Supplementary Materials II: "A Comparison Study of Fixed Effect Models and Mixed Effect Models for Gene Level Association Studies"

Appendix D Results of European Lipid Studies

D.1 Information Of the Eight European Cohorts

We analyzed lipid traits from eight European cohorts, where five are from Finland [Finland United States Investigation of NIDDM Genetics (FUSION Stage 2) [Scott et al., 2007], FIN-D2D 2007 (D2d-2007) [Kotronen et al., 2010], The Finnish Diabetes Prevention Study (DPS) [Tuomilehto et al., 2001], METabolic Syndrome in Men (METSIM) [Stancakova et al., 2009], and The Dose Responses to Exercise Training Study (DRs EXTRA) [Kouki et al., 2012], two are from Norway [Nord-Trondelag Health Study 2 and Tromso 4 (HUNT and Tromso) [Holmen et al., 2003; Jacobsen et al., 2012], and one from Germany [The DIAbetes GENetic Study (DIAGEN)] [Schwarz et al., 2006]. The two Norwegian cohorts were combined into one study for a joint analysis. The genotype data were from Metabochip genotyping, which was designed to fine map regions that have been associated with metabolic traits [Altshuler et al., 2010]. For each cohort, 54,741 genetic variants were genotyped, located in 97 genetic regions across the 22 autosomes. For our analysis, we utilized the existing literature as a reference for gene selection and found that 22 gene regions were fine mapped [Li et al., 2014; Liu et al., 2014; Morris et al., 2012; Scott et al., 2012; Voight et al., 2010; Zeggini et al., 2008]. We used Builder Mar. 2006 (NCBI36/hg18) to determine gene positions and 5kb was used to extend the gene region on each side of a gene. The summary of 22 genes and the number of genetic variants in each gene region are given in Table S.19.

Four lipid traits were analyzed: high-density lipoprotein (HDL) levels, low-density lipoprotein (LDL) levels, triglycerides (TG), and total cholesterol (CHOL). The sample sizes for each combination of seven studies and four trait are provided in Table S.20.

D.2 Lipid Traits in Eight European Cohorts

For each trait, inverse normal rank transformation was performed to ensure that the normality assumption was valid. For all studies except for METSIM, age, sex, and type 2 diabetes status were used as covariates. For METSIM, age and type 2 diabetes status were used as covariates since no women were included in the study. A significance threshold of $P < 3.1 \times 10^{-6}$ was taken from Liu et al. [2014] (corresponding to 0.05/16,153 based on the number of genes tested therein).

Table S.21 reports significant results of association analysis of 5 European studies in the regions of *APOE* and *LDLR* genes [Table 1 of Wang et al., 2015]. At the significance threshold of $P < 3.1 \times 10^{-6}$, we detected association at *APOE* in the five European studies: D2d-2007, FUSION Stage 2, Norway, DIAGEN, and METSIM. At *LDLR*, association was detected in one study of METSIM. For the studies of D2d-2007 and FUSION Stage 2, two traits (LDL and CHOL) and their bivariate combination (LDL, CHOL) showed association with *APOE* by our F-approximation tests as well as SKAT-O. For the studies of Norway, DIAGEN, and METSIM, LDL and the tri-variate combination (LDL, TG, CHOL) were associated with *APOE*. For the study of Norway, CHOL and bivariate combinations of (LDL, TG), (LDL, CHOL), and (TG, CHOL) were associated with *APOE*.

For the studies of DIAGEN and METSIM, neither TG nor CHOL showed significant association with *APOE* at the significance threshold of $P < 3.1 \times 10^{-6}$. However, the bivariate combinations and tri-variate combinations were significantly associated with *APOE*. The bivariate combination (TG, CHOL) also showed association with *APOE* in the DIAGEN study despite the fact that neither TG nor CHOL was significant in the univariate analysis. For the gene *LDLR*, CHOL showed a significant association while LDL did not; the bivariate combination (LDL, CHOL) also was significantly associated with *LDLR*.

