

Evidence for Age As a Modifier of Genetic Associations for Lipid Levels

Logan Dumitrescu^{1,2}, Kristin Brown-Gentry¹, Robert Goodloe¹, Kimberly Glenn¹, Wenjian Yang³, Nancy Kornegay³, Ching-Hon Pui⁴, Mary V. Relling³ and Dana C. Crawford^{1,2*}

¹Center for Human Genetics Research, Vanderbilt University, Nashville, TN 37232, USA

²Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA

³Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

⁴Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Summary

In order to identify novel genetic variants that influence plasma lipid concentrations, we performed a genome-wide association study (GWAS) comprised of 411 children under 18 years of age, ascertained at St. Jude Children's Research Hospital, all of whom were of European, African, or Mexican descent. Promising associations ($p < 10^{-5}$) were subsequently examined in 1040 additional youths and 3508 adults from the Third National Health and Nutrition Examination Survey (NHANES III), a diverse population-based study. Three genotype–phenotype associations replicated in NHANES III youths and three associated in NHANES III adults at $p < 0.05$; however, no single association was significant in both youths and adults. The most significant association ($p = 0.009$) in NHANES III youths was between low-density lipoprotein cholesterol (LDL-C) and intronic rs2429917 among participants of African descent. Given the known age dependency of lipid levels, we also tested for gene–age interactions in NHANES III participants across all ages. We identified a significant ($p = 0.024$) age-dependent association between *SGSM2* rs2429917 and LDL-C. This finding illustrates the utility of using children to discover novel variants associated with complex phenotypes and the importance of considering age-dependent genetic effects in association studies of lipid levels.

Keywords: Lipids, age, children, NHANES III, interaction, genome-wide association study

Introduction

Genome-wide association studies (GWAS) have identified many common genetic variants that contribute to normal variation in lipid traits. The largest GWAS meta-analysis to date, containing greater than 100,000 individuals of European descent, identified 95 loci that were independently associated with total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) (Teslovich et al., 2010). Combined, these loci explained only ~25–30% of the genetic variance, leaving a majority of the genetic variance unexplained.

Some have argued that the “missing heritability” observed for most complex human traits is due, in part, to unidentified genetic and nongenetic modifiers (McCarthy & Hirschhorn,

2008). Very few GWAS have tested for and identified genetic modifiers as the statistical methods and approaches have yet to be developed that will fully exploit these data (Cordell, 2009; Thomas, 2010). Also, few GWAS have collected the data necessary to test for modifiers, particularly environmental exposures that modify genetic associations.

Age, a potential modifier of the lipid trait distribution due to genetics, has been relatively unexplored. It is known that children and adolescents have different lipid distributions compared with adults (Hickman et al., 1998; Jolliffe & Janssen, 2006). Indeed, the National Cholesterol and Education Program (NCEP) provides a set of lipid and lipoprotein guidelines specific for children and adolescents (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992). Furthermore, the increase in the means and variances of most lipid parameters as humans age has been well established (Reilly et al., 1990; Ericsson et al., 1991; Heller et al., 1993; Boomsma et al., 1996; Snieder et al., 1997). How much of that age dependency is due to increases in environmental or genetic variance, or both, is still left to be determined. Preliminary

*Corresponding author: Dana C. Crawford, PhD, Center for Human Genetics Research, Vanderbilt University, 2215 Garland Avenue, 515B Light Hall, Nashville, Tennessee 37232. Tel: 615-343-7852; Fax: 615-322-6974; E-mail: crawford@chgr.mc.vanderbilt.edu

studies of complex traits such as body mass index (BMI) and systolic blood pressure suggest that, over time, while the contribution of genes to a phenotype remains relatively constant, environmental exposures are thought to increase the phenotypic variance (Brown et al., 2003). Thus, the accumulation effects of environmental exposures (i.e., diet, exercise, or smoking) may increase the phenotypic variance of complex traits such as BMI and lipids over time, making it more difficult to detect genotype–phenotype associations.

All GWAS studies of the lipid traits (HDL-C, LDL-C, and TG) have been performed in adults (>18 years of age) (Hindorff et al., 2010), a population exposed to environmental factors that influence lipid trait distributions for at least two decades. While these GWAS have been successful in identifying genetic variants associated with lipid traits, we propose that the study of younger participants such as children will identify additional variants masked by the age dependency well documented for these traits. To identify these novel variants associated with HDL-C, LDL-C, and TG levels, we performed a GWAS in 411 children of European, African, or Mexican descent ascertained at St. Jude Children's Research Hospital followed by replication in an additional dataset of youths from the Third National Health and Nutrition Examination Survey ($n = 1040$; NHANES III). Replicated genetic variants were formally tested for an interaction with age in the larger NHANES III dataset with adults ($n = 3508$). Gene discovery in children followed by testing for interactions in adults has identified one lipid trait-associated locus potentially modified by age and represents a modest step toward identifying the full genetic architecture of the lipid trait distributions in human populations.

