Comparative analysis of genome-wide-association studies signals for lipids, diabetes, and coronary heart disease: Cardiovascular Biomarker Genetics Collaboration

Aspasia Angelakopoulou††, Tina Shah†††, Reecha Sofat†, Sonia Shah†, Diane J. Berry†, Jackie Cooper†, Jutta Palmen†, Ioanna Tzoulaki‡‡, Andrew Wong§, Barbara J. Jefferis§, Nikolas Maniatis§§, Fotios Drenos¶¶, Bruna Gigante||, Rebecca Hardy¶¶, Ross C. Laxton‡‡, Karin Leander‡‡, Anna Motterle‡‡, Iain A. Simpson¶¶, Liam Smeeth¶¶, Andy Thomson¶¶, Claudio Verzilli†, Diana Kuh§, Helen Ireland§, John Deanfield¶¶¶, Mark Caulfield¶¶¶, Chris Wallace¶¶¶¶, Nilesh Samani¶¶¶¶, Patricia B. Munroe¶¶¶¶, Mark Lathrop¶¶¶¶, F. Gerry R. Fowkes¶¶¶¶, Michael Marmot¶¶¶¶, Peter H. Whincup¶¶¶¶, John C. Whittaker¶¶¶¶, Ulf de Faire††, Mika Kivimaki¶¶¶¶, Meena Kumari¶¶¶¶, Elina Hypponen¶¶¶¶, Chris Power¶¶¶¶, Steve E. Humphries§, Philippa J. Talmud§, Jackie Price¶¶¶¶, Richard W. Morris§, Shu Ye12, Juan P. Casas†¶, and Aroon D. Hingorani12

Aims

To evaluate the associations of emergent genome-wide-association study-derived coronary heart disease (CHD)-associated single nucleotide polymorphisms (SNPs) with established and emerging risk factors, and the association of genome-wide-association study-derived lipid-associated SNPs with other risk factors and CHD events.

Methods and results

Using two case–control studies, three cross-sectional, and seven prospective studies with up to 25 000 individuals and 5794 CHD events we evaluated associations of 34 genome-wide-association study-identified SNPs with CHD.
Using a collaboration of 12 studies, involving ~25 000 individuals, we typed the first SNPs to be identified from GWAS (from 2007 to 2008) that were associated either with myocardial infarction (MI) or with an intermediate phenotype previously associated with MI risk including LDL-C, HDL-cholesterol (HDL-C), triglycerides (TG), body mass index (BMI), or type-2 diabetes mellitus (T2DM).

We studied the associations of these SNPs with a wider range of intermediate phenotypes, to elucidate shared points of regulation, and with CHD risk to assess if the biomarkers altered are likely to mediate or mark changes in the causal pathway to CHD.

First, we hypothesized that comparative analysis of SNP associations with cardiovascular disease (CVD) risk factors/biomarkers and CVD endpoints would help delineate the mechanisms linking genetic variation to CVD events. Identifying which biomarker-associated SNPs also alter disease risk should help clarify which biomarkers lie in the causal pathway. Conversely, understanding which risk factors/biomarkers are altered by disease-associated SNPs should help elucidate the mechanisms underlying the disease association where these are uncertain. Second, we hypothesized that since non-genetic biomarkers are frequently correlated, SNPs identified for an index association with one CVD risk factor/biomarker would frequently be associated with a diverse array of other risk factors/biomarkers. Studying the effect of a SNP on many phenotypes (a phenome scan) would define common points of regulation for diverse risk factors and help disentangle the causal from non-causal interconnections between correlated lipid and inflammation and coagulation markers. By helping to evaluate which SNP associations are exclusive to a single risk factor and which are more extensive, these analyses should also inform on the specificity of individual biomarkers for Mendelian randomization analyses of biomarkers and risk factors.

**Methods**

**Data sets**

**Prospective studies**

Seven UK prospective studies contributed to the collaboration: Northwick Park Heart Study II (NPHS II), British Regional Heart Study
Cross-sectional studies

Two cross-sectional studies contributed prevalent cases of CHD: the Southampton Atherosclerosis Study (SAS) and the Stockholm Heart Epidemiology Program (SHEEP). The SHEEP study included contemporaneous controls, while the representative, population-based cohort study NPHS II, in which participant recruitment overlapped geographically with that for SAS, acted as a control data set for SAS. Details of design, matching criteria, recruitment, main demographic details, definition of outcomes, and other measures are described in Supplementary material online, Methods 1 and Table S1. One cross-sectional study, the MRC-BHF British Genetics of Hypertension (BRIGHT) study included individuals without CHD but with high blood pressure (BP) (Table 1).

Measures

Demographic and other variables

Age, gender, BMI, and systolic and diastolic BP were available from all studies (Table 1).

Blood biomarkers

The availability of blood lipids and apolipoproteins (ApoBs), indices of glycaemic control, as well as inflammation, coagulation, and metabolic markers are listed in Table 1. All measurements were made using validated assays and protocols whose details have been reported previously (Supplementary material online, Methods 1). All markers were assessed in subjects free from CHD at the time of sampling. For prospective studies, these samples were obtained either from the baseline survey or from a subsequent resurvey closest to the assessment at which the study DNA repository was established.

Clinical outcomes

We defined coronary heart disease as a composite endpoint of non-fatal MI, CHD death, or coronary revascularization procedure using prevalent and incident events. We used a similar approach to that adopted by the CARDioGRAM consortium for defining CHD endpoints. We defined incident events in cohort studies as occurring after establishment of the DNA repository and prevalent events as those non-fatal events preceding the establishment of a DNA repository. Subjects from the SAS study were sampled on the condition of having coronary stenosis ≥50% of the diameter in at least one major coronary artery (defined by coronary angiography rather than clinical CHD event).

