A small region on chromosome 9p21.3, discovered in parallel by three groups in the year 2007, is typical of the new understanding of the genetic basis of myocardial infarction (MI). The finding emerged from the application of novel high-throughput genome-wide approaches, the risk-associated allele is frequent, acts independently of traditional risk factors, and confers a modest yet highly reproducible hazard. Since then, another 10 chromosomal regions have been identified to affect the risk of MI or coronary artery disease (CAD). Although the number of risk alleles is growing rapidly, several conclusions can already be drawn from the findings to date. First, it appears that multiple hitherto unknown molecular mechanisms—initiated by these chromosomal variants—ultimately precipitate CAD. Secondly, essentially all Caucasians carry a variable number of risk alleles such that disease manifestation is affected to some extent by these inherited factors in basically all individuals. This means that a better understanding of underlying functional genomic mechanisms may offer novel opportunities to neutralize a broadly based genetic susceptibility for CAD in a large proportion of the population. In parallel, the newly discovered genes open novel opportunities for disease prediction. In summary, modern MI genetics carries the promise to identify individuals at high risk and to improve prevention and therapy of this important disease.

The long road to the identification of myocardial infarction gene loci

A positive family history is one of the strongest cardiovascular risk factors. The European Guidelines utilize this information already today in suggesting preventive measures in first-degree relatives of myocardial infarction (MI) patients. However, an intensive search of the molecular mechanisms explaining the inherited predisposition of MI and coronary artery disease (CAD) has remained largely elusive for more than two decades.

Initial attempts to unravel the genetics of MI were based on the prevalent understanding of the disease process. Candidate gene studies tested the hypothesis that proteins known to be involved in the pathogenesis of atherosclerosis carry mutations or variants that affect their function and ultimately the risk of developing CAD. However, despite over 5000 publications on this topic, variants in only a limited number of genes mainly affecting LDL cholesterol (LDL-C) were convincingly shown to be associated with disease risk (Table 1).

A limitation of multiple non-reproducible candidate gene studies was the restriction to a single or rather few genetic variants tested for association with disease in a given gene. Thus, given the enormous variability found in the human genome, genetic variants of candidate genes with strong effects at the transcriptional level or others affecting the functionality of the protein may have escaped the test for association with disease risk. Moreover, from a statistical point of view, many of the study samples were too small and too heterogeneous to provide reliable information. Thus, in retrospect, it is not surprising that the candidate gene approach resulted in only limited success in the elucidation of MI genes.

On the other hand, the failure to associate candidate genes reproducibly with CAD risk exemplifies how little of the genetic risk can be predicted by currently known pathways. This may be less surprising, given the fact that the predictive information of a
positive family history for MI is largely independent of other traditional risk factors.3

**Family-based studies**

Consequently, novel strategies were developed in order to interrogate the entire human genome without a priori hypothesis on which genes may be responsible for disease risk. Genome-wide affected sib-pair linkage analyses were the principle instrument of the initial wave of such studies. The methodology is based on the assumption that siblings who are both affected by a phenotype are more likely to share the chromosomal region on which the responsible gene is located than would be expected by chance (50%). Given the high mortality of MI, great efforts had to be undertaken in order to identify sufficient numbers of such affected sibling pairs. A few studies were ultimately successful in identification of putative chromosomal regions harbouring MI or CAD genes, but again fine mapping of the regions for identification of the underlying mutations proved to be difficult.4,5

Rarely, MI appears to be inherited in an autosomal-dominant fashion. In families with similar Mendelian inheritance patterns, linkage analysis had successfully identified mutations causing complex cardiovascular disorders including hypertension or cardiomyopathies.6,7 However, linkage analyses carried out on CAD families have turned out to be more difficult.8,9 A reason for this failure may be the high frequency of the disease in the general population resulting in ‘phenocopies’, i.e. disease manifestations due to reasons other than the single-gene defect responsible in the reminder of the family. From the few successful linkage analyses carried out on MI families, it appears that the disease was either caused by private mutations affecting only a single family or related to variants that were without functional relevance in other studies.10

