

# Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p

Helen M. Broadbent<sup>1,†</sup>, John F. Peden<sup>1,†</sup>, Stefan Lorkowski<sup>3</sup>, Anuj Goel<sup>1</sup>, Halit Ongen<sup>1</sup>, Fiona Green<sup>4</sup>, Robert Clarke<sup>2</sup>, Rory Collins<sup>2</sup>, Maria Grazia Franzosi<sup>5</sup>, Gianni Tognoni<sup>6</sup>, Udo Seedorf<sup>3</sup>, Stephan Rust<sup>3</sup>, Per Eriksson<sup>7</sup>, Anders Hamsten<sup>7</sup>, Martin Farrall<sup>1,\*</sup> and Hugh Watkins<sup>1</sup> for the PROCARDIS consortium<sup>‡</sup>

<sup>1</sup>Department of Cardiovascular Medicine and <sup>2</sup>Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, UK, <sup>3</sup>Leibniz-Institut für Arterioskleroseforschung an der Universität Münster, Münster, Germany, <sup>4</sup>Faculty of Health and Medical Sciences, University of Surrey, UK, <sup>5</sup>Department of Cardiovascular Research, 'Mario Negri' Institute for Pharmacological Research, Milano, Italy, <sup>6</sup>Consorzio Mario Negri Sud, Santa Maria Imbaro, Italy and <sup>7</sup>Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

Received November 9, 2007; Revised and Accepted November 28, 2007

Genome-wide association studies have identified a region on chromosome 9p that is associated with coronary artery disease (CAD). The region is also associated with type 2 diabetes (T2D), a risk factor for CAD, although different SNPs were reported to be associated to each disease in separate studies. We have undertaken a case-control study in 4251 CAD cases and 4443 controls in four European populations using previously reported ('literature') and tagging SNPs. We replicated the literature SNPs ( $P = 8 \times 10^{-13}$ ; OR = 1.29; 95% CI: 1.20–1.38) and showed that the strong consistent association detected by these SNPs is a consequence of a 'yin-yang' haplotype pattern spanning 53 kb. There was no evidence of additional CAD susceptibility alleles over the major risk haplotype. CAD patients without myocardial infarction (MI) showed a trend towards stronger association than MI patients. The CAD susceptibility conferred by this locus did not differ by sex, age, smoking, obesity, hypertension or diabetes. A simultaneous test of CAD and diabetes susceptibility with CAD and T2D-associated SNPs indicated that these associations were independent of each other. Moreover, this region was not associated with differences in plasma levels of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, fibrinogen, albumin, uric acid, bilirubin or homocysteine, although the CAD-high-risk allele was paradoxically associated with lower triglyceride levels. A large antisense non-coding RNA gene (ANRIL) collocates with the high-risk haplotype, is expressed in tissues and cell types that are affected by atherosclerosis and is a prime candidate gene for the chromosome 9p CAD locus.

## INTRODUCTION

Coronary artery disease (CAD) is the commonest cause of death worldwide and has a rapidly increasing incidence in

nations undergoing industrial development (1). CAD is a multifactorial disease in which multiple genes and environmental factors are believed to influence disease risk. Candidate gene, positional cloning and comparative mapping techniques have

\*To whom correspondence should be addressed at: Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. Tel: +44 1865287601; Email: martin.farrall@cardiov.ox.ac.uk

<sup>†</sup>To authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

<sup>‡</sup>See www.procarris.org for full membership of PROCARDIS consortium.

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

identified a number of susceptibility genes for CAD, but the identity of the genes explaining much of the heritability of CAD remains unknown (1).

Recent genome-wide association (GWA) studies have tested the common disease/common variant (CD/CV) hypothesis and successfully identified susceptibility genes for a number of multifactorial diseases including CAD. The most striking CAD result involves a genomic region on chromosome 9 approximately 22 million base pairs from the 9p telomere (2–5). Single nucleotide polymorphisms (SNPs) showing the strongest association with CAD did not map within an annotated gene sequence leading to the nomination of neighbouring genes (CDKN2A, CDKN2B and MTAP) as candidate susceptibility genes.

Contemporaneous GWA studies of type 2 diabetes (T2D) susceptibility identified the same region on chromosome 9 in several (6–8) but not all (9,10) populations. Diabetes mellitus (DM) has long been recognized as a major risk factor for atherosclerosis with a complex pathogenesis involving multiple biological processes. For instance, increased reactive oxidant stress in diabetic arterial walls contributes to endothelial damage, the earliest stage in the development of an atherosclerotic plaque. Subsequently, features of the insulin resistance syndrome associated with T2D such as hyperlipidaemia and hypertension exacerbate the damage and lead to disease progression. The temporal relationship of DM with CAD is complex; DM may precede the onset of CAD symptoms, or be diagnosed *de novo* as a consequence of a CAD clinical presentation or appear shortly after a major CAD event such as myocardial infarction (MI) (11,12). Consequently, the co-localization of CAD and T2D susceptibility genes would seem, at first glance, as likely to point to a tightly coupled pathophysiological mechanism. However, the reports of CAD and T2D susceptibility (in non-overlapping patient populations) apparently involve different genetic markers. Moreover, quantitative genetic analysis of intermediate phenotypes associated with the pro-atherogenic insulin resistance phenotype detected no obvious QTL effects with CAD-associated SNPs. These intriguing findings motivate studies that aim to disentangle the susceptibility effects encoded by this chromosome 9p CAD/T2D susceptibility locus.

The PROCARDIS consortium was conceived to study the complex genetics of CAD susceptibility in Europeans and has assembled clinical resources to undertake genetic association studies. We have undertaken a genetic analysis of markers spanning the chromosome 9p susceptibility region in 4251 cases and 4443 controls to replicate the CAD locus, interpret the pattern of association in the context of European haplotype diversity and investigate if established risk factors (in particular diabetes) modify the CAD susceptibility effect.

