



Relative Contributions of Genes, Environment, and Interactions to Blood Lipid Concentrations in a General Adult Population

Michael C. Costanza¹, Eftihia Cayanis², Barbara M. Ross², Maria Sol Flaherty², Gabriela B. Alvin², Kamna Das², and Alfredo Morabia¹

¹ Division of Clinical Epidemiology, Geneva University Hospitals, Geneva, Switzerland.

² Columbia Genome Center and Departments of Genetics and Development and of Psychiatry, Columbia University, New York, NY.

Received for publication September 15, 2004; accepted for publication December 9, 2004.

The authors evaluated the contributions of nine genetic (G) variants (selected from 275 single nucleotide polymorphisms in 11 reverse cholesterol transport pathway genes), five environmental (E) factors (selected from 10), and $G \times G$, $E \times E$, and $G \times E$ interactions in explaining population variance of blood lipid concentrations. Total cholesterol, triglycerides, and high density lipoprotein (HDL) cholesterol were measured, and low density lipoprotein (LDL) cholesterol and HDL cholesterol/LDL cholesterol ratio were calculated in a population-based random sample of 1,543 men and women in Geneva, Switzerland, aged 35–74 years in 1999–2001. Explained variances (R^2) for HDL cholesterol/LDL cholesterol ratio, HDL cholesterol, and LDL cholesterol, respectively, were 34%, 33%, and 19%, decomposed into main effects of G (6%, 4%, and 5%) and E (25%, 28%, and 11%), with just 3%, 2%, and 3% due to $G \times G$, $E \times E$, and $G \times E$ interactions, respectively. Risk factor clustering was only moderate: 70% of study subjects had ≤ 3 variants, 75% had ≤ 2 environmental exposures, and 69% had ≤ 5 of both types of factors. Multiple genes with weak associations, together with more dominating environmental factors, are involved in determining blood lipid concentrations. Interactions added little explained variance. Increasing trends in hypercholesterolemia are attributable to environmental changes affecting populations as a whole. Reducing obesity and smoking and moderating alcohol intake in entire populations should remain the primary strategies for lipid control.

body mass index; environment; genes; genetics; hypercholesterolemia; lipids; metabolism; risk factors

Abbreviations: BMI, body mass index; E, environment or environmental; G, gene or genetic; HDL, high density lipoprotein; LDL, low density lipoprotein; SNP, single nucleotide polymorphism; T, tertile.

Both genetic (G) and environmental (E) factors control the concentrations of blood lipids, but the relative importance of these two broad categories of determinants has not been assessed systematically. Evidence also exists that synergistic or antagonistic interactions between determinants play a role in the etiology of common traits, but their actual contribution to the population variation of a complex phenotype, such as blood lipids, remains elusive (1). The presence of interactions would be of high clinical and public health relevance because it implies that specific subgroups of the population may benefit more than others from targeted prevention (e.g., physical activity (2)) or therapeutic (e.g., statins (3)) strategies.

Assessing the relative importance of interactions regarding common disorders is currently hampered by their formidable number between even a limited set of determinants (4) and by the paucity of biologic knowledge relative to their mechanisms. The reverse transport of cholesterol from peripheral tissues to the liver may be an exception in that it lends itself well to assessment of the relative roles of G and E main and interaction effects, for two reasons. There exists a rich biochemical and physiologic literature on the architecture of this pathway (5–8), and its environmental determinants have also been studied extensively (2, 9–22).

Correspondence to Prof. Alfredo Morabia, Division of Clinical Epidemiology, Geneva University Hospitals, 25 rue Micheli-du-Crest, 1211 Geneva 14, Switzerland (e-mail: a.morabia@hcuge.ch).

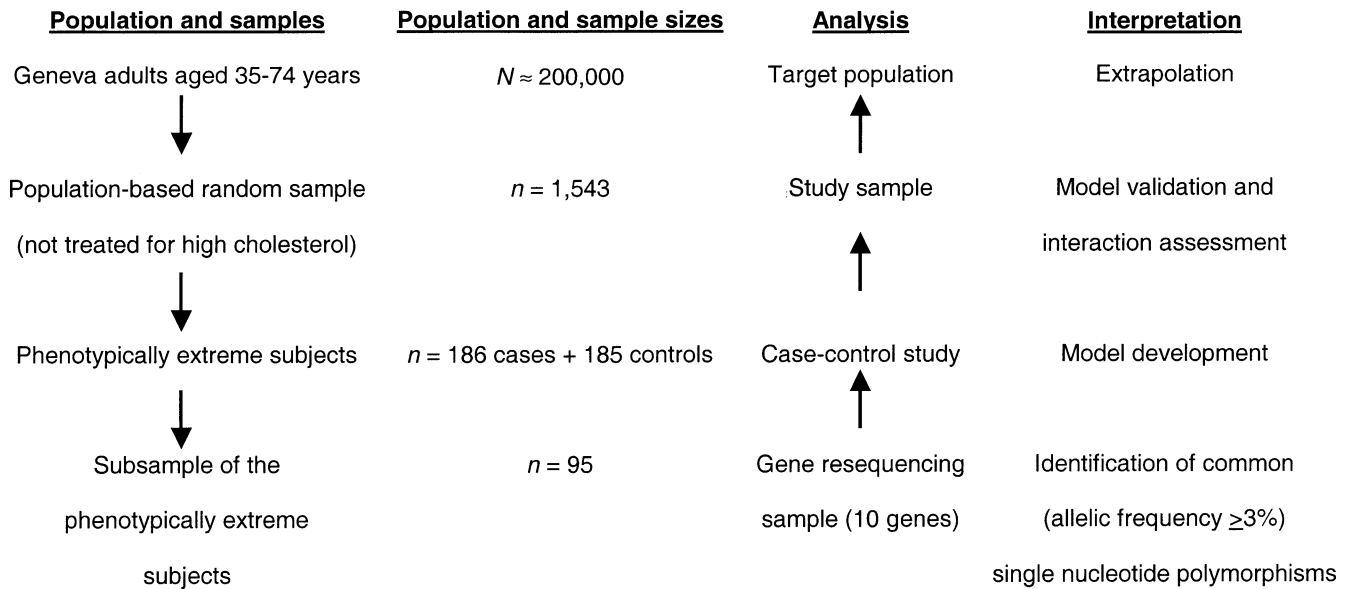


FIGURE 1. Sampling design and analytical strategy used to assess the relative contributions of genetic (G) variants, environmental (E) factors, and $G \times G$, $E \times E$, and $G \times E$ interactions to the variance in blood lipid concentrations in the Geneva, Switzerland, general adult study population. The population-based random sample of residents was obtained from January 1999 through February 2001. Refer to the introductory paragraphs of the text for further details.

