Large scale association analysis for identification of genes underlying premature coronary heart disease: cumulative perspective from analysis of 111 candidate genes

J J McCarthy, A Parker, R Salem, D J Moliterno, Q Wang, E F Plow, S Rao, G Shen, W J Rogers, L K Newby, R Cannata, K Glatt, E J Topol, for the GeneQuest Investigators*

Background: to date, only three groups have reported data from large scale genetic association studies of coronary heart disease using a case control design.

Methods and results: to extend our initial report of 62 genes, we present data for 210 polymorphisms in 111 candidate genes genotyped in 352 white subjects with familial, premature coronary heart disease (onset age for men, 45; for women, 50) and a random sample of 418 population based whites. Multivariate logistic regression analysis was used to compare the distributions of genotypes between cases and the comparison group while controlling for age, sex, body mass, diabetes, and hypertension. Significant associations were found with polymorphisms in thrombospondin-4 (THBS4), thrombospondin-2 (THBS2) and plasminogen activator inhibitor-2 (PAI2), the strongest being with the A387P variant in THBS4 (p<0.002). The THBS2 and THBS4 associations have since been replicated. We evaluated polymorphisms in 40 genes previously associated with coronary heart disease and found significant associations with 10: ACE, APOE, F7, FGB, GP1BA, IL1RN, LRP1, MTHFR, SELP, and THPO. For five of these genes, the polymorphism associated in our study was different from that previously reported, suggesting linkage disequilibrium as an explanation for failure to replicate associations consistently across studies. We found strong linkage disequilibrium between polymorphisms within and between genes, especially on chromosome 1q22-2q5, a region containing several candidate genes.

Conclusions: despite known caveats of genetic association studies, they can be an effective means of hypothesis generation and complement classic linkage studies for understanding the genetic basis of coronary heart disease.
Candidate gene choice, polymorphism selection, and genotyping

A total of 243 candidate genes were initially chosen for analysis based on previously reported genetic associations or knowledge of their involvement in coronary heart disease pathways of endothelial cell biology, vascular biology (thrombosis), lipid metabolism, the coagulation cascade, and other risk factors (diabetes, obesity). To cast a wide net to find suggestive associations with coronary heart disease or myocardial infarction in as many genes as possible, we focused on 1–3 common polymorphisms, the majority single nucleotide polymorphisms in the coding region per gene. Single nucleotide polymorphisms were identified from two sources: a proprietary database, the result of screening the coding region of several thousand genes by three methods in and around the gene were typed. This report is based on an analysis of 111 of these genes where validated coding region single nucleotide polymorphisms were readily available.

High throughput genotyping was carried out by one of two methods: single base extension with detection by fluorescence energy transfer or fluorescence polarisation or the 5’ nuclease assay with allele specific TaqMan probes.18 PCR conditions, oligonucleotide primers, and probes are available from the authors upon request.

RESULTS

Descriptive statistics

The characteristics of the study population, which included 352 white patients with coronary heart disease and a random population based sample of 418 whites, are shown in table 1. At enrolment, cases ranged in age from 29 to 72 years (mean 48.1 years) and the random population sample from 20–70 (mean 43.2 years). Cases were retrospectively ascertained with an average time from their qualifying event to enrollment of 6.8 years (range 0–30 years) for women and 9.3 years (range 0–42 years) for men. Thus, the cases were probably overrepresented by coronary heart disease survivors. Because all cases were chosen from families originally ascertained for coronary heart disease, they are enriched for coronary heart disease risk factors including hypertension, high body mass index, and diabetes (types 1 and 2 combined). No differences were found in current smoking status. Due to the retrospective ascertainment of cases, history of smoking prior to coronary heart disease onset could not reliably be obtained. No difference was found in the proportion of cases with diabetes, high body mass index, hypertension, current smoking, or of the male sex when comparing myocardial infarction to other qualifying events, nor between subjects enrolled at the Cleveland Clinic versus other sites.
Polymorphisms

