

Table S4. Comparison between area under the ROC curve (AUC) comparing the five different prediabetes prediction models with and without three selected metabolites in 143 adolescents from the SOLAR (discovery) cohort and 56 young adults from the CHS (validation) cohort by fitting the models independently in each cohort. The three predictive metabolites (caprylic acid, allylphenol sulfate, and taurocholic acid) were identified using LASSO regression and a stability selection procedure in the SOLAR cohort, and prediction models with and without the identified metabolites were developed independently in the SOLAR and CHS cohorts. In the CHS cohort, only one participant developed elevated HbA1c, so prediction models were not fit for this outcome. The P-value between the models compares the differences between the AUC without the metabolites versus the AUC with the metabolites. GRF: general risk factors only (sex, age, BMI, and in CHS, and in CHS, Hispanic/Latino Ethnicity); SPF: specific risk factors including fasting glucose, two-hour glucose, total cholesterol, and triglycerides.

Clinical Outcome at Follow up	Risk Factors	SOLAR (discovery)			CHS (validation)		
		AUC Without Metabolites	AUC With Three Metabolites	P-value Between Models	AUC Without Metabolites	AUC With Three Metabolites	P-value Between Models
Prediabetes/ type 2 diabetes	GRF	0.59	0.81	7.8×10^{-5}	0.68	0.80	0.079
	GRF + SRF	0.63	0.84	2.6×10^{-5}	0.74	0.88	0.032
	No clinical risk factors	-	0.80	-	-	0.78	-
Impaired fasting glucose (fasting glucose ≥ 100 mg/dL)	GRF	0.71	0.72	0.68	0.90	0.93	0.25
	GRF + SRF	0.88	0.89	0.14	0.97	0.97	0.87
	No clinical risk factors	-	0.67	-	-	0.75	-
Impaired glucose tolerance (Two-hour glucose ≥ 140 mg/dL)	GRF	0.60	0.78	0.018	0.70	0.89	0.036
	GRF + SRF	0.63	0.80	0.012	0.80	0.94	0.083
	No clinical risk factors	-	0.77	-	-	0.81	-
HbA1c $\geq 5.7\%$	GRF	0.67	0.89	0.021	-	-	-
	GRF + SRF	0.71	0.89	0.013	-	-	-
	No clinical risk factors	-	0.85	-	-	-	-