Figure 1 depicts the overall study design strategy used to determine the relative contributions of G and E main and interaction effects to the population variance of blood lipid concentrations. A subset of subjects with extreme phenotypes was selected from the total random sample of 1,543 untreated subjects because their respective high density lipoprotein (HDL) cholesterol and low density lipoprotein (LDL) cholesterol concentrations were in the lowest (or highest) tertile (T1, 33.3rd percentile) and highest (or lowest) tertile (T3, 66.7th percentile), separately by gender. The tertile boundaries (mmol/liter) for HDL cholesterol and LDL cholesterol were, for men/women, as follows: HDL cholesterol—T1 = 1.06/1.33, T3 = 1.32/1.63; LDL cholesterol—T1 = 3.54/3.22, T3 = 4.30/4.04 (to convert to mg/dl, multiply by 38.6). For 186 subjects, HDL cholesterol was $<T1$ and LDL cholesterol was $>T3$ (low HDL, high LDL), considered an atherogenic phenotype (“cases”). At the other extreme were 185 subjects whose HDL cholesterol was $>T3$ and LDL cholesterol was $<T1$ (high HDL, low LDL), considered a nonatherogenic phenotype (“controls”). The case-control study subjects also tended to score in the corresponding extreme tertiles of triglycerides because HDL cholesterol, LDL cholesterol, and triglyceride levels are highly correlated in the extreme tails of their distributions. The entire coding DNA and the immediately adjacent noncoding DNA, including both 5′ and 3′ untranslated regions and a portion of each intron, were assayed first in a purposefully selected resequencing subsample of 95 of the case-control study subjects in whom the largest genetic variability was anticipated. Forty-eight subjects had a non-atherogenic lipid profile and were sedentary; hence, they

were expected to have a “protective” genetic constitution. The 47 other subjects had an atherogenic lipid profile but were physically active; hence, they were expected to have a “deleterious” genetic constitution.

Eleven genes and 10 environmental determinants of blood lipids involved in the reverse transport of cholesterol pathway were analyzed. The 11 genes were adenosine 5′-triphosphate-binding cassette protein 1 (*ABCA1*), apolipoprotein A1 (*APOA1*), apolipoprotein E (*APOE*), cholesteryl-ester transfer protein (*CETP*), endothelial lipase (*EL*), hepatic lipase (*HL*), lecithin-cholesterol acetyl transferase (*LCAT*), lipoprotein lipase (*LPL*), low density lipoprotein receptor (*LDLR*), phospholipid transfer protein (*PLTP*), and scavenger receptor class B type I (*SR-BI*). The 10 environmental factors were gender, age, educational level, country of birth, dietary fiber, dietary fat, physical activity, body mass index (BMI), cigarette smoking, and alcohol intake. Even though these factors also have genetic determinants, they are referred to as “environmental” because, as opposed to polymorphisms, their assessment is not DNA based.

A previous work established that a subset of nine genetic variants (mostly single nucleotide polymorphisms (SNPs), out of a total of 275), spanning seven candidate genes, together with five of the 10 environmental factors (BMI, alcohol intake, current cigarette smoking, gender, and age) explained the largest amount (65 percent) of the lipid variance for the subjects with extreme phenotypes (23). The present study extends this previous work by 1) implementing the model developed for the phenotypically extreme subjects to the whole population-based sample and 2) estimating the absolute and relative contributions of the

genetic, environmental, and interaction effects on blood lipid levels in a general European adult population.

MATERIALS AND METHODS

Subject selection

Geneva (city and canton, Switzerland) has a population of about 400,000 distributed over an area of 242 km². Potential survey subjects were randomly selected uniformly over the 26-month period from January 1999 through February 2001 to represent the 200,000 men and women noninstitutionalized residents aged 35–74 years, and they were asked to participate in a general population health survey. About 0.5 percent of the target population was sampled each year from the official residents' register. The study was approved by the relevant institutional ethics committees in Geneva and New York, and all participants provided informed written consent.

The intensive and standardized recruitment of a potential subject lasted from 2 weeks to 2 months. Eligible subjects were identified by using a standardized procedure from an annual list of residents established by the local government. All legal residents of the canton are registered. The only information from the list used in the survey (i.e., gender, age, and whether the person is of Swiss origin) is highly accurate. Stratified random sampling, based on the list by gender within 10-year age strata, was proportional to the corresponding population distributions. Selected subjects were mailed an invitation to participate; if they did not respond, as many as seven telephone attempts at different times on various days of the week were made. If telephone contact was unsuccessful, two more letters were mailed. Subjects not reached (15 percent of men and 19 percent of women) were replaced by using the same selection protocol. Routine survey quality monitoring had shown that these subjects no longer resided in the canton, so they were not eligible for the study. Subjects who refused to participate were not replaced, and participating subjects were not eligible for future surveys. The participation rate was 61 percent. The population-based sample used in the analyses comprised 1,543 subjects not under treatment for hypercholesterolemia.

Survey measurements

Each participant filled out several self-administered, standardized questionnaires covering risk factors for the major lifestyle chronic diseases, sociodemographic characteristics, educational and occupational histories, and reproductive history (women only). A semiquantitative food frequency questionnaire, previously developed and tested in the Geneva general adult population (24), asked about serving sizes and consumption frequencies of 80 food items organized by food groups during the 4 previous weeks, which were converted into daily energy, nutrient, and alcohol intakes in the analyses. A physical activity frequency questionnaire, also developed previously in the same target population and validated by using a heart rate monitor (25), measured total and activity-specific energy expendi-

tures based on which of 70 physical activities grouped by general type (e.g., occupational, housework, leisure time, sports) were performed in the past 7 days.

During a scheduled appointment at a mobile epidemiology clinic (housed in a special bus), the questionnaires were checked for completion by trained interviewers, and a physical examination was performed. After participants removed their shoes and heavy outerwear, their weight was measured by using a medical scale (precision, 0.5 kg), and height was measured with a medical gauge (precision, 1 cm), from which BMI was calculated.

