

Published in final edited form as:

Circ Cardiovasc Genet. 2012 April 1; 5(2): 217–225. doi:10.1161/CIRCGENETICS.111.961243.

A Genome Wide Association Study for Coronary Artery Disease Identifies a Novel Susceptibility Locus in the Major Histocompatibility Complex

Robert W. Davies, MSc¹, George A. Wells, MSc, PhD¹, Alexandre F.R. Stewart, PhD², Jeanette Erdmann, PhD³, Svati H. Shah, MD, MHS⁴, Jane F. Ferguson, PhD⁵, Alistair S. Hall, MD, PhD⁶, Sonia S. Anand, MD, PhD⁷, Mary S. Burnett, PHD⁸, Stephen E. Epstein, MD⁸, Sonny Dandona, MD², Li Chen, MSc¹, Janja Nahrstaedt, MSc^{3,9}, Christina Loley, MSc^{3,9}, Inke R. König, PhD^{3,9}, William E. Kraus, MD⁴, Christopher B. Granger, MD⁴, James C. Engert, PhD¹⁰, Christian Hengstenberg, MD¹¹, H.-Erich Wichmann, MD, PhD¹², Stefan Schreiber, MD¹³, W. H. Wilson Tang, MD¹⁴, Stephen G. Ellis, MD¹⁴, Daniel J. Rader, MD¹⁵, Stanley L. Hazen, MD, PhD¹⁴, Muredach P. Reilly, MB⁵, Nilesh J. Samani, MD, FMedSci¹⁶, Heribert Schunkert, MD³, Robert Roberts, MD², and Ruth McPherson, MD, PhD^{2,17}

¹Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Canada

²John and Jennifer Ruddy Canadian Cardiovascular Research Centre, University of Ottawa Heart Institute, Ottawa, Canada

³Medizinische Klinik II, Lübeck, Germany

⁴Department of Medicine, Duke University Medical Center, Durham, NC

⁵Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA

⁶Division of Cardiovascular and Neuronal Remodeling, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, UK

⁷Population Health Research Institute, McMaster University, Hamilton Health Sciences, Hamilton, ON, Canada

⁸Cardiovascular Research Institute, Medstar Health Research Institute, Washington Hospital Center, Washington, DC

⁹Institut für Medizinische Biometrie und Statistik, Lübeck, Germany

¹⁰Departments of Medicine and Human Genetics, McGill University, Montreal, Canada

¹¹Klinik und Poliklinik für Innere Medizin II, Universität Regensburg, Regensburg, Germany

¹²Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Klinikum Grosshadern, Munich, Institute of Medical Information Science, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Germany

Address for correspondence: Robert W Davies Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, 40 Ruskin Street - H2267-A, Ottawa, Canada, K1Y 4W7 rw Davies@ottawaheart.ca t:613-798-5555x19124, f: 613-761-4412. Ruth McPherson Atherogenomics Laboratory University of Ottawa Heart Institute 40 Ruskin St - H4203, Ottawa, Canada, K1Y 4W7 rmcpherson@ottawaheart.ca, t:613-761-5256, f: 613-761-4676.

Disclosures Genotyping of PennCath and MedStar was supported by GlaxoSmithKline. There are no other disclosures to report.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

¹³Institut für Klinische Molekularbiologie, Christian-Albrechts Universität, Kiel, Germany

¹⁴Cleveland Clinic, Cleveland, OH

¹⁵Institute for Translational Medicine and Therapeutics, School of Medicine University of Pennsylvania, Philadelphia, PN

¹⁶Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing, Glenfield Hospital; Leicester National Institute for Health Research, Biomedical Research Unit in Cardiovascular Disease, Leicester, UK

¹⁷Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada

Abstract

Background—Recent genome-wide association studies (GWAS) have identified several novel loci that reproducibly associate with CAD and/or MI risk. However, known common CAD risk variants explain only 10% of the predicted genetic heritability of the disease, suggesting that important genetic signals remain to be discovered.

Methods and Results—We performed a discovery meta-analysis of 5 GWASs involving 13,949 subjects (7123 cases, 6826 controls) imputed at approximately 5 million SNPs using pilot 1000 Genomes based haplotypes. Promising loci were followed up in an additional 5 studies with 11,032 subjects (5211 cases, 5821 controls). A novel CAD locus on chromosome 6p21.3 in the major histocompatibility complex (MHC) between *HCG27* and *HLA-C* was identified and achieved genome wide significance in the combined analysis (rs3869109; $p_{\text{discovery}}=3.3\times 10^{-7}$, $p_{\text{replication}}=5.3\times 10^{-4}$ $p_{\text{combined}}=1.12\times 10^{-9}$). A sub-analysis combining discovery GWASs showed an attenuation of significance when stringent corrections for European population structure were employed ($p=4.1\times 10^{-10}$ versus 3.2×10^{-7}) suggesting the observed signal is partly confounded due to population stratification. This gene dense region plays an important role in inflammation, immunity and self cell recognition. To determine whether the underlying association was driven by MHC class I alleles, we statistically imputed common *HLA* alleles into the discovery subjects; however, no single common *HLA* type contributed significantly or fully explained the observed association.

Conclusion—We have identified a novel locus in the MHC associated with CAD. MHC genes regulate inflammation and T cell responses that contribute importantly to the initiation and propagation of atherosclerosis. Further laboratory studies will be required to understand the biological basis of this association and identify the causative allele(s).

