# Supplementary Material

## Investigating the associations Between Polygenic Risk Scores, Environmental Exposure, and Type II Diabetes in Parkinson's Disease

**Supplementary Table 1**: Logistic Regression Models of PD Against 4 Different PRS Calculated Using PD -Specific Variants (Nalls et al., 2019) in the MPBC Cohort
*PRS polygenic score; AGE = AAI; PC principal component; glm generalized linear model.*
**: Significant p-value*
*glm[1] (formula = PD ~ PRS + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10, family = binomial, data = data)*
*glm[2] (formula = AAD ~ PRS + SEX + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10, family = binomial, data = data)*
*PRS are normalized z-Scores.*

| | **Estimate** | **Standard Error** | ***P V*alue** |
|---|---|---|---|
| **90 Significant SNPs using PLINK[1]** | | | |
| Intercept | 1.07461 | 0.45323 | 0.01774* |
| PRS | 0.61482 | 0.05162 | $< 2 \times 10^{-16}$* |
| SEX | -0.03091 | 0.10240 | 0.76277 |
| AGE | -0.01711 | 0.00595 | 0.00404* |
| PC1 | -2.98263 | 2.20179 | 0.17553 |
| PC2 | 3.91841 | 2.17333 | 0.07140 |
| PC3 | -5.87932 | 2.14219 | 0.00606* |
| PC4 | 3.24137 | 2.11891 | 0.12608 |
| PC5 | -2.42583 | 2.12000 | 0.25252 |
| **1,805 SNPs using PLINK[1]** | | | |
| Intercept | 1.01609 | 0.45014 | 0.02399* |

| | | | |
|---|---|---|---|
| PRS | 0.56303 | 0.05076 | $< 2 \times 10^{-16}$* |
| SEX | -0.02698 | 0.10168 | 0.79077 |
| AGE | -0.01598 | 0.00590 | 0.00678* |
| PC1 | -2.45671 | 2.18908 | 0.26175 |
| PC2 | 3.56576 | 2.15169 | 0.09748 |
| PC3 | -5.12995 | 2.11861 | 0.01546* |
| PC4 | 4.77251 | 2.10194 | 0.02318* |
| PC5 | -1.52038 | 2.10612 | 0.47036 |
| **7,862,087 SNPs using PRSice with LD-clumping ($r^2 \leq 0.1$, 250kb window, and $p \leq 1$) [1]** | | | |
| Intercept | 1.19006 | 0.43658 | 0.00641* |
| PRS | 0.30294 | 0.04894 | $5.99 \times 10^{-10}$* |
| SEX | -0.01445 | 0.09904 | 0.88403 |
| AGE | -0.01707 | 0.00573 | 0.00290* |
| PC1 | -3.50557 | 2.13481 | 0.10057 |
| PC2 | 3.40332 | 2.10089 | 0.10524 |
| PC3 | -3.05910 | 2.05382 | 0.13636 |
| PC4 | 5.12566 | 2.04263 | 0.01210* |
| PC5 | -3.20231 | 2.04564 | 0.11748 |
| **7,862,087 SNPs using PRSice without LD-clumping[1]** | | | |
| Intercept | 1.23863 | 0.43970 | 0.00485* |
| PRS | 0.40830 | 0.04917 | $< 2 \times 10^{-16}$* |
| SEX | -0.02828 | 0.09994 | 0.77721 |

| | | | |
|---|---|---|---|
| AGE | -0.01799 | 0.00578 | 0.00184* |
| PC1 | -2.45784 | 2.14874 | 0.25269 |
| PC2 | 3.89408 | 2.11876 | 0.06608 |
| PC3 | -3.42548 | 2.07315 | 0.09847 |
| PC4 | 4.52895 | 2.06400 | 0.02822* |
| PC5 | -3.53600 | 2.06287 | 0.08651 |
| **PRS-AAD Model[2]** | | | |
| Intercept | 65.1632 | 1.0752 | < 2e-16* |
| zSCORE | -0.9976 | 0.3630 | 0.00613* |
| SEX | 0.1805 | 0.7423 | 0.80799 |
| PC1 | -38.9462 | 18.6077 | 0.03667* |
| PC2 | -11.7065 | 16.6587 | 0.48244 |
| PC3 | -7.0433 | 15.9747 | 0.65940 |
| PC4 | -12.4637 | 15.4129 | 0.41896 |
| PC5 | -11.8135 | 14.9366 | 0.42924 |

**Supplementary Table 2**: Confusion Matrices Showing Model Predictive Performance for PD cases and Controls for all 4 PRS Logistic Regressions.

| **Metric** | **90 Significant SNPs** | **1,805 SNPs** | **LD-Clumped SNPs** | **Non-LD-Clumped SNPs** |
|---|---|---|---|---|
| Sensitivity | 0.5737 | 0.5791 | 0.5942 | 0.5511 |
| Specificity | 0.6920 | 0.6492 | 0.5807 | 0.6578 |

| | | | | |
|---|---|---|---|---|
| Positive Predictive Value (PPV) | 0.6492 | 0.6212 | 0.5847 | 0.6154 |
| Negative Predictive Value (NPV) | 0.6203 | 0.6082 | 0.5902 | 0.5959 |
| Prevalence | 0.4894 | 0.4984 | 0.4984 | 0.4984 |
| Detection Rate | 0.2859 | 0.2886 | 0.2961 | 0.2747 |
| Detection Prevalence | 0.4405 | 0.4646 | 0.5064 | 0.4464 |
| *Balanced Accuracy* | 0.6329 | 0.6142 | 0.5875 | 0.6044 |
| 95% Confidence Interval | (0.6107, 0.655) | (0.5917, 0.6364) | (0.5647, 0.6099) | (0.582, 0.6269) |

**Supplementary Table 3**: T2D Logistic Regression Models of PRS Calculations from 4 Different GWAS (Ge et al., 2022; Khera et al., 2018; Lin et al., 2023; Mars et al., 2020) in the MPBC Cohort.
*PRS polygenic score; AGE = AAI; PC principal component; glm generalized linear model.*
**: Significant p-value*
*glm (formula = T2D ~ T2D PRS + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)*
*PRS are normalized z-Scores and values varies based on the GWAS summary statistics used.*

| | **Estimate** | **Standard Error** | ***P V*alue** |
|---|---|---|---|
| **Ge et al, 2022** | | | |
| Intercept | -2.96200 | 0.95680 | 0.00196* |
| PRS | 0.74551 | 0.10097 | $1.54 \times 10^{-13}$* |
| SEX | -0.78082 | 0.23637 | 0.00096* |
| AGE | 0.01339 | 0.01253 | 0.28532 |
| PC1 | 4.25079 | 3.09973 | 0.17027 |
| PC2 | 4.45538 | 5.20255 | 0.39179 |

| | | | |
|---|---|---|---|
| PC3 | -5.17284 | 4.62050 | 0.26291 |
| PC4 | -7.68378 | 4.29430 | 0.07357 |
| PC5 | 1.31320 | 4.42274 | 0.76653 |
| **Khera et al, 2018** | | | |
| Intercept | -2.85747 | 0.94510 | 0.00250* |
| PRS | 0.57779 | 0.10060 | $9.28 \times 10^{-9}$* |
| SEX | -0.73895 | 0.23454 | 0.00163* |
| AGE | 0.01271 | 0.01237 | 0.30408 |
| PC1 | 1.17085 | 3.25050 | 0.71869 |
| PC2 | 6.16150 | 5.07918 | 0.22510 |
| PC3 | -6.72567 | 4.64248 | 0.14742 |
| PC4 | -5.74890 | 4.24925 | 0.17608 |
| PC5 | 1.06388 | 4.35581 | 0.80704 |
| **Lin et al, 2023** | | | |
| Intercept | -3.10641 | 0.95826 | 0.00119* |
| PRS | 0.64482 | 0.10255 | $3.21 \times 10^{-10}$* |
| SEX | -0.75594 | 0.23452 | 0.00127* |
| AGE | 0.01609 | 0.01247 | 0.19686 |
| PC1 | 5.95945 | 3.09605 | 0.05425 |
| PC2 | 9.20425 | 5.06268 | 0.06905 |
| PC3 | -7.15387 | 4.61151 | 0.12083 |
| PC4 | -6.66797 | 4.28654 | 0.11981 |
| PC5 | 1.81897 | 4.37053 | 0.67727 |
| **Mars et al, 2020** | | | |
| Intercept | -2.78790 | 0.93281 | 0.00280* |
| PRS | 0.50939 | 0.09675 | $1.40 \times 10^{-7}$* |
| SEX | -0.77410 | 0.23396 | 0.00094* |
| AGE | 0.01282 | 0.01222 | 0.29398 |
| PC1 | 6.35329 | 3.05722 | 0.03770* |

| | | | |
|---|---|---|---|
| PC2 | 4.34839 | 5.16739 | 0.40007 |
| PC3 | -4.52878 | 4.56272 | 0.32092 |
| PC4 | -6.49283 | 4.28559 | 0.12976 |
| PC5 | 2.61149 | 4.36647 | 0.54979 |

