ARTICLE

The genetic architecture of fasting plasma triglyceride response to fenofibrate treatment

Jennifer A Smith^{*,1}, Donna K Arnett², Reagan J Kelly¹, Jose M Ordovas³, Yan V Sun¹, Paul N Hopkins⁴, James E Hixson⁵, Robert J Straka⁶, James M Peacock⁷ and Sharon LR Kardia¹

¹Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA; ²Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA; ³Nutrition and Genomics Laboratory, Jean Mayer-United States Department of Agriculture Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA; ⁴Cardiovascular Genetics Research, University of Utah, Salt Lake City, UT, USA; ⁵Human Genetics Center, University of Texas Health Science, Houston, TX, USA; ⁶Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, Minneapolis, MN, USA; ⁷Department of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Metabolic response to the triglyceride (TG)-lowering drug, fenofibrate, is shaped by interactions between genetic and environmental factors, yet knowledge regarding the genetic determinants of this response is primarily limited to single-gene effects. Since very low-density lipoprotein (VLDL) is the central carrier of fasting TG, identifying factors that affect both total TG and VLDL–TG response to fenofibrate is critical for predicting individual fenofibrate response. As part of the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study, 688 individuals from 161 families were genotyped for 91 single-nucleotide polymorphisms (SNPs) in 25 genes known to be involved in lipoprotein metabolism. Using generalized estimating equations to control for family structure, we performed linear modeling to investigate whether single SNPs, single covariates, SNP–SNP interactions, and/or SNP–covariate interactions had a significant association with the change in total fasting TG and fasting VLDL–TG after 3 weeks of fenofibrate treatment. A 10-iteration fourfold cross-validation procedure was used to validate significant associations and quantify their predictive abilities. More than one-third of the significant, cross-validated SNP–SNP interactions predicting each outcome involved just five SNPs, showing that these SNPs are of key importance to fenofibrate response. Multiple variable models constructed using the top-ranked SNP–covariate interactions explained 11.9% more variation in the change in TG and 7.8% more variation in the change in VLDL than baseline TG alone. These results yield insight into the complex biology of fenofibrate response, which can be used to target fenofibrate therapy to individuals who are most likely to benefit from the drug.

European Journal of Human Genetics (2008) 16, 603–613; doi:10.1038/sj.ejhg.5202003; published online 23 January 2008

Keywords: fenofibrate; triglyceride; VLDL; gene–drug interaction; epistasis; gene–gene interaction

*Correspondence: JA Smith, Department of Epidemiology, School of Public Health, University of Michigan, 109 Observatory, Room 4605, Ann Arbor, MI 48109-2029, USA.

Tel: +1 734 647 3721; Fax: +1 734 998 6837;

E-mail: smjenn@umich.edu

Received 29 September 2007; revised 12 December 2007; accepted 13 December 2007; published online 23 January 2008

Introduction

Elevated serum triglyceride (TG) level is an underlying risk factor for coronary heart disease (CHD),¹ even after controlling for other known lipoprotein risk factors such as low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C).^{2,3} It contributes

to CHD directly by triggering vascular endothelial cell dysfunction and inflammation due to increased production of atherogenic TG-rich lipoproteins⁴ and indirectly through its association with type II diabetes and abdominal obesity, both major risk factors for CHD.⁵

Reducing high serum TG in individuals through diet or drug therapies has been shown to reduce adverse outcomes (ie, death or myocardial infarction) due to CHD.² Current TG-lowering therapy includes fibrates, niacin, and omega-3 fatty acids (fish oil).⁶ Statins, which are often prescribed to lower LDL-C, may also have a mild TG-lowering effect in some patient populations.⁷ Fibrates (which include fenofibrate, gemfibrozil, and bezafibrate among others) are the primary treatment for high TG and have a very low frequency of adverse events,⁸ but there is significant interindividual variation in response to these drugs.⁹ Fibrates bind to transcription factors called PPARs, which then migrate to the nucleus and affect transcription of target genes that are involved in the lipid metabolic pathway.⁸

Fenofibrate reduces fasting TG level and induces a shift to a less atherogenic TG-rich lipoprotein profile.¹⁰ The effects of fenofibrate on the levels of total plasma TG and circulating very low-density lipoprotein (VLDL) particles, the primary carrier of TG during fasting,¹¹ are mediated through its binding to *PPAR- α* .⁸ The resulting up- and downregulation of a number of genes involved in lipoprotein metabolism lowers TG by several mechanisms: (1) reducing the availability of free fatty acids for TG formation, (2) increasing the catabolism of VLDL, and (3) reducing neutral lipid (cholesterol ester and TG) exchange between VLDL and HDL.^{10,12} Genes that are upregulated in response to *PPAR- α* include *LPL*, which is involved in TG lipolysis,⁸ *APOA5*, which enhances *LPL* action and may restrict VLDL secretion from the liver,^{8,13} and several genes related to HDL-C metabolism including *APOA1*, *APOA2*, and *FABP*. *PPAR- α* also downregulates *APOC3*, an apolipoprotein that inhibits VLDL clearance.⁸

Although the mechanism of fenofibrate's action is well characterized, significant interindividual variability in response to fenofibrate⁹ limits the ability to predict individual response to the drug. This variation in fenofibrate response, like other complex phenotypes, is likely due to a combination of genetic and biological risk factors¹⁴ and interactions among these factors.¹⁵ In this study, the largest to date, we begin to characterize the genetic architecture of fasting TG response to fenofibrate drug therapy through the examination of independent genetic effects, interactions between genetic and demographic or biochemical factors, and interactions between pairs of genetic factors that are associated with changes in total plasma TG (mg/dl) and TG content of its largest constituent, VLDL particles, in response to fenofibrate.

We identified 91 single-nucleotide polymorphisms (SNPs) in 25 genes that have been previously identified as

playing a role in lipoprotein metabolism.^{16,17} The genes and polymorphisms tested are shown in Table 1, and a more in-depth description of the functional role of the genes is found in Supplementary Table 1. The polymorphisms chosen for this analysis have been previously shown to have a functional effect on lipid metabolism either alone or through interaction with dietary factors.^{18–23} Variations in these genes have also been shown to affect response to pharmacological intervention with fibrates and other drugs.^{24–26}

We examine the effects of these variants on fenofibrate response using a comprehensive, systematic approach that evaluates the role of many genes, many demographic and clinical covariates, and their interactions. Identifying the constellation of genetic and demographic and/or biochemical factors that interact to affect response to fenofibrate will provide insight into the complex mechanism of action of fenofibrate and may improve physicians' ability to identify individuals likely to respond to fenofibrate treatment.