## RESULTS

### SNP selection

SNPs were selected on the basis of earlier reports of association to CAD or to tag additional genetic variation based on the HapMap CEU (CEPH European ancestry) database. First, the chromosome 9p region was partitioned into several sections

spanning 403.8 kb (21 790 000–22 193 800 bp): a 68 kb core section containing SNPs reported to show strong association with CAD and T2D (22 057 000–22 125 000 bp); a broader 185 kb region containing SNPs showing moderate association to CAD (21 940 000–22 125 000); two sections incorporating adjacent linkage disequilibrium (LD) blocks (150 kb upstream and 68.8 kb downstream). All published associated SNPs were selected for genotyping (Table 1) except for two where an alternative in strong LD ( $r^2 \geq 0.9$ ) was used instead (rs10811650 was replaced by rs10811647 and rs6475608 was replaced by rs4977756). HaploView (13) was used to select additional tagging SNPs with MAF > 0.2 and  $r^2 \geq 0.9$  for the core region and  $r^2 \geq 0.5$  across the 403.8 kb region. The list of SNPs genotyped is shown in Supplementary Material, Table S1.

### Replication of SNP associations with CAD

Table 1 shows the results of an analysis of 11 SNPs previously reported to show strong association to CAD or T2D in single marker tests (literature SNPs). The PROCARDIS cases ( $n = 4251$ ) were patients with documented diagnoses of MI, symptomatic acute coronary syndrome, intervention for coronary revascularization, or chronic stable angina with an age at diagnosis of 65 years or less. The controls ( $n = 4,443$ ) were self-reportedly free of coronary disease at age 65 and without any sibling history of CAD. Two CAD literature SNPs (rs518394 and rs1333040) were genotyped but showed marked distortion of Hardy–Weinberg equilibrium in the control group ( $P \leq .0005$ ) and were eliminated from further analysis. The seven remaining CAD literature SNPs showed strong evidence of association ( $8 \times 10^{-13} \leq P \leq 3 \times 10^{-11}$ ) with strikingly similar odds ratio estimates ranging from 1.29 to 1.26. There was no evidence of non-additivity ( $P \geq 0.36$ ) on the logit scale consistent with an allelic-effects-only association model. There was also no evidence of heterogeneity of the association across the four European populations ( $P \geq 0.86$ ). Four T2D literature SNPs were genotyped and passed the QC criteria, one SNP (rs564398) showed significant ( $P = 4 \times 10^{-8}$ ) association to CAD with no evidence of heterogeneity ( $P = 0.27$ ) across the countries. SNP rs564398 was in moderate LD with the 7 CAD literature SNPs ( $0.29 \leq r^2 \leq 0.41$ ).

### Identification of a parsimonious SNP association model

A further 21 tagging SNPs were genotyped in the complete PROCARDIS panel of 4251 cases and 4443 controls. These SNPs were combined with the literature SNPs described above to form a model selection data set; the overall genotype call-rate across the 32 SNPs was 98.8%. However, 19% of the genotypes were incomplete for one SNP and 7% for between two and five SNPs. The strong LD between some of the SNPs (details shown below) motivated a combined haplotyping/genotype imputation analysis to infer missing genotypes using information from flanking markers. Following haplotype and genotype imputation with the MACH program using a quality score threshold of 0.85, the proportion of incomplete genotypes was halved to 13% so overall, 99.5% of genotypes were available for analysis.

**Table 1.** Association analysis of CAD with literature SNPs

SNP	Position (bp)	HWE $\chi^2$	Risk allele	freq.	OR	95% CI	Additive effects $-\log_{10}(P\text{-value})$	Non-additive effects $P\text{-value}$	het. $P\text{-value}$	Selection criterion of reported association	Publication
rs564398	22 019 547	7.0	T	0.58	1.21	1.13–1.30	7.4	0.07	0.27	T2D	Scott&Zeggini
rs10116277	22 071 397	0.5	T	0.48	1.28	1.19–1.37	11.6	0.36	0.94	CAD	Helgadottir
rs6475606	22 071 850	0.4	T	0.48	1.27	1.19–1.36	11.5	0.41	0.93	CAD	McPherson&WTCCC
rs10757274	22 086 055	0.3	G	0.48	1.29	1.20–1.38	12.1	0.69	0.86	CAD	McPherson
rs2383206	22 105 026	0.2	G	0.51	1.26	1.18–1.35	10.5	0.50	0.97	CAD	McPherson
rs2383207	22 105 959	0.0	G	0.51	1.27	1.18–1.36	10.8	0.37	0.97	CAD	Helgadottir
rs10757278	22 114 477	1.2	G	0.47	1.27	1.19–1.37	11.3	0.62	0.89	CAD	Helgadottir
rs1333049	22 115 503	1.1	C	0.48	1.27	1.19–1.36	11.1	0.72	0.86	CAD	McPherson&WTCCC
rs2383208	22 122 076	0.4	G	0.20	1.02	0.93–1.11	0.2	0.51	0.43	T2D	Scott
rs10811661	22 124 094	1.0	T	0.82	1.02	0.93–1.11	0.2	0.62	0.59	T2D	Scott&Zeggini
rs10757283	22 124 172	0.0	T	0.43	1.07	1.00–1.15	1.3	0.67	0.73	T2D	Zeggini

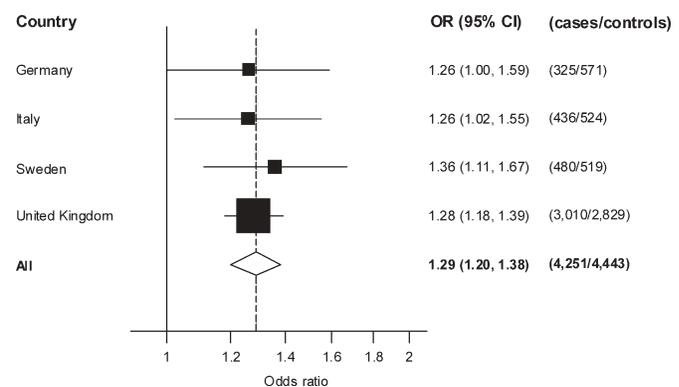
HWE  $\chi^2$ , Hardy–Weinberg equilibrium  $\chi^2$  statistic; freq., frequency of risk allele; het.  $P\text{-value}$ , fixed-effects heterogeneity test  $P\text{-value}$ .