**Supplementary Table 4**: Comparison of Model Fitness (T2D Logistic Regression Models of PRS Calculations from 4 Different GWAS in the MPBC Cohort)

| **Model** | **AIC** | **BIC** | **McFadden $R^2$** | **AUC** | **Odds Ratio** | **P Value** |
|---|---|---|---|---|---|---|
| Ge | 809.3293 | 859.1037 | 0.0892 | 0.7300 | 2.1075 | 1.5412e-13 |
| Khera | 833.0398 | 882.8141 | 0.0619 | 0.6874 | 1.7821 | 9.2781e-09 |
| Mars | 825.8700 | 875.6443 | 0.0702 | 0.7015 | 1.9056 | 3.2146e-10 |
| Lin | 839.1382 | 888.9126 | 0.0549 | 0.6752 | 1.6643 | 1.4007e-07 |

**Supplementary Table 5**: Confusion Matrices and Classification Metrics for Predicting T2D Status Using 4 Different PRS Calculations

| **Metric** | **Ge** | **Khera** | **Mars** | **Lin** |
|---|---|---|---|---|
| Sensitivity | 0.62931 | 0.66379 | 0.55172 | 0.58621 |
| Specificity | 0.72883 | 0.62071 | 0.74027 | 0.65789 |
| Positive Predictive Value (PPV) | 0.13346 | 0.10405 | 0.12355 | 0.10210 |
| Negative Predictive Value (NPV) | 0.96735 | 0.96530 | 0.96137 | 0.95993 |
| Prevalence | 0.06223 | 0.06223 | 0.06223 | 0.06223 |
| Detection Rate | 0.03916 | 0.04131 | 0.03433 | 0.03648 |

| | | | | |
|---|---|---|---|---|
| Detection Prevalence | 0.29345 | 0.39700 | 0.27790 | 0.35730 |
| *Balanced Accuracy* | 0.67907 | 0.64225 | 0.64600 | 0.62205 |
| 95% Confidence Interval | (0.7017, 0.7429) | (0.6009, 0.6454) | (0.7077, 0.7486) | (0.6534, 0.675) |

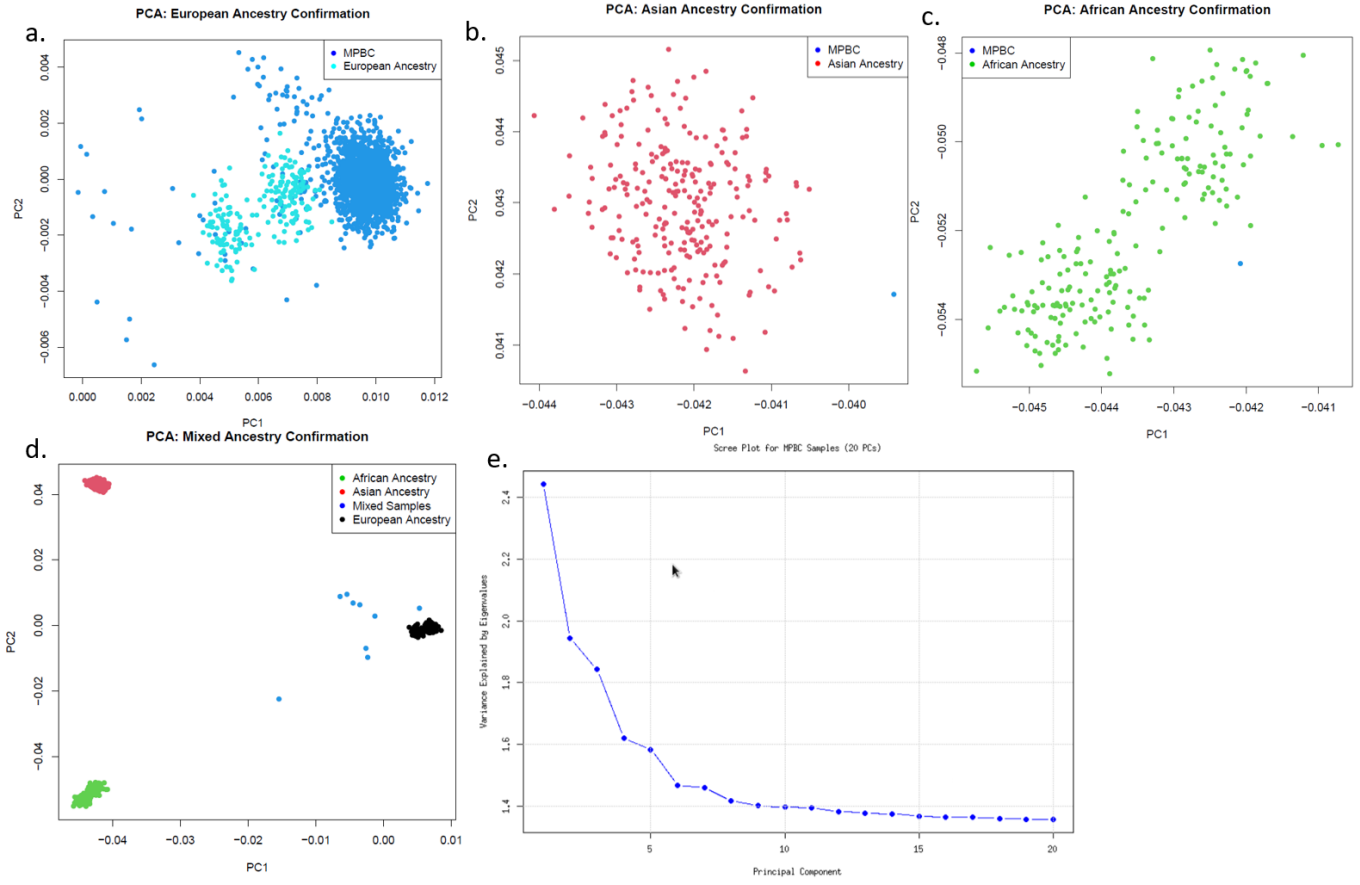
**Supplementary Table 6**: Logistic Regression Modelling PD as an Outcome of T2D PRS, PD PRS, age, sex, and PCs.
*PRS polygenic score; AGE = AAI; PC principal component; glm generalized linear model.*
**: Significant p-value*
*glm (formula = PD ~ T2D PRS + PD PRS + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)*
*PRS are normalized z-Scores.*

| **Term** | **Estimate** | **Standard Error** | **P Value** |
|---|---|---|---|
| Intercept | 1.0657 | 0.4521 | 0.0184* |
| T2D PRS | 0.0112 | 0.0487 | 0.8187 |
| PD PRS | 0.6115 | 0.0513 | $< 2 \times 10^{-16}$* |
| AGE | -0.0171 | 0.0059 | 0.0040* |
| SEX | -0.0251 | 0.1022 | 0.8063 |
| PC1 | -3.0417 | 2.2065 | 0.1680 |
| PC2 | 3.9118 | 2.1606 | 0.0702 |
| PC3 | -5.8704 | 2.1409 | 0.0061* |
| PC4 | 3.1784 | 2.1411 | 0.1377 |
| PC5 | -2.4190 | 2.1193 | 0.2537 |