Genotyping

We typed 21 SNPs associated with LDL-C, HDL-C, TG, or BMI and a total of 16 SNPs previously shown to exhibit an association with CHD or T2DM from GWAS reported in 2007 and 2008. All SNPs that were significant with a P-value < 10^-8 identified from GWAS were included for the analysis unless they were in linkage disequilibrium (LD). Where SNPs were in LD, the best proxy was chosen based on feasibility for genotyping to provide the maximum data available for the analysis. Two SNPs from the chromosome 9p21 region were selected as they had both been identified from GWAS. However, since rs10757274 had not been typed in HapMap at the time this analysis was initiated, and the degree of LD with other SNPs on 9p21 was not known, both SNPs were included. Details are provided in Table 1. New genotyping was conducted using validated, high throughput genotyping platforms at the Genome Centre, Queen Mary University of London, using Kaspar technology or the ABI TaqMan platform; Medical Solutions, Nottingham, using the ABI SNPlex platform; or the Centre for Cardiovascular Genetics UCL, using the ABI TaqMan platform (Table 1). The Whitehall II study provided genotypes from the ITMAT/Broad Institute CARE consortium (IBC) Human CVD Beadchip (Illumina) while 1998BC and BRIGHT provided genotypes from the Affymetrix 500K whole genome array.

Analysis

We regarded a SNP association in CBGC with the same endpoint as that reported in the discovery GWAS as a replication analysis. Single nucleotide polymorphism associations in CBGC with outcomes distinct from the original discovery GWAS e.g. evaluation of associations of CHD-associated SNPs with risk factors and biomarkers were regarded as a discovery analysis.

Single nucleotide polymorphism associations with coronary heart disease

Seven studies including NPHS II, BRHS, ELSA, EAS, WHII, SAS, and SHEEP contributed to the meta-analysis of the association of genotype with CHD (Table 1). The meta-analysis utilized summary data from individual studies using a protocol agreed jointly by a central analysis subgroup in conjunction with principal investigators and statisticians from the participating studies. All analyses were restricted to individuals of Caucasian ethnicity and limited to subjects with complete data for gender, age, and genotype.

For the purposes of quality control and to allow evaluation of any genetic heterogeneity between studies, each study provided details of genotyping platform, call-rate, minor allele frequency (MAF), the exact P-value for a test of departure from Hardy–Weinberg equilibrium (HWE) in subjects without clinical evidence of CHD, and concordance rates for duplicate genotyping. We pre-specified a threshold call rate of 90%, but included SNPs with call rates >80% provided the MAF was concordant with other studies, genotype error rates were <1%, and the P-value for deviation for HWE exceeded 0.001. The pre-specified analysis plan is included as Supplementary material online, Methods 2. Genotypes were coded using a standardized designation for homozygous and heterozygous individuals. Individuals homozygous for the common allele served as the reference group for all the comparisons.