Ninety percent of polymorphisms typed came from a proprietary database of coding region variants discovered through direct sequencing efforts. A total of 210 polymorphisms, including 207 single nucleotide polymorphisms and three insertion/deletions were successfully evaluated in 111 candidate cardiovascular disease genes in our case control study. A complete list of the polymorphisms and their flanking sequences can be found in the supplemental table (online at http://jmg.bmjournals.com/supplemental). An additional 76 single nucleotide polymorphisms were evaluated but not analysed. These included 13 single nucleotide polymorphisms for which working assays could not be developed; one single nucleotide polymorphism that was not in Hardy-Weinberg equilibrium (p<0.0001), 51 that typed as monomorphic, suggesting they were either not real or too infrequent in our population to be detected; and two single nucleotide polymorphisms for which the location in the gene could not be confirmed.

Of those 210 polymorphisms successfully typed (supplemental table, online at http://jmg.bmjournals.com/supplemental), most occurred in the coding region of the gene (44% missense and 36% silent). The remaining polymorphisms were in the untranslated regions or intronic regions immediately flanking the exons. For 62 candidate genes, only one polymorphism was typed. There were two polymorphisms typed in 22 genes, three in 12 genes, and 4–8 in 15 genes. Only 25 polymorphisms had minor allele frequencies ≤5%. There were 27 polymorphisms with minor allele frequencies between 6–10%, 77 between 11–25%, and 84 between 26–50%. All of the single nucleotide polymorphisms were within the limits of Hardy-Weinberg equilibrium, taking into account the multiple testing that was done (all p values >0.0008).

Missing genotype data can be an issue in studies employing high throughput genotyping. For seven polymorphisms, >20% of subjects were missing genotypes; for 40 polymorphisms, 11–19% were missing; for 69 polymorphisms, 6–10% were missing; for 97 polymorphisms, <5% of subjects were missing genotypes. Genotypes for all but five single nucleotide polymorphisms were in Hardy-Weinberg equilibrium in the controls. This number was within the range of expected deviations, given multiple testing. Uncorrected p values for the five single nucleotide polymorphisms ranged from 0.05 to 0.009. The test for population stratification resulted in a mean χ² of 1.2 for the 72 single nucleotide polymorphisms typed, suggesting no significant stratification exists in our population (p>0.05).

Replication of previous genetic associations

In our study, we evaluated associations between coronary heart disease or myocardial infarction and polymorphisms in 40 genes for which prior associations with coronary heart disease have been described (“replication genes”). For 30 of these genes, we examined the exact variants previously associated (table 2). For the remaining 10 genes (ACE, CD14, IL1A, IL1RN, F13A1, LIPC, PON2, TGFBI, THBD, THPO, VWF), the polymorphisms examined were not the same as those associated previously. Polymorphisms in a total of 10 genes were significantly associated with coronary heart disease or myocardial infarction after controlling for covariates (table 3). In five of these genes, it was the exact same variant as previously reported: APOE, F7, FGB, GP1BA, and MTHFR. For the remaining five genes, associations were found between coronary heart disease or myocardial infarction and a polymorphism different from that previously reported: ACE, IL1RN, THPO, LR1P1, and SELP. The SELP_3 single nucleotide polymorphism associated with coronary heart disease and myocardial infarction in our study was only in moderate linkage disequilibrium (D’ = 0.43) with the SELP_1 (T715P) single nucleotide polymorphism previously associated. The LR1P1_3 single nucleotide polymorphism associated in our study was in strong linkage disequilibrium with the LR1P1_5 single nucleotide polymorphism previously associated. For the remaining three single nucleotide polymorphisms, linkage disequilibrium with the previously associated variant was unknown.

Restricting the cases to those with myocardial infarction resulted in enhanced associations for single nucleotide polymorphisms in four genes—APOE, F7, GP1BA, and MTHFR—all consistent with a recessive mode of inheritance. The single nucleotide polymorphism in FGB, on the other hand, was associated only with the full set of coronary heart disease cases and was consistent with a dominant or codominant mode of inheritance.