Materials and Methods

Subjects in the discovery GWAS were drawn from the Total Therapy Study XV, a prospective study of children undergoing treatment for acute lymphoblastic leukemia (ALL), initiated in 2000 at St. Jude Children's Research Hospital to investigate outcomes of a new treatment approach (Pui et al., 2009). From June 2000 to October 2007, a total of 501 newly diagnosed patients aged 1–18 years were enrolled in the study, 411 of whom were evaluable for both serum lipids and genome-wide genotyping. Race/ethnicity was inferred using germ-line genotypes as previously described (Yang et al., 2009) based on genotype-based hierarchical clustering of patients and using data from the International HapMap cell lines [$n = 90$ Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain collection (CEU) (European); 90 Yoruba in Ibadan, Nigeria (YRI); 30 Han Chinese in Beijing, China (CHB); and 30 Japanese in Tokyo, Japan (JPT) (Asians)] as references. Patients exhibiting >90%

European, African, or Asian ancestries were classified as White, Black, or Asian, respectively; "Hispanic" status was inferred for those patients who were self-declared Hispanics who also had less than 90% of European, African, and Asian ancestries. The remaining patients were labeled as "other" and were not included in this analysis. Serum levels of HDL, LDL, and TG were measured directly using standard enzymatic techniques (Roche Diagnostics, IN) as described in Kawedia et al. (2011). All lipid measurements were taken on consolidation day 15 of treatment (at least 4 weeks from the last dose of glucocorticoid or asparaginase) and were nonfasting.

Participants in the St. Jude study were genotyped with either the Affymetrix GeneChip 500K and 100K or 6.0 array. Only common SNPs genotyped across both platforms were considered for analysis. SNPs were excluded from the analysis based on genotyping call rates per SNP ($\leq 98\%$) and minor allele frequency (≤ 0.01). From a total of 532,546 SNPs genotyped, 420,005 (79%) passed quality control thresholds. Tests of association between each genetic variant assuming an additive genetic model and natural log-transformed lipid trait were implemented in PLINK (Purcell et al., 2007) using linear regression, stratified by race/ethnicity.

Ascertainment of the NHANES III and method of DNA collection have been previously described (Centers for Disease Control and Prevention, 1994). Briefly, NHANES III is a cross-sectional survey that was conducted from 1988 to 1994 by the National Center for Health Statistics (NCHS) at the Centers for Disease Control (CDC) and Prevention. Like all the NHANES, NHANES III is a complex survey design that oversampled minorities (non-Hispanic Blacks and Mexican Americans), the young, and the elderly. Blood samples were obtained at a central location known as the Mobile Examination Center (MEC). Serum TC, TG, and HDL cholesterol were measured using standard enzymatic methods and LDL cholesterol was calculated using the Friedewald equation, with missing values assigned for samples with TG levels greater than 400 mg/dL (Center for Disease Control and Prevention, 1996). Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older. Race/ethnicity was self-identified as non-Hispanic White, non-Hispanic Black, Mexican American, or other. All NHANES III procedures were approved by the CDC Ethics Review Board and written informed consent was obtained from all participants. The present study was approved by the CDC Ethics Review Board. Because the study investigators did not have access to personal identifiers, this study was considered nonhuman subjects research by the Vanderbilt University Institutional Review Board.

A total of 65 SNPs were carried forward for replication in NHANES III and were targeted for genotyping using Sequenom's iPLEX Gold assay on the MassARRAY platform (San Diego, CA) according to manufacturer's instructions

(www.sequenom.com). One SNP, rs4811011, failed assay design and was subsequently genotyped using Applied Biosystem's TaqMan (Foster City, CA). A total of 57 SNPs were genotyped successfully (genotyping call rate >95%); however, five SNPs failed blinded duplicates quality control measures as required by CDC and were excluded from further analyses. The remaining SNPs were tested for deviations from Hardy–Weinberg Equilibrium (HWE) within each racial/ethnic group and all had HWE p -values >0.01 in at least two populations, as required by CDC. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through CDC.

Genotype–phenotype tests of associations for the 52 SNPs and the three natural log-transformed lipid traits were performed in SAS version 9.2 (SAS Institute, Cary, NC) using linear regression, assuming an additive genetic model, and unweighted for selection and nonresponse biases. Single-SNP analyses were stratified by race/ethnicity, adjusted for age and sex and, for the adult cohort (>18 years), were limited to fasting individuals not on lipid-lowering medications. Tests for interactions between age and genotype were also considered in these regression models, excluding the adjustment for age. Replication and interaction associations were deemed significantly significant at $p < 0.05$. Data were accessed remotely from the CDC's Research Data Center (RDC) in Hyattsville, Maryland using Analytic Data Research by Email (ANDRE).

Differences in lipid levels between the St. Jude and NHANES III cohort and between genotype groups were tested in STATA 10 using a standard two-sample t -test with unequal variances. Manhattan plots were produced using code provided by <http://gettinggeneticsdone.blogspot.com>.

Results

Study Population Characteristics

Across all three racial/ethnic subpopulations, St. Jude children with ALL had consistently lower mean LDL-C, and higher

mean HDL-C and mean TG compared to NHANES III youths (Table 1; Supplementary Table S1). The means of some lipid levels (HDL-C in Whites and Hispanics LDL-C in Whites and Hispanics) were statistically significantly different between the two studies of children, with differences in means of up to 18.8 mg/dL (Table 1: Supplementary Table S1). These differences in lipid levels are most likely related to the older mean age and/or the larger percentage of females in NHANES III youths compared with St. Jude children rather than health status, as none of the medications used during the consolidation stage of therapy for ALL are known to affect lipid levels, *per se*.

When comparing the two NHANES III replication datasets, not surprisingly (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992; Hickman et al., 1998; Snieder et al., 1999; Jolliffe & Janssen, 2006), adults had different lipid profiles compared to youths with statistically higher mean LDL-C and TG levels across all subpopulations (Table 1 and Supplementary Table S2). Mean HDL-C levels were significantly higher in White adults and significantly lower in Hispanic adults compared to youths, although the differences were small (2.3 and 1.7 mg/dL, respectively).