Materials and methods

Sample

The sample for this study consisted of 688 Caucasian participants from 161 families that participated in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study, which was approved by the Institutional Review Boards at the University of Alabama, the University of Minnesota, the University of Utah, the University of Michigan, and Tufts University. Participants were recruited from two NHLBI Family Heart Study field centers, Minneapolis, MN and Salt Lake City, UT, as described by Lai *et al.*²⁷ In each case, only families with at least two siblings were recruited, and only participants who did not take lipid-lowering agents (pharmaceuticals or nutraceuticals) for at least 4 weeks prior to the initial visit were included.

Protocol

Participants took part in two clinical visits, at least 3 weeks apart. During both visits, blood was drawn after a 12 h fast and analyzed for TG level, lipoprotein levels, total cholesterol, insulin, and glucose. Age, sex, blood pressure, weight, height, waist-to-hip ratio, and smoking status were ascertained at the initial clinical visit. TG was measured by the glycerol-blanked enzymatic method (Roche Diagnostics Corporation, Indianapolis, IN, USA). VLDL-TG was measured by NMR, a method that has been shown to have a high degree of correlation with the more traditional method of ultracentrifugation separation in the GOLDN population.²⁸ Between the initial and follow-up clinical visits, participants took 160 mg of fenofibrate per day (TriCor[®], Abbott Laboratories, Chicago, IL, USA). Fenofibric acid, the active moiety of fenofibrate, was measured by HPLC and the area under the serum fenofibric acid

Table 1 Functions of the 25 lipoprotein metabolism genes investigated in this study

Category	Function	Genes (gene symbols)
Apolipoproteins	Found on lipid membrane of circulating lipoproteins and act as ligands for cellular receptors	Apolipoprotein A-I (APOA1) Apolipoprotein A-II (APOA2) Apolipoprotein A-IV (APOA4) Apolipoprotein A-V (APOA5) Apolipoprotein B (APOB) Apolipoprotein C-III (APOC3) Apolipoprotein E (APOE)
Cellular receptors and cofactors	Mediate cellular TG and/or cholesterol exchange with circulating lipoproteins or act as a receptor for circulating lipoproteins themselves	ATP-binding-cassette transporter A1 (ABCA1) Low-density lipoprotein-related receptor protein 1 (LRP1) PDZ domain-containing 1 (PDZK1) Scavenger receptor, class B, member 1 (SCARB1)
Lipases	Hydrolyze TG	Hepatic lipase (LIPC) Endothelial lipase (LIPG) Lipoprotein lipase (LPL)
Transcription factors	Regulate transcription of lipid metabolism genes	Peroxisome proliferate-activated receptor α (PPARA) Peroxisome proliferate-activated receptor γ (PPARG) Peroxisome proliferate-activated receptor γ , cofactor 1, α (PPARGC1A)
Cholesterol metabolism proteins	Mediate cholesterol metabolism	ATP-binding cassette transporter, subfamily G, member 5 (ABCG5) ATP-binding cassette transporter, subfamily G, member 8 (ABCG8) Cholesterol-7- α -hydroxylase (CYP7A1)
Fatty acid binding-proteins	Bind fatty acid and affect TG-rich lipoprotein production	Fatty acid-binding protein 1 (FABP1) Fattyacid-binding protein 2 (FABP2)
Lipoprotein assembly proteins	Mediate synthesis of chylomicrons and VLDL	Microsomal triglyceride transport protein (MTP)
Lipoprotein oxidation	Synthesize nitric oxide in endothelial cells	Nitric oxide synthase 3 (NOS3)
Adiposity/dyslipidemia	Unknown	WD and tetratricopeptide repeats 1 (WDR1)

concentration *versus* time curve was calculated from three measurements over a 6-h period following the last dose of fenofibrate.²⁹ Plasma insulin and glucose levels were measured before and after fenofibrate treatment, and the change in insulin and glucose levels after fenofibrate treatment was calculated.

Genotyping

Ninety-one SNPs from 25 genes known to be involved in lipoprotein metabolism were identified through the genetic association literature and in some cases additional SNPs within a gene were identified using public databases such as dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). SNPs were genotyped using the Applied Biosystems TaqMan SNP genotyping system according to the laboratory protocol (Applied Biosystems, Foster City, CA, USA). Internal controls and repetitive sampling methods were used to ensure genotyping quality. The overall genotyping error and missing genotype call rate were approximately 1%. SNP allele and genotype frequencies and *P*-values for tests of the Hardy–Weinberg equilibrium are provided in Supplementary Table 2.

Statistics

All analyses were carried out using the *R* statistical language, v.2.4.³⁰ Correlations between clinical covariates

were estimated using Pearson's product moment correlation. Allele and genotype frequencies were calculated using standard gene counting methods. Linkage disequilibrium (LD), as measured by r^2 , was estimated using an expectation maximization algorithm.³¹ The Hardy–Weinberg equilibrium was assessed using a χ^2 test or Fisher's exact test if a genotype class had less than five individuals.³² Covariates showing a large deviation from a normal distribution in diagnostic plots, such as fasting TG, were log₁₀-transformed. Five individuals were considered outliers (>4 standard deviations from the mean) and were removed from the analysis due to extreme values of Δ TG. If an individual had a missing value for a single SNP or covariate, the individual was excluded from association tests involving that variable.

TG or VLDL–TG response to fenofibrate was defined as the difference between fasting TG and VLDL–TG concentration in mg/dl at the first and second visit, denoted Δ TG and Δ VLDL. In all models, Δ TG and Δ VLDL were adjusted for age, sex, waist-to-hip ratio and smoking status (ever *versus* never) because these variables have known associations with serum lipoprotein levels^{33–35} or because they were statistically significant predictors of TG and lipoprotein variables in this data set. A person was considered as having ever smoked if they had smoked more than 100 cigarettes during their lifetime.

