

ORIGINAL ARTICLE

Concordance of two multiple analytical approaches demonstrate that interaction between BMI and *ADIPOQ* haplotypes is a determinant of LDL cholesterol in a general French population

Francis Vasseur¹, Thomas Caeyseele², Mouna Barat-Houari³, Stéphane Lobbens³, Aline Meirhaeghe⁴, David Meyre³, Philippe Froguel³, Philippe Amouyel⁴ and Nicole Helbecque⁴

Genetic and environmental factors are involved in insulin resistance (IR). IR and dyslipidemia associate with increased risk of cardiovascular diseases. Plasma low-density lipoprotein cholesterol (LDL-C) level is a marker of cardiovascular risk. In a Caucasian general population we aimed at determining the multifactorial components of LDL-C levels using 10 genes and 3 phenotypes. In the *PPARG*, *UCP3*, *ADIPOQ*, *TNF*, *LIPC*, *CARTPT*, *PCSK9*, *SCAP*, *SCARB1* and *ENPP1* genes known to be associated with IR or dyslipidemia we genotyped 19 single nucleotide polymorphisms (SNPs) in 846 subjects. When several SNPs were genotyped for a given gene we constructed haplotypes. Including genetic and environmental variables (gender, body mass index (BMI) and adiponectin level) we used (1) the multifactor dimensionality reduction method to explain clusters of high and low LDL-C, and (2) the restricted partition method to explain LDL-C levels. Both methods showed that BMI and haplotypes at the *ADIPOQ* adiponectin encoding gene but not adiponectin level itself, were discriminant regarding to LDL-C. Subjects bearing an at-risk combination of BMI and *ADIPOQ* genotypes were prone to have a higher LDL-C (OR=3.13, 95% CI=2.20–4.46, $P<0.0001$). Our results suggest that in interaction with BMI, *ADIPOQ* haplotypes capture genetic variation(s) from neighboring gene(s) that would modulate LDL-C level.

Journal of Human Genetics (2010) 55, 227–231; doi:10.1038/jhg.2010.10; published online 26 February 2010

Keywords: adiponectin; cardiovascular risk; genetic variants; LDL-C; multifactorial analysis

INTRODUCTION

Genetic and nongenetic (so-called environmental) factors are involved in the insulin resistance (IR) metabolic syndrome. IR and dyslipidemia are associated with increased risk of cardiovascular diseases. Among the various blood variables, plasma low-density lipoprotein cholesterol (LDL-C) is used as a marker of cardiovascular risk. Numerous papers report associations between genetic variants and IR, features of the metabolic syndrome and dyslipidemia. Several recent genome-wide association studies reported many loci that may modulate dyslipidemia.^{1–5} In this study, we selected genetic variants distributed on various human chromosomes and previously reported to be associated with either IR or dyslipidemia and searched for those explaining LDL-C level. *PPARG* P12A is associated with IR,⁶ and LDL-C.⁷ *UCP3* –55C>T has been reported to be associated with type 2 diabetes and atherogenic lipid profile.⁸ Haplotypes of the promoter (–11391G>A, –11377C>G) and of the exon 2 intron 2 region

(+45T>G, +276G>T) of the *ADIPOQ* gene (also called *ACDC*, *APMI*, *GBP28*) encoding the adiponectin protein are associated with IR and adiponectin level,^{9,10} and decreased adiponectin levels were reported to contribute to the atherogenic lipid profile.¹¹ *TNF* –308G>A has been reported to be associated with IR.¹² The functional *LIPC* –514C>T single nucleotide polymorphism (SNP) has been associated with plasma lipids levels.¹³ *CARTPT* –3608C>T has been associated with obesity¹⁴ and plasma lipid levels.¹⁵ *PCSK9* E670G was reported to be associated with plasma LDL-C.¹⁶ The functional *ENPP1* K121Q has been associated with IR and atherogenic phenotypes.¹⁷ Moreover, the K121Q and the IVS8 +27T>G, IVS20 –11 delT, +828 13 bp insertion, +1044A>G define haplotypes associated with IR and obesity.¹⁸ The *SCAP* I796V is involved in cholesterol homeostasis.¹⁹ Several SNPs (G2S, c.795C>T, c.1119C>T) from *SCARB1* are associated with plasma lipid levels and LDL-C.²⁰ Atherogenesis, for which LDL-C level may be considered as a marker, is a

¹Biostatistics Department EA2694, University Hospital, University Lille Nord de France, UDSL, Lille cedex, France; ²Medical Information Department, University Hospital, Lille, France; ³CNRS, UMR8090, Institute of Biology, Lille, France and ⁴INSERM, U744, Pasteur Institute, University Lille Nord de France, UDSL, Lille cedex, France

Correspondence: Dr F Vasseur, Biostatistics Department EA2694, University Hospital, University Lille Nord de France, Pôle de Santé Publique Parc Eurasanté CHRU, Lille cedex 59037, France.