Keywords

Coronary Artery Disease; Myocardial Infarction; Meta-Analysis; Genetics

Coronary artery disease is a complex trait resulting from a combination of environmental and genetic factors. The age specific incidence of CAD is increased twofold in subjects with a family history of premature disease and the contribution of family history cannot be fully explained by known CAD risk factors.¹ The identification of genetic variants contributing to CAD may provide insight into novel biological pathways affecting atherosclerosis initiation or progression and ultimately assist in risk assessment. Several genome-wide association studies (GWAS) have identified novel loci that reproducibly associate with coronary artery disease (CAD) and/or myocardial infarction (MI) risk.²⁻¹³ However, in aggregate, the confirmed associations explain a small proportion of known risk. In the recent report by the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDioGRAM) consortium, it was estimated that 13 newly discovered and 10 previously identified loci

associated with CAD account for only 10% of additive genetic heritability of the disease,¹¹ which suggests that important CAD associated loci remain to be discovered.

Most GWASs have benefited from the use of imputation, a statistical technique which uses reference haplotypes from a source such as HapMap along with mathematical models to infer genotypes at single nucleotide polymorphisms (SNPs). Imputation is a reliable way to boost power in GWAS by facilitating meta-analyses of studies genotyped on different arrays and allowing for the interrogation of SNPs not directly genotyped on an array but indirectly tagged through combinations of SNPs.^{14, 15} However, the process of imputation is limited by both the number of reference haplotypes available and the number of SNPs present on the haplotypes. Previous GWAS for CAD have employed imputation using approximately 2.5 million SNPs from the HapMap Phase II populations. More recently, low pass whole genome sequencing by the 1000 Genomes project has allowed for reference haplotypes panels with approximately 10 million SNPs to become available. To what extent undiscovered risk variants for CAD are better tagged using these newer panels is largely unknown.

Furthermore, previous GWAS for CAD have focused solely on the consideration of a single model of inheritance, the additive model. However, recent evidence suggests that the risk imparted by CAD variants is often better modeled by a dominant or recessive mode of inheritance. In the recent CARDIoGRAM meta-analysis, of the 13 novel loci discovered, seven were better suited to either a recessive or dominant mode of inheritance, as opposed to an additive mode.¹¹ This suggests that considering non-additive discovery screens may be a promising method by which to boost power in GWASs.

The goal of the present study was to search for novel variants associated with CAD by performing a meta-analysis of 5 GWAS studies for CAD, making use of 1000 Genomes based imputation and considering additive, recessive and dominant modes of inheritance.

Materials and Methods

Research ethics committee approval and individual informed consent was obtained for all subjects in each study cohort.

Stage 1 Studies

The discovery analysis was comprised of the Ottawa Heart Genomics Study in collaboration with the Cleveland Clinic Gene Bank, (OHGS_A and OHGS_CCGB_B), the CAD GWAS from the Wellcome Trust Case Control Consortium (WTCCC), a subset of the INTERHEART case control study (ITH), and the Duke CATHGEN study (DUKE). Genotyping was performed using the Affymetrix 500K array (OHGS_A, WTCCC), 6.0 array (OHGS_CCGB_B, ITH) or Axiom array (DUKE). Further details on recruitment methodology and phenotypic criteria is given in the Supplementary Methods, while brief summary phenotypic details are available in Table 1.

Stage 1 Analysis

Removal of subjects of admixed or non-European ancestry was performed using the smartPCA program from EIGENSOFT v3.0.¹⁶ Study subjects were processed with 270 HapMap2 subjects for PCA (90 CEU, 90 JPT+CHB, 90 YRI). In the resulting first two dimensions from PCA, k-means was used to ascertain the center of each of the CEU, JPT+CHB and YRI clusters, and the original two PC dimensions were projected onto these axes. Subjects were removed if they fell outside an oval whose major axes were 10 times the

standard deviation of the CEU cluster along the two transformed axes. Results are shown in Supplemental Figure 1.

Genotyping quality control (QC) was synchronized across studies wherever possible, leaving as few as 367,036 SNPs (OHGS_A) or as many as 648,636 SNPs (ITH) to inform imputation (Supplementary Table 1). Pre-imputation QC was set meeting each of a) Hardy-Weinberg Equilibrium (HWE) p -value $>1e-6$; b) Call Rate (CR) $>95\%$ and Minor Allele Frequency (MAF) $>5\%$ or CR $>99\%$ and MAF $<5\%$.

Imputation for each study was performed separately using IMPUTE2 using 112 available CEU haplotypes from the 1000 Genomes project (August 2009 release) as well as 298 haplotypes from a combined CEU/TSI reference panel with CEU subjects not used in the 1000 Genomes panel.¹⁴ Post-imputation QC was set at meeting each of a) HWE $p>1e-6$; b) CR $>90\%$; c) INFO >0.5 . Post-imputation, the number of SNPs ranged from 4,101,281 SNPs (DUKE) to 5,501,436 (OHGS_CCGB_B). All analyses were adjusted for gender and the first two principal components of ancestry; age was not used due to non-overlapping case/control age distributions in several cohorts.

Stage 1 Meta Analysis

A fixed-effects genomic-control inverse-variance meta-analysis using META (v1.2) was employed, where the standard error for each cohort was multiplied by the square root of the lambda as estimated by the genotyped SNPs for that inheritance model for that cohort.¹⁷ Only SNPs with data from at least 2 studies and with an I2 less than 50% were considered. For any locus where a SNP met $p<5\times 10^{-6}$, the SNP with the best p -value across each of the 3 models was taken as the lead SNP for the locus. A separate method of inheritance procedure was then performed to determine whether it showed additive, dominant or recessive inheritance, as outlined below

SNPs were excluded for replication if they fell within 250 KBp of a locus previously reported to be associated with CAD. The exact number of SNPs for each study and model is provided in Supplemental Table 2. Each SNP representing a locus was tested for replication using only its discovery determined inheritance model.

