was measured on 500 data samples.

To clarify the confusion about n and N, Theorem 2 requires fresh samples to be used at each iteration of S-LPSR. Given N samples, we divide into s batches of size n, corresponding to the s iteration steps. We apologize for the confusing presentation, and will clarify this in the revision.

References

[1] Qiyang Han, Tengyao Wang, Sabyasachi Chatterjee, and Richard J. Samworth. Isotonic regression in general dimensions. *arXiv 1708.0946v1*, 2017.

[2] Ramamurthy Mani, Robert P. St. Onge, John L. Hartman IV, Guri Giaever, and Frederick P. Roth. Defining genetic interaction. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*,