

File S3. Analysis of the DCCT Genetics Study data

Characterization of direct and/or indirect genetic associations for multiple traits in longitudinal studies of disease progression

Myriam Brossard,^{*,1} Andrew D. Paterson,^{†,‡} Osvaldo Espin-Garcia,^{‡,**,§§} Radu V. Craiu,^{**} Shelley B. Bull^{*,‡,1}

^{*}Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, M5T 3L9, Ontario, Canada;

[†]Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, M5G 1X8, Toronto, Ontario, Canada;

[‡]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, M5T 3M7, Toronto, Ontario, Canada;

[§]Department of Biostatistics, Princess Margaret Cancer Centre, M5G 2C1, Toronto, Ontario, Canada;

^{**}Department of Statistical Sciences, University of Toronto, M5S 3G3, Toronto, Ontario, Canada;

^{§§}Department of Epidemiology and Biostatistics, Western University, N6A 5C1, London, Ontario, Canada.

¹**Corresponding authors:** Lunenfeld-Tanenbaum Research Institute, 60 Murray Street, Box #18, M5T 3L9, Toronto, Ontario, Canada. E-mails: bull@lunenfeld.ca, brossard@lunenfeld.ca. Phone number: +1 416-586-5052.

[Table of Contents](#)

1. Time-to-event traits analyzed	2
2. Selection of the 307 candidate SNPs from the literature	2
3. Analysis of the DCCT data	3
4. Association structures for HbA1c on T1DC traits	3
5. Joint model diagnostics and sensitivity analyses to modeling assumptions	8
6. Sample size/power analyses based on parametric resampling	19
Cited Literature	30

1. Time-to-event traits analyzed

We use time to mild DR and time to persistent microalbuminuria, for DR and DN outcomes respectively, as previously defined in the motivating GWAS of HbA1c (Paterson et al. 2010) and summarized in **Table S1-1**.

Table S1-1 Time-to-event traits analyzed in the DCCT Genetics Study

Time-to-event outcome	Name	Outcome definition
DR	Time to mild DR	Time from DCCT baseline to mild non-proliferative diabetic retinopathy (EDTRS step 4, patient level 35/<35)
DN	Time to persistent microalbuminuria	Time from DCCT baseline to the first of two consecutive visits with Albumin Excretion Rate >30 mg/day

Out of the 667 DCCT individuals, we analyze $N=516$ subjects with genotype data, without mild retinopathy at DCCT baseline (or prior) or without DN event at DCCT baseline. By the time of the DCCT close-out visit, 297 (57.6%) experienced a DR event, 61 (11.8%) a DN event, including 47 subjects (9.1%) that experienced both events.

2. Selection of the 307 candidate SNPs from the literature

We identified 322 independent SNPs ($r^2 < 0.8$, $MAF \geq 5\%$ in European ancestry individuals) associated with HbA1c, SBP, DR and/or DN outcomes as reported in the literature (Paterson et al. 2010; Grassi et al. 2011; Sandholm et al. 2012; Hosseini et al. 2015; Wheeler et al. 2017; Evangelou et al. 2018; Pollack et al. 2019). For HbA1c, DR or DN we selected SNPs reported at the suggestive significance level of $P^* = 10^{-6}$ by GWAS in T1D individuals or SNP associations reported by GWAS conducted in T2D individuals or general populations and confirmed in T1D individuals of European ancestry at the nominal significance level (Paterson et al. 2010; Grassi et al. 2011; Sandholm et al. 2012; Hosseini et al. 2015; Wheeler et al. 2017; Evangelou et al. 2018; Pollack et al. 2019). For the SBP SNP list, since large-scale GWAS results of SBP in individuals with diabetes were lacking at the time of our analyses, we selected SNPs reported in the largest meta-analysis conducted in the general population of European ancestry at the conventional genome-wide significance level $P^* = 5 \times 10^{-8}$. In total, we analyzed in DCCT 307 biallelic SNPs with

imputation quality score $R^2 \geq 0.50$ and $MAF \geq 5\%$, and after pruning on linkage disequilibrium ($r^2 < 0.8$) using LDlink (Machiela and Chanock 2015) in 1000 Genomes phase 3 European-ancestry population (See **File S4** for a full list of the SNPs and their marginal association results with the discovery trait and with the other investigated traits in $N = 516$ DCCT individual).

3. Analysis of the DCCT data

We fit the joint model (JM) for each SNP, one at a time, including baseline covariates (age at diagnosis, T1D duration, cohort, sex, and year of entry in the DCCT study). For the covariate year of entry in DCCT, we group the patients into four consecutive strata with homogeneous number of individuals (*i.e.* 1983-1984, 1985-1986, 1987, using 1988-1989 as the reference category). To account for non-linear trends observed in HbA1c measures, we include an indicator variable in the model that captures short-term effects between DCCT entry and 3-month visit. In Stage 1 of JM, bivariate longitudinal mixed-effects models for HbA1c and SBP are fitted using all available measures at quarterly visits from DCCT baseline to the close-out visit; HbA1c and SBP trajectories are fitted for each individual. In Stage 2 of JM, trajectory values interpolated to the start time of each risk interval are then used as time-varying covariates in the Cox PH frailty model. Time-to-event sub-models are fitted using annual records of DR and DN events. Because the DR and DN outcomes were assessed with different frequency of visits in DCCT (semi-annually for DR and annually for DN), we assign each DR event to the one-year interval visit that include the observed time-to-event. We use $B = 500$ bootstraps of DCCT individuals to compute empirical variance-covariance matrices for parameters estimated by the joint model.

4. Association structures for HbA1c on T1DC traits

Given established cumulative effects of HbA1c on T1DC traits (Lind et al. 1995; Lind et al. 2010), we compare joint model results obtained with contemporaneous HbA1c value to joint model results obtained using time-weighted cumulative and updated cumulative mean HbA1c effects on T1DC (Lind et al. 1995; Lind et al. 2010). Under each cumulative association profile, the time-to-event sub-model in Stage 2 of JM is fitted by substituting the fitted trajectory of the HbA1c by a summary function of the prior fitted values from DCCT baseline up to the beginning of each risk interval (**Table S4-1**). While the updated mean association structure assumes an equal weighting for all fitted HbA1c values at prior visits from baseline, the time-weighted cumulative HbA1c

effect association structure assumes different weights for each visit. Here, we use a time-weighted formulation that considers all HbA1c values from DCCT baseline up to 5 years prior to the start of each risk interval following previous DCCT data analysis (Lind et al. 2010). We extracted weights from (Lind et al. 2010), as presented in **Fig. S4-1**, and recalibrated them for each risk interval such that the sum of the weights is equal to 1 (**Table S4-1**). Joint model results fitted with each alternative association structure for HbA1c on T1DC traits are presented in **Table S4-2**. In the main paper, we present results under the time-weighted cumulative association structure which exhibits the stronger prior association with T1DC and in the DCCT individuals analyzed here (**Table S4-2**); but we obtain similar results for tests of SNP effects using the two alternative association structures (**Fig. S4-2**).

Table S4-1. Three alternative association structures used to account for time-dependent HbA1c effects on T1DC traits in DCCT

Parametrization	Time-dependent association structure in the two-stage approach
Contemporaneous (current value)	$f_{1,k}(\widehat{y}_{1,t}^*(t_{i,j})) = \widehat{y}_{1,t}^*(t_{i,j})$
Updated cumulative mean	$f_{1,k}(\widehat{y}_{1,t}^*(t_{i,j})) = \frac{1}{j} \sum_{s=1}^j \widehat{y}_{1,t}^*(t_{i,s})$, with $s \leq j$
Time-weighted cumulative	$f_{1,k}(\widehat{y}_{1,t}^*(t_{i,j})) = \begin{cases} \sum_{s=1}^j v_k \widehat{y}_{1,t}^*(t_{i,j-s}) & \text{if } t_{i,j} < 5 \\ \sum_{s=1}^S v_k \widehat{y}_{1,t}^*(t_{i,j-s}) & \text{if } t_{i,j} \geq 5 \end{cases}$ <p>Here, we use $S=5$ to account for the fitted HbA1c values up to 5 years prior to the current time point $t_{i,j}$, with weights v_k based on (Lind et al. 2010), see Fig. S4-1.</p>

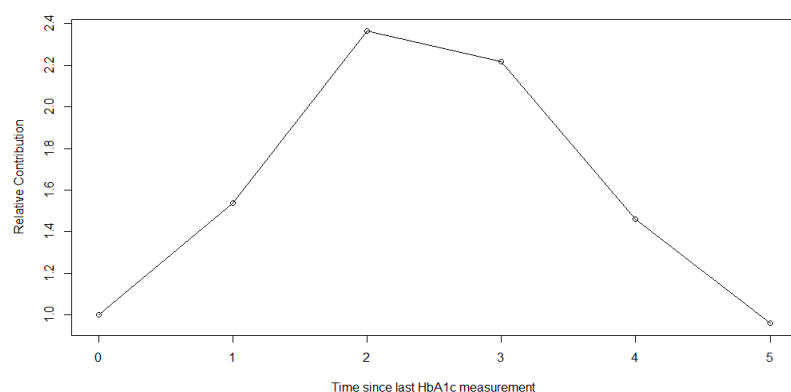


Fig. S4-1. Relative contribution of past HbA1c measures to the risk of mild retinopathy at current time point, based on (Lind et al. 2010). The X axis represents the time (in year) since the current HbA1c measurement.

Table S4-2. Results from the joint model fitted in DCCT subjects with alternative association structures for HbA1c and without genetic variable Each sub-model is adjusted for sex, age at diagnosis, T1D duration and year of entry in DCCT. Year of entry in DCCT is treated in the models as indicator variables for each category 1983-1984 ($N = 86$), 1985-1986 ($N = 131$), 1987 ($N = 127$) and 1988-1989 ($N = 171$). The category 1988-1989 was used as reference category. t0t1hba_1 is the indicator variable used to account for the non-linear trend of HbA1c between the baseline and the 3 month-visit.

(see table on next page)

Longitudinal sub-model				
	HbA1c (<i>l</i>=1)		SBP (<i>l</i>=2)	
	Effect	P	Effect	P
Intercept	9.51		111.28	
t0t1hba_1	-0.24	7.3E-06		
visits (years)	0.04	5.8E-03	0.21	7.5E-04
T1D duration at baseline (years)	-0.05	3.9E-02	0.21	0.14
Age at baseline (years)	-0.02	3.3E-02	0.15	2.2E-03
Cohort	-0.16	0.32	1.12	0.27
Gender (Female)	0.11	0.35	-6.39	4.4E-22
Year of entry in DCCT				
1983-1984	0.22	0.18	0.06	0.95
1985-1986	-0.02	0.86	-0.3	0.73
1987-1988	0.18	0.27	-1.05	0.2
Residual variances	0.72		69.43	

Time-to-event sub-model		Association structures of HbA1c with Time-to-DR (<i>k</i>=1)					
		Contemporaneous		Updated cumulative mean		Time-weighted cumulative	
		log HR	P	log HR	P	log HR	P
		0.44	1.0E-15	0.52	2.8E-18	0.53	1.3E-18
HbA1c Trajectory							
T1D duration at baseline (years)		0.19	3.8E-10	0.20	6.6E-10	0.20	7.7E-10
Age at baseline (years)		0.02	0.11	0.02	4.1E-02	0.02	3.8E-02
Cohort		1.10	2.0E-07	1.15	7.1E-08	1.16	4.0E-08
Sex (Female)		-0.37	1.2E-02	-0.40	8.4E-03	-0.39	9.2E-03
Year of entry in DCCT							
1983-1984		-0.23	0.31	-0.29	0.23	-0.30	0.21
1985-1986		-0.14	0.50	-0.20	0.35	-0.21	0.32
1987-1988		-0.61	4.9E-03	-0.67	2.4E-03	-0.69	1.9E-03
		Association structures of HbA1c with Time-to-DN (<i>k</i>=2)					
		Contemporaneous		Updated cumulative mean		Time-weighted cumulative	
		log HR	P	log HR	P	log HR	P
		0.46	1.6E-05	0.55	1.2E-06	0.54	8.2E-07
HbA1c Trajectory							
SBP Trajectory		0.07	1.0E-04	0.07	1.0E-04	0.07	1.0E-04
T1D duration at baseline (years)		0.07	0.20	0.07	0.17	0.08	0.16
Age at baseline (years)		-0.05	2.6E-02	-0.05	0.05	-0.05	4.8E-02
Cohort		0.22	0.59	0.25	0.54	0.25	0.55
Sex (Female)		-0.20	0.54	-0.23	0.47	-0.23	0.47
Year of entry in DCCT							
1983-1984		0.54	0.25	0.51	0.27	0.50	0.28
1985-1986		0.04	0.94	-0.01	0.99	-0.02	0.97
1987-1988		0.01	0.98	-0.05	0.92	-0.06	0.91

