NPDR Supplementary Material

Trang T. Le[1], Bryan A. Dawkins[2] and Brett A. McKinney[2,3*]

[1]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104
[2]Department of Mathematics, University of Tulsa, Tulsa, OK 74104
[3]Tandy School of Computer Science, University of Tulsa, Tulsa, OK 74104

March 7, 2019

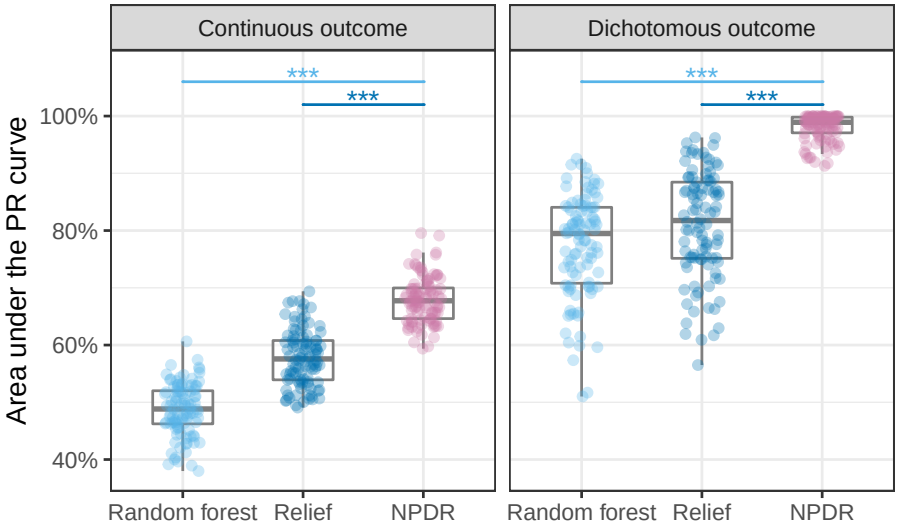

Figure S1: **Relief, NPDR and random forest comparison of area under the PRC (auPRC)** for 100 replicate simulations of continuous outcome data with main effect (left) and dichotomous outcome data with interaction effect (right). All simulations use $m = 200$ samples and $p = 1,000$ attributes with 100 functional. NPDR yields significantly higher auPRC in both simulation types.

