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 10, 2019

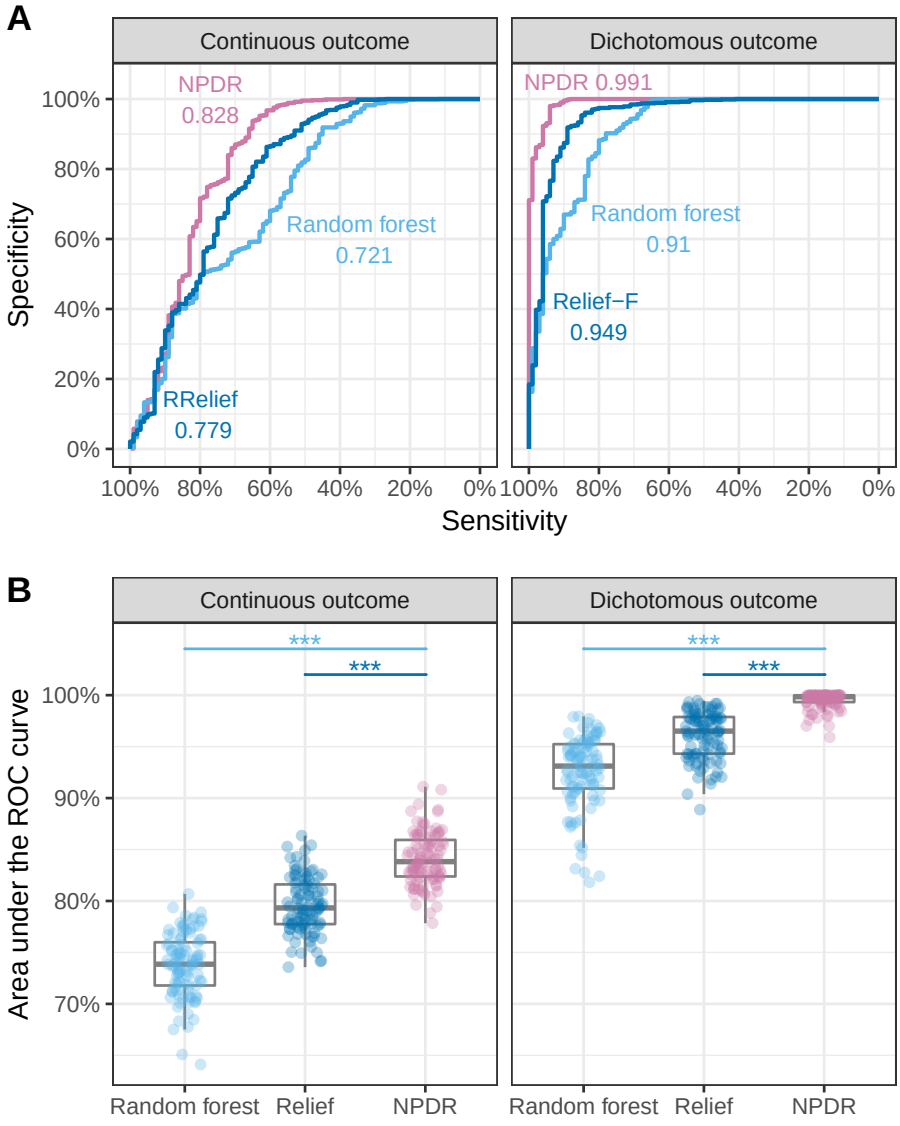

Figure S1: **Receiver Operating Characteristic (ROC) curves for Relief, NPDR and random forest feature selection.** For one replicate simulation (A), ROCs for continuous outcome data with main effects (left) and dichotomous outcome data with interaction effects (right). The auROC value is given for each method. For 100 replicate simulations of both simulation types (B), NPDR yields statistically significant higher auROC than Relief or random forest (blue *** indicate P<.0001). All simulations use $m = 200$ samples and $p = 1,000$ attributes with 100 functional.