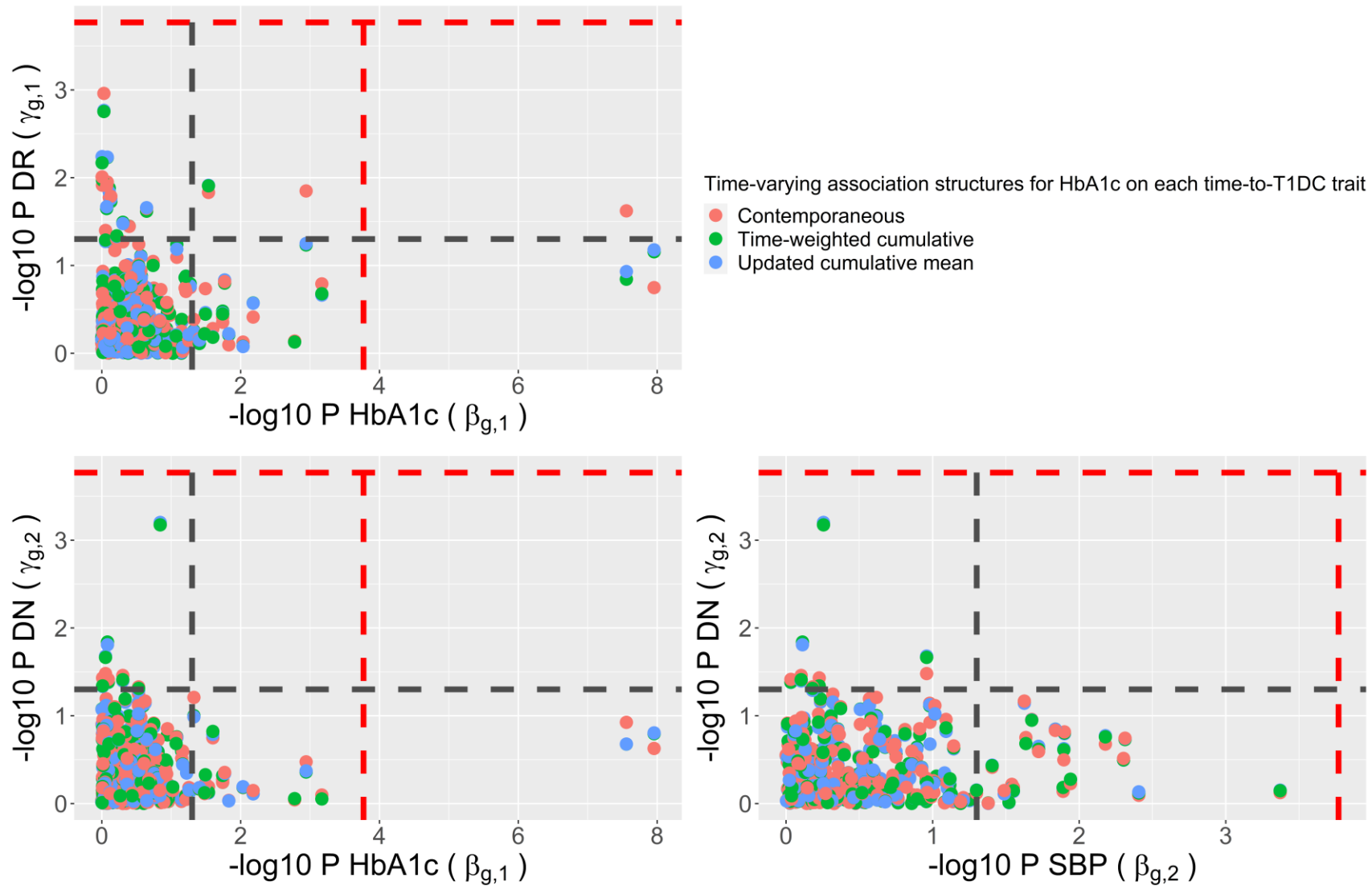


Fig. S4-2. Classification of direct and/or indirect SNP association based on the joint model fitted for each of the 307 candidate SNPs tested in $N = 516$ DCCT subjects using the contemporaneous association structure and each of the two alternative cumulative association structures from [Table S4-1](#) (updated cumulative mean and time-weighted cumulative) for HbA1c effects on time-to-T1DC traits. Scatter plots represent the P -values ($-\log_{10}$) for tests of $\beta_{g,l}$ (X axis) and $\gamma_{g,k}$ (Y axis) for HbA1c/DR, HbA1c/DN and SBP/DN trait pairs respectively. Significance levels $P^* = 1.7 \times 10^{-4}$ and $P^* = 0.05$ are indicated by red and dark grey lines.

5. Joint model diagnostics and sensitivity analyses to modeling assumptions

Because the joint model integrates longitudinal and time-to-event sub-models, model misspecification can occur in multiple ways and can lead to invalid classification of direct and/or indirect genetic associations. For example, (Arisido et al. 2019) show by simulation studies that misspecification of the shape of the longitudinal trajectory, or the hazard function, can severely bias the QT effect estimate on the time-to-event trait in a shared-random effect joint model; however misspecification of the normal assumption for the random effects appears to have a negligible effect, particularly as the study sample size increases.

As described by (Rizopoulos 2012), joint model diagnostics can be based on standard approaches for residual analyses from the longitudinal and the time-to-event sub-models. Here, we apply diagnostics to the joint model of HbA1c, SBP, DR and DN fitted in DCCT individuals with the Two-Stage approach (see **section 3** for details). To illustrate, we present diagnostic analyses for rs1358030 and for the time-weighted cumulative association structure for HbA1c effects on both T1DC traits (we obtained similar conclusions under the two alternative association structures described in **Table S4-1**). We summarize the conclusions from the residual analysis applied to the longitudinal sub-model and time-to-event components of the joint model as well as from sensitivity analyses of the classification results to model assumptions in **Table S5-1**.

Table S5-1. Summary of the joint model diagnostic results in DCCT

Assumptions	Diagnosis tools	Joint model diagnosis results in DCCT ¹	
		Results	Main conclusions
Longitudinal sub-model for both HbA1c and SBP QTs: Bivariate longitudinal mixed-effects models for HbA1c and SBP			
Homoscedasticity	Plot of the standardized subject-specific residuals (conditional residuals) of each QT versus the corresponding subject-specific fitted QT values	<u>Fig. S5-1</u> (Panels A & C)	No deviation from the homoscedasticity assumption for both QTs.
Normality	Q-Q plot of the conditional residuals	<u>Fig. S5-1</u> (Panels B & D) <u>Fig. S5-2</u>	No systematic deviation from the expected normal distribution for both QTs, except for the subjects in the tails of the Q-Q plot for HbA1c. Assessment of conditional residuals for a random selection of subjects showing deviation from normality assumption on the Q-Q plot could be explained by poor model fit at some visits due to high within-subject HbA1c variability.
Specified mean structures for each QT	Scatter plot of the standardized marginal residuals (population averaged) against the subject-specific fitted QT values.	<u>Fig. S5-3</u> (Panels A & C)	Null horizontal trends observed for the loess curve of the scatterplot for each QT which does not indicate misspecification of the mean structures in the longitudinal sub-model.
Linearity	Scatter plot of the marginal residuals versus the visit times for each QT	<u>Fig. S5-3</u> (Panels B & D)	No evidence for non-linear time effects on each QT trajectory.
Time-to-event sub-model for DR and DNs: Cox PH frailty time-to-event sub-model			
Functional form of the longitudinal QT trajectory effect on each T1DC trait	Scatter plots of the martingale residuals from the Cox PH frailty model for both T1DC traits fitted alternatively without the QT being assessed	<u>Fig. S5-4</u>	No systematic deviation of the loess curves of each fitted QT trait against the martingale residuals from the horizontal line expected under the linear assumption assumed for each QT effect on each T1DC trait.
Frailty distribution	Sensitivity analysis to alternative specification of frailty distributions	<u>Table S5-2</u>	No change of the conclusions when the Gaussian distribution is assumed for the frailty term (instead of the Gamma distribution) in the Cox PH frailty sub-model.
PH	PH test based on the formal score test for the null slope of the time-dependent covariate effect on the time-to-event	<u>Table S5-3</u>	No indication for global deviation from the PH assumption, but four covariates show evidence of significant time-dependent effect on time-to-DR at the

¹For the joint model fitted in DCCT with the two-stage approach (see section 3 for details) for rs1358030 and time-weighted cumulative effects of HbA1c on both time-to-T1DC outcomes.

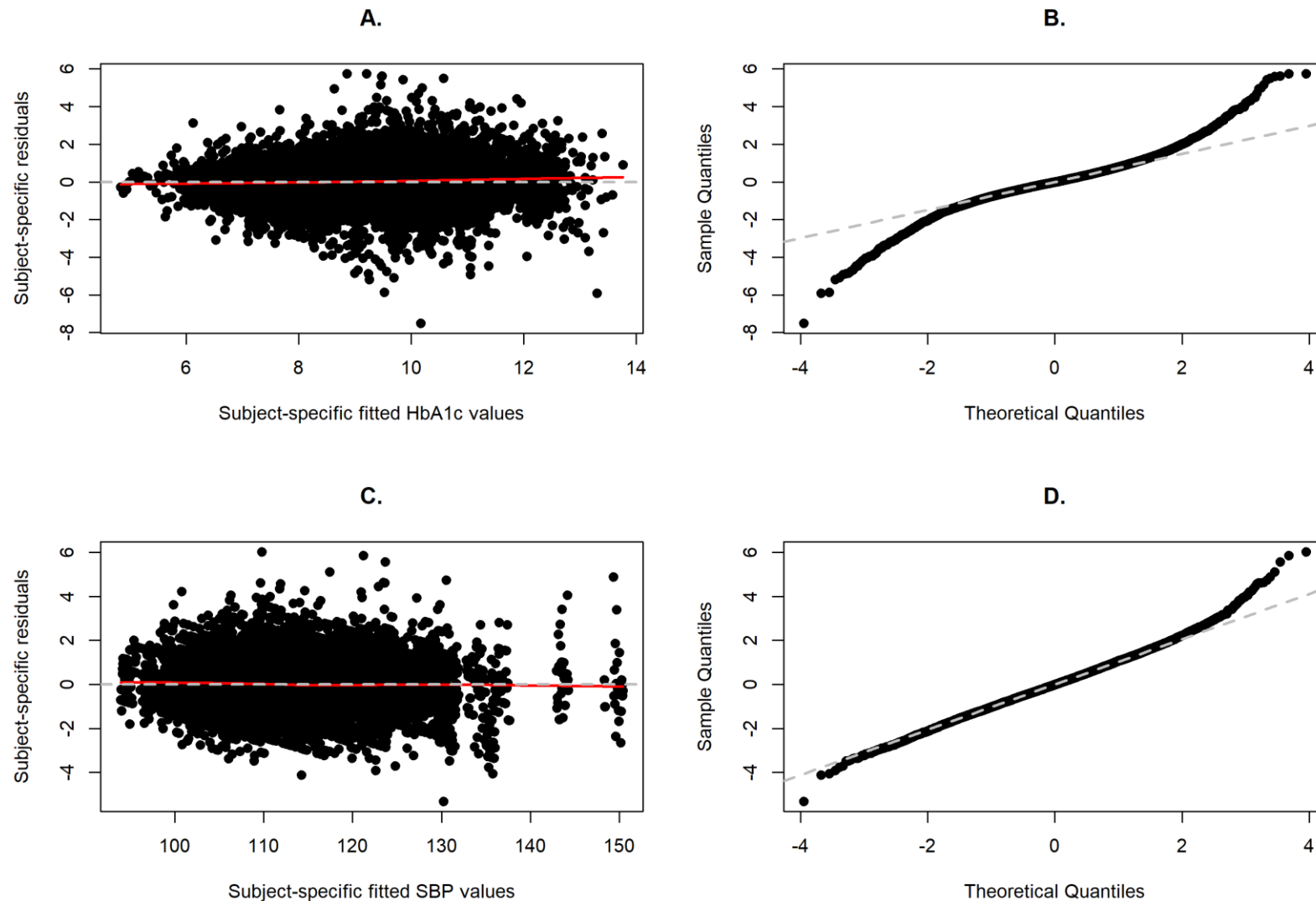


Fig. S5-1. Diagnostic analysis for the homoscedasticity and normality assumptions based on the standardized subject-specific residuals extracted for HbA1c (panels A-B) and SBP (panels C-D) from the bivariate mixed model fitted for HbA1c and SBP fitted at Stage2. Panels A and D correspond to the scatterplots of the standardized subject-specific residuals for each QT versus the subject-specific fitted QT values; Panels C and D represent the Q-Q plots of the standardized subject-specific residuals. The loess curves of the scatterplots on panels A and C are shown in red.