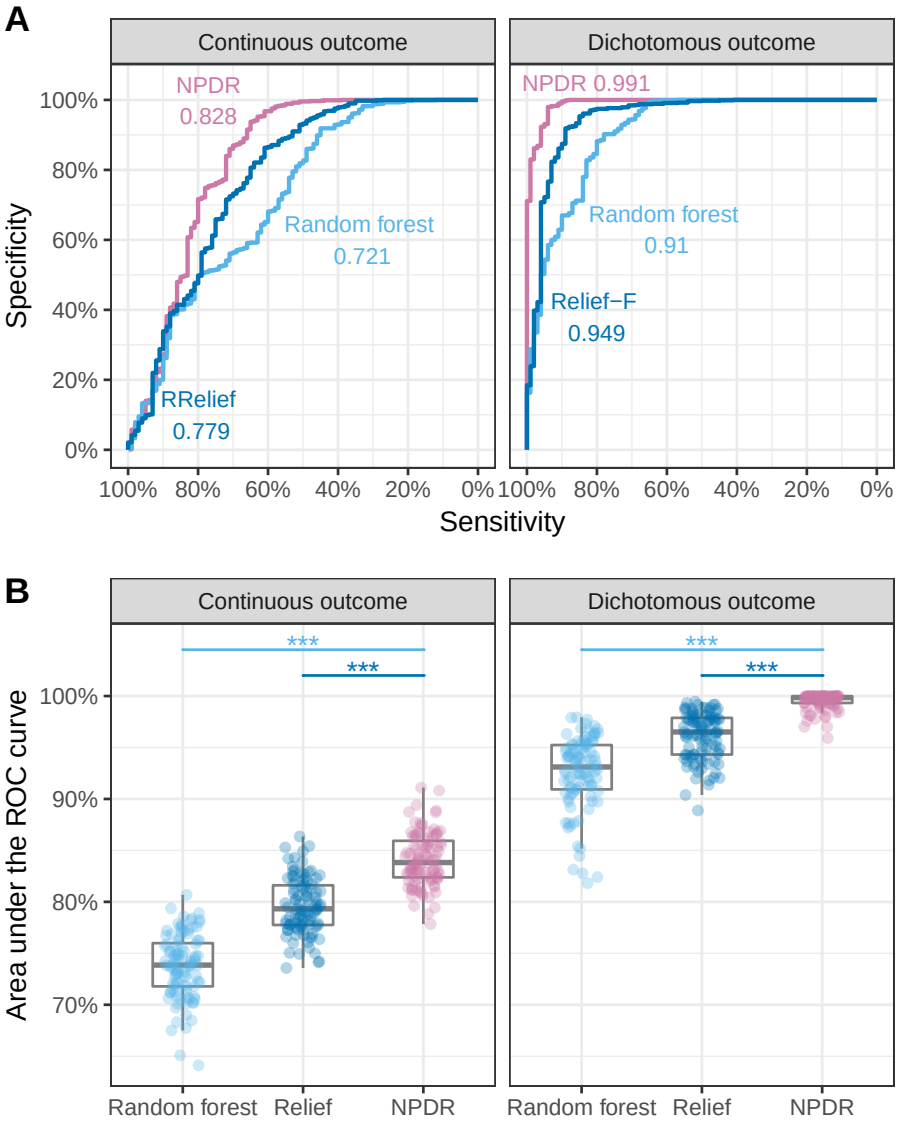

Figure S2: **Receiver Operating Characteristics (ROC) curves of Relief, NPDR and random forest** in one replicate simulation (A) and comparison of their area under the ROC (auROC) for 100 replicate simulations (B) of continuous outcome data with main effect (left) and dichotomous outcome data with interaction effect (right). All simulations use $m = 200$ samples and $p = 1,000$ attributes with 100 functional.

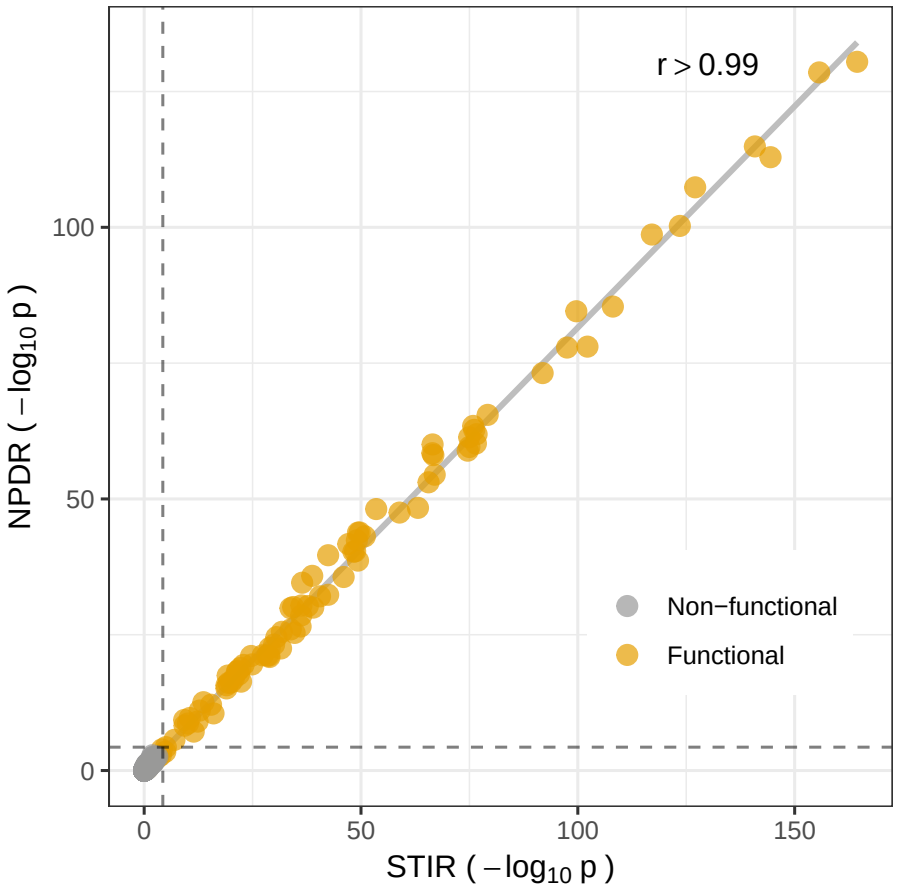

Figure S3: **Similarity between NPDR and STIR** for one simulation of $m = 200$ samples and $p = 1000$ attributes. In 100 replications, correlation of the P values produced by the two methods r ranges from 0.9827 to 0.9994. STIR is based on a t-test of projected distances and NPDR is based on a logistic regression of projected distances.

