While family history is a strong risk factor for coronary artery disease (CAD), the actual molecular basis has not been characterized in most cases. In 2007, four genome-wide association studies (GWAS) reported strong association (up to $p=10^{-22}$) between a region of chromosome 9p21 and CAD. The high risk genotype was found in up to 30% of people, creating the potential for a clinical genetic test to assist in the prediction of a patient’s risk for CAD. However, the reported effect size of the association is modest (odds ratio ~1.3) and incorporation into a clinical setting may be premature. This paper reviews the advances in human molecular genetics allowing for the production of GWAS, and a result of the four GWAS of CAD released in 2007. In addition, these four GWAS are meta-analyzed and the pooled significance and effect size estimate are calculated. As of yet, there is no explanation for the basis of the association of this region of chromosome 9 with CAD. The consistency of the association across populations suggests a biological mechanism is responsible. Perhaps the greatest contribution from the association is yet to come, as further studies identify the mechanistic basis for the strong association, possibly leading to additional insights into the progression, prevention, and treatment of CAD.

Completion of the sequencing of the human genome was a monumental achievement. Molecular researchers now take for granted the information provided by the sequence, however the clinical applications are not immediately obvious. A limitation of the Human Genome Project was that it produced only a single “reference” sequence. But in order to identify new disease causing mechanisms and cures for disease, we need to go beyond the “reference” and characterize the differences between our genomes, and in turn the effect that these differences have. The Human HapMap consortium and recent genome-wide association studies (GWAS) have set out to capture the inter-individual differences that are associated with disease processes, including coronary artery disease (CAD).

**Basics of Genetic Variation**

Every human nucleus contains 46 chromosomes organized into 23 homologous pairs: 22 autosomes plus a sex chromosome inherited from the mother and 22 autosomes plus a sex chromosome inherited from the father. Chromosomes are made of DNA and can be divided into genes, which are areas that are transcribed into mRNA then translated into proteins, and intergenic regions, which can contain transcription regulating elements. Only ~5% of the genome is thought to be translated ultimately into proteins; the remainder is silent. Most of the genomic sequence (~99.5%), whether coding or silent, is invariant between individuals.

The main form of genomic difference between people is the single nucleotide polymorphism (SNP, pronounced “snip”). A SNP is a single base pair change in the DNA sequence that, by convention, occurs in at least 5% of the population. With the human genome sequence known, the location of any polymorphism can be precisely determined. Each SNP is represented by two alleles, together called a genotype, with one allele residing on each homologous chromosome. SNPs are the most basic and easily measured form of genetic variation that might lead to inter-individual differences at the level of the phenotype. A gene can contain zero to many SNPs: on average SNPs occur about 1 in every 300 to 400 base pairs, and genes range in size from less than a thousand to more than two million base pairs. DNA is replicated extremely effectively with very few errors. Thus, SNPs are generally inherited from our parents rather than occurring spontaneously. Occasionally, a de novo meiotic mutation increases in frequency in successive generations to become a SNP, especially if it confers a survival or fitness advantage. Because they are inherited, SNPs can give insight into the history of the surrounding block of the
chromosome. If an unknown sequence variant that alters disease risk resides within the same block of DNA as a SNP, known as a haplotype block, we can indirectly identify the risk sequence variant by examining the cosegregation of the SNP with the disease. In association studies, researchers compare the prevalence of alleles of SNPs in cases and controls to potentially identify a risk block.

Early in the twenty first century a new technology called “SNP chips”, also called arrays or oligonucleotide microarrays, were introduced. These SNP chips permit simultaneous analysis of up to a million SNP genotypes across the genome in a single experiment for a single DNA sample. Less than five years ago, using older technology a highly motivated graduate student might produce 400 genotypes in a day. Using SNP chips, large genetics centres can now perform GWAS to measure millions of variants in thousands of cases and controls in a few days. In 2007, SNP chips were combined with the GWAS approach to discover new chromosomal regions that were associated with a range of common diseases including rheumatoid arthritis, Crohn’s disease, type 1 diabetes, type 2 diabetes, and CAD.