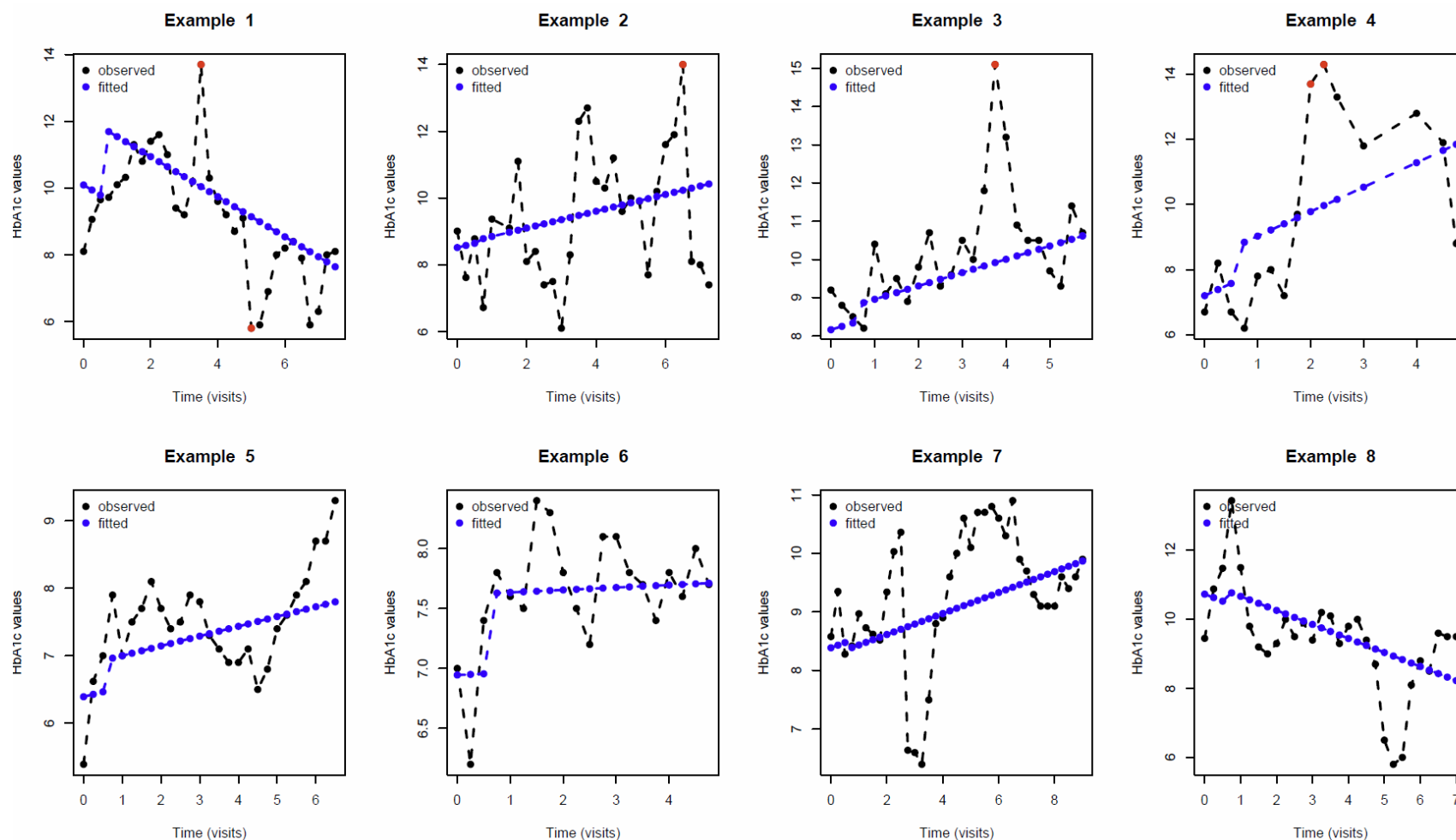


Fig. S5-2. Examples of *observed* and *fitted* HbA1c trajectories values for eight individuals randomly chosen from the analyzed DCCT individuals. Examples 1-4 are for individuals selected among those exhibiting a standardized subject-specific residual larger than 4 (in absolute value) in the Stage 1 longitudinal sub-model for at least one visit; values with standardized subject-specific residual larger than 4 (in absolute value) are shown in red. Examples 5-8 show in comparison HbA1c values for individuals selected among those that do not exhibit a large, standardized subject-specific residual (absolute value < 2).

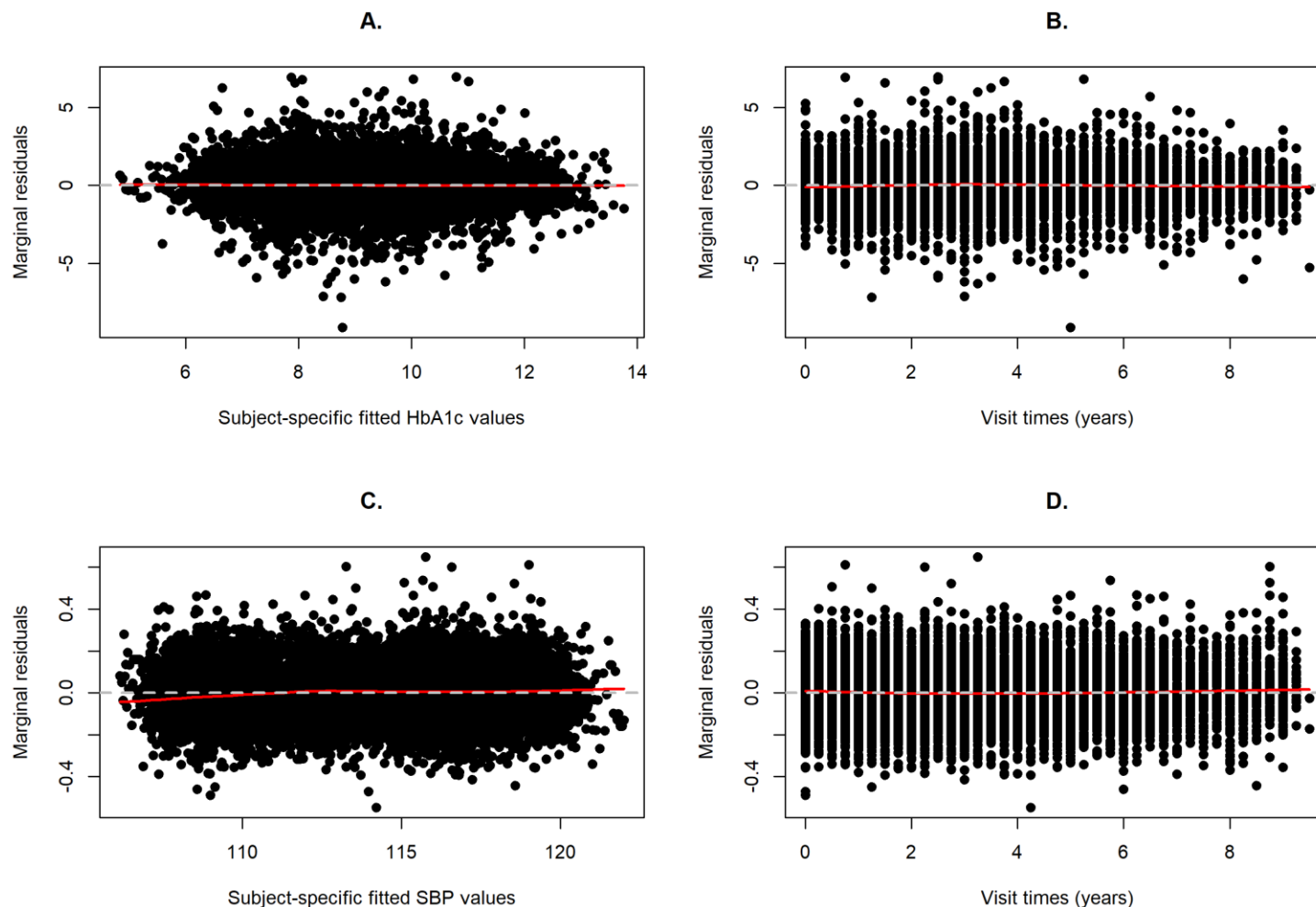


Fig. S5-3. Diagnostic analysis based on the marginal residuals (population averaged) extracted for HbA1c (panels A-B) and SBP (panels C-D) from the bivariate mixed model fitted for HbA1c and SBP at Stage2. Panels A and C correspond to the scatterplots of the marginal residuals for each QT versus the subject-specific fitted QT values; Panels B and D represent the scatterplots of the marginal residuals against the visit times. The loess curves of the scatterplots on panels A-D are shown in red. These plots do not indicate any evidence of misspecification of the design matrix for the fixed effects for both QTs (A, C); and no deviation for the linear trajectory for each QT in time (B, D).