**Genome-wide association studies**

The scientific breakthrough for genetic studies of complex traits came with the sequencing of the 3 billion base pairs of the human genome, followed by the cataloguing of common single-nucleotide polymorphisms (SNPs) at these bases by SNP consortium, and the establishment of the relationship (linkage disequilibrium) between adjacent SNPs by the HapMap Consortium. These consecutive efforts allowed selection of a limited set of SNPs for genotyping to provide information on the vast mass of common SNP variation in the genome. At the same time, arrays were developed which allowed in parallel typing of hundreds of thousands and now over a million SNPs in an individual. These breakthroughs led to the advent of genome-wide association (GWA) studies. The GWA study design does not require multiple affected family members but works best in large case/control samples of unrelated subjects (Figure 1). As a consequence, GWA studies offered better statistical power and increased spatial resolution for identification of chromosomal regions carrying MI genes. In 2007, three studies employed this approach successfully and identified in parallel variants on chromosome 9p21.3 to be associated with risk of MI and CAD.11–13 Since then, several studies have confirmed the exceptional role of the chromosome 9p21.3 region on risk of CAD.14 At the same time, additional GWA and follow-up studies have expanded the list of loci and variants associated with risk of MI and CAD (Figure 2 and Table 2).

**Novel loci associated with risk of myocardial infarction and coronary artery disease identified through genome-wide association studies**

A summary statement on recent high-ranking publications on the discovery of MI/CAD loci could read as follows.

(i) The strongest and most replicated genetic effect on the MI risk known today is located on chromosome 9p21.3, a region without known protein-coding gene.14

(ii) Only 2 out of 11 chromosomal regions, affecting risk of MI or CAD, also affect traditional risk factors. Thus, the majority of novel loci modulate disease risk via hitherto unknown mechanisms.11–13,15–18

(iii) In addition to MI risk, some chromosomal loci affect multiple other seemingly unrelated disease phenotypes.17,19 It is unclear which mechanisms explain these chromosomal hot spots of genetic diseases (pleiotropy).

(iv) All currently identified risk alleles are relatively frequent. For example, in Europeans, the probability to carry one or two risk alleles at the chromosome 9p21.3 locus is 50 and 25%,

---

**Table 1** Genetic variants in the low-density lipoprotein cholesterol metabolism known to affect the risk of myocardial infarction as well

<table>
<thead>
<tr>
<th>Gene</th>
<th>Risk-allele frequency (%)</th>
<th>Increase in LDL per risk allele (%)</th>
<th>Effect of risk allele on MI risk</th>
<th>P-value for association with MI</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCSK9</td>
<td>96</td>
<td>+15</td>
<td>OR 1.5</td>
<td>0.0003</td>
<td>24</td>
</tr>
<tr>
<td>Apo E</td>
<td>2.8</td>
<td>+14</td>
<td>OR 1.29</td>
<td>0.0001</td>
<td>24</td>
</tr>
<tr>
<td>LDLR</td>
<td>11</td>
<td>+6</td>
<td>OR 1.18</td>
<td>0.0001</td>
<td>36</td>
</tr>
<tr>
<td>Apo B</td>
<td>33</td>
<td>+5</td>
<td>OR 1.1</td>
<td>0.004</td>
<td>37</td>
</tr>
</tbody>
</table>

OR, odds ratio for each allele; PCSK9, proprotein convertase subtilisin/kexin type 9; Apo E, apolipoprotein E; LDLR, low-density lipoprotein receptor; Apo B, apolipoprotein B.
respectively. Thus, only 25% of our population is free of this genetic risk factor for MI.\textsuperscript{14}

(v) Each risk allele increases the probability of MI by a relatively small margin, i.e. 10–30% per allele. In other words, individuals who are homozygous for the risk allele on chromosome 9p21.3 carry a 60% risk increase when compared with the 25% of our population, who do not carry this allele.\textsuperscript{14}

(vi) The high frequency of risk alleles, on the other hand, explains why the implications of the recently identified genetic factors at the population level are substantial, even though an affected individual carries only a relatively moderate risk increase. In fact, the genetic epidemiological data demonstrate that basically, all Europeans carry a variable number of risk alleles from the genes listed in Table 2.\textsuperscript{14}

(vii) The genetic risk conferred by the newly discovered genes is independent of the risk signalled by a positive family history. Thus, the molecular-genetic information for risk prediction goes beyond that of all traditional risk factors (Figure 3).