Discovery GWAS Results

To identify novel common variants associated with HDL-C, LDL-C, and TG levels in children, we performed a genome-wide association screen in all three subpopulations of the St. Jude cohort. No SNP surpassed genome-wide statistical significance ($p < 5 \times 10^{-7}$) (Wellcome Trust Case Control Consortium, 2007) for any of the analyses; however, four SNPs approached that threshold and were significantly associated at $p < 1 \times 10^{-6}$ (Fig. 1 and Supplementary Table S3). In Whites, two SNPs (intergenic rs4742455 and *USH2A* rs17026635) were associated with transformed HDL-C ($p = 5.18 \times 10^{-6}$ and $p = 6.33 \times 10^{-6}$, respectively). In Hispanics, intergenic rs7790255 and *GBAS/MRPS17* rs15892 were associated with transformed TG ($p = 2.38 \times 10^{-6}$ and $p = 9.42 \times 10^{-6}$, respectively).

Table 1 Participant characteristics.

Characteristic	GWAS cohort St. Jude children			Replication cohort 1 NHANES III youths			Replication cohort 2 NHANES III adults		
	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics
Total n	282	66	63	241	439	360	1386	1039	1083
Age (years)	7.0 \pm 5	7.2 \pm 4	6.4 \pm 4	14.6 \pm 2	14.8 \pm 2	14.8 \pm 2	52.2 \pm 21	39.8 \pm 16	40.0 \pm 17
Female (%)	41.1	43.9	47.6	54.4	54.0	52.2	60.6	58.7	48.9
HDL-C (mg/dL)	55.1 \pm 30	59.0 \pm 27	67.9 \pm 50	47.2 \pm 10	52.6 \pm 12	49.1 \pm 11	50.3 \pm 15	53.6 \pm 17	47.4 \pm 13
LDL-C (mg/dL)	84.5 \pm 27	81.3 \pm 37	87.4 \pm 35	92.9 \pm 28	98.5 \pm 27	94.4 \pm 24	129.4 \pm 38	123.0 \pm 40	120.1 \pm 34
Triglycerides (mg/dL)	108.2 \pm 79	88.3 \pm 47	121.7 \pm 101	97.6 \pm 56	80.0 \pm 43	99.0 \pm 57	145.3 \pm 111	107.6 \pm 82	151.8 \pm 110

Values are listed as mean \pm SD unless otherwise indicated.