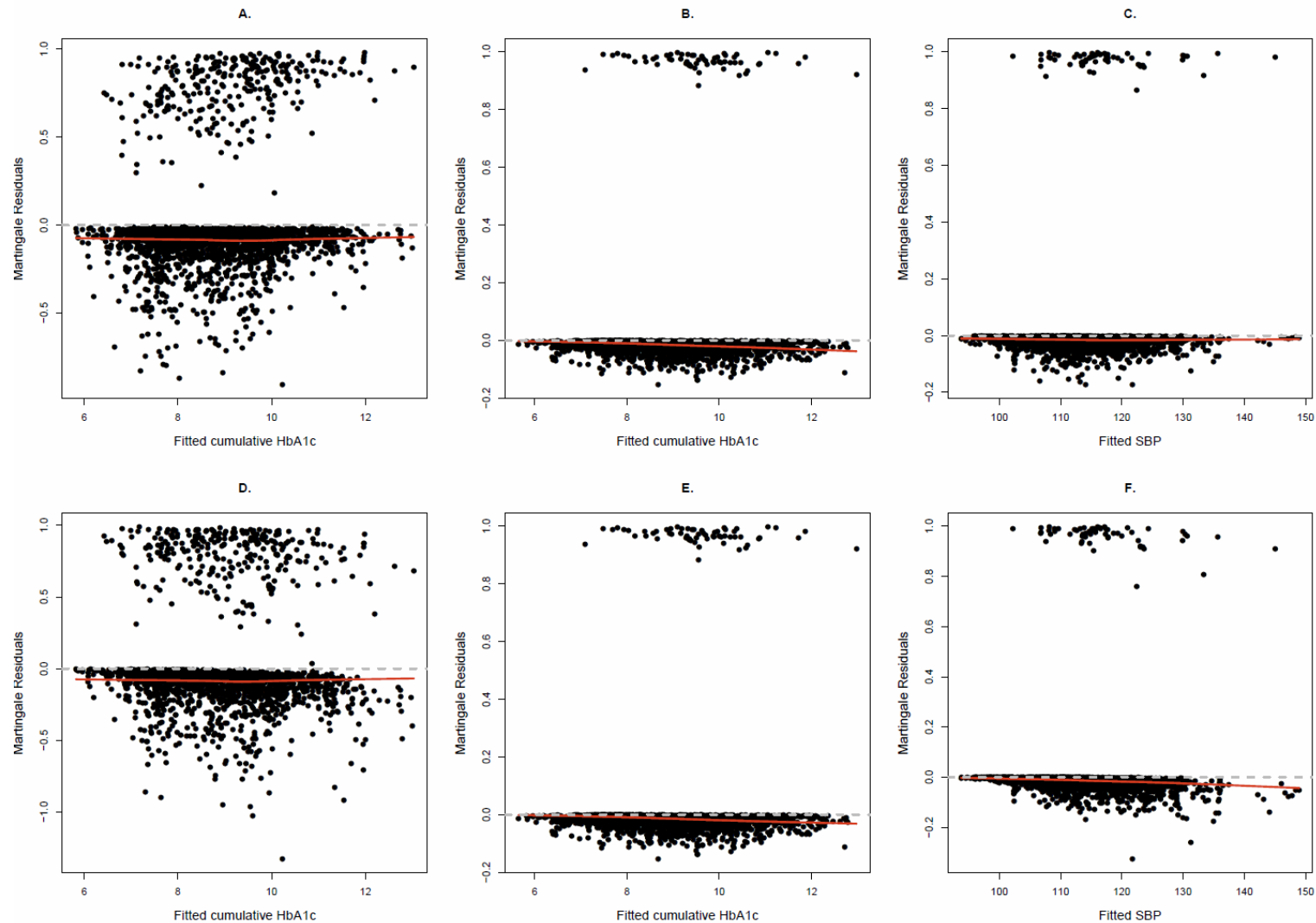


Fig. S5-4. Functional form of QT effects on each T1DC trait. Panels A-C show the martingale results from the Stage 2 time-to-event sub-model when each longitudinal QT is excluded from the model as suggested by (Therneau and Grambsch 2000) to display form of the QT effect on the time-to-event traits. Panels D-F, show the martingale results from the Stage 2 time-to-event sub-model when the QTs are included in the Cox frailty models as used by (Rizopoulos 2012) to visually inspect the functional form of the QT on the time-to-event is correct. Panels A and D show the martingale residuals for the DR outcome and Panels B, C, E and F show the martingale residuals for the DN outcome.

	Cox PH frailty model with <i>Gamma distributed frailty</i> (as used in the paper)		Cox PH frailty model with <i>Gaussian distributed frailty</i> (for comparisons)	
	Log HR	P-value	Log HR	P-value
Time-to-DR				
SNP (rs1358030)	-0.2	7.0E-02	-0.19	8.2E-02
HbA1c trajectory (Time-weighted cumulative)	0.55	1.0E-18	0.54	8.3E-19
Time-to-DN				
SNP (rs1358030)	-0.31	1.6E-01	-0.30	1.7E-01
HbA1c trajectory (Time-weighted cumulative)	0.58	2.6E-07	0.58	2.9E-07
SBP trajectory (contemporaneous)	0.07	1.2E-04	0.07	1.3E-04

Table S5-2. Comparisons of the results for rs1358030 and QT effects on each T1DC trait obtained from the Cox PH frailty time-to-event sub-model fitted at Stage 2 assuming either a Gamma or a Gaussian distribution for the frailty term. For each coefficient, the *P*-values are obtained using a 1df Wald test based on the variances estimated by 500 bootstraps. The results for the adjusting baseline covariates were similar between the two models but are not shown here to simplify the table.

	Cox frailty model as presented in the paper (<i>adjusted</i> for the cohort covariate)	Cox frailty model with baseline hazard <i>stratified</i> on the cohort variable
GLOBAL	9.9E-01	8.5E-01
Time-to-DR		
SNP (rs1358030)	2.4E-02	6.9E-02
HbA1c trajectory (Time-weighted cumulative)	9.9E-03	2.1E-01
T1D duration at baseline (years)	3.7E-02	5.1E-01
Age at baseline (years)	2.5E-01	2.3E-01
Sex (Female)	3.7E-01	7.6E-01
Year of entry in DCCT (1983-1984)	5.8E-01	4.9E-01
Year of entry in DCCT (1985-1986)	7.3E-01	7.1E01
Year of entry in DCCT (1987-1988)	4.7E-01	5.0E-01
Cohort	7.2E-03	NA
Time-to-DN		
SNP (rs1358030)	2.0E-01	1.7E-01
HbA1c trajectory (Time-weighted cumulative)	2.3E-01	1.6E-01
SBP trajectory (Contemporaneous)	8.4E-01	8.9E-01
T1D duration at baseline (years)	8.7E-01	4.2E-01
Age at baseline (years)	7.5E-01	8.9E-01
Sex (Female)	7.5E-01	6.2E-01
Year of entry in DCCT (1983-1984)	2.7E-01	2.2E-01
Year of entry in DCCT (1985-1986)	3.1E-01	2.7E-01
Year of entry in DCCT (1987-1988)	9.6E-01	9.7E-01
Cohort	4.8E-01	.

Table S5-3. *P*-values for assessment of the PH assumption both globally and for each covariate entered in the Cox PH frailty time-to-event sub-model fitted at Stage 2. The PH assumption was tested using `cox.zph()` from the R “survival” package. This function returns the results of a formal score test for the null slope of the time-dependent covariate effect on the time-to-event outcome as described in (Grambsch and Therneau 1994). The frailty term is treated as a fixed offset in the model and is not tested for PH assumption.

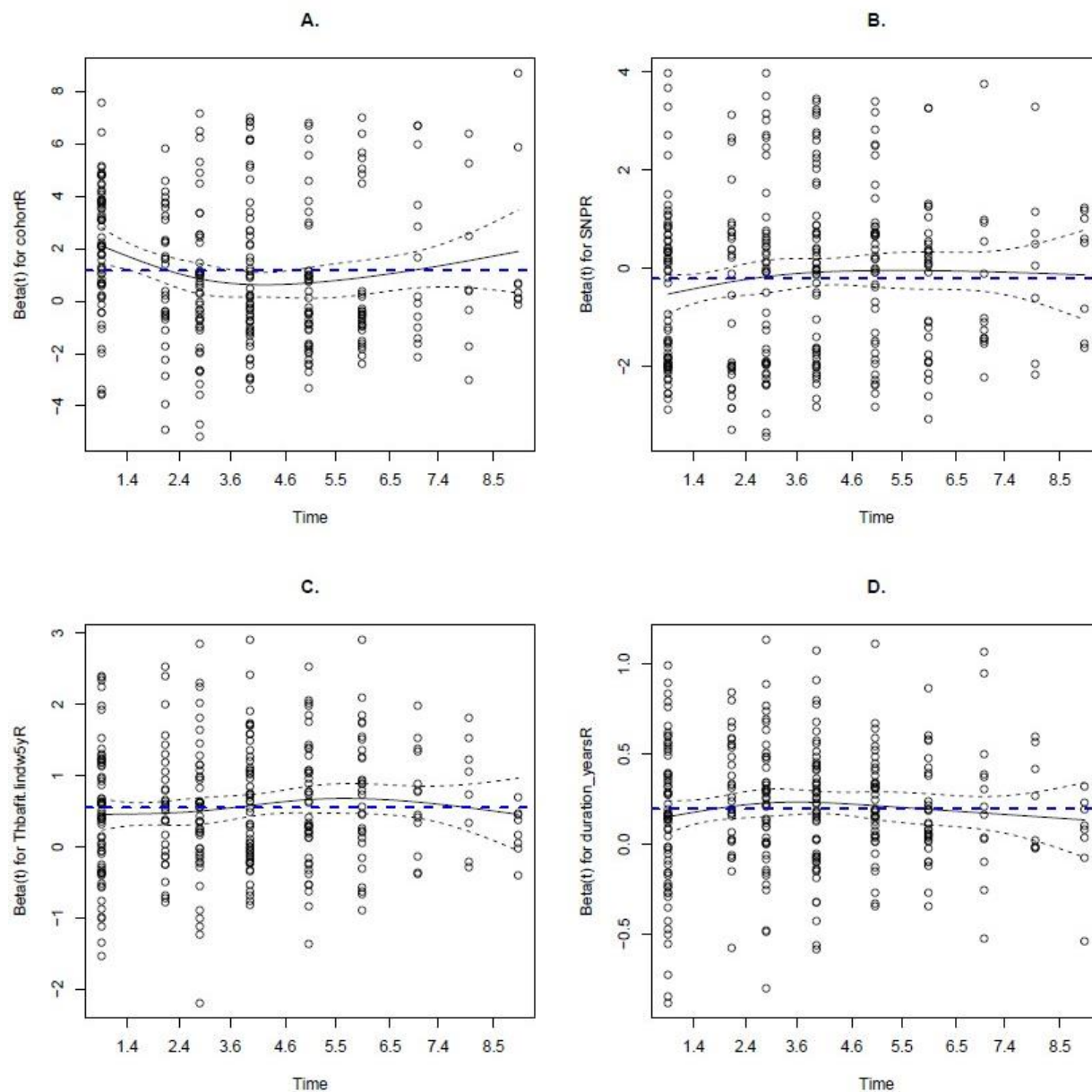


Fig. S5-5. Schoenfeld residuals for the time-dependent covariate effects on time-to-DR for each of the four covariates that violates the PH assumption at the nominal significance level ($P < 0.05$ in [Table S5-2](#)) in the Cox PH frailty time-to-event sub-model fitted at Stage 2. Panels A. cohort, B. rs1358030, C. Time-weighted cumulative effects of HbA1c, D. T1D duration (years). The blue dashed lines show the covariate effect estimated by the Cox PH frailty time-to-event sub-model. These plots were produced using the output from `cox.zph()` from the R “survival” package.

	Cox frailty model as presented in the paper (<i>including</i> cohort as a covariate)		<i>Stratified</i> Cox frailty model for the cohort covariate	
	Log HR	P-value	Log HR	P-value
Time-to-DR				
SNP (rs1358030)	-0.20	7.0E-02	-0.14	1.6E-01
HbA1c trajectory (Time-weighted cumulative)	0.55	1.0E-18	0.47	1.0E-15
Time-to-DN				
SNP (rs1358030)	-0.31	1.6E-01	-0.29	1.7E-01
HbA1c trajectory (Time-weighted cumulative)	0.58	2.6E-07	0.55	4.4E-07
SBP trajectory (Contemporaneous)	0.07	1.2E-04	0.07	4.3E-05

Table S5-4. Results from sensitivity analysis for rs1358030 and QT effects on each T1DC trait using the Cox PH frailty time-to-event sub-model fitted at Stage2 and when the cohort variable introduced as a stratification factor in the baseline hazard. For each coefficient, *P*-values are obtained using a 1df Wald test based on the variances estimated by 500 bootstraps of individuals. The adjusting covariates are not presented but were added in the models.

	Cox frailty model as presented in the paper (<i>including</i> cohort as a covariate)		<i>Stratified</i> Cox frailty model for the cohort covariate	
	Log HR	P-value	Log HR	P-value
Time-to-DR				
SNP (rs10810632)	0.3	1.4E-01	0.28	1.4E-01
HbA1c trajectory (Time-weighted cumulative)	0.48	5.0E-16	0.43	8.2E-15
Time-to-DN				
SNP (rs10810632)	0.5	2.1E-01	0.47	2.3E-01
HbA1c trajectory (Time-weighted cumulative)	0.5	6.0E-06	0.48	9.2E-06
SBP trajectory (Contemporaneous)	0.07	6.5E-05	0.07	2.2E-05

Table S5-5. Results from sensitivity analysis for rs10810632 and QT effects on each T1DC trait using the Cox PH frailty time-to-event sub-model fitted at Stage2 and when the cohort variable introduced as a stratification factor in the baseline hazard. For each coefficient, the *P*-values are obtained using a 1df Wald test based on the variances estimated by 500 bootstraps of individuals. The adjusting covariates are not presented but were added in the models.