We employed a two-step testing strategy to reduce the number of false-positive results. In step 1, we tested for associations between each of the predictors (SNPs and clinical covariates) with Δ TG and Δ VLDL in the full data set using least-squares linear regression methods.³⁶ We also tested for association between each single SNP and each covariate to discern whether a SNP's effect on the outcome was being mediated through a particular covariate. Results were eligible for step 2 if the predictor was significant at a P -value ≤ 0.1 . To determine if interactions among predictors explained additional variation in either outcome, we tested pairwise interactions among all possible pairs of predictors (ie, SNP–SNP, SNP–covariate, and covariate–covariate interactions) in the full data set. Associations involving interactions were assessed with a partial F -test,³⁷ which compares a full model that includes both the interaction terms and the main effects of the variables comprising the interaction terms to a reduced model that includes only the main effects. Interaction models that had a partial F P -value ≤ 0.1 were eligible for step 2.

In step 2, associations (main effects or interactions) with P -values < 0.1 in step 1 were evaluated for their predictive ability in independent test samples using a 10-repetition fourfold cross-validation strategy.^{38,39} Cross-validation is an internal validation strategy that seeks to reduce false-positive results by eliminating associations that lack predictive ability in independent test sets.⁴⁰ Because complex traits are likely influenced by a number of moderate genetic effects, P -value correction can be overly conservative and can lead to the removal of important associations. Cross-validation provides a direct assessment of the predictive ability of an association in a new, independent sample of individuals and does not depend strictly (ie, monotonically) on the P -value of an association. Results can be prioritized based on cross-validation, and only those that meet a preset threshold of predictive ability are considered potential 'true positives.' Cross-validation methods have been widely used over the last decade in genomic,^{41,42} metabolomic,⁴³ proteomic,^{44,45} and transcriptomic^{46–48} studies as a method for discriminating between true associations and false-positive associations.

For each association, the fourfold cross-validation procedure began by randomly dividing the sample into four equally sized groups without regard to family structure. Three of the four groups were combined into a training data set, and the modeling strategy outlined above (in step 1) was carried out to estimate model coefficients for the training data set. Models were then applied to the fourth group, the testing data set, to make predictions about the value of the outcome variable of each individual in the independent test sample. This process was repeated for each of the four testing sets. Predicted values for all individuals in the test set were then subtracted from their observed values, yielding the total predictive error variability (SSPE), $\sum_{i=1}^n (y_i - \hat{y}_i)^2$. The total variability in the

outcome (SST)—the difference between each individual's observed value and the mean value for the outcome—was then calculated, $\sum_{i=1}^n (\bar{y} - y_i)^2$. To estimate the proportion of variation in the outcome predicted in the independent test samples the cross-validated R^2 (CV R^2) was calculated as follows: $CV R^2 = \frac{SST - SSPE}{SST}$.

Cross-validation provides a more accurate measure of the predictive ability of the genetic models than the R^2 values from linear modeling and will be negative when the model's predictive ability is poor. Because random variations in the sampling of the four mutually exclusive test groups can potentially impact the estimates of CV R^2 , this fourfold cross-validation procedure was repeated 10 times for each association and the CV R^2 values were averaged.³⁸ Univariate associations and interaction effects were considered cross-validated if the average percentage of variation predicted in independent test samples was greater than 0.5.

Each association that was validated in independent test samples in step 2 was then analyzed to determine whether the family structure in the sample influenced the observed association in the full data set. Generalized estimating equations were used to adjust the standard errors of the coefficients in both univariate and pairwise interaction models.⁴⁹ We considered cross-validated associations with a P -value > 0.1 on a global Wald test to be significantly influenced by family structure and therefore did not report them as cross-validated associations.

To visualize the complex genetic architecture, we applied a novel data visualization scheme, the KGraph, described by Kelly *et al.*⁵⁰ The KGraph was developed for the visualization of multilocus genetic association results in the context of the underlying relationships among the predictors. We used the KGraph to simultaneously display significant univariate associations and pairwise interactions with the outcomes of interest, Δ TG and Δ VLDL, as well as the relationships among the predictors including SNP–SNP frequency correlations (ie, LD), SNP–covariate associations, and covariate–covariate correlations.

Understanding that total TG and VLDL–TG response to fenofibrate treatment is a complex system unlikely to be adequately explained by a single polymorphism or interaction, we constructed several prediction models using the four top-ranked significant cross-validated SNP–SNP and SNP–covariate interactions (ie, those that had the greatest predictive utility for the outcomes, Δ VLDL and Δ TG). As a point of reference, we compared these multivariable prediction models to a model including only baseline TG.

Results

Descriptive statistics of the clinical covariates and outcomes are shown in Table 2. The mean age of the

participants was 47.8 years, 51.9% of participants were female, and 17.6% of participants were classified as an ever-smoker (smoked ≥ 100 cigarettes lifetime). Participants had a mean BMI of 28.5 kg/m², waist-to-hip ratio of 0.9, and fasting total cholesterol level of 189 mg/dl. The mean fasting TG level was 132 mg/dl prior to fenofibrate treatment and 88 mg/dl after fenofibrate treatment. The mean Δ TG and Δ VLDL levels were -46.3 and -46.4 mg/dl respectively. Allele and genotype frequencies, rs numbers from dbSNP (build 127), SNP type (synonymous, non-synonymous, intron, and so on), and Hardy–Weinberg equilibrium test results are reported in Supplementary Table 2.

As described in the Methods section, association results were subjected to a two-step analysis process intended to reduce the occurrence of false-positive results and to identify associations with predictive ability in independent test samples. Table 3 shows the number of associations meeting the criteria for each of the steps (significant association and cross-validation) and the percentage of tests from the association step that cross-validated. For example, the 91 SNPs were evaluated for their association with Δ VLDL and Δ TG and in each case, 13 SNPs (14%) showed significant association with the outcome but only

one SNP (*APOA5_S19W*) was validated in independent test samples by predicting $>0.5\%$ of the variation in each of these traits. In contrast, there were 1183 tests of SNP–covariate interactions for each outcome, and 156 (13%) tests were significant in each case, with 34 (3%) validated interactions predicting Δ VLDL and 25 (2%) validated interactions predicting Δ TG. Similarly, there were 4095 tests of SNP–SNP interactions for each outcome; 54 (1%) interactions were validated as predictors in independent test samples for Δ TG and 54 (1%) interactions were validated as predictors in independent test samples for Δ VLDL. While *P*-values and CV R^2 were moderately correlated in tests with *P*-value ≤ 0.01 ($R^2 = 0.13$), Figure 1 demonstrates that there is also a moderate amount of discrepancy between the two methods.