A forward stepwise model selection procedure was applied to identify a subset of SNPs with strong main-effects association to CAD. In this conditional association analysis, modelling additive (logit scale) genetic effects, a single SNP, rs2891168 was identified in the combined-countries data using an entry  $P\text{-value}$  threshold of 0.01. Association between CAD and rs2891168 was strongly supported ( $P = 6 \times 10^{-13}$ ) and the per-G-allele odds ratio = 1.29 (95% CI 1.20–1.38). There was no evidence for non-additive (logit scale) susceptibility effects ( $P = 0.69$ ); there was also no evidence of heterogeneity of susceptibility across the four European populations (Cochrane's  $Q = 0.37$ , 3 d.f.,  $P = 0.95$ ) (Fig. 1).

In order to extend the search for SNPs that complement the susceptibility defined by rs2891168, 32 additional tagging/fine-mapping SNPs were genotyped in a subset of the PROCARDIS cases ( $n = 1699$ ) and controls ( $n = 1896$ ). One of these SNPs, rs12379111, was excluded from further analysis on the basis of a low (72%) genotype call-rate. A combined haplotype and genotype imputation analysis was again used to maximize the genetic information as 26% of genotypes were incomplete for 1–5 SNPs; this was reduced to 9% after imputation. A forward stepwise model selection procedure was then applied with rs2891168 forcibly included as an additive-effect in the model. This analysis failed to identify any further associated SNPs with significant main-effects ( $P > 0.01$ ).

### Common haplotypes delimit CAD susceptibility

Supplementary Material, Figure S1 shows a representation of the pairwise LD coefficients (scaled as  $r^2$ ) for 62 SNPs spanning the chromosome 9p CAD locus based on a combined (cases and controls) data set. There is a region of strong LD with flanking SNPs rs10116277 (22 071 397 bp) and rs1333049 (22 115 503 bp) in which 13 SNPs show  $r^2 \geq 0.77$  with rs2891168. Haplotype analysis (Table 2) shows that these 14 SNPs form four common haplotypes (cumulative frequency = 89%) in which a perfect complementary yin–yang pattern is maintained (14). All seven CAD literature



**Figure 1.** CAD association with rs2891168 in four European populations. Horizontal lines indicate 95% confidence intervals (CI) for odds ratio (OR) estimates in each country. The overall (fixed-effects) estimate of the OR is shown by a vertical dashed line, the 95% CI for this estimate is shown by an unshaded rhombus. Solid squares are centered on the OR estimate for each country and scaled in proportion to the number of participants in each country (number of cases and controls shown in parentheses).

SNPs reported in Table 1 comply with the yin–yang pattern. An analysis of this region in the CEPH European HapMap data shows that this striking LD pattern is maintained across 37 SNPs from rs10757269 (22 062 264 bp) to rs1333049 (22 115 503 bp) (Supplementary Material, Table S2). There is also no evidence of long range LD on chromosome 9 with an  $r^2 \geq 0.5$  criterion (Dr Robert Lawrence, WTCHG, personal communication).

### Susceptibility in disease and risk factor subgroups

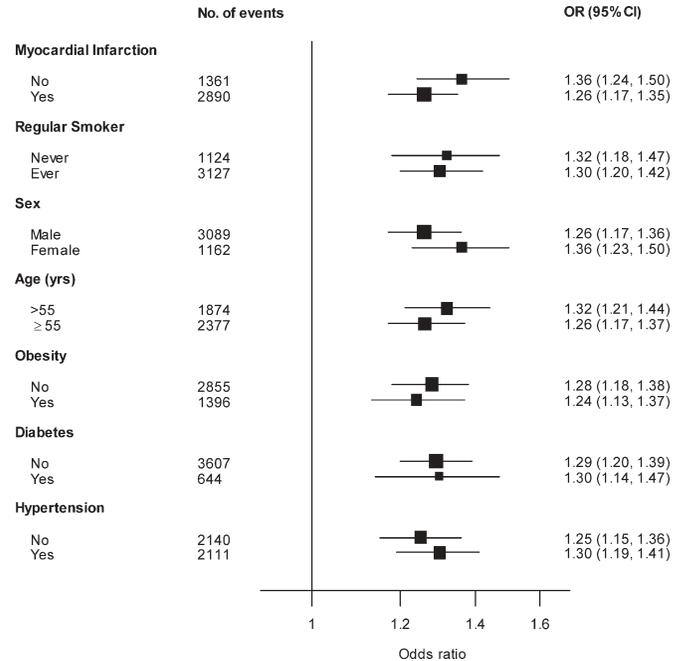
A multinomial logistic regression model was used to test if susceptibility to CAD detected by rs2891168 might vary in CAD patients with different clinical diagnostic or risk factor profiles (Fig. 2). CAD patients ( $n = 1361$ ) with no clinical history of MI showed a trend towards stronger association than CAD patients who had suffered MI ( $n = 2890$ ) (OR = 1.36 versus 1.26;  $\chi^2 = 3.1$ , 1 d.f.;  $P = 0.08$ ). There were no

**Table 2.** Haplotype analysis of SNPs in strong LD with rs2891168

rs10116277	rs10965227	rs6475606	rs1547705	rs10738607	rs10757272	rs10757274	rs4977574	rs2891168	rs1333042	rs2383206	rs2383207	rs1333045	rs10757278	rs1333048	rs1333049	freq.
T	A	T	A	G	T	G	G	G	G	G	C	C	C	C	C	0.35
G	A	C	A	A	C	A	A	A	A	A	T	A	A	G	G	0.26
G	G	C	A	A	C	A	A	A	A	A	T	A	A	G	G	0.16
T	A	T	C	G	T	G	G	G	G	G	C	G	C	C	C	0.12

SNPs coupled to two complementary (yin-yang) haplotypes are shown in bold black type, SNPs in grey type define subhaplotypes. The uppermost (most frequent) haplotype denotes the CAD risk allele.

freq., maximum likelihood estimate of the haplotype frequency.