E-mail: francis.vasseur@univ-lille2.fr

Received 25 September 2009; revised 8 January 2010; accepted 28 January 2010; published online 26 February 2010

complex trait resulting from potential multiple gene–gene and gene–environment interactions. In dissecting multifactorial traits, conventional statistical approaches as multivariate logistic regression have been proved to be limited by a lack of power, or would require huge populations.²¹ To circumvent these limitations, we used both the multifactor dimensionality reduction (MDR)²¹ and the restricted partition methods (RPM).²² MDR is designed to detect high-order gene–gene or gene–environment interactions in relatively small samples. The method defines a new Boolean (high risk/low risk) variable summarizing information related to the multiloci and environmental informations. The new Boolean variable is then evaluated for its ability to classify and predict high- or low-risk status using cross validation testing. MDR analysis already showed its ability to detect susceptibility loci in various diseases.^{20,23–25} The RPM method uses a partitioning algorithm for determining predictors of a quantitative trait. In this study including 846 subjects from a general population, among 19 SNPs from 10 genes and 3 phenotypic traits (gender, adiponectinemia and body mass index (BMI)) we aimed at determining the relevant genetic and environmental variables combinations involved in the modulation of plasma LDL-C level.

MATERIALS AND METHODS

Subjects

Participants were recruited in the framework of the World Health Organization-Multinational mONItoring of trends and determinants of Cardiovascular diseases (WHO-MONICA) population survey conducted from 1995 to 1997 in the urban community of Lille in the North of France. The sample included 1195 representative subjects (601 men/594 women) aged 35–64 years, stratified by town size and randomly selected from the electoral rolls to obtain 200 participants for each gender and 10-year age group (WHO-MONICA Project protocol). The Ethical Committee of Lille University Hospital (CHRU) approved the protocol. After signing an informed consent, participants were administered a standard questionnaire including personal medical history. Physical measurements were taken by a specially trained nurse. A fasting blood sample was drawn for 1170 participants (590 men and 580 women). The studied sample consisted of a subgroup of 846 subjects (424 men and 422 women; mean age=49 ± 8 years, range 35–66; mean BMI=25.7 ± 4.4 kg m⁻², range 16.1–44.1) who were not treated for hypercholesterolemia, hypertension or type 2 diabetes mellitus. Adiponectinemia was measured using a commercial assay kit (LINCO Research, St Charles, MO, USA) as previously reported.¹⁰ Plasma LDL-C concentrations were calculated by the Friedewald formula.²⁶ The population was divided into four BMI classes: class 1—BMI < 25 kg m⁻²; class 2—25 ≤ BMI < 27 kg m⁻²; class 3—27 ≤ BMI < 30 kg m⁻²; and class 4—BMI ≥ 30 kg m⁻².

Genotyping

We genotyped SNPs with LightCycler LightTyper (Roche Diagnostics, Basel, Switzerland), or with TaqMan (Applied Biosystems, Foster City, CA, USA) or by direct sequencing. To avoid SNP genotyping errors, we systematically re-genotyped 10% of DNA samples for further verification. We found concordance rates of 100% for all SNPs. All genotypes fitted the Hardy–Weinberg equilibrium.

Haplotyping

Linkage disequilibrium (LD) and haplotype blocks were determined with the Haploview software.²⁷ Phase 2.1^{28,29} was used to construct haplotypes. Phase 2.1 implements a Bayesian statistical method to infer phase and to construct haplotypes from population genetics by Markov Chain–Monte Carlo algorithm and coalescent theory. It was shown to infer haplotypes more accurately than other Bayesian-based methods in real data sets.²⁹

