Supplementary Material

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

Supplementary Tables

Supplementary Table 1: Logistic Regression Models of PD Against 4 Different PRS Calculated Using PD -Specific Variants [18] in the MPBC Cohort

	Estimate	**Standard Error**	***P V*alue**
90 Significant SNPs using PLINK[1]			
Intercept	1.0746	0.4532	0.0177*
PRS	0.6148	0.0516	$< 2 \times 10^{-16}$*
SEX	-0.0309	0.1024	0.7628
AGE	-0.0171	0.006	0.004*
PC1	-2.9826	2.2018	0.1755
PC2	3.9184	2.1733	0.0714
PC3	-5.8793	2.1422	0.0061*
PC4	3.2414	2.1189	0.1261
PC5	-2.4258	2.12	0.2525
1,805 SNPs using PLINK[1]			
Intercept	1.0161	0.4501	0.024*
PRS	0.563	0.0508	$< 2 \times 10^{-16}$*
SEX	-0.027	0.1017	0.7908

AGE	-0.016	0.0059	0.0068*
PC1	-2.4567	2.1891	0.2618
PC2	3.5658	2.1517	0.0975
PC3	-5.13	2.1186	0.0155*
PC4	4.7725	2.1019	0.0232*
PC5	-1.5203	2.1061	0.4704
7,862,087 SNPs using PRSice with LD-clumping ($r^2 \leq 0.1$, 250kb window, and $p \leq 1$) [1]			
Intercept	1.1901	0.4366	0.0064*
PRS	0.3029	0.0489	5.99×10^{-10}*
SEX	-0.0145	0.099	0.884
AGE	-0.0171	0.0057	0.0029*
PC1	-3.5056	2.1348	0.1006
PC2	3.4033	2.1009	0.1052
PC3	-3.0591	2.0538	0.1364
PC4	5.1257	2.0426	0.0121*
PC5	-3.2023	2.0456	0.1175
7,862,087 SNPs using PRSice without LD-clumping[1]			
Intercept	1.2386	0.4397	0.0049*
PRS	0.4083	0.0492	$< 2 \times 10^{-16}$*
SEX	-0.0283	0.0999	0.7772
AGE	-0.018	0.0058	0.0018*
PC1	-2.4578	2.1487	0.2527

PC2	3.8941	2.1188	0.0661
PC3	-3.4255	2.0732	0.0985
PC4	4.529	2.064	0.0282*
PC5	-3.536	2.0629	0.0865
PRS-AAD Model[2]			
Intercept	65.1632	1.0752	< 2e-16*
zSCORE	-0.9976	0.3630	0.0061*
SEX	0.1805	0.7423	0.808
PC1	-38.9462	18.6077	0.0367*
PC2	-11.7065	16.6587	0.4824
PC3	-7.0433	15.9747	0.6594
PC4	-12.4637	15.4129	0.419
PC5	-11.8135	14.9366	0.4292

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.

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 [32-35] in the MPBC Cohort.

	Estimate	**Standard Error**	***P V*alue**
Ge et al, 2022			
Intercept	-2.962	0.9568	0.002*
PRS	0.7455	0.101	1.54×10^{-13}*
SEX	-0.7808	0.2364	0.0009*
AGE	0.0134	0.0125	0.2853
PC1	4.2508	3.0997	0.1703
PC2	4.4554	5.2026	0.3918
PC3	-5.1728	4.6205	0.2629
PC4	-7.6838	4.2943	0.0736
PC5	1.3132	4.4227	0.7665

Khera et al, 2018			
Intercept	-2.8575	0.9451	0.0025*
PRS	0.5778	0.1006	9.28×10^{-9}*
SEX	-0.739	0.2345	0.0012*
AGE	0.0127	0.0124	0.3041
PC1	1.1709	3.2505	0.7187
PC2	6.1615	5.07918	0.2251
PC3	-6.7257	4.6425	0.1474
PC4	-5.7489	4.2493	0.1761
PC5	1.0639	4.3558	0.807
Lin et al, 2023			
Intercept	-3.1064	0.9583	0.0012*
PRS	0.6448	0.1026	3.21×10^{-10}*
SEX	-0.7559	0.2345	0.0013*
AGE	0.0161	0.0125	0.1969
PC1	5.9595	3.0961	0.0542
PC2	9.2043	5.0627	0.0691
PC3	-7.1538	4.6115	0.1208
PC4	-6.668	4.2865	0.1198
PC5	1.819	4.3705	0.6773
Mars et al, 2020			
Intercept	-2.7879	0.9328	0.0028*
PRS	0.5094	0.0968	1.40×10^{-7}*
SEX	-0.7741	0.234	0.0009*
AGE	0.0128	0.0122	0.294
PC1	6.3533	3.0572	0.0377*
PC2	4.3484	5.1674	0.4001
PC3	-4.5288	4.5627	0.3209
PC4	-6.4928	4.2856	0.1298

PC5	2.6115	4.3665	0.5498

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.

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	0.6293	0.6638	0.5517	0.5862
Specificity	0.7288	0.6207	0.7403	0.6579
Positive Predictive Value (PPV)	0.1335	0.1041	0.1236	0.1021
Negative Predictive Value (NPV)	0.9674	0.9653	0.9614	0.9599
Prevalence	0.0622	0.0622	0.0622	0.0622

Detection Rate	0.0392	0.0413	0.0343	0.0365
Detection Prevalence	0.2935	0.397	0.2779	0.3573
Balanced Accuracy	0.6791	0.6423	0.646	0.6221
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, adjusting for PD PRS, age, sex, and PCs.

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

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.

Supplementary Table 7: Combined Logistic Regression Results for T2D PRS and Lifestyle Factors Associated with T2D Risk.

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)