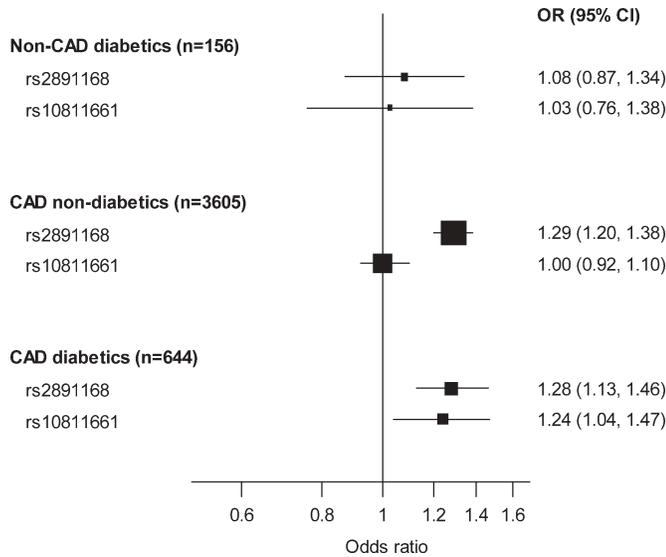


**Figure 2.** Association between CAD and rs2891168 in clinical and risk factor subgroups. Solid squares centered on the OR estimate and scaled in proportion to sample size with 95% CI (horizontal bar) are shown for subgroups contrasting CAD with or without a history of MI, history of regular smoking, sex, age at first clinical presentation, obesity, clinical history of diabetes or hypertension.

significant differences in CAD susceptibility detected by rs2891168 contrasting those who had regularly smoked with those who had never regularly smoked ( $P = 0.98$ ), male versus female CAD patients ( $P = 0.14$ ), CAD patients with age-of-onset earlier or later than age 55 ( $P = 0.34$ ), with BMI greater or less than  $30 \text{ kg/m}^2$  ( $P = 0.53$ ), self-reported history of diabetes ( $P = 0.82$ ) or hypertension ( $P = 0.47$ ).

A further multinomial logistic regression model was fitted to simultaneously assess association between CAD and rs2891168 and diabetes status and rs10811661 (Fig. 3). SNP rs10811661 was selected as it showed the strongest and most consistent association in a meta-analysis of T2D GWA studies (8) and maps outside of the CAD-associated haplotype (Fig. 4). CAD and diabetes status are strongly positively correlated (tetrachoric correlation =  $0.45 \pm 0.02$ ) in the PROCARDIS populations and rs2891168 and rs10811661 are in LD ( $r^2 = 0.008$ ). Using a reference group of non-CAD/non-diabetics, the rs10811661/CAD/diabetics logit was significantly ( $P = 0.016$ ) higher than the rs10811661/CAD/non-diabetics logit consistent with the reported association of the T-allele with T2D risk. A nested model with linear constraints designed to specify marginal rs2891168-CAD and rs10811661-diabetes effects only, was not rejected in a likelihood ratio test ( $\chi^2 = 1.8, 4 \text{ d.f.}, P = 0.77$ ). Thus the two SNPs are independently associated with the two separate diseases in the PROCARDIS population.

Biochemical intermediate phenotypes measured in CAD cases were tested for QTL effects with rs2891168 (Table 3). There was no evidence of association to plasma levels of low-density lipoprotein (LDL)-cholesterol, high-density



**Figure 3.** Association to CAD and diabetes with rs2891168 and rs10811661. Solid squares centered on the relative risk ratio estimate and scaled in proportion to sample size with 95% CI (horizontal bar) for subgroups contrasting CAD with or without a clinical history of diabetes. The reference group is non-CAD, non-diabetics ( $n = 4305$ ). The number of individuals in each subgroup ( $n$ ) is shown.

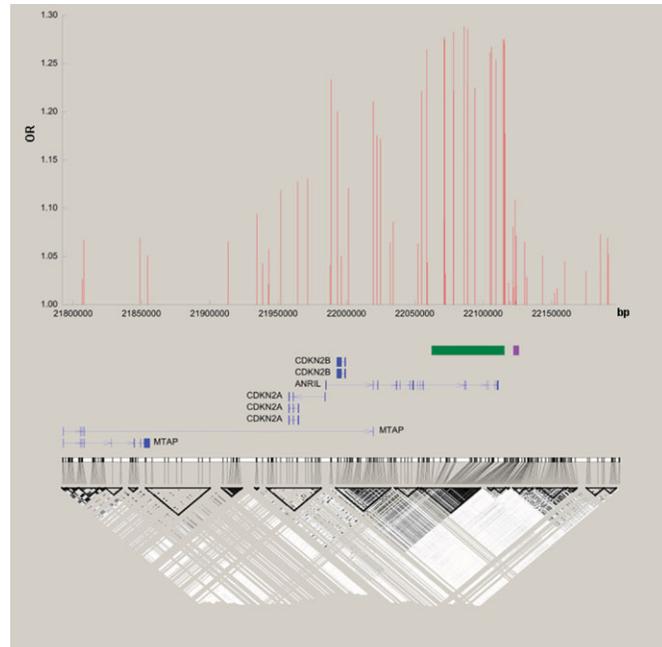
lipoprotein (HDL) cholesterol, lipoprotein(a), albumin, fibrinogen, uric acid or homocysteine (all  $P > 0.11$ ). There was evidence of association with triglyceride levels ( $P = 0.009$ ); the high-CAD risk G-allele was associated with lower levels of this lipid.

### ANRIL is expressed in cells and tissues affected by atherosclerosis

Primers were designed to PCR amplify exons 17, 18 and 19 of the 3'-UTR of the reported ANRIL transcript DQ485453 (15). These exons fall within the CAD-associated haplotype (Fig. 4) and are represented by three ESTs (BX100299, A1825844, AW169296). Total RNA was extracted, and amplified by RT-PCR, from human cells and tissues of potential relevance to CAD: human umbilical vein endothelial (HUVEC) and human microvascular endothelial (HMEC) cell lines; HepG2, Jurkat and THP1 cell lines; primary coronary smooth muscle and human monocyte-derived macrophage cell cultures and carotid endarterectomy (CEA) and abdominal aortic aneurysm (AAA) samples. Expression was shown in the CEA and AAA tissue samples, and also the HUVEC and HMEC endothelial cells, macrophages, HepG2 and coronary smooth muscle cells but not the Jurkat and THP1 cell lines (Fig. 5).