Laboratory measurements

Total plasma cholesterol, HDL cholesterol, and triglycerides were assayed (mmol/liter) in fasting blood (Bayer Technicon Diagnostics, Brussels, Belgium, with monthly quality control checks performed by the Swiss Center for Quality Control in Clinical Chemistry and Hematology (Geneva)). Total genomic DNA was extracted from ethylenediaminetetraacetic acid (EDTA) blood (Gentra Puregene blood kit; BioConcept, Allschwil, Switzerland). LDL cholesterol (mmol/liter) was calculated as (cholesterol – HDL cholesterol – triglycerides/2.2) (26).

SNPs were assayed by using modified template-directed dye-terminator incorporation with fluorescence polarization (TDI-FP) detection (Acycloprime-FP SNP Detection Kit; PerkinElmer Life and Analytical Sciences, Inc., Shelton, Connecticut). Information on primers to amplify and detect SNPs is available from the authors on request.

Statistical analyses

Linear regression models were used to evaluate the joint contributions (measured by the squared multiple correlation coefficient, R^2) of the nine genetic variants, five environmental factors, and their interaction effects to the variation in log(HDL cholesterol/LDL cholesterol ratio), HDL cholesterol, and LDL cholesterol (log data reported as geometric mean). The nine variants (across seven genes) comprised two *ABCA1* SNPs (exons 32b. + 30 and 50b.3038), *APOE2* ($\epsilon 2/\epsilon 2$ or $\epsilon 2/\epsilon 3$), two *HL* SNPs (exon 1b. – 280 promoter and 3b. – 279), *LPL S447X* (exon 9.b99), *LDLR2* (rs2228671 on exon 2), *PLTP* exon 1b. + 26, and *SR-BI A350A* (exon 8b.41). Each genetic variant was coded 0 if the rare variant was absent and 1 if present (as heterozygote/homozygote). This coding assumes that the rare allele effect is dominant. Because homozygotes for the rare allele were extremely rare, it is essentially equivalent to an additive model, consistent with most evidence about the genetics of many quantitative traits.

The five environmental factors were gender (reference = women vs. men), cigarette smoking (reference = (never smoker (<100 lifetime cigarettes) + former smoker (quit ≥ 1 year preinterview)) vs. current smoker), alcohol intake (grams of alcohol/day) (reference = none vs. low intake (men/women: 1–40/1–20) vs. (medium + high intakes) (men/women: $\geq 41/\geq 21$)), age (years), and BMI (weight (kg)/height (m)²). The first three covariates were

expressed with four dummy variables, and the last two were continuous. In some analyses, BMI was recoded as normal weight (BMI <25) vs. overweight (only) ($25 \leq \text{BMI} < 30$) vs. obese (BMI ≥ 30), and age was recoded as 1) <55 vs. ≥ 55 years or 2) <50 vs. ≥ 50 years.

A systematic analytical strategy based on stepdown selections of $G \times G$, $E \times E$, and $G \times E$ interaction effects first within and then between each of the three types of interactions for each lipid outcome was used to obtain the final models. Details of these procedures are described in the Appendix.

All final p values, regression coefficients, and cumulative and total R^2 's were obtained by using Monte Carlo bootstrap procedures (resampling with replacement) (27). A final $p \leq 0.05$ ($p = 0.10$) value was considered statistically (or borderline) significant. Sampling variations in the R^2 estimates were also assessed with 95 percent bias-corrected and accelerated percentile confidence intervals (28). All bootstrapped estimates were based on 2,000 replicates each, twice the minimum number recommended for bootstrap confidence intervals (28).

The final models were further augmented with the other five environmental factors that had been *removed* in the modeling process performed for only the extreme-phenotype subjects (23) and were bootstrapped as above. These other five environmental factors were 1) educational level (reference group = primary (<9 years of schooling), secondary, university (≥ 13 years of schooling and a Swiss baccalaureate degree)); 2) country of birth (reference = (Switzerland + all other (\approx one third France, remainder (mostly northern Europe) <5 percent each)), Mediterranean (Italy, Spain, Portugal)); 3) total dietary fat (%); 4) dietary fiber (g/day); and 5) daily energy expenditure (reference = active (≥ 10 percent total energy expended in physical activities requiring four or more times the basal metabolic rate, sedentary (<10 percent expended in physical activities requiring four or more times the basal metabolic rate) (29)). The first two covariates were expressed with three dummy variables, and the last three were continuous (in the descriptive analyses, dietary fat and dietary fiber were coded as below the overall median (percentile (P)50) vs. $\geq P50$). These additional analyses were used to assess any residual contributions of the other environmental factors, which reasonably might have been expected to play a role in determining blood lipid concentrations in the general population despite contraindications when focusing on only the extreme-phenotype subjects.

RESULTS

HDL cholesterol, LDL cholesterol, and log(HDL cholesterol/LDL cholesterol ratio) outcome measures

HDL cholesterol and LDL cholesterol were only modestly negatively correlated in the total sample ($r = -0.15$) and by gender (men: $r = -0.07$; women: $r = -0.15$). As expected, both HDL cholesterol and LDL cholesterol, respectively, were strongly correlated with log(HDL cholesterol/LDL cholesterol ratio) (total: $r = 0.77$, -0.73 ; men: $r = 0.74$, -0.70 ; women: $r = 0.73$, -0.77).

Associations between environmental factors and blood lipids

Table 1 describes in detail all 10 environmental characteristics of the population-based sample and their effects on blood lipid measurements. For the five environmental factors retained in the modeling process in the Morabia et al. (23) study, age and gender differences in lipid levels were in the expected directions; overweight or obesity and cigarette smoking were individually associated with higher serum total cholesterol, LDL cholesterol, and triglycerides but with a lower HDL cholesterol and HDL cholesterol/LDL cholesterol ratio, whereas alcohol intake was associated with higher serum total cholesterol, HDL cholesterol, and HDL cholesterol/LDL cholesterol ratio.

For the five environmental factors removed by the modeling process in the Morabia et al. (23) study, sedentarity was associated with lower HDL cholesterol and HDL cholesterol/LDL cholesterol ratio but higher triglycerides; more dietary fiber was associated with lower triglycerides; Mediterranean birth was associated with lower HDL cholesterol and HDL cholesterol/LDL cholesterol ratio; university education was associated with higher HDL cholesterol but lower triglycerides; and dietary fat was not associated with any of the lipid outcomes.