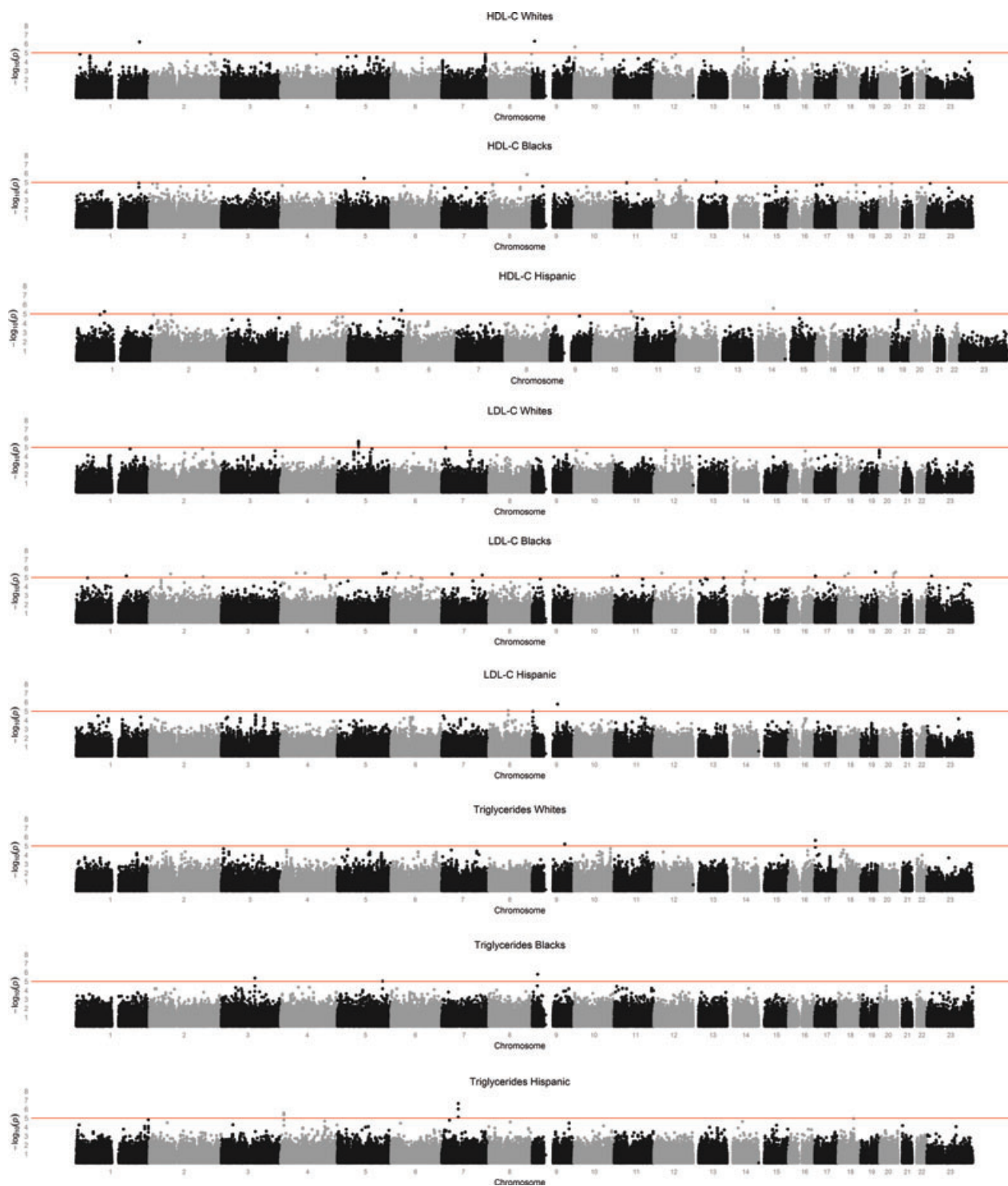


Figure 1 Discovery GWAS association results. The figure summarizes genome-wide association results for three lipid traits, HDL-C, LDL-C, and triglycerides in the three populations, Whites, Blacks, and Hispanics. Values are plotted as $-\log_{10}P$ value on the y -axis. Chromosomal location is designated on the x -axis. The red line corresponds to a $p = 1 \times 10^{-5}$, the threshold used for determining which associations would be carried further for replication.

The latter two SNPs are in very high linkage disequilibrium (LD; $r^2 = 0.91$) in the Hispanic subpopulation and therefore likely represent the same association. Across all three lipid traits and all three racial/ethnic groups, 65 associations were significantly associated at $p < 1 \times 10^{-5}$ and carried forward for replication in NHANES III youths.

Replication in NHANES III Youths and Adults

Next, we tested if the novel associations observed in the discovery GWAS would replicate in an independent study of children. For the replication study, a total of 1040 youths (12–18 years of age) in NHANES III were genotyped for 65 SNPs, 52 of which passed assay design and quality control measures (see Methods and Materials). Three of the 52 associations replicated at $p < 0.05$ in NHANES youths (Table 2). Two SNPs, *SGSM2* rs2429917 and intergenic rs12190789, were associated with transformed LDL-C at $p = 0.009$ and $p = 0.047$, respectively, and *CD96* rs16858329 was associated with transformed TG at $p = 0.048$. All three of the significant associations were in Blacks and the directions of effects were concordant with those of the discovery cohort, although they were of smaller magnitude.

We also conducted a second replication study in a large cohort of fasting adults (>18 years of age) in NHANES III to determine if associations initially discovered in children would generalize to adults. Despite the considerable increase in sample size, none of the three associations that replicated in NHANES youths were also associated in NHANES adults (Table 2). There were, however, three significant associations in NHANES adults that were not replicated in NHANES youths (TG and intergenic rs6477578 in Whites, LDL and *FRMD3* rs10868008 in Hispanics, and LDL and *FRMD3* rs1140077 in Hispanics; Table 2). The former

association was the most significant at $p = 0.006$, yet the direction of effect was in the opposing direction compared with the GWAS discovery study of St. Jude children. The latter two associations most likely represent the same association as rs10868008 and rs114007 are in complete LD ($r^2 = 1.00$) in our discovery GWAS among Hispanic samples.

Age as a Potential Modifier

As mentioned previously and shown in Table 1 and Supplementary Table S2, mean HDL-C, LDL-C, and TG levels differ significantly between adults and children/youths. Given that three associations were discovered and replicated in children but failed to generalize to adults, we explored the hypothesis that genetic associations observed for the lipid traits are modified by age. After testing for SNP-age interactions, we observed a significant interaction ($p = 0.024$) between age and *SGSM2* rs2429917 with transformed LDL-C in Blacks.

Up until about the fourth decade of life, participants homozygous for the major allele of rs2429917 (C, frequency = 0.96) had consistently lower LDL-C levels compared to CT heterozygotes (Table 3) with the largest significant difference in the 12–21 age group (mean difference = 17.25 mg/dL, $p = 0.002$). Any trend in mean differences is harder to detect later in life due to smaller sample sizes for participants with the CT genotype (n ranges from 0 to 7; Table 3). However, we did note that participants with the CC genotype had significantly lower mean LDL-C (mean difference = 24.14 mg/dL, $p = 0.016$) in the older 72–81 age group, a reverse of what we observed in the 12- to 21-year-olds. These results lead us to speculate that early in life (and possibly much later in life) one's LDL-C concentration is dependent, in some small part, on one's genotype at the rs2429917 locus.

Table 2 Significant replication results in NHANES III.

Trait	Population	SNP	Gene/ location	Coded allele (Freq)	GWAS cohort St. Jude children		Replication cohort 1 NHANES III youths		Replication cohort 2 NHANES III adults	
					Beta (SE)	<i>p</i> -value	Beta (SE)	<i>p</i> -value	Beta (SE)	<i>p</i> -value
LDL-C	Blacks	rs2429917	<i>SGSM2</i> intron	T (0.04)	−1.45 (0.30)	7.01×10^{-6}	−0.20 (0.08)	0.009	0.02 (0.05)	0.662
LDL-C	Blacks	rs12190789	Intergenic	T (0.03)	−1.46 (0.29)	3.32×10^{-6}	−0.17 (0.08)	0.047	−0.04 (0.05)	0.413
TG	Blacks	rs16858329	<i>CD96</i> intron	T (0.41)	−0.36 (0.07)	4.29×10^{-6}	−0.06 (0.03)	0.048	0.03 (0.02)	0.127
TG	Whites	rs6477578	Intergenic	G (0.27)	0.25 (0.05)	6.13×10^{-6}	0.03 (0.05)	0.425	−0.06 (0.02)	0.006
LDL	Hispanics	rs10868008	<i>FRMD3</i> intron	C (0.04)	−0.69 (0.13)	1.66×10^{-6}	−0.06 (0.06)	0.368	−0.08 (0.05)	0.049
LDL	Hispanics	rs11140077	<i>FRMD3</i> intron	A (0.06)	−0.69 (0.13)	1.66×10^{-6}	−0.06 (0.06)	0.368	−0.08 (0.05)	0.049