Genetic associations described in the GeneQuest study for the first time

We have found significant (p<0.05) associations between coronary heart disease or myocardial infarction and single nucleotide polymorphisms in eight genes, which have not been previously described by others: ECE1, HRG, PAI2, PLCG1, SDC4, THBS1, THBS2, and THBS4. Only the THBS genes were published in our initial report.6 For an additional three genes—ANXA4, PLOD2, and PROC—the 95% confidence interval for one of the genotype groups excluded 1.0. Among these 11 top associations, only three were significantly associated with coronary heart disease or myocardial infarction in adjusted analyses: THBS4, THBS2, and PAI2 (table 4). Restricting the cases to those with myocardial infarction resulted in enhanced associations for all single nucleotide polymorphisms. The THBS4 variant conferred a greater than twofold increased odds of myocardial infarction in both heterozygotes and homozygotes and, of all the 210 polymorphisms examined in this study, was the strongest association remaining after adjustment for covariates (p = 0.002).

Linkage disequilibrium

We also undertook an analysis of the extent of pairwise linkage disequilibrium between single nucleotide polymorphisms within a gene. To do this, we calculated the normalised disequilibrium parameter, D’, whose values range from −1 to +1. For the 185 pairs of single nucleotide polymorphisms examined, approximately 80% of the single nucleotide polymorphisms were in moderate linkage disequilibrium (D’ = 0.43) with the single nucleotide polymorphism previously associated. The LR1P1_3 single nucleotide polymorphism associated in our study was in strong linkage disequilibrium with the LR1P1_5 single nucleotide polymorphism previously associated. For the remaining three single nucleotide polymorphisms, linkage disequilibrium with the previously associated variant was unknown.

### Table 1 Characteristics of the study population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Coronary heart disease cases</th>
<th>Random population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>352</td>
<td>418</td>
</tr>
<tr>
<td>% male</td>
<td>70%</td>
<td>44%</td>
</tr>
<tr>
<td>Mean current age (SD)</td>
<td>48.1 (7.4)</td>
<td>43.2 (14.3)</td>
</tr>
<tr>
<td>(range)</td>
<td>(29–79)</td>
<td>(20–70)</td>
</tr>
<tr>
<td>Mean age at diagnosis (SD)</td>
<td>39.3 (5.0)</td>
<td>Not applicable</td>
</tr>
<tr>
<td>(range)</td>
<td>(22–51)</td>
<td></td>
</tr>
<tr>
<td>% diabetes</td>
<td>14%</td>
<td>4%</td>
</tr>
<tr>
<td>% hypertension</td>
<td>44%</td>
<td>15%</td>
</tr>
<tr>
<td>Mean body mass index (SD)</td>
<td>29.3 (5.7)</td>
<td>26.7 (6.2)</td>
</tr>
<tr>
<td>(range), kg/m²</td>
<td>(16–61)</td>
<td>(15–58)</td>
</tr>
<tr>
<td>% current smokers</td>
<td>30%</td>
<td>25%</td>
</tr>
<tr>
<td>Qualifying events</td>
<td>Not applicable</td>
<td></td>
</tr>
<tr>
<td>Angiogram &gt;70% stenosis</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Coronal artery bypass graft</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>54%</td>
<td></td>
</tr>
<tr>
<td>Angioplasty</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

All variables differed significantly (p<0.0001) between cases and controls, except current smoking status.
polymorphism pairs had “useful” linkage disequilibrium (D’ > 0.30) and 50% of the single nucleotide polymorphism pairs gave values of D’ > 0.90. As in other reports, disequilibrium was highly variable but in general it was strongest for single nucleotide polymorphisms in close proximity. The median linkage disequilibrium dropped off substantially for single nucleotide polymorphisms separated by >20 kb.