## DISCUSSION

It is encouraging to report that the associations between SNPs on chromosome 9p and CAD risk are convincingly and independently confirmed in the PROCARDIS samples (Table 1).



**Figure 4.** The chromosome 9p CAD region. The upper panel shows odds ratio (OR) estimates in red plotted against physical location (bp) for 62 SNPs tested in the PROCARDIS case-control study. The exon-intron structure of annotated genes are shown in blue; the location of the CAD-associated haplotype is shown in green; the location of the T2D-associated SNP (rs10811661) is shown in purple. The lower panel summarizes the linkage disequilibrium (LD) structure in the HapMap database (CEU European ancestry). Regions of LD are shaded in grey (moderate LD) or black (strong LD). Predicted haplotype blocks are framed by triangles. The physical location of the HapMap SNPs is shown above the haplotype blocks.

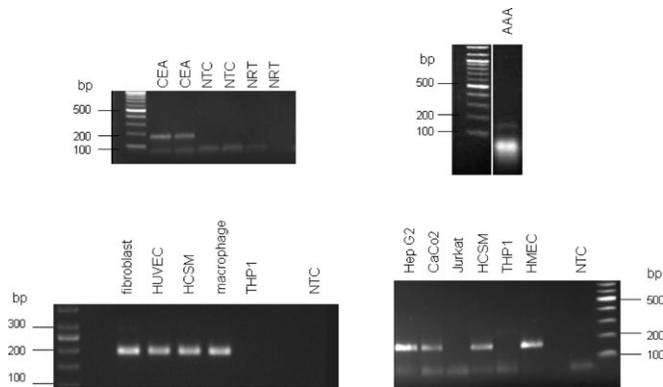
Replication of candidate gene associations can prove very difficult (16) and consistent susceptibility effects are needed to provide a firm rationale for further biological experimentation. The magnitude of the susceptibility effect in the four European populations ( $OR = 1.29$ ) is very similar to that reported in the GWA studies. The absolute magnitude of the susceptibility effect is modest in comparison to classic risk factors (e.g. smoking,  $OR > 3$ ) noting that this locus conferred the largest effect detected in three GWA studies to date.

The results of the different GWA studies that identified the chromosome 9p CAD locus included partially overlapping sets of associated SNPs. We show here that the strongest associations are coupled to a pair of high and low risk haplotypes that constitute a yin-yang pattern (Table 2 and Fig. 4) (14). An analysis of the HapMap database suggests that this haplotype pattern is extended across a region of at least 53 kb. Given the apparent low frequency of minor haplotypes that would be informative for further fine-mapping, the prospects for further resolving the susceptibility within CAD populations of European ancestry are limited. Indeed even large-scale re-sequencing is unlikely to contribute much. Using models of conditional association, we sought evidence for secondary risk alleles in moderate or low LD with the major susceptibility region. This was unsuccessful using a P-entry criterion of 0.01. We applied a forward stepwise procedure which implements a 'greedy' algorithm that is prone to

**Table 3.** QTL analysis of biochemical risk factors for CAD

Phenotype	Transformation or variance adjustment	rs2891168 (per G allele effect)					Covariates included in model
		coefficient	S.E.	F	d.f.	P-value	
LDL-C	Log	$-1.7 \times 10^{-2}$	$2.6 \times 10^{-2}$	0.4	750	$5.1 \times 10^{-1}$	Sex
HDL-C	Log	$3.3 \times 10^{-2}$	$2.1 \times 10^{-2}$	2.5	748	$1.1 \times 10^{-1}$	Sex, age, BMI
TG	Log	$-1.0 \times 10^{-1}$	$3.9 \times 10^{-2}$	6.9	749	$9.1 \times 10^{-3}$	BMI
Lp(a)	Log	$-3.5 \times 10^{-3}$	$5.4 \times 10^{-2}$	0.0	751	$9.5 \times 10^{-1}$	Sex
Albumin	Bootstrap	$1.7 \times 10^1$	$1.9 \times 10^1$	0.9	752	$3.5 \times 10^{-1}$	Sex, age
Fibrinogen	Bootstrap	$1.3 \times 10^{-2}$	$2.3 \times 10^{-2}$	0.3	1817	$5.8 \times 10^{-1}$	Sex, age, BMI, current smoking, cotinine
Homocysteine	1/sqrt	$-3.6 \times 10^{-4}$	$2.3 \times 10^{-3}$	0.0	750	$8.9 \times 10^{-1}$	Sex, age, cotinine
Bilirubin	Log	$8.0 \times 10^{-3}$	$1.3 \times 10^{-2}$	0.4	1830	$5.3 \times 10^{-1}$	Sex, age, BMI, current smoking, cotinine
Urate	Sqrt	$5.4 \times 10^{-3}$	$7.3 \times 10^{-3}$	0.6	1839	$4.6 \times 10^{-1}$	Sex, age, BMI, current smoking

coefficient, regression coefficient scaled in transformed or original units; S.E., standard error of regression coefficients; F, F-statistic (numerator degrees of freedom=1); d.f., denominator degrees of freedom.



**Figure 5.** RT-PCR analysis of ANRIL expression. NRT, no-reverse transcriptase control; NTC, no-template control; CEA, carotid endarterectomy; AAA, abdominal aortic aneurysm; HUVEC, human umbilical vein endothelial cells; HCSM, human coronary artery smooth muscle; Hep G2, human hepatocellular carcinoma cell lines; CaCo2, human colon cancer cell line; Jurkat, immortalized T-cell line; THP1, human acute monocytic leukaemia cell line; HMEC, human microvascular endothelial cell line.

overestimate the magnitude of regression parameters of predictors included in a model, with the consequence that secondary predictors might be overlooked (i.e. type II error). However, relaxing the P-entry criterion could easily lead to over-fitting (type I error), a serious possibility as 62 SNPs were available for testing. We therefore interpret our results to show that at least the great majority of the CAD susceptibility is encoded by two common haplotypes spanning a 53 kb region and that any linked minor variants will be associated with relatively small effects.