Significant ($p \leq 0.05$) associations in the replication cohorts are bolded and italicized. Coded allele frequencies are based on the whole genetic NHANES III in the designated racial/ethnic population.

rs2429917 Genotype					
Age group (years)	CC <i>n</i> = 734		CT <i>n</i> = 52		<i>p</i> -value
	<i>N</i>	LDL-C mg/dL (mean ± SD)	<i>N</i>	LDL-C mg/dL (mean ± SD)	
12 to <21	157	100.25 ± 26.8	16	82.75 ± 19.0	0.003
≤21 to <30	126	113.79 ± 37.8	8	103.25 ± 32.1	0.398
≤30 to <39	169	115.80 ± 35.3	12	111.41 ± 24.7	0.574
≤39 to <48	125	124.80 ± 40.4	7	139.71 ± 52.5	0.486
≤48 to <57	42	132.12 ± 42.7	2	152.50 ± 37.5	0.580
≤57 to <63	40	151.00 ± 47.1	3	120.67 ± 63.5	0.497
≤63 to <72	43	138.55 ± 44.1	2	153.50 ± 55.9	0.770
≤72 to <81	22	145.86 ± 42.2	2	170.00 ± 2.8	0.016
≥81	10	126.20 ± 41.6	0	–	–

Significant ($p \leq 0.05$) differences in mean are bolded and italicized.

Discussion

The aim of this study was to discover novel variants associated with lipid levels in children and to test if those associations were also significant in adults. We performed a GWAS of children undergoing treatment for ALL ascertained at St. Jude Children's Research Hospital, and attempted replication in an independent population of youths and adults from NHANES III. Three of the 52 lipid-genotype associations tested in NHANES III children replicated at $p < 0.05$, including intronic *SGSM2* rs2429917 at $p = 0.009$. However, these associations did not generalize to NHANES III adults. We also identified a genotype \times age interaction with *SGSM2* rs2429917 for transformed LDL-C in Blacks, supporting a genetic basis for the differences observed in lipid levels in children compared to older individuals.

Age, as a modifier of genetic association studies, has only recently been highlighted in the literature. In one study of the 100K data of the longitudinal Framingham Heart Study, Lasky-Su et al. (2008) describe an age-dependent association between *ROBO1* and obesity where the association was stronger among the pediatric cohorts compared with adult cohorts. Although Lasky-Su et al. do not speculate on the mechanism behind the age-dependent interaction, it is interesting to note that heritability estimates for obesity in children tend to be higher (Haworth et al., 2008) compared with estimates in adults (Brown et al., 2003). Somewhat consistent with these observations in obesity are the observations of heritability estimates for the lipid traits. That is, some studies have found that heritability of select lipid levels tends to decrease with age (Heller et al., 1993; Beekman et al., 2002). However, the review by Snieder et al. (1999) concluded that heritability estimates for HDL-C, LDL-C, and TG remain relatively stable with age.

While the magnitude of the genetic influence on lipid metabolism may not change with age, the importance of dif-

ferent genes may. In other words, different genes may be expressed in childhood and adolescence compared to adulthood. In regards to lipid metabolism, longitudinal twin studies support this possibility (Williams & Wijesiri, 1993; Friedlander et al., 1997; Nance et al., 1998), and an extended parent-twin study determined that different genes are expressed in adolescence compared to adulthood (Snieder et al., 1997). It is also possible that the same genes function throughout life, but are expressed at different levels depending on the decade of life. Supporting this latter hypothesis is the observation that younger patients heterozygous for *ABCA1* mutations that cause Tangier disease have significantly higher HDL-C levels than older patients heterozygous for *ABCA1* mutations (Clee et al., 2000). There is evidence that normal *ABCA1* function increases over time (Clee et al., 2000), which suggests that pronounced HDL-C deficiency between age groups may be highlighting the inability of heterozygotes to increase their *ABCA1* activity with increasing age.

Given the proposed and observed differences between children and adults for these traits, we purposefully performed a discovery study in children as this subset may allow for discovery of novel genes associated with lipid levels compared with previously published GWAS from adults. The most promising novel candidate as a result of this study is rs2429917, located in the intron of *SGSM2*, or small G protein signaling modulator 2. *SGSM2* is ubiquitously expressed in various tissues, including the liver, and as the name implies, acts as a modulator of G-protein signaling through its interaction with a subfamily of RAS proteins (Yang et al., 2007). Proteins involved in G protein-mediated signal transduction are associated with a number of cellular mechanisms, including differentiation and proliferation. It is also important to note that rs2429917 is located in a fairly gene-dense region of chromosome 17, including *SMG6*, *SRR*, *TSR1*, *MNT*, and *METTL16*, all within ~100 kb flanking *SGSM2*. Based