We examined disequilibrium between genes that cluster together on a chromosome. For the IL1 gene cluster including IL1RN, IL1B, and IL1A on q12-q22, linkage disequilibrium is strong between the two polymorphisms within the IL1RN gene (D’ = 0.76) and between IL1A and IL1B, separated by less than 5 kb (D’ = 0.81), but weak between IL1RN and IL1B (D’ = 0.22) and between IL1RN and IL1A (D’ = 0.20). Due to the high linkage disequilibrium in the region, the association we found between IL1RN and coronary heart disease may reflect haplotypes previously associated with coronary heart disease.

Three fibrinogen genes FGA, FGB, and FGG are clustered in a region of ~50 kb on chromosome 4q31 (fig 1). Within the FGB gene, strong disequilibrium exists between four polymorphisms typed (all pairwise D’ > 0.99), but not between single nucleotide polymorphisms in FGA, FGB, and FGG. Because of its close proximity (~12 kb) to FGB, Pleiotropic Regulator 1 (PLRG1) might be considered a positional candidate gene for coronary heart disease. The selectin genes, SELP and SELL, and the factor V gene (F5) are clustered in an ~220 kb region on 1q22-q25 (fig 2). Significant disequilibrium exists between single nucleotide polymorphisms within each gene, as well as between the SELP and SELL genes and the SELP and F5 genes, the latter pair being separated by ~2000 bp. Significant associations found with single nucleotide polymorphisms in SELP and F5 single nucleotide polymorphisms in our and other studies may reflect common haplotypes spanning these genes.

Results of the analysis of haplotypes within 9 of the 13 genes where novel associations were uncovered were inconclusive because of the cumulative effects of missing data on sample size and power.

DISCUSSION

The current report is one of only three published large scale genetic association studies of coronary heart disease, and the only one among white Americans. Subsequent to our interim report on 62 candidate genes among genetically enriched coronary heart disease cases and population controls, Yamada and colleagues assessed 112 candidate gene polymorphisms in Japanese individuals with myocardial infarction and Ozaki et al examined over 90 000 gene based single nucleotide polymorphisms in Japanese patients who had had myocardial infarction. In our initial report, we described association between variants in three thrombospondin genes and myocardial infarction. Since then, various other groups have replicated the association with two of these genes. The thrombomodulin 4 (THBS4) A387P single nucleotide polymorphism was confirmed to be significantly associated with myocardial infarction in men in the study from Yamada and colleagues, a European study of premature coronary heart disease, a population of myocardial infarction cases and controls from the Cleveland Clinic, and in the Atherosclerosis Risk in Communities study. The THBS2 association has also been replicated in the European study of premature coronary heart disease and the Atherosclerosis Risk in Communities study. Furthermore, we have undertaken functional genomic studies and demonstrated that the A387P single nucleotide polymorphism is a gain of function mutation that interferes with endothelial cell adhesion and proliferation, which may account for predisposition to myocardial infarction. Now that we have expanded our assessment of candidate vascular biology genes from 62 to