Method of Inheritance Determination

Following a similar procedure to that used in CARDIoGRAM,¹¹ inheritance was estimated for each top SNP as follows. Merging raw data for all discovery studies which passed QC at that SNP, a dominant and recessive model were fit, with the one giving the lower deviance chosen as the most likely method of inheritance. SNPs were then determined to be additive if the likelihood ratio test p -value for the model with both recessive and dominant effects had $p<0.05$ compared to the recessive or dominant model alone.

Stage 2 Studies

Confirmatory evidence was sought in five studies: German Myocardial Infarction Family Studies 1 and 2 (GerMIFS1, GerMIFS2);^{5, 8} PennCath;⁴ MedStar; and Ottawa Heart Genomics Study in collaboration with the Cleveland Clinic GeneBank Supplementary (OHGS_CCGB_S). The latter cohort used controls from the WTCCC2¹⁸ not included in the WTCCC CAD analysis and cases from Ottawa and Cleveland for which genotypes and phenotypes became available after performing Stage 1 of the study. Genotyping was performed using the Affymetrix 500K (GerMIFS1) and 6.0 (GerMIFS2, PennCath, MedStar, OHGS_CCGB_S) arrays. Further descriptions of these cohorts are available in the Supplementary Methods, while brief summary phenotypic details are available in Table 2.

We *a priori* defined SNPs to be significant following the second stage of analysis if they met both a Bonferroni adjusted alpha of 0.05 and if when meta-analyzed with the results from the first stage of the analysis, they achieved a genome-wide significant p-value threshold. Since three tests were employed for each SNP, and the 4,916,498 SNPs tested in this analysis correspond to no more than 548,447 semi-independent SNPs (pairwise removal of one SNP in a pair with $r^2 > 0.8$ using a sliding window along the genome), we defined a genome-wide significant p-value threshold at $0.05/(3 \times 548,447) = 3.0 \times 10^{-8}$.

Ancestry Plots

PCA plots which are superimposed with ancestry were made available based on subject's self-reported grandparental ethnicity as recorded for a subset of OHGS cases. European ethnicity was divided broadly into Northern Europe (Britain, Denmark, Estonia, Iceland, Ireland, Latvia, Lithuania, Scandinavia, Scotland, Wales), Central Europe (Austria, Belgium, France, Germany, Luxembourg, Monaco, Netherlands, Switzerland), Southern Europe (Albania, Andorra, Bosnia, Croatia, Greece, Italy, Portugal, Serbia, Spain), Eastern European (Belarus, Bulgaria, Czech Republic, Hungary, Poland, Romania, Russia, Ukraine), French Canadian, Arabic, Ashkenazi Jewish, Other (specify) and Unknown. The unknown, other and mixed (<3 grandparents from one ethnic region) were merged into one category.

MHC Class I imputation

MHC Class I alleles (*HLA-A*, *HLA-B* and *HLA-C*) were imputed using HLA*IMP, a method similar to SNP imputation which infers *HLA* alleles using genotyped SNP data and a reference dataset with both genotyped SNPs and *HLA* alleles.^{19, 20} Imputation with HLA*IMP is performed using predefined sets of SNPs for each of several common commercial microarrays, including the Affymetrix 500K and 6.0, but not the Axiom array. For each of the discovery cohorts, SNPs surrounding the MHC were phased using IMPUTE2, with 1000 Genomes haplotypes as reference haplotypes. SNPs which were required by HLA*IMP but which failed QC and were not in the 1000 Genomes panel were substituted using their next best tagging variant, either genotyped or imputed. Imputation of *HLA* alleles for subjects genotyped with the Axiom array were generated using a combination of Axiom genotype data and imputed SNPs using the SNPs selected for the 6.0 array.

Post-imputation, we used a posterior probability threshold of 0.7 to define successful *HLA* imputation of an allele, as described in the original HLA*IMP paper. Subjects were considered to have been successfully imputed if both alleles of the relevant *HLA* type met a posterior probability threshold of 0.7. As such, each Class I allele could be treated as a SNP with each subject having 0, 1 or 2 copies of the *HLA* allele. Quality control was set as being similar to SNPs with a post-imputation cutoff of $HWE > 1 \times 10^{-6}$ and missing data rate <10%.

Alleles with a control frequency of less than 2% were removed, corresponding to leaving 88.4%, 81.8% and 93.8% of available genotypes for analysis for *HLA-A*, *HLA-B* and *HLA-C*, respectively. Posterior probability estimates suggested that the procedure was highly accurate, with most alleles achieving a median posterior estimate of >0.98 (Supplementary Table 4).

The effect of an *HLA* allele *vis-à-vis* rs3869109 was assessed using a likelihood ratio test, by assessing a model with both rs3869109 and the relevant *HLA* allele against a model with one or the other. All analyses were performed on the entire discovery panel adjusted for gender and the first two principal components of ancestry.