MDR analysis

For evaluation of high-order interactions among genetic and environmental variables with a relatively small sample size and a large variable number we used

the MDR method.²¹ It includes a combined cross-validation procedure dividing the data into a training set and a testing set and thus minimizes false-positive results by multiple examinations of the data. With 10-fold cross-validations, the data are divided into 10 equal parts, and the model is developed on 9/10 of the data (training set). A set of n candidate variables is selected representing data in an n -dimensional space. The ratio number of cases/number of control subjects is evaluated within each multifactor cell in the n -dimensional space and thus cells are labeled as high or low risk according to the ratio of cases and controls. This reduces the n -dimensional model to one dimension (that is one variable with two multifactor classes: low risk and high risk). All possible combinations of n factors are evaluated sequentially for their ability to classify affected and unaffected subjects in the training set and the best n -variable model is then tested on the remaining 1/10 of the data (testing set). Steps are repeated 10 times with the data split into 10 different training and testing sets. In addition to the prediction error it is of interest to use cross-validation consistency as a measure of evidence for a particular model; that is how often were the same variables selected across the 10 cross-validations. We developed Perl scripts to perform repeated MDR analyses with numerous random partitions of the data between training and testing sets to better evaluate the reproducibility of the results.

RPM analysis

Conversely to MDR, RPM is an iterative algorithm devoted to the study of a quantitative trait. As reported by Culverhouse *et al.*³⁰ RPM searches among multiple genetic and environmental factors the combination that explains the highest part of the variance of a quantitative trait. It attempts to find the most reasonable partition for evaluation, balancing maximization of the between-group variation with minimization of the number of groups and the within-group variation. In RPM analyses, interaction between variables may be evidenced when R -square of a combined model is at least 15% higher than the sum of R -squares of the variables taken individually (R Culverhouse, personal communication).

RESULTS

Summarization of genetic variables using haplotype structures

In the general population studied here SNPs of the *ADIPOQ* gene showed strong LD defining two haplotype blocks: $-11391G > A / -11377C > G$ ($D' = 0.997$) and $+45T > G / +276G > T$ ($D' = 0.990$) as previously reported in other populations.^{10,31–33} For each block, the haplotypes including both rare variant alleles were never observed (theoretical frequencies of the $-11391A_-11377G$ and $+45G_+276T$ haplotypes were 0.000029 and 0.000072, respectively). To include haplotypic information in further analyses, haplotypes at the *ADIPOQ* locus were inferred to each subject using the Phase 2.1 software. Only diplotypes configurations inferred with a posterior probability > 0.95 were retained defining six diplotypes configurations for each haplotype block, thus, defining two variables. In agreement with previous studies reporting for the *ADIPOQ* gene, the presence of two LD blocks (in the promoter and the exon 2–intron 2) showing independent associations with adjusted adiponectin level,^{31,32,34} in our sample from a general population adiponectin level was modulated by *ADIPOQ* haplotypes from the above mentioned LD blocks ($P = 0.0013$ and 0.0058 , respectively). Among the other genes included in our study only *ENPP1* was genotyped for several SNPs. The five *ENPP1* SNPs defined two haplotype blocks, $K121Q/IVS8+27T > G$ ($D' = 0.999$) and $+828$ 13 bp insertion/ $+1044A > G$ ($D' = 0.999$) and one ‘independent’ *IVS20-11delT* not included in any of the haplotype blocks. Haplotypes were inferred for each block using the Phase 2.1 software. Diplotypes configurations inferred with a posterior probability > 0.95 were retained defining six diplotypes configurations for each haplotype block. Thus, three variables were used for the *ENPP1* gene in analyses: SNP *IVS20-11delT*, diplotypes of the $K121Q/IVS8+27T > G$ and $+828$ 13 bp insertion/ $+1044A > G$ blocks. Regarding the *SCARB1* gene, LD between the three SNPs was considered as weak (D' between 0.2 and

0.7) thus they were included as three independent variables in the analyses.

MDR analysis of the LDL-C level

We divided our population in two clusters using a combination of HDL-C and LDL-C as a marker of dyslipidemia. High- and low-risk clusters for dyslipidemia were separately defined for men and for women according to the median for each subgroup. Subjects were considered as at risk if they had simultaneously HDL-C \leq median of the subgroup and LDL-C \geq median of the subgroup. In a less stringent way we classified the subjects in two clusters according to the median of LDL-C alone in our population. Both classifications were in total concordance ($\kappa=0.999$), thus we opted to use the LDL-C criteria alone as a marker of dyslipidemia. For a greater contrast, instead using two groups according to the median, analyses were performed between the 1st and 3rd tertiles of LDL-C level as MDR requires a binary trait.