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Common name</th>
<th>P value coronary heart disease unadjusted/adjusted</th>
<th>P value myocardial infarction unadjusted/adjusted</th>
<th>Representative reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA1</td>
<td>ABCA1_1</td>
<td>R219K</td>
<td>0.97/0.74</td>
<td>0.83/0.25</td>
<td>14</td>
</tr>
<tr>
<td>ADR2B</td>
<td>ADR2B_1</td>
<td>I/D</td>
<td>0.84/0.66</td>
<td>0.96/0.82</td>
<td>19</td>
</tr>
<tr>
<td>AQP</td>
<td>AQP_1</td>
<td>M237</td>
<td>0.50/0.93</td>
<td>0.45/0.70</td>
<td>11</td>
</tr>
<tr>
<td>AGTR1</td>
<td>AGTR1_1</td>
<td>1166 A/C</td>
<td>0.89/0.78</td>
<td>0.92/0.93</td>
<td>21</td>
</tr>
<tr>
<td>APOA1</td>
<td>APOA1_1</td>
<td>−75 G/A</td>
<td>0.05/0.17</td>
<td>0.10/0.37</td>
<td>22</td>
</tr>
<tr>
<td>APOB</td>
<td>APOB_1</td>
<td>R361I</td>
<td>0.21/0.24</td>
<td>0.37/0.28</td>
<td>23</td>
</tr>
<tr>
<td>APOB</td>
<td>APOB_2</td>
<td>Xbal</td>
<td>0.23/0.16</td>
<td>0.51/0.24</td>
<td>24</td>
</tr>
<tr>
<td>APOC3</td>
<td>APOC3_1</td>
<td>Sat1</td>
<td>0.34/0.98</td>
<td>0.12/0.47</td>
<td>24</td>
</tr>
<tr>
<td>APOE</td>
<td>APOE_1</td>
<td>*</td>
<td>0.02/0.03</td>
<td>0.02/0.04</td>
<td>25</td>
</tr>
<tr>
<td>CD14</td>
<td>CD14_1</td>
<td>−159 T/C</td>
<td>0.73/0.54</td>
<td>0.94/0.84</td>
<td>26</td>
</tr>
<tr>
<td>CETP</td>
<td>CETP_4</td>
<td>Taq1B</td>
<td>0.28/0.28</td>
<td>0.60/0.69</td>
<td>27</td>
</tr>
<tr>
<td>CYBA</td>
<td>CYBA_1</td>
<td>H272Y</td>
<td>0.10/0.54</td>
<td>0.08/0.33</td>
<td>28</td>
</tr>
<tr>
<td>EDN1</td>
<td>EDN1_2</td>
<td>S665 G/T</td>
<td>0.16/0.58</td>
<td>0.43/0.58</td>
<td>9</td>
</tr>
<tr>
<td>F2</td>
<td>F2_2</td>
<td>20210 G/A</td>
<td>0.98/0.99</td>
<td>0.84/0.61</td>
<td>29</td>
</tr>
<tr>
<td>F5</td>
<td>F5_1</td>
<td>Leiden</td>
<td>0.31/0.29</td>
<td>0.17/0.11</td>
<td>30</td>
</tr>
<tr>
<td>F7</td>
<td>F7_2</td>
<td>R335EQ</td>
<td>0.03/0.06</td>
<td>0.05/0.04</td>
<td>31</td>
</tr>
<tr>
<td>FGB</td>
<td>FGB_3</td>
<td>−455 G/A</td>
<td>0.02/0.04</td>
<td>0.08/0.23</td>
<td>32</td>
</tr>
<tr>
<td>GPIBA</td>
<td>GPIBA_1</td>
<td>T145M</td>
<td>0.05/0.04</td>
<td>0.05/0.05</td>
<td>33</td>
</tr>
<tr>
<td>HFE</td>
<td>HFE_1</td>
<td>C282Y</td>
<td>0.15/0.37</td>
<td>0.35/0.39</td>
<td>34</td>
</tr>
<tr>
<td>IL6</td>
<td>IL6_1</td>
<td>−174 G/C</td>
<td>0.85/0.91</td>
<td>0.83/0.78</td>
<td>35</td>
</tr>
<tr>
<td>ITGB3</td>
<td>ITGB3_1</td>
<td>Pl-A1/A2</td>
<td>0.27/0.07</td>
<td>0.21/0.15</td>
<td>36</td>
</tr>
<tr>
<td>LPL</td>
<td>LPL_1</td>
<td>*</td>
<td>0.25/0.50</td>
<td>0.47/0.71</td>
<td>37</td>
</tr>
<tr>
<td>LRP1</td>
<td>LRP1_1</td>
<td>4012 C/T</td>
<td>0.09/0.06</td>
<td>0.14/0.10</td>
<td>38</td>
</tr>
<tr>
<td>MTHFR</td>
<td>MTHFR_1</td>
<td>677 C/T</td>
<td>0.07/0.10</td>
<td>0.07/0.02</td>
<td>39</td>
</tr>
<tr>
<td>NO3S</td>
<td>NO3S_1</td>
<td>D298E</td>
<td>0.29/0.27</td>
<td>0.25/0.29</td>
<td>40</td>
</tr>
<tr>
<td>PAI1</td>
<td>PAI1_1</td>
<td>I/D</td>
<td>0.71/0.78</td>
<td>0.77/0.82</td>
<td>41</td>
</tr>
<tr>
<td>PCG1</td>
<td>PCG1_1</td>
<td>M355</td>
<td>0.69/0.53</td>
<td>0.56/0.33</td>
<td>42</td>
</tr>
<tr>
<td>PPP2G</td>
<td>PPP2G_2</td>
<td>161 C/T</td>
<td>0.22/0.35</td>
<td>0.25/0.25</td>
<td>43</td>
</tr>
<tr>
<td>SCARB1</td>
<td>SCARB1_2</td>
<td>*</td>
<td>0.87/0.52</td>
<td>0.82/0.48</td>
<td>44</td>
</tr>
<tr>
<td>SELE</td>
<td>SELE_1</td>
<td>S128R</td>
<td>0.12/0.17</td>
<td>0.06/0.09</td>
<td>45</td>
</tr>
<tr>
<td>SELP</td>
<td>SELP_1</td>
<td>T171P</td>
<td>0.27/0.29</td>
<td>0.28/0.46</td>
<td>46</td>
</tr>
</tbody>
</table>