URLs

META v1.2 <http://www.stats.ox.ac.uk/~jsliu/meta.html>; IMPUTE v2.1.0

<https://mathgen.stats.ox.ac.uk/impute/impute.html>; SNPTEST v2.1.1

<http://www.stats.ox.ac.uk/~marchini/software/gwas/snpstest.html>; GTOOL v0.6.1

<http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html>; EIGENOPT v3.0

<http://genepath.med.harvard.edu/~reich/Software.htm>; LocusZOOM

<http://csg.sph.umich.edu/locuszoom/>; R <http://www.r-project.org/>; HLA*IMP

<https://oxfordhla.well.ox.ac.uk/hla/>

Results

A total of 13,949 subjects (7,123 cases and 6,826 controls) of European ethnicity from 5 CAD case control GWASs were included in the discovery meta-analysis (Table 1). Approximately 5 million SNPs were imputed and tested for each of the three models, and inflation of test statistics was low (additive: 4,916,498 SNPs, $\lambda=1.032$; dominant: 4,889,746 SNPs, $\lambda=1.036$; recessive: 4,637,828 SNPs, $\lambda<1$; see Supplementary Table 1). A total of 37 loci met the stage 1 significance threshold ($p<5\times 10^{-6}$), including 16 known CAD loci and 21 putative loci (Supplemental Table 2). Of the lead SNPs carried forward for each of the 21 novel loci, 2 were identified by an additive model, 12 by a dominant model and 7 by a recessive model. Three of the lead SNPs were 1000 Genomes specific, with no genotyped or pre-1000 Genomes imputation reference panel with $r^2>0.8$.

Follow up was carried out in 5 additional CAD case control studies of European ethnicity with a total sample size of 11,032 (5,211 cases and 5,821 controls). We identified a single novel association with CAD at chromosome 6p21.3 in the major histocompatibility complex (MHC) which met our *a priori* significance requirements and reached a genome wide significant level in the combined analysis (discovery $p=3.3\times 10^{-7}$, replication $p=5.3\times 10^{-4} < 0.05/21=2.3\times 10^{-3}$, global $p=1.12\times 10^{-9} < 3.0\times 10^{-8}$ (Table 3); see Table 1 for more details of rs3869109 and Figure 1 for the forest plot). The lead SNP, rs3869109, lies in an intergenic region between *HCG27* and *HLA-C*. The gene density of the surrounding region is quite high and the discovery signal spans a large region, as shown in Figure 2. The extended MHC, specifically the *HLA* genes, has been implicated in a wide variety of disorders, including autoimmune thyroid disease, multiple sclerosis, psoriasis, celiac disease, systemic lupus erythematosus and type 1 diabetes.²¹ Results for SNPs which did not replicate are shown in Supplemental Table 3.

Correlation with Ethnicity

Due to its role in immunological processes, the MHC, and the region surrounding *HLA-C* and *HLA-B* in particular, is under strong evolutionary pressure. Several studies attempting to determine ancestry-informative markers have noted that SNPs in the MHC are highly correlated with geographic background in European populations.^{22, 23} As susceptibility to CAD varies among different ethnic populations, we sought to confirm that the observed association was not a false positive driven by population stratification by combining all genotyped SNPs in discovery subjects into a single cohort and repeating PCA. Ethnicity was collected from a subset of subjects and superimposed. Results are illustrated in Figure 3 and Supplemental Figure 2, and show strong agreement with past reports demonstrating that the primary ethnic difference between Europeans largely falls into a Northern versus Southern

and Eastern versus Western gradient.²⁴ Figure 3C illustrates the result of predicting the fraction of rs3869109 dosage on case/control status and PC1 using linear regression, and demonstrates that there exists a slight difference among Europeans. We note that the allele frequency of the G allele of rs3869109 is 56.2 % in the CEU population (equivalent to a dosage of 1.12) and 65.2 % in the TSI population (dosage of 1.30), which closely matches these results.

After correcting for known within Europe ethnicity, the significance of the original result remained but was slightly attenuated, with a p-value of 4.1×10^{-10} for the entire merged discovery cohort reduced to 6.2×10^{-8} when adjusting for the first two PC's. Further correcting for more subtle population structure by including the top 10 PC's again slightly tapered the results, leading to a p-value of 3.2×10^{-7} , but nonetheless the result remained highly significant and this value was similar to the meta-analyzed value from the discovery analysis of 3.3×10^{-7} . This indicates that the observed signal is only partly explained by confounding due to European population structure.

Class I MHC types

Association signals in intergenic regions near MHC class I genes may be due to differences in disease susceptibility risk among underlying *HLA* alleles. For example, a psoriasis association at rs10484554 was attributed to HLA-C*0602.¹⁹ Given the proximity of rs3869109 to *HLA-B* and *HLA-C*, we imputed MHC class I alleles into the discovery subjects using HLA*IMP to determine whether the association at rs3869109 could be refined.^{19, 20} We were able to assign bi-allelic *HLA-A*, *HLA-B* and *HLA-C* genotypes to 93.3%, 93.1% and 95.4% of subjects, respectively, using the recommended posterior threshold cutoff of 0.7, which has been previously shown to predict 4 digit *HLA* alleles with >96% accuracy. Not considering lower frequency *HLA* alleles (removal of <2% frequency), no individual *HLA* allele appears to have a significantly greater effect than rs3869109; in fact, no common *HLA* allele significantly improved a model containing rs3869109 (min $p=0.03$), whereas rs3869109 added significantly to every *HLA* allele (max $p=8 \times 10^{-5}$) (Figure 4). Furthermore, repeating the discovery meta-analysis correcting for rs3869109 did not indicate an association beyond rs3869109 ($p>1 \times 10^{-5}$), suggesting no single underlying *HLA* allele accounted for the observed association.