The CAD association appears remarkably uniform with no evidence of gene-environment interactions in that it does not appear to be modified by age, sex, smoking history, obesity or presence of hypertension or diabetes. The susceptibility effect in CAD cases with no clinical history of MI showed a trend towards stronger association compared with those who had survived an infarct. This result would seem to exclude the chromosome 9p CAD locus as a 'plaque rupture', or thrombosis, gene. On the other hand, genes acting at a very early stage in the development of an

atheromatous plaque might be predicted to show similar risks in CAD patients with or without MI. So speculatively, this locus might promote the development of stable, obstructive, plaques and it will be interesting to see if the subtle trend reported here is confirmed in other collections of CAD patients.

The reports of T2D susceptibility detected by different SNPs to those reported in CAD studies were tested in a multinomial logit model in which a CAD associated SNP (rs2891168) and a T2D associated SNP (rs10811661) were simultaneously fitted to a joint CAD/diabetes outcome. The results of this analysis are consistent with the hypothesis that the two susceptibility effects are independent; this is the first time this has been formally tested. A limitation of the present study was the inability to distinguish sub-types of diabetes, but the proportion of diabetes in the PROCARDIS population attributable to T2D is likely to exceed 90% based on prevalence data and the age-range of the population. Furthermore, although there are a considerable number of informative patients with CAD and diabetes in the PROCARDIS collection, the numbers of non-CAD diabetics are relatively low which inevitably limits the power of this study to reject the null hypothesis of independence. Case-control collections designed to assess association of susceptibility genes with CAD such as PROCARDIS are typically based on the disease criterion of primary interest to the investigators; it would be very difficult to ascertain and assemble large numbers of CAD cases and controls in a balanced design across (potentially multiple) risk factor subgroups. Progress therefore may be dependent on large-scale prospective studies such as UK Biobank ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)) which have the potential to provide nested-case-control samples (for example CAD cases and controls with or without T2D or other risk factors) to test such hypotheses.

There was no evidence of an influence of the CAD risk haplotype on quantitative variation for LDL-cholesterol or HDL-cholesterol levels, lipoprotein(a), fibrinogen or albumin, homocysteine or bilirubin in PROCARDIS cases. Consequently, this genetic susceptibility to CAD appears not to be mediated by influences on known pathogenic mechanisms such as altered lipid metabolism or inflammation. There was some evidence of a QTL effect for triglyceride levels noting that the CAD high-risk allele was associated with lower

lipid levels. Triglyceride has been reported as an independent risk factor for CAD and is a component of the metabolic syndrome; higher levels of triglyceride are associated with increased CAD risk (17–20). Thus the risk conferred by this locus is not easily explained by a lowering of triglyceride levels. Further, published mapping studies have not detected an association between triglyceride levels and chromosome 9p (2), suggesting that this result may be a false-positive and should be regarded as tentative. In aggregate, the lack of association between variation at this locus and known intermediate phenotypes for CAD suggests that the underlying gene(s) may operate by a novel mechanism.

Figure 4 summarizes the association results in the PROCARDIS populations in the context of known genes in the chromosome 9p region. The high-risk CAD haplotype overlaps with exons 13–19 of ANRIL, a newly annotated gene which encodes a large antisense non-coding RNA (ncRNA) that was identified through a deletion analysis of an extended French family with hereditary melanoma-neural system tumours (15). Our RT-PCR studies show that ANRIL is expressed in atheromatous human vessels (abdominal aortic aneurysm and carotid endarterectomy samples), tissues with comparable cell type profiles to atherosclerotic coronary arteries. ANRIL was expressed in vascular endothelial cells, monocyte-derived macrophages and coronary smooth muscle cells, all of which are involved in atherosclerosis. As is typical of most ncRNAs at present, very little is currently known of the function of ANRIL, a class of gene which is thought to add to a cell's transcriptional control repertoire (21). A survey of the dbSNP database revealed no SNPs that map within exons that collocate with the risk haplotype. However, multiple SNPs coupled to the high-risk haplotype map to intronic or downstream sequences; these variants are plausible candidates for influencing levels of ANRIL expression. However, the downstream targets of ANRIL remain to be discovered as do any interactions with neighbouring genes.

To conclude, the discovery and replication of a convincing susceptibility locus for CAD and T2D on chromosome 9p has opened unanticipated lines of enquiry into molecular pathways underlying these diseases. The apparent independence of variants for each disease is intriguing, as is the coincidence of high-risk haplotype for CAD and an ncRNA sequence of unknown function. Further experiments that will stretch the resolution of currently available techniques for studying the complex interplay of *cis*- and/or *trans*-acting factors will be needed to identify details of the molecular basis of the susceptibility.

## MATERIALS AND METHODS

### Samples

A detailed description of the PROCARDIS population can be found elsewhere (22). In brief CAD cases were collected using physician-confirmed diagnoses of MI, symptomatic acute coronary syndrome, chronic stable angina or intervention for coronary revascularization at or before the age of 65 years. Controls with no personal or sibling history of CAD (at or before age 65 years) were identified through the same

recruitment infrastructure. The age-at-recruitment was similar for cases and controls; 92.8% of the cases and 90.0% of the controls were born between 1930 and 1960. Recruitment of participants was carried out in Germany, Italy, Sweden and the UK with an approximate 1:1 ratio of cases: controls from each country. All participants reported having white European ancestry. The protocol was approved by the Ethics Committees of the participating institutions and all participants gave written, informed consent. DNA was extracted and quantified as previously described (22).

### Multiplex genotyping

SNP genotyping was performed by primer extension and MALDI-TOF mass spectrometry, using Sequenom iPLEX™ technology (Sequenom Inc. San Diego, CA). Sequenom Mass Array Design Software (v.3.0) was used to design 70 SNP assays (in two multiplex reactions, iPLEX1 and iPLEX2) across the chromosome 9p region. PCR was performed using 20 ng of DNA in a reaction volume of 10 µl containing 2U of HotStar Taq (Qiagen, Valencia, CA, USA), 1 mM each dNTP and 200 nM each primer. PCR was performed at 94°C for 15 min followed by 45 cycles of 94°C for 20 s, 56°C for 30 s and 72°C for 1 min and finally 1 cycle of 72°C for 3 min. 0.3 U of shrimp alkaline phosphatase was added and the samples incubated at 37°C for 20 min followed by 85°C for 5 min to inactivate the enzyme. Assays were divided into four groups according to mass and extension primer concentrations adjusted to reduce signal-to-noise ratio. A 5 µl extension reaction containing 0.041 µl DNA polymerase, 0.2 µl termination reaction and either 7, 9, 11 or 14 µM of extension primer was added. Reactions were incubated at 94°C for 30 s followed by 40 cycles of annealing (94°C for 5 s) and extension (5 cycles of 52°C 5 s and 80°C 5 s), followed by 72°C for 3 min. The samples were desalted and dispensed onto a 384 element SpectroCHIP bioarray. Mass ARRAY Workstation version 3.3 software was used to process and analyse the SpectroCHIP bioarrays.