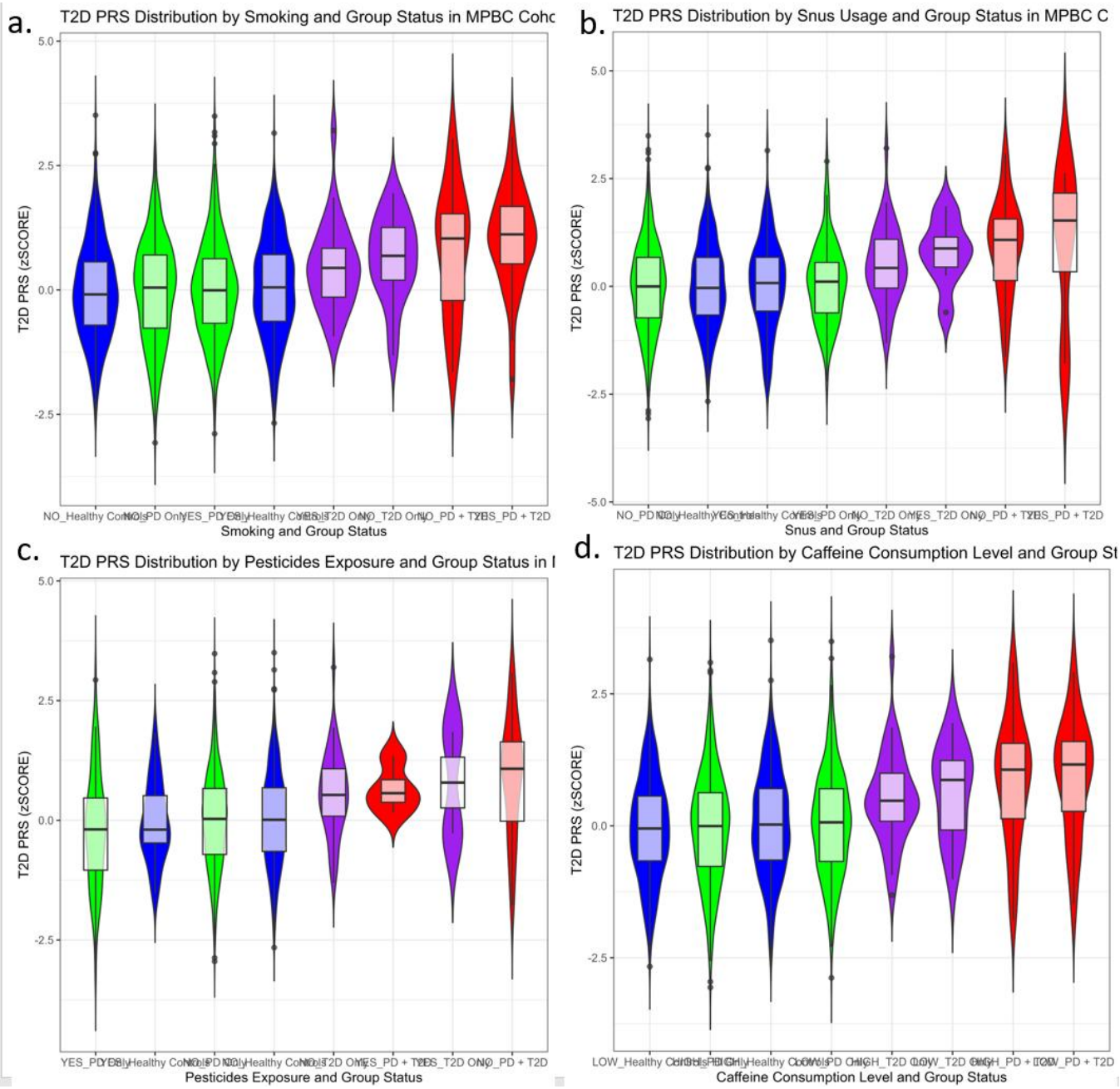
**Supplementary Table 7:** Combined Logistic Regression Results for T2D PRS and Lifestyle Factors Associated with T2D Risk.
*PRS polygenic score; Age = AAI; PC principal component; glm generalized linear model.*
*[1]glm (formula = T2D ~ T2D PRS + Smoking/Snus/Pesticides/Caffeine + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)*
*[2]glm (formula = T2D ~ T2D PRS * Smoking/Snus/Pesticides/Caffeine + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)*

*PRS are normalized z-Scores.*

| Variable | Smoking (Add)[1] | Smoking (Int)[2] | Snus (Add)[1] | Snus (Int)[2] | Pesticides (Add)[1] | Pesticides (Int)[2] | Caffeine (Add)[1] | Caffeine (Int)[2] | Caffeine Levels (Add)[1] | Caffeine Levels (Int)[2] |
|---|---|---|---|---|---|---|---|---|---|---|
| **Intercept** | -2.859 (p=0.003) | -2.862 (p=0.003) | -2.709 (p=0.006) | -2.702 (p=0.006) | -3.139 (p=0.002) | -3.138 (p=0.002) | -3.145 (p=0.007) | -4.160 (p=0.027) | -3.225 (p=0.001) | -3.251 (p=0.001) |
| **T2D PRS** | 0.745 ($p<1\times10^{-13}$) | 0.753 ($p<1\times10^{-7}$) | 0.749 ($p<1\times10^{-13}$) | 0.741 ($p<1\times10^{-12}$) | 0.736 ($p<1\times10^{-11}$) | 0.738 ($p<1\times10^{-12}$) | 3.012 ($p<1\times10^{-13}$) | 7.773 (p=0.146) | 0.736 ($p<1\times10^{-13}$) | 0.778 ($p=2\times10^{-5}$) |
| **Estimate** | -0.209 (p=0.291) | -0.200 (p=0.397) | -0.434 (p=0.187) | -0.488 (p=0.237) | -0.076 (p=0.866) | -0.059 (p=0.911) | 0.242 (p=0.749) | 1.254 (p=0.447) | 0.268 (p=0.221) | 0.308 (p=0.242) |
| **Interaction Term** | N/A | -0.015 (p=0.941) | N/A | 0.074 (p=0.822) | N/A | -0.028 (p=0.949) | N/A | -4.803 (p=0.370) | -0.060 (p=0.780) | -0.060 (p=0.780) |
| **Sex** | -0.799 (p=0.001) | -0.800 (p=0.001) | -0.837 (p<0.001) | -0.836 (p<0.001) | -0.726 (p=0.004) | -0.726 (p=0.004) | -0.761 (p=0.001) | -0.763 (p=0.001) | -0.759 (p=0.001) | -0.759 (p=0.001) |
| **Age** | 0.014 (p=0.265) | 0.014 (p=0.265) | 0.012 (p=0.353) | 0.012 (p=0.356) | 0.015 (p=0.263) | 0.015 (p=0.263) | 0.012 (p=0.331) | 0.012 (p=0.326) | 0.014 (p=0.263) | 0.014 (p=0.265) |
| **PC1** | 4.103 (p=0.186) | 4.084 (p=0.189) | 4.084 (p=0.190) | 4.101 (p=0.188) | 4.802 (p=0.123) | 4.810 (p=0.123) | 4.602 (p=0.136) | 4.366 (p=0.160) | 4.710 (p=0.127) | 4.647 (p=0.133) |
| **PC2** | 4.419 (p=0.395) | 4.430 (p=0.394) | 4.304 (p=0.411) | 4.309 (p=0.410) | 4.806 (p=0.373) | 4.803 (p=0.373) | 5.284 (p=0.326) | 5.344 (p=0.322) | 3.838 (p=0.454) | 3.880 (p=0.449) |
| **PC3** | 5.084 (p=0.270) | 5.075 (p=0.271) | 4.879 (p=0.291) | 4.858 (p=0.293) | 2.994 (p=0.537) | 3.010 (p=0.535) | 5.057 (p=0.277) | 5.183 (p=0.266) | 5.061 (p=0.274) | 5.004 (p=0.279) |