Figure 2 is a visual representation of the complex genetic and clinical covariate associations underlying the Δ TG and Δ VLDL outcomes. Using both color and spatial relationships, the KGraph presents both associations with outcomes of interest and the correlation structure of the predictors that underlie those associations. SNPs without at least one significant and cross-validated association either univariately or interacting with another SNP or covariate in either outcome were excluded from display on the KGraph.

Table 2 Descriptive statistics of clinical covariates and outcome variables

Clinical covariates and outcome variables	Total (N = 688)	Mean values (mean $\pm$ SD)		P-value ^a
		Men (N = 331)	Women (N = 357)	
Gender, Male	48.1%			
Age, years	47.8 $\pm$ 16.2	47.9 $\pm$ 16.2	47.7 $\pm$ 16.3	0.8710
BMI, kg/m ²	28.5 $\pm$ 5.5	28.3 $\pm$ 4.5	28.6 $\pm$ 6.3	0.5395
Waist-to-hip ratio	0.90 $\pm$ 0.10	0.95 $\pm$ 0.09	0.85 $\pm$ 0.10	<0.0001
Ever-smoker	17.6%	19.9%	15.4%	0.1204
Total fasting cholesterol, mg/dl	189.1 $\pm$ 38.7	186.7 $\pm$ 35.7	191.3 $\pm$ 41.2	0.1110
Fenofibric acid area under the curve	69.4 $\pm$ 32.1	65.1 $\pm$ 27.8	73.6 $\pm$ 35.2	0.0022
<i>Insulin</i>				
Baseline, mU/l	14.3 $\pm$ 8.3	14.3 $\pm$ 7.4	14.1 $\pm$ 8.4	0.2692
After fenofibrate, mU/l	13.7 $\pm$ 8.6	13.7 $\pm$ 8.0	13.7 $\pm$ 8.3	0.2932
Δ , mU/l	-0.57 ± 5.6	-0.63 ± 5.4	-0.44 ± 5.0	0.9930
<i>Glucose</i>				
Baseline, mg/dl	101.9 $\pm$ 18.8	103.9 $\pm$ 18.2	98.5 $\pm$ 17.2	<0.0001
After fenofibrate, mg/dl	99.4 $\pm$ 18.6	101.0 $\pm$ 18.9	96.3 $\pm$ 16.7	<0.0001
Δ , mg/dl	-2.5 ± 8.4	-3.0 ± 7.5	-2.2 ± 6.9	0.0880
<i>Triglyceride</i>				
Baseline, mg/dl	132.2 $\pm$ 83.3	139.0 $\pm$ 84.1	125.9 $\pm$ 79.5	0.0363
After fenofibrate, mg/dl	87.7 $\pm$ 50.2	94.2 $\pm$ 52.2	81.7 $\pm$ 47.5	0.0011
Δ , mg/dl	-44.5 ± 51.8	-44.7 ± 53.7	-44.2 ± 50.0	0.8875
Adjusted Δ , mg/dl ^b	-46.3 ± 50.6	-46.3 ± 53.7	-46.3 ± 47.6	1
<i>VLDL—TG</i>				
Baseline, mg/dl	101.1 $\pm$ 79.2	113.1 $\pm$ 83.6	89.9 $\pm$ 73.2	0.0001
After fenofibrate, mg/dl	56.8 $\pm$ 47.5	66.7 $\pm$ 51.8	47.7 $\pm$ 41.3	<0.0001
Δ , mg/dl	-44.5 ± 53.7	-46.7 ± 57.5	-42.4 ± 49.8	0.2974
Adjusted Δ , mg/dl ^b	-46.4 ± 52.4	-46.4 ± 57.3	-46.4 ± 47.5	1

^aT-test *P*-value between men and women.

^bAdjusted for age, sex, waist-to-hip ratio, and smoking status (ever *versus* never).

Table 3 Distribution of test results before and after cross-validation

Model	Number of tests	$P \leq 0.1^a$ N (%)	$P \leq 0.05^a$ N (%)	Validated in independent samples ^b N (%)
Δ VLDL (mg/dl) ^c				
Single SNP	91	13 (14)	6 (7)	1 (1)
SNP–covariate interaction	1183	156 (13)	100 (8)	34 (3)
SNP–SNP interaction	4095	474 (12)	245 (6)	54 (1)
Δ TG (mg/dl) ^c				
Single SNP	91	13 (14)	4 (4)	1 (1)
SNP–covariate interaction	1183	156 (13)	79 (7)	25 (2)
SNP–SNP interaction	4095	468 (11)	252 (6)	54 (1)

^aModel P -value for single SNP models; partial F -test P -value for the full and reduced models (interaction terms only) for the interaction models.

^bMore than 0.5% variability explained in independent test samples. Cross-validation was performed on all associations with $P \leq 0.1$. This column includes only associations that remained significant with GEE modeling.

^cAdjusted for age, sex, waist-to-hip ratio, and smoking status (ever *versus* never).

Figure 1 Partial F -test P -values *versus* $CV R^2$ for SNP–covariate and epistatic (SNP–SNP) interactions predicting Δ TG and Δ VLDL. Percent variation explained in independent test samples ($CV R^2$) *versus* $-\log_{10}$ partial F -test P -value for SNP–covariate and epistatic (SNP–SNP) interactions that were significant in Δ TG or Δ VLDL (partial F -test P -value ≤ 0.1). R^2 for the correlation between $-\log P$ -value and $CV R^2$ is 0.13. Lines are shown at $-\log P = 0.05$ and $CV R^2 = 0.5\%$.