Discussion

Although cholesterol accumulation in the arterial wall is a *sine qua non* for atherosclerosis, it is well accepted that atherosclerosis is driven by a chronic inflammatory process²⁵ in response to subendothelial lipoprotein retention^{26, 27} and involving innate and adaptive immune responses.²⁸ Here, we have identified a novel CAD risk allele at 6p21.3 within the major histocompatibility locus. Relevant to the immunological origins of atherosclerosis²⁹, the observed association signal at rs3869109 spans a large region containing numerous genes in addition to *HLA-B* and *HLA-C*, many with known functions in immune mediated processes. Class II MHC molecules are expressed on antigen presenting cells including dendritic cells within the atherosclerotic plaque and present the peptide fragments of oxidized LDL to helper T cells, key participants in atherosclerosis.²⁹

Although this is the first robust GWAS signal for the MHC locus in CAD, smaller candidate gene studies have found associations with *MHC2TA* encoding the MHC class II transactivator³⁰, and genes involved in pro-inflammatory leukotriene synthesis.³¹ Previous studies have also non-convincingly implicated *LTA* (lymphotoxin A), a MHC class III gene, in myocardial infarction. Ozaki *et al.* identified T26N (rs1041981) in *LTA* which lies ~350Kbp away from rs3869109.³² However, this association was not confirmed by other studies in European and Japanese populations.^{33, 34} Of note, rs1041981 and rs3869109 are

not closely linked as assessed using European ($r^2=0.01$, $D'=0.14$) or Asian ($r^2=0.00$, $D'=0.01$) 1000 Genomes data. Furthermore, rs1041981 is well imputed using Affymetrix 6.0 and Axiom data, and in this study shows no evidence with CAD ($p=0.46$).

We employed a number of novel strategies in an attempt to discover important SNP associations that may have been missed in previous large GWAS for CAD.¹¹ Specifically, this study marks the first attempt to use 1000 Genomes based imputation and non-additive discovery screens in a GWAS for CAD or MI. The non-additive discovery screen enabled us to identify 4 known loci which would have otherwise been missed, indicating that this technique may be an important screening tool. Although the 1000 Genomes based imputation interrogated twice as many SNPs as other recent analyses, including that of the CARDIoGRAM consortium¹¹, only three of the loci brought to replication were based on haplotypes not present in previous imputation reference sets. Thus, the utility of this approach will likely be contingent on the progressive availability of a larger number of reference haplotypes, so as to better enable imputation of lower frequency variants.

Of note, these two approaches do not appear to have contributed substantially to the novel association in the MHC locus, since rs3869109 was genotyped directly in most of the studies included in this analysis and was identified by an additive inheritance model. The fact that the MHC was identified here but not in previous studies likely reflects several factors. One can subdivide previous GWAS efforts into those which featured large scale discovery meta-analyses, most notably the recent CARDIoGRAM and C4D reports, and previous investigations. Studies published before CARDIoGRAM and C4D either featured smaller discovery populations than analyzed here, with a consequent reduction in power, or staged designs, which, although successful, can certainly miss SNPs. Consider for example the highly successful meta-analysis by the Myocardial Infarction Genomics Consortium (MIGC), which, despite identifying 5 new loci for CAD/MI, failed to identify several loci with rather large effect sizes, such as PPAP2B and ADAMTS7. In the C4D meta-analysis, it is worth noting that their European discovery population was smaller than analyzed here (8,424 versus 13,949), and while their combined South-East Asian and European discovery population was significantly larger than analyzed here, C4D was by intent a non-ethnically homogeneous population. If the association is driven by haplotypes of varying frequency among different populations, it is not surprising that it was not evident in a GWAS meta analysis including multi-ethnic populations .

It is also important to note that the design of studies which made up the discovery metaanalysis in this study generally featured early-onset CAD cases with a family history of CAD whereas controls were recruited on the basis of normal angiograms and/or age greater than 70 years with no history or symptoms of vascular disease. This was a consequence of the decision to base the primary (discovery) analysis of this study on subjects for whom genotype level data was available in Ottawa. This was primarily intended to facilitate imputation and analysis in a timely and uniform fashion but also secondarily led to the aforementioned more rigorously defined CAD case/control phenotype. Consequently, replication samples were sought with similar phenotypic definitions at a sample size which would ensure replication of genuinely associated putative loci. Thus, in comparison to CARDIoGRAM, for which there exists substantial overlap in subjects considered here, a more uniform study description was used, which would be expected to decrease effect heterogeneity among studied studies. Most notably, CARDIoGRAM included approximately 60,000 out of 85,000, or two thirds of its subjects from population studies with incident disease and population (unscreened controls). Although CARDIoGRAM was highly successful, the efforts taken here in selecting more similar studies (both ethnically with North-Western Europeans and phenotypically with extreme case control designs) could be expected to decrease effect heterogeneity between studies and facilitate identification of

new loci. Finally, it is worth noting that even in CARDIoGRAM, loci identified elsewhere, including LIPA locus^{12, 35} and PDGFD¹² did not reach significance.

Conclusion

We have identified a novel association with CAD at 6p21.3 within the MHC locus. MHC genes play roles in both innate and adaptive immunity, regulating inflammation and T cell responses that contribute importantly to the initiation and propagation of atherosclerosis.^{28, 29} Further laboratory studies will be required to understand the biological basis of this association and identify the causative allele(s).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the assistance of the staff of the University of Ottawa Information Technology Department, particularly Hana Pika, Pierre Lefebvre and Marc Charbonneau; the staff of the High Performance Computing Virtual Laboratory, particularly Dr. Peter Taillon; and the staff of the John and Jennifer Ruddy Canadian Cardiovascular Genetics Centre, particularly Heather Doelle and Melody Dallaire.