### Genotyping quality control

Genotype cluster plots were manually examined. Four SNPs were eliminated from further analysis due to poor clustering (rs1333046, rs2069422, rs7866410, rs9632884). For the remaining assays, the genotypes of outlying samples were manually called following examination of the mass spectrum profiles. Within each iPLEX experiment, samples which had a genotype call rate of <85% (iPLEX1) or <87.5% (iPLEX2) were discarded. Three additional SNPs were eliminated with Hardy–Weinberg equilibrium statistics in controls >9 (rs518394, rs1333040, rs2811712).

All SNPs, with the exception of rs12379111, had a genotype call rate >90% and a difference in genotype call rate between cases and controls <2%. Concordance with HapMap genotype data was >99.5%.

### Statistical analysis

Hardy–Weinberg equilibrium was tested using the *genhw* Stata™ package (23). Haplotyping and genotype imputation

was performed using the MACH (version 1.0.10) computer program (24). MACH implements a Markov chain algorithm for extended haplotyping of unrelated individuals; the resulting haplotype frequency information can then be used to drive a genotype imputation procedure. Two hundred haplotype states (for updating each individual) were examined to balance the desire to explore the (potentially enormous) space of possible haplotypes as comprehensively as possible with finite computer resources. Five hundred rounds of the Markov sampler were performed; a comparison of results based on 100 and 500 rounds were similar (data not shown), suggesting that an acceptable convergence was attained. MACH reports quality scores to quantify the accuracy of imputed missing genotypes; imputed genotypes with scores  $>0.85$  were included in the analysis.

Association analyses were performed using regression models (logistic and multinomial logistic for categorical outcomes, linear for continuous quantitative traits). A mixture of unrelated (some cases and all controls) and related (affected siblings) individuals were included in the analysis; this familial clustering was modelled by using a robust sandwich estimation of the variance (25) as implemented in Stata™ version 9.2. Country-of-origin was included as a categorical main-effect to indirectly model differences in SNP allele frequencies across the populations, a technique which efficiently absorbs this source of over-dispersion (26). Additive and non-additive effects were modelled by defining continuous variables with levels 0, 1 & 2 or 0, 1 & 0, respectively, corresponding to genotypes AA, AB & BB. In the context of logistic regression, an additive-effects-only model predicting categorical (disease status) outcomes in terms of logit [i.e. log(odds ratio)] effects is equivalent to a multiplicative odds ratio model (i.e. a model of allelic association); inclusion of non-additive effects was used to model non-multiplicative associations (i.e. a model of genotype association). Meta-analysis was performed using fixed-effects (inverse variance weighting) methods implemented in the *metan* Stata™ package (27).

### Biochemical phenotypes

Total plasma cholesterol, HDL cholesterol, triglycerides, albumin, urate and bilirubin were measured at baseline using commercial standard clinical chemistry kits (Roche Diagnostics GmbH, Mannheim, Germany) on a Hitachi 917 Roche automated clinical chemistry analyser in CAD cases. LDL cholesterol was calculated using the Friedewald formula if triglycerides were  $\leq 4.5$  mmol/L (400 mg/dL). Lp(a) was measured by a latex-enhanced immunoturbidimetric test from Immuno (Vienna, Austria). (28). Determinations of cholesterol, HDL-cholesterol and triglycerides were controlled by the NHLBI Lipid Standardization Program organized by the Centers for Disease Control (CDC, Atlanta, GA, USA). Homocysteine in CAD cases was measured by reverse-phase high-performance liquid chromatography with a commercial assay kit (Chromsystems) on an isocratic liquid chromatograph (Kontron, Neufahrn, Germany) interfaced with a fluorescence detector (model RF-535, Shimadzu, Kyoto, Japan). Cotinine levels in CAD cases were measured using a serum cotinine microplate enzyme immunoassay (Cozart Bioscience

Ltd, Abingdon, UK). Plasma fibrinogen concentration in CAD cases was determined with the Clauss assay (29) using the Fibrinogen-C kit and calibration plasma from Instrumentation Laboratory SpA (Milano, Italy), in the ACL 9000 coagulation analyser (Instrumentation Laboratory). Coefficients of variation for a normal control plasma for coagulation analysis from Global Hemostasis Institute (Linköping, Sweden) were 2% (intra-assay) and 7% (inter-assay,  $n = 106$ ).

For those biochemical phenotypes that are known to be modified by statins (LDL-cholesterol, HDL-cholesterol and triglycerides), CAD patients taking this medication were excluded from the lipid association analysis. The distribution of the continuous phenotypes was examined by use of quantile–quantile plots and histograms, a few extreme outliers (which are likely to be due to measurement or data entry errors) were eliminated. Markedly non-normally distributed traits were log, reciprocal-square-root or square-root transformed before analysis (and then asymptotic *P*-values are reported) or variance–covariance estimates were derived empirically by bootstrapping (200 replicates). Covariates (age, sex, BMI, current smoking status and cotinine level (to assess exposure to passive smoking) were selected on the basis of prior expectation and exploratory regression analyses (data not shown) on a trait-by-trait basis.