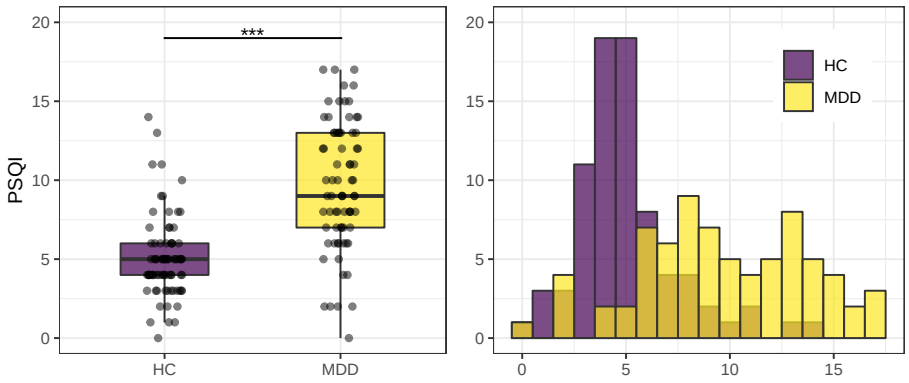

Figure S4: **The distribution of the Pittsburgh Sleep Quality Index (PSQI)** among individuals with and without MDD in Le et al.'s RNASeq dataset [1].

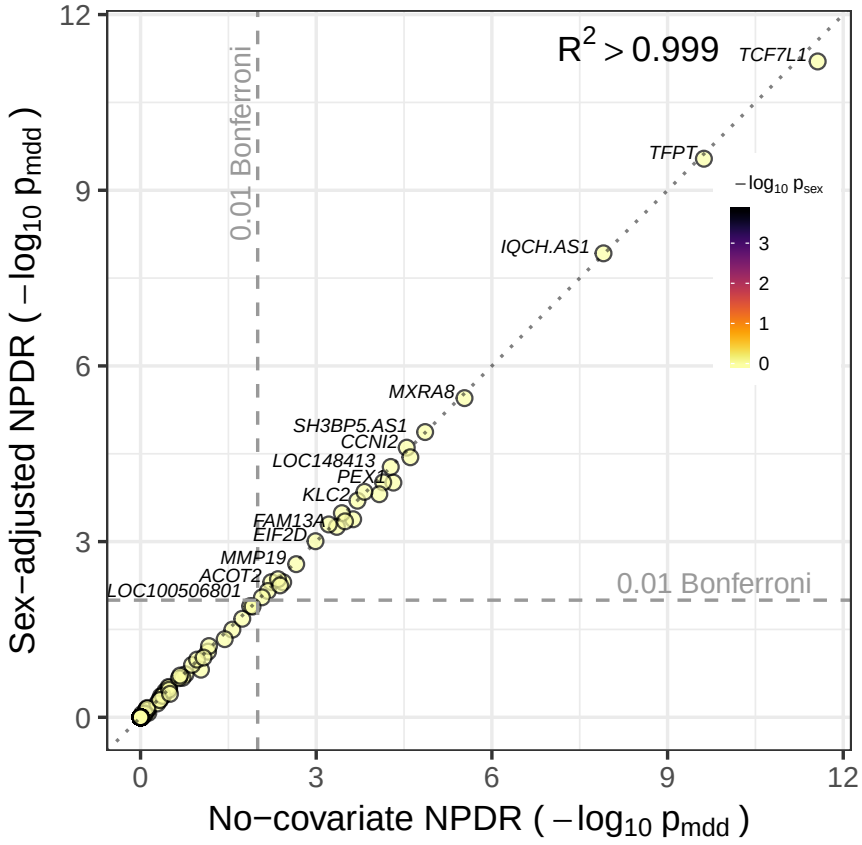

Figure S5: **NPDR with and without sex adjustment for analysis of MDD-associated genes** in Le et al.'s RNASeq dataset [1]. Adjustment for the sex covariate has a negligible effect on the resulting P values for each important gene because of the balanced study design. Both methods yield consistent results with STIR from previous study (Fig. 4 of Ref. [2]), not shown.

References

[1] Trang T. Le, Jonathan Savitz, Hideo Suzuki, Masaya Misaki, T. Kent Teague, Bill C. White, Julie H. Marino, Graham Wiley, Patrick M. Gaffney, Wayne C. Drevets, Brett A. McKinney, and Jerzy Bodurka. Identification and replication of RNA-Seq gene network modules associated with depression severity. *Translational Psychiatry*, 8(1):180, September 2018.

[2] Trang T Le, Ryan J Urbanowicz, Jason H Moore, and Brett A McKinney. Statistical inference relief (stir) feature selection. *Bioinformatics*, 2018.