*one of two single nucleotide polymorphisms that define the E2/E3/E4 haplotypes.
In the current study, the persistent finding of THBS4 as the most significant further anchors its potential of being clinically meaningful. The ability of various diverse groups to replicate these findings in population studies adds validity to clinically meaningful. The ability of various diverse groups to replicate some, but not all, previously reported associations.

Furthermore, our study provides strong support for the contribution of linkage disequilibrium in the failure to replicate genetic associations. In our study, we were able to replicate some, but not all, previously reported associations. Replication failure may be the result of testing not the underlying causal variant, but rather a variant in linkage disequilibrium with the causal variant. Our data support strong linkage disequilibrium between single nucleotide polymorphisms within a gene, and the presence of significant disequilibrium between genes in close proximity. Furthermore, while we were not able to replicate some associations directly, we did find evidence for association with other single nucleotide polymorphisms in the same gene. While examining a single polymorphism in a gene may be an efficient strategy for hypothesis generation, especially in large scale or genome wide studies, follow up should include a more comprehensive analysis of individual polymorphisms and haplotypes in the region to identify the “causative” variant prior to attempting to replicate the association in independent populations.

There are key limitations to acknowledge with our report which could have lead to either type I (false positive) or type II (false negative) errors affecting our results. Most associations found in our study were only nominally significant, a function of both the complex aetiology of coronary heart disease and a relatively small sample size. The p values presented here were not adjusted for multiple testing, which increases the likelihood of false positive associations. In addition, uncontrolled confounding is another possible source of spurious associations. While many important confounders were controlled for in our analysis, some potential confounders, such as lipid levels, were not. The retrospective nature of our study prohibited accurate temporal assessment of other possible confounding factors such as smoking, where only current smoking status was reliably obtained.