Table 3. Mean LDL-C levels in Blacks, stratified by rs2429917 genotype and age group. Age was categorized into age groups spanning 10 years, beginning at age 12 (the age at which NHANES III began collecting genetic data on participants). For each age group, mean and standard deviations of LDL-C concentrations were calculated for participants with a CC genotype or a CT genotype at the rs2429917 locus, separately. Data for the TT genotype are not presented here due to small sample size ($n = 3$). *t*-Tests with unequal variances were calculated to test for differences in mean LDL-C between genotypes, within the same age group.

on their biological functions, none of these neighboring genes are compelling candidates for association with lipid metabolism. However, *SMG6* is an intriguing candidate given its essential association with telomerase activity (Reichenbach et al., 2003; Snow et al., 2003) and, thereby, aging. Deletion of *Est1p* (the *S. cerevisiae* homolog to human *SMG6*) in yeast leads to ever-shorter telomeres over time despite functional telomerase activity (Lundblad & Szostak, 1989). Telomere shortening occurs in all mitotic tissues (excluding germline tissue) as humans age and has been shown to contribute to mortality in many age-related diseases, including heart disease (Cawthon et al., 2003). Although these findings require further study, it is interesting to speculate that these data may point to the involvement of previously unsuspected pathways contributing to lipid metabolism.

This study had several limitations, including that the discovery GWAS was underpowered due to its small sample size. Even with our largest population ($n = 282$ in Whites), and an allele frequency of 5%, we had 80% power to detect only large effect sizes ($R^2 > 11\%$) at genome-wide significance. The majority of published lipid GWAS-identified variants have small effect sizes and explain only a small percentage of the variance of lipid traits in the population (Manolio, 2009; Teslovich et al., 2010). However, to our knowledge, no GWAS has been performed on children with lipid levels; therefore, it is unknown whether the effect size and/or the significance of these well-known variants remain constant over a lifetime.

Another limitation is that the discovery cohort consisted of children undergoing treatment for ALL. Although medications administered for treatment of ALL are not known to affect lipid levels, side effects (such as loss of appetite or nausea) of these medications may cause nutritional differences that affect lipid levels. While this possibility was not directly measured and, therefore, cannot be completely ruled out, we observed that the children in the St. Jude cohort maintained healthy appetites (data not shown).

Despite the small discovery sample size, we were able to detect nominally significant associations, of which three replicated at $p < 0.05$ in an independent dataset. Furthermore, examination of genetic variation known to influence lipids in European-descent populations demonstrates that true associations can be detected in spite of the low power of the study. That is, of the 26 established lipid-associated SNPs in 23 genes (including *CETP*, *LPL*, *GCKR*, *APOB*, etc.; Table S4), we detected seven associations with p -values ≤ 0.05 . As an example, rs328 is a nonsynonymous SNP in *LPL* and has been shown to have a reproducible effect (~ 19 mg/dL in one study) (Kathiresan et al., 2008) on lowering TG. In our GWAS of children, rs1741102, a proxy for rs328 ($r^2 = 1$ in HapMap CEU), was associated at $p = 0.02$ with $\beta = -0.17$, corresponding to -21.3 mg/dL, which is consistent in both the previously reported magnitude and direction of effect.

A benefit of using NHANES III data is that it allows for genetic studies in a large, diverse population with a wide age range. However, it is a cross-sectional study. Since our data suggest that there may be age-specific genetic influences, longitudinal data are necessary to derive further conclusions and to replicate this interaction.

Postmortem studies on young adults and children have shown that atherosclerosis starts early in life (Expert Panel Blood Cholesterol Levels Child Adolesc, 1992), even though clinical symptoms usually do not manifest until decades later. The potential temporal nature of factors, both genetic and nongenetic, which contributes to cardiovascular disease, is important for better understanding of the etiology of the disease. While it is often assumed in genotype–phenotype association studies that genetic effects are stable over a lifetime, the possibility of important age-effects should not be ignored when studying the genetics of lipid metabolism.

Acknowledgements

Genotyping for replication in NHANES III was supported in part by the Vanderbilt Institute for Clinical and Translational Research (VICTR), as well as a Clinical and Translational Science Award (CTSA) grant (1UL1 RR024975–01 from the National Center for Research Resources and the National Institutes of Health [NIRR/NIH]). We would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data and Dr. Jean Cai at St. Jude for serum lipids measurements. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The NHANES DNA samples are stored and plated by the Vanderbilt DNA Resources Core. Genotyping was performed by Ping Mayo and Dr. Nathalie Schnetz-Boutaud in the laboratory of Dr. Jonathan Haines and Ping Mayo under the direction of Cara Sutcliffe in the Vanderbilt DNA Resources Core. This study was supported by NCI grant CA 21765, the NIH/NIGMS Pharmacogenomics Research Network (U01 GM92666), and the American Lebanese Syrian Associated Charities (ALSAC).

References

- Beekman, M., Heijmans, B. T., Martin, N. G., Pedersen, N. L., Whitfield, J. B., DeFaire, U., van Baal, G. C., Snieder, H., Vogler, G. P., Slagboom, P. E. & Boomsma, D. I. (2002) Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res* **5**, 87–97.
- Boomsma, D. I., Kempen, H. J., Gevers Leuven, J. A., Havekes, L., de Knijff, P. & Frants, R. R. (1996) Genetic analysis of sex and generation differences in plasma lipid, lipoprotein, and apolipoprotein levels in adolescent twins and their parents. *Genet Epidemiol* **13**, 49–60.
- Brown, W. M., Beck, S. R., Lange, E. M., Davis, C. C., Kay, C. M., Langefeld, C. D. & Rich, S. S. (2003) Age-stratified heritability estimation in the Framingham Heart Study families. *BMC Genet* **4**(Suppl 1), S32.
- Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A., & Kerber, R. A. (2003) Association between telomere length in