Clustering of genetic and environmental determinants of blood lipids

The symmetrical distributions of the numbers of SNPs, of (relevant) environmental exposures (counting one each for age ≥ 55 years, BMI ≥ 25 kg/m² (overweight), current smoking, or any alcohol intake, but excluding gender), and of both types of factors harbored by the study subjects are shown in Web Appendix figure A1. (This figure and four supplementary Appendix tables (each referred to as "Web Appendix table" in the text) are posted on the *Journal's* website (<http://aje.oupjournals.org/>.) There was practically no correlation ($r = -0.004$) between the numbers of SNPs and the numbers of environmental factors. The median numbers were three out of a maximum of nine SNPs, two out of four environmental exposures (apart from gender), and five out of 13 genetic and environmental factors combined. About 70 percent of the study subjects had three SNPs or fewer, 75 percent had two or fewer environmental exposures, and 69 percent had five or fewer of both types of factors.

Cumulative genetic, environmental, and interaction effects on blood lipids

Consistent with the model developed for the phenotypically extreme subjects, the nine SNPs and five environmental factors had significant effects (bootstrapped $p < 0.05$ either individually or through interaction) on all or most of the three lipid outcomes (table 2). For HDL cholesterol/LDL cholesterol ratio, all but one (*SR-BI A350A*) of the nine SNPs and all five environmental factors had significant effects. For HDL cholesterol alone, all nine SNPs and all five environmental factors had significant effects. For LDL cholesterol alone, six

TABLE 1. Environmental factors and their effects on blood lipids in the population-based study sample (n = 1,543), Geneva, Switzerland, 1999–2001

Environmental factor and category	No.	Serum total cholesterol (mmol/liter)* (mean (SE†))	HDL cholesterol (mmol/liter)* (mean (SE))	LDL cholesterol (mmol/liter)* (mean (SE))	HDL cholesterol/LDL cholesterol ratio (mean (SE)‡)	Triglycerides (mmol/liter)* (mean (SE)‡)
<i>Included in the final logistic model (23)</i>						
Gender						
Men (a)	751	5.79 (0.04)	1.21 (0.01)	3.92 (0.03)	0.31 (0.01)	1.25 (0.02)
Women (b)	792	5.70 (0.03)	1.50 (0.01)	3.71 (0.03)	0.41 (0.01)	0.96 (0.02)
p_{b-a} §		0.073	<0.0001	<0.0001	<0.0001	<0.0001
Age group (years)						
35–54 (a)	1,051	5.60 (0.03)	1.35 (0.01)	3.70 (0.03)	0.36 (0.01)	1.19 (0.02)
55–74 (b)	492	6.04 (0.04)	1.38 (0.01)	4.06 (0.04)	0.34 (0.02)	1.06 (0.02)
p_{b-a} ¶		<0.0001	0.067	<0.0001	<0.0001	<0.0001
Body mass index#						
Normal (a)	884	5.66 (0.03)	1.43 (0.01)	3.74 (0.03)	0.38 (0.01)	0.96 (0.02)
Overweight (b)	520	5.85 (0.04)	1.27 (0.01)	3.91 (0.04)	0.32 (0.02)	1.27 (0.02)
Obese (c)	139	5.88 (0.08)	1.20 (0.03)	3.94 (0.07)	0.30 (0.03)	1.42 (0.04)
p_{b-a}		0.0010	<0.0001	0.0008	<0.0001	<0.0001
p_{c-b}		0.74	0.013	0.71	0.037	0.015
p_{c-a}		0.015	<0.0001	0.013	<0.0001	<0.0001
Cigarette smoking						
Current (a)	399	5.84 (0.05)	1.30 (0.02)	3.90 (0.04)	0.33 (0.02)	1.22 (0.02)
Not current (b)	1,144	5.71 (0.03)	1.37 (0.01)	3.78 (0.03)	0.36 (0.01)	1.06 (0.01)
p_{b-a}		0.027	<0.0001	0.025	<0.0001	<0.0001
Alcohol intake#						
None (a)	191	5.70 (0.07)	1.24 (0.02)	3.85 (0.06)	0.32 (0.03)	1.15 (0.04)
Low (b)	1,177	5.72 (0.03)	1.36 (0.01)	3.80 (0.03)	0.36 (0.01)	1.08 (0.01)
Medium + high (c)	175	5.93 (0.07)	1.45 (0.02)	3.85 (0.07)	0.38 (0.03)	1.17 (0.04)
p_{b-a}		0.77	<0.0001	0.49	0.0003	0.13
p_{c-b}		0.027	0.0007	0.49	0.032	0.037
p_{c-a}		0.0094	<0.0001	0.99	<0.0001	0.65
<i>Removed from the final logistic model (23)</i>						
Educational level#						
University (a)	601	5.69 (0.04)	1.37 (0.01)	3.78 (0.04)	0.36 (0.01)	1.03 (0.02)
Secondary (b)	850	5.80 (0.03)	1.35 (0.01)	3.84 (0.03)	0.35 (0.01)	1.14 (0.02)

Table continues

of the nine SNPs and four environmental factors (excluding alcohol intake) had significant effects. Of the nine genetic variants, four (*APOE2*, the *PLTP* SNP, and the two *ABCA1* SNPs) had significant effects on all three lipid outcomes; a fifth, *SR-BI A350A*, had significant effects on HDL cholesterol and LDL cholesterol and a borderline effect on HDL cholesterol/LDL cholesterol ratio. Another three variants (*LPL S447X* and the two *HL* SNPs) had significant effects on the HDL cholesterol/LDL cholesterol ratio and on HDL cholesterol. The ninth *LDLR* SNP had significant effects on HDL cholesterol and LDL cholesterol.

On the whole, because the SNPs had been sampled from a metabolic pathway primarily for HDL cholesterol, the model predicted LDL cholesterol less well. Remarkably, of the 19

additional interactions (excluding hierarchically required terms) across all three lipid outcomes, all but two were confirmed in the final bootstrapped models. The two non-confirmed interactions were of borderline statistical significance (*SR-BI* × alcohol medium + high intakes × gender for HDL cholesterol/LDL cholesterol ratio ($p = 0.098$) and *SR-BI A350A* × age for HDL cholesterol ($p = 0.061$)).