6. Sample size/power analyses based on parametric resampling

We assessed the dependence on study sample size of the classification results for associations of rs1358030 and rs10810632 detected in DCCT, using five datasets with increasing sample size drawn from DCCT individuals using a parametric resampling approach under a specified alternative hypothesis (which maintains the relationships between the SNP, the traits and the covariates as observed in DCCT). Parametric resampling approaches have been used for example in clinical trials to estimate sample size requirements (Walters and Campbell 2005) or assess statistical power (Kleinman and Huang 2017) under a specified alternative hypothesis, and in genetic association studies to construct an empirical null distribution for complex joint gene-gene interaction testing (Chen et al. 2007) or for gene-set association testing of gene-environment interactions (Coombes and Biernacka 2018).

Briefly, we *resample* the individual covariate vectors (including the SNPs, longitudinal QTs and baseline covariates) observed in DCCT to generate five datasets with sample sizes up to 5 times the DCCT dataset (from $N=516$ to $N=2580$ individuals), and we *simulate* the time-to-T1DC traits based on the joint model parameters estimated in DCCT (parametric part).

Resampled datasets with increasing sample size for each SNP

1. **Resampling of the longitudinal data:** We combine successively five duplicates of $N=516$ DCCT individuals of observed longitudinal and baseline covariates (including the SNP), that is: Data 1: 1 duplicate, Data 2: 2 duplicates, Data 3: 3 duplicates, Data 4: 4 duplicates, Data 5: 5 duplicates.
2. **Simulation of the time-to-event traits for Data 1 to Data 5:**
 - a. Simulation of five data replicates of $N=516$ individuals of time-to-T1DC traits (DR, DN). For each replicate and each DCCT individual, we generate each k^{th} time-to-event trait similarly as presented in **Fig. 4** and described in **Step3** (File S2, section 3):
 - i. We generate the *uncensored* $T_{i,k}^*$ for each k^{th} time-to-event outcome ($k=1$ for DR and $k=2$ for DN) by calculating the inverse of the cumulative hazard function for each time-to-event trait k with parameters set to the estimated

values from the joint model time-to-event sub-model fitted in DCCT (see section 3 for details about the joint model fitted in DCCT). Here, the unexplained dependency between the simulated time-to-T1DC traits is induced by the fitted subject-specific frailty term.

- ii. To generate event rates as observed in DCCT (*ie* 57.6% DR events and 11.8% DN events), constant across the five replicates, we define for each k^{th} time-to-event trait the **censoring** time $C_{i,k}$ such that $P(T_{i,1}^* \leq C_{i,1}) = 0.576$ for DR and $P(T_{i,2}^* \leq C_{i,2}) = 0.118$ for DN. We attribute an event ($\delta_{i,k} = 1$) to all the subjects with $T_{i,k}^* \leq C_{i,k}$ or a censored event to those with $T_{i,k}^* > C_{i,k}$.
- b. We combine successively the individuals from the 5 data replicates of time-to-event traits such that:
 - Data 1, $N=516$ (replicate 1 of time-to-T1DC traits)
 - Data 2, $N=1032$ (replicates 1 and 2 of time-to-T1DC traits combined)
 - Data 3, $N=1548$ (replicates 1 to 3 of time-to-T1DC traits combined)
 - Data 4, $N=2064$ (replicates 1 to 4 of time-to-T1DC traits combined)
 - Data 5, $N=2580$ (replicates 1 to 5 of time-to-T1DC traits combined)

3. We treat the duplicated individuals in each dataset as independent individuals.

Similarly, to assess the impact of the Winner's curse bias on the classification results for rs1358030 and rs10810632, we generate five additional datasets (named as Data 6: $N=515$, Data 7: $N=1032$, Data 8: $N=1548$, Data 9: $N=2064$ and Data 10: $N=2580$) under the above parametric resampling approach, but we replace the observed genotypes for each SNP in (1) by simulated genotypes with a specified SNP effect on HbA1c ($\beta_{g,1}^W$) equal to 50% of its joint model estimate in DCCT ($\beta_{g,1}^W = 0.5\widehat{\beta}_{g,1}$) using the procedure described in **File 2 (section 3)** for the simulation of the SNPs with indirect effects.

Analysis of each resampled dataset

We apply to each dataset (Data1 to Data10) the same joint model as described in the paper for joint analysis of HbA1c, SBP, DR and DN for each SNP (see Section 3, for details), with the cumulative time-weighted association structure for HbA1c effects on T1DC traits. We obtain empirical estimates of the joint variance-covariance matrices using $B=500$ bootstraps. We classify each SNP as direct and/or indirect association using the same level of significance as used in the manuscript for the classification of SNPs in DCCT (that is $P^*=1.7\times 10^{-4}$). For each SNP and each dataset, we compute 95% and 99% confidence intervals of the SNP effects estimate and present joint model results, as well as the dependence of SNP classification results on increasing sample size. Joint model results, confidence intervals of the SNP effect estimates as well as classification results in each dataset are shown, respectively, in **Table S6-1** and **Fig. S6-1 to 6-3** for rs10810632 (MAF=7%) and in **Table S6-2, Fig. S6-4 to 6-5** for rs1358030 (MAF=36%).

Table S6-1. Dependence of joint model results for rs10810632 (MAF=7%) on increasing sample size, investigated by parametric resampling under a specified alternative hypothesis, and adjusted for winner's curse bias. We generated Data 1 to Data 10 with increasing sample size, based on the parametric resampling approach applied to DCCT individuals. For Data 6 to Data 10, we specified the effect size for rs10810632 on HbA1c equal to 50% of its estimate in DCCT. Only the SNP effects on HbA1c and T1DC traits and the effects of HbA1c on the T1DC traits are presented here.

Traits	Parameters		DCCT N=516	Sample size analyses					Sample size analyses with a reduced SNP effect on HbA1c				
				Data 1 N=516	Data 2 N=1032	Data 3 N=1548	Data 4 N=2064	Data 5 N=2580	Data 6 N=516	Data 7 N=1032	Data 8 N=1548	Data 9 N=2064	Data 10 N=2580
HbA1c (l=1)	$\beta_{g,1}$	Effect ¹	0.93	0.94	0.94	0.94	0.94	0.95	0.51	0.51	0.52	0.52	0.52
		sd ²	0.17	0.17	0.12	0.10	0.08	0.07	0.15	0.11	0.08	0.07	0.07
		P-value ³	2.8E-08	2.5E-08	3.2E-15	6.2E-22	4.7E-30	3.0E-38	5.2E-04	1.0E-06	7.7E-10	3.1E-13	1.1E-14
DR (k=1)	$\alpha_{1,1}$	Effect ¹	0.48	0.41	0.49	0.45	0.43	0.46	0.49	0.54	0.54	0.54	0.50
		sd ²	0.06	0.06	0.04	0.03	0.03	0.02	0.06	0.04	0.03	0.03	0.03
		P-value ³	4.7E-16	4.9E-13	8.2E-32	9.0E-45	1.6E-53	2.7E-79	1.9E-16	3.6E-37	5.6E-57	1.3E-64	6.2E-71
	$\mu_{g,1,1}$ $= \alpha_{g,1,1}\beta_{g,1}$	Effect ¹	0.45	0.39	0.46	0.43	0.41	0.43	0.25	0.28	0.28	0.28	0.26
		sd ²	0.10	0.07	0.06	0.04	0.04	0.03	0.07	0.06	0.05	0.04	0.03
		P-value ³	1.9E-05	5.3E-06	3.4E-11	2.5E-16	1.9E-20	1.1E-27	1.8E-03	5.6E-06	1.8E-08	6.4E-11	1.1E-12
	$\gamma_{g,1}$	Effect ¹	0.30	0.17	0.23	0.33	0.40	0.35	0.17	0.29	0.19	0.20	0.21
		sd ²	0.21	0.19	0.13	0.10	0.10	0.09	0.24	0.17	0.13	0.12	0.10
		P-value ³	1.4E-01	3.6E-01	8.4E-02	9.9E-04	2.9E-05	6.8E-05	4.7E-01	9.3E-02	1.5E-01	8.2E-02	3.1E-02
	$\theta_{1,1} = \mu_{g,1,1}$ $+ \gamma_{g,1}$	Effect ¹	0.75	0.56	0.69	0.76	0.81	0.79	0.42	0.57	0.47	0.48	0.47
		sd ²	0.22	0.07	0.06	0.04	0.04	0.03	0.07	0.06	0.05	0.04	0.03
		P-value ³	5.3E-04	3.1E-03	1.4E-06	1.7E-12	6.4E-15	2.3E-16	9.6E-02	1.8E-03	8.2E-04	1.2E-04	1.1E-05
DN (k=2)	$\alpha_{1,2}$	Effect ¹	0.50	0.50	0.46	0.47	0.50	0.52	0.46	0.58	0.59	0.59	0.59
		sd ²	0.11	0.11	0.07	0.06	0.05	0.04	0.13	0.08	0.06	0.05	0.05
		P-value ³	6.2E-06	1.3E-05	3.5E-10	8.7E-16	5.4E-25	1.0E-31	2.4E-04	4.3E-13	3.9E-20	2.0E-28	2.0E-35
	$\mu_{g,1,2}$ $= \alpha_{g,1,2}\beta_{g,1}$	Effect ¹	0.46	0.47	0.44	0.45	0.48	0.49	0.23	0.30	0.31	0.31	0.31
		sd ²	0.14	0.14	0.09	0.07	0.06	0.05	0.10	0.08	0.06	0.05	0.05
		P-value ³	8.9E-04	6.9E-04	1.8E-06	3.3E-10	3.9E-15	2.8E-19	1.5E-02	7.7E-05	8.7E-07	1.2E-09	4.9E-10
	$\gamma_{g,2}$	Effect ¹	0.49	0.43	0.64	0.60	0.54	0.54	0.36	0.57	0.53	0.41	0.34
		sd ²	0.40	0.40	0.28	0.22	0.18	0.16	0.37	0.25	0.20	0.16	0.15
		P-value ³	2.1E-01	2.9E-01	2.2E-02	5.7E-03	2.6E-03	7.7E-04	3.3E-01	2.1E-02	7.2E-03	1.1E-02	1.9E-02
	$\theta_{1,2} = \mu_{g,1,2}$ $+ \gamma_{g,2}$	Effect ¹	0.96	0.90	1.07	1.05	1.02	1.03	0.59	0.87	0.84	0.72	0.65
		sd ²	0.39	0.08	0.06	0.05	0.04	0.04	0.07	0.06	0.05	0.04	0.04
		P-value ³	1.5E-02	1.5E-02	2.6E-05	2.1E-07	2.3E-09	1.9E-11	1.1E-01	5.5E-04	3.3E-05	1.6E-05	6.9E-06

¹Joint model effect size estimates.

²Empirical standard deviation estimated using 500 bootstraps.

³One degree of freedom Wald test P-values for the SNP effect; P-values indicated in bold satisfy the corrected significance level of $P^*=1.7E-04$.