In addition, a number of factors could result in type II error, leading to the inability to detect a true underlying

### Table 3

<table>
<thead>
<tr>
<th>Gene (polymorphism)</th>
<th>Genotype</th>
<th>Random population</th>
<th>Coronary heart disease cases</th>
<th>Myocardial infarction cases</th>
<th>p value (1-tailed)</th>
<th>Adjusted odds ratio</th>
<th>95% confidence interval</th>
<th>Adjusted odds ratio</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1RN (IL1RN_3)</td>
<td>CC</td>
<td>337</td>
<td>268</td>
<td>150</td>
<td>1.00</td>
<td>1.00</td>
<td>0.97 (0.67, 1.40)</td>
<td>1.00</td>
<td>0.97 (0.67, 1.40)</td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>44</td>
<td>53</td>
<td>24</td>
<td>1.00</td>
<td>1.00</td>
<td>1.69 (1.03, 2.77)</td>
<td>1.69 (1.03, 2.77)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>31</td>
<td>22</td>
<td>13</td>
<td>1.00</td>
<td>1.00</td>
<td>1.45 (0.75, 2.80)</td>
<td>1.45 (0.75, 2.80)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>148</td>
<td>139</td>
<td>69</td>
<td>1.00</td>
<td>1.00</td>
<td>1.61 (1.12, 2.33)</td>
<td>1.61 (1.12, 2.33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>201</td>
<td>213</td>
<td>77</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>LRP1 (LRP1_3)</td>
<td>TT</td>
<td>40</td>
<td>21</td>
<td>11</td>
<td>1.00</td>
<td>1.00</td>
<td>0.89 (0.47, 1.69)</td>
<td>0.89 (0.47, 1.69)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>172</td>
<td>157</td>
<td>85</td>
<td>1.00</td>
<td>1.00</td>
<td>1.57 (1.09, 2.26)</td>
<td>1.57 (1.09, 2.26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>189</td>
<td>115</td>
<td>57</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MTHFR (MTHFR_1)</td>
<td>TT</td>
<td>46</td>
<td>47</td>
<td>31</td>
<td>1.00</td>
<td>1.00</td>
<td>1.69 (1.01, 2.83)</td>
<td>1.69 (1.01, 2.83)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>172</td>
<td>165</td>
<td>65</td>
<td>1.00</td>
<td>1.00</td>
<td>0.97 (0.67, 1.40)</td>
<td>0.97 (0.67, 1.40)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>191</td>
<td>151</td>
<td>73</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SELP (SELP_3)</td>
<td>AA</td>
<td>8</td>
<td>14</td>
<td>11</td>
<td>1.00</td>
<td>1.00</td>
<td>3.82 (1.43, 10.2)</td>
<td>3.82 (1.43, 10.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>90</td>
<td>100</td>
<td>54</td>
<td>1.00</td>
<td>1.00</td>
<td>1.38 (0.93, 2.04)</td>
<td>1.38 (0.93, 2.04)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>270</td>
<td>207</td>
<td>108</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>THPO (THPO_1)</td>
<td>GG</td>
<td>79</td>
<td>55</td>
<td>33</td>
<td>1.00</td>
<td>1.00</td>
<td>0.52 (0.31, 0.89)</td>
<td>0.52 (0.31, 0.89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>176</td>
<td>146</td>
<td>84</td>
<td>1.00</td>
<td>1.00</td>
<td>0.55 (0.36, 0.85)</td>
<td>0.55 (0.36, 0.85)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>75</td>
<td>87</td>
<td>42</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

*Measure of linkage disequilibrium between the polymorphism listed and the polymorphism with a prior association with coronary heart disease or myocardial infarction in other studies. N/A, not applicable, indicates this is the exact polymorphism with a prior association; unknown, indicates that linkage disequilibrium could not be measured since the prior polymorphism was not evaluated in our study.
association. Among these are limited polymorphism and haplotype analysis within a gene, low allele frequencies, small effect sizes, and the relatively small sample size of our study population. In addition, since our control group was not selected to be free of coronary heart disease, the resulting misclassification of controls could bias the results toward the null. Therefore, we cannot rule out that an association does exist between any of the genes examined in this study and susceptibility to coronary heart disease or myocardial infarction.