Region 1 of Figure 2, shown in green, displays the association between the SNPs and clinical covariates. Though this step is often overlooked in association studies, understanding the relationship between SNPs and the covariates can be informative in discerning how the SNPs' effect on the outcome is mediated. In this region, there are a number of cross-validated associations between SNPs in the apolipoprotein gene cluster, which contains *APOA1*, *APOA4*, *APOA5*, and *APOC3*, and fasting baseline TG or cholesterol. Region 2, shown in gray, illustrates the correlations between the covariates. The majority of the covariates are not significantly correlated ($|r| < 0.3$), with only two pairs (glucose before and after fenofibrate treatment and insulin before and after fenofibrate treatment) highly correlated ($|r| > 0.6$). LD between SNPs is shown in red in Region 3. Most of the LD observed occurs

between SNPs within the same gene, primarily in *APOC3*, and among other SNPs in the apolipoprotein gene cluster.

The remaining regions are colored in blue, and they represent associations with Δ TG and Δ VLDL. In these regions, the upper right section of the cells represents results for Δ TG and the lower left section represents results for Δ VLDL. Cross-validated associations are indicated by a black bar. Region 4, which displays the univariate association between clinical covariates and the outcomes, shows that baseline TG, cholesterol, and both insulin and glucose before fenofibrate treatment have cross-validated associations with both Δ TG and Δ VLDL, while fenofibric acid area under the curve, glucose level after fenofibrate, and waist-to-hip ratio have cross-validated associations with Δ VLDL but not Δ TG. Region 5, which illustrates univariate association between the SNPs and the outcomes, reveals that only one SNP in *APOA5* (*APOA5_S19W*) has a cross-validated association with both Δ TG and Δ VLDL.

Region 6 displays the covariate–covariate interactions significantly associated with Δ TG and/or Δ VLDL. The majority of the cross-validated interactions were observed between fasting baseline TG and measures of glucose and insulin level before and after fenofibrate and between fenofibric acid area under the curve and insulin level before and after fenofibrate.

Region 7 displays the interactions between the SNPs and covariates that were associated with Δ TG and/or Δ VLDL. A complete listing of all significant and cross-validated SNP–covariate interactions and the corresponding relationships between genotypes and phenotypes are provided in Supplementary Tables 3, 5, and 6. Five of the clinical covariates (sex, baseline TG, fenofibric acid area under the curve, baseline glucose, and change in glucose after fenofibrate) accounted for a large portion of SNP–covariate interactions: 15 (60%) of the validated interactions predicted Δ TG and 16 (47%) of the validated interactions predicted Δ VLDL. The predictive abilities of these interactions ranged from 0.50 to 1.90% (mean = 0.89%) for Δ TG and 0.51–1.47% (mean = 0.92%) for Δ VLDL. The $\Delta R^2 \times 100$

Figure 2 KGraph: fasting total TG and VLDL–TG and response to fenofibrate. Using 8 regions, the KGraph shows the relationships between the SNPs, clinical covariates, and outcomes by displaying the results from tests of correlation, linkage disequilibrium, association and cross-validation. The key at the bottom of the graphic shows the test criterion for each region and the colors associated with the test result. The region number key in the lower left corner shows the location of each region, and indicates whether the results in the region were assessed using cross-validation (shaded regions). Cross-validation of a specific test result is indicated by a black bar. Results for Δ TG are displayed in the top/right half of the cell and results for Δ VLDL are displayed in the bottom/left half. Region 1 displays the association between the SNPs and the clinical covariates, region 2 displays the correlation between the clinical covariates, and region 3 displays the linkage disequilibrium between the SNPs. Region 4 displays covariate association with the outcomes (Δ TG and Δ VLDL), region 5 displays SNP association with the outcomes, region 6 displays covariate–covariate interactions predicting the outcomes, region 7 displays SNP–covariate interactions predicting the outcomes, and region 8 displays SNP–SNP (epistatic) interactions predicting the outcomes.

Table 4 Multiple variable modeling results

Model	Percent variation explained in independent test samples	
	Δ TG ^a	Δ VLDL ^a
Baseline triglyceride only	48.1	46.7
Baseline triglyceride+four top-ranked epistatic interactions	48.5 ^b	47.1 ^c
Baseline triglyceride+four top-ranked SNP–covariate interactions	60 ^d	54.5 ^e

^aAdjusted for age, sex, waist-to-hip ratio, and smoking status (ever *versus* never).

^bModel includes interactions between *APOA5_S19W* and *PDZK1_i4201*, *APOA1_M2630* and *FABP1_T94A*, *APOA1_M3012* and *APOB_E4181K*, and *MTP_M493* and *PPARG_P12AEB*, as well as all main effects of the SNPs.

^cModel includes interactions between *APOA5_S19W* and *PDZK1_i4201*, *FABP1_m2353* and *PPARGC1A_i27289*, *LIPC_V95M* and *MTP_M493*, and *ABCA1_I27943* and *LPL_D9N*, as well as all main effects of the SNPs.

^dModel includes interactions between *LIPC_i67180* and baseline triglyceride (log), *ABCA1_I27943* and age, *APOA4_M35* and sex, and *LIPG_T1111* and fenofibric acid area under the curve, as well as all main effects of the SNPs and the clinical covariates. The main effects only model explains 55.3% of the variation in Δ TG.¹

^eModel includes interactions between *FABP1_m2353* and baseline glucose (log), *LIPC_i67180* and baseline triglyceride (log), *ABCA1_I27943* and age, and *LIPG_i13576* and fenofibric acid area under the curve, as well as all main effects of the SNPs and the clinical covariates. The main effects only model explains 52.8% of the variation in Δ VLDL.¹

values (adjusted R^2 from the model including interactions minus adjusted R^2 from the main effects only model) for the same interactions, which range from 0.58 to 2.29% (mean = 1.35%) for Δ TG and 0.67–2.40% (mean = 1.39%) for Δ VLDL, are generally overly optimistic. Baseline glucose and the area under the fenofibric acid concentration curve, a measure of an individual's plasma level of the active form of fenofibrate over a 6 h time period, were most sensitive to genetic context. They each interacted with five SNPs in their effect on Δ VLDL and also interacted with three SNPs in their effect on Δ TG. Sex interacted with three SNPs in its effects on Δ TG and on Δ VLDL, and baseline TG (log) and change in glucose after fenofibrate had three interactions with SNPs in their effects on Δ TG.