In general, the F-approximation tests of multivariate functional linear models are more sensitive than the F-approximation tests of the multivariate additive model which in turn is more sensitive than SKAT-O in the univariate case. SKAT-O only detected association of two traits (LDL and CHOL) with *APOE* in

two studies, D2d-2007 and FUSION Stage 2. In comparison, the F-approximation tests of the multivariate additive linear model detected more association than SKAT-O in the univariate case between two traits (LDL and CHOL) and *APOE* in the study of Norway. Generally, the p-values o F-approximation tests of multivariate functional linear models are smaller than those of the F-approximation tests of the multivariate additive model. In the study of DIAGEN, the F-approximation tests of the multivariate additive model did not detect any association between LDL [or (TG, CHOL)] and APOE. In the METSIM study, the F-approximation tests of the multivariate additive model did not detect any association between LDL [or (LDL, CHOL) or (LDL, TG, CHOL)] and *APOE*, and between CHOL and *LDLR*.

Appendix E Results of The Trinity Students Study

We performed a pleiotropy analysis of 36 SNP variants in one enzyme gene region on three biochemical traits (denoted by A, B, and C) in a sample of 2232 individuals from the Trinity Students Study. Since the raw traits were not normally distributed, we transformed the three traits by inverse normal rank transformation. We adjusted for three factors: gender, another chemical compound known to affect these biochemical traits as a continuous covariate, and a dichotomous covariate to indicate if supplements containing these biochemical factors was used.

In Fan et al. [2013], the three traits were analyzed individually and the results were compared with both SKAT and SKAT-O. In Wang et al. [2015], we analyzed four combinations of the three traits: three bivariate combinations (A, B), (A, C), (B, C), and one tri-variate combination (A, B, C). We tested the association between the transformed individual traits and the 36 SNPs by approximate F-test statistics of bivariate and tri-variate linear models using B-spline basis, Fourier basis, and linear spline basis functions.

Table S.22 presents the p-values of the F-approximation tests based on the Pillai-Bartlett trace for the SNP data of the enzyme gene of the Trinity Students Study [Table 2 of Wang et al., 2015]. We present the results of four combinations of the three traits on the bottom of the Table S.22: (A, B), (A, C), (B, C), and (A, B, C). The four combinations of (A, B), (A, C), (B, C), and (A, B, C) provided much stronger

results than those of univariate analysis individually since the p-values of the approximate F-distribution test statistics in the bottom four columns of Table S.22 were much smaller than the F-test statistics of the individual univariate analyses of the three traits, A, B, and C. For all three traits, A, B, and C, the results of the univariate F-distributed tests are far better than those of SKAT-O [Table S.22 and Fan et al., 2013]. Again, the p-values of our F-approximation tests are smaller than those of the F-approximation tests of the multivariate additive model.

Reference

Altshuler DM, Lander ES, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL. 2010. A map of human genome variation from population scale sequencing. *Nature* 467:10611073.

Fan R, Wang Y, Mills JL, Wilson AF, Bailey-Wilson JE, Xiong MM. 2013. Functional linear models for association analysis of quantitative traits. *Genetic Epidemiology* 37:726-742.

Holmen J, Midthjell K, Krüger O, Langhammer A, Holmen TL, Bratberg GH, Vatten L, Lund-Larsen PG et al. 2003. The Nord-Trndelag Health Study 1995-97 (HUNT 2): Objectives, contents, methods and participation. *Nor J Epidemiol* 13:19-32.

Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. 2012. Cohort profile: the Tromso Study. *Int J Epidemiol* 41:961-967.

Kotronen A, Yki-Järvinen H, Männistö S, Saarikoski L, Korpi-Hyövälti E, Oksa H, Saltevo J, Saaristo T, Sundvall J, Tuomilehto J et al. 2010. Non-alcoholic and alcoholic fatty liver disease - two diseases of affluence associated with the metabolic syndrome and type 2 diabetes: the FIN-D2D survey. *BMC Public Health* 10:237.

Kouki R, Schwab U, Lakka TA, Hassinen M, Savonen K, Komulainen P, Krachler B, Rauramaa R et al.

2012. Diet, fitness and metabolic syndrome-the DR's EXTRA study. *Nutr Metab Cardiovasc Dis* 22:553-560.

Li S, Mukherjee B, Taylor JMG, Rice KM, Wen X, Rice JD, Stringham HM, and Boehnke M. 2014. The role of environmental heterogeneity in meta-analysis of gene-environment interactions with quantitative traits. *Genetic Epidemiology* 38:416-429.

Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H et al. 2014. Meta-analysis of gene-level tests for rare variant association. *Nat Genet* 46:200-204.

Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, et al. 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* 44:981-990.

Schwarz PE, Towers GW, Fischer S, Govindarajalu S, Schulze J, Bornstein SR, Hanefeld M, Vasseur F et al. 2006. Hypoadiponectinemia is associated with progression toward type 2 diabetes and genetic variation in the ADIPOQ gene promoter. *Diabetes Care* 29:1645-1650.

Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Magi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al. 2012. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics* 44:991-1005.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU et al. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341-1345.

Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M et al. 2009. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* 58:1212-1221.

Tuomilehto J, Lindströmb J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Keinänen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M et al. 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 344:1343-1350.

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics* 42:579-589.

Wang YF, Liu AY, Mills JL, Boehnke M, Wilson AF, Bailey-Wilson JE, Xiong MM, Wu CO, Fan RZ. 2015. Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models. *Genetic Epidemiology* 39:259-275.

Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PIW, Abecasis GR, Almgren P, Andersen G, et al. 2008. Wellcome Trust Case Control Consortium. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* 40:638-645.

Table S.19: **Summary of 22 Genes and the Number of Genetic Variants in Each Gene Region by Mar. 2006 (NCBI36/hg18).** The number of variants is the number of genetic variants in a region of **Start (-5Kb) - End (+5Kb) Positions**. * The gene region of *PCSK9* is (55277737, 55303114), and (55271537, 55286109) is the region in the database. # The length is the length of the region in bp.

Gene	Chromosome Region	Gene Positions (bp)	Start (-5Kb) - End (+5Kb) Positions (Length#)	Number of Variants
*PCSK9**	1	55277737 - 55303114	55271537 - 55286109 (14572)	74
APOB	2	21077806 - 21120450	21072806 - 21125450 (52644)	223
IGF2BP2	3	186844221 - 187025521	186839221 - 187030521 (191300)	231
CDKAL1	6	20642667 - 21340613	20637667 - 21345613 (707946)	560
JAZF1	7	27836718 - 28186962	27831718 - 28191962 (360244)	384
LPL	8	19840862 - 19869050	19835862 - 19874050 (38188)	212
CDKN2B	9	21992902 - 21999312	21987902 - 22004312 (16410)	64
CDC123	10	12277971 - 12332593	12272971 - 12337593 (64622)	265
IDE	10	94201421 - 94323832	94196421 - 94328832 (132411)	327
KIF11	10	94342805 - 94405132	94337805 - 94410132 (72327)	216
HHEX	10	94439661 - 94445388	94434661 - 94450388 (15727)	30
TCF7L2	10	114699999 - 114917426	114694999 - 114922426 (227427)	258
KCNQ1	11	2422797 - 2826916	2417797 - 2831916 (414119)	660
MTNR1B	11	92342437 - 92355596	92337437 - 92360596 (23159)	106
HMGA2	12	64504507 - 64646338	64499507 - 64651338 (151831)	214
TSPAN8	12	69805144 - 69838046	69800144 - 69843046 (42902)	54
HNF1A	12	119900932 - 119924697	119895932 - 119929697 (33765)	71
OASL	12	119942478 - 119961428	119937478 - 119966428 (28950)	108
FTO	16	52295376 - 52705882	52290376 - 52710882 (420506)	191
LDLR	19	11061038 - 11105505	11056038 - 11110505 (54467)	43
APOE	19	50100879 - 50104490	50095879 - 50109490 (13611)	35
GIPR	19	50863342 - 50877557	50858342 - 50882557 (24215)	37

Table S.20: **Sample Sizes of the Four Lipid Traits for Each of the Seven Studies.**

Study	HDL	LDL	TG	CHOL
D2d-2007	2075	2074	2075	2075
DIAGEN	1470	1454	1470	1471
DPS	412	410	412	412
DRs EXTRA	1157	1157	1157	1157
FUSION Stage 2	2496	1892	2062	2500
METSIM	1346	1345	1346	1346
Norway	2484	2320	2487	2476

Table S.21: **Results of Association Analysis of Lipid Traits in 5 European Studies in the Regions of *APOE* and *LDLR* Genes Using the F-approximation Based on Pillai-Bartlett Trace.** The associations that attain a threshold significance of $P < 3.1 \times 10^{-6}$ are highlighted in red [Liu et al. 2014].