- blood and mortality in people aged 60 years or older. *Lancet* **361**, 393–395.
- Centers for Disease Control and Prevention. (1994) Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988–94. Centers for Disease Control and Prevention. Ref Type: Report, Bethesda, MD.
- Center for Disease Control and Prevention. (1996) The Third National Health and Nutrition Examination Survey, 1988–94. Reference manuals and reports (CD ROM). *US Department of Health and Human Services National Center for Health Statistics*.
- Clee, S. M., Kastelein, J. J., van Dam, M., Marcil, M., Roomp, K., Zwarts, K. Y., Collins, J. A., Roelants, R., Tamasawa, N., Stulc, T., Suda, T., Ceska, R., Boucher, B., Rondeau, C., DeSouich, C., Brooks-Wilson, A., Molhuizen, H. O., Frohlich, J., Genest, J., Jr. & Hayden, M. R. (2000) Age and residual cholesterol efflux affect HDL cholesterol levels and coronary artery disease in ABCA1 heterozygotes. *J Clin Invest* **106**, 1263–1270.
- Cordell, H. J. (2009) Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* **10**, 392–404.
- Ericsson, S., Eriksson, M., Vitols, S., Einarsson, K., Berglund, L. & Angelin, B. (1991) Influence of age on the metabolism of plasma low density lipoproteins in healthy males. *J Clin Invest* **87**, 591–596.
- Expert Panel Blood Cholesterol Levels Child Adolesc. (1992) National Cholesterol Education Program Report of the Expert Panel on blood cholesterol levels in children and adolescents. *Pediatrics* **89**, 495–501.
- Friedlander, Y., Austin, M. A., Newman, B., Edwards, K., Mayer-Davis, E. I. & King, M. C. (1997) Heritability of longitudinal changes in coronary-heart-disease risk factors in women twins. *Am J Hum Genet* **60**, 1502–1512.
- Haworth, C. M., Butcher, L. M., Docherty, S. J., Wardle, J. & Plomin, R. (2008) No evidence for association between BMI and 10 candidate genes at ages 4, 7 and 10 in a large UK sample of twins. *BMC Med Genet* **9**, 12.
- Heller, D. A., de, F. U., Pedersen, N. L., Dahlen, G., & McClearn, G. E. (1993) Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med* **328**, 1150–1156.
- Hickman, T. B., Briefel, R. R., Carroll, M. D., Rifkind, B. M., Cleeman, J. I., Maurer, K. R. & Johnson, C. L. (1998) Distributions and trends of serum lipid levels among United States children and adolescents ages 4–19 years: Data from the Third National Health and Nutrition Examination Survey. *Prev Med* **27**, 879–890.
- Hindorf, L. A., Junkins, H. A., Mehta, J. P., & Manolio, T. A. (2010) A catalog of published genome-wide association studies. *Natl Hum Genome Res Inst*. Available at <http://www.genome.gov/gwastudies>.
- Jolliffe, C. J. & Janssen, I. (2006) Distribution of lipoproteins by age and gender in adolescents. *Circulation* **114**, 1056–1062.
- Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burt, N. P., Rieder, M. J., Cooper, G. M., Roos, C., Voight, B. F., Havulinna, A. S., Wahlstrand, B., Hedner, T., Corella, D., Tai, E. S., Ordoas, J. M., Berglund, G., Vartiainen, E., Jousilahti, P., Hedblad, B., Taskiran, M. R., Newton-Cheh, C., Salomaa, V., Peltonen, L., Groop, L., Altshuler, D. M. & Orho-Melander, M. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189–197.
- Kawedia, J. D., Kaste, S. C., Pei, D., Panetta, J. C., Cai, X., Cheng, C., Neale, G., Howard, S. C., Evans, W. E., Pui, C. H. & Relling, M. V. (2011) Pharmacokinetic, pharmacodynamic and pharmacogenetic determinants of osteonecrosis in children with acute lymphoblastic leukemia. *Blood* **117**, 2340–2347.
- Lasky-Su, J., Lyon, H. N., Emilsson, V., Heid, I. M., Molony, C., Raby, B. A., Lazarus, R., Klanderma, B., Soto-Quiros, M. E., Avila, L., Silverman, E. K., Thorleifsson, G., Thorsteinsdottir, U., Kronenberg, F., Vollmert, C., Illig, T., Fox, C. S., Levy, D., Laird, N., Ding, X., McQueen, M. B., Butler, J., Ardlie, K., Papoutsakis, C., Dedoussis, G., O'Donnell, C. J., Wichmann, H. E., Celedón, J. C., Schadt, E., Hirschhorn, J., Weiss, S. T., Stefansson, K. & Lange, C. (2008) On the replication of genetic associations: Timing can be everything! *Am J Hum Genet* **82**, 849–858.
- Lundblad, V. & Szostak, J. W. (1989) A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633–643.
- Manolio, T. A. (2009) Cohort studies and the genetics of complex disease. *Nat Genet* **41**, 5–6.
- McCarthy, M. I. & Hirschhorn, J. N. (2008) Genome-wide association studies: Potential next steps on a genetic journey. *Hum Mol Genet* **17**, R156–R165.
- Nance, W. E., Bodurtha, J., Eaves, L. J., Hewitt, J., Maes, H., Segrest, J., Meyer, J., Neale, M. & Schieken, R. (1998) Models for the longitudinal genetic analysis of same-age twins: Application to HDL cholesterol. *Twin Res* **1**, 3–8.
- Pui, C. H., Campana, D., Pei, D., Bowman, W. P., Sandlund, J. T., Kaste, S. C., Ribeiro, R. C., Rubnitz, J. E., Raimondi, S. C., Onciu, M., Coustan-Smith, E., Kun, L. E., Jeha, S., Cheng, C., Howard, S. C., Simmons, V., Bayles, A., Metzger, M. L., Boyett, J. M., Leung, W., Handgretinger, R., Downing, J. R., Evans, W. E. & Relling, M. V. (2009) Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med* **360**, 2730–2741.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.
- Reichenbach, P., Höss, M., Azzalin, C. M., Nabholz, M., Bucher, P. & Lingner, J. (2003) A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. *Curr Biol* **13**, 568–574.
- Reilly, S. L., Kottke, B. A., & Sing, C. F. (1990) The effects of generation and gender on the joint distributions of lipid and apolipoprotein phenotypes in the population at large. *J Clin Epidemiol* **43**, 921–940.
- Snieder, H., van Doornen, L. J., & Boomsma, D. I. (1997) The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. *Am J Hum Genet* **60**, 638–650.
- Snieder, H., van Doornen, L. J., & Boomsma, D. I. (1999) Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: Lessons from twin studies. *Arterioscler Thromb Vasc Biol* **19**, 2826–2834.
- Snow, B. E., Erdmann, N., Cruickshank, J., Goldman, H., Gill, R. M., Robinson, M. O. & Harrington, L. (2003) Functional conservation of the telomerase protein Est1p in humans. *Curr Biol* **13**, 698–704.
- Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin Go, M., Jin Kim, Y., Lee, J. Y., Park,

