Despite the recent success of genome-wide association studies (GWASs) in identifying loci consistently associated with coronary artery disease (CAD), a large proportion of the genetic components of CAD and its metabolic risk factors, including plasma lipids, type 2 diabetes and body mass index, remain unattributed. Gene-gene and gene-environment interactions might produce a meaningful improvement in quantification of the genetic determinants of CAD. Testing for gene-gene and gene-environment interactions is thus a new frontier for large-scale GWASs of CAD. There are several anecdotal examples of monogenic susceptibility to CAD in which the phenotype was worsened by an adverse environment. In addition, small-scale candidate gene association studies with functional hypotheses have identified gene-environment interactions. For future evaluation of gene-gene and gene-environment interactions to achieve the same success as the single gene associations reported in recent GWASs, it will be important to pre-specify agreed standards of study design and statistical power, environmental exposure measurement, phenomic characterization and analytical strategies. Here we discuss these issues, particularly in relation to the investigation and potential clinical utility of gene-gene and gene-environment interactions in CAD.
Broadly defined, interactions are differences in the strength of association between a gene and phenotype on the basis of the presence of, absence of or quantitative differences in an additional factor, which could be another genetic variant or an environmental exposure. There are several putative models for gene-environment interactions, including synergy, modification of effects and redundancy (Figure 2). For a gene-gene interaction, the additional factor might be dichotomous, such as carrier versus non-carrier status, or additive, such as zero, one or two copies of the minor allele. For a gene-environment interaction, the additional factor can similarly be dichotomous, such as presence or absence of smoking history, or it can be a continuous variable, such as number of pack-years smoked.

Types of interactions
Broadly defined, interactions are differences in the strength of association between a gene and phenotype on the basis of the presence of, absence of or quantitative differences in an additional factor, which could be another genetic variant or an environmental exposure. There are several putative models for gene-environment interactions, including synergy, modification of effects and redundancy (Figure 2). For a gene-gene interaction, the additional factor might be dichotomous, such as carrier versus non-carrier status, or additive, such as zero, one or two copies of the minor allele. For a gene-environment interaction, the additional factor can similarly be dichotomous, such as presence or absence of smoking history, or it can be a continuous variable, such as number of pack-years smoked.

Role of interactions in genetic association studies
Recent advances in cost-effective, array-based, high-throughput genotyping platforms have led to a flood of investigations of common single nucleotide polymorphisms (SNPs) in various diseases. Genome-wide association studies (GWASs) have successfully identified genetic determinants of CAD and its component risk factors [3-16]. For instance, several investigations found a region of chromosome 9p21 that was associated with CAD independently of traditional risk factors [3-6]. Furthermore, multiple genetic associations for T2DM [7,17] and body mass index (BMI) [18] have been discovered. However, most associated loci from GWASs have been reported for lipoprotein traits, including over 30 loci associated with plasma concentrations of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol and triglyceride [7-16]. The success in finding genetic associations with lipoprotein phenotypes was due to methodological standardization (accuracy and precision) in trait measurement and to evaluation of large sample sizes, allowing detection of relatively subtle effects. Meta-analyses and collaborative consortia with large sample sizes have allowed GWASs to detect risk variants with low minor allele frequencies (<5%) and small effect sizes (odds ratio of about 1.1 to 1.7) (Box 1); SNP association studies may have already reached their limit to detect clinically or
biologically relevant loci with such effect sizes [8,11,13,17,18].

Despite recent success in identifying CAD-associated SNPs, much of the genetic component of CAD and its risk factors remains unattributed. Forcing additional genetic markers with small effect sizes into predictive models only marginally improves prediction over traditional risk factors [19]. However, accounting for gene-gene and gene-environment interactions might produce a meaningful increase in the combined effect of the genetic determinants [1,2]. To ensure a valid assessment of gene-gene and gene-environment interactions, standards are required for sample sizes, accuracy and precision for continuous data, specificity and sensitivity for discrete data and appropriate statistical methods. Phenomics, defined as the comprehensive characterization of phenotype and environmental exposure [20], is also of key importance.

Identification of small effect genetic and environmental factors

So far, most genetic association studies have evaluated effects on intermediate phenotypes or pathogenic mechanisms, which can themselves be considered disease processes. For CAD, these intermediate phenotypes include blood coagulability, hypertension, altered lipid metabolism, cell proliferation and inflammation. When a new gene or locus is discovered, such as the chromosome 9p21 region associated with early CAD [3-6], and its association is subsequently replicated in multiple study samples [21-24], the basis of the association with CAD is assumed to be mediated through a pathogenic pathway [22]. This assumption will guide the design of subsequent functional experiments. Similarly, newly identified environmental determinants might exert their influence through one or even several pathogenic mechanisms and might even help identify previously unappreciated pathways.

Although the effect sizes of SNP associations identified in GWASs of CAD are modest, they are still important because: (i) individual associations can be combined to obtain larger cumulative effects; (ii) genes with small effects in GWASs can point to targets for drug-based or other interventions; (iii) genes with small effects in GWASs might contain rare, large-effect mutations in more severely affected patients; (iv) some GWAS loci with no previous CAD association might unveil new pathways; and (v) the effects of a GWAS locus could be amplified by gene-gene or gene-environment interactions.

These principles can be extended to the study of gene-environment interactions. For instance: (i) individual environmental interactions could be combined to obtain a cumulatively larger effect; (ii) rare extreme environmental exposures may display larger effects on the CAD phenotype than more common or typical environmental variation; (iii) identification of gene-environment interactions might suggest new hypotheses to evaluate disease-causing mechanisms. These principles could direct the design of future studies of gene-environment interactions in CAD.

Minor versus major alleles as a risk factor for CAD

How do alleles affecting CAD susceptibility arise? DNA mutagenesis could provide a basis for understanding the generation of risk alleles. Several mutagenic mechanisms have been identified [25]. If a DNA error escapes repair and becomes embedded in the genome, it could, by affecting the expression or function of a protein, modify CAD risk either positively or negatively. If the recently mutated allele increases CAD risk, it is possible that genetic drift, inbreeding, pleiotropy, heterozygote advantage or small effects on reproductive fitness could be responsible for the allele reaching appreciable frequencies in the population [26]. For CAD, mortality typically occurs after the reproductive years, thus reducing selection pressure against deleterious alleles. Another possibility is that an environmental change might cause an allele that once had a neutral or beneficial effect to become deleterious.

Alternatively, if the mutated allele is beneficial, reducing CAD risk, one would expect the allele to increase in frequency to become the major allele. If the mutation occurred relatively recently, it is possible the minor allele is gradually becoming more prevalent. Such ‘protective’ minor alleles, or conversely major alleles that increase CAD risk, are possibly important from a public health perspective, since defining a gene-environment interaction might suggest an environmental intervention with a potentially large impact, due to the high population prevalence of the risk allele.

Analytical detection strategies

Gene-gene and gene-environment investigations have included family-based and population-based samples in retrospective and prospective designs. Statistical methods have included methods modifying regression and chi-squared analyses, as well as statistical classification techniques, such as neural networks, support vector machines or Bayesian networks (Table 1). Although the statistical methods used in GWASs are fairly consistent and include regression and chi-squared analysis [3-8,10-18,27-30], the statistical approaches to detect gene-gene and gene-environment interactions are somewhat less standardized at present.

Investigators have tested for association between the cumulative number of risk alleles at multiple independent loci and disease [11,27,28]. Absolute allele counts [28] and relative weighting of alleles on the basis of their effect size [11,27] have both been reported. Although this showed that the alleles were independent and their effects could be
added, no interaction between the alleles was measured. Subgroup analyses, in which the strength or effect size of the association is compared between sample subgroups, have substantially less power to detect an association than the original intact sample, increasing the risk of false negative results. For example, assuming 80% power to detect a difference in allele frequencies between cases and controls within one subgroup, the second equally sized subgroup will yield disparate results about 30% of the time just by chance. The clinical trial literature contains many examples of inappropriate subgroup analyses [31], and one excellent review examines the lack of consistency of sex-specific subgroup genetic associations [32].

Regression techniques can be modified to test for gene-gene or gene-environment interactions, either by including additional interaction terms in the model or testing association with or without an additional covariate. Careful reviews of regression approaches to study interactions show the multitude and complexity of these techniques [33,34]. Finally, sophisticated statistical classification techniques, including but not limited to neural networks [29], support vector machines [35] and Bayesian networks [30], are being updated to accommodate analysis of interactions.

Multiple comparisons
If N genetic variants are entered into an analysis, N(N-1)/2 potentially interacting pairs can be constructed. Selecting a priori known functional SNPs, or SNPs with coinciding spatial or temporal expression patterns, is one approach to reduce the number of tests. An alternative approach is first to test for marginal main effects in a primary hypothesis-generating analysis and then to test for interactions between those significant effects in a second analysis in which the nominal level of significance has not been substantially adjusted [33]. In GWASs, permutation testing, control of false discovery rates and Bonferroni correction have been used to determine appropriate significance thresholds. Whatever approach is used, care will be required for selecting the nominal level of significance in gene-gene and gene-environment investigations.

Potential biases in gene-gene and gene-environment investigations of CAD
Many types of biases can affect gene-gene and gene-environment interactions (Table 2). The accuracy and precision of genotyping technologies render genetic investigations relatively resistant to measurement bias, compared to other sources of potential bias. Unequivocal disease phenotypes, such as myocardial infarction or coronary bypass surgery, are least susceptible to measurement bias. New imaging techniques, such as ultrasound-based intima-media thickness or magnetic resonance imaging (MRI)-based plaque volume calculations, are more susceptible to systematic errors of measurement. Self-reported measures of environmental exposure, such as caloric intake, energy expenditure or alcohol use, are most vulnerable to biases. Strategies to maximize the sensitivity and specificity of environmental factor measurement will improve the likelihood of detecting a significant association signal for interactions with genetic determinants [36].

Study design can affect bias, because prospective cohort studies are generally more resistant to bias than retrospective case-control designs [1]. Survivorship bias and population stratification are less common in prospective studies, assuming a truly representative cross-sectional cohort. Survivorship bias is a potential liability of retrospective studies of CAD, because patients with a fatal first myocardial infarction (up to 30% of cases) cannot be included in future studies. Recall bias, in which the study participant is more likely to remember an environmental exposure if it is

---

Table 1

| Analytical strategy                                                                 | Advantages                                           | Disadvantages                                               | References |
|------------------------------------------------------------------------------------|------------------------------------------------------|-------------------------------------------------------------|------------|
| Examine the effect of the cumulative number of risk alleles at multiple loci       | Simple; shows independence of loci                   | No interaction measured                                     | [11,27,28] |
| Compare effect of risk allele in sample subgrouped by environmental exposure or    | Simple                                               | Substantial loss of power in subgroups                      | [13,31,32] |
| additional genotype                                                                |                                                      |                                                             |            |
| Identify risk allele whose association with phenotype is modulated by inclusion of | Easy to implement                                    | Multiple comparisons                                         | [11,33,34] |
| environmental or genetic covariate                                                 |                                                      |                                                             |            |
| Inclusion of interaction term in regression model                                 | Direct modeling of gene-gene or gene-environment    | Need to define multiple terms in model; possibility of over-fitting; multiple comparisons | [13,33,34,55,56] |
| Non-linear statistical classification techniques, including Bayesian networks,     | Large volumes of data in model-free manner           | Difficult to interpret; require validation datasets         | [29,30,34,35,57,58] |
| neural networks, neural networks and support vector machines                       |                                                      |                                                             |            |
associated with a negative outcome, respondent bias, in which patients alter their answers to exposure questions following a negative outcome, and exposure suspicion biases, in which investigators query individuals who have a negative outcome more thoroughly, are all reduced in prospective designs, as long as environmental exposure information is collected from all study participants irrespective of CAD outcomes.

### Statistical power

Statistical power is directly proportional to the number of study participants and to the size of the effect under study. Factors to be included in power calculations of all genetic investigations include the minor allele frequency, the degree of linkage disequilibrium between the queried marker and the hypothetical disease locus, the genotype error rate and the genetic or phenotypic heterogeneity (Box 2). Fortunately, high-throughput genotyping platforms have a negligible genotype error rate [37]. Correction for multiple comparisons and the measurement error of environmental exposures also influence study power [1,2]. As a result of the greater accuracy of genotyping compared with the measurement or report of environmental exposures, there is theoretically more power to detect a gene-gene interaction than a gene-environment interaction for the same sized sample. Studies with inaccurate or imprecise measurement of phenotype or environmental exposure may require up to 20 times larger samples to detect an association signal above background noise [36]. However, the power advantage of gene-gene investigations resulting from their higher measurement accuracy is diminished by the need to correct for multiple comparisons and by the potentially increased complexity of interactions compared with gene-environment investigations. How large a sample is required for adequate power to find gene-gene and gene-environment interactions? A rule of thumb is that a four-fold increment in sample size is required to test for a multiplicative interaction of two main effects [2,38]. This may overestimate the sample size requirement, especially if the effect of the interaction is larger than the main effects, but it illustrates the general requirement for a larger sample size when interactions are introduced into hypothesis testing. Given that many previous candidate gene studies, and even many GWASs, were powered to detect only main effects, testing these samples for gene-gene and gene-environment interactions has the potential for false positive and false negative results [2,3]. Higher-order interactions will require even larger samples to attain suitable power and may not be possible even among the largest current association studies [1].

### Examples of interactions in monogenic CAD

Studies of monogenic susceptibility to CAD have revealed several gene-gene and gene-environment interactions. For

| Table 2 |
| --- |
| **Potential biases in gene-gene and gene-environment investigations of coronary artery disease (CAD)** |
| Bias | General description | Application to CAD |
| --- | --- | --- |
| Selection bias | Skewed in the selection of study participants | Patients with strong family history may self-select for study participation; patients with strong family history may be more likely to be referred to tertiary care and research centers |
| Survivor bias | Selection of study participants may miss mild disease or severe fatal cases | Patients whose first myocardial infarction is fatal are less likely to be studied |
| Recall bias | Patients are more likely to recall an environmental exposure if it was linked to a negative outcome | Patients with CAD may be more likely to remember an environmental exposure because of its negative consequences |
| Respondent bias | Patients answer in the way they believe they should answer, not the true answer | Patients with CAD and knowledge of potential CAD risk factors will be more motivated to report those exposures |
| Family information bias | Individuals become more aware of exposure if it is prevalent in their family | Many CAD risk factors and environmental exposures cluster in families |
| Exposure suspicion bias | Disease status can affect the amount of environmental exposure history collected | If data collection is not standardized, investigators may more thoroughly query patients with CAD |
| Publication bias | Statistically significant findings are more likely to be published | Gene-gene and gene-environment interaction findings in CAD are more likely to be published if significant |
| Measurement bias | Systematic errors of measurement | Platform- or laboratory-dependent genotyping errors; errors of laboratory values; errors of environmental exposure measurement |
| Population stratification | Differences in allele frequencies between groups resulting from ancestry not outcome status | CAD prevalence varies between ethnicities; but this can be tested and corrected for using methodological and statistical techniques |
instance, age at death from CAD was studied in large Mormon families with familial hypercholesterolemia (FH) attributable to rare heterozygous mutations in the \textit{LDLR} gene [39,40]. Carriers of \textit{LDLR} mutations who lived in the 19th century had survived to the eighth and ninth decades of life, whereas carriers of \textit{LDLR} mutations who lived in the 20th century died early with CAD, often in the third and fourth decades of life [39,40]. The most likely explanation for this observation was a healthier environment in past times, including higher physical activity and lower saturated fat consumption compared with the contemporary environment [39,40]. Similar conclusions were reached with multigenerational studies of FH patients in the Netherlands [41]. Other investigators found that Chinese FH heterozygotes who had immigrated to North America had worsened biochemical and clinical phenotypes than carriers of the same \textit{LDLR} mutations living in China [42]. The difference in disease severity was ascribed to differences in dietary fat consumption; these circumstantial observations strongly suggested that environmental factors, such as diet and activity level, modulated the phenotype of heterozygous FH.

From our personal experience, there are other examples of monogenic illnesses whose severity can be significantly modulated by the environment - mainly diet and activity. For instance, we have seen the severity of expression of the disease phenotype made worse by an adverse environment in patients with hypertriglyceridemia due to apo CII-T [43], with analphalipoproteinemia due to \textit{APOA1} Q[-2]X [44], with T2DM due to \textit{HNF1A} G319S [45] and with metabolic complications and CAD in familial partial lipodystrophy due to \textit{LMNA} R482Q [46].

\textbf{Examples of interactions with common SNPs}

Although interactions between environment and disease penetrance in rare monogenic disorders are instructive, a much larger potential impact could be seen in common
complex CAD susceptibility because of small-effect common SNPs. The effect of the environment might be even more pronounced in patients whose phenotypes are caused by the aggregation of small contributions from many genetic and non-genetic factors. Examples of replicated gene-gene and gene-environment interactions identified in investigations of common SNPs in candidate genes are shown in Table 3. For instance, increased CAD risk has been observed in smokers with null genotypes for glutathione S-transferases, which are involved in the detoxification of carcinogens and products of oxidative stress [47,48]. Furthermore, smokers who are carriers of at least one APOE ε4 allele seem to have significantly higher concentrations of oxidized LDL cholesterol compared with non-carriers, potentially further increasing CAD risk [49,50]. Humphries and colleagues report a robust association between the -455G>A SNP of the fibrinogen beta chain (FGB) gene and elevated post-exercise fibrinogen levels [51,52]. Elevated fibrinogen levels may modulate the myocardial infarction risk associated with the Leu34 allele of blood coagulation factor XIII (F13A1), a tetrameric zymogen that protects the fibrin clot from proteolytic degradation [53,54]. These candidate gene-environment interactions were examined because of plausible biological relationships, but large-scale replications are still required, with careful attention to the issues raised in this article.

Examples from GWASs
Gene-gene or gene-environment interactions are not yet routinely evaluated in GWASs, but two recent reports include exploratory examinations. Kathiresan and colleagues performed a two-stage GWAS of plasma lipoproteins [11]. The first stage identified over 1,000 associated SNPs in 25 loci (p < 5 × 10−8) [11]. The second stage analysis re-tested all SNPs using 36 of the significantly associated SNPs from the first stage as covariates in the regression. The number of associated SNPs was reduced to 105 in 7 loci (p < 5 × 10−8) [11]. All loci identified in the second stage had been identified in the first stage of analysis, suggesting that additional SNPs in known loci - that are not in linkage disequilibrium with the SNPs used as covariates - are associated with lipoprotein traits.

Sabatti and colleagues examined genome-wide gene-environment interactions, with the caveat that the work was under-powered to confidently identify interactions [13]. They examined four dichotomized environmental variables (sex, use of oral contraceptives, BMI over 25 kg/m2 and gestational age), comparing differences in effect size between the two subgroups and two variables separated into quintiles (birth BMI and early growth), which were tested by regression using an interaction term [13]. At least one interaction SNP was identified (p < 5 × 10−7) for five out of six environment variables, although none of the SNPs were in genes with a main effect or with known biological relevance [13].

These findings represent possible novel associations with metabolic CAD risk factors, but replication in larger samples is required. The issues discussed above in relation to study design, power and analytic strategies to detect gene-gene

---

### Table 3

Examples of replicated gene-gene and gene-environment interactions in CAD

| Gene   | Environment | Interaction                                                                 | Independently associated with CAD? | References |
|--------|-------------|-----------------------------------------------------------------------------|-------------------------------------|------------|
| LDLR   | Lifestyle   | Rare mutations have larger effect in less active people with high-fat diet    | LDLR: yes; lifestyle: yes           | [40,42]    |
| GSTM1, GSTT1 | Smoking | Elevated CAD risk in smokers with null mutations | GSTM1, GSTT1: weak; smoking: yes | [47,48] |
| APOE   | Smoking     | Exaggerated smoking-associated CAD risk in carriers of APOE ε4              | APOE: yes; smoking: yes             | [49,50]    |
| ADHIC  | Alcohol consumption | Slow-metabolizing γ2 allele homozygotes have the greatest CAD protection | ADHIC: weak; alcohol: yes           | [57,58]    |
| FGB    | Strenuous exercise | Carriers of 455A allele have exaggerated increase in fibrinogen after exercise | FGB: no; exercise: yes              | [51,52]    |
| F13A1  | Plasma fibrinogen | Leu34 is protective for CAD in people with high fibrinogen levels | F13A1: no; fibrinogen: yes         | [53,54]    |
| ACE, AGT | | Unclear multi-locus epistatic interactions | ACE: no; AGT: no | [59,60] |
| LPL, APOE | | Greater negative effect of rare LPL alleles in APOE ε4 carriers | LPL: yes; APOE: yes | [55,56] |

Abbreviations: ACE, acetylcholine esterase; ADH1C, alcohol dehydrogenase 1C; AGT, angiotensinogen; APOE, apolipoprotein E; FGB, fibrinogen beta chain; F13A1, coagulation factor XIII, subunit A1; GSTM1, glutathione S-transferase mu 1; GSTT1, glutathione S-transferase theta 1; LDLR, low-density lipoprotein receptor; LPL, lipoprotein lipase.
Box 2. Factors affecting the statistical power of a study of gene-gene or gene-environment interactions

- Number of study participants
  - ratio of cases to controls
- Effect size of risk factor
  - proportional to standard deviation of trait
- Measurement error of disease trait or environmental exposure
  - accuracy and precision of quantitative trait measurement
  - sensitivity and specificity for dichotomous measurements
- Genotyping error rate
  - especially pertinent when examining copy number variation
- Minor allele frequency of genetic variant
  - affected by disease model if grouped into carrier versus non-carrier
- Linkage disequilibrium between queried variant and ‘true’ functional variant
- Significance threshold
  - needed to correct for multiple tests, to an even greater degree in multi-way tests for gene-gene or gene-environment interactions
- Genetic heterogeneity
  - different variants in the same locus, or in multiple loci, lead to the same phenotype
- Phenotypic heterogeneity
  - different phenotypes are produced as a result of the same genetic variation
- Type of interaction and size of interaction effect

and gene-environment interactions are relevant to these large multi-center population studies, as these studies will form the precedent for future investigations.

Clinical Implications
Accounting for gene-gene and gene-environment interactions will probably be important for future strategies of diagnosis, prognosis and management of CAD. For instance, current treatment guidelines for CAD prevention require risk stratification of the patient. CAD risk strata in a currently disease-free patient are calculated using traditional epidemiological risk factors, such as older age, male sex, the presence of cigarette smoking, diabetes, hypertension, dyslipidemia and, in some models, a family history of early CAD. Quantification of the patient’s CAD risk using these variables guides the intensity of evidence-based drug treatment of modifiable risk factors, such as hypertension and dyslipidemia. It certainly seems feasible that reliable molecular genetic information can be included in future risk stratification models, improving precision over simply documenting a family history of CAD. Furthermore, combinations of specific genetic variables in the context of specific environmental variables - reflecting both gene-gene and gene-environment interactions - could help to re-stratify an individual between risk strata derived using non-molecular data. Also, given that such environmental factors as diet, activity level, stress, smoking and air quality are known to be important determinants of CAD risk, the first line of cost-effective and safe intervention for an individual with a high genetic risk burden would include modulation of such environmental factors instead of more costly, high-tech approaches, such as gene-based biological therapies.

Conclusions
In the context of GWAS datasets, gene-gene and gene-environment interactions are a new frontier for CAD association studies. GWASs have been extremely successful in identifying individual loci for CAD susceptibility, but the practical limits of sample size and array resolution for the identification of biologically valid loci will soon be reached.
As a result of the high prevalence of CAD and the presence of large, multi-center prospective cohort initiatives with genotyping on high-density DNA genotyping arrays, gene-gene and gene-environment interaction studies of CAD will be possible in the future. Rigorous testing for gene-gene and gene-environment interactions should be built into the experimental study design. To ensure that testing for interactions enjoys the same success as GWASs of CAD, precise standards, including suitable sample sizes, reliable methods for measurement of environmental exposures, phenomic characterization and statistical analyses, will be required to minimize both false negative and false positive findings and to allow findings to be compared across samples and reports. The increment in the understanding of CAD susceptibility provided through systematic study and replication of gene-gene and gene-environment interactions will permit a more complete set of tools for diagnosis, disease prediction and prognosis and tailored therapy, perhaps using appropriate environment-based interventions.

**Abbreviations**

BMI, body mass index; CAD, coronary artery disease; FGB, fibrinogen beta chain; FH, familial hypercholesterolemia; F13A1, blood coagulation factor XIII subunit A1; GWAS, genome-wide association study; HMGCR, 3-hydroxyl-3-methylglutaryl-CoA reductase; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor; SNP, single nucleotide polymorphism; T2DM, type 2 diabetes.

**Competing interests**
The authors report no competing interests.

**Author contributions**
Both authors contributed to the conception and production of the manuscript and approved the final version.

**Acknowledgements**

MIBL is supported by the Canadian Institute of Health Research MD/PhD Studentship Award and the University of Western Ontario MD/PhD Program. RAH holds the Edith Schulich Venet Canada Research Chair (Tier I) in Human Genetics and the Jacob J Wolfe Distinguished Medical Research Chair at the University of Western Ontario. This work was supported by operating grants from the Heart and Stroke Foundation of Ontario (NA 6018), the Canadian Institutes for Health Research (MOP supported by operating grants from the Heart and Stroke Foundation of Canada). RAH holds the Edith Schulich Venet Canada Research Chair and the Jean Davignon Distinguished Cardiovascular Research Chair at the University of Western Ontario, the Edith Schulich Venet Canada Research Chair and the University of Western Ontario MD/PhD Studentship Award and the University of Western Ontario MD/PhD Program. RAH holds the Jean Davignon Distinguished Cardiovascular Research Chair and the Canadian Institutes for Health Research (MOP supported by operating grants from the Heart and Stroke Foundation of Canada). RAH holds the Edith Schulich Venet Canada Research Chair and the University of Western Ontario MD/PhD Studentship Award and the University of Western Ontario MD/PhD Program.

**References**

1. Manolio TA, Bailey-Wilson JE, Collins FS: Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006, 7:812-820.
2. Hunter DJ: Gene-environment interactions in human diseases. Nat Rev Genet 2005, 6:287-298.
3. Welcome Trust Care Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007, 447:661-678.
4. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir S, Jonsson L, Sigurdsson A, Baker A, Palsson Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Mathiasdottir S, Jonsson T, Palsson S, Einarsson H, Gunnarsdottir S, Glyfason A, Valfannar C, Hooper WC, Reilly MP, Granger CB, Austin H, Shah SH, Quyyumi AA, et al.: A common variant on chromosome 9q21 affects the risk of myocardial infarction. Science 2007, 316:1491-1493.
5. McPherson R, Pertsinidis A, Kavalas N, Steward A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hoobs HH, Cohen JC: A common allele on chromosome 9 associated with coronary heart disease. Science 2007, 316:1488-1491.
6. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JC, Konig IR, Stevens SE, Szymbczak S, Tregouet DA, Isles MM, Pike F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom T, Bransine I, Gieger C, Deloukas P, et al.: Genomewide association analysis of coronary artery disease. N Engl J Med 2007, 357:443-453.
7. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PJ, Chen H, Cox JJ, Kathiresan S, Jhun BS, Jonsson L, Hengstenberg C, Maller J, Mangino M, Mayer B, Almgren P, Florez JC, Meyer A, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Ophus-Melander M, Rastam L, Spielhofer EK, Taskinen M, Tuomilehto J, et al.: Genomewide association study loci for type 2 diabetes and triglyceride levels. Science 2007, 316:1331-1336.
8. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Piraestor PP, Penninx BW, Jansen AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kiyvil KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marnoni F, Hayward C, Vitart V, Jonsson I, Pastor C, Wright A, Halsey N, Pichler I, Hicks AA, et al.: loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 2009, 41:47-55.
9. Chasman DI, Pare G, Zee RY, Parker AN, Cook NR, Buring JE, Kwiatkowski DJ, Rose LM, Smith JD, Williams PT, Rieder MJ, Rotter Jl, Nickerson DA, Krauss RM, Milesip JG, Ridker PM: Genetic loci associated with plasma concentrations of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and apolipoprotein B among 6382 white women in genomewide analysis with replication. Circ Cardiovasc Genom Cardiometab Dis 2009, 2:121-31.
10. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlandr B, Hedner T, Corella D, Tai ES, Ordovas J, Berglund G, Vartiainen E, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marnoni F, Hayward C, Vitart V, Jonsson I, Pastor C, Wright A, Halsey N, Pichler I, Hicks AA, et al.: loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 2009, 41:47-55.
11. Kathiresan S, Willer CJ, Peluso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanla T, Voight BF, Bynonnae CY, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubaiana KA, Morken MA, Scott J, Rimpling HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, et al.: Common variants at 3 loci contribute to polygenic dyslipidemia. Nat Genet 2009, 41:56-65.
12. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, Gomez Perez FJ, Frazer KA, Elliott P, Scott J, Milos PM, Cox DR, Thompson JR: Genomewide scan identifies variation in MXIPIXI associated with plasma triglycerides. Nat Genet 2008, 40:149-151.
13. Sabatti C, Service SK, Hartikainen AL, Pouza A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varlio T, Kaakinen M, Sovio U, Ruokone A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliott P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB, Peltonen L: Genome-wide association study of metabolic traits in a Finnish cohort from a founder population. Nat Genet 2009, 41:435-46.
14. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, Inouye M, Luben R, Sims M, Hyde D, Macarlie W, Barter P, Kanesami YA, Mahley RW, McPherson R, Grundy SM, Bingham SA, Khaw KT, Loos RJ,
20. Joy T, Hegele RA: Genomics of metabolic syndrome: is there a role for phenomics? Curr Atheroscler Rep 2010, 12:201-208.

21. Anderson JL, Horrie BD, Kolek MJ, Mulfstein JB, Mower CP, Park J, May HT, Camp NJ, Carlquist JF: Genetic variation at the 9p21 locus predicts angiographic coronary artery disease prevalence but not extent and has clinical utility. Am Heart J 2008, 156:1155-1162.e2.

22. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, Brown M, Connell JM, Dominiczak A, Lathrop GM, Webster J, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Con- stin LS, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Heath SC, Timpson NJ, Najjar SS, Stringham HM,. Nat Genet 2008, 40:1155-1162.e2.

23. Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Kostulas K, Gottsater A, Flex A, Stefansson H, Hansen T, Andersen G, Weinsheimer S, Borch-Johnsen K, Jorgensen T, Shah SH, Quyyumi AA, Granger CB, Reilly MP, Austin H, Levey AI, Vaccarino V, et al.: Phenotypic variation in heterozygous familial hypercholesterolemia: a comparison between family and non-family hypercholesterolemia. J Clin Invest 1994, 93:223-229.

24. Triggs-Raine BL, Kirkpatrick RD, Kelly SL, Norquay LD, Cattini PA, Yamagata H, Kanlay A, Zimm B, Harris BS, Barnett PH, Hegele RA: HNF-1alpha G319S, a transcription-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. Proc Natl Acad Sci USA 2002, 99:164-169.

25. Meisinger C, Ouwehand W, Deloukas P, et al.: Replicated replication and a prospective meta-analysis of the association between chromosome 17q21.3 and coronary artery disease. Circulation 2008, 117:675-684.

26. Fleck O, Nielsen O: DNA repair, J Cell Sci 2004, 117:515-517.

27. Crespi B: The evolution of maladaptation. Heredity 2000, 84:623-629.

28. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Pukkonen-olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Con- stin LS, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Heath SC, Timpson NJ, Najjar SS, Stringham HM,. Nat Genet 2008, 40:1155-1162.e2.

29. Cathiresan S, Melander O, Anevis D, Guiducci C, Burtt NP, Roos C, Hirschhorn NJ, Berglund H, Hedblad B, Groop L, Altshuler DM, Newman-Cheetham C, Orho-Melander M: Polygenes associated with cholesterol and risk of cardiovascular events. N Engl J Med 2008, 358:1240-1249.

30. Motsinger-Reif AA, Dudek SM, Hahn LW, Ritchie MD: Comparison of approaches for machine-learning optimization of neural networks for detecting genetic interactions in epigenetic epidemiology. Genet Epidemiol 2008, 32:350-340.

31. Mukherjee B, Ahn J, Gruber SB, Rennert G, Moreno V, Chatterjee N: Tests for gene-environment interaction from case-control data: a novel study of type I error, power and designs. Genet Epidemiol 2008, 32:65-66.

32. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ: Subgroup analysis in randomized trials: the role of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004, 57:229-236.

33. Lanktree M, Oh J, Hegele RA: Genes testing for atherosclerosis risk in non-involved or pipe dream? Circ Res 2008, 103:851-854.

34. Joy T, Hegele RA: Genetics of metabolic syndrome: is there a role for phenomics? Curr Atheroscler Rep 2010, 12:201-208.

35. Chen SH, Sun J, Dimitrov L, Turner AR, Adams TS, Meyers DA, Allison DB: Detection of gene x gene interactions in genome-wide association studies of human population data. Hum Hered 2007, 63:87-84.

36. Smith PG, Day NE: The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 1984, 13:356-365.

37. Williams RR, Hasstedt SJ, Wilson DE, Ash KO, Yanowitz FF, Reiber GE, Kuida H: Evidence that men with familial hypercholesterolemia can avoid early coronary death. An analysis of 77 carrier genes in four Utah pedigrees. JAMA 1986, 255:219-224.

38. Hegele RA, Emi M, Wu LL, Hopkins PN, Williams RR, Laluol JJ, Kastelein JJ: Mortality over two centuries in large pedigrees with familial hypercholesterolemia: family tree mortality study. BMJ 2001, 322:1019-1023.

39. Pimstone SN, Sun XM, du Souich C, Frohlich JJ, Hayden MR, Soutar AK: Phenotypic variation in heterozygous familial hypercholesterolemia: a comparison between Chinese patients with the same or similar mutations in the LDL receptor gene in China or Canada. Arterioscler Thromb Vasc Biol 1998, 18:309-315.

40. Hegele RA, Breckenridge WC, Cox DW, Maguire GF, Little JA, Connelly PW: Interaction between variants of apolipoproteins C-III and E that affects plasma lipoprotein concentrations. Arterioscler Thromb 1991, 11:1303-1309.

41. Ng DS, Leiter LA, Vezina C, Connelly PW, Hegele RA: Apolipoprotein A-I Q122X causing isolated apolipoprotein A-I deficiency in a family with familial hypercholesterolemia. Am J Cardiol 1998, 81:160-169.

42. Triggs-Raine BL, Kirkpatrick RD, Kelly SL, Norquay LD, Cattini PA, Yamagata H, Kanlay A, Zimm B, Harris BS, Barnett PH, Hegele RA: HNF-1alpha G319S, a transcription-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. Proc Natl Acad Sci USA 2002, 99:4614-4619.

43. Hegele RA: Premature atherosclerosis associated with monogenic insulin resistance. Circulation 2001, 103:2225-2229.

44. Abu-Amero KK, Al-Boudari OM, Mohamed GH, Dzimiri N: The null and M null genotypes of the glutathione S-transferase gene are risk factor for CAD independent of smoking. BMC Med Genet 2006, 7:353.

45. Manfredi S, Federici C, Picano E, Botto N, Rizza A, Andreassi MG, GSTM1, GSTTI and CYP1A1 detoxification gene polymorphisms and susceptibility to smoking-related coronary artery disease: a case-only study. Mutat Res 2007, 621:106-112.
49. Lahoz C, Schaefer EJ, Cupples LA, Wilson PW, Levy D, Osgood D, Parpos S, Pedro-Bolet J, Daly JA, Ordovas JM: Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. Atherosclerosis 2001, 154:529-537.

50. Talmud PJ, Stephens JW, Hawe E, Demissie S, Cupples LA, Hurel SJ, Humphries SE, Ordovas JM: The significant increase in cardiovascular disease risk in APOEpsilon4 carriers is evident only in men who smoke: potential relationship between reduced antioxidant status and APOE. Ann Hum Genet 2005, 69:613-622.

51. Brull DJ, Dhamrait S, Moulding R, Rumley A, Lowe GD, World MJ, Humphries SE, Montgomery HE: The effect of fibrinogen genotype on fibrinogen levels after strenuous physical exercise. Thromb Haemost 2002, 87:37-41.

52. Humphries SE, Henry JA, Montgomery HE: Environment interaction in the determination of levels of haemostatic variables involved in thrombosis and fibrinolysis. Blood Coagul Fibrinolysis 1999, 10(Suppl 1):S17-21.

53. Bereczky Z, Balogh E, Katona E, Czuriga I, Karpati L, Shemirani AH, Edes I, Muszbek L: Decreased factor XII levels in factor XII A subunit Leu34 homozygous patients with coronary artery disease. Thromb Res 2008, 121:469-476.

54. Bereczky Z, Balogh E, Katona E, Pocsai Z, Czuriga I, Szeles G, Karpati L, Adany R, Edes I, Muszbek L: Modulation of the risk of coronary sclerosis/myocardial infarction by the interaction between factor XII subunit A Val34Leu polymorphism and fibrinogen concentration in the high risk Hungarian population. Thromb Res 2007, 120:567-573.

55. Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Felch J, Ordovas JM: Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. J Lipid Res 2002, 43:416-427.

56. Perron P, Brisson D, Sartore M, Blackburn P, Bergeron J, Yohl MC, Despres JP, Gaudet D: Apolipoprotein E and lipoprotein lipase gene polymorphisms interaction on the atherogenic combined expression of hypertriglycerideremia and hyperapolipoproteinemia phenotypes. J Endocrinol Invest 2007, 30:551-557.

57. Hines LM, Stumpfier MJ, Ma J, Gaziano JM, Ridker PM, Hankinson SE, Sacks F, Rimn EB, Hunter DJ: Genetic variation in alcohol dehydrogenase and the beneficial effect of moderate alcohol consumption on myocardial infarction. N Engl J Med 2001, 344:549-555.

58. Younis J, Cooper JA, Miller GJ, Humphries SE, Talmud PJ: Genetic variation in alcohol dehydrogenase 1C and the beneficial effect of alcohol intake on coronary heart disease risk in the Second Northwick Park Heart Study. Atherosclerosis 2003, 180:225-232.

59. Tsai CT, Hwang JJ, Ritchie MO, Moore JH, Chiang FT, Lai LP, Hsu KL, Tseng CD, Lin JL, Tseng YZ: Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort detection of high order gene-gene interaction. Atherosclerosis 2007, 192:172-180.

60. Kardia SL, Bielak LF, Lange LA, Cheverud JM, Boerwinkle E, Turner ST, Sheedy PF, 2nd, Peyser PA: Epistatic effects between two genes in the renin-angiotensin system and systolic blood pressure and coronary artery calcification. Med Sci Monit 2006, 12:CR150-CR158.