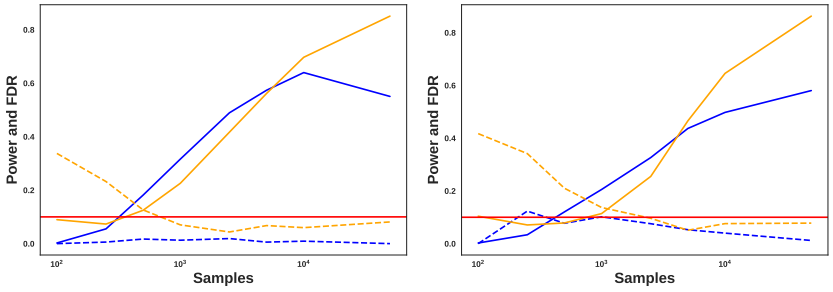
Black Box FDR - Supplement

1 Additional Stage 2 details and results

For BB-FDR, we fit an XGBoost Chen & Guestrin (2016) model for each of the three folds and each of the 50 covariates. We used the XGBoost holdout predictions and the corresponding BB-FDR fold-specific model to calculate null posterior entropies. Each covariate p-value was estimated based on 200 IID Monte Carlo draws and Benjamini-Hochberg correction was applied to each set of 50 p-values at a 10% FDR threshold.

Figure 1 shows the analogous results for the Stage 2 benchmarks on the well-separated alternative distribution.



(a) Linear (WS) (b) Nonlinear (WS)

Figure 1: Variable selection results at a 10% FDR threshold. Results are similar to that of the poorly-separated alternative in Section 4 of the main text.

2 Experimental details for cancer drug screening case study

We select two drugs in the GDSC dataset with known biological targets: Lapatinib, an EGFR-HERR2 inhibitor; and Nutlin-3, an MDM2 inhibitor. The genomic targets of both drugs are well understood, making them useful drugs on which to validate BB-FDR. Each drug was tested on a large number of cancer cell lines (294 for Lapatinib and 528 for Nutlin-3) to determine treatment efficacy. The results were measured by immunofluorescence; positive (untreated) and

negative (unpopulated) controls were decorrelated and temporal batch effects were removed. Positive controls are populations of cells which were never treated and thus their measurements serve as a null distribution. We take the maximum concentration dosage for both drugs and calculate z-scores for the treatment response, yielding a series of test statistics.

Each cell line has also been profiled via WES and we followed Iorio et al. (2016) in filtering down to a set of relevant, protein-changing gene mutations. We throw out any cell lines that were not sequenced and any genes whose mutation state was constant across all cell lines. For each cell line, we then get a binary feature vector representing the genomic profile of the experiment (66 features for Lapatinib and 71 for Nutlin-3). These features can then serve as an important source of prior information for the sensitivity or resistance of each cell line to the given drug.

A standard approach is to build a linear model of the genomic features to predict the outcome measurement. However, cellular processes are driven by complex, non-linear interactions known as gene regulatory networks Karlebach & Shamir (2008). A linear model of the interaction of genes is unlikely to be well-specified and will likely be underpowered, as in the non-linear prior scenario from Section 4.3. BB-FDR is a natural fit given its flexibility to model highly non-linear covariate dependencies.

We kept the majority of the BB-FDR settings the same as in Section 4. However, to account for the smaller sample sizes here, we changed the batch size to 10 and created 10 folds instead of 3 to maximize training data available to each model.

References

Chen, Tianqi and Guestrin, Carlos. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794. ACM, 2016.

Iorio, Francesco, Knijnenburg, Theo A, Vis, Daniel J, Bignell, Graham R, Menden, Michael P, Schubert, Michael, Aben, Nanne, Gonçalves, Emanuel, Barthorpe, Syd, Lightfoot, Howard, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

Karlebach, Guy and Shamir, Ron. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770, 2008.