ORIGINAL ARTICLE

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*

....

See end of article for authors' affiliations
....

Correspondence to:
E J Topol, MD,
The Cleveland Clinic
Foundation, Department of
Cardiovascular Medicine,
9500 Euclid Ave., Desk
F-25, Cleveland, Ohio
44195; Topole@ccf.org

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 (p<0.05) 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-q25, 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.

C oronary heart disease is the leading cause of death and disability in the Western world. Coronary heart disease is a complex disease with both environmental and genetic determinants. Risk factors known to increase the likelihood of developing coronary heart disease include hypertension, diabetes, obesity, hypercholesterolaemia, and diet. In addition, a positive family history of coronary heart disease in first degree relatives is a strong independent risk factor for coronary heart disease.[1] The risk of developing coronary heart disease is 2–12 times higher for individuals with a first degree family history compared with those without a family history of coronary heart disease.[1] The risk is highest for early age of onset in the affected family members and the greater the number of affected first degree relatives.[2 3]

Despite years of research, the genetic basis of coronary heart disease remains to be fully elucidated. Many of the classic risk factors are themselves under genetic control (blood pressure, lipids, obesity), but they account for only a portion of familial aggregation of coronary heart disease.[1] To date, four published studies have been carried out in high risk families employing genome wide scans to identify regions of the genome linked to coronary heart disease. Two significant linkages were identified for coronary heart disease on chromosomes 2q21.1-22 and Xq23-26 using a Finnish study population.[4] A study among Indo-Mauritians[5] identified a region on chromosome 16p13 linked to coronary heart disease. Another study[6] found evidence for linkage to myocardial infarction in a region on chromosome 14q11.2-12. A study of acute coronary syndrome[7] found a suggestive

linkage to 2q36-37.3. The specific genes underlying these linkages have yet to be found.

Most studies of coronary heart disease genetics to date have focused on a candidate gene approach in unrelated cases and controls, a design that is more powerful than linkage analysis in families when examining a specific gene of interest. Three large scale studies have been published to date, two in Japanese populations[8 9] and an initial report of the current study by our group.[10] Despite numerous positive findings from these and other studies, consistent evidence for any single gene associated with coronary heart disease is lacking. Several recent meta-analyses suggest that the effect of individual genes on the risk of complex traits such as coronary heart disease may be weak.[11 12] Therefore, attempts should be made to improve the power to detect genetic associations. Besides using a large sample size, selecting cases that are genetically loaded may also improve power.

This study was undertaken to extend substantively our first high throughput candidate gene association study of coronary heart disease[10] by comparing genetically enriched cases to ethnically matched adult general population controls. The goal of our study was twofold: first, to attempt to replicate associations previously reported in the literature and second, to survey genetic variants in a large number of candidate genes to generate novel hypotheses regarding their association with coronary heart disease.

....

Abbreviations: PAI2, plasminogen activator inhibitor-2; THBS2, thrombospondin-2; THBS4, thrombospondin-4

METHODS

Study population

Cases were drawn from the GeneQuest population, a study of affected sibling pair families ascertained for early onset familial coronary heart disease from 15 sites in the United States,[10] the majority coming from Ohio, Alabama, and North Carolina. Families were enrolled if the probands had premature coronary heart disease (age of onset in men, 45; in women, 50) and a sibling of theirs met the same criteria. Coronary heart disease was defined as having either myocardial infarction, angiography with >70% stenosis, or surgical or percutaneous coronary revascularisation and was confirmed by review of medical records. Of the 762 cases from the GeneQuest family study, 365 unrelated white individuals were chosen (the sibling with the earlier age of onset was selected). Thirteen cases had insufficient DNA and were removed, leaving a final sample size of 352 cases.

Comparisons were made with a random population sampling of 447 American whites, ascertained through random digit dialling in the greater Atlanta area, and ranging in age from 20–70 years. The comparison group was not selected based on disease status or family history and thus may include individuals with either overt coronary heart disease or at high risk of coronary heart disease. Twenty nine individuals were removed from further analysis after finding that they were missing >50% of their genotypes due to insufficient quantity or quality of DNA. The final sample included 418 individuals.

At enrolment, anthropometric measures, health and lifestyle questionnaires, medication usage, and family history data were collected by a trained interviewer for both cases and controls. Non-fasting blood samples were drawn for measurement of serum markers and extraction of DNA. Genomic DNA was isolated from peripheral blood lymphocytes of the controls using the Puregene kit (Gentra Systems) according to the manufacturer's suggested protocol at a commercial laboratory. Diabetes status (type I and type II) was based on self report. Hypertension was defined based on a single measure of blood pressure as systolic blood pressure >140 or diastolic blood pressure >90 or on the use of antihypertensive medication. Body mass index was calculated in kg/m^2.

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 ethnically diverse cell lines,[13] and the public database, dbSNP. For genes having possible associations with coronary heart disease, additional common single nucleotide polymorphisms 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.[14] PCR conditions, oligonucleotide primers, and probes are available from the authors upon request.

Association analysis

Analyses were performed using the SAS statistical package version 6.12 (SAS Institute, Inc.). Differences in the distribution of genotypes and other covariates between cases and the random population sample were assessed with a Wald χ^2 statistic (2 degrees of freedom test) implemented in the PROC LOGISTIC procedure in SAS. Two outcomes were examined: coronary heart disease, where the entire set of cases was compared with the general population, and myocardial infarction, where a subset of coronary heart disease cases whose qualifying event for enrolment was myocardial infarction, were compared with the random population. For each single nucleotide polymorphism, odds ratios and Wald 95% confidence intervals were calculated for homozygous variant (minor allele) and heterozygous genotypes with the homozygous wild-type genotype as the reference. Men and women were analysed separately for single nucleotide polymorphisms in two X-linked genes. Multivariate analysis was performed for individual single nucleotide polymorphisms that showed statistically significant associations ($p<0.05$) to control for potential confounding effects of age, sex, body mass index, hypertension, and type 2 diabetes. Hardy-Weinberg equilibrium was assessed with a χ^2