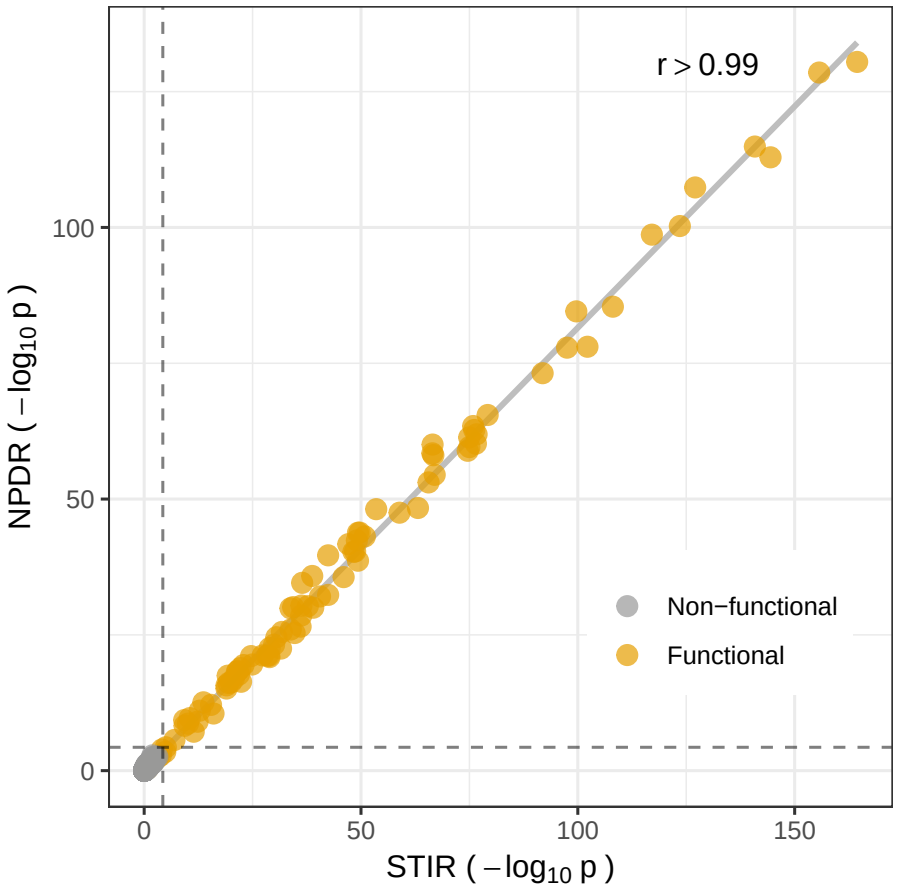

Figure S2: **Similarity between NPDR and STIR (dichotomous outcomes).** Comparison of -$\log_{10}$ P values for one interaction simulation of $m = 200$ (100 cases and 100 controls) and $p = 1000$ attributes with 100 functional. In 100 replicate simulations, correlation, r, between the two methods ranges 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. NPDR has the added benefit of handling continuous outcomes and covariate correction.

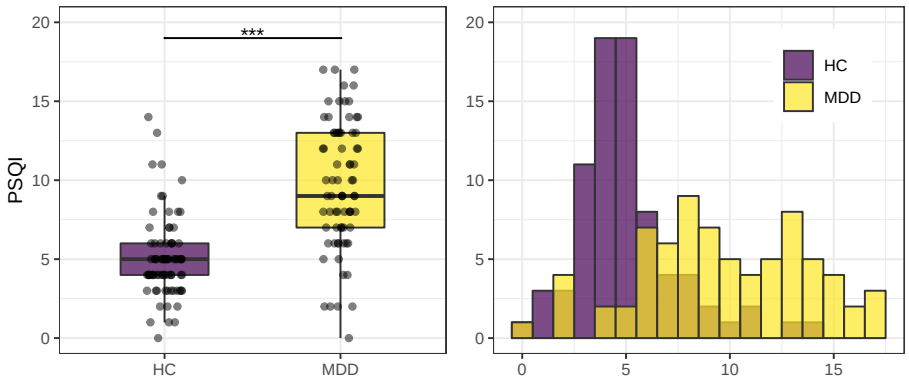

Figure S3: **The distribution of the Pittsburgh Sleep Quality Index (PSQI)** among individuals with and without MDD in Le et al.'s RNASeq dataset [1].

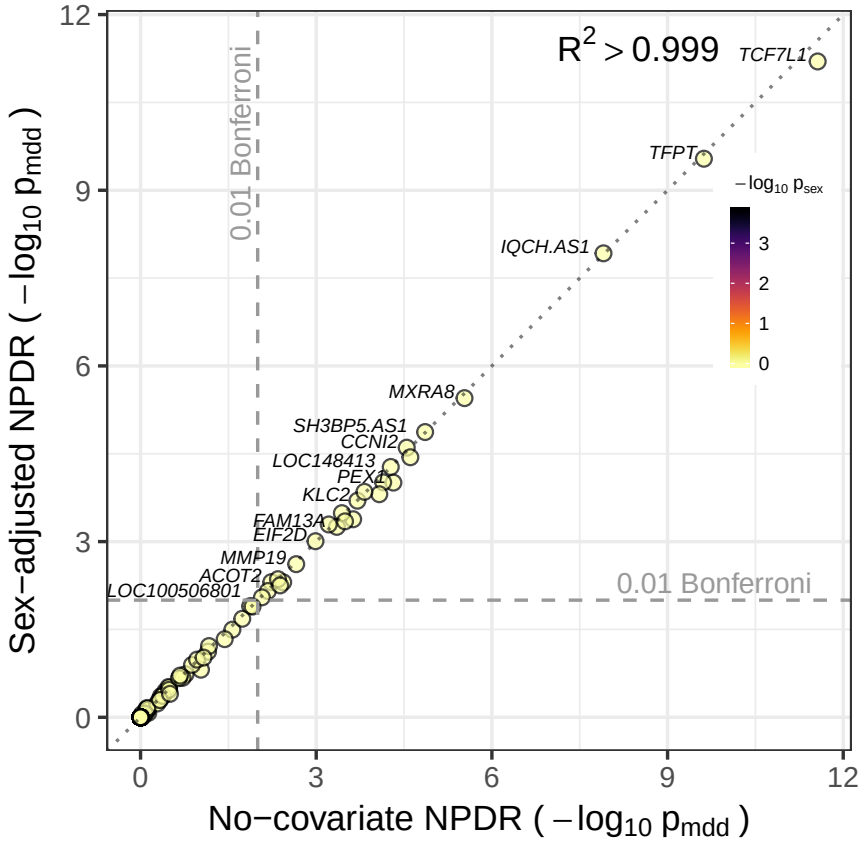

Figure S4: **NPDR with and without sex adjustment for analysis of MDD-associated genes** in Le et al.'s RNASeq dataset [1]. Adjustment for the sex covariate has a negligible effect on the resulting P values for each important gene because of the balanced study design. Both methods yield consistent results 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.