Table 1 summarizes the results of an exhaustive MDR analysis that evaluated all possible one, two and three variables model from gender, BMI, adiponectin level and the 15 genetic variables summarizing 19 SNPs from 10 genes. Although the model including BMI as only variable showed a cross-validation consistency of 10/10, the overall best model included BMI and diplotypes of *ADIPOQ* +45T>G/+276G>T with a cross-validation consistency of 10/10 and a slightly better testing accuracy than BMI alone (0.590 vs 0.575). This two variables model was significant at the 0.0004 level. The OR was 3.13 (95% CI, 2.20–4.46). One thousand repeated MDR analyses with random partitions of the data between training and testing sets showed BMI as the best one variable model, and BMI diplotypes of *ADIPOQ* +45T>G/+276G>T, as the best two variables model with 100% reproducibility. Moreover, the interaction dendrogram of the MDR software showed an interaction between BMI and diplotypes of *ADIPOQ* +45T>G/+276G>T (data not shown).

RPM analysis of the LDL-C level

To analyze the determinants of LDL-C with another method we opted for the RPM algorithm able to use a quantitative trait. We first explored the models including one explaining variable. Among the 18 variables tested only 5 significantly partitioned our population in groups of different mean LDL-C (Table 2). As expected, BMI was the best explaining variable as the proportion of variation attributable to the partition (*R*-square) was the highest observed. We further explored

Table 1 Results of MDR analyses with 3 phenotypic and 15 genetic variables summarizing 19 SNPs from 10 genes

Model	CVC	Testing accuracy	P-value
BMI	10	0.575	0.0009
BMI Diplotypes of <i>ADIPOQ</i> +45T>G/+276G>T	10	0.590	0.0004
BMI Diplotypes of <i>ADIPOQ</i> +45T>G/+276G>T Diplotypes of <i>ADIPOQ</i> -11391G>A/-11377C>G	7	0.576	0.002
BMI Diplotypes of <i>ADIPOQ</i> +45T>G/+276G>T diplotypes of <i>ADIPOQ</i> -11391G>A/-11377C>G Diplotypes of <i>ENPP1</i> ins13bp+828 /A>G+1044	6	0.500	0.119

Table 2 Results of univariate RPM analyses with 3 phenotypic and 15 genetic variables summarizing 19 SNPs from 10 genes

Variable	R-square	P-value
BMI	0.0365	0.000001
Gender	0.0188	0.000083
<i>UCP3</i>	0.0180	0.001
<i>CARTPT</i>	0.0113	0.008
Adiponectin level (tertiles)	0.0080	0.034

Only variables that significantly partitioned our population in groups of different mean LDL-C are displayed. Empirical *P*-values were determined following 10 000 Monte Carlo permutations.

Table 3 Results of bivariate RPM analyses with 3 phenotypic and 15 genetic variables summarizing 19 SNPs from 10 genes

Variables in model	R-square	P-value
BMI, diplotypes of <i>ADIPOQ</i> +45T>G/+276G>T	0.0601	0.00003
BMI, diplotypes of <i>ENPP1</i> +828 13 bp ins /+1044A>G	0.0592	0.00010
BMI, diplotypes of <i>ADIPOQ</i> -11391G>A/-11377C>G	0.0573	0.00004
BMI, gender	0.0529	0.00002
BMI, <i>CARTPT</i> -3608C>T	0.0517	0.00002
BMI, <i>SCARB1</i> G2S	0.0501	0.00002

From the 153 possible two variable models only the six best ones that significantly partitioned our population in groups of different mean LDL-C are displayed. Empirical *P*-values were determined following 100 000 Monte Carlo permutations.

the two variable models. Among the 153 possible models including two variables, the combination of BMI and diplotypes of *ADIPOQ* +45T>G/+276G>T showed the best *R*-square and was therefore assumed to be the best two variable model (Table 3). This model partitioned our population between a low (mean=3.57, s.d.=0.95) and a high LDL-C (mean=4.12, s.d.=1.02) group. The *R*-square of each variable alone was 0.0365 for BMI (Table 2) and was not different from zero for the diplotypes of *ADIPOQ* +45T>G/+276G>T (data not shown). Thus, *R*-square of the combined two variable model (*R*-square=0.0601) was 64% higher than the sum of *R*-squares (0.0365+0.000) of the variables taken individually, in agreement with an interaction between BMI and diplotypes of *ADIPOQ* +45T>G/+276G>T. A three variable model implies 816 combinations of variables for the 18 variables included. As our population consists of 846 subjects there is an obvious overfitting of the data and the results of the three variable models should not be taken into account.