Genomics of CAD
CAD is the leading cause of death among North Americans. Smoking cessation, weight and diabetes management, low-dose aspirin, hypertension control, and lipid-lowering therapy can substantially reduce the risk of CAD. Many rare single-gene disorders have been discovered that substantially increase the risk of CAD but are only present in a small proportion of the population (eg. familial hypercholesterolemia due to point mutations found in 1 in 500 people). Heritability studies of CAD and myocardial infarction (MI) indicate that common susceptibility gene variants (perhaps found in 30-50% of people) are an important part of CAD risk.

Synthesis of large GWAS in CAD from 2007
Four GWAS from 2007, performed in samples from several different populations, all showed that a region on the short arm of chromosome 9 (namely 9p21) was associated with increased CAD risk. The four studies reviewed in this paper will be referred to as Helgadottir, McPherson, WTCCC and Samani for the remainder of the paper. The chromosome 9p21 results from these four GWAS will be meta-analyzed.

Study samples: Helgadottir studied patients from Iceland, an isolated population, to ensure genetic homogeneity between cases and controls. McPherson studied European Caucasian participants from the Ottawa Heart Study, African American and Caucasian participants

<table>
<thead>
<tr>
<th>Study</th>
<th>Date</th>
<th>Number of cases</th>
<th>Number of controls</th>
<th>Population</th>
<th>Genotyping method</th>
<th>Number of SNPs examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helgadottir et al.</td>
<td>June 2007</td>
<td>4587</td>
<td>12767</td>
<td>Icelandic</td>
<td>Illumina Hap300 array</td>
<td>305 953</td>
</tr>
<tr>
<td>McPherson et al.</td>
<td>June 2007</td>
<td>3989</td>
<td>18808</td>
<td>Mixed</td>
<td>Custom array</td>
<td>75 000</td>
</tr>
<tr>
<td>WTCCC</td>
<td>June 2007</td>
<td>1926</td>
<td>2938</td>
<td>Great Britain Caucasian</td>
<td>Affymetrix 500k array</td>
<td>377 857†</td>
</tr>
<tr>
<td>Samani et al.</td>
<td>Aug. 2007</td>
<td>875</td>
<td>1644</td>
<td>German Caucasian</td>
<td>Affymetrix 500k array</td>
<td>272 602†</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>11337</td>
<td>36157</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Studies described in this paper. Abbreviations - WTCCC: Welcome Trust Case Control Consortium; SNP: Single nucleotide polymorphism. *initial sample of 322 cases and 312 controls, replication of results p<0.025 in 1658 cases and 9380 controls, 2 9p21 SNPs performed in remaining subjects. †SNPs that did not meet Hardy Weinberg equilibrium, had a minor allele frequency <1%, <98% SNP call rate, or were on X chromosome were removed.
from the Atherosclerosis Risk in Communities Study, Danish Caucasian participants from the Copenhagen City Heart Study, and a multi-ethnic sample from the Dallas Heart Study. Controls were carefully selected in order to ensure equal representation in case and control groups. Samani cases and controls were collected from a genetically homogenous region of Bavaria in southern Germany. WTCCC extensively tested for differences in background allele frequencies due to the ethnicity of study subjects. Twelve regions from across the genome were found to be heavily influenced by geographic variation and were thus removed from the WTCCC and Samani studies.

**Study subjects:** All studies included males and females. The Helgadottir cases were diagnosed with a MI before the age of 70 in males and 75 in females. The McPherson cases underwent coronary artery bypass grafting, coronary artery angiography, or care for acute MI before the age of 60. The WTCCC cases had a documented history of MI or coronary revascularization before their 66th birthday. Finally, the Samani cases were diagnosed with a MI prior to the age of 60.