The total R^2 was 33.7 percent for the HDL cholesterol/LDL cholesterol ratio, 33.0 percent for HDL cholesterol, and 18.5 percent for LDL cholesterol. These three total R^2 's, respectively, were decomposed into the main effects of the nine genetic variants (6 percent, 4 percent, and 5 percent) and the five environmental factors (25 percent, 28 percent, and 11 percent), with the remainder (3 percent,

TABLE 1. Continued

Environmental factor and category	No.	Serum total cholesterol (mmol/liter)* (mean (SE))	HDL cholesterol (mmol/liter)* (mean (SE))	LDL cholesterol (mmol/liter)* (mean (SE))	HDL cholesterol/LDL cholesterol ratio (mean (SE)‡)	Triglycerides (mmol/liter)* (mean (SE)‡)
Primary (c)	92	5.60 (0.10)	1.29 (0.03)	3.74 (0.09)	0.34 (0.04)	1.12 (0.05)
ρ_{b-a}		0.046	0.26	0.27	0.099	<0.0001
ρ_{c-b}		0.072	0.061	0.23	0.55	0.72
ρ_{c-a}		0.42	0.018	0.62	0.17	0.12
Country of birth#						
Switzerland (a)	895	5.76 (0.03)	1.37 (0.01)	3.82 (0.03)	0.36 (0.01)	1.08 (0.02)
Other (b)	436	5.75 (0.05)	1.36 (0.02)	3.81 (0.04)	0.35 (0.02)	1.12 (0.02)
Mediterranean (c)	210	5.68 (0.07)	1.28 (0.02)	3.80 (0.06)	0.34 (0.02)	1.12 (0.03)
ρ_{b-a}		0.91	0.35	0.93	0.33	0.18
ρ_{c-b}		0.42	0.0065	0.92	0.14	0.99
ρ_{c-a}		0.33	0.0002	0.86	0.017	0.30
Dietary fat (P†50 = 34.7%)						
<P50 (a)	771	5.75 (0.03)	1.36 (0.01)	3.80 (0.03)	0.36 (0.01)	1.11 (0.02)
≥P50 (b)	772	5.74 (0.03)	1.35 (0.01)	3.82 (0.03)	0.35 (0.01)	1.08 (0.02)
ρ_{b-a}		0.91	0.54	0.72	0.51	0.32
Dietary fiber (P50 = 14.9 g/day)						
<P50 (a)	772	5.75 (0.03)	1.36 (0.01)	3.80 (0.03)	0.36 (0.01)	1.12 (0.02)
≥P50 (b)	771	5.73 (0.03)	1.35 (0.01)	3.82 (0.03)	0.35 (0.01)	1.07 (0.02)
ρ_{b-a}		0.71	0.79	0.60	0.55	0.048
Sedentarity#						
Active (a)	581	5.72 (0.04)	1.40 (0.01)	3.78 (0.04)	0.37 (0.01)	1.03 (0.02)
Sedentary (b)	962	5.76 (0.03)	1.33 (0.01)	3.83 (0.03)	0.35 (0.01)	1.14 (0.01)
ρ_{b-a} **		0.40	<0.0001	0.27	0.0003	<0.0001
Gender-/age-group adjusted mean		5.74	1.36	3.81	0.35	1.09

* To convert from mmol/liter to mg/dl, multiply serum total cholesterol, high density lipoprotein (HDL) cholesterol, and low density lipoprotein (LDL) cholesterol by 38.6, triglycerides by 88.

† SE, standard error; P, percentile.

‡ Geometric mean (SE of log data).

§ p values were adjusted for age (continuous).

¶ p values were adjusted for gender, body mass index, cigarette smoking, alcohol intake, educational level, country of birth, dietary fat, and dietary fiber.

Refer to the text for more information.

** p values were adjusted for gender and age (continuous).

2 percent, and 3 percent) being jointly due to the $G \times G$, $G \times E$, and $E \times E$ interaction effects.

Detailed 2×2 and $2 \times 2 \times 2$ tables specifically illustrating the additional two- and three-way interaction effects on the lipid outcome measurements are available in Web Appendix tables A1, A2, and A3.

Stability of the final model after augmentation by the other five environmental factors

When the final model with the interactions shown in table 2 was augmented with the other five environmental factors that had been removed during the modeling process in the

Morabia et al. (23) study, the previous regression coefficients and p values remained essentially unchanged, as shown in Web Appendix table A4. Although sedentarity was significantly associated with the HDL cholesterol/LDL cholesterol ratio ($p = 0.044$) and HDL cholesterol alone ($p = 0.017$), and (log)fiber intake showed borderline significant associations with the HDL cholesterol/LDL cholesterol ratio ($p = 0.051$) and LDL cholesterol ($p = 0.065$), all of the additional contributions to the explained variances were marginal. Specifically, total R^2 increased by only 0.7 (to 34.4) percent for the HDL cholesterol/LDL cholesterol ratio, by only 0.8 (to 33.8) percent for HDL cholesterol, and by only 0.7 (to 19.2) percent for LDL cholesterol.

TABLE 2. Contributions of nine G* variants, five E* factors, and G × G, E × E, and G × E interactions to prediction of log(HDL* cholesterol/LDL* cholesterol ratio), HDL cholesterol, and LDL cholesterol in the population-based study sample (n = 1,543), Geneva, Switzerland, 1999–2001