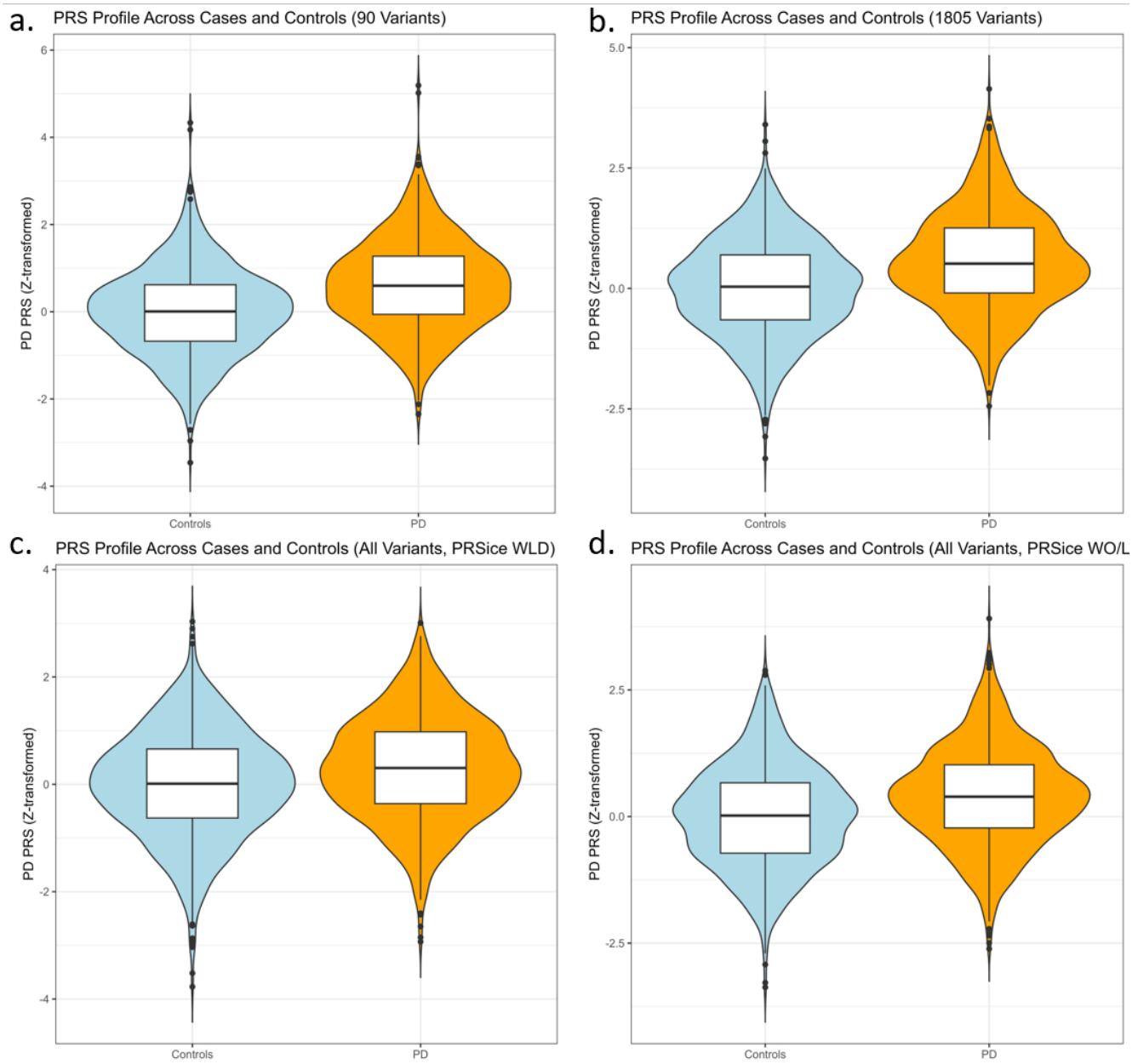


**Supplementary Figure 1:** Clustering of MPBC samples with different populations A. Identification of Individuals with European Ancestry Based on PCA Clustering B. Identification of Individuals with Asian Ancestry Based on PCA Clustering C. Identification of Individuals with African Ancestry Based on PCA Clustering D. Identification of Genetically Admixed Individuals (Mixed Ancestry) Based on PCA E. Scree plot of Principal Components in the MPBC Cohort.

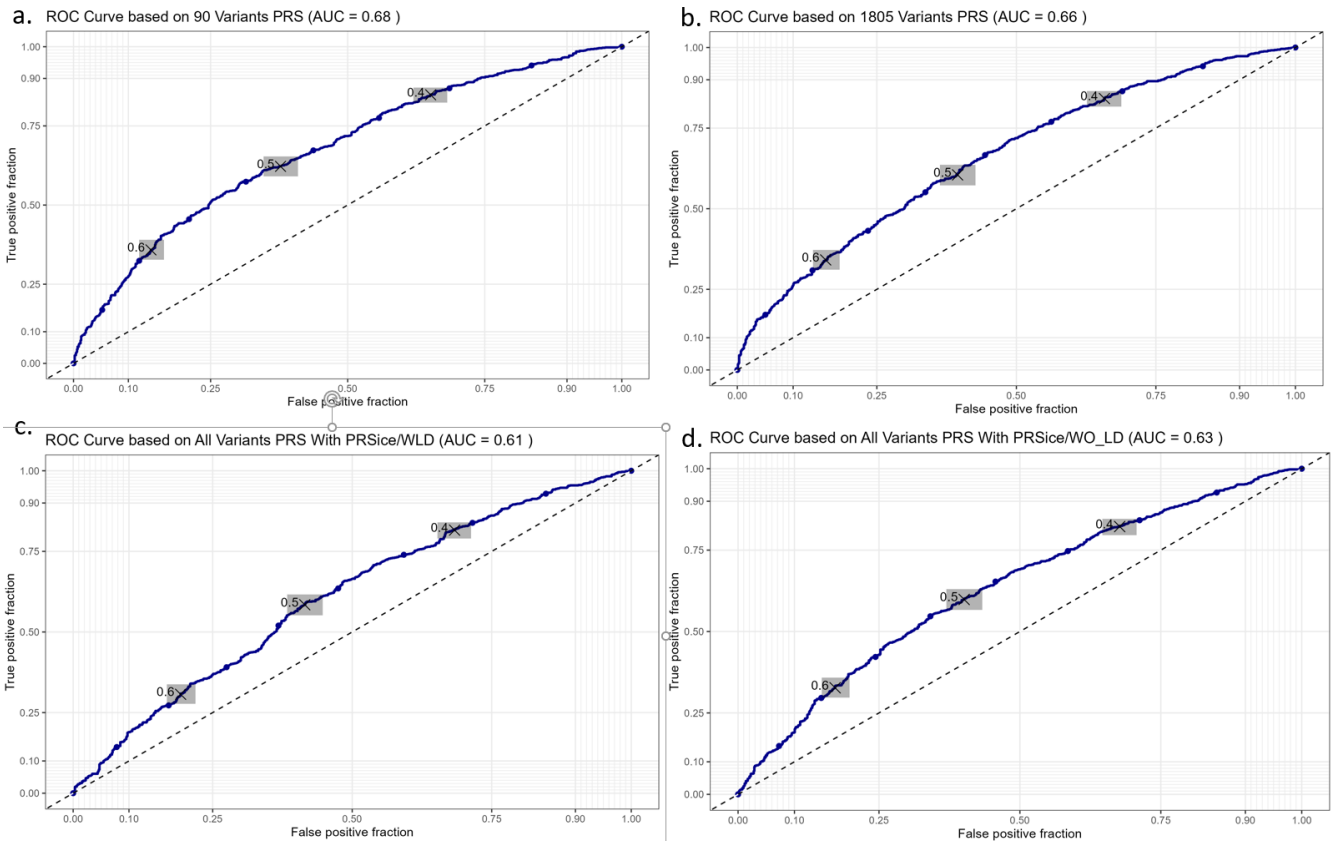



**Supplementary Figure 2:** T2D PRS profiles stratified by each factor across the four disease groups (Controls, PD Only, T2D Only, and PD + T2D). A. Stratification by Smoking Status B.

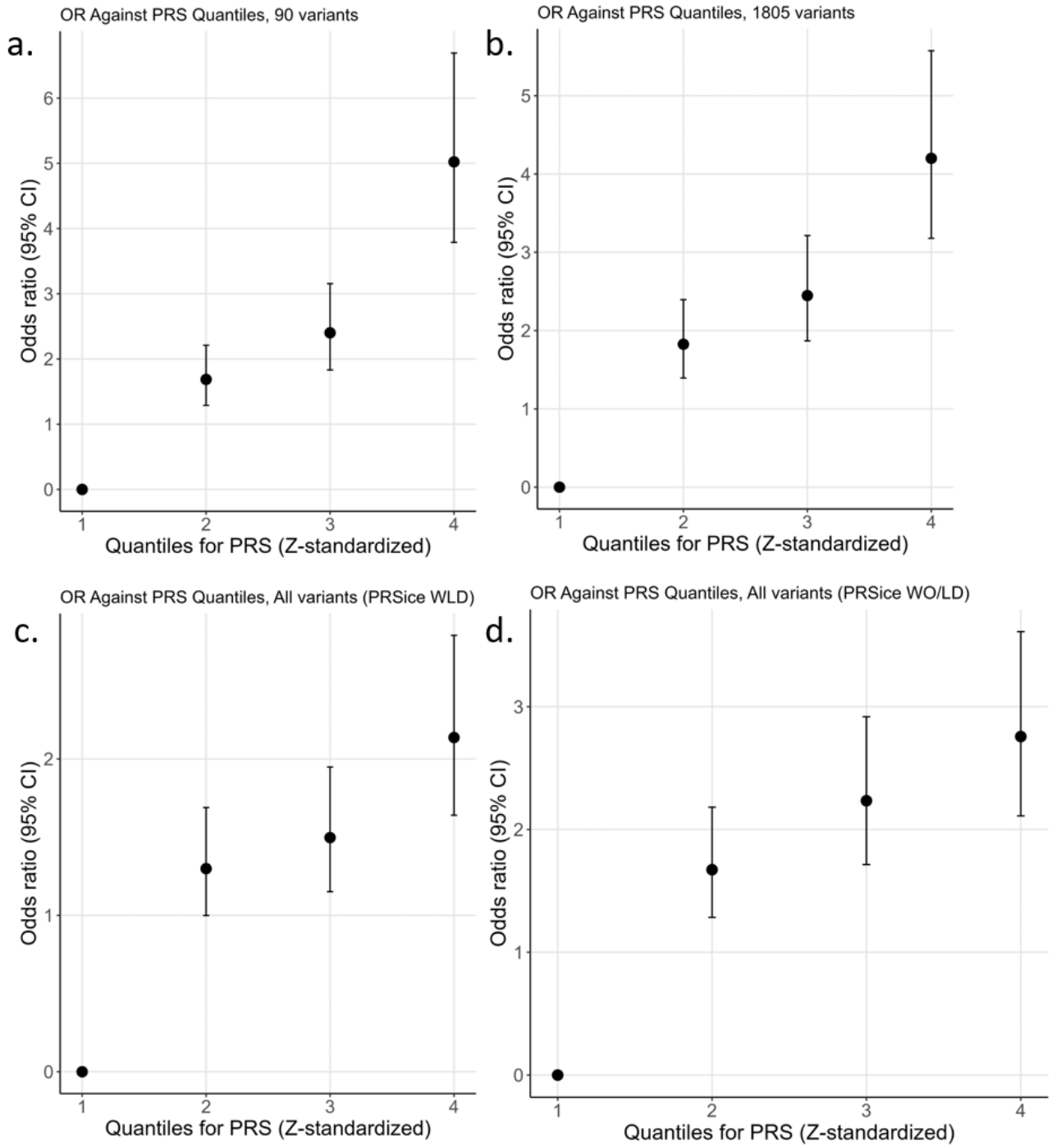
Stratification by Snus Usage c. Stratification by Pesticides Exposure D. Stratification by Caffeine Consumption Level.