Each contributing study estimated an unadjusted and adjusted OR (and standard error) for CHD, for each additional rare allele carried (i.e. a trend analysis using an additive model on the logarithmic scale) as well as for subjects heterozygous or homozygous for the rare allele compared with those homozygous for the common allele. Variables used in the adjustment were age (in 5 year bands, e.g. 50–55, etc.), gender (male vs. female), and smoking (ever vs. never). A summary odds ratio (95% confidence interval) for the risk of CHD for each SNP was calculated by fixed effects meta-analysis using the Mantel–Hanzel method as well as by random effects meta-analysis using the DerSimonian and Laird method. Where available, we included estimates from the Wellcome Trust Case Control Consortium 1 (WTCCC1) study of CHD in the meta-analysis. A
<table>
<thead>
<tr>
<th>Study design</th>
<th>NPHS-II</th>
<th>BRHS</th>
<th>ELSA</th>
<th>EAS</th>
<th>WH-II</th>
<th>1958BC</th>
<th>MRC 1946</th>
<th>SAS</th>
<th>SHEEP</th>
<th>UDACS</th>
<th>EDS</th>
<th>BRIGHT-cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
<td>Prospective cohort</td>
<td>Prospective birth cohort</td>
<td>Prospective birth cohort</td>
<td>Cases with angiographic coronary disease</td>
<td>CHD patients</td>
<td>Cases of CHD and controls</td>
<td>Cross-sectional</td>
<td>Cross-sectional</td>
<td>Cases from case–control study of hypertension</td>
<td></td>
</tr>
<tr>
<td>Sampling frame</td>
<td>General practices</td>
<td>General practices</td>
<td>Respondents of HSE</td>
<td>General practices</td>
<td>Workplace</td>
<td>Birth register</td>
<td>Birth register</td>
<td>Swedish citizens</td>
<td>Diabetic patients</td>
<td>Diabetic patients</td>
<td>General practices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N with DNA</td>
<td>2775</td>
<td>3947</td>
<td>5274</td>
<td>940</td>
<td>5500</td>
<td>1480</td>
<td>2700</td>
<td>1164</td>
<td>2698</td>
<td>575</td>
<td>331</td>
<td>1759</td>
<td></td>
</tr>
<tr>
<td>Genotyping method</td>
<td>TaqMan</td>
<td>TaqMan</td>
<td>TaqMan</td>
<td>TaqMan</td>
<td>IBC 50k CVD chip</td>
<td>Affy 500k</td>
<td>SNPlex</td>
<td>KASPAR</td>
<td>TaqMan</td>
<td>TaqMan</td>
<td>Affy 500k</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% men</td>
<td>100</td>
<td>100</td>
<td>48</td>
<td>50</td>
<td>77</td>
<td>50</td>
<td>50</td>
<td>76</td>
<td>69</td>
<td>59</td>
<td>60</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Years follow-up</td>
<td>17</td>
<td>26</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>51</td>
<td>63</td>
<td>NA</td>
<td>NA</td>
<td>—</td>
<td>—</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) age of participants</td>
<td>56.1 (3.4)</td>
<td>48.9 (5.50) from 1978 to 1980; 68.8 (5.49) from 1998 to 2000</td>
<td>64</td>
<td>64.3 (5.63)</td>
<td>60.9 (6.0)</td>
<td>45</td>
<td>53 at year of collection now 63</td>
<td>59.6 (7.15)</td>
<td>66.7 (11.0)</td>
<td>63.5 (13.8)</td>
<td>54.6 (10.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases of CHD</td>
<td>273(^{a})</td>
<td>724(^{b})</td>
<td>140(^{a})</td>
<td>117(^{b})</td>
<td>241(^{c})</td>
<td>NA</td>
<td>NA</td>
<td>1164(^{d})</td>
<td>1213(^{e})</td>
<td>NA</td>
<td>NA</td>
<td>3872</td>
<td></td>
</tr>
<tr>
<td>Cases of T2D</td>
<td>229(^{a})</td>
<td>595(^{b})</td>
<td>249(^{a})</td>
<td>336(^{b})</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>600</td>
<td>331</td>
<td>NA</td>
<td>2340</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>2746</td>
<td>3863</td>
<td>5257</td>
<td>893</td>
<td>4789</td>
<td>1436</td>
<td>2455</td>
<td>NA</td>
<td>1052</td>
<td>554</td>
<td>301</td>
<td>1732</td>
<td>25 078</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>2747</td>
<td>3860</td>
<td>5342</td>
<td>892</td>
<td>4803</td>
<td>1431</td>
<td>2439</td>
<td>NA</td>
<td>1055</td>
<td>556</td>
<td>320</td>
<td>1759</td>
<td>25 204</td>
</tr>
<tr>
<td>Diastolic BP</td>
<td>2747</td>
<td>3860</td>
<td>5341</td>
<td>890</td>
<td>4803</td>
<td>1431</td>
<td>2439</td>
<td>NA</td>
<td>1050</td>
<td>556</td>
<td>319</td>
<td>1759</td>
<td>25 195</td>
</tr>
<tr>
<td>Pulse P</td>
<td>2747</td>
<td>3860</td>
<td>5341</td>
<td>890</td>
<td>4803</td>
<td>1431</td>
<td>2439</td>
<td>NA</td>
<td>NA</td>
<td>556</td>
<td>319</td>
<td>1759</td>
<td>24 145</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>2742</td>
<td>3845</td>
<td>5406</td>
<td>892</td>
<td>4799</td>
<td>1416</td>
<td>2321</td>
<td>NA</td>
<td>1496</td>
<td>556</td>
<td>311</td>
<td>1502</td>
<td>25 286</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>1735</td>
<td>3732</td>
<td>5264</td>
<td>887</td>
<td>4741</td>
<td>1338</td>
<td>2145</td>
<td>NA</td>
<td>1482</td>
<td>541</td>
<td>269</td>
<td>1502</td>
<td>23 636</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>1836</td>
<td>3735</td>
<td>5404</td>
<td>887</td>
<td>4799</td>
<td>1413</td>
<td>2155</td>
<td>NA</td>
<td>1487</td>
<td>556</td>
<td>289</td>
<td>1503</td>
<td>24 064</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>2742</td>
<td>2745</td>
<td>5406</td>
<td>892</td>
<td>4799</td>
<td>1415</td>
<td>2319</td>
<td>NA</td>
<td>1496</td>
<td>556</td>
<td>289</td>
<td>1502</td>
<td>24 161</td>
</tr>
<tr>
<td>ApoAI</td>
<td>2344</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4637</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1495</td>
<td>NA</td>
<td>NA</td>
<td>8476</td>
<td></td>
</tr>
<tr>
<td>ApoB</td>
<td>2344</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4637</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1495</td>
<td>NA</td>
<td>NA</td>
<td>8476</td>
<td></td>
</tr>
<tr>
<td>Homocysteine</td>
<td>1361</td>
<td>3776</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1116</td>
<td>NA</td>
<td>NA</td>
<td>6253</td>
<td></td>
</tr>
<tr>
<td>HbA1c</td>
<td>NA</td>
<td>3792</td>
<td>5405</td>
<td>NA</td>
<td>4759</td>
<td>1414</td>
<td>2333</td>
<td>NA</td>
<td>NA</td>
<td>553</td>
<td>303</td>
<td>18 559</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>NA</td>
<td>3843</td>
<td>3224</td>
<td>891</td>
<td>4791</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1363</td>
<td>556</td>
<td>315</td>
<td>14 983</td>
<td></td>
</tr>
<tr>
<td>Insulin</td>
<td>NA</td>
<td>2717</td>
<td>NA</td>
<td>NA</td>
<td>4259</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1124</td>
<td>NA</td>
<td>NA</td>
<td>8100</td>
<td></td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>2733</td>
<td>3834</td>
<td>5436</td>
<td>874</td>
<td>4357</td>
<td>1416</td>
<td>NA</td>
<td>1403</td>
<td>NA</td>
<td>NA</td>
<td>20 053</td>
<td>8507</td>
<td></td>
</tr>
<tr>
<td>IL-6</td>
<td>NA</td>
<td>2258</td>
<td>NA</td>
<td>628</td>
<td>4274</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>801</td>
<td>546</td>
<td>NA</td>
<td>8507</td>
<td></td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>2279</td>
<td>3833</td>
<td>5404</td>
<td>605</td>
<td>4663</td>
<td>1418</td>
<td>NA</td>
<td>NA</td>
<td>1115</td>
<td>545</td>
<td>82</td>
<td>NA</td>
<td>19 944</td>
</tr>
</tbody>
</table>

\(^{a}\)Incident cases only.