SNPs* and covariates	log(HDL cholesterol/LDL cholesterol ratio)			HDL cholesterol (mmol/liter)			LDL cholesterol (mmol/liter)		
	Cumulative R ² †,‡	Final model		Cumulative R ² †,‡	Final model		Cumulative R ² †,‡	Final model	
		Coefficient‡	p value‡		Coefficient‡	p value‡		Coefficient‡	p value‡
Intercept		0.275	0.40		1.62	≈0		0.521	0.51
Body mass index (kg/m ²)	13.8	-0.056	1.8 × 10 ⁻⁶	11.6	-0.024	≈0	4.4	0.117	0.00014
APOE2*	16.8	0.863	1.4 × 10 ⁻⁶	12.4	0.086	0.00078	7.4	-0.490	7.2 × 10 ⁻¹⁴
PLTP* 1b. + 26 ("P1 + 26")	16.9	-0.039	0.14	12.5	0.184	0.078	7.4	0.227	0.036
Alcohol intake: medium + high	17.1	0.125	0.083	13.4	0.203	5.4 × 10 ⁻¹¹	7.6	-0.058	0.51
Alcohol intake: low	17.1	0.093	0.0023	13.5	0.101	4.3 × 10 ⁻⁶	7.7	-0.051	0.44
Current cigarette smoker	18.6	-0.113	2.2 × 10 ⁻⁹	15.6	-0.120	5.3 × 10 ⁻¹⁰	7.9	0.148	0.0037
ABCA1* 50b.3038 ("ABC50")	19.0	0.067	0.0042	15.8	0.042	0.043	8.2	0.734	0.022
LPL* S447X ("S447X")	19.4	0.062	0.0017	16.4	0.069	0.0017	8.3	-0.006	0.90
Age (years)	20.5	-0.019	0.0012	16.9	0.0003	0.82	13.3	0.050	0.00083
Gender (women)	28.8	0.514	1.5 × 10 ⁻⁸	29.3	0.339	≈0	13.9	-1.232	1.3 × 10 ⁻⁷
ABCA1 32b. + 30 ("ABC32")	29.1	-0.102	5.5 × 10 ⁻⁵	29.7	-0.057	0.0040	14.1	0.176	0.0087
LDLR2* (rs2228671)	29.5	0.021	0.34	29.7	0.098	0.0056	14.6	-0.089	0.12
HL* 1b. - 280 ("HL1-280")	29.8	0.022	0.25	30.4	0.039	0.020	14.6	0.020	0.65
HL 3b. - 279 ("HL3-279")	30.0	0.068	0.0075	30.6	0.046	0.024	14.7	-0.010	0.13
HL 3b. - 279 × gender	30.2	-0.071	0.040	30.8	-0.051	0.092	14.8	0.093	0.31
SR-B* A350A ("A350A")	30.3	-0.153	0.12	30.9	-0.132	0.12	15.0	0.051	0.49
SR-BI A350A × gender	30.6	-0.063	0.090	31.1	-0.063	0.056	15.1	0.131	0.17
<i>Additional interactions (required terms)§</i>									
(Gender × alcohol intake medium + high)	30.6	0.168	0.19						
(A350A × alcohol intake medium + high)	30.7	0.087	0.31						
A350A × alcohol intake medium + high × gender	30.9	-0.236	0.098						
ABC32 × LDLR2	31.5	0.178	0.00054						
HL1-280 × P1 + 6	32.0	0.127	0.0027						
Gender × age	32.5	-0.004	0.025						
APOE2 × age	32.9	-0.006	0.045						
Age × body mass index	33.2	0.001	0.0072						
APOE2 × body mass index	33.5	-0.015	0.025						
A350A × age	33.7	0.003	0.061						
HL1-280 × P1 + 26				31.7	0.123	0.0025			
A350A × age				31.9	0.004	0.016			
LDLR2 × S447X				32.2	-0.107	0.0047			

S447X × smoker	32.5	0.096	0.026	
LDLR2 × A350A	32.8	-0.095	0.013	
P1 + 26 × age	33.0	-0.004	0.031	
Gender × age				16.7
P1 + 26 × A350A				17.2
ABC32 × LDLR2				17.7
Age × body mass index				18.1
ABC50 × body mass index				18.5
95% BC _s * confidence limits for total R ²	28, 35			13, 20

* G, gene or genetic; E, environment or environmental; HDL, high density lipoprotein; LDL, low density lipoprotein; SNPs, single nucleotide polymorphisms; APOE2, ε2 allele(s) of the apolipoprotein gene present; PLTP, phospholipid transfer protein gene; ABCA1, adenosine 5'-triphosphate-binding cassette protein 1 gene; LPL, lipoprotein lipase gene; LDLR2, rs2228671 SNP on exon 2 of the low density lipoprotein receptor gene; HL, hepatic lipase gene; SR-BI, scavenger receptor class B type I gene; BC_s, bias corrected and accelerated.

† Percentage. Variables listed above the additional interactions were entered into the linear regression models in the same order as in Morabia et al. (23). The additional interactions were then entered in order of importance based on their increases to R².

‡ Bootstrapped R²'s and final model regression coefficients were based on 2,000 replicates. Two-tailed, test-based p value (assuming H₀: regression coefficient = 0) was derived from the mean and standard error of the bootstrap replicates.

§ Two-way interactions required to preserve the hierarchical structure of three-way interactions.

DISCUSSION

This study used a statistical model developed a priori that related genetic and nongenetic factors to blood lipid levels and produced three main findings. First, altogether, the nine genetic variants, five environmental factors, and their interactions explained about one third of the total population variance for the HDL cholesterol/LDL cholesterol ratio and HDL cholesterol, which is substantial and indicates that the selection of factors was on target. Second, the environmental factors were the most important determinants of the HDL cholesterol/LDL cholesterol ratio and HDL cholesterol. Finally, interactions contributed only a few extra percentage points beyond the genetic and environmental main effects.

In addition to either HDL cholesterol or LDL cholesterol alone, the combined phenotype based on the HDL cholesterol/LDL cholesterol ratio was used as an outcome measurement because biochemical, physiologic, and pharmacologic data all support the notion that the 11 genes originally targeted by Morabia et al. (23) converge to moderate both HDL cholesterol and LDL cholesterol levels. Moreover, although HDL cholesterol and LDL cholesterol are relatively independent of each other in the general population, they are highly (negatively) correlated in the tails of their distributions, where the working model for the present study was developed.

There is a growing body of evidence that combinations of multiple genes harboring predisposing alleles can be identified that, together, contribute significantly to the population variance of lipid levels (30). Mootha et al. (31) have shown that combining the expressions from 22,000 genes of previously defined pathways may increase signal relative to noise and improve statistical power. Yang et al. (32) used simulations and empirical data to demonstrate that the information from several genetic and environmental factors could be combined to improve prediction of a multifactorial disease. From an evolutionary perspective, adaptive pressures may select against any single gene variant (allele) explaining a large fraction of the population phenotypic variance. Instead, alleles explaining the variance of a population trait are more likely to be distributed among several, or numerous, genes within a pathway or larger network of gene products. In addition, substantial exposure heterogeneity in the present population also supported the idea that different combinations of determinants modulated blood lipid levels. Typically, a person tended to be exposed to about a median of only five of the 13 total factors (apart from gender) included in the models. The present work of estimating and separating out the relative roles of some pathway-related genetic and environmental factors and their epistatic effects in determining blood lipid levels demonstrates the success of the latter approaches when applied in a general population setting.