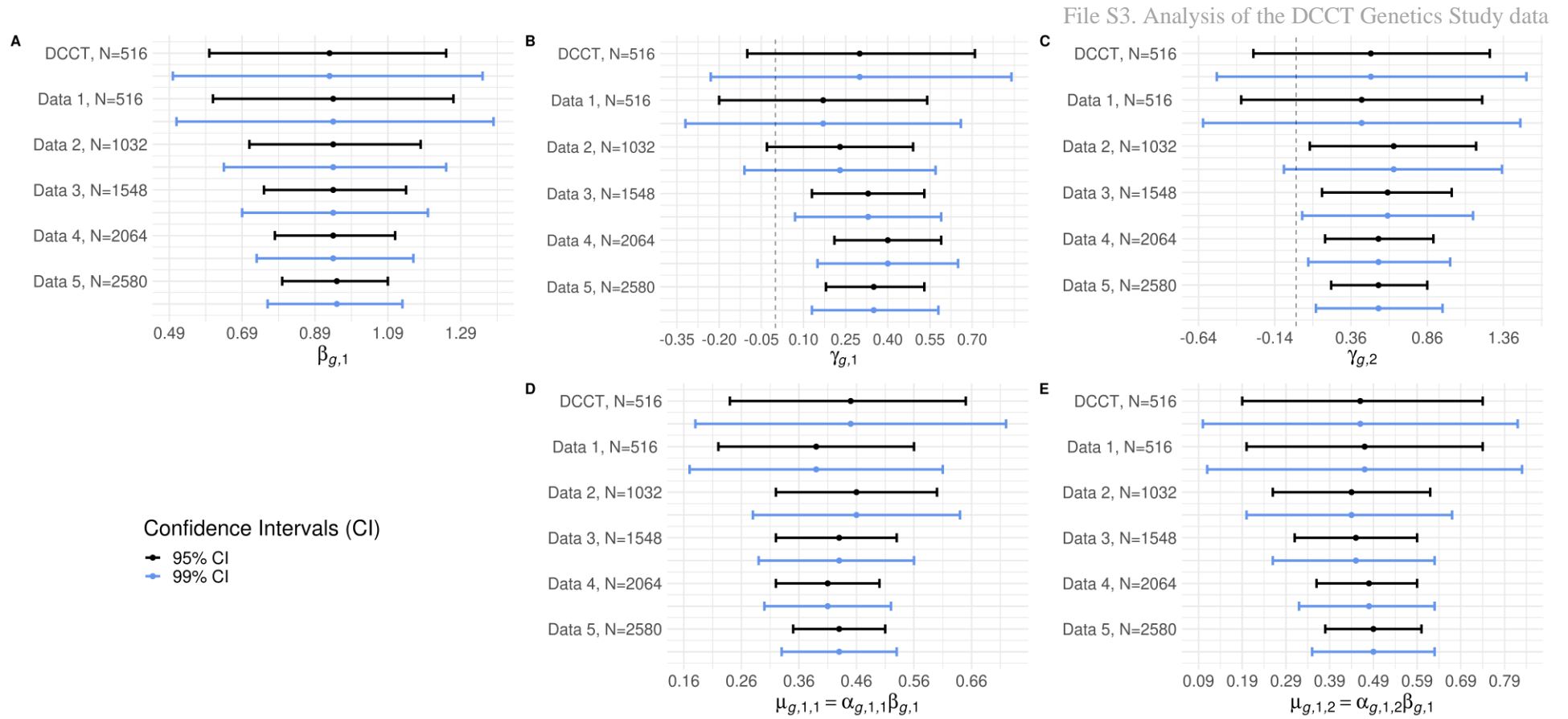


Fig. S6-1. Confidence Intervals for the effects of rs10810632 (MAF=7%) on HbA1c and on each T1DC trait estimated by the joint model in datasets generated with increasing sample sizes by the parametric resampling approach under a specified alternative hypothesis, as described above. Each plot corresponds to: (A) effect of rs10810632 on HbA1c, (B) direct effect of rs10810632 on DR, (C) direct effect of rs10810632 on DN, (D) indirect effect of rs1358030 on DR; and (E) indirect effect of rs10810632 on DN. For each plot, “DCCT, N=516” corresponds to the Confidence Intervals calculated in DCCT; while “Data 1, N=516”, to “Data 5, N=2580”, correspond to the Confidence Intervals calculated in each of the generated dataset. We calculate confidence intervals as $\text{effect_estimate} \pm z_{\alpha} \times \text{bootstrap_sd}$, where z_{α} correspond to the $1 - \alpha/2$ quantile of a standard normal distribution; with $\alpha = 95\%$ and 99% . The vertical dashed lines indicate the reference line for a null SNP effect. One degree of freedom Wald tests for each of these SNP effects are shown in **Table S6-1** and the change of the classification results for rs10810632 according to the sample size is shown in **Fig S6-3**.

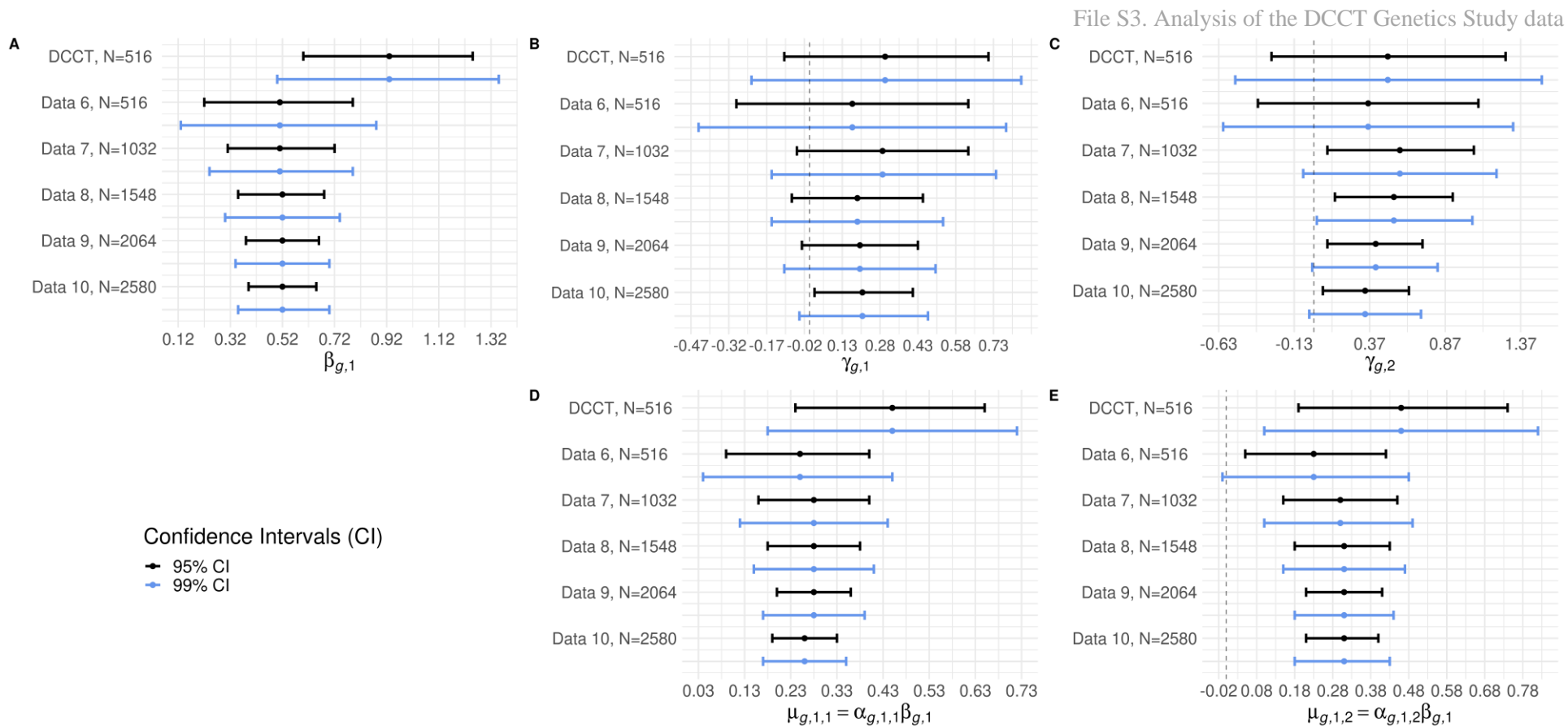


Fig. S6-2. As **Fig. S6-1** but using Data 6 to 10 generated with a Winner's curse adjusted effect of rs10810632 (MAF=7%) on HbA1c (equal to 50% of its estimate in DCCT).

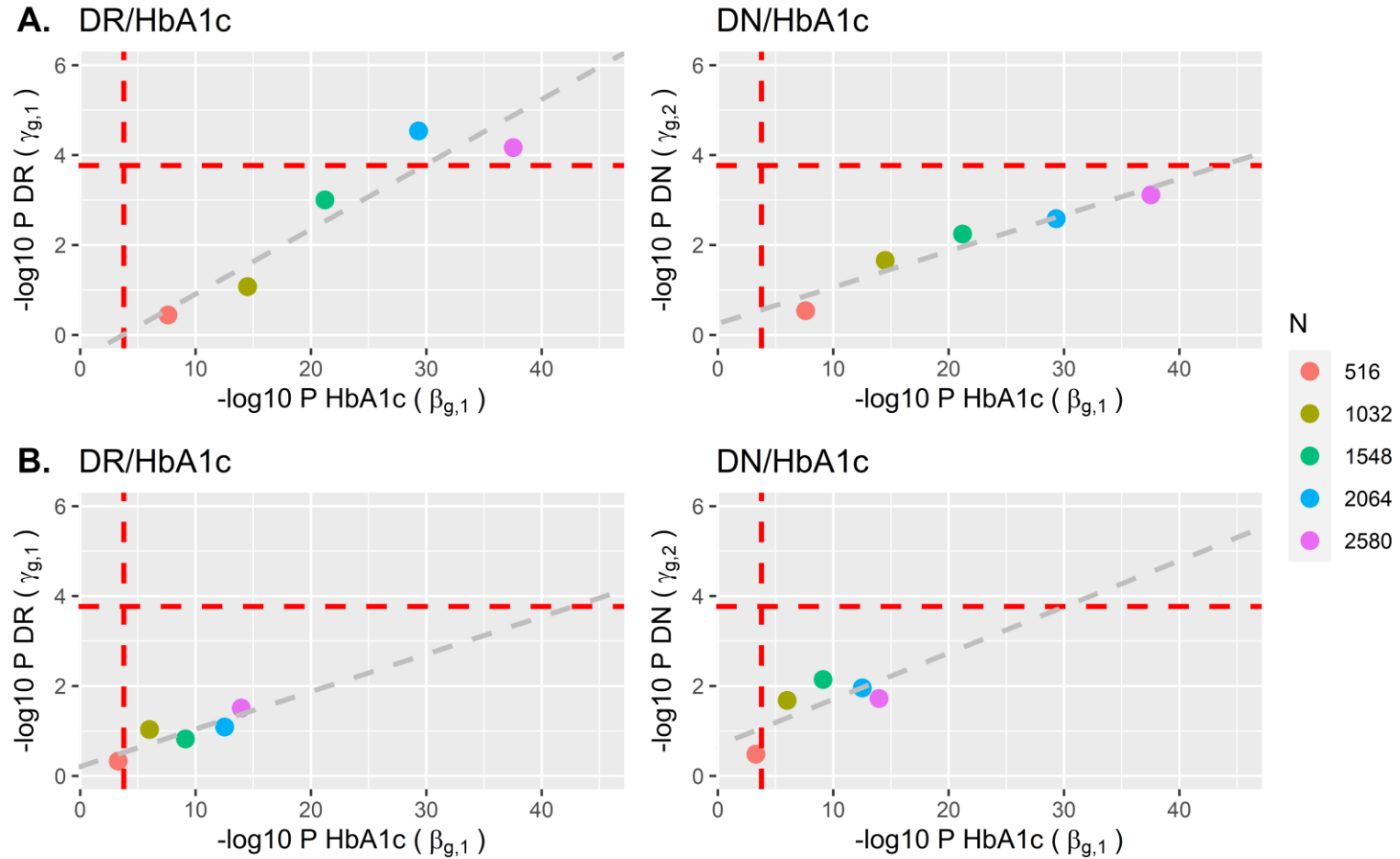


Fig. S6-3. Dependence of SNP classification results for rs10810632 (MAF=7%) with DR/HbA1c and DN/HbA1c on increasing sample size, investigated by parametric resampling (A) under a specified alternative hypothesis, and (B) adjusted for winner's curse bias. Panel A show results for Data 1 ($N=516$) to Data 5 ($N=2580$), generated using the parametric resampling approach described above, while Panel B, shows results for Data 6 ($N=516$) to Data 10 ($N=2580$), generated as for panel A but with a specified $\beta_{g,1}$ adjusted for the Winner's curse effect. Vertical and horizontal red dashed lines denote the significance threshold used for the SNP classification procedure ($P^* = 1.7 \times 10^{-4}$). For each plot, we fitted a regression line (represented by a grey dashed line) to visualize the trend of the classification results with increasing DCCT sample sizes.