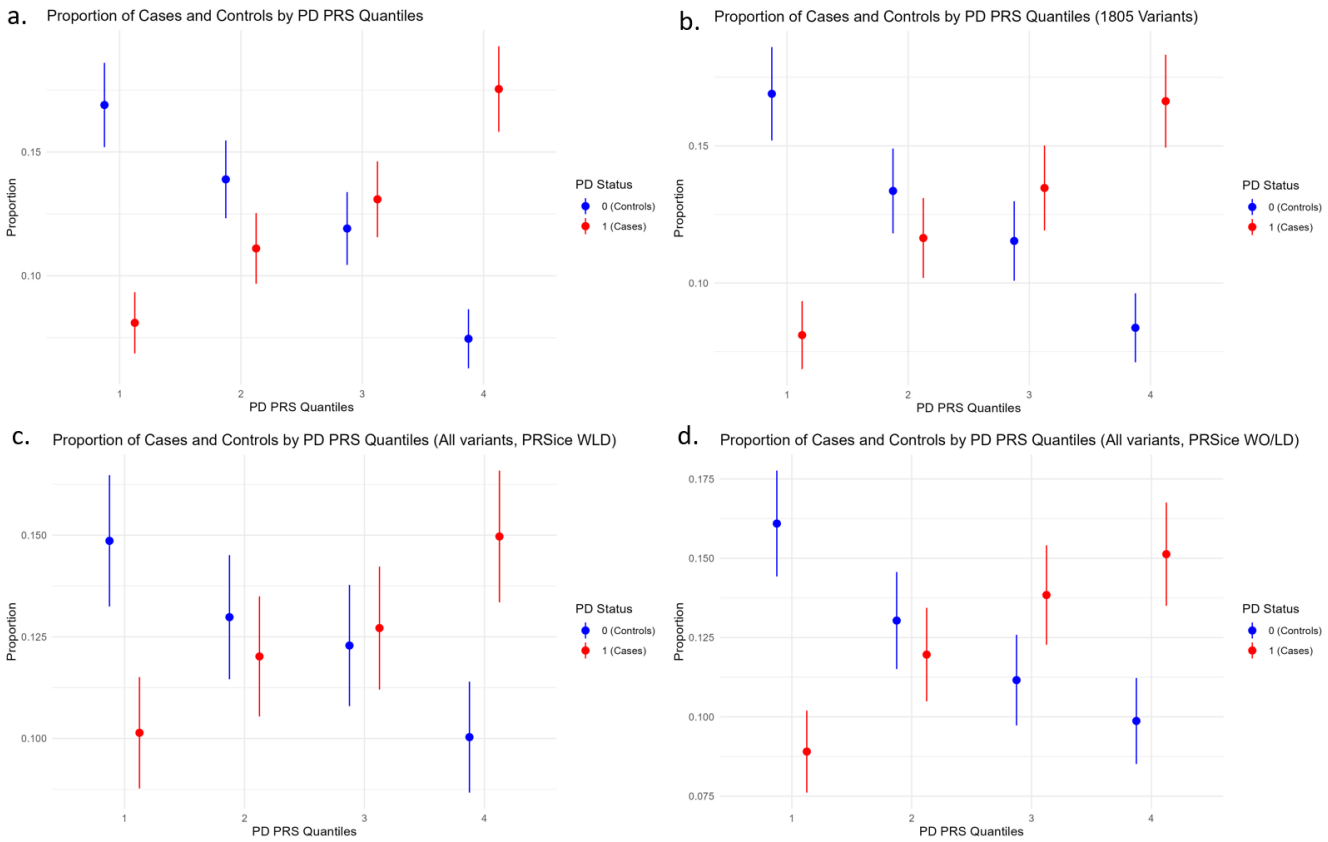

**Supplementary Figure 3:** Violin plots of PD PRS profiles for all 4 PRS construction methods using variants studied by Nalls et al. (2019). A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.



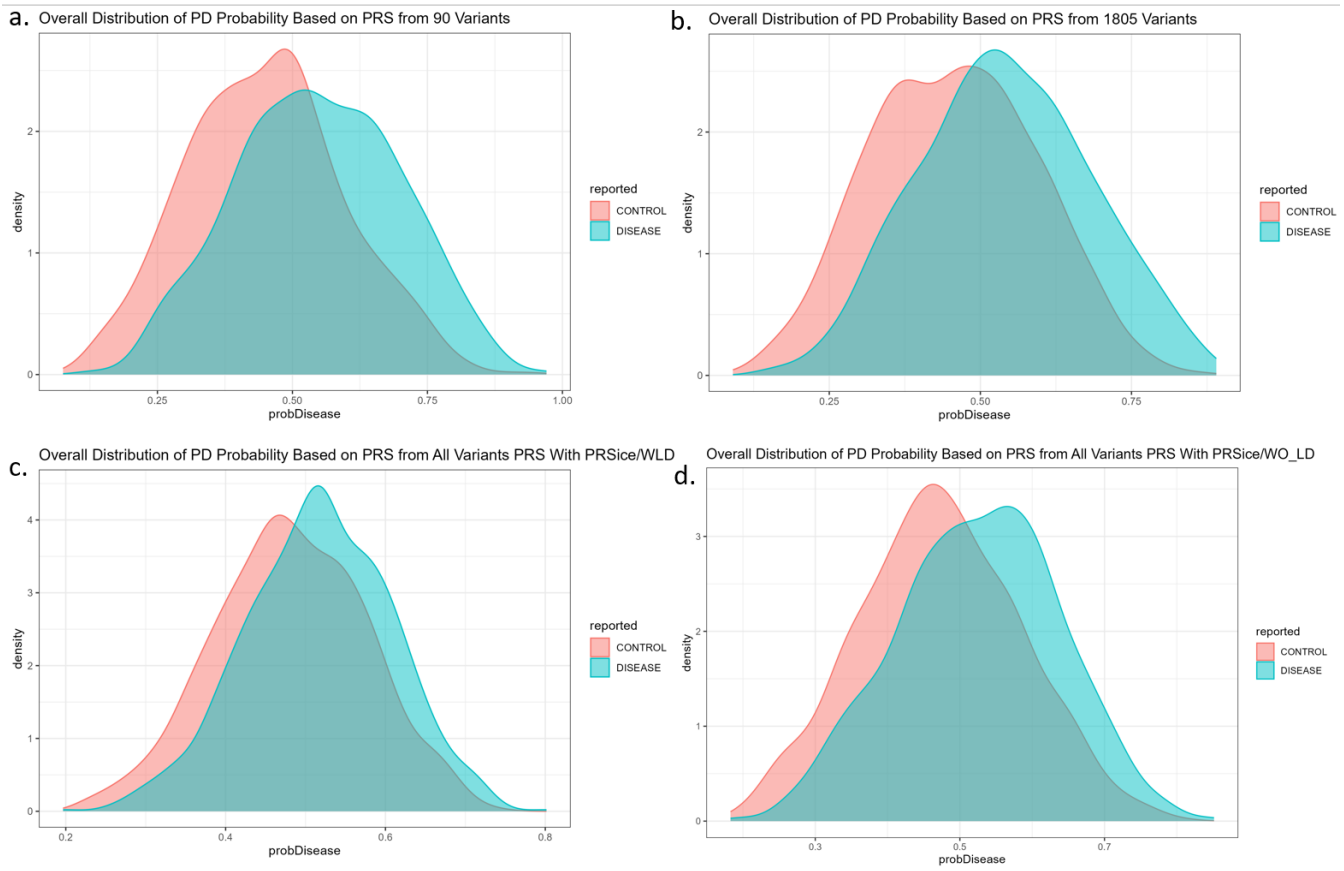

**Supplementary Figure 4:** RcOC curve plots showing logistic model predictive accuracy for all PD PRS construction methods using variants studied by Nalls et al. (2019). A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.