The more dominant role of the environmental factors in determining blood lipid concentrations makes good sense. Blood cholesterol levels have been fluctuating over the last decades in almost all populations of the world (33, 34), and this phenomenon can be explained only by changes in exposure to factors totally or in part "environmental," such

as diet, physical activity, smoking, alcohol intake, or overweight/obesity. The very modest contributions of interactions are unexpected, however. One explanation could be that their effects were underestimated in the analysis. The particular genetic variants identified in the final model represent only a small fraction of all variants that operate in the population as a whole. Interactions may also involve factors that are outside of the studied model and that therefore could not have been accounted for in the analysis. Up to three-way interactions were feasible to analyze given the sample size, but there may be important higher-order interactions. Finally, the set of genetic and nongenetic factors analyzed were all predictors of lipid levels before interactions were investigated. Therefore, some determinants, which are related to lipid levels only when interacting with other factors, may have been missed.

However, there are also reasons why the results probably do reflect the relative contributions of genes, environment, and interactions to HDL cholesterol levels. The main analytical strategy in the population-based random sample began with nine variants across seven of 11 reverse transport of cholesterol pathway genes and five of 10 environmental factors, which are considered the main determinants of blood lipid levels (23). All polymorphisms in the exons and flanking regions of the 11 genes were surveyed, and state-of-the-art methods were used to assess environmental exposures such as diet (including alcohol intake) (24), physical activity (25), and smoking (35). The final model implicated BMI, gender, and *APOE*, which are the strongest individual determinants of HDL cholesterol. Remarkably, the cumulative effect of the other eight trait-related gene variants managed to approach that of only *APOE* alone. Lastly, the findings are consistent with the literature: even though work exists suggesting that some genetically defined subgroups of the population may benefit more than others from specific interventions or therapies (2, 3), on the whole no strong signal is emerging from the wealth of available studies that cholesterol-reducing preventive strategies should target subgroups of the population defined by specific $G \times G$, $G \times E$, or $E \times E$ interactions.

It is of note that sedentarity and fiber intake were related to lipid levels in the full population-based sample but not for the phenotypically extreme subjects alone. To check the possibility that these associations might have been confounded by the strong effect of BMI in the case-control study of Morabia et al. (23), the final logistic model for the phenotypically extreme subjects was augmented by sedentarity and log(fiber intake), but *BMI was excluded*. The results continued to indicate that neither sedentarity ($p = 0.18$) nor fiber intake ($p = 0.60$) was associated with case-control status in the absence of BMI. The absence of effects of lipid intake, country of birth, and educational level in the full sample can probably be explained by the extreme homogeneity of this entirely urban population, in which migrants do not maintain the ethnic characteristics of their population of origin.

The results of this systematic investigation of genes, environmental factors, and their interactions in explaining the variance of lipoprotein levels indicate that 1) BMI, smoking, and alcohol intake are strong predictors of blood

lipid levels; and 2) their effects are only marginally modified by genetic background. These findings have tremendous clinical and public health implications. They imply that the documented trends of the growing prevalence of hypercholesterolemia in the population of origin of the study sample (34) are attributable to modifiable lifestyle behaviors affecting the population as a whole and that they are not driven mainly by genetic subgroups. Confirmation of these results by further studies in other parts of the world would provide evidence that, at least at present, prevention strategies should target every individual patient, or the whole population—that is, not specific genetic subgroups. These strategies should focus on reducing obesity, in particular through physical activity (36), elimination of cigarette smoking, and moderation of alcohol intake.

ACKNOWLEDGMENTS

This project was supported by grants from the Swiss National Science Foundation (3200-054097 and 3200-068275.02) and by funding from the Columbia Genome Center. This study was part of the Geneva Program for Metabolic Disorders.

REFERENCES

1. Templeton AR. Epistasis and complex traits. In: Wolf J, Brodie EI, Wade M, eds. *Epistasis and the evolutionary process*. New York, NY: Oxford University Press, 2000:41–57.
2. Bernstein MS, Costanza MC, James RW, et al. Physical activity may modulate effects of ApoE genotype on lipid profile. *Arterioscler Thromb Vasc Biol* 2002;22:133–40.
3. Chasman DI, Posada D, Subrahmanyam L, et al. Pharmacogenetic study of statin therapy and cholesterol reduction. *JAMA* 2004;291:2821–7.
4. Wade M. Epistasis as a genetic constraint within populations and an accelerant of adaptive divergence among them. In: Wolf J, Brodie EI, Wade M, eds. *Epistasis and the evolutionary process*. New York, NY: Oxford University Press, 2000:213–31.
5. Fielding CJ, Fielding PE. Molecular physiology of reverse cholesterol transport. *J Lipid Res* 1995;36:211–28.
6. Reichl D, Miller NE. Pathophysiology of reverse cholesterol transport. Insights from inherited disorders of lipoprotein metabolism. *Arteriosclerosis* 1989;9:785–97.
7. Rader DJ, Maugeais C. Genes influencing HDL metabolism: new perspectives and implications for atherosclerosis prevention. *Mol Med Today* 2000;6:170–5.
8. Tall AR, Jiang X, Luo Y, et al. 1999 George Lyman Duff Memorial Lecture: lipid transfer proteins, HDL metabolism, and atherogenesis. *Arterioscler Thromb Vasc Biol* 2000;20:1185–8.
9. Ordovas JM. Gene-diet interaction and plasma lipid responses to dietary intervention. *Biochem Soc Trans* 2002;30:68–73.
10. Ye SQ, Kwiterovich PO Jr. Influence of genetic polymorphisms on responsiveness to dietary fat and cholesterol. *Am J Clin Nutr* 2000;72:1275S–1284S.