Despite this recent progress in identification of MI/CAD genes, only a relatively limited fraction (<10%) of the overall genetic risk (heritability) of the disease is explained by the currently identified loci. A part explanation for this relates to the limited power of individual GWA studies to detect such loci. Currently, a global consortium (CARDioGRAM) is analysing genome-wide information from more than 20,000 cases of CAD and over 60,000 controls and this will undoubtedly identify additional loci harbouring even more common variants. Other studies are investigating the role of other forms of genomic variation such as copy number variation. Moreover, an increasing effort is being made on elucidating the role of rare variants, which will be aided by novel information on such variants coming out of the 1000 genome project (http://www.1000genomes.org). Finally, parallel genome-wide studies are characterizing a large number of genes affecting the risk of cardiovascular risk factors including dyslipidaemia, hypertension, diabetes mellitus, smoking addiction, and obesity. These findings need to be integrated with loci associated directly with risk of MI/CAD to obtain the fullest picture.

This wealth of new information on heritable aspects of CAD and its risk factors obviously opens multiple windows for scientific exploration. From a clinical point of view, the main focus is on risk prediction and (preventive) therapy for CAD.
Is genetic risk prediction feasible?

From recent discoveries on MI genetics, it is obvious that basically, all individuals of European ancestry carry a variable number of risk alleles (Table 2). For risk prediction, in practical terms, the challenge is to utilize the genomic information for the refinement of existing clinically utilized risk scores. These scores are largely dominated by the predictive information of age and gender and based on prediction of short-term risks. It is obvious that a man in his 70s has a higher risk than a young woman, over the next 10 years, whatever genetic risk burden in these two subjects may carry. The clinically relevant question is what difference genetic factors will make in refining risk prediction in, for example, two middle-aged men in assessing life-long risk in order to better target future preventative measures. Epidemiological studies with prospective DNA and data collection are ongoing to address these clinically important issues. One statistical question that these studies will need to tackle is the handling and integration of information from multiple risk variants to compute the overall “genetic” risk.

A simple approach could be to count the number of risk alleles similar to the quantitative assessment of cholesterol levels in a population. However, this simplification does not take into account the fact that the biological mechanisms at the various loci as well as their impact on risk are likely to be different. Fairly, sophisticated algorithms also considering interactions with lifestyle factors will be required to provide a reliable assessment of the genetic risk carried by an individual.

Presumed mechanisms of myocardial infarction genes

In order to utilize the new genetic information for treatment and prevention of CAD, it will be necessary to understand the functions of the gene(s) at the disease-associated loci and the mechanisms through which they affect coronary risk. As mentioned before, most genes discovered so far do not fit into the clichés of traditional risk mechanisms.

The most prominent example is chromosome 9p21.3. It appears that at this locus, a large antisense non-coding RNA gene (ANRIL) affects the regulation of several other genes. ANRIL is expressed in tissues and cell types that are affected by atherosclerosis. Liu et al. analysed the expression of 9p21.3 transcripts in purified peripheral blood T cells from healthy probands and found a significantly reduced expression of ANRIL and other genes at this locus including p15[INK4b], p16[INK4a], and ARF in patients with CAD, stroke, and aortic aneurysm. A more detailed analysis by Jarinova et al. revealed that a conserved sequence within the 9p21.3 locus has enhancer activity shown in primary aortic smooth muscle. Furthermore, in healthy individuals homozygous for the risk allele, whole-blood RNA expression of short ANRIL variants was increased by 2.2-fold, whereas expression of the long ANRIL variant was decreased by 1.2-fold. Moreover, relevant to CAD, genome-wide expression profiling demonstrated up-regulation of gene sets modulating cellular proliferation in carriers of the risk
Previous work has shown that M-ras is involved in tumor necrosis factor-alpha-stimulated lymphocyte function-associated antigen 1 activation in splenocytes by using mice deficient in this process.23 These findings suggest a plausible role for M-ras in adhesion signaling, which is important in CAD.