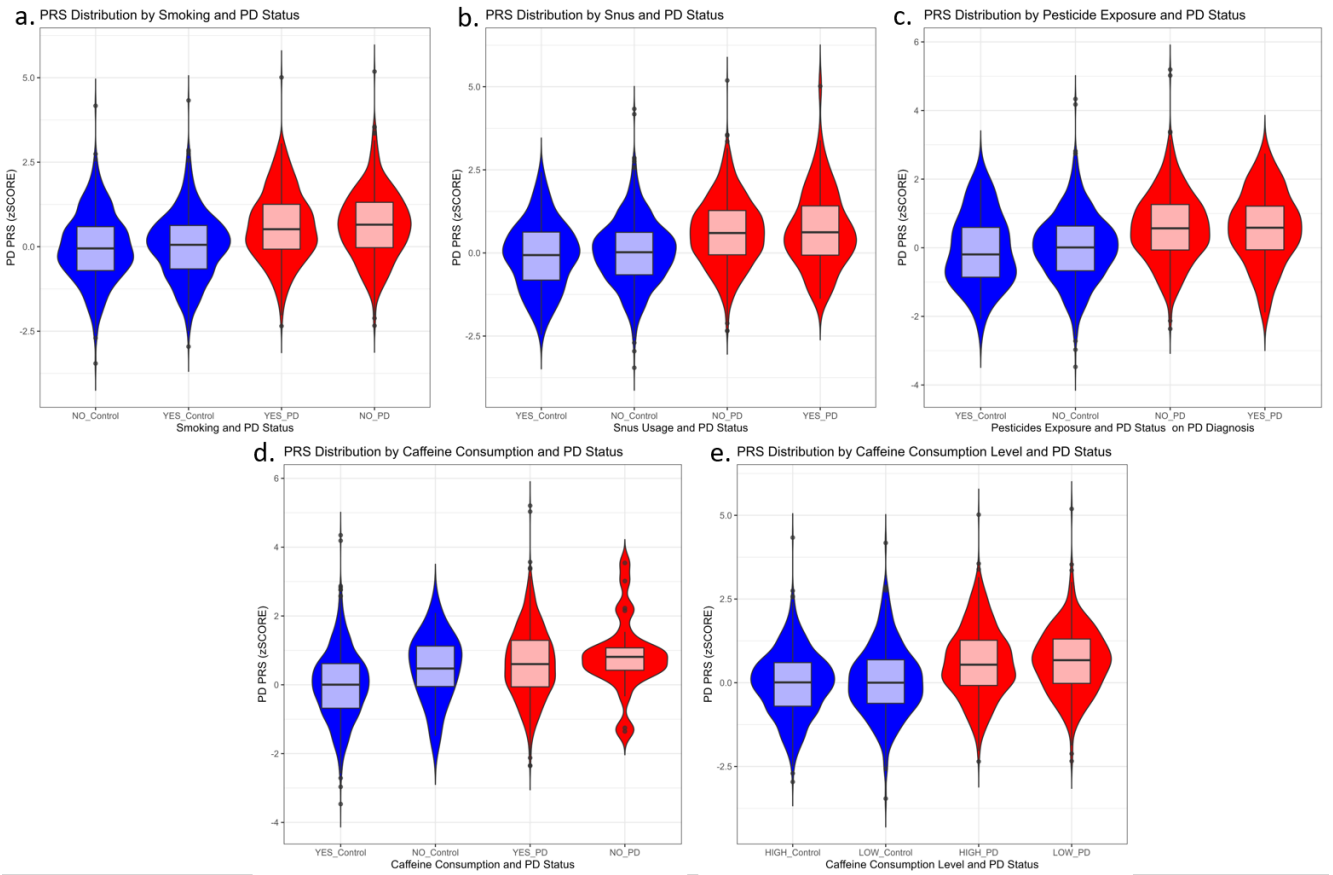

**Supplementary Figure 5:** Quantile plots of odds ratio against normalized PRS (zSCORE) for all PRS construction methods using variants studied by Nalls et al. (2019). A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.



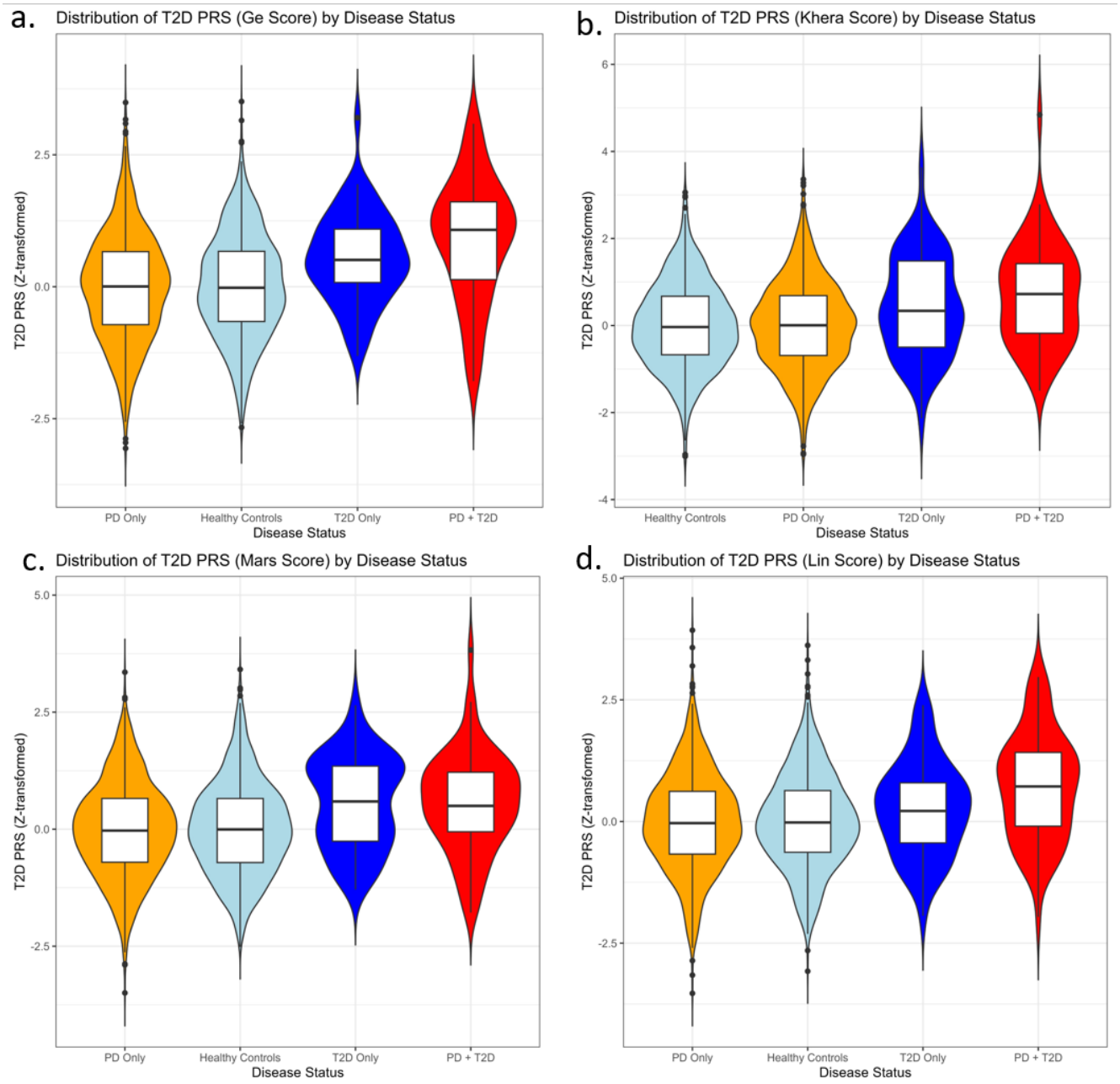

**Supplementary Figure 6:** Quantile plots showing proportion of cases and controls by PD PRS quantiles for all PRS construction methods using variants studied by Nalls et al. (2019). A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.