### RT–PCR analysis

Primers were designed using Primer3 (30) in exon 17 (CAGAGCAATTCCAGTGCAAG) and exon 18 (GATTTGCAAAAACAGCTG) or exon 18 (AGCTTTCTGCTACATGGAGGCTAG) and exon 19 (GGCAAATCACTTTTCATCTTTCTGTAT) to give a product size of 176 or 177 bp by reverse transcription PCR. RNA was obtained by extraction with TRIzol (Invitrogen) from carotid endarterectomy and abdominal aortic aneurysm tissue samples and human cell cultures and lines according to the manufacturer's instructions. In the case of human monocyte-derived macrophages and human skin fibroblasts, RNA was isolated employing the RNeasy Mini kit (Qiagen) according to the supplier's instructions. Contaminating genomic DNA was removed by treatment with DNase I (Qiagen). RT–PCR was (performed on 10–100 ng RNA) using the Quantitect one step RT–PCR kit (Qiagen, Valencia, CA, USA) using the Corbett Life Science Rotor-Gene 3000 system or the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) and the QuantiTect SYBR®Green PCR kit (Qiagen). Products were size fractionated on a 1% agarose gel.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

### ACKNOWLEDGEMENTS

We are grateful to the participants and to the medical and nursing staff who assisted in this project. We are grateful to the core genotyping service at the Wellcome Trust Centre for Human Genetics for technical support and Dr Christina

Bursill and Dr Haris A. Khwaja for gifts of RNA, carotid endarterectomy and aortic aneurysm samples.

*Conflict of Interest statement.* None declared.

## FUNDING

This work was funded by the British Heart Foundation, EC Sixth Framework Programme (LSHM-CT- 2007- 037273) and AstraZeneca AB. R.Cl. and R.C. acknowledge support from the MRC; A.H. obtained support for this project from the Swedish Heart-Lung Foundation, the Swedish Medical Research Council (8691), the Knut and Alice Wallenberg Foundation, the Karolinska Institute and the Stockholm County Council (560183). 'Funding to pay Open Access publication Charges for this article was provided by the EC Sixth Framework Programme (LSHM-CT-2007-037273)'.

## REFERENCES

- Watkins, H. and Farrall, M. (2006) Genetic susceptibility to coronary artery disease: from promise to progress. *Nat. Rev. Genet.*, **7**, 163–173.
- McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. *et al.* (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science*, **316**, 1488–1491.
- Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A. *et al.* (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*, **316**, 1491–1493.
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
- Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E. *et al.* (2007) Genome-wide association analysis of coronary artery disease. *N. Engl. J. Med.*, **2357**, 443–453.
- Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, **316**, 1331–1336.
- Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341–1345.
- Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. *et al.* (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**, 1336–1341.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. *et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, **445**, 881–885.
- Salonen, J.T., Uimari, P., Aalto, J.M., Pirskanen, M., Kaikkonen, J., Todorova, B., Hypponen, J., Korhonen, V.P., Asikainen, J., Devine, C. *et al.* (2007) Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am. J. Hum. Genet.*, **81**, 338–345.
- Pajunen, P., Koukkunen, H., Ketonen, M., Jerkkola, T., Immonen-Raiha, P., Karja-Koskenkari, P., Mahonen, M., Niemela, M., Kuulasmaa, K., Palomaki, P. *et al.* (2005) Five-year risk of developing clinical diabetes after first myocardial infarction; the FINAMI study. *Diabet. Med.*, **22**, 1334–1337.
- Mozaffarian, D., Marfisi, R., Levantesi, G., Silletta, M.G., Tavazzi, L., Tognoni, G., Valagussa, F. and Marchioli, R. (2007) Incidence of new-onset diabetes and impaired fasting glucose in patients with recent myocardial infarction and the effect of clinical and lifestyle risk factors. *Lancet*, **370**, 667–675.
- Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.*, **21**, 263–265.
- Zhang, J., Rowe, W.L., Clark, A.G. and Buetow, K.H. (2003) Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am. J. Hum. Genet.*, **73**, 1073–1081.
- Pasman, E., Laurendeau, I., Heron, D., Vidaud, M., Vidaud, D. and Bieche, I. (2007) Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.*, **67**, 3963–3969.
- Colhoun, H.M., McKeigue, P.M. and Davey, S.G. (2003) Problems of reporting genetic associations with complex outcomes. *Lancet*, **361**, 865–872.
- Assmann, G., Schulte, H. and von Eckardstein, A. (1996) Hypertriglyceridemia and elevated lipoprotein(a) are risk factors for major coronary events in middle-aged men. *Am. J. Cardiol.*, **77**, 1179–1184.
- Sarwar, N., Danesh, J., Eiriksdottir, G., Sigurdsson, G., Wareham, N., Bingham, S., Boekholdt, S.M., Khaw, K.T. and Gudnason, V. (2007) Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation*, **115**, 450–458.
- Bansal, S., Buring, J.E., Rifai, N., Mora, S., Sacks, F.M. and Ridker, P.M. (2007) Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. *JAMA*, **298**, 309–316.
- Nordestgaard, B.G., Benn, M., Schnohr, P. and Tybjaerg-Hansen, A. (2007) Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA*, **298**, 299–308.
- Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15** (Spec no. 1), R17–R29.
- Farrall, M., Green, F.R., Peden, J.F., Olsson, P.G., Clarke, R., Hellenius, M.L., Rust, S., Lagercrantz, J., Franzosi, M.G., Schulte, H. *et al.* (2006) Genome-wide mapping of susceptibility to coronary artery disease identifies a novel replicated locus on chromosome 17. *PLoS Genet.*, **2**, e72.
- Cleves, M.A. (1999) Hardy-Weinberg Equilibrium Tests and Allele Frequency Estimation. *STATA Technical Bulletin*, **48**, 34–37.
- Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.*, **579**, 2290.
- Williams, R.L. (2000) A note on robust variance estimation for cluster-correlated data. *Biometrics*, **56**, 645–646.
- Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. *et al.* (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, **448**, 470–473.
- Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557–560.
- Siekmeier, R., Marz, W., Scharnagl, H., Nauck, M., Mayer, H., Wieland, H., Gross, W. and Seiffert, U.B. (1996) Determination of lipoprotein(a): Comparison of a novel latex enhanced immunoturbidimetric assay and an immunoradiometric assay [In:German]. *Lab. Medizin*, **20**, 294–298.
- Clauss, A. (1957) Rapid new physiological method for measurement of fibrinogen [German]. *Acta. Haematol.*, **17**, 237–246.
- Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. In Misener, S. and Krawetz, S.A. (eds). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.