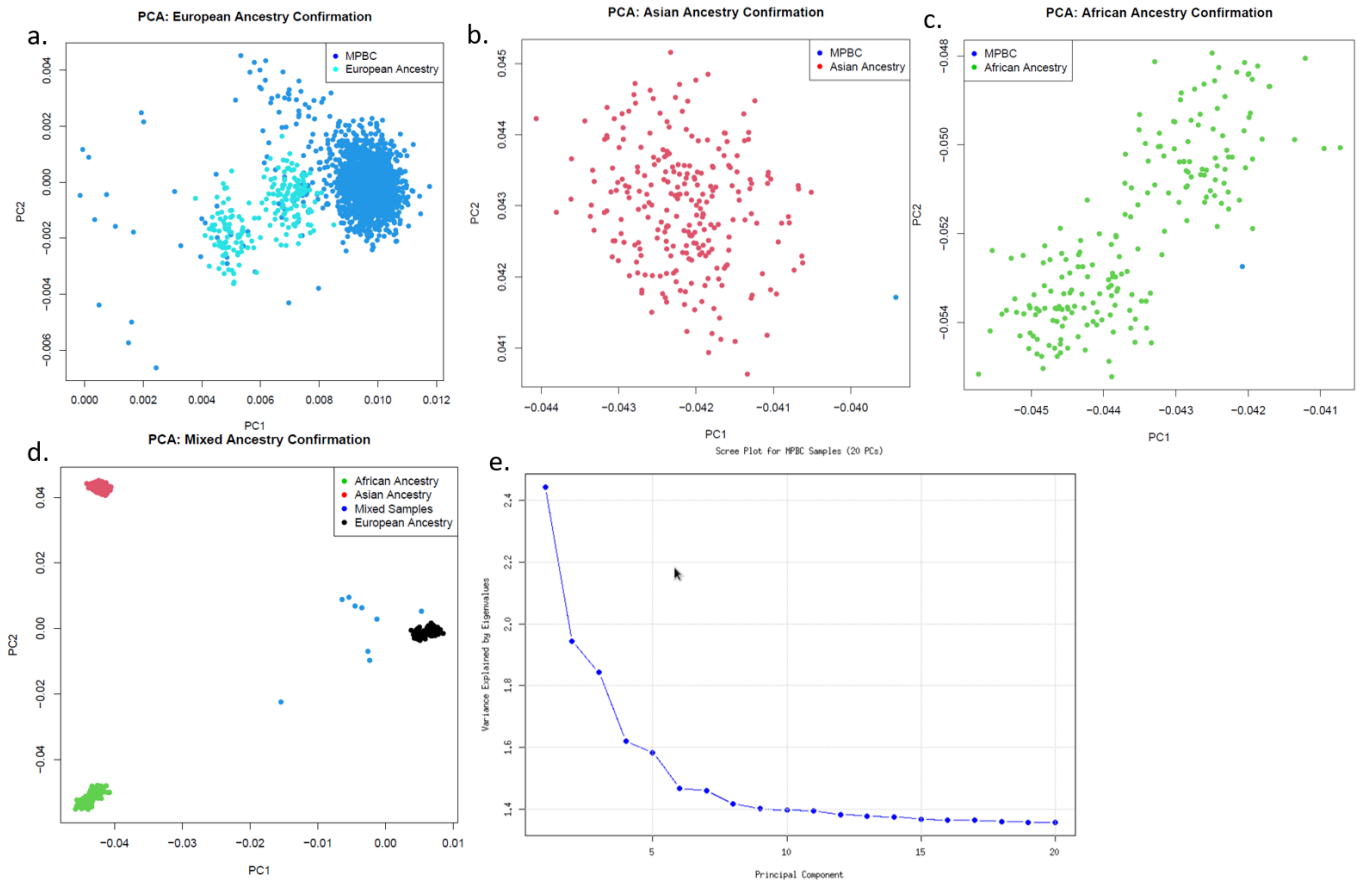
PRS polygenic score; Age = AAI; PC principal component; glm generalized linear model.

[1]glm (formula = T2D ~ T2D PRS + Smoking/Snus/Pesticides/Caffeine + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)

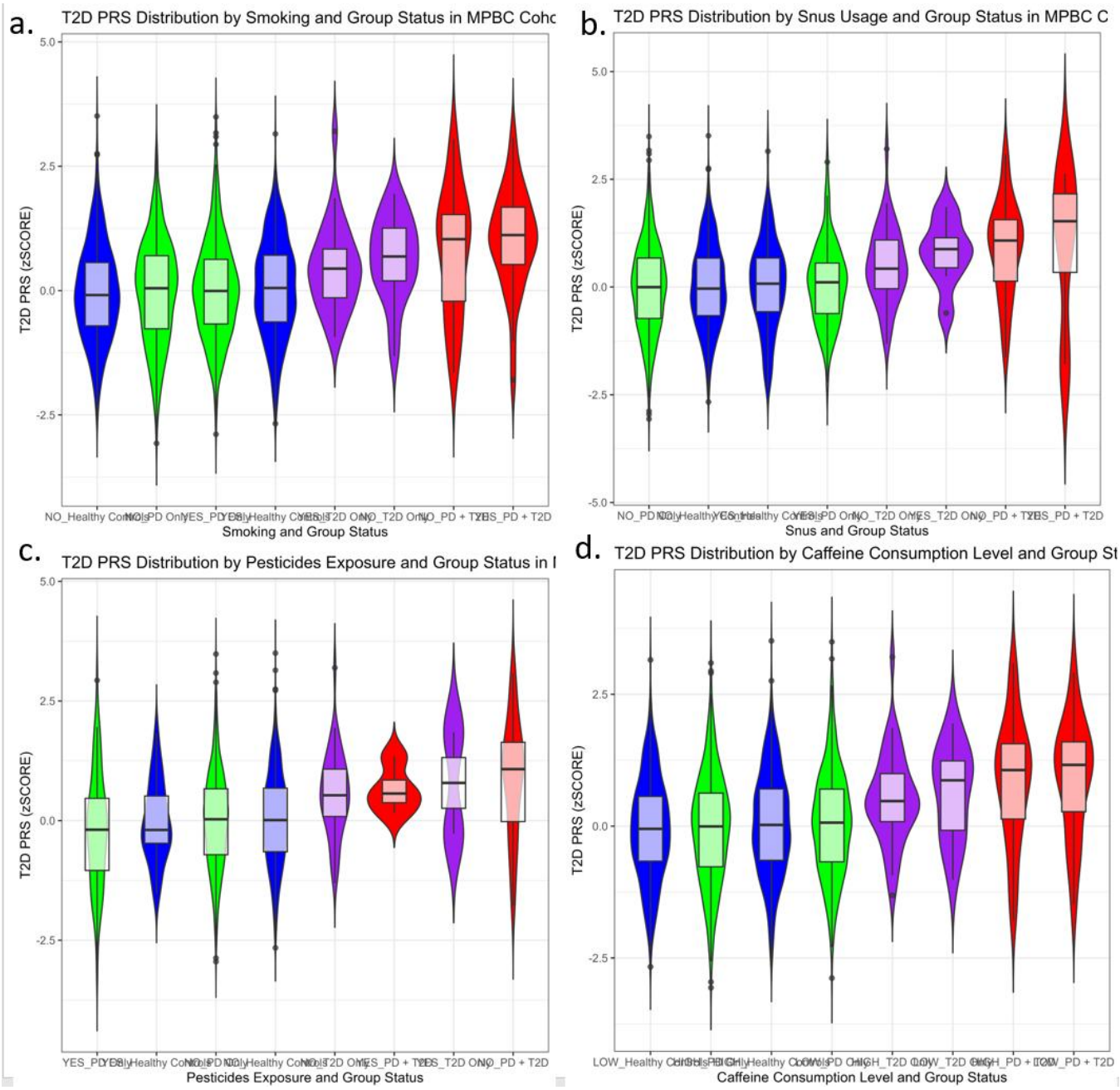
*[2]glm (formula = T2D ~ T2D PRS * Smoking/Snus/Pesticides/Caffeine + SEX + AGE + PC1 + PC2 + PC3 + PC4 + PC5, family = binomial, data = data)*

PRS are normalized z-Scores.