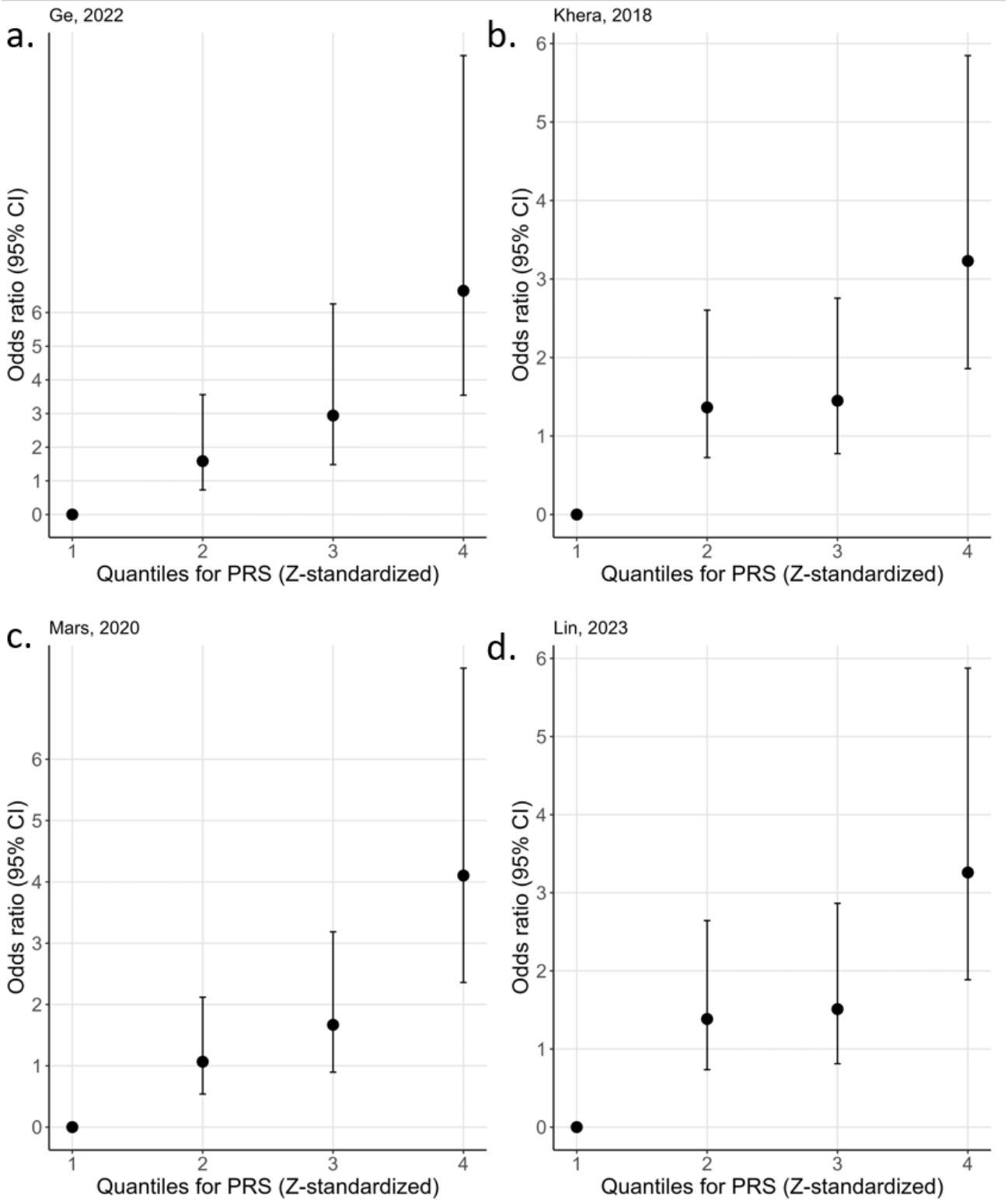

**Supplementary Figure 7:** Density plots of PRS distribution in PD cases and controls against disease probability for all PRS construction methods using variants studied by Nalls et al. (2019). A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.



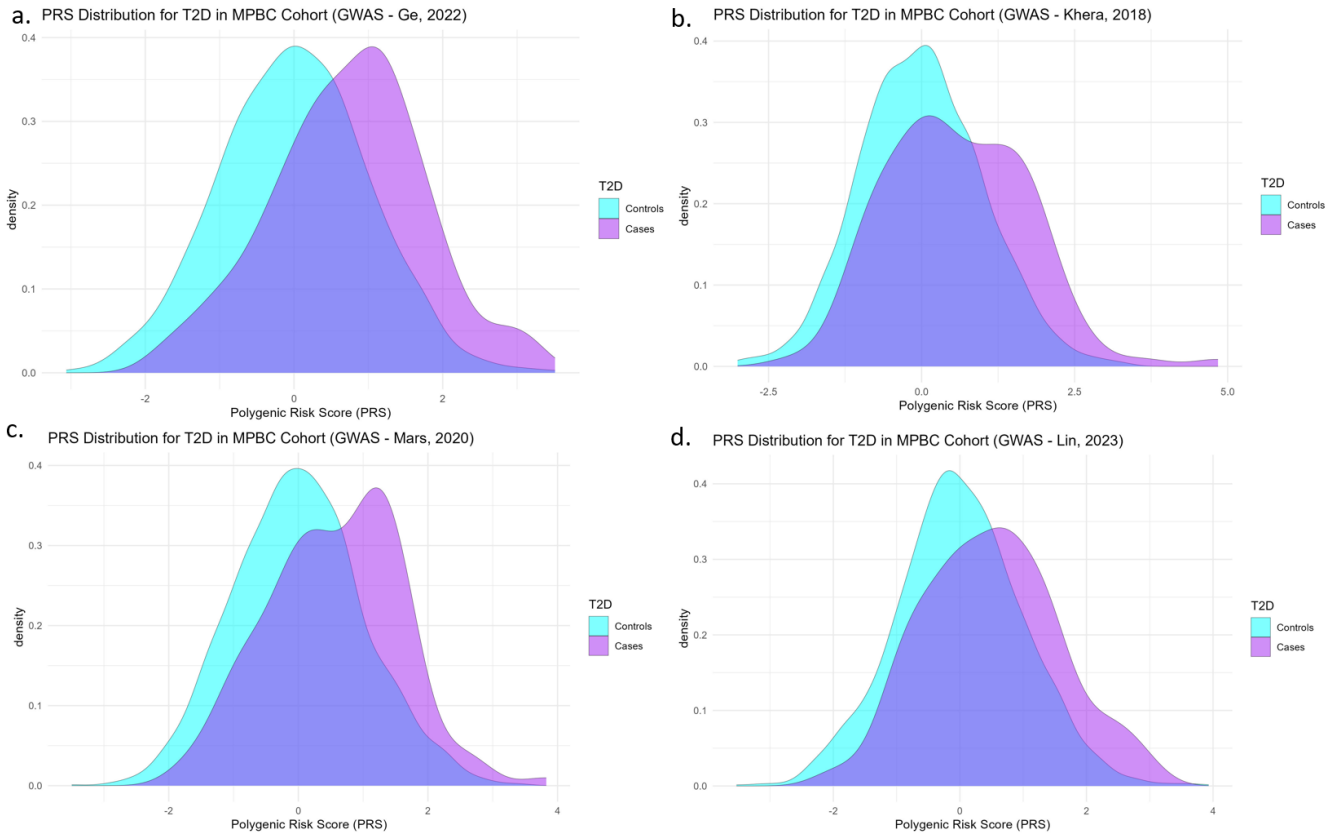

**Supplementary Figure 8:** Stratification of PD PRS distribution by environmental factors and PD status. A. Smoking B. Snus Usage C. Pesticide Exposure D. Caffeine Consumption (Ever/Never) E. Caffeine Consumption Level (High/Low).