\(^{b}\)Both prevalent and incident cases.

\(^{c}\)Prevalent cases only; NA: not available.
false discovery rate (FDR)-adjusted P-value was calculated based on the number of hypotheses tested. Defining FDR as the proportion of falsely rejected hypotheses, i.e. for which the null was actually true, this new P-value, known as the q-value, is the minimum FDR when rejecting a null hypothesis from a list of tested null hypotheses, conditioned on at least one positive finding having occurred.

### Table 2 Categories of single nucleotide polymorphisms typed by the Cardiovascular Biomarkers Genetics Collaboration

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>Initial discovery study</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs599839</td>
<td>SORT1</td>
<td>1</td>
<td>A</td>
<td>G</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs17465637</td>
<td>MIA3</td>
<td>1</td>
<td>C</td>
<td>A</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs17672135</td>
<td>FMN2</td>
<td>1</td>
<td>T</td>
<td>C</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs2943634</td>
<td>intergenic</td>
<td>2</td>
<td>C</td>
<td>A</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs383830</td>
<td>APC</td>
<td>5</td>
<td>T</td>
<td>A</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs6922669</td>
<td>MTHFD1L</td>
<td>6</td>
<td>G</td>
<td>A</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs1333049</td>
<td>CDKN2B (d)</td>
<td>9</td>
<td>G</td>
<td>C</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs10757274</td>
<td>CDKN2B (d)</td>
<td>9</td>
<td>A</td>
<td>G</td>
<td>McPherson (2007)</td>
</tr>
<tr>
<td>rs501120</td>
<td>CXCL12</td>
<td>10</td>
<td>C</td>
<td>T</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs1994016</td>
<td>ADAMTS7</td>
<td>15</td>
<td>C</td>
<td>T</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs17228212</td>
<td>SMAD3</td>
<td>15</td>
<td>T</td>
<td>C</td>
<td>Samani (2007)</td>
</tr>
<tr>
<td>rs8055236</td>
<td>CDH13 (d)</td>
<td>16</td>
<td>T</td>
<td>G</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs7250581</td>
<td>VSTM2B</td>
<td>19</td>
<td>A</td>
<td>G</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs688034</td>
<td>SEZ6L</td>
<td>22</td>
<td>C</td>
<td>T</td>
<td>WTCCC (2007)</td>
</tr>
<tr>
<td>rs10811661</td>
<td>CDKN2B (d)</td>
<td>9</td>
<td>C</td>
<td>T</td>
<td>Saxena (2007)</td>
</tr>
<tr>
<td>rs9939609</td>
<td>FTO</td>
<td>16</td>
<td>T</td>
<td>A</td>
<td>Frayling (2007)</td>
</tr>
<tr>
<td>rs17411031</td>
<td>LPL</td>
<td>8</td>
<td>C</td>
<td>G</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs3890182</td>
<td>ABCA1</td>
<td>9</td>
<td>G</td>
<td>A</td>
<td>Kathiresan (2008)</td>
</tr>
<tr>
<td>rs261332</td>
<td>LPC</td>
<td>15</td>
<td>G</td>
<td>A</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs9989419</td>
<td>CETP</td>
<td>16</td>
<td>G</td>
<td>A</td>
<td>Willer (2008)</td>
</tr>
<tr>
<td>rs16979595</td>
<td>CLPTM1</td>
<td>19</td>
<td>G</td>
<td>A</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs599839</td>
<td>SORT1</td>
<td>1</td>
<td>A</td>
<td>G</td>
<td>Willer (2008)</td>
</tr>
<tr>
<td>rs562338</td>
<td>APOB</td>
<td>2</td>
<td>A</td>
<td>G</td>
<td>Willer (2008)</td>
</tr>
<tr>
<td>rs688</td>
<td>LDLR</td>
<td>19</td>
<td>C</td>
<td>T</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs4420638</td>
<td>APOC1/APOE</td>
<td>19</td>
<td>A</td>
<td>G</td>
<td>Willer (2008)</td>
</tr>
<tr>
<td>rs12042319</td>
<td>ANGPTL3</td>
<td>1</td>
<td>G</td>
<td>A</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs3917820</td>
<td>SELP</td>
<td>1</td>
<td>G</td>
<td>A</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs12140698</td>
<td>FAM5B (d)</td>
<td>1</td>
<td>C</td>
<td>T</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs780094</td>
<td>GCKR</td>
<td>2</td>
<td>C</td>
<td>T</td>
<td>Willer (2008)</td>
</tr>
<tr>
<td>rs1471233</td>
<td>intergenic</td>
<td>4</td>
<td>C</td>
<td>T</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs2074755</td>
<td>BAZ1B</td>
<td>7</td>
<td>T</td>
<td>C</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs17482753</td>
<td>LPL</td>
<td>8</td>
<td>G</td>
<td>T</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs4740635</td>
<td>intergenic</td>
<td>9</td>
<td>G</td>
<td>C</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs7861175</td>
<td>LPAR1 (d)</td>
<td>9</td>
<td>T</td>
<td>C</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs6589566</td>
<td>APOA5</td>
<td>11</td>
<td>A</td>
<td>G</td>
<td>Wallace (2008)</td>
</tr>
<tr>
<td>rs7229921</td>
<td>CIDEA (d)</td>
<td>18</td>
<td>A</td>
<td>G</td>
<td>Wallace (2008)</td>
</tr>
</tbody>
</table>

(d) refers to the nearest downstream gene.