Funding Sources RWD is supported by a CANNeCTIN Biostatistics Fellowship. R.M. is supported by a Merck Frosst Canada/University of Ottawa Chair in Atherosclerosis Research. The Ottawa Heart Genomics Study is supported by the Canadian Institutes of Health Research #MOP82810, #MOP77682; the Canada Foundation for Innovation CFI #11966; the Heart and Stroke Foundation of Ontario #NA6001, #NA6650. **OHGS_CCGB_B and OHGS_CCGB_S cases – CCGB:** The Cleveland Clinic GeneBank is supported by National Institutes of Health grants P01 HL076491, P01 HL098055, R01 DK080732, R01 HL103866 and the Cleveland Clinic Clinical Research Unit of the Cleveland Clinic/Case Western Reserve University CTSA (1UL1RR024989). **DUKE:** WEK is supported by 1RC2-HL101612-01 and SHS is supported by R01-HL095987. **ITH:** A subset of INTERHEART cases and controls of European ancestry was used for this analysis. DNA samples were processed, prepared by the Population Health Research Institute and Clinical Trial Research Laboratory in Hamilton (Canada Foundation for Innovation CFI #88056), and genotyping was performed at the Ottawa Heart Institute. **WTCCC-CAD and OHGS_CCGB_S controls (WTCCC2):** N.J.S. holds a British Heart Foundation Chair of Cardiology and is supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. This study makes use of data generated by the Wellcome Trust Case-Control Consortium 2. A full list of the investigators who contributed to the generation of the data is available from <http://www.wtccc.org.uk>. Funding for the project was provided by the Wellcome Trust under award 076113 and award 085475. **GerMIFS1, GerMIFS2:** Supported by the Deutsche Forschungsgemeinschaft and the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus), the FP6 and FP7 EU funded integrated projects Cardiogenics (LSHM-CT-2006-037593) and ENGAGE (201413), and the bi-national BMBF/ANR funded project CARDomics (01KU0908A). **PennCath/MedStar:** Supported by the Cardiovascular Institute of the University of Pennsylvania and a research grant from GlaxoSmithKline. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Reference List

- (1). Lloyd-Jones DM, Nam BH, D'Agostino RB Sr, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA*. 2004; 291:2204–11. [PubMed: 15138242]
- (2). McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316:1488–91. [PubMed: 17478681]
- (3). Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007; 316:1491–3. [PubMed: 17478679]
- (4). Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction

- in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet*. 2011; 377:383–92. [PubMed: 21239051]
- (5). Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007; 357:443–53. [PubMed: 17634449]
 - (6). Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009; 41:334–41. [PubMed: 19198609]
 - (7). Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–78. [PubMed: 17554300]
 - (8). Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*. 2009; 41:280–2. [PubMed: 19198612]
 - (9). Tregouet DA, König IR, Erdmann J, Munteanu A, Braund PS, Hall AS, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet*. 2009; 41:283–5. [PubMed: 19198611]
 - (10). Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, König IR, Baumert J, et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J*. 2011; 32:158–68. [PubMed: 21088011]
 - (11). Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011; 43:333–8. [PubMed: 21378990]
 - (12). Peden JF, Hopewell JC, Saleheen D, Chambers JC, Hager J, Soranzo N, et al. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011; 43:339–44. [PubMed: 21378988]
 - (13). Wang F, Xu CQ, He Q, Cai JP, Li XC, Wang D, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet*. 2011; 43:345–9. [PubMed: 21378986]
 - (14). Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5:e1000529. [PubMed: 19543373]
 - (15). Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genomewide association studies by imputation of genotypes. *Nat Genet*. 2007; 39:906–13. [PubMed: 17572673]
 - (16). Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–9. [PubMed: 16862161]
 - (17). Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55:997–1004. [PubMed: 11315092]
 - (18). Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet*. 2009; 41:1330–4. [PubMed: 19915572]
 - (19). Dilthey A, Moutsianas L, Leslie S, McVean G. HLA*IMP - an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics*. 27:968–72. 2011. [PubMed: 21300701]
 - (20). Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet*. 2008; 82:48–56. [PubMed: 18179884]
 - (21). de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet*. 2006; 38:1166–72. [PubMed: 16998491]
 - (22). Paschou P, Drineas P, Lewis J, Nievergelt CM, Nickerson DA, Smith JD, et al. Tracing substructure in the European American population with PCA-informative markers. *PLoS Genet*. 2008; 4:e1000114. [PubMed: 18797516]

- (23). Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet.* 2008; 4:e236. [PubMed: 18208327]
- (24). Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. *Nature.* 2008; 456:98–101. [PubMed: 18758442]
- (25). Ross R. Atherosclerosis—an inflammatory disease. *N Engl J Med.* 1999; 340:115–26. [PubMed: 9887164]
- (26). Duewell P, Kono H, Rayner KJ, Sirois CM, Vladimer G, Bauernfeind FG, et al. NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature.* 2010; 464:1357–61. [PubMed: 20428172]
- (27). Tabas I, Williams KJ, Boren J. Subendothelial lipoprotein retention as the initiating process in atherosclerosis: update and therapeutic implications. *Circulation.* 2007; 116:1832–44. [PubMed: 17938300]
- (28). Hansson GK, Hermansson A. The immune system in atherosclerosis. *Nat Immunol.* 2011; 12:204–12. [PubMed: 21321594]
- (29). Lahoute C, Herbin O, Mallat Z, Tedgui A. Adaptive immunity in atherosclerosis: mechanisms and future therapeutic targets. *Nat Rev Cardiol.* 2011; 8:348–58. [PubMed: 21502963]
- (30). Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, Jagodic M, et al. MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat Genet.* 2005; 37:486–94. [PubMed: 15821736]
- (31). Dwyer JH, Allayee H, Dwyer KM, Fan J, Wu H, Mar R, et al. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *N Engl J Med.* 2004; 350:29–37. [PubMed: 14702425]
- (32). Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet.* 2002; 32:650–4. [PubMed: 12426569]
- (33). Koch W, Hoppmann P, Michou E, Jung V, Pfeufer A, Mueller JC, et al. Association of variants in the BAT1-NFKBIL1-LTA genomic region with protection against myocardial infarction in Europeans. *Hum Mol Genet.* 2007; 16:1821–7. [PubMed: 17517687]
- (34). Yamada A, Ichihara S, Murase Y, Kato T, Izawa H, Nagata K, et al. Lack of association of polymorphisms of the lymphotoxin alpha gene with myocardial infarction in Japanese. *J Mol Med.* 2004; 82:477–83. [PubMed: 15175864]
- (35). Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, et al. A Genome-wide Association Study Identifies LIPA as a Susceptibility Gene for Coronary Artery Disease. *Circ Cardiovasc Genet.* 2011; 4:403–12. [PubMed: 21606135]
- (36). Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26:2336–7. [PubMed: 20634204]