Table S6-2. Joint model test results for rs1358030 (MAF=36%) according to increasing sample size, investigated by parametric resampling under a specified alternative hypothesis, and adjusted for winner's curse bias. Only the SNP effects on HbA1c and T1DC traits and the effects of HbA1c on the T1DC traits are presented here.

Traits	Parameters		DCCT N=516	Sample size analyses					Sample size analyses accounting for Winner's cure bias				
				Data 1 N=516	Data 2 N=1032	Data 3 N=1548	Data 4 N=2064	Data 5 N=2580	Data 6 N=516	Data 7 N=1032	Data 8 N=1548	Data 9 N=2064	Data 10 N=2580
HbA1c (l=1)	$\beta_{g,1}$	Effect ¹	0.44	0.44	0.44	0.44	0.44	0.44	0.24	0.24	0.24	0.24	0.24
		sd ²	0.08	0.08	0.06	0.04	0.04	0.03	0.08	0.06	0.05	0.04	0.04
		P-value ³	1.1E-08	8.6E-09	1.3E-15	2.1E-25	1.1E-32	2.6E-38	2.7E-03	1.8E-05	3.1E-07	2.0E-09	5.9E-11
DR (k=1)	$\alpha_{1,1}$	Effect ¹	0.55	0.52	0.60	0.56	0.54	0.55	0.52	0.58	0.61	0.61	0.61
		sd ²	0.06	0.06	0.05	0.03	0.03	0.03	0.06	0.04	0.04	0.03	0.03
		P-value ³	1.0E-18	5.3E-16	2.6E-39	1.6E-58	1.6E-72	2.5E-103	8.5E-19	4.7E-39	2.2E-62	5.1E-83	1.2E-90
	$\mu_{g,1,1}$ $= \alpha_{g,1,1}\beta_{g,1}$	Effect ¹	0.24	0.23	0.27	0.25	0.24	0.24	0.12	0.14	0.14	0.15	0.15
		sd ²	0.05	0.04	0.03	0.02	0.02	0.02	0.04	0.03	0.03	0.02	0.02
		P-value ³	3.8E-06	2.5E-06	1.1E-11	3.5E-18	1.7E-22	3.6E-27	4.8E-03	6.9E-05	2.0E-06	2.6E-08	3.7E-10
	$\gamma_{g,1}$	Effect ¹	-0.20	-0.30	-0.30	-0.25	-0.22	-0.19	-0.07	-0.16	-0.21	-0.23	-0.22
		sd ²	0.11	0.12	0.08	0.06	0.05	0.05	0.10	0.08	0.06	0.05	0.05
		P-value ³	7.0E-02	1.4E-02	2.5E-04	5.2E-05	2.5E-05	7.0E-05	4.8E-01	3.5E-02	1.4E-03	2.1E-05	5.1E-06
	$\theta_{1,1} = \mu_{g,1,1}$ $+ \gamma_{g,1}$	Effect ¹	0.04	-0.07	-0.03	0.00	0.02	0.06	0.05	-0.02	-0.06	-0.09	-0.07
		sd ²	0.11	0.04	0.03	0.02	0.02	0.02	0.04	0.03	0.03	0.02	0.02
		P-value ³	7.5E-01	5.4E-01	6.9E-01	9.7E-01	6.9E-01	2.3E-01	6.4E-01	7.7E-01	3.6E-01	1.3E-01	1.7E-01
DN (k=2)	$\alpha_{1,2}$	Effect ¹	0.58	0.57	0.55	0.57	0.59	0.61	0.76	0.71	0.71	0.68	0.69
		sd ²	0.11	0.12	0.08	0.06	0.05	0.04	0.11	0.08	0.06	0.05	0.05
		P-value ³	2.6E-07	1.1E-06	3.5E-13	9.2E-21	1.5E-31	1.5E-42	1.5E-11	4.5E-21	8.8E-31	9.1E-42	2.7E-44
	$\mu_{g,1,2}$ $= \alpha_{g,1,2}\beta_{g,1}$	Effect ¹	0.254	0.253	0.244	0.251	0.259	0.271	0.180	0.169	0.168	0.164	0.165
		sd ²	0.069	0.065	0.045	0.037	0.031	0.028	0.065	0.043	0.036	0.030	0.028
		P-value ³	2.5E-04	9.8E-05	7.8E-08	1.5E-11	1.8E-16	3.0E-22	5.8E-03	8.3E-05	3.6E-06	7.2E-08	4.3E-09
	$\gamma_{g,2}$	Effect ¹	-0.31	-0.29	-0.25	-0.38	-0.31	-0.39	-0.34	-0.21	-0.27	-0.31	-0.29
		sd ²	0.22	0.26	0.16	0.12	0.10	0.09	0.24	0.17	0.13	0.11	0.10
		P-value ³	1.6E-01	2.6E-01	1.2E-01	1.6E-03	2.7E-03	1.5E-05	1.5E-01	2.1E-01	3.0E-02	2.9E-03	5.5E-03
	$\theta_{1,2} = \mu_{g,1,2}$ $+ \gamma_{g,2}$	Effect ¹	-0.05	-0.04	-0.01	-0.13	-0.05	-0.12	-0.16	-0.04	-0.10	-0.15	-0.12
		sd ²	0.21	0.04	0.03	0.02	0.02	0.02	0.06	0.04	0.03	0.03	0.03
		P-value ³	8.0E-01	8.9E-01	9.7E-01	2.7E-01	6.3E-01	1.6E-01	5.1E-01	8.1E-01	4.1E-01	1.7E-01	2.5E-01

¹Joint model effect size estimates.

²Empirical standard error estimated using 500 bootstraps.

³One degree of freedom Wald test P-values for the SNP effect; P-values indicated in bold satisfy the corrected significance level of $P^*=1.7E-04$

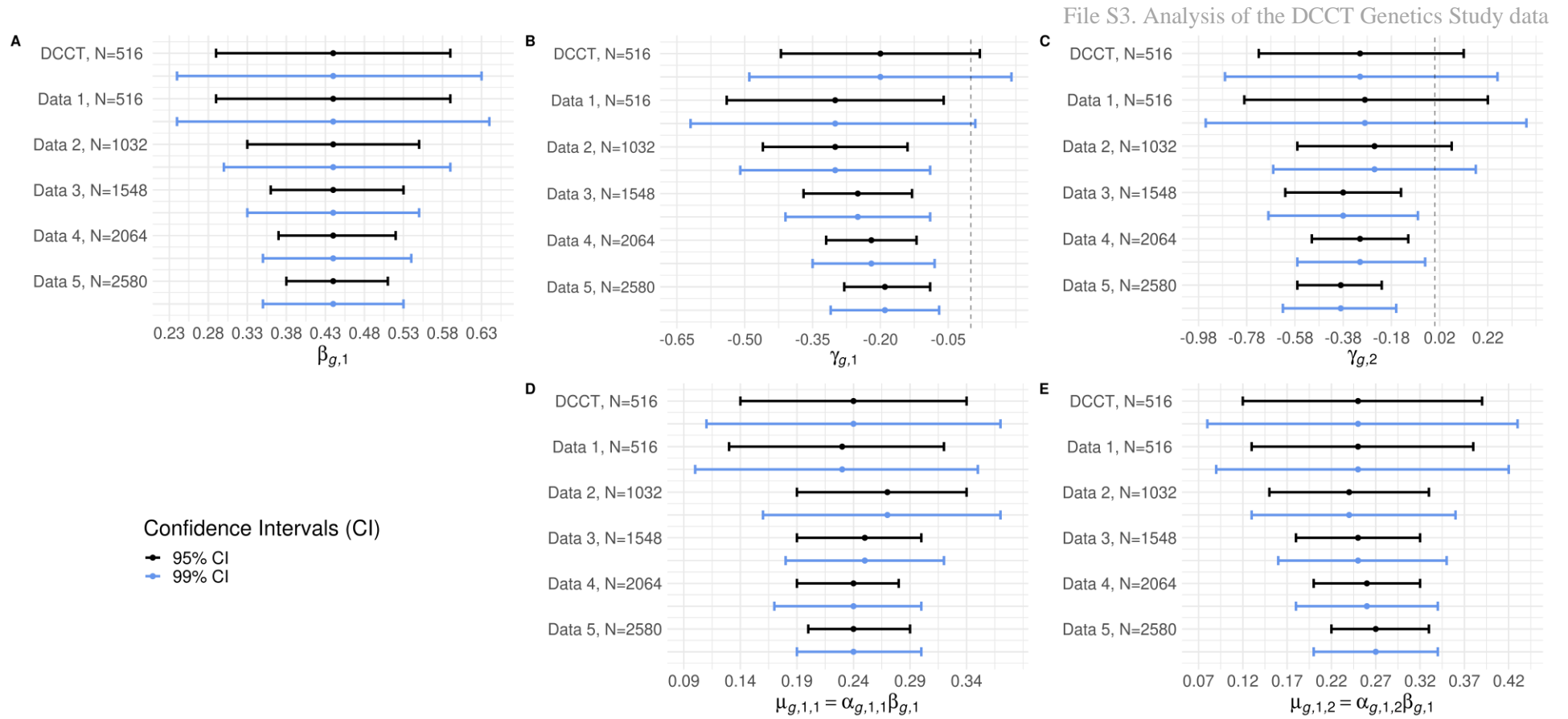


Fig. S6-4. Confidence Intervals for the effects of rs1358030 (MAF=36%) estimated by the joint model to datasets with increasing sample sizes generated by the parametric resampling approach under a specified alternative hypothesis. Each plot corresponds to: (A) effect of rs1358030 on HbA1c, (B) direct effect of rs1358030 on DR, (C) direct effect of rs1358030 on DN, (D) indirect effect of rs1358030 on DR; and (E) indirect effect of rs1358030 on DN. For each plot, “DCCT, N=516” corresponds to the Confidence Intervals calculated in DCCT; while “Data 1, N=516”, to “Data 5, N=2580”, correspond to the Confidence Intervals calculated in each of the generated dataset. We calculate confidence intervals as $\text{effect_estimate} \pm z_\alpha \times \text{bootstrap_sd}$, where z_α correspond to the $1 - \alpha/2$ quantile of a standard normal distribution; with $\alpha = 95\%$ and 99% . The vertical dashed lines indicate the reference line for a null SNP effect.