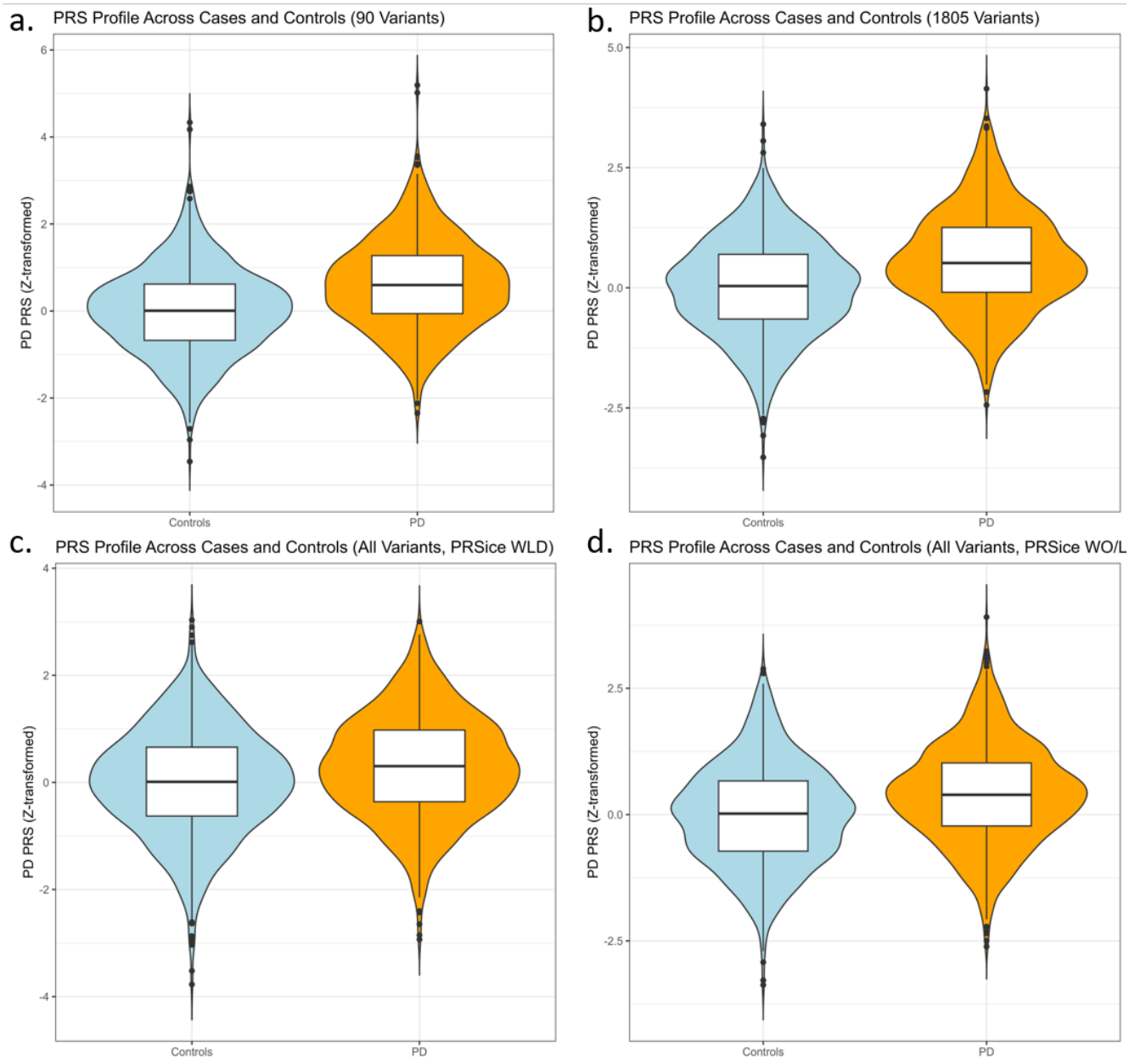
Supplementary Figures



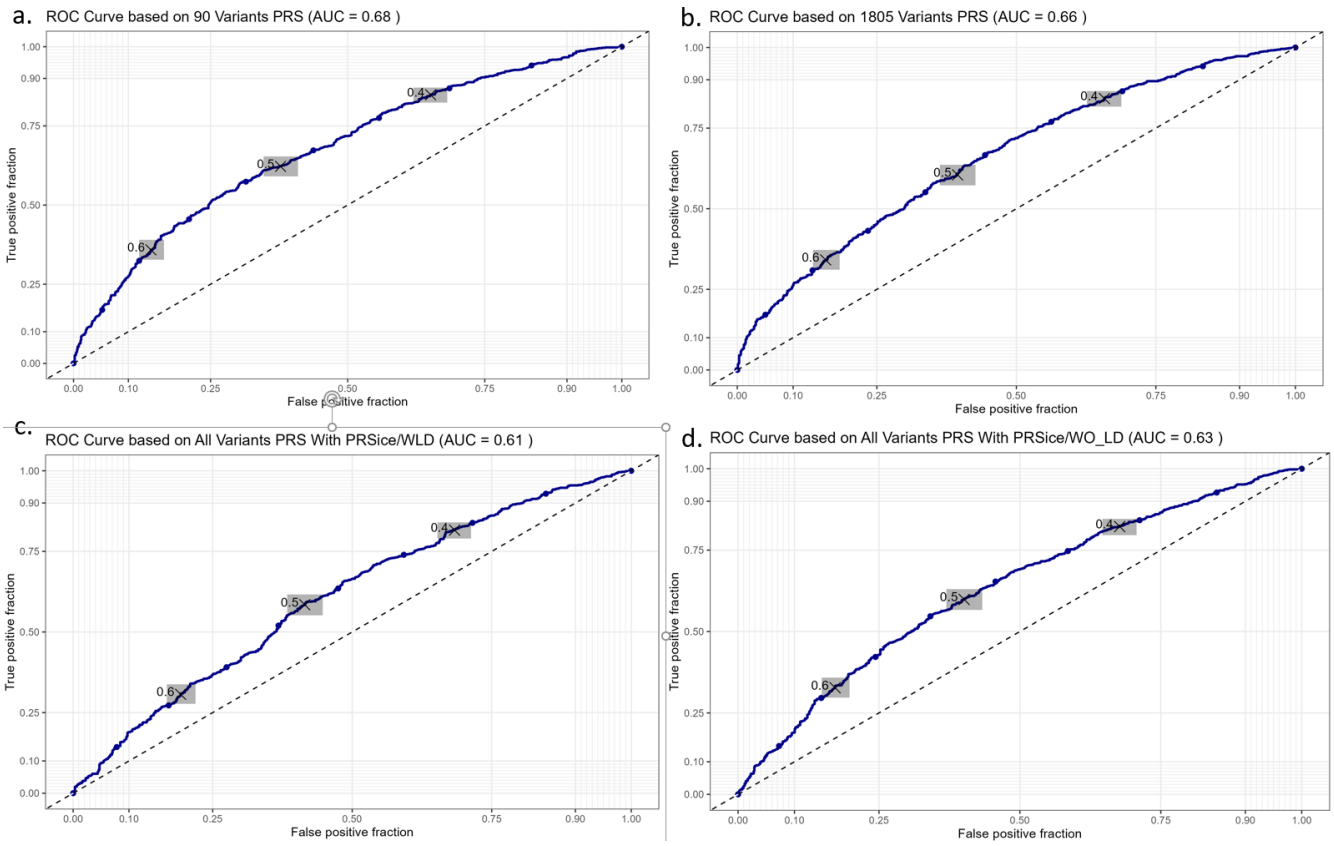
Supplementary Figure 1: Clustering of MPBC samples with different populations. *A. Identification of individuals with European ancestry based on PCA clustering. B. Identification of individuals with Asian ancestry based on PCA clustering. C. Identification of individuals with African ancestry based on PCA clustering. D. Identification of genetically admixed individuals (Mixed ancestry) based on PCA. E. Scree plot of the calculated principal components in the MPBC Cohort.*