**Statistics:** The WTCCC and Samani studies reported odds ratios comparing individuals with two risk alleles verses no risk alleles and individuals with one risk allele verses no risk alleles (genotype odds ratio). The Helgadottir and McPherson studies reported odds ratios by comparing the number of the risk alleles present in cases to the number of risk alleles present in controls (allelic odds ratio). It is possible to calculate allelic odds ratio from genotype odds ratio (each homozygote contributes two alleles and each heterozygote contributes one of each allele) but not vice versa; hence, the current meta-analysis uses allelic odds ratio. The four GWAS did not assess the exact same SNPs due to different genotyping technologies, but all significantly associated SNPs were located in the same haplotype block (Helgadottir, rs10757278; McPherson, rs10757274; WTCCC, rs6475606; Samani, rs4977574). Meta-analysis was performed using the Mantel-Haenszel method, which creates an estimate of the pooled odds ratio assuming a fixed effects model using the raw data from the individual studies. Haplotype block structure was obtained from the International HapMap Consortium via Haploview.

**Synthesis findings:** The odds ratio, confidence interval and chi-square p-value found for the SNPs on chromosome 9p21 are shown in figure 1. The Mantel-Haenszel summary effect estimate across all four studies is an odds ratio of 1.30 (95% C.I. 1.25-1.36).

![Figure 1. Forest plot of odds ratios obtained comparing 9p21 SNPs with CAD from the four GWAS and the combined effect estimate using Mantel-Haenszel meta-analysis. A vs. B indicates allelic odds ratio and AA vs. BB indicates genotypic odds ratio.](image-url)

**Discussion**

This meta-analysis of 4 large GWAS emphasizes the strength, significance, and replicability of the association on chromosome 9p21 with CAD. The significance level from the combined analysis, $p = 4 \times 10^{-69}$, is a result rarely seen in statistical analysis of biological systems, and is approximately equivalent to the probability of flipping a non-loaded coin and obtaining heads 227 times in a row. However, despite the overwhelming statistical relationship, it is important to recall the difference between statistical significance and clinical significance. The level of confidence that a difference exists between the allele frequencies in CAD cases and controls implies little regarding the clinical implication of this difference. An odds ratio of 1.3 is hardly sufficient for a clinician to introduce any new risk factor test or imaging
method into his/her decision making for a single patient. In comparison, smoking confers an increased risk of MI with an odds ratio of 3.0.\textsuperscript{11} Simply taking a family history verbally and finding out that a parent had heart disease under age 60 confers a 2-fold increased risk.\textsuperscript{12} It is extremely difficult to plumb the importance of a risk of an odds ratio of 1.3 derived from a SNP-based assay on an individual patient.

So why is the genetic finding important? Perhaps this discovery will ultimately contribute to our understanding of the pathogenesis of CAD. What explains the incredibly strong statistically significant association of the 9p21 SNPs with CAD? The 9p21 haplotype block is in the middle of a stretch of DNA with no known function. The closest genes to the reported SNPs are: 1) MTAP (methylthioadenosine phosphorylase), which is involved in polyamine metabolism and the salvage of adenine and methionine; 2) CDKN2A and CDKN2B (cyclin-dependent kinase inhibitor 2A and B), which are tumour suppressor genes; 3) and DMRTA1 (doublesex and mab-3 related transcription factor like family A1), which is important in sexual differentiation. However, these genes are approximately 45 kilobases (kb), 70 kb and 200 kb away from the strongest SNP signal respectively. It is possible that variants at 9p21 could affect these genes via regulatory elements, but this is difficult to determine using current technologies. Moreover, it is difficult to even hypothesize how these genes could be involved in CAD risk. Clearly, there is much more to learn and discover.

In conclusion, four GWAS published in 2007 that enrolled large and non-overlapping study populations have identified a region of 9p21 to be strongly associated with CAD risk. The effect size is too small for the results to be easily integrated into clinical decision-making for a single patient, but the strength of the association suggests that there must be a shared underlying biological mechanism, although the actual mechanism remains unclear at this time. Further studies to identify the reason for the strong association may lead to additional insights into the progression, prevention, and treatment of CAD.

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