Despite a relatively small sample size, we were able to replicate some prior associations reported in the literature and substantiate our previously reported association with THBS4 that withstood replication and has proved to be biologically relevant. Thus, the feasibility of our study design, which employed an enriched source of cases for detecting genetic associations, was demonstrated. By selecting cases with very early onset disease and a strong family history, our cases were weighted toward those individuals whose disease has a strong genetic aetiology. While our study design generated some interesting hypotheses related to genetic variation associated with coronary heart disease, further studies are required to demonstrate both the reproducibility and generalisability of these findings to non-familial, late onset cases.

The optimal approach to understanding the genetic basis of a complex disease such as myocardial infarction or coronary heart disease has been debated. Despite the vast collective efforts of many investigators to demonstrate reproducible associations between specific gene polymorphisms and coronary heart disease or myocardial infarction, no clear cut, reliable associations have been found. In a review of genetic association studies, Hirschhorn and colleagues have pointed out that only 6 of 166 putative associations between genetic variants and complex diseases were consistently replicated. Subsequent work by this group also indicated the problems of false negative, underpowered studies and highlighted the need for very large sample sizes to assess the modest but real risk of a polymorphism for a common disease. In addition, disease heterogeneity and linkage disequilibrium with nearby loci, as illustrated in our work must be considered. Recently, Colhoun et al have expressed their pessimistic concerns "that association approaches will always be hopelessly simplistic and reductionist". On the other hand, with the exception of the recent identification of the myocardial infarction gene, MEF2A, the genome wide...
linkage analysis approach to coronary heart disease has thus far only identified putative loci but has not homed in on causative genes. While some may argue that individual results from small studies such as this only add to the confusion in the literature, the accumulation of both positive and negative findings will stimulate initiatives by independent groups to replicate novel hypotheses, minimise publication bias, and facilitate meta-analysis. Ultimately, either traditional meta-analysis or pooling of raw data across a number of similar observational studies with a thorough analysis of the effects of ascertainment criteria will enable investigators to identify subsets of individuals for whom a particular genetic marker may have the greatest impact on risk of coronary heart disease.

Authors' affiliations
J J McCarthy, R Salem, University of Alabama, Birmingham, AL 35203, USA
W J Rogers, University of Kentucky, Lexington, KY 40536, USA
J P Parker, K Glott, Millennium Pharmaceuticals, Inc., Cambridge, MA 02139, USA

REFERENCES

ECHO

Duodenal adenomatosis in familial adenomatous polyposis

S Bülow, J Björk, I J Christensen, O Fausa, H Järvinen, F Moesgaard, H F A Vasen, the DAF Study Group

Background: The prevalence of duodenal carcinoma is much higher in familial adenomatous polyposis (FAP) than in the background population, and duodenal adenomatosis is found in most polyposis patients.

Aims: To describe the long term natural history of duodenal adenomatosis in FAP and evaluate if cancer prophylactic surveillance of the duodenum is indicated.

Methods: A prospective five nation study was carried out in the Nordic countries and the Netherlands.

Patients: A total of 368 patients were examined by gastroduodenoscopy at two year intervals during the period 1990–2001.

Results: At the first endoscopy, 238 (65%) patients had duodenal adenomas at a median age of 38 years. Median follow up was 7.6 years. The cumulative incidence of adenomatosis at age 70 years was 90% (95% confidence interval (CI) 79–100%), and of Spigelman stage IV 52% (95% CI 28–76%). The probability of an advanced Spigelman score increased during the study period (p<0.0001) due to an increasing number and size of adenomas. Two patients had asymptomatic duodenal carcinoma at their first endoscopy while four developed carcinoma during the study at a median age of 52 years (range 26–58). The cumulative incidence rate of cancer was 4.5% at age 57 years (95% CI 0.1–8.9%) and the risk was higher in patients with Spigelman stage IV at their first endoscopy than in those with stages 0–III (p<0.01).

Conclusions: The natural course of duodenal adenomatosis has now been described in detail. The high incidence and increasing severity of duodenal adenomatosis with age justifies prophylactic examination, and a programme is presented for upper gastrointestinal endoscopic surveillance.


www.jmedgenet.com