Overall, four SNPs are responsible for a large portion of the SNP–covariate interactions that predict Δ TG (*APOA4_M35*, *APOC3_3U386*, *ABCA1_I27943*, and *LIPC_T224T*). Together, these SNPs account for eight (32%) of the validated SNP–covariate interactions in their effect on Δ TG, with predictive ability ranging from 0.52 to 1.29% (mean = 0.82%) and $\Delta R^2 \times 100$ values ranging from 0.73 to 1.33% (mean = 1.02%). With respect to predicting Δ VLDL, four SNPs are responsible for a substantial proportion of the SNP–covariate interactions (*ABCG8_C54Y*, *LIPC_i67180*, *FABP1_m2353*, and *FABP1_T94A*). These SNPs account for 10 (29%) of the validated SNP–covariate interactions for Δ VLDL, with predictive abilities ranging from 0.56 to 1.47% (mean = 0.90%) and $\Delta R^2 \times 100$ values ranging from 0.87 to 2.40% (mean = 1.41%). For both outcomes, the remaining SNPs were involved in at most one validated interaction with a covariate (see Supplementary Table 3 for details).

Region 8 displays the epistatic interactions among the SNPs that have a significant and validated predictive ability for Δ TG and Δ VLDL. As with SNP–covariate interactions, a large proportion of epistatic interactions are accounted for

by a small number of SNPs. The SNPs that have the most interactions that were validated as predictors of Δ TG are *APOA5_S19W* with six interactions, *PDZK1_i4201* with five interactions, and *APOA1_M2630* and *FABP1_T94A* with four interactions each. Together, these SNPs account for 16 (30%) of the validated epistatic interactions that predict Δ TG, have predictive abilities ranging from 0.53 to 2.19% (mean = 0.85%), and have $\Delta R^2 \times 100$ values ranging from 0.71 to 2.80 (mean = 1.58%). The SNPs with the most interactions that were validated as predictors of Δ VLDL are *APOA5_S19W* with six interactions, *APOB_E4181K* with five interactions, and *APOA1_M3012*, *APOB_T2515T*, and *APOC3_3U386* with four interactions each. These SNPs account for 18 (33%) of the validated epistatic interactions that predict Δ VLDL, have predictive abilities ranging from 0.51 to 2.42% (mean = 0.81%), and have $\Delta R^2 \times 100$ values ranging from 0.52 to 2.56% (mean = 1.18%). A complete listing of all significant and cross-validated epistatic interactions is given in Supplementary Table 6.

Results for the multivariable modeling using the top-ranked SNP–SNP and SNP–covariate interactions are shown in Table 4. Only two of the SNPs used in the models had cross-validated associations with any of the covariates (*PPARG_P12AEB* with fenofibric acid area under the curve and *PPARG_C1A_i27289* with insulin after fenofibrate), which indicates that the associations between these top-ranked SNPs and the outcome is not primarily a result of their relationship with the covariates. A model that includes the four top-ranked epistatic (SNP–SNP) interactions minimally increases the percentage of variation explained beyond that explained by baseline TG alone (0.4% for Δ TG and 0.4% for Δ VLDL). However, a model that includes the four top-ranked SNP–covariate interactions significantly increases the percentage of variation explained beyond baseline TG (11.9% for Δ TG and 7.8% for Δ VLDL).

Discussion

Understanding the genetic architecture that underlies a person's TG response to fenofibrate will help identify those who will most likely benefit from fenofibrate treatment and shed light on the complexities of the network that controls the metabolic response to drugs such as fenofibrate. By looking beyond simple, single SNP effects and examining interactions between SNPs and clinical covariates and pairs of SNPs, this study takes an important first step toward the mapping of this genetic architecture.

The key factor in an analysis such as this is to attempt to appropriately model the system's complexity to the extent possible. Fasting TG response to fenofibrate treatment is influenced by environmental changes (at both the cellular and organismal levels) and large-scale gene expression networks that are controlled in part by genetic variations. A study of this size is unlikely to be adequately powered to reconstruct this network in its entirety, but can be used to begin constructing a better map of the gene-covariate interactions and associations that play a role in inter-individual variation.

As the number of SNPs considered in an association study increases, so does the problem of separating true results from spurious results caused by random chance.⁵¹ Attempts to combat this problem often focus on *P*-values, with the rationale that by choosing the models least likely to be found by random chance alone, the probability of including a false positive will be greatly reduced.⁵² Unfortunately, when considering a complex trait, the effect sizes of individual SNPs and pairwise interactions are generally fairly small. This limits the significance level that these models are able to attain in typical epidemiologic studies, which means that true positives may be removed during *P*-value adjustment. Additionally, the very low *P*-values that are required for an association to remain significant after adjustment (eg, by Bonferroni correction or false discovery rate adjustment) are likely to be obtained by SNPs with low minor allele frequencies that capture groups on the extreme of the phenotypic tail (sometimes simply by chance alone).

We approached the problem from a different direction. The utility of the associations being investigated lies in their ability to make a prediction about the outcome of interest, so it is natural to select the best models based on predictive ability. By using a cross-validation procedure that successively leaves one quarter of the data out and constructs genetic models in the remaining three quarters of the data, we can estimate the performance of the model on a new, independent sample of individuals from the same population.

Interestingly, the correlation between *P*-value and CV R^2 is moderate at best (Figure 1), and many associations with only marginally significant *P*-values have a non-negative CV R^2 . With the movement toward genome-wide association studies that measure hundreds of thousands of SNPs,

the problems in correcting for multiple testing will be magnified. Cross-validation is an attractive option to genetic epidemiologists both because it offers a direct assessment of an association's predictive utility and because of the relative simplicity and stability of the cross-validation estimate.³⁸ In addition, replication in an independently collected sample has become the standard tool to validate genetic association results. However, no comparable samples exist for specialized intervention trials such as this one. In the absence of replication samples, cross-validation gives us the ability to validate our results while simultaneously assessing their predictive capacity. Without a replicate sample, however, we cannot estimate whether the predictive ability of the multiple variable models established in this data set would be the same in other data sets.