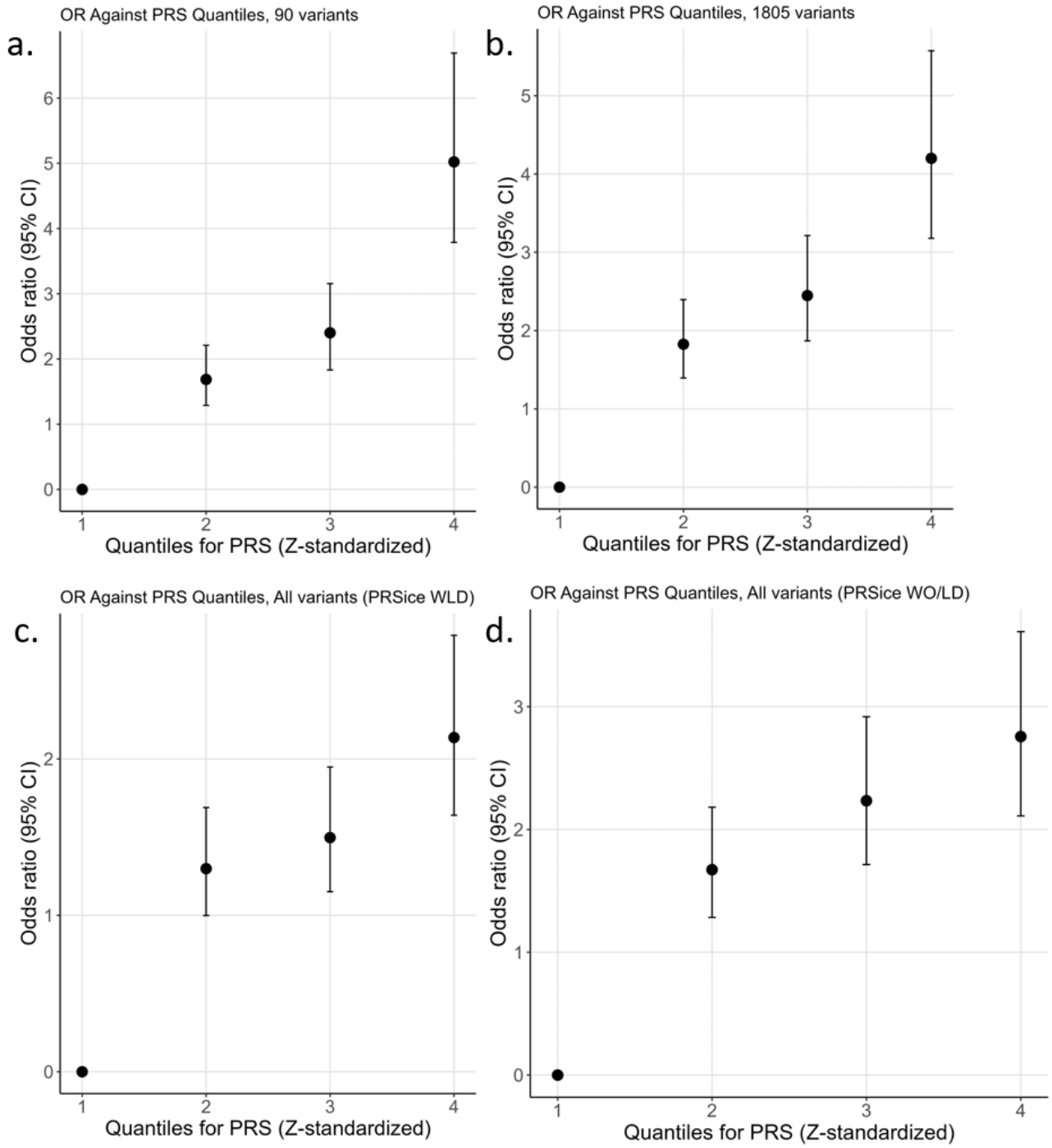
Supplementary Figure 2: T2D PRS profiles stratified by each factor across the four disease groups (Controls, PD Only, T2D Only, and PD + T2D). *A. Stratification by smoking status B. stratification by snus usage c. Stratification by pesticides exposure D. Stratification by caffeine consumption Level.*



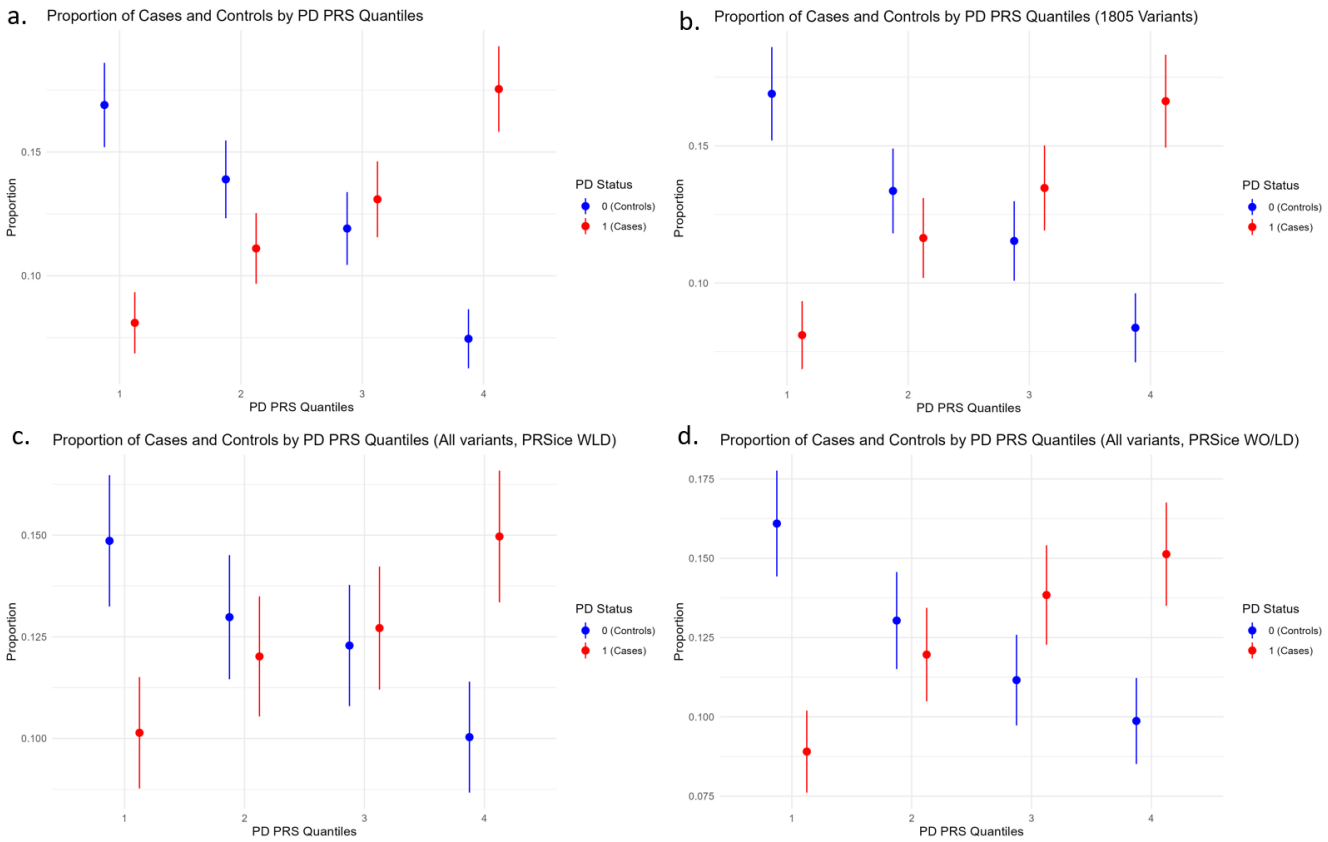
Supplementary Figure 3: Violin plots of PD PRS profiles for all 4 PRS construction methods using PD-specific SNPs. *A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.*