- T., Kim, K., Sim, X., Tzee-Hee Ong, R., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y., Wright, A. F., Witteman, J. C., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R., McCarthy, M. I., McArdle, W., Masson, D., Martin, N. G., Marioni, F., Mangino, M., Magnusson, P. K., Lucas, G., Luben, R., Loos, R. J., Lokki, M. L., Lettre, G., Langenberg, C., Launer, L. J., Lakatta, E. G., Laaksonen, R., Kyvik, K. O., Kronenberg, F., König, I. R., Khaw, K. T., Kaprio, J., Kaplan, L. M., Johansson, A., Jarvelin, M. R., Janssens, A. C., Ingelsson, E., Igl, W., Kees Hovingh, G., Hottenga, J. J., Hofman, A., Hicks, A. A., Hengstenberg, C., Heid, I. M., Hayward, C., Havulinna, A. S., Hastie, N. D., Harris, T. B., Haritunians, T., Hall, A. S., Gyllenstein, U., Guiducci, C., Groop, L. C., Gonzalez, E., Gieger, C., Freimer, N. B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K. G., Döring, A., Dominiczak, A. F., Demissie, S., Deloukas, P., de Geus, E. J., de Faire, U., Crawford, G., Collins, F. S., Chen, Y. D., Caulfield, M. J., Campbell, H., Burtt, N. P., Bonnycastle, L. L., Boomsma, D. I., Boekholdt, S. M., Bergman, R. N., Barroso, I., Bandinelli, S., Ballantyne, C. M., Assimes, T. L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T. Y., Tai, E. S., Feranil, A. B., Kuzawa, C. W., Adair, L. S., Taylor, H. A. Jr., Borecki, I. B., Gabriel, S. B., Wilson, J. G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R. M., Mohlke, K. L., Ordovas, J. M., Munroe, P. B., Kooner, J. S., Tall, A. R., Hegele, R. A., Kastelein, J. J., Schadt, E. E., Rotter, J. I., Boerwinkle, E., Strachan, D. P., Mooser, V., Stefansson, K., Reilly, M. P., Samani, N. J., Schunkert, H., Cupples, L. A., Sandhu, M. S., Ridker, P. M., Rader, D. J., van Duijn, C. M., Pelttonen, L., Abecasis, G. R., Boehnke, M. & Kathiresan, S. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713.
- Thomas, D. (2010) Gene-environment-wide association studies: Emerging approaches. *Nat Rev Genet* **11**, 259–272.
- Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678.
- Williams, C. J. & Wijesiri, U. W. (1993) Lipid data from NHLBI veteran twins: Interpreting genetic analyses when model assumptions fail. *Genet Epidemiol* **10**, 551–556.
- Yang, H., Sasaki, T., Minoshima, S., & Shimizu, N. (2007) Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. *Genomics* **90**, 249–260.
- Yang, J. J., Cheng, C., Yang, W., Pei, D., Cao, X., Fan, Y., Pounds, S. B., Neale, G., Treviño, L. R., French, D., Campana, D., Downing, J. R., Evans, W. E., Pui, C. H., Devidas, M., Bowman, W. P., Camitta, B. M., Willman, C. L., Davies, S. M., Borowitz, M. J., Carroll, W. L., Hunger, S. P. & Relling, M. V. (2009) Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393–403.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 Differences in mean HDL-C, LDL-C, and triglycerides among St. Jude Children and NHANES III youths, stratified by race/ethnicity.

Table S2 Differences in mean HDL-C, LDL-C, and triglycerides among NHANES III adults and NHANES III youths, stratified by race/ethnicity.

Table S3 SNPs selected for replication based on GWAS in St. Jude children.

Table S4 Association results of lipid candidate genes and HDL-C, LDL-C, and TG in the St. Jude discovery GWAS, Whites.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 20 February 2011

Accepted: 13 May 2011