Study	Gene	Traits	P-values of the F-approximation Based on Pillai-Bartlett Trace					P-values of SKAT-O
			Basis of both GVF and $\beta_\ell(t)$		Basis of beta-Smooth Only		Multivariate Model (1)	
			B-sp Basis	Fourier Basis	B-sp Basis	Fourier Basis		
D2d-2007	APOE	LDL	1.89×10^{-25}	9.02×10^{-25}	1.89×10^{-25}	9.02×10^{-25}	2.85×10^{-24}	5.87×10^{-13}
		CHOL	9.09×10^{-18}	3.01×10^{-17}	9.09×10^{-18}	3.01×10^{-17}	7.97×10^{-17}	1.72×10^{-9}
		LDL, CHOL	1.21×10^{-20}	2.08×10^{-19}	1.21×10^{-20}	2.08×10^{-19}	7.91×10^{-19}	X
FUSION Stage 2	APOE	LDL	4.34×10^{-10}	2.24×10^{-11}	4.34×10^{-10}	2.24×10^{-11}	3.42×10^{-11}	8.61×10^{-14}
		CHOL	1.34×10^{-12}	4.92×10^{-13}	1.34×10^{-12}	4.92×10^{-13}	8.70×10^{-13}	1.64×10^{-12}
		LDL,CHOL	1.20×10^{-7}	1.29×10^{-8}	1.20×10^{-7}	1.29×10^{-8}	1.75×10^{-8}	X
Norway	APOE	LDL	3.79×10^{-28}	1.90×10^{-27}	3.79×10^{-28}	1.90×10^{-27}	6.05×10^{-27}	6.21×10^{-6}
		TG	5.69×10^{-4}	3.94×10^{-4}	5.69×10^{-4}	3.95×10^{-4}	6.55×10^{-4}	5.55×10^{-2}
		CHOL	2.12×10^{-14}	6.15×10^{-14}	2.12×10^{-14}	6.15×10^{-14}	1.35×10^{-13}	3.00×10^{-3}
		LDL,TG	1.42×10^{-25}	8.16×10^{-25}	1.42×10^{-25}	8.16×10^{-25}	4.72×10^{-24}	X
		LDL,CHOL	8.12×10^{-29}	1.64×10^{-27}	8.12×10^{-29}	1.64×10^{-27}	6.70×10^{-27}	X
		TG,CHOL	5.32×10^{-20}	1.46×10^{-19}	5.32×10^{-20}	1.46×10^{-19}	6.08×10^{-19}	X
		LDL,TG,CHOL	1.18×10^{-24}	3.06×10^{-23}	1.18×10^{-24}	3.06×10^{-23}	1.68×10^{-22}	X
DIAGEN	APOE	LDL	7.84×10^{-7}	3.31×10^{-6}	7.84×10^{-7}	3.31×10^{-6}	5.76×10^{-6}	2.37×10^{-1}
		TG	3.51×10^{-3}	8.53×10^{-3}	3.51×10^{-3}	8.53×10^{-3}	1.23×10^{-2}	7.59×10^{-2}
		CHOL	1.91×10^{-3}	5.61×10^{-3}	1.91×10^{-3}	5.61×10^{-3}	7.38×10^{-3}	4.73×10^{-1}
		LDL,TG	1.78×10^{-8}	1.76×10^{-7}	1.78×10^{-8}	1.76×10^{-7}	4.47×10^{-7}	X
		LDL,CHOL	1.24×10^{-9}	1.44×10^{-8}	1.24×10^{-9}	1.44×10^{-8}	3.24×10^{-8}	X
		TG,CHOL	2.99×10^{-6}	2.49×10^{-5}	2.99×10^{-6}	2.49×10^{-5}	4.51×10^{-5}	X
		LDL,TG,CHOL	1.81×10^{-10}	4.43×10^{-9}	1.81×10^{-10}	4.43×10^{-9}	1.19×10^{-8}	X
METSIM	APOE	LDL	1.85×10^{-5}	1.98×10^{-5}	1.85×10^{-5}	1.98×10^{-5}	3.45×10^{-5}	1.25×10^{-4}
		TG	2.80×10^{-2}	3.43×10^{-2}	2.80×10^{-2}	3.43×10^{-2}	3.96×10^{-2}	4.04×10^{-1}
		CHOL	1.87×10^{-2}	1.84×10^{-2}	1.87×10^{-2}	1.84×10^{-2}	2.73×10^{-2}	5.43×10^{-2}
		LDL,TG	2.70×10^{-7}	3.45×10^{-7}	2.70×10^{-7}	3.45×10^{-7}	7.77×10^{-7}	X
		LDL,CHOL	3.87×10^{-5}	5.63×10^{-5}	3.87×10^{-5}	5.63×10^{-5}	9.45×10^{-5}	X
		LDL,TG,CHOL	1.09×10^{-6}	2.08×10^{-6}	1.09×10^{-6}	2.08×10^{-7}	3.91×10^{-6}	X
	LDLR	LDL	1.72×10^{-4}	2.20×10^{-5}	1.72×10^{-4}	2.20×10^{-5}	4.01×10^{-5}	1.50×10^{-2}
		CHOL	3.47×10^{-4}	2.97×10^{-6}	3.47×10^{-4}	2.97×10^{-6}	5.67×10^{-6}	5.79×10^{-3}
		LDL,CHOL	3.24×10^{-5}	2.99×10^{-7}	3.24×10^{-5}	2.99×10^{-7}	7.83×10^{-7}	X