The 91 SNPs investigated in this study span genes known to be involved in many facets of lipid metabolism. Interestingly, 78 of them (86% of the SNPs) have at least one significant cross-validated association with a clinical covariate or have at least one epistatic or SNP-covariate interaction that predicted either outcome. It is sobering but not completely unexpected that individual polymorphisms and interactions make relatively small contributions to explaining the variation in Δ TG and Δ VLDL. However, the percentage of trait variation explained by SNPs in this study is similar to that observed in studies of other complex phenotypes. For example, a recent genome-wide association study of diabetes and associated traits showed that the single SNPs most strongly associated with TG level explained between 0.5 and 1.2% of the residual variance of TG level after adjustment for traditional risk factors.⁵³ A candidate gene study of blood pressure phenotypes found that the CV R^2 of the most predictive SNPs for systolic blood pressure ranged between 1.9 and 6.2%.⁵⁴

Combining these individual interactions into a multiple variable model did increase the amount of variation explained, even if only moderately in the case of epistatic interactions. We do note that, in general, non-cross-validated genetic models all overestimate the contribution of each predictor or interaction, so it is likely that the percentage of variation explained using cross-validation, while small, is a more realistic estimate than R^2 values obtained without cross-validation.

The large number of SNPs involved in significant cross-validated interactions illustrates the complexity of the network underlying these traits, comprising single gene effects and interactions among multiple variations in many genes as well as their interaction with clinical factors. Also, while there is substantial overlap between the clinical factors and interactions that predicted Δ TG and Δ VLDL, there is also evidence of associations that are unique to each outcome, which shows that there are genetic and environmental interactions that affect the TG response to

fenofibrate during fasting that lie outside of the metabolism of VLDL particles.

As genotyping has become increasingly less expensive, the scale of genetic association studies has risen dramatically,⁵⁵ but most studies, even genome-wide studies, concentrate on single, univariate associations. There is no denying that some single polymorphisms can significantly influence physiology; many diseases are the result of a single genetic error with serious phenotypic consequences. Complex, non-Mendelian traits, however, are likely influenced less by single polymorphisms than by interacting effects, particularly epistatic interactions.⁵⁶ Our results indicate that a wealth of pairwise interactions can be found even in this fairly limited subset of polymorphisms. Methods, such as the multivariable modeling approach that we describe, must now be developed that will allow the integration of these results into a predictive framework, so that the understanding of the genetic architecture of TG response to fenofibrate treatment can be refined and harnessed.

Acknowledgements

We thank the families that participated in the GENOA study and Jian Chu, Douglas Jacobsen, Kristin Meyers, and Todd Greene for editorial and computing assistance. This work was supported by grant U 01 HL72524 from the National Heart, Lung, and Blood Institute and training grant T32 HG00040 from the National Human Genome Research Institute (JAS). The authors have no conflicts of interest to disclose relating to the subject of this paper.

References

- Haymore BR, Parks JR, Oliver TG, Glister BC: Hypertriglyceridemia. *Hosp Physician* 2005; **41**: 17–24.
- Cullen P: Evidence that triglycerides are an independent coronary heart disease risk factor. *Am J Cardiol* 2000; **86**: 943–949.
- Hopkins PN, Wu LL, Hunt SC, Brinton EA: Plasma triglycerides and type III hyperlipidemia are independently associated with premature familial coronary artery disease. *J Am Coll Cardiol* 2005; **45**: 1003–1012.
- Ooi TC, Ooi DS: The atherogenic significance of an elevated plasma triglyceride level. *Crit Rev Clin Lab Sci* 1998; **35**: 489–516.
- Yoshino G, Hirano T, Kazumi T: Dyslipidemia in diabetes mellitus. *Diabetes Res Clin Pract* 1996; **33**: 1–14.
- McKenney JM, Sica D: Role of prescription omega-3 fatty acids in the treatment of hypertriglyceridemia. *Pharmacotherapy* 2007; **27**: 715–728.
- Stein EA, Lane M, Laskarzewski P: Comparison of statins in hypertriglyceridemia. *Am J Cardiol* 1998; **81**: 66B–69B.
- Chapman MJ: Fibrates in 2003: therapeutic action in atherogenic dyslipidaemia and future perspectives. *Atherosclerosis* 2003; **171**: 1–13.
- Keating GM, Ormrod D: Micronised fenofibrate: an updated review of its clinical efficacy in the management of dyslipidaemia. *Drugs* 2002; **62**: 1909–1944.
- Tsimihodimos V, Miltiadous G, Daskalopoulou SS, Mikhailidis DP, Elisaf MS: Fenofibrate: metabolic and pleiotropic effects. *Curr Vasc Pharmacol* 2005; **3**: 87–98.
- Mason TM: The role of factors that regulate the synthesis and secretion of very-low-density lipoprotein by hepatocytes. *Crit Rev Clin Lab Sci* 1998; **35**: 461–487.
- Staels B, Dallongeville J, Auwerx J, Schoonjans K, Leitersdorf E, Fruchart JC: Mechanism of action of fibrates on lipid and lipoprotein metabolism. *Circulation* 1998; **98**: 2088–2093.
- Merkel M, Heeren J: Give me A5 for lipoprotein hydrolysis!. *J Clin Invest* 2005; **115**: 2694–2696.
- Shah RR: Pharmacogenetics in drug regulation: promise, potential and pitfalls. *Philos Trans R Soc Lond B Biol Sci* 2005; **360**: 1617–1638.
- Sing CE, Stengard JH, Kardina SL: Genes, environment, and cardiovascular disease. *Arterioscler Thromb Vasc Biol* 2003; **23**: 1190–1196.
- Ordovas JM, Corella D: Nutritional genomics. *Annu Rev Genomics Hum Genet* 2004; **5**: 71–118.
- Hachem SB, Mooradian AD: Familial dyslipidaemias: an overview of genetics, pathophysiology and management. *Drugs* 2006; **66**: 1949–1969.
- Corella D, Ordovas JM: Single nucleotide polymorphisms that influence lipid metabolism: interaction with dietary factors. *Annu Rev Nutr* 2005; **25**: 341–390.
- Ordovas JM: HDL genetics: candidate genes, genome wide scans and gene-environment interactions. *Cardiovasc Drugs Ther* 2002; **16**: 273–281.
- Ye SQ, Kwiterovich Jr PO: Influence of genetic polymorphisms on responsiveness to dietary fat and cholesterol. *Am J Clin Nutr* 2000; **72**: 1275S–1284S.
- Brisson D, Ledoux K, Bosse Y *et al*: Effect of apolipoprotein E, peroxisome proliferator-activated receptor alpha and lipoprotein lipase gene mutations on the ability of fenofibrate to improve lipid profiles and reach clinical guideline targets among hypertriglyceridemic patients. *Pharmacogenetics* 2002; **12**: 313–320.
- Talmud PJ, Hawe E, Robertson K, Miller GJ, Miller NE, Humphries SE: Genetic and environmental determinants of plasma high density lipoprotein cholesterol and apolipoprotein AI concentrations in healthy middle-aged men. *Ann Hum Genet* 2002; **66**: 111–124.
- Lai CQ, Parnell LD, Ordovas JM: The APOA1/C3/A4/A5 gene cluster, lipid metabolism and cardiovascular disease risk. *Curr Opin Lipidol* 2005; **16**: 153–166.
- Yamada M: Influence of apolipoprotein E polymorphism on bezafibrate treatment response in dyslipidemic patients. *J Atheroscler Thromb* 1997; **4**: 40–44.
- Zhang X, Chen ZQ, Wang Z, Mohan W, Tam SP: Protein-DNA interactions at a drug-responsive element of the human apolipoprotein A-I gene. *J Biol Chem* 1996; **271**: 27152–27160.
- Nemeth A, Szakmary K, Kramer J *et al*: Apolipoprotein E and complement C3 polymorphism and their role in the response to gemfibrozil and low fat low cholesterol therapy. *Eur J Clin Chem Clin Biochem* 1995; **33**: 799–804.
- Lai CQ, Arnett D, Straka RJ *et al*: Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study. *Arterioscler Thromb Vasc Biol* 2007; **27**: 1417–1425.
- Tsai MY, Georgopoulos A, Otvos JD *et al*: Comparison of ultracentrifugation and nuclear magnetic resonance spectroscopy in the quantification of triglyceride-rich lipoproteins after an oral fat load. *Clin Chem* 2004; **50**: 1201–1204.
- Straka RJ, Burkhardt RT, Fisher JE: Determination of fenofibric acid concentrations by HPLC after anion exchange solid-phase extraction from human serum. *Ther Drug Monit* 2007; **29**: 197–202.
- R Core Development Team: R: A language and environment for statistical computing, 2008.
- Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Massachusetts: Sinauer Associates, Inc., 1998.
- Weir B: *Genetic Data Analysis II*. Massachusetts: Sinauer Associates, Inc., 1996.
- Freedman DS, Otvos JD, Jeyarajah EJ *et al*: Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham study. *Clin Chem* 2004; **50**: 1189–1200.