Although cholesterol accumulation in the arterial wall is a requirement for the initiation of atherosclerosis, it is well accepted that atherosclerosis is further driven by a chronic inflammatory process in response to lipid deposition and involving innate and adaptive immune responses. In this study, a two-stage GWAS for coronary artery disease including 24,981 subjects, a novel association with CAD has been identified. The signal resides between HLA-C and HCG27 in the major histocompatibility (MHC) locus. MHC genes regulate inflammation and T cell responses that are believed to contribute importantly to the initiation and propagation of atherosclerosis. While the MHC has previously been implicated in atherosclerosis, this marks the first time that an association has met a genome wide significant threshold. The discovery will serve as the basis for future research to understand the causal alleles and genes involved, which will further understanding of the link between inflammation and coronary artery disease.

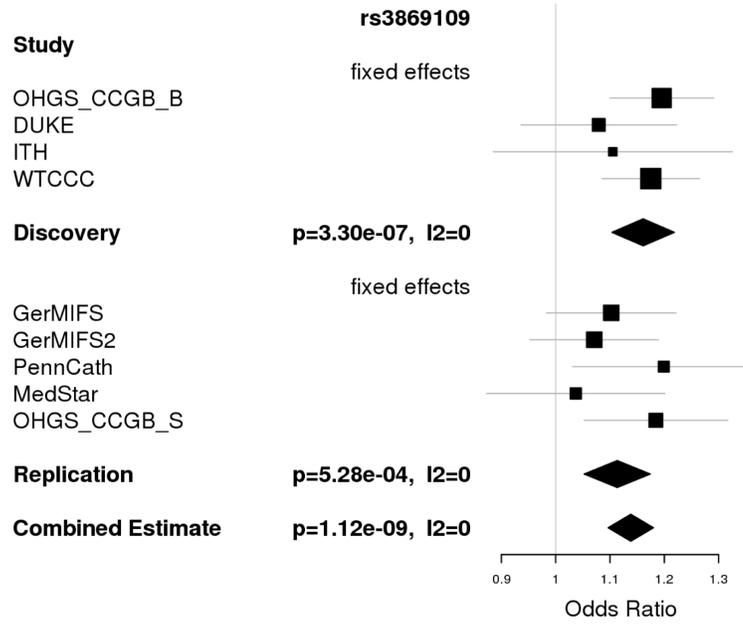


Figure 1. Forest plot for association of rs3869109 with CAD

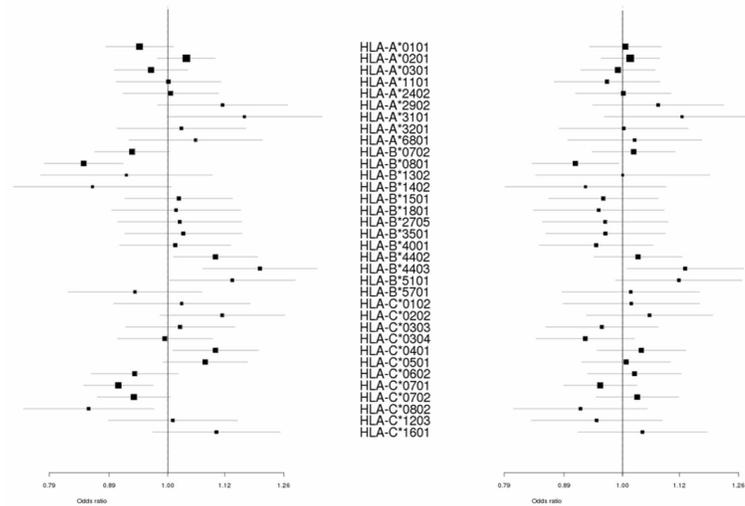


Figure 2. Manhattan plot of rs3869109 and surrounding genes at 6p21.3
 Local Manhattan plot generated using LocusZoom.³⁶ Shown is rs3869109 and flanking region. The final meta-analysis p-value is given for rs3869109, while for all other SNPs, the p-value indicated is from the discovery analysis.