Table S.22: **Results of Association Analysis of Three Traits of the Trinity Students Study in the Region of An Enzyme Gene Using the F-approximation Based on Pillai-Bartlett Trace.** The associations that attain a threshold significance of $P < 3.1 \times 10^{-6}$ are highlighted in red [Liu et al. 2014]. The results of "Basis of both GVF and $\beta_\ell(t)$" were based on smoothing both GVF and genetic effect functions $\beta_\ell(t)$ of model (S.4), the results of "FPCA Approach" were based on functional principal component analysis (FPCA) model [Fan et al., 2013], the results of "Basis of beta-Smooth Only" were based on smoothing $\beta_\ell(t)$ only approach of model (7), and the p-values of SKAT-O Were Based of R Package SKAT. Abbreviation: GVF = Genetic Variant Function.

Traits	**P-values of the F-approximation Based on Pillai-Bartlett Trace**						**P-values of SKAT-O**
	Basis of both GVF and $\beta_\ell(t)$		**FPCA**	**Basis of beta-Smooth Only**		**Multivariate Model (1)**	
	B-spline Basis	**Fourier Basis**		**B-spline Basis**	**Fourier Basis**		
A	1.73×10^{-13}	7.89×10^{-13}	1.54×10^{-15}	1.73×10^{-13}	7.89×10^{-13}	2.84×10^{-12}	2.16×10^{-10}
B	3.44×10^{-13}	1.80×10^{-11}	1.58×10^{-13}	3.44×10^{-13}	1.80×10^{-11}	1.23×10^{-10}	2.72×10^{-5}
C	1.11×10^{-11}	9.91×10^{-10}	8.67×10^{-11}	1.11×10^{-11}	9.91×10^{-10}	3.78×10^{-9}	1.25×10^{-5}
(A, B)	2.14×10^{-20}	3.14×10^{-18}	3.00×10^{-21}	2.14×10^{-20}	3.14×10^{-18}	7.67×10^{-17}	X
(A, C)	1.08×10^{-17}	9.53×10^{-16}	9.29×10^{-18}	1.08×10^{-17}	9.53×10^{-16}	4.46×10^{-15}	X
(B, C)	6.54×10^{-15}	9.51×10^{-12}	1.19×10^{-14}	6.54×10^{-15}	9.51×10^{-12}	1.05×10^{-10}	X
(A, B, C)	2.30×10^{-21}	5.87×10^{-18}	3.74×10^{-21}	2.30×10^{-21}	5.87×10^{-18}	1.56×10^{-16}	X