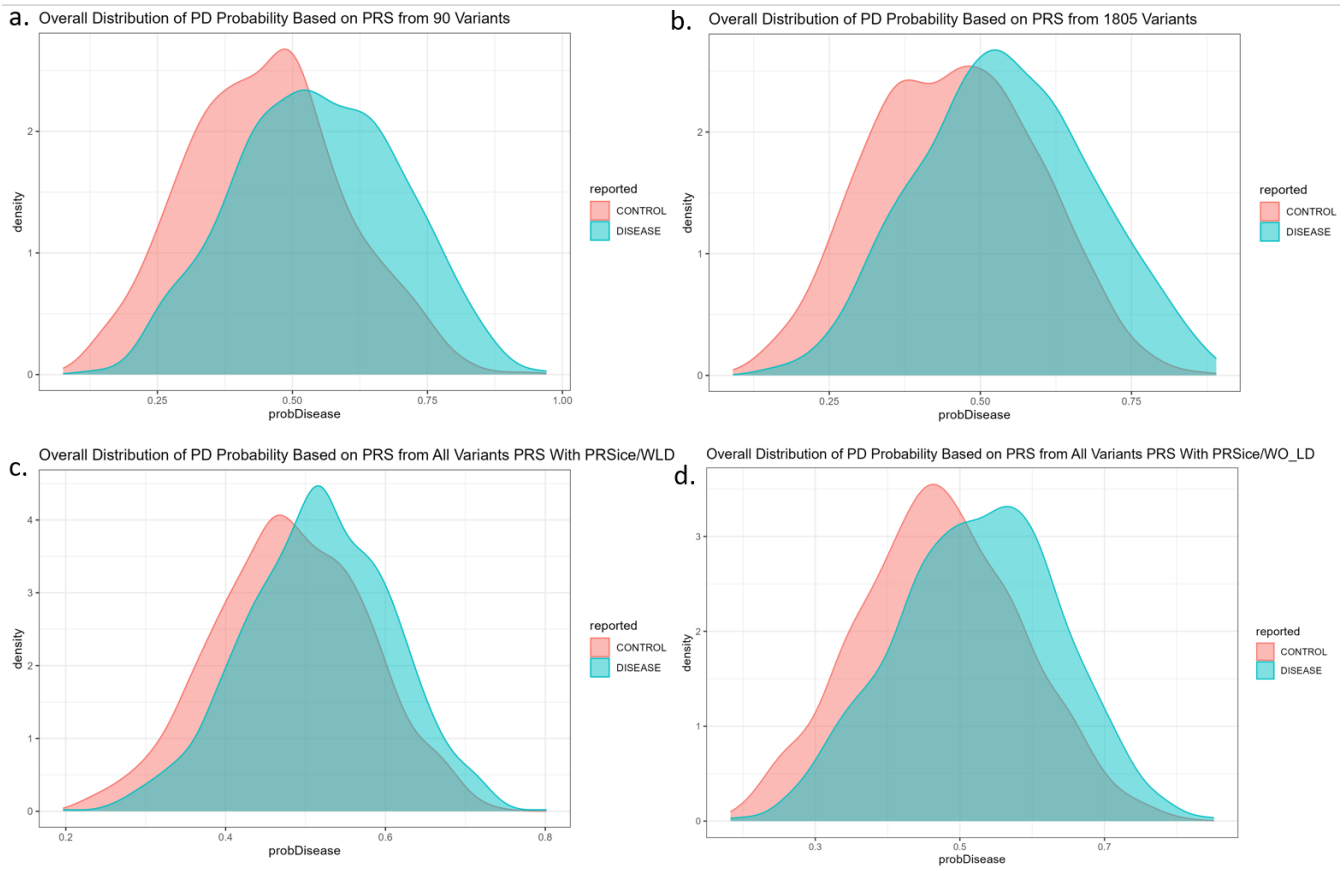
Supplementary Figure 4: ROC curve plots showing logistic model predictive accuracy for all PD PRS construction methods using PD-specific SNPs. *A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.*



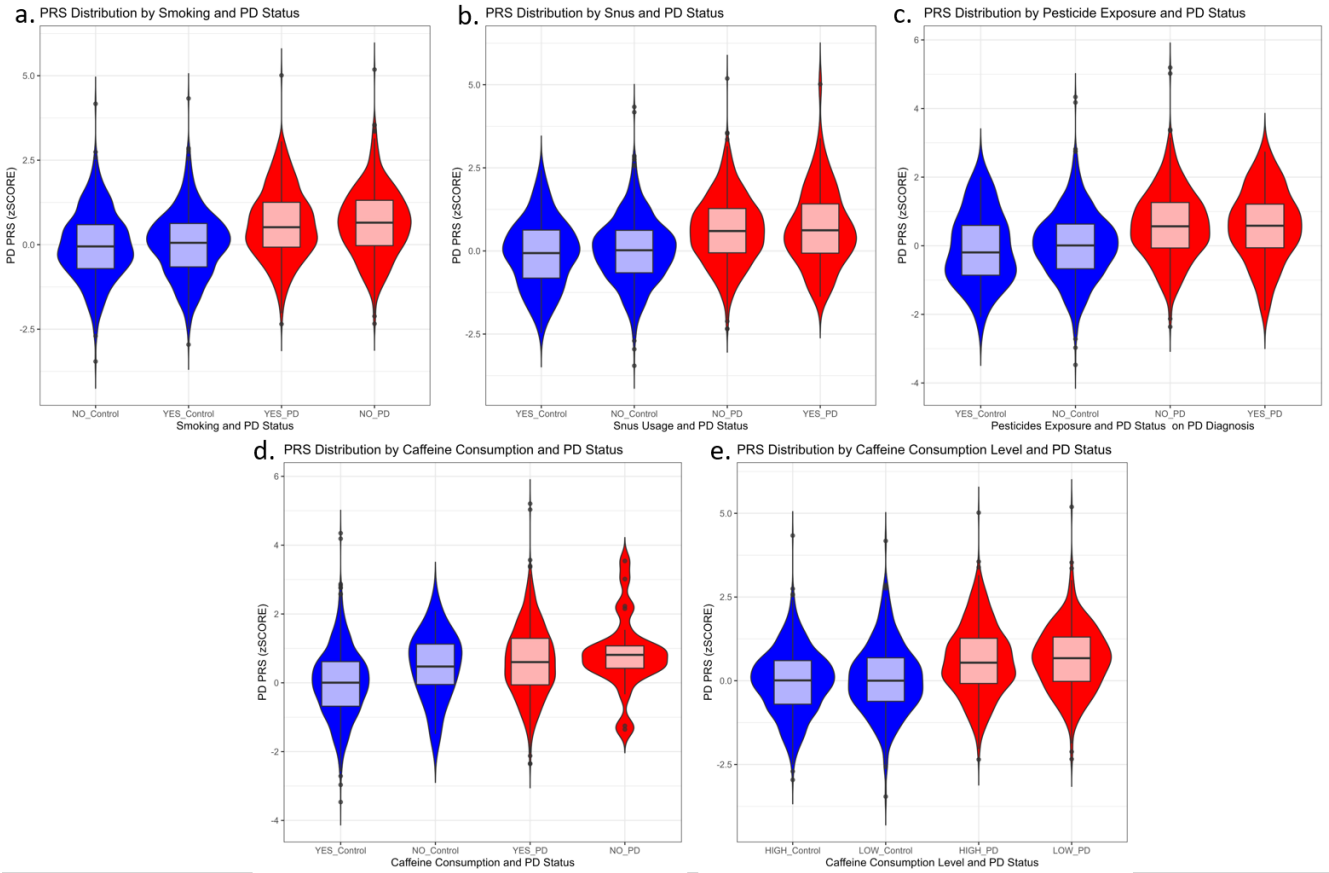
Supplementary Figure 5: Quantile plots of odds ratio against normalized PRS (zSCORE) for all PRS construction methods using PD-specific SNPs. *A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.*