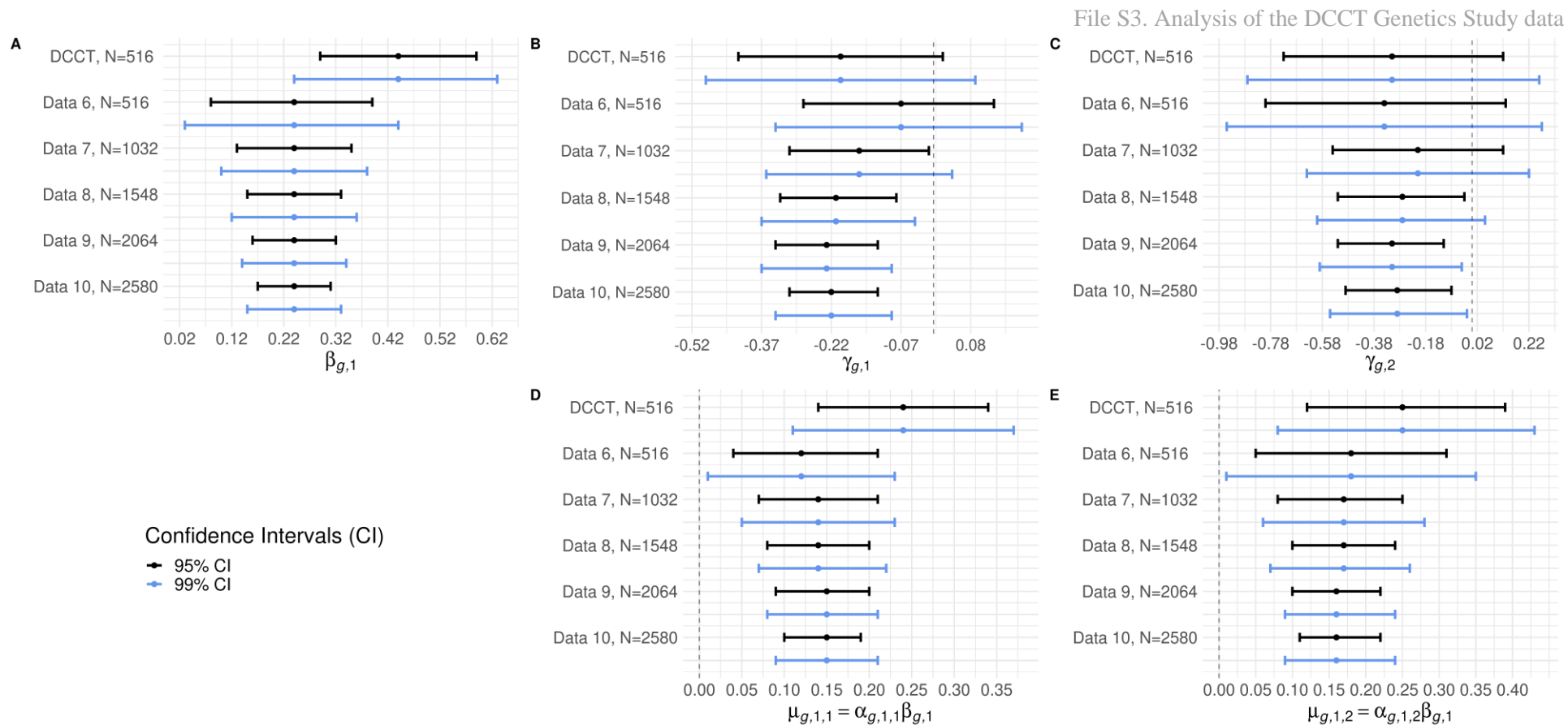


Fig. S6-5. As **Fig. S6-4** but Data 6 to 10 are generated with an effect of rs1358030 (MAF=36%) on HbA1c adjusted for the Winner's curse bias

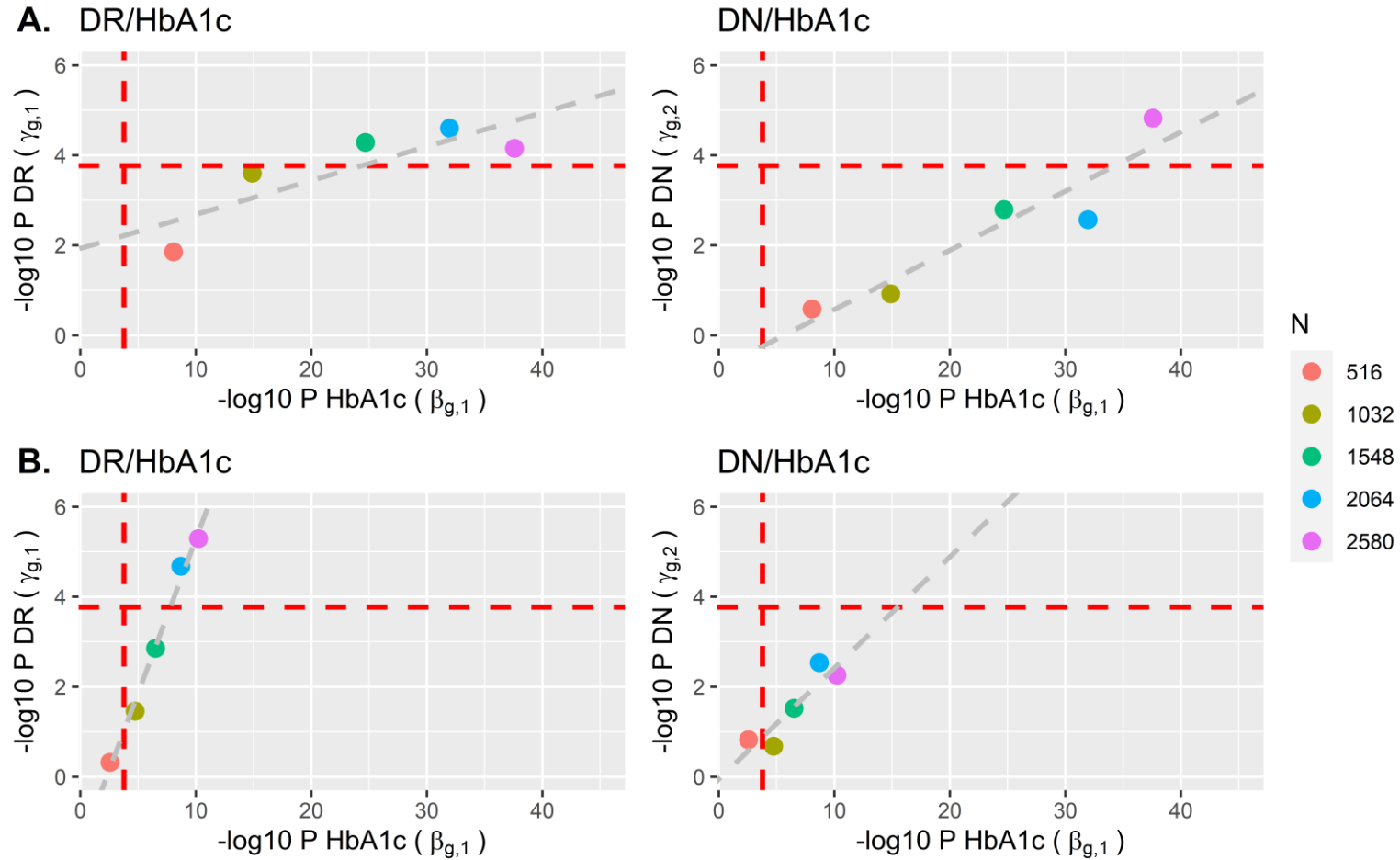


Fig. S6-6. Dependence of SNP classification results for rs1358030 (MAF=36%) with DR/HbA1c and DN/HbA1c on increasing sample size, investigated by the parametric resampling approach (A) under a specified alternative hypothesis, and (B) adjusted for winner's curse bias. Panel A show results for Data 1 ($N=516$) to Data 5 ($N=2580$), generated using the parametric resampling approach described above, while Panel B, shows results for Data 6 ($N=516$) to Data 10 ($N=2580$), generated as for panel A but with a specified $\beta_{g,1}$ adjusted for the Winner's curse effect (set to 50% of its estimate in DCCT). Vertical and horizontal red dashed lines denote the significance threshold used for the SNP classification procedure ($P^* = 1.7 \times 10^{-4}$). For each plot, we fitted a regression line (represented by the grey dashed line) to visualize the trend of the classification results with increasing DCCT sample sizes.

Cited Literature

- Arisido MW, Antolini L, Bernasconi DP, Valsecchi MG, Rebora P. 2019. Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. *BMC Med Res Methodol.* 19(1):222. doi:10.1186/s12874-019-0873-y. <http://www.ncbi.nlm.nih.gov/pubmed/31795933>.
- Chen J, Yu K, Hsing A, Therneau TM. 2007. A partially linear tree-based regression model for assessing complex joint gene–gene and gene–environment effects. *Genet Epidemiol.* 31(3):238–251. doi:10.1002/gepi.20205. <https://onlinelibrary.wiley.com/doi/10.1002/gepi.20205>.
- Coombes BJ, Biernacka JM. 2018. Application of the parametric bootstrap for gene-set analysis of gene–environment interactions. *Eur J Hum Genet.* 26(11):1679–1686. doi:10.1038/s41431-018-0236-x. <http://www.nature.com/articles/s41431-018-0236-x>.
- Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al. 2018. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet.* 50(10):1412–1425. doi:10.1038/s41588-018-0205-x. <http://www.ncbi.nlm.nih.gov/pubmed/30224653>.
- Grambsch PM, Therneau TM. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 81(3):515–526. doi:10.1093/biomet/81.3.515. <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/81.3.515>.
- Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. 2011. Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet.* 20(12):2472–2481. doi:10.1093/hmg/ddr121.
- Hosseini SM, Boright AP, Sun L, Canty AJ, Bull SB, Klein BEK, Klein R, Paterson AD. 2015. The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy. *Hum Genet.* 134(2):247–257. doi:10.1007/s00439-014-1517-2.
- Kleinman K, Huang SS. 2017. Calculating power by bootstrap, with an application to cluster-randomized trials. *eGEMs (Generating Evid Methods to Improv patient outcomes).* 4(1):32. doi:10.13063/2327-9214.1202. <https://egems.journal.ubiquity.website//article/10.13063/2327-9214.1202/>.
- Lind M, Odén A, Fahlén M, Eliasson B. 1995. The Relationship of Glycemic Exposure (HbA1c) to the Risk of Development and Progression of Retinopathy in the Diabetes Control and Complications Trial. *Diabetes.* 44(8):968–983. doi:10.2337/diab.44.8.968. <https://diabetesjournals.org/diabetes/article/44/8/968/8802/The-Relationship-of-Glycemic-Exposure-HbA1c-to-the>.
- Lind M, Odén A, Fahlén M, Eliasson B. 2010. The shape of the metabolic memory of HbA1c: Re-analysing the DCCT with respect to time-dependent effects. *Diabetologia.* 53(6):1093–1098. doi:10.1007/s00125-010-1706-z.

Machiela MJ, Chanock SJ. 2015. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 31(21):3555–3557. doi:10.1093/bioinformatics/btv402.

Paterson AD, Waggott D, Boright AP, Hosseini SM, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary PA, Lachin JM, et al. 2010. A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. *Diabetes*. 59(2):539–549. doi:10.2337/db09-0653.

Pollack S, Igo RP, Jensen RA, Christiansen M, Li X, Cheng C-Y, Ng MCY, Smith A V, Rossin EJ, Segrè A V, et al. 2019. Multiethnic genome-wide association study of diabetic retinopathy using liability threshold modeling of duration of diabetes and glycemic control. *Diabetes*. 68(2):441–456. doi:10.2337/db18-0567. [accessed 2019 Oct 29]. <http://www.ncbi.nlm.nih.gov/pubmed/30487263>.

Rizopoulos D. 2012. Joint Models for Longitudinal and Time-to-Event Data. Chapman and Hall/CRC. <https://www.taylorfrancis.com/books/9781439872871>.

Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, McKay GJ, Williams WW, Sadlier DM, Mäkinen V-P, et al. 2012. New Susceptibility Loci Associated with Kidney Disease in Type 1 Diabetes. Böger CA, editor. *PLoS Genet*. 8(9):e1002921. doi:10.1371/journal.pgen.1002921. <http://www.ncbi.nlm.nih.gov/pubmed/23028342>.

Therneau TM, Grambsch PM. 2000. Modeling survival data: extending the Cox model. New York, NY: Springer New York (Statistics for Biology and Health). [accessed 2019 Aug 8]. <https://books.google.ca/books?id=oj0mBQAAQBAJ&printsec=frontcover&dq=therneau+and+grambsch&hl=fr&sa=X&ved=0ahUKEwjZ96antfPjAhWXHM0KHbxCCoAQ6AEILDAA#v=onepage&q=age1&f=false>.

Walters SJ, Campbell MJ. 2005. The use of bootstrap methods for estimating sample size and analysing health-related quality of life outcomes. *Stat Med*. 24(7):1075–1102. doi:10.1002/sim.1984. <https://onlinelibrary.wiley.com/doi/10.1002/sim.1984>.

Wheeler E, Leong A, Liu C-TT, Hivert M-FF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, et al. 2017. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med*. 14(9):e1002383. doi:10.1371/journal.pmed.1002383. <http://www.ncbi.nlm.nih.gov/pubmed/28898252>.