We compared the classifications obtained using both methods. The results are summarized in Table 4. There was a good concordance between both methods ($\kappa=0.83$; 95% CI, 0.61–1.00). Moreover, knowing the best combination of variables and their interaction, we included BMI, diplotypes of *ADIPOQ* +45T>G/+276G>T and their interaction term in a general linear model and observed a statistically significant model ($P<0.0001$) and a significant interaction ($P=0.04$). There was no significant difference between adiponectin levels adjusted by gender and BMI between the clusters of high and low LDL-C defined by MDR or RPM ($P>0.05$).

DISCUSSION

Our data using two mining methods show that among the variables included, BMI and a particular genetic status at the *ADIPOQ* adiponectin encoding gene locus appear as the best variables explaining LDL-C level. LDL-C was chosen as it is, among others, a well established marker of atherogenesis. Moreover, in most

Table 4 Comparison of the classifications obtained with the MDR and the RPM methods as low and high LDL-C with the two variable models

BMI class	diplotypes of <i>ADIPOQ</i> +45T>G/+276G>T	RPM classification	MDR classification	Number of subjects
1	T_G / T_G	LOW	LOW	175
1	T_G / T_T	LOW	LOW	126
2	T_G / T_G	LOW	LOW	65
4	T_G / T_G	LOW	HIGH ^a	47
2	T_G / T_T	LOW	LOW	47
1	T_T / G_G	LOW	HIGH ^a	25
1	T_G / G_G	LOW	LOW	40
1	T_T / T_T	LOW	LOW	39
4	T_T / T_T	LOW	LOW	8
4	T_T / G_G	LOW	LOW	8
3	T_G / G_G	LOW	LOW	8
4	T_G / T_T	HIGH	HIGH	27
3	T_G / T_G	HIGH	HIGH	63
4	G_G / G_G	HIGH	HIGH	4
1	G_G / G_G	HIGH	HIGH	6
2	T_T / T_T	HIGH	HIGH	14
3	T_G / T_T	HIGH	HIGH	34
4	T_G / G_G	HIGH	HIGH	18
2	T_G / G_G	HIGH	HIGH	14
2	T_T / G_G	HIGH	HIGH	7
3	T_T / G_G	HIGH	HIGH	10
3	T_T / T_T	HIGH	HIGH	12
2	G_G / G_G	HIGH	HIGH	2
3	G_G / G_G	HIGH	HIGH	2

^aThe discrepancies only rely on 3 subjects differently classified in MDR.

epidemiologic studies related to cardiovascular disorders or to dyslipidemia, the LDL-C phenotype is systematically investigated in view of its potential association with genetic determinants. Therefore, in this study besides gender, BMI and adiponectinemia we focused our attention on 19 SNPs from the *PPARG*, *UCP3*, *ADIPOQ*, *TNF*, *LIPC*, *CARTPT*, *PCSK9*, *SCAP*, *SCARB1* and *ENPP1* genes that have all been reported to be associated with IR and/or plasma lipid levels. For the *ADIPOQ* and *ENPP1* genes where several SNPs were genotyped, we opted to use haplotypic structures as they better reflect the genetic architecture of the genes.^{18,32} In haplotype analyses the population is partitioned into a larger number of strata than in SNP analyses and could be associated with a lower power. Nevertheless, when haplotypes better capture the genetic variation at a given locus they are more efficient in analyses than SNPs alone. It is commonly admitted and largely previously reported that haplotypes at the *ADIPOQ* and at the *ENPP1* loci used in our study, better capture the genetic information than SNPs alone.^{9,18} LDL-C phenotype is a complex trait resulting from potential multiple gene-gene and gene-environment interactions. Most of previously reported association studies report investigations of one SNP or one gene at a time. Multigenic and multifactorial conventional approaches as multivariate logistic regression including many explaining variables have been proved to be limited by a lack of statistical power,²¹ or would require huge populations and only investigate a limited number of interactions. MDR and RPM methods allow determining the best combination of variables and interactions that explain either a binary trait (MDR) or a quantitative trait (RPM). The MDR method is nonparametric and does not presuppose any mathematical relation between the variables (that is linear relation in the logistic regression). The RPM method is a

robust method to examine quantitative phenotypes even if the loci have no single locus effect. Both independent methods showed quite similar results both on explaining variables and on the results of classification in low and high LDL-C groups. The best models with MDR and RPM classifications both included BMI and the genetic status at the *ADIPOQ* locus and their interaction. This model was ascertained a posteriori using a conventional statistical method. Regarding MDR analyses, improvement in testing accuracies scores when adding *ADIPOQ* diplotypes in the model was weak; this is not surprising as LDL-C is a multifactorial trait modulated among other by many genetic factors, each of them exhibiting a weak effect. Nevertheless, this weak effect was ascertained by RPM and by conventional statistical methods.