- 34 Mansfield E, McPherson R, Koski KG: Diet and waist-to-hip ratio: important predictors of lipoprotein levels in sedentary and active young men with no evidence of cardiovascular disease. *J Am Diet Assoc* 1999; **99**: 1373–1379.
- 35 Billimoria JD, Pozner H, Metselaar B, Best FW, James DC: Effect of cigarette smoking on lipids, lipoproteins, blood coagulation, fibrinolysis and cellular components of human blood. *Atherosclerosis* 1975; **21**: 61–76.
- 36 Kleinbaum D, Kupper L, Muller K, Nizam A: *Applied Regression Analysis and Other Multivariate Methods*. California: Brooks/Cole Publishing Company, 1998.
- 37 Rosner BA: *Fundamentals of Biostatistics*. United States: Thomson Brooks/Cole, 2006.
- 38 Molinaro AM, Simon R, Pfeiffer RM: Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005; **21**: 3301–3307.
- 39 Browne MW: Cross-validation methods. *Journal of Mathematical Psychology* 2000; **44**: 108–132.
- 40 Stone M: Cross-validatory choice and assessment of statistical predictions. *J R Statist Soc B* 1974; **36**: 111–147.
- 41 Motsinger AA, Ritchie MD: Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene–gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics* 2006; **2**: 318–328.
- 42 Gong R, Liu Z, Li L: Epistatic effect of plasminogen activator inhibitor 1 and beta-fibrinogen genes on risk of glomerular microthrombosis in lupus nephritis: interaction with environmental/clinical factors. *Arthritis Rheum* 2007; **56**: 1608–1617.
- 43 Pohjanen E, Thysell E, Jonsson P *et al*: A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum. *J Proteome Res* 2007; **6**: 2113–2120.
- 44 Agranooff D, Fernandez-Reyes D, Papadopoulos MC *et al*: Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* 2006; **368**: 1012–1021.
- 45 Mertens BJ, De Noo ME, Tollenaar RA, Deelder AM: Mass spectrometry proteomic diagnosis: enacting the double cross-validatory paradigm. *J Comput Biol* 2006; **13**: 1591–1605.
- 46 Feng Z, Prentice R, Srivastava S: Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. *Pharmacogenomics* 2004; **5**: 709–719.
- 47 Zhu J, Hastie T: Classification of gene microarrays by penalized logistic regression. *Biostatistics* 2004; **5**: 427–443.
- 48 Wood IA, Visscher PM, Mengersen KL: Classification based upon gene expression data: bias and precision of error rates. *Bioinformatics* 2007; **23**: 1363–1370.
- 49 Liang KY, Zeger SL: Regression analysis for correlated data. *Annu Rev Public Health* 1993; **14**: 43–68.
- 50 Kelly RJ, Jacobsen DM, Sun YV, Smith JA, Kardia SL: KGraph: a system for visualizing and evaluating complex genetic associations. *Bioinformatics* 2007; **23**: 249–251.
- 51 Thomas DC, Haile RW, Duggan D: Recent developments in genomewide association scans: a workshop summary and review. *Am J Hum Genet* 2005; **77**: 337–345.
- 52 Page GP, George V, Go RC, Page PZ, Allison DB: Are we there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet* 2003; **73**: 711–719.
- 53 Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research Saxena R, Voight BF, Lyssenko V *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
- 54 Kardia SL, Sun YV, Hamon SC, Barkely RA, Boerwinkle E, Turner ST: Interactions between the adducin 2 gene and antihypertensive drug therapies in determining blood pressure in people with hypertension. *BMC Med Genet* 2007; **8**: 61.
- 55 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
- 56 Wagner GP, Laubichler MD, Bagheri-Chaichian H: Genetic measurement of theory of epistatic effects. *Genetica* 1998; **102–103**: 569–580.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)