11. Denke MA. Connections between obesity and dyslipidaemia. *Curr Opin Lipidol* 2001;12:625–8.
12. Rossouw JE. Hormones, genetic factors, and gender differences in cardiovascular disease. *Cardiovasc Res* 2002;53:550–7.
13. Betteridge DJ. Lipid-lowering trials in diabetes. *Curr Opin Lipidol* 2001;12:619–23.
14. Couillard C, Despres JP, Lamarche B, et al. Effects of endurance exercise training on plasma HDL cholesterol levels depend on levels of triglycerides: evidence from men of the Health, Risk Factors, Exercise Training and Genetics (HERITAGE) Family Study. *Arterioscler Thromb Vasc Biol* 2001;21:1226–32.
15. Pearson TA. Alcohol and heart disease. *Circulation* 1996;94:3023–5.
16. Craig WY, Palomaki GE, Haddow JE. Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *BMJ* 1989;298:784–8.
17. Leon AS. Effects of exercise conditioning on physiologic precursors of coronary heart disease. *J Cardiopulm Rehabil* 1991;11:46–57.
18. Marrugat J, Elosua R, Covas MI, et al. Amount and intensity of physical activity, physical fitness, and serum lipids in men. *Am J Epidemiol* 1996;143:562–9.
19. Ponjee GA, Janssen EM, Hermans J, et al. Effects of long-term exercise of moderate intensity on anthropometric values and serum lipids and lipoproteins. *Eur J Clin Chem Clin Biochem* 1995;33:121–6.
20. Pronk NP, Crouse SF, O'Brien BC, et al. Acute effects of walking on serum lipids and lipoproteins in women. *J Sports Med Phys Fitness* 1995;35:50–8.
21. Ready AE, Drinkwater DT, Ducas J, et al. Walking program reduces elevated cholesterol in women postmenopause. *Can J Cardiol* 1995;11:905–12.
22. Ekstedt B, Jonsson E, Johnson O. Influence of dietary fat, cholesterol and energy on serum lipids at vigorous physical exercise. *Scand J Clin Lab Invest* 1991;51:437–42.
23. Morabia A, Cayanis E, Costanza MC, et al. Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors. *Hum Mol Genet* 2003;12:2733–43.
24. Morabia A, Bernstein M, Kumanyika S, et al. Development and validation of a semi-quantitative food questionnaire based on a population survey. (In French). *Soz Praventivmed* 1994;39:345–69.
25. Bernstein M, Sloutskis D, Kumanyika S, et al. Data-based approach for developing a physical activity frequency questionnaire. *Am J Epidemiol* 1998;147:147–54.
26. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972;18:499–502.
27. Insightful Corporation. S-PLUS 2000 guide to statistics. Seattle, WA: Insightful Corporation, 1999.
28. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York, NY: Chapman & Hall, 1993.
29. Bernstein MS, Costanza MC, Morabia A. Association of physical intensity levels with overweight and obesity in a population-based sample of adults. *Prev Med* 2004;38:94–104.
30. Knoblauch H, Bauerfeind A, Toliat MR, et al. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet* 2004;13:993–1004.
31. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267–73.
32. Yang Q, Khoury MJ, Botto L, et al. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. *Am J Hum Genet* 2003;72:636–49.
33. Ford ES, Mokdad AH, Giles WH, et al. Serum total cholesterol concentrations and awareness, treatment, and control of hypercholesterolemia among US adults: findings from the National Health and Nutrition Examination Survey, 1999 to 2000. *Circulation* 2003;107:2185–9.
34. Galobardes B, Costanza MC, Bernstein MS, et al. Trends in risk factors for the major “lifestyle-related diseases” in Geneva, Switzerland, 1993–2000. *Ann Epidemiol* 2003;13:537–40.
35. Morabia A, Bernstein MS, Curtin F, et al. Validation of self-reported smoking status by simultaneous measurement of carbon monoxide and salivary thiocyanate. *Prev Med* 2001;32:82–8.
36. Morabia A, Costanza MC. Does walking 15 minutes per day keep the obesity epidemic away? Simulation of the efficacy of a populationwide campaign. *Am J Public Health* 2004;94:437–40.

APPENDIX

Analytical Strategy for Investigating $G \times G$, $E \times E$, and $G \times E$ Interactions

$G \times G$, $E \times E$, and $G \times E$ interaction effects on log(HDL cholesterol/LDL cholesterol ratio), HDL cholesterol, and LDL cholesterol were investigated with a systematic analytical strategy based on stepdown selections of interaction terms augmenting the variables previously identified as statistically significant determinants of blood lipid concentrations among subjects with jointly extreme HDL cholesterol and LDL cholesterol phenotypes (23). For convenience, age was recoded as <50 years versus ≥ 50 years in these analyses.

First, each of the three types of interaction effects was analyzed separately. The following sections describe these effects.

$G \times G$ interactions

The nine SNPs were first divided into two subgroups, G_2 and G_7 . G_2 comprised the *HL3b* – 279 and *SR-BI A350A* SNPs that showed significant two-way interactions with gender (23); G_7 comprised the remaining seven SNPs.

Step 1: For *each* SNP, the previously identified variables were augmented with all possible three-way $G_7 \times G_7 \times$ gender and $G_7 \times G_2 \times$ gender interactions + any two-way $G_7 \times$ gender and $G_7 \times G_2$ interactions required for a hierarchical model + all possible remaining two-way $G_7 \times G_2$, $G_7 \times G_7$, and $G_2 \times G_2$ interactions.

Step 2: $p = 0.10$ stepdown selection was applied *only* to the step 1 model *three-way* interactions.

Step 3: $p = 0.10$ stepdown selection was applied *only* to the step 2 model *two-way* interactions.

Step 4: Interactions retained in *all* step 3 models were fitted in an overall model.

Step 5: $p = 0.10$ stepdown selection was applied *only* to the step 4 model interactions.

E × E interactions

Step 1: The previously identified variables were augmented with all possible two-way E × E interactions.

Step 2: $p = 0.10$ stepdown selection was applied *only* to the step 1 model interactions.

G × E interactions

Step 1: For *each* of the four E factors of age, BMI, smoking, and alcohol intake separately, the previously identified variables were augmented with all nine possible

two-way G × E interaction terms (gender and the two corresponding three-way G2 × E × gender interactions were automatically included).

Step 2: $p = 0.10$ stepdown selection was applied *only* to the interactions for each step 1 model.

Step 3: Interactions retained in *all* step 1 models were fitted in an overall model.

Step 4: $p = 0.10$ stepdown selection was applied *only* to the step 3 model interactions.

Finally, the simultaneous effects of the G × G, E × E, and G × E interactions were assessed by augmenting the previously identified variables with all of the interaction terms retained as described above *pooled together* and applying $p = 0.05$ stepdown selection to *only* those interactions.