Besides these genes with apparently no direct link to traditional risk factors, a few genes have been identified with very evident relationship between traditional risk factor, gene function, and MI risk. The most often replicated locus with such a causal relationship is located on chromosome 1p13.3 and was initially identified through a genome-wide study on CAD patients.11 Since then, this locus has also consistently been associated with LDL-C in several studies.24,25 In European populations, the minor allele is associated with lower levels of LDL-C and lower risk of CAD. Hereby, SNP rs599839 explains about 1% of the variation in circulating LDL-C levels, equivalent to more established genes for LDL regulation, particularly apolipoprotein E (Apo E). The lead SNP rs599839 representing this locus lies intergenic within an ∼97 kb haplotype block on 1p13.3. This chromosomal region harbors four genes: proline-serine-rich coiled coil protein 1 (PSRC1); cadherin, EGF LAG seven-pass G-type receptor 2 (CELSR2), myosin-binding protein H-like (MYBPHL), and sortilin 1 (SORT1). The hepatic mRNA expression levels of PSRC1, CELSR2, and SORT1 have been shown to correlate with LDL-C plasma levels in a mouse model of cardiovascular disease and in a human cohort.24,26 The CAD risk allele (A) is associated with lower levels of CELSR2 and SORT1 expression and with higher LDL-C levels. Both genes fall into the category of cell surface receptor linked signal transduction.26 SORT1 is a transmembrane receptor protein that binds to a variety of different ligands and has been shown to be involved in the endocytosis and intracellular degradation of lipoprotein lipase.27,28 a rate-limiting enzyme of triglyceride hydrolysis in lipoproteins. Recently, SORT1 has also been linked to the endocytosis of apolipoprotein A-V-containing chylomicrons.28

A second example of a surprising new player in the pathogenesis of MI is the locus on chromosome 3q22.3 spanning the muscle RAS oncogene homolog (MRAS) gene.15 The M-ras protein belongs to the ras superfamily of guanosine triphosphate-binding proteins and is widely expressed in all tissues, with a very high expression in the cardiovascular system, especially in the heart. Previous work has shown that M-ras is involved in tumor necrosis factor-alpha-stimulated lymphocyte function-associated antigen 1 activation in splenocytes by using mice deficient in this process.23 These findings suggest a plausible role for M-ras in adhesion signaling, which is important in CAD.

Table 2 Coronary artery disease/myocardial infarction loci identified by genome-wide association studies

<table>
<thead>
<tr>
<th>Band</th>
<th>SNP</th>
<th>Risk-allele frequency in Europeans (%)</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>Genes</th>
<th>Underlying mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p13.3</td>
<td>rs599839</td>
<td>77</td>
<td>1.13 (1.08–1.19)</td>
<td>1.1 x 10^{-14}</td>
<td>PSCR1, CELSR2, SORT1, MYBPHL</td>
<td>LDL</td>
<td>11,38</td>
</tr>
<tr>
<td>1q41</td>
<td>rs3008621</td>
<td>72</td>
<td>1.10 (1.04–1.17)</td>
<td>1.4 x 10^{-9}</td>
<td>MIA3</td>
<td>Unknown</td>
<td>11,16,38</td>
</tr>
<tr>
<td>2q33</td>
<td>rs6725887</td>
<td>14</td>
<td>1.17 (1.11–1.23)</td>
<td>1 x 10^{-8}</td>
<td>WDR12, ALSC2RT3</td>
<td>Unknown</td>
<td>15</td>
</tr>
<tr>
<td>3q22.3</td>
<td>rs9818870</td>
<td>15</td>
<td>1.15 (1.11–1.19)</td>
<td>7.4 x 10^{-13}</td>
<td>MRAS</td>
<td>Unknown</td>
<td>15</td>
</tr>
<tr>
<td>6p24</td>
<td>rs12526453</td>
<td>65</td>
<td>1.13 (1.08–1.17)</td>
<td>1 x 10^{-9}</td>
<td>PHACTR1</td>
<td>Unknown</td>
<td>16</td>
</tr>
<tr>
<td>6q26-27</td>
<td>rs2048327 rs3127599 rs7767084 rs10755578</td>
<td>18</td>
<td>1.20 (1.13–1.28)</td>
<td>1.2 x 10^{-9}</td>
<td>SLC22A3, LPA2, LPA</td>
<td>Lp(a)</td>
<td>18</td>
</tr>
<tr>
<td>9p21.3</td>
<td>rs1333049</td>
<td>52</td>
<td>1.20 (1.16–1.25)</td>
<td>2.8 x 10^{-21}</td>
<td>MTAP, CDKN2A, CDKN2B, ANRIL</td>
<td>Unknown</td>
<td>11–13,39</td>
</tr>
<tr>
<td>10q11</td>
<td>rs501120</td>
<td>84</td>
<td>1.11 (1.05–1.18)</td>
<td>9.5 x 10^{-8}</td>
<td>SDF1</td>
<td>Unknown</td>
<td>11,38</td>
</tr>
<tr>
<td>12q24</td>
<td>rs11065987</td>
<td>34</td>
<td>1.14 (1.10–1.19)</td>
<td>5.2 x 10^{-11}</td>
<td>SH2B3</td>
<td>Unknown</td>
<td>17,40</td>
</tr>
<tr>
<td>12q24.3</td>
<td>rs2259816</td>
<td>36</td>
<td>1.08 (1.05–1.11)</td>
<td>4.8 x 10^{-7}</td>
<td>HNF1A, C12orf43</td>
<td>Unknown</td>
<td>15</td>
</tr>
<tr>
<td>21q22</td>
<td>rs9982601</td>
<td>13</td>
<td>1.19 (1.14–1.27)</td>
<td>6 x 10^{-11}</td>
<td>SLC5A3, MRPS6, KCNE2</td>
<td>Unknown</td>
<td>16</td>
</tr>
</tbody>
</table>