Allele 1 is the reference allele for all comparisons.

Initial discovery studies are Samani (2007); WTCCC (2007); McPherson (2007); Saxena (2007); Frayling (2007); Wallace (2008); Kathiresan (2008); Willer (2008).

### Single nucleotide polymorphism associations with continuous risk factors and biomarkers

Eleven studies including NPHS II, BRHS, ELSA, EAS, WHII, 1958BC, MRC 1946, SHEEP, UDACS, EDS, and BRIGHT contributed to the analysis of SNP associations with continuous risk factors and biomarkers. The distributions of TG, homocysteine, C-reactive
protein, and interleukin-6 (IL-6) were skewed in all studies and these variables were log transformed and analysed on the log-scale. For each SNP, we performed a linear regression analysis within study for each continuous biomarker assuming an additive effect of each variant allele. The per-allele regression coefficient, which is equal to the weighted per-allele mean difference, was then divided by the pooled standard deviation of the trait of interest, derived using information from each of the three genotypes categories; to calculate the standardized mean difference (and its corresponding standard error). Further details are provided in Supplementary material online, Methods 2 with information on the STATA do-files. Separate analyses were conducted unadjusted and adjusted for all of age (in 5 year bands, e.g. 50–55, etc.), gender (male vs. female), and smoking (ever vs. never). We used random and fixed effect meta-analysis to pool the within-study estimates for each trait to generate a summary per-allele standardized mean difference (and 95% confidence interval) to allow comparison of SNP effects across different traits on a common scale. For all traits, the mean of the standardized mean differences is 0 and the standard deviation 1. In the absence of consensus regarding statistical significance level for reporting SNP-associations with multiple traits or disease endpoints (pleiotropy), we reported all results as point estimates and 95% confidence limits (although P-values are available in the Supplementary material online). An FDR-adjusted P-value (q-value) was also calculated based on the number of hypotheses tested. Supplementary material online, Methods 3 provides a detailed discussion of sample size and study power in the context of the range of confirmatory and exploratory analyses that we conducted.

Results

Study populations and measures

The age range of the study populations was 44–67 years. Two studies were of men only (NPHS II and BRHS) while in the remainder, the proportion of men was between 48 and 72% of participants. Among the prospective studies with incident events, the length of follow-up was between 10 and 26 years. The two case–control studies (SHEEP and SAS) (2377 cases), and five of the seven prospective studies (1495 cases) contributed to the analysis of CHD outcomes (Table 1). The control subjects from SHEEP, participants with BP but no CHD from BRIGHT, and the participants from all cross-sectional and cohort studies contributed to the analyses of SNP effects on risk factors and biomarkers (Table 1). Information was available on BMI from 25 078 participants, and systolic and diastolic BP from 25 204 and 25 195 participants, respectively. The blood biomarkers measured spanned lipids and lipoproteins, indices of glycaemic control, inflammation, and coagulation including: total cholesterol (TC; n = 25 286), HDL-C (n = 24 064), LDL-C (n = 23 636), fasting, or non-fasting TGs (n = 24,161), ApoA1 (n = 8476), ApoB (n = 8476), C-reactive protein (n = 19 944), IL-6 (n = 8507), fibrinogen (n = 20 053), fasting glucose (n = 14 983), fasting insulin (n = 8100), glycated haemoglobin (HbA1c; n = 18,559), and homocysteine (n = 6253).

Single nucleotide polymorphisms previously identified for an association with coronary heart disease

There was a strong concordance of MAFs across studies; see Supplementary material online, Tables S1–S3 for call rates, MAF and tests for departure from HWE. Of the 14 SNPs selected through an association with CHD, the following nine SNPs were associated with CHD events in a fixed effects meta-analysis: two at chromosome 9p21 (rs1333049, OR 1.17; 1.11–1.24 and rs10757274, OR 1.17; 1.09–1.26), and one each near SORT1 (rs599839 (OR 1.20; 1.15–1.26), MIA3 (rs17465637 (OR 1.10; 1.04–1.15), Ch2q36 rs2943634 (OR 1.08; 1.03–1.14), APC rs383830 (OR 1.10; 1.02–1.18), MTHFD1L rs6922269 (OR 1.10; 1.03–1.16), CXCL12 rs501120 (OR 1.12; 1.04–1.20), and SMAD3 rs17228212 (OR: 1.11; 1.05–1.17) (Figure 1 and Supplementary material online, Table S4). Supplementary material online, Figure S1 provides estimates from a random effects model. However, aside from rs599839, none of these CHD SNPs was associated with any of the wide range of risk factors and biomarkers analysed, despite available information from 20 000 or more participants for TC, LDL-C, HDL-C, TG, BMI, BP, fibrinogen, and C-reactive protein, over 15 000 for HbA1c, 10 000 for fasting glucose, and just under 10 000 for IL-6 and fasting insulin (Figures 2 and 3).

Single nucleotide polymorphisms previously identified for an association with blood lipids

Seventeen of 20 SNPs selected because of an initial association with a blood lipid component were associated with the same lipid fraction in the CBGC studies (Figure 3 and Supplementary material online, Table S4). The overlap with SNPs identified by the recent Global Lipids Genetics Consortium (GLGC) meta-analysis is shown in Supplementary material online, Table S7. However, in all cases the effect of this category of SNPs was found to extend beyond the initial reported lipid fraction. For example, ANGPTL3 rs12042319 was associated with BMI, TC, LDL-C, and IL-6 in addition to the reported association with TG (Supplementary material online, Figure S2). Another SNP, LPC rs261332 whose initial reported association was with HDL-C, also showed additional associations with ApoAI, TC, fasting insulin, and TG (Supplementary material online, Figure S3).