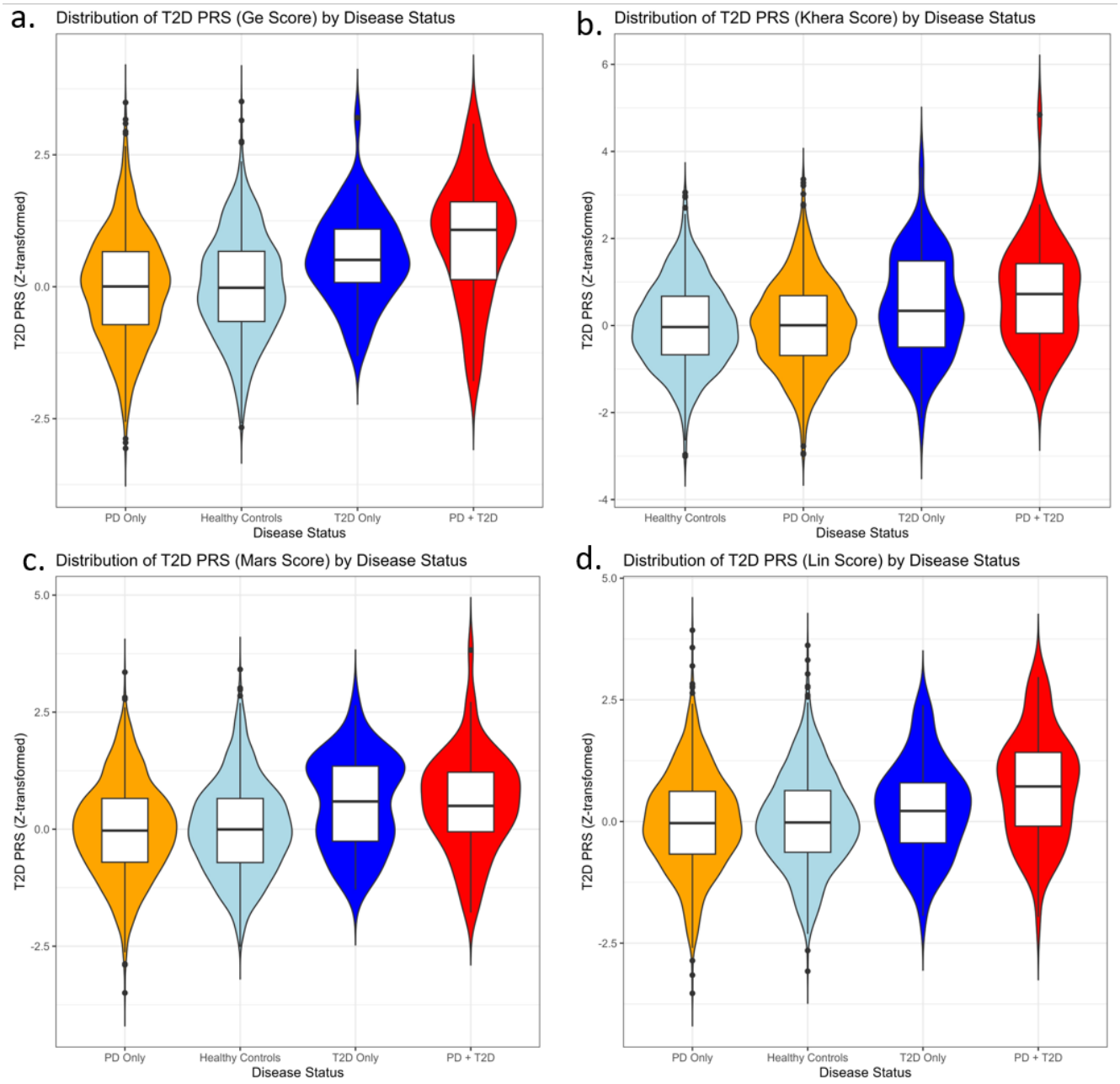
Supplementary Figure 6: Quantile plots showing proportion of cases and controls by PD PRS quantiles for all PRS construction methods using PD-specific SNPs. *A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.*