A slight discrepancy between MDR and RPM methods occurred for subjects with a BMI > 30 kg m⁻² (class 4) and a T_G/T_G *ADIPOQ* +45 +276 diplotype. It is noteworthy that this discrepancy only relies on 3 out of 47 subjects differently classified in MDR. Likewise, a discrepancy involving 3 subjects out of 25 occurred for subjects with a BMI under 25 kg m⁻² (class 1) and a T_T/G_G *ADIPOQ* +45 +276 diplotype. Regarding the variance explained as expected in one variable models, BMI was the best predictor with twice as more variance explained as the best genetic variable (*UCP3*). It is noteworthy that *ADIPOQ* +45 +276 diplotypes alone had no detectable effect on LDL-C. This was ascertained in classical ANOVA (data not shown). Including this genetic variable (diplotypes of *ADIPOQ*) and BMI in the same model allows to quite double the variance explained. This increase only relies on interaction between the two variables. To disclose similar findings with classical statistical methods would have required to test 306 analyses: 153 including two explaining variables and 153 including the interaction term as additional variable. In this context it is unlikely that a significant result would have been disclosed according to corrections for multiple testing. In this context these RPM and MDR methods appear as suitable opportunities to perform data mining to dissect complex diseases where multiple genetic determinants, environmental factors and their interactions may be involved.

Together with gender, BMI and genetic variables, adiponectin level was included in analyses. Surprisingly, if genotypes at the *ADIPOQ* adiponectin encoding gene were among the factors that associated with LDL-C level, adiponectin level itself was not. This is not amazing as only 3% of the variance of serum adiponectin level is explained by haplotypes at the *ADIPOQ* locus.⁹ As adiponectin levels adjusted by gender and BMI were similar in the clusters of high and low LDL-C defined by MDR or RPM, *ADIPOQ* genotypes classified as at risk for LDL-C level do not reflect a significant variation of adjusted adiponectin level and we can exclude that genotypes at the *ADIPOQ* locus influence LDL-C level through a modulation of adiponectin level. This is in agreement with previous results showing that if genetic variants of the *ADIPOQ* gene modulate adiponectin level, they are not contributing to the genetic linkage with the metabolic syndrome at the 3q27 locus.^{9,35} In addition, as adiponectin level is not a determinant of LDL-C level, it seems logical that the *PPARG* gene was not among the factors that modulated LDL-C level although the adiponectin gene contains a PPARgamma responsive element. Genetic variations in the remaining genes included in this study (*UCP3*, *TNF*, *LIPC*, *CARTPT*, *PCSK9*, *SCAP*, *SCARB1* and *ENPP1*), although individually associated with IR, metabolic syndrome and/or plasma LDL-C level, are not discriminant for LDL-C level in our multifactorial analyses.

As genome-wide scans and one meta-analysis reported highly significant linkage with coronary heart disease and LDL-C level at

the 3q27 locus^{36–39} we hypothesize that haplotype blocks of the *ADIPOQ* gene would capture genetic variation(s) from neighboring gene(s) that would be modulating metabolic syndrome, coronary heart disease and LDL-C level. More than 20 genes map to the 3q27 locus. If most do not appear as putative candidate (*FETUB*, *DNAJB11*, *CRYGS*, *RTP4*, *DGKG*...) several have already been reported to be associated with phenotypes of the metabolic syndrome; that is the alpha 2 Heremans–Schmid glycoprotein (*AHSG*) gene is associated with type 2 diabetes,⁴⁰ kininogen encoded by the *KNKI* gene is involved in insulin sensitivity at least in rodents,⁴¹ the eukaryotic translation initiation factor 4 alpha 2 (*EIF4A2*) gene contributes to the linkage with type 2 diabetes.⁴² Additional investigations including at least genetic variants from these genes will be required to better define the 3q27 contribution to metabolic syndrome, coronary heart disease and LDL-C level. Moreover, a similar analysis in another independent population sample would be very instructive. Anyway our data show that data mining methods such as MDR and RPM are quite suitable to dissect multifactorial traits in relatively small samples, and detect the most prominent determinants among numerous genetic and environmental variables and their complex interactions.