Figure 3 Chromosome 9p21.3 and family history for myocardial infarction. Each risk allele from this locus increases the risk of myocardial infarction by a factor of about 1.3 or 30%, irrespectively whether the family history is positive or not.
might therefore be increased sortilin expression leading to enhanced LDL uptake into the liver, which in turn results in lower LDL-C levels and subsequently lower risk of CAD. 29

Are there different genetic forms of coronary artery disease?

The risk allele on chromosome 9p21.3 appears to increase the susceptibility of CAD, stroke, as well as peripheral arterial disease and also induce aneurysmal disease of the aorta and cerebral vessels.19,30 Such variable patterns of disease make it likely that there are distinct molecular-genetic mechanisms underlying different forms of atherosclerotic disease. Indeed, patients with left main disease display a higher heritability than patients with CAD in general.31 Likewise, calcification of the coronary arteries or the aorta displayed high recurrence rates within families.32 In mouse models, such forms of calcifications could be explained by single-gene mutations.33,34 Consequently, ongoing studies are evaluating as to whether specific anatomical features of CAD or MI are related to specific risk alleles as well.

Outlook

Modern genetics open up an entirely new view on the biology of CAD. It appears that its genetically triggered pathogenesis is largely independent of that mediated through traditional risk factors. Nevertheless, it may be that genetic risk variants require a specific environment in order to come into effect. Indeed, it is likely that genetic factors are embedded in a network that also includes potentially modifiable co-factors (Figure 4).35 A better knowledge of these interactions will be vital to gain the greatest benefit from this emerging information on the genetic predisposition to CAD.

This progress report also shows that rapid scientific developments in the discovery of hazardous molecular variants make better disease prediction possible. Clinicians may expect that addition of genetic information can refine the predictive accuracy of risk scores. However, data from prospective studies need to be awaited to learn which specific groups within our population benefit from determination of genetic risk. Any clinical implementation of genetic testing will require that individuals receive not only more precise estimations of risk but also profit from specific interventions that lower the overall risk of CAD.

Figure 4 Integrative view of genetic risk variants affecting gene expression or function in the context of traditional risk factors and hitherto unspecified environmental co-factors. Ultimately, biological networks may malfunction resulting in the precipitation of coronary artery disease.
Funding
N.J.S. holds a Chair funded by the British Heart Foundation and is supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease. Preparation of this review was undertaken under the EU-funded Integrated Project Cardiogenics (LISHM-CT-2006-037593) and the BMBF-funded German National Genome Network (NGFN-Plus) Project Atherogenomics (FKZ: 01GS0831).

Conflict of interest: none declared.

References
3. Erdmann J, Grosshening A, Brandt PS, Konig IR, Hengstenberg C, Hall AS, N. J. S. holds a Chair funded by the British Heart Foundation and is supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease. Prevention of this review was undertaken under the EU-funded Integrated Project Cardiogenics (LISHM-CT-2006-037593) and the BMBF-funded German National Genome Network (NGFN-Plus) Project Atherogenomics (FKZ: 01GS0831).
4. Funding
5. N.J.S. holds a Chair funded by the British Heart Foundation and is supported by the Leicester NIHR Biomedical Research Unit in Cardiovascular Disease. Preparation of this review was undertaken under the EU-funded Integrated Project Cardiogenics (LISHM-CT-2006-037593) and the BMBF-funded German National Genome Network (NGFN-Plus) Project Atherogenomics (FKZ: 01GS0831).
6. Funding
7. References
8. Funding
9. Funding
10. Funding
11. Funding
12. Funding
Genetics of myocardial infarction


30. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.