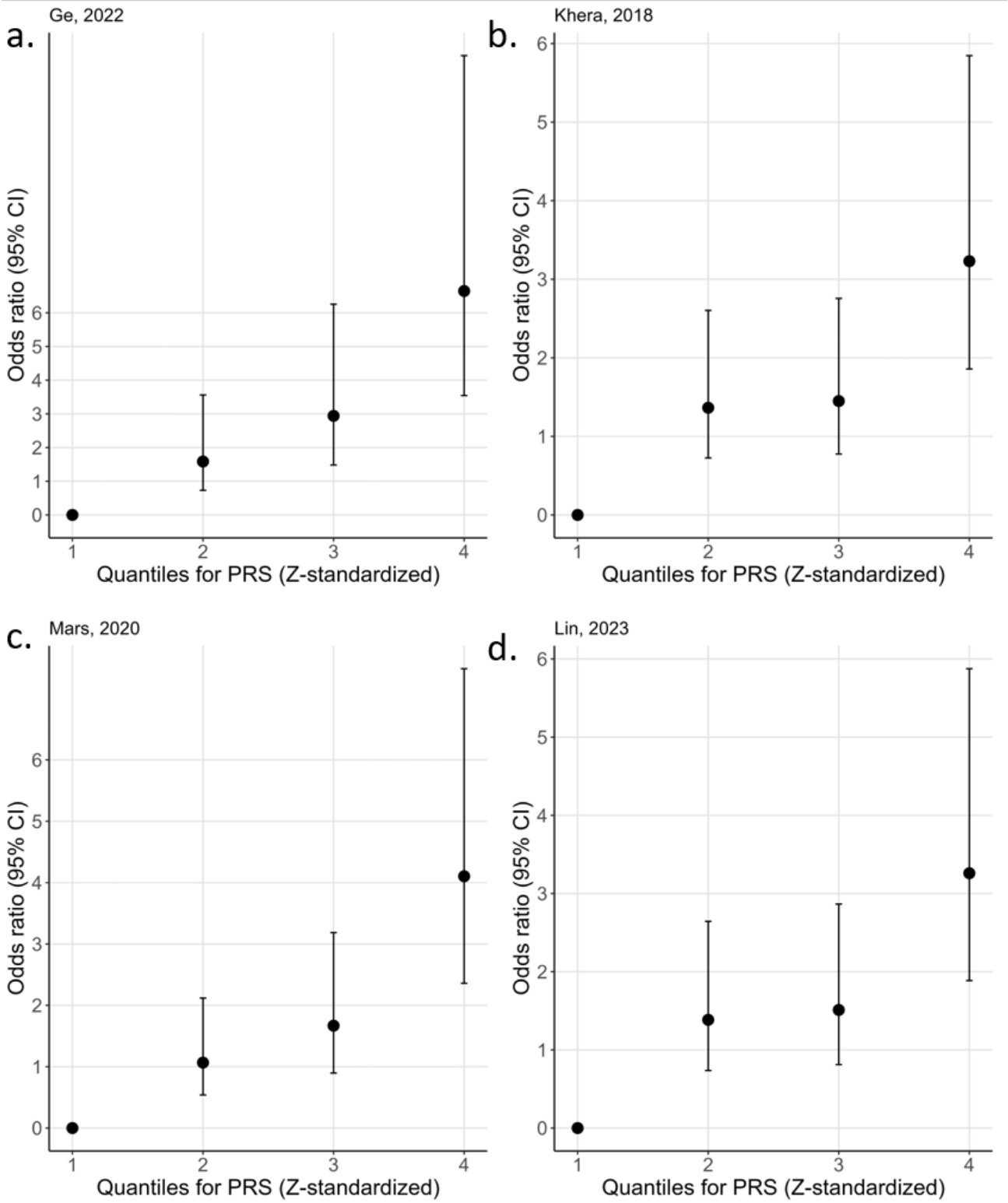
Supplementary Figure 7: Density plots of PRS distribution in PD cases and controls against disease probability for all PRS construction methods using PD-specific SNPs. *A. 90 significant SNPs B. 1805 SNPS. C. All SNPs with LD clumping D. All SNPs without LD clumping.*



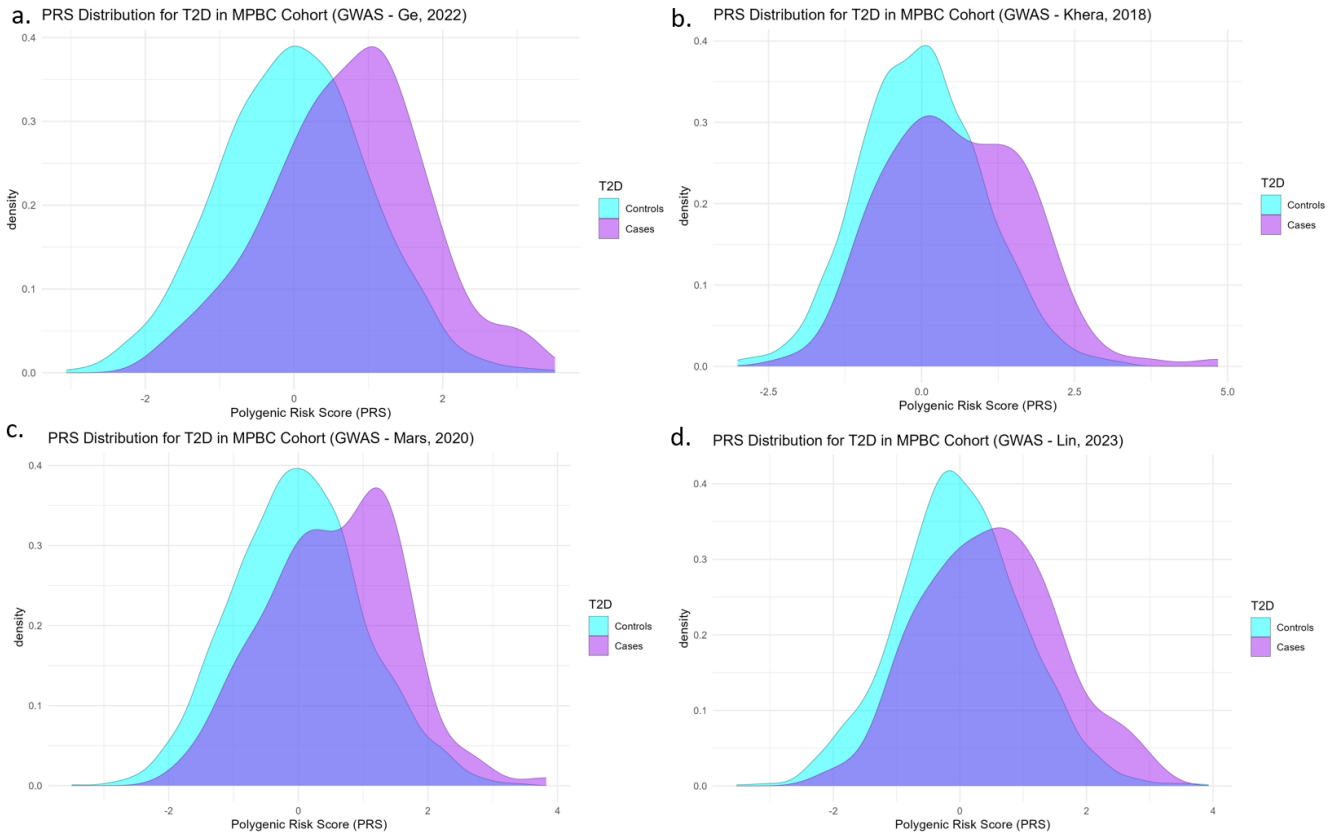
Supplementary Figure 8: Stratification of PD PRS distribution by environmental factors and PD status. *A. Smoking B. Snus usage C. Pesticide exposure D. Caffeine consumption (Ever/Never) E. Caffeine consumption level (High/Low).*