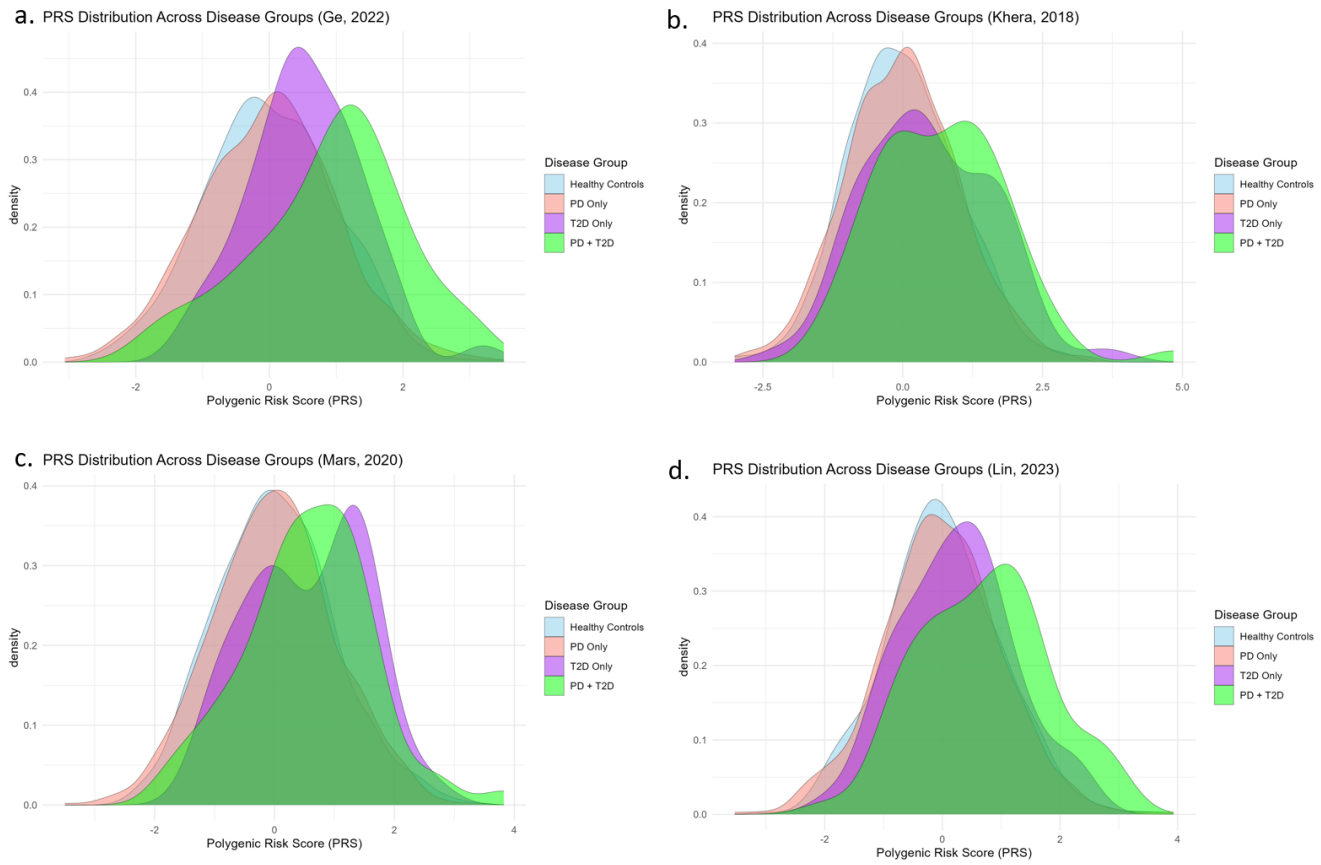

**Supplementary Figure 9:** Violin and Box Plots of T2D PRS Distributions by Disease Status Using 4 GWAS Summary Statistics. A. Ge et al. (2022) B. Khera et al. (2018) C. Mars et al. (2020) D. Lin et al. (2023).



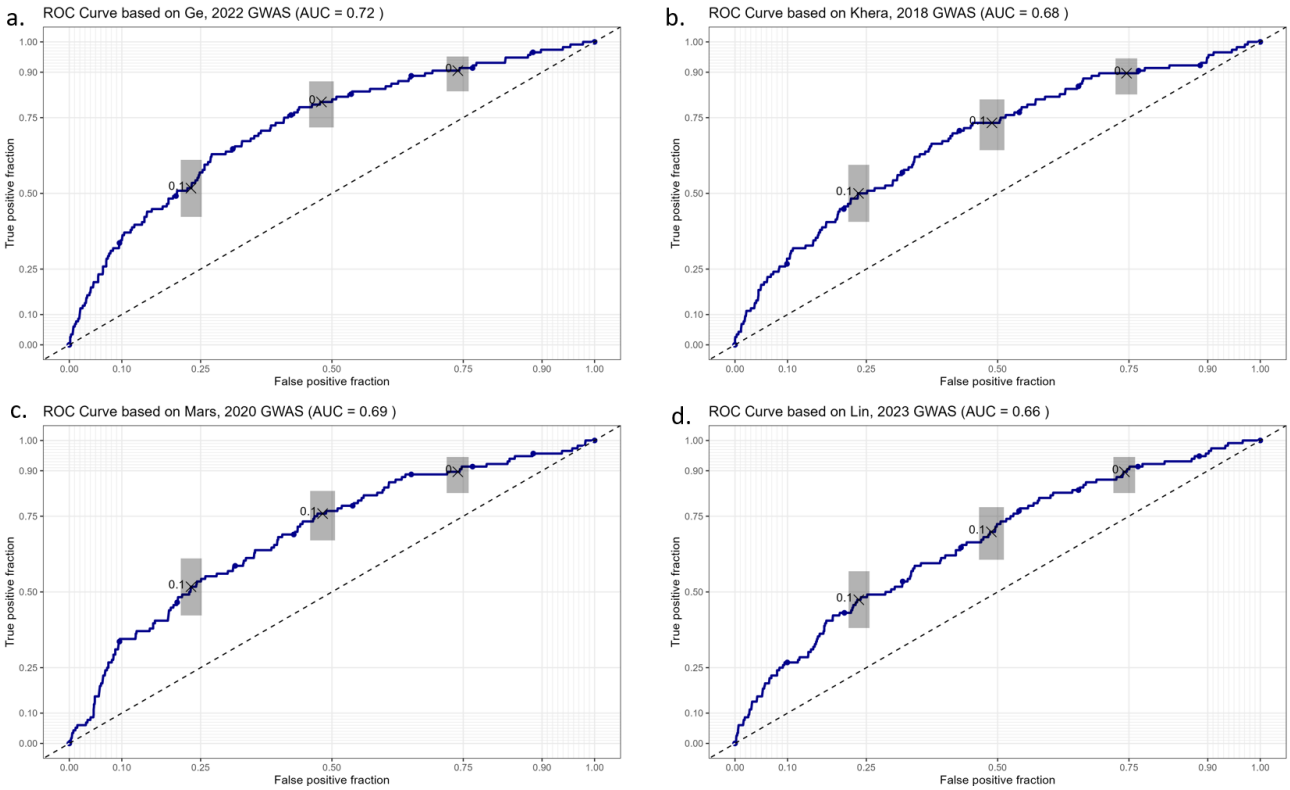

**Supplementary Figure 10:** Quantile Plots Showing Odds Ratio Estimates Across PRS Quantiles for T2D Risk Using 4 GWAS Summary Statistics. A. Ge et al. (2022) B. Khera et al. (2018) C. Mars et al. (2020) D. Lin et al. (2023).




**Supplementary Figure 11:** Density Plots of T2D PRS Distribution for Cases and Controls in the MPBC Cohort Using 4 GWAS Summary Statistics. A. Ge et al. (2022) B. Khera et al. (2018) C. Mars et al. (2020) D. Lin et al. (2023).




**Supplementary Figure 12:** Density Plots of T2D PRS Distribution Across Disease Groups in the MPBC Cohort Using 4 GWAS Summary Statistics. A. Ge et al. (2022) B. Khera et al. (2018) C. Mars et al. (2020) D. Lin et al. (2023).




**Supplementary Figure 13:** ROC Curves for T2D Prediction Using PRS from 4 GWAS Summary Statistics. A. Ge et al. (2022) B. Khera et al. (2018) C. Mars et al. (2020) D. Lin et al. (2023).

Ge, T., Irvin, M. R., Patki, A., Srinivasasainagendra, V., Lin, Y.-F., Tiwari, H. K., Armstrong, N. D., Benoit, B., Chen, C.-Y., Choi, K. W., Cimino, J. J., Davis, B. H., Dikilitas, O., Etheridge, B., Feng, Y.-C. A., Gainer, V., Huang, H., Jarvik, G. P., Kachulis, C., . . . Karlson, E. W. (2022). Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Medicine*, *14*(1), 70. https://doi.org/10.1186/s13073-022-01074-2

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219-1224. https://doi.org/10.1038/s41588-018-0183-z

Lin, J., Yang, H., Zhang, Y., Zhou, L., Chen, Y., Xu, W., & Wang, Y. (2023). Association of time spent in outdoor light and genetic susceptibility with the risk of type 2 diabetes. *Science of The Total Environment*, *888*, 164253. https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.164253

Mars, N., Koskela, J. T., Ripatti, P., Kiiskinen, T. T. J., Havulinna, A. S., Lindbohm, J. V., Ahola-Olli, A., Kurki, M., Karjalainen, J., Palta, P., Palotie, A., Daly, M., Jacob, H., Matakidou, A., Runz, H., John, S., Plenge, R., McCarthy, M., Hunkapiller, J., . . . Analysis, T. (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nature Medicine*, *26*(4), 549-557. https://doi.org/10.1038/s41591-020-0800-0

Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D. A., Noyce, A. J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrøm, L., Siitonen, A., . . . Singleton, A. B. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*, *18*(12), 1091-1102. https://doi.org/10.1016/s1474-4422(19)30320-5