A summary profile of associations for all the SNPs analysed with all the biomarkers measured is shown in Figure 3. The range and direction of associations was distinctive for each SNP. For example, of the five SNPs associated with HDL-C (CLPTM1 rs16979595, LPL rs17411031, LPC rs261332, ABCA1 rs3890182, and CETP rs9989419), all had a different pattern of association with other risk factors and biomarkers. For some SNPs, the associations extended across a wider range of CHD biomarkers e.g. for GCKR rs780094, associations encompassed lipids and ApoB (TC and TG) as well as the inflammation markers C-reactive protein and fibrinogen (Supplementary material online, Figure S4).
Of the SNPs in this category, only three were associated with CHD (Figure 1): LPL rs17411031 whose initial association was with HDL-C (OR 0.91; 0.84–0.97) (Figure 4), SORT1 rs599839 whose initial association was with LDL-C (OR 1.20; 1.15–1.26), and ANGPTL3 rs12042319 whose initial association was with TG (OR 1.11; 1.03–1.19) (Supplementary material online, Figure S2).

**Single nucleotide polymorphisms previously identified for an association with type-2 diabetes or adiposity**

Neither of the two SNPs typed whose initial association was with T2DM (Chromosome 9p21 rs10811661 and FTO rs9939609) or adiposity (FTO rs9939609) was associated with CHD in the current analysis.
However, we observed associations of the FTO SNP with variables incorporated in one or more definitions of the metabolic syndrome including systolic BP, TG and HDL-C (but not LDL-C), and C-reactive protein, in addition to fasting insulin and glucose, HbA1c, and BMI (Figures 3 and 5).

Discussion

Genome wide association studies have had resounding success in identifying genetic variants contributing to individual differences in the levels of established and emerging risk factors and CHD events. \(^5,6,11,19,35,36\) Genome wide association studies thus far have typically been designed to assess associations of many hundreds of thousands of SNPs usually with a single risk factor or disease endpoint at a time. However, many cardiovascular risk factors and biomarkers are correlated, and scores of alleles are thought to contribute to any common disease or traits, \(^37\) so it has been hypothesized that overlapping genetic associations with multiple phenotypes are likely to be frequent. For example, SNPs in the SH2B3 gene has been associated with celiac disease, rheumatoid arthritis, eosinophil count, high BP, and MI. \(^5,6,38–40\)

In the current analysis, we systematically tested the extent to which SNPs identified through an initial association with a lipid fraction, diabetes, or CHD risk are related to a wider range of intermediate phenotypes. The aim was to identify common points of regulation among correlated risk factors and biomarkers and to evaluate which intermediate phenotypes are likely to lie in or mark changes in the causal pathways involved in CHD.

Single nucleotide polymorphisms initially identified for an association with coronary heart disease events

The well-studied Ch9p21 SNPs (rs1333049 and rs10757274) were significantly associated with the endpoint of CHD events, defined as prevalent MI (in SHEEP), incident/prevalent CHD events (in the prospective observational studies) or angiographic CAD (in SAS). However, neither of these SNPs nor the lead SNPs in SMAD3, MIA3, CXCL12, MTHFD1L, or Ch2q36 were associated with T2DM or any of the CHD biomarkers including BMI, BP, blood lipids, and apoB, the inflammation markers IL-6, fibrinogen, and C-reactive protein, or a number of markers of glycaemic status including fasting glucose and insulin as well as HbA1c. Given that
the combined data set included over 20 000 observations for BMI, BP, TC, HDL-C and LDL-C, TG, C-reactive protein, and fibrinogen and over 10 000 for fasting glucose and HbA1c and the fact that these traits are continuous rather than dichotomous measures, the effect estimates we obtained are likely to be precise and based on adequately powered analysis. The FDR for many of the identified trait associations have a q-value < 10^{-5}, suggesting that the majority of these are likely to be true. However, we could not exclude very minor effects on these or other traits. Prior studies of these CHD-associated SNPs (including the initial GWAS) have reported a lack of association with blood lipids and BP or with clinically defined hypertension or hyperlipidaemia, but the breadth and detail of intermediate phenotypes studied previously has not been as great as in the current study.5,12,13,41,42

Figure 3 (A) Matrix of associations between the 34 single nucleotide polymorphisms analysed, continuous risk factors and biomarkers, and coronary heart disease events. Each cell is colour coded according to the P-value for the relevant association. (B) Matrix of associations between the 34 single nucleotide polymorphisms analysed, continuous risk factors and biomarkers, and coronary heart disease events. Each cell is colour coded according to the beta-coefficient from the pooled regression analysis (i.e. effect size) for the relevant association.
The association of these SNPs with clinical events despite the absence of association with a wide range of established and emerging cardiovascular risk factors suggests their effects are mediated through a previously unsuspected disease mechanism. Additional fine mapping analyses will be required to demarcate the likely causal variants and functional studies to help identify the disease mechanisms.

Lead SNPs in genomic regions including \textit{SEZ6L}, \textit{VSTM2B}, \textit{CDH13}, \textit{ADAMTS7}, and \textit{FMN2} that were associated with CHD in initial GWAS\cite{5,12} were not associated with CHD events in the current data set, broadly consistent with a recent analysis from the Coronary Artery Disease Consortium which included \textasciitilde11,000 cases of CHD or MI\cite{41}.

Single nucleotide polymorphisms initially identified for an association with lipids

With the exception of three SNPs identified previously to be associated with TG (\textit{SELP} rs3917820, Ch9p24 rs4740635, \textit{QDEA} rs7229921), we replicated associations of 17 other SNPs with LDL-C, HDL-C, or TG. Each of these 17 variants had additional effects on other lipids or ApoBs, at least equal in size to the index association, and some also had effects on other phenotypes such as inflammation markers or glycaemic indices.

Although many of these phenotypes are inter-correlated, the profile of associations was distinctive for each SNP (Figure 3), arguing that SNPs associated with several phenotypes have a true biological basis. For example, of the five SNPs associated in the same direction with HDL-C, only two were associated with TG; however, the direction of the effect was different. A similar situation was observed for SNPs originally associated with TG. These marked differences in the patterns of SNP associations contrast with the observed almost invariable association between HDL-C and TG levels\cite{43}. Genetic studies in populations have been likened to natural randomized trials and we have previously reported on concordant effects on blood lipids and lipoproteins of \textit{CETP} SNPs and treatment with a CETP-inhibitor\cite{44}. Genome wide association studies have also reported associations of SNPs in the \textit{HMGCR} gene (which encodes the target for statins) with LDL-C and CHD risk\cite{16,17}. Single nucleotide polymorphisms in the gene \textit{PPARG} that encodes the target for glitazone drugs have also been shown to influence the risk of T2DM\cite{19,45}. This suggests that the SNPs

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\textbf{Trait} & \textbf{Subjects} & \textbf{SMD (95\% CI)} \\
\hline
Lipid Traits & & \\
ApoA1 & 6969 & 0.08 (0.04, 0.12) \\
ApoB & 6969 & -0.02 (-0.10, 0.05) \\
TC & 20696 & -0.01 (-0.05, 0.02) \\
HDL-C & 19649 & 0.12 (0.10, 0.15) \\
LDL-C & 19252 & -0.01 (-0.04, 0.02) \\
Triglycerides & 19586 & -0.14 (-0.18, -0.11) \\
\hline
Adiposity and Glycaemic Control & & \\
BMI & 20835 & 0.01 (-0.01, 0.03) \\
Glu & 13128 & -0.01 (-0.03, 0.02) \\
HbA1c & 15468 & -0.00 (-0.04, 0.03) \\
Insulin & 6931 & -0.00 (-0.08, 0.08) \\
\hline
Blood Pressure & & \\
DBP & 20951 & -0.00 (-0.03, 0.03) \\
PulseP & 20950 & -0.02 (-0.06, 0.01) \\
SBP & 20954 & -0.01 (-0.05, 0.02) \\
\hline
Emerging Biomarkers & & \\
Fibr & 17882 & 0.01 (-0.01, 0.04) \\
CRP & 18004 & 0.00 (-0.03, 0.04) \\
Hcy & 5080 & 0.01 (-0.04, 0.05) \\
IL-6 & 7675 & -0.00 (-0.05, 0.05) \\
\hline
\end{tabular}
\caption{Association of rs17411031 in \textit{LPL} with triglycerides, HDL-cholesterol, and Apolipoprotein Al.}
\end{table}
The ANGPTL3 SNP rs12042319 was associated with increased CHD risk but with lower LDL-C, TG, and IL-6 levels, all of which have themselves been associated with increased risk of CHD.46 – 48 Although ANGPTL3 has been associated with TG and LDL-C levels,35 it has not been previously associated with CHD risk in GWAS analysis or in the recent pooled analysis.49 Therefore, the CHD association we observed should be considered hypothesis generating, and any relevant mechanism is deserving of further investigation. The LPL SNP rs17411031 was associated with a lower risk of CHD and with lower TG and higher HDL-C values. Other SNPs in this gene have previously been shown to affect TG, HDL-C, and CHD risk,14– 17,19,50 and a recent analysis of a common variant in the APOA5 gene that is functionally linked with LPL, is also associated with TG, HDL-C, and CHD risk.51,52,47 These findings suggest further studies of the ApoAV/LPL pathway should be performed to evaluate it as a possible therapeutic target for coronary prevention.

Single nucleotide polymorphisms initially identified for an association with body mass index or diabetes

We studied two SNPs identified with an index association with T2DM, one of which (FTO rs9939609) is thought to act through a primary effect on adiposity and BMI. Neither of these SNPs was associated with CHD in this study, however, the FTO SNP was associated with higher BMI, systolic BP, fasting insulin and glucose, Hba1c, TG, and C-reactive protein, as well as a lower HDL-C. The International Diabetes Federation (http://www.idf.org/metabolic_syrndrome) defines metabolic syndrome as the presence of central obesity (indexed by waist circumference) together with the presence of two of the following: raised BP, raised TGs, raised fasting glucose, and reduced HDL-C. Our findings are also in keeping with those reported previously from a meta-analysis of ~17 000 participants.53 Taken together, the findings suggest that targeting FTO itself or FTO-mediated effects may be effective...
in reducing the risk of metabolic syndrome and, perhaps, its downstream consequences on disease risk as recently suggested.\textsuperscript{54}

**Limitations**

Our analysis was limited to SNPs identified by the first wave of GWAS, and although the list of variants influencing CHD, blood lipids, diabetes, and BMI has since increased substantially, the first SNPs to be identified are likely to represent the largest effect sizes. Our analysis represents one of the first attempts to study the effect of SNPs from GWAS relevant to cardiovascular disease on a wide range of cardiovascular phenotypes as well as CHD. The study is well powered to achieve this, and for example, a sample size of \( \sim 7000 \) subjects would be required to be able to discover SNPs that explain as little as 0.5% of the variance with 80% power, and for the majority of traits we far exceed this number. The approach we have taken could now be extended to incorporate both a wider range of SNPs from subsequent GWAS as well as a wider range of phenotypes, to build a more comprehensive picture of the repertoire of SNPs affecting each of the cardiovascular risk factors and biomarkers as well as the repertoire of traits affected by any given SNP, and to integrate this information with the risk of clinical disease endpoints.

The GLGC analysis was a recent hugely important and successful effort to discover and replicate additional loci for four lipid traits: total-, LDL-, and HDL-C, as well as TGs. For the SNPs that were present in both our study and the GLGC, the directions of effect were concordant as shown in Supplementary material online, Tables S5–S7. However, the GLGC did not have the opportunity to study the wide range of inflammation, coagulation, and glycemic markers that we have been able to report on in the current analysis (see Supplementary material online, Table S7), many of the SNPs we studied were already identified for their association with CHD risk. Nevertheless, our findings should be interpreted within the context of the available power of the study and the potential for false positive association. There may also have been heterogeneity of effect estimates among the different sample collections evidenced by the high \( \hat{\tau}^2 \) values for many of the pooled analyses, and the reasons for this potential heterogeneity are worthy of further investigation, though some could be attributed to the differences in study design. Ongoing collaborative GWAS in CHD should help to clarify the role of these loci and also unveil additional smaller effect loci underlying susceptibility to CHD.

**Conclusions**

The unique properties of genotype, which are distinct from other natural differences between individuals, provide new opportunities for evaluating causal links between associated intermediate phenotypes and between phenotypes and disease. Our findings demonstrate that there are likely to be important unsuspected disease mechanisms and therapeutic targets for CHD; and that there may be points of regulation for diverse cardiovascular risk factors and biomarkers. In future, enhancing the level of detail at both genetic and phenotypic level, incorporating transcriptomics, metabolomics, as well as structural imaging should provide a more comprehensive understanding of the mechanisms linking genome variation with disease. In turn, this should help the development of additional effective therapies for cardiovascular disease prevention.

**Supplementary material**

Supplementary material is available at *European Heart Journal* online.

**Funding**

*NPHS II* was supported by the British Medical Research Council, the US National Institute of Health (grant NHLBI 33014) and Du Pont Pharma, Wilmington, USA. The British Regional Heart Study is a British Heart Foundation Research Group and is supported by British Heart Foundation (RG/04/003). The views expressed in this publication are those of the authors and not necessarily those of the funding bodies. Samples from the English Longitudinal Study of Ageing (ELSIA) DNA Repository (EDNAR), received support under a grant (AG176440651) awarded by the National Institute on Ageing (NIA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here. The Edinburgh Artery Study (EAS) was funded by the British Heart Foundation. The Whitehall II study has been supported by grants from the Medical Research Council; Economic and Social Research Council; British Heart Foundation; Health and Safety Executive; Department of Health; National Heart Lung and Blood Institute (HL36310), US, NIH: National Institute on Aging (AG13196), US, NIH: Agency for Health Care Policy Research (HS06516); and the John D and Catherine T MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health. Analyses of the 1958 birth cohort data were funded by the Medical Research Council (G0601653) and undertaken at GOSH/UCL Institute of Child Health, which received a proportion of funding from the Department of Health’s NIHR Biomedical Research Centres funding scheme. DNA collection was funded by MRC grant G0000934 and genotyping was funded by Wellcome Trust grant 068545/Z/02. The 1946 British birth cohort is funded by the UK Medical Research Council. The Southampton Atherosclerosis Study (SAS) was funded by the British Heart Foundation PG/98183. This work was facilitated by the Barts and The London National Institute for Health Research Cardiovascular Biomedical Research Unit. The Stockholm Heart Epidemiology Program Study (SHEEP) was supported by grants from the Research Council of Sweden (09533), the Swedish Heart and Lung Foundation and Stockholm County Council (ALF). The BRIGHT study is supported by the Medical Research Council (G9521010D) and the British Heart Foundation (PG/02/128). The Barts and The London Charity funded the Barts and The London Genome Centre. This work forms part of the research themes contributing to the translational research portfolio of Barts and the London Cardiovascular Biomedical Research Unit which is supported and funded by the National Institute of Health Research. The BRIGHT study is also extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. The Wellcome Trust Case Control Consortium was funded by the Wellcome Trust (grant number: 076113/B/04/Z). A.D.H. holds a British Heart Foundation Senior Fellowship (F5/05/125). L.S. holds a Wellcome Trust Senior Research Fellowship. R.S. is supported by a British Heart Foundation (Schillingford).
Clinical Training Fellowship (FS/07/01). S.E.H. holds a British Heart Foundation Chair in Cardiovascular Genetics, he, P.J.T., J.C. and J.P. are supported by British Heart Foundation RG08/014. E.H. is funded by the National Heart, Lung, and Blood Institute, NIH, USA (RO1HL036310). M.C. and P.B.M. are funded by the Barts and The London National Institute for Health Research Cardiovascular Biomedical Research Unit and the MRC Programme grant G9521010. Funding to pay the Open Access publication charges for this article will be provided by the British Heart Foundation.

Conflict of interest: A.D.H. is on the editorial board of Drug and Therapeutics Bulletin, a BMJ Group Publication; has provided non-renumerated advice to GlaxoSmithKline, and London Genetics; and has received honoraria for speaking at educational meetings relating to risk factor management and primary prevention. J.W. is 90% employed at GlaxoSmithKline while retaining a 10% appointment at London School of Hygiene and Tropical Medicine.

References

5. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.