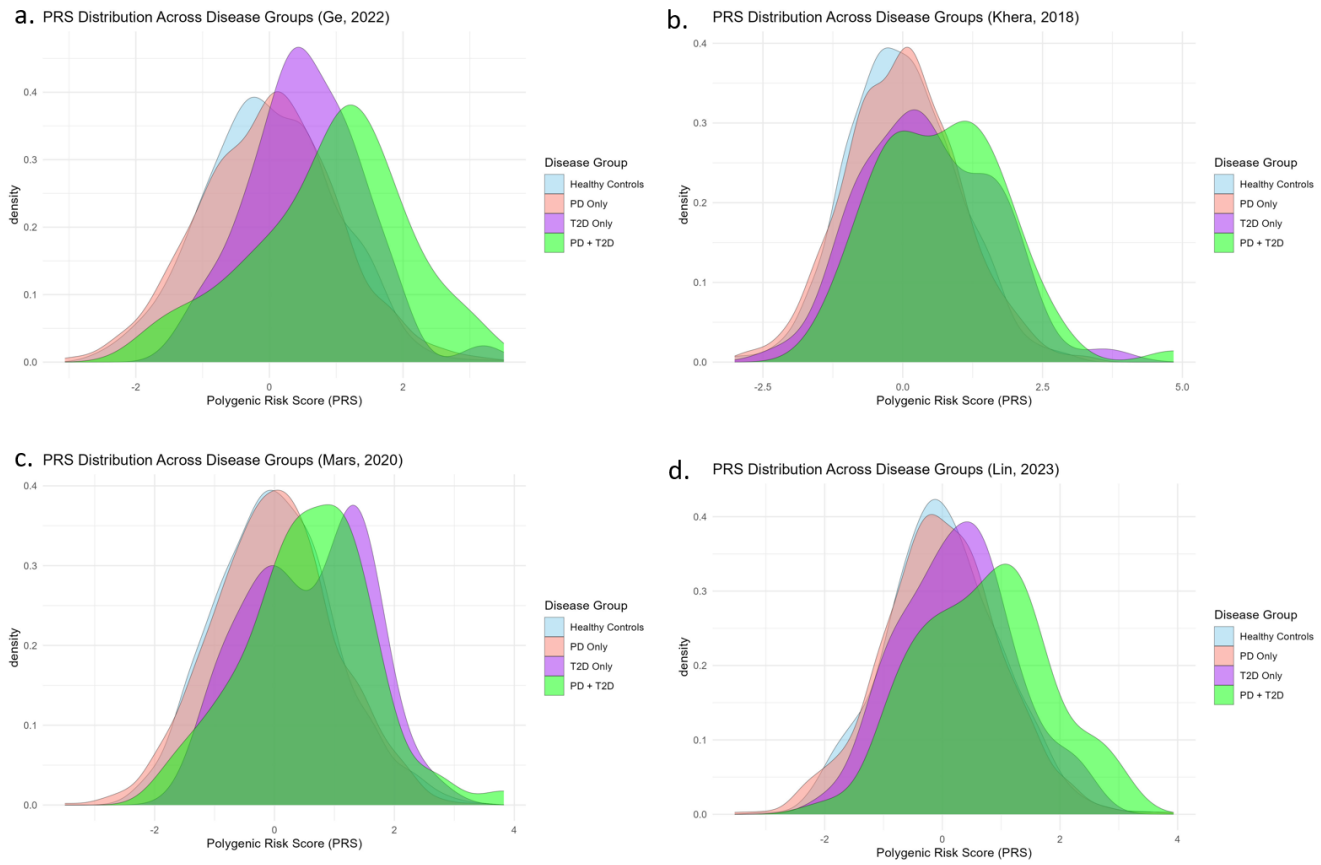
Supplementary Figure 9: Violin and Box Plots of T2D PRS Distributions by Disease Status Using 4 GWAS Summary Statistics. *A. Ge et al (33) B. Khera et al (34) C. Mars et al (32) D. Lin et al (35).*



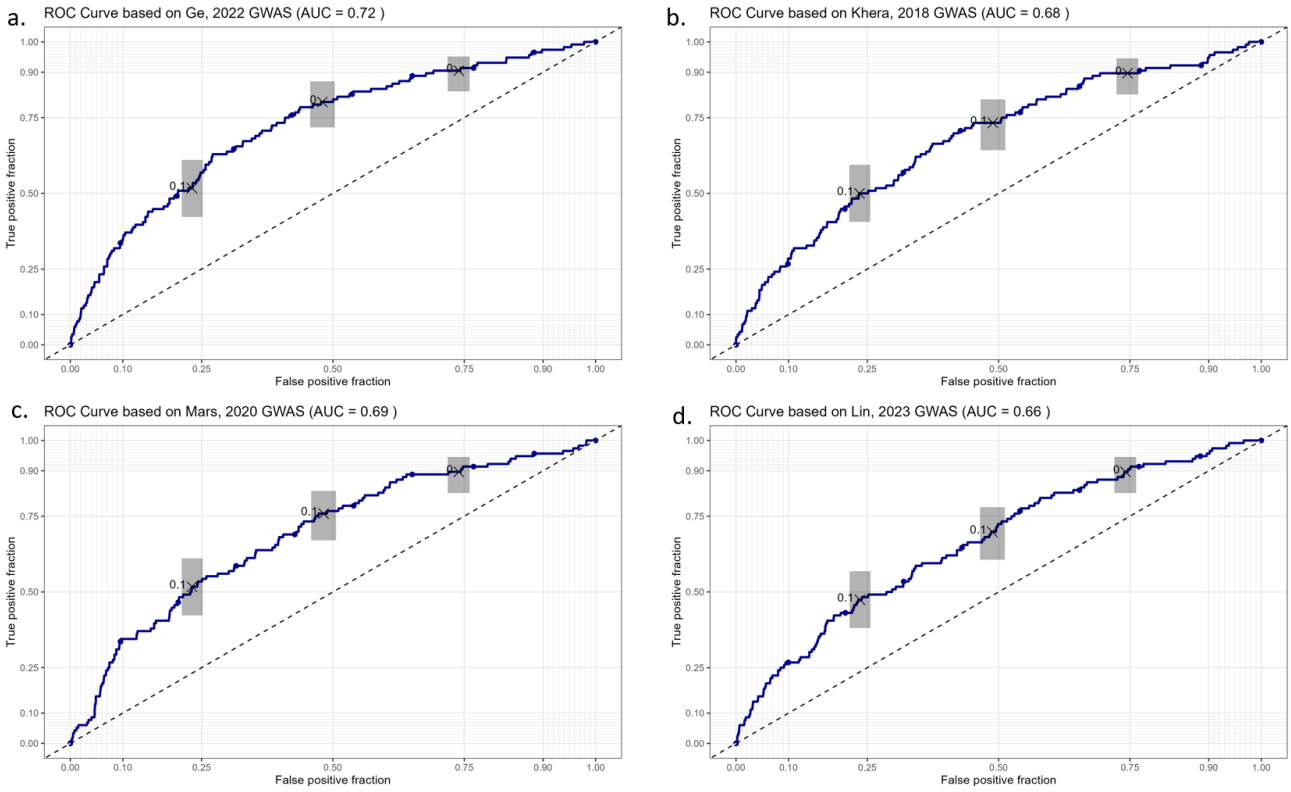
Supplementary Figure 10: Quantile Plots Showing Odds Ratio Estimates Across PRS Quantiles for T2D Risk Using 4 GWAS Summary Statistics. *A. Ge et al (33) B. Khera et al (34) C. Mars et al (32) D. Lin et al (35).*



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 (33) B. Khera et al (34) C. Mars et al (32) D. Lin et al (35).*



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 (33) B. Khera et al (34) C. Mars et al (32) D. Lin et al (35).*



Supplementary Figure 13: ROC Curves for T2D Prediction Using PRS from 4 GWAS Summary Statistics. *A. Ge et al (33) B. Khera et al (34) C. Mars et al (32) D. Lin et al (35).*