- 1 Sandhu, M. S., Waterworth, D. M., Debenham, S. L., Wheeler, E., Papadakis, K., Zhao, J. H. et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**, 483–491 (2008).
- 2 Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R. et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
- 3 Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P. et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
- 4 Pollex, R. L. & Hegele, R. A. Genetic determinants of plasma lipoproteins. *Nat. Clin. Pract. Cardiovasc. Med.* **4**, 600–609 (2007).
- 5 Wallace, C., Newhouse, S. J., Braund, P., Zhang, F., Tobin, M., Falchi, M. et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.* **82**, 139–149 (2008).
- 6 Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J. et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
- 7 Meirhaeghe, A., Fajas, L., Helbecque, N., Cottel, D., Auwerx, J., Deeb, S. S. et al. Impact of the Peroxisome Proliferator Activated Receptor gamma2 Pro12Ala polymorphism on adiposity, lipids and non-insulin-dependent diabetes mellitus. *Int. J. Obes. Relat. Metab. Disord.* **24**, 195–199 (2000).
- 8 Meirhaeghe, A., Amouyel, P., Helbecque, N., Cottel, D., Otabe, S., Froguel, P. et al. An uncoupling protein 3 gene polymorphism associated with a lower risk of developing Type II diabetes and with atherogenic lipid profile in a French cohort. *Diabetologia* **43**, 1424–1428 (2000).
- 9 Vasseur, F., Helbecque, N., Dina, C., Lobbens, S., Delannoy, V., Gaget, S. et al. Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum. Mol. Genet.* **11**, 2607–2614 (2002).
- 10 Vasseur, F., Helbecque, N., Lobbens, S., Vasseur-Delannoy, V., Dina, C., Clement, K. et al. Hypoadiponecตินaemia and high risk of type 2 diabetes are associated with adiponectin-encoding (ACDC) gene promoter variants in morbid obesity: evidence for a role of ACDC in diabetes. *Diabetologia* **48**, 892–899 (2005).
- 11 van der Vleuten, G. M., van Tits, L. J., den Heijer, M., Lemmers, H., Stalenhoef, A. F. & de Graaf, J. Decreased adiponectin levels in familial combined hyperlipidemia patients contribute to the atherogenic lipid profile. *J. Lipid Res.* **46**, 2398–2404 (2005).
- 12 Sookoian, S., Garcia, S. I., Gianotti, T. F., Dieuzeide, G., Gonzalez, C. D., Pirola, C. J. The G-308A promoter variant of the tumor necrosis factor-alpha gene is associated with hypertension in adolescents harboring the metabolic syndrome. *Am. J. Hypertens.* **18**, 1271–1275 (2005).
- 13 Isaacs, A., Sayed-Tabatabaei, F. A., Njajou, O. T., Wittman, J. C. & van Duijn, C. M. The -514 C->T hepatic lipase promoter region polymorphism and plasma lipids: a meta-analysis. *J. Clin. Endocrinol. Metab.* **89**, 3858–3863 (2004).
- 14 Guéradel, A., Barat-Houari, M., Vasseur, F., Dina, C., Vatin, V., Clement, K. et al. Analysis of sequence variability in the CART gene in relation to obesity in a Caucasian population. *BMC Genet.* **6**, 19 (2005).
- 15 Vasseur, F., Guéradel, A., Barat-Houari, M., Cottel, D., Amouyel, P., Froguel, P. et al. Impact of a CART promoter genetic variation on plasma lipid profile in a general population. *Mol. Genet. Metab.* **90**, 199–204 (2007).

- 16 Chen, S. N., Ballantyne, C. M., Gotto, Jr A. M., Tan, Y., Willerson, J. T. & Marian, A. J. A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. *J. Am. Coll. Cardiol.* **45**, 1611–1619 (2005).
- 17 Bacci, S., Ludovico, O., Prudente, S., Zhang, Y. Y., Di Paola, R., Mangiacotti, D. et al. The K121Q polymorphism of the ENPP1/PC-1 gene is associated with insulin resistance/atherogenic phenotypes, including earlier onset of type 2 diabetes and myocardial infarction. *Diabetes* **54**, 3021–3025 (2005).
- 18 Meyre, D., Bouatia-Naji, N., Tounian, A., Samson, C., Lecoecur, C., Vatin, V. et al. Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat. Genet.* **37**, 863–867 (2005).
- 19 Iwaki, K., Nakajima, T., Ota, N. & Emi, M. A common Ile796Val polymorphism of the human SREBP cleavage-activating protein (SCAP) gene. *J. Hum. Genet.* **44**, 421–422 (1999).
- 20 Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F. et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* **69**, 138–147 (2001).
- 21 Hahn, L. W., Ritchie, M. D. & Moore, J. H. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* **19**, 376–382 (2003).
- 22 Culverhouse, R., Klein, T. & Shannon, W. Detecting epistatic interactions contributing to quantitative traits. *Genet. Epidemiol.* **27**, 141–152 (2004).
- 23 Meng, Y., Ma, Q., Yu, Y., Farrell, J., Farrer, L. A. & Wilcox, M. A. Multifactor-dimensionality reduction versus family-based association tests in detecting susceptibility loci in discordant sib-pair studies. *BMC Genet.* **6** (Suppl 1), S146 (2005).
- 24 Cho, Y. M., Ritchie, M. D., Moore, J. H., Park, J. Y., Lee, K. U., Shin, H. D. et al. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia* **47**, 549–554 (2004).
- 25 Williams, S. M., Ritchie, M. D., Phillips, III J. A., Dawson, E., Prince, M., Dzhura, E. et al. Multilocus analysis of hypertension: a hierarchical approach. *Hum. Hered.* **57**, 28–38 (2004).
- 26 Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
- 27 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2004).
- 28 Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
- 29 Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
- 30 Culverhouse, R., Hinrichs, A. L., Jin, C. H. & Suarez, B. K. Gene × gene and gene × environment interactions for complex disorders. *BMC Proc.* **1** (suppl 1), S72 (2007).
- 31 Hara, K., Boutin, P., Mori, Y., Tobe, K., Dina, C., Yasuda, K. et al. Genetic variation in the gene encoding adiponectin is associated with an increased risk of type 2 diabetes in the Japanese population. *Diabetes* **51**, 536–540 (2002).
- 32 Heid, I. M., Wagner, S. A., Gohlke, H., Iglseider, B., Mueller, J. C., Cip, P. et al. Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1727 healthy caucasians. *Diabetes* **55**, 375–384 (2006).
- 33 Menzaghi, C., Ercolino, T., Di Paola, R., Berg, A. H., Warram, J. H., Scherer, P. E. et al. A haplotype at the adiponectin locus is associated with obesity and other features of the insulin resistance syndrome. *Diabetes* **51**, 2306–2312 (2002).
- 34 Woo, J. G., Dolan, L. M., Deka, R., Kaushal, R. D., Shen, Y., Pal, P. et al. Interactions between noncontiguous haplotypes in the adiponectin gene ACDC are associated with plasma adiponectin. *Diabetes* **55**, 523–529 (2006).
- 35 Gibson, F. & Froguel, P. Genetics of the APM1 locus and its contribution to type 2 diabetes susceptibility in French Caucasians. *Diabetes* **53**, 2977–2983 (2004).
- 36 Chiodini, B. D. & Lewis, C. M. Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1863–1868 (2003).
- 37 Francke, S., Manraj, M., Lacquemant, C., Lecoecur, C., Lepretre, F., Passa, P. et al. A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum. Mol. Genet.* **10**, 2751–2765 (2001).
- 38 Harrap, S. B., Zammit, K. S., Wong, Z. Y., Williams, F. M., Bahlo, M., Tonkin, A. M. et al. Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler. Thromb. Vasc. Biol.* **22**, 874–878 (2002).
- 39 Rainwater, D. L., Almasy, L., Blangero, J., Cole, S. A., VandeBerg, J. L., MacCluer, J. W. et al. A genome search identifies major quantitative trait loci on human chromosomes 3 and 4 that influence cholesterol concentrations in small LDL particles. *Arterioscler. Thromb. Vasc. Biol.* **19**, 777–783 (1999).
- 40 Siddiq, A., Lepretre, F., Hercberg, S., Froguel, P. & Gibson, F. A synonymous coding polymorphism in the alpha2-Heremans-schmid glycoprotein gene is associated with type 2 diabetes in French Caucasians. *Diabetes* **54**, 2477–2481 (2005).
- 41 Damas, J., Bourdon, V. & Lefebvre, P. J. Insulin sensitivity, clearance and release in kininogen-deficient rats. *Exp. Physiol.* **84**, 549–557 (1999).
- 42 Cheyssac, C., Dina, C., Lepretre, F., Vasseur-Delannoy, V., Dechaume, A., Lobbens, S. et al. EIF4A2 is a positional candidate gene at the 3q27 locus linked to type 2 diabetes in French families. *Diabetes* **55**, 1171–1176 (2006).