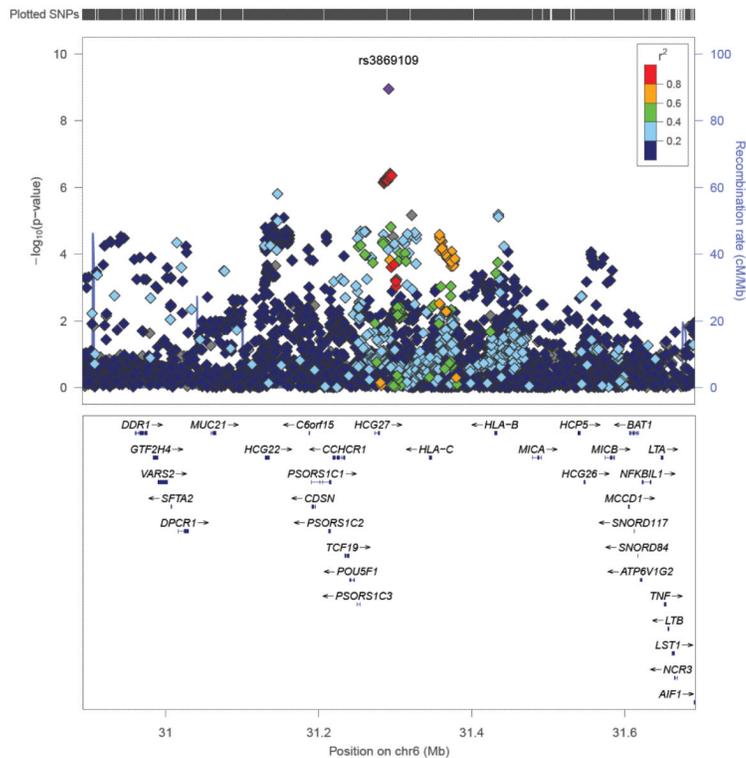


Figure 3. The effect of ethnicity of rs3869109

A) A map of Europe with countries colored to reflect collected ethnic information B) First main principal component for PCA performed on SNPs genotyped in the entire discovery cohort, broadly corresponding to Northern versus Southern geographic background, with a subset colored according to ethnicity. Horizontal axis variation is uniform normal random distribution for ease of visual interpretation. See Figure S2 for the second main principal component and more details. C) Predicted frequency of rs3869109 in cases and controls with respect to first main principal component. This graph is aligned vertically with Panel C.

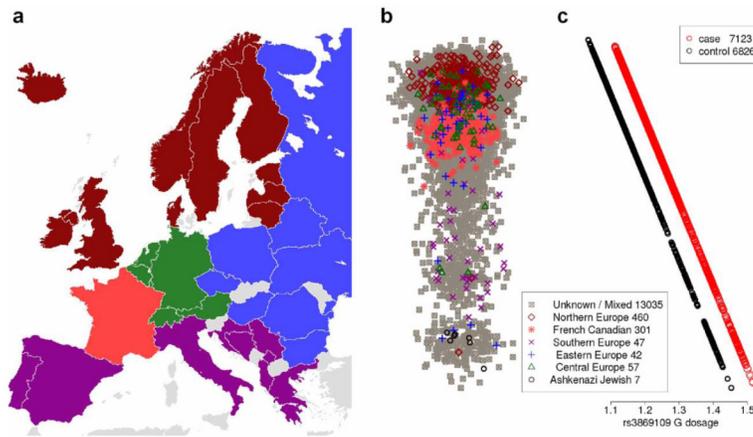


Figure 4. Association of common MHC Class I alleles with CAD
 Association of common (>2% frequency) MHC Class I alleles with CAD in discovery population unadjusted (left) and adjusted (right) for rs3869109.

Table 1

Discovery population characteristics

	N	Age [*] (Mean± SD)	Male (%)	MI (%)	Diabetics excluded	Proportion new [†]
OHGS_A	Case	48.2±7.0	78.1	54.6	Y	6.3
	Control	74.9±4.9	54.6	0	N	3.6
OHGS_CCGB_B	Case	49.8±7.7	75.1	59.8	Y	79.1
	Control	74.8±5.4	49	0	N	76.3
DUKE	Case	56.7±9.7	69.4	48	Y	100
	Control	63.3±8.7	42	0	Y	100
WTCCC	Case	49.8±7.7	79.3	71.5	N	0
	Control	N/A	50	N/A	N	0
ITH	Case	63.3±10.6	45.1	100	Y	100
	Control	64.3±10.3	45.5	0	y	100

* Age refers to age at diagnosis (cases) and age at recruitment (controls)

† Proportion new refers to proportion which have not previously been used for a discovery genome wide association study

Table 2

Replication population characteristics

		N	Age* (Mean± SD)	Male (%)	MI (%)	Diabetics excluded
GerMIFS1	Case	894	50.2±7.8	67.0	100	N
	Control	1604	62.6±10.0	49.2	0	N
GerMIFS2	Case	1218	51.4±11.9	79.6	100	N
	Control	1284	51.2±11.9	52.1	0	N
PennCath	Case	933	52.7±7.6	76.3	50.1	N
	Control	468	61.7±9.6	48.1	0	N
MedStar	Case	874	48.9±6.4	67	48.1	N
	Control	447	59.7±8.9	45.4	0	N
OHGS_CCGB_S	Case	582	54.4±8.6	79.7	N/A	Y
	Control	2728	N/A	52	N/A	N

* Age refers to age at diagnosis (cases) and age at recruitment (controls)

Table 3

Association of rs3869109 with CAD

Band	rsID	Nearby genes	RAF* (RA)	Discovery OR	Discovery p	Replication OR	Replication p	Global OR	Global p	Global I ²
6p21.32	rs3869109	<i>HLA-C</i> , <i>HLA-B</i> , <i>HCC27</i>	0.55 (G)	1.16	3.3×10 ⁻⁷	1.10	5.28×10 ⁻⁴	1.14	1.12×10 ⁻⁹	0

RAF = risk allele frequency, RA = risk allele, OR = odds ratio.

* RAF taken from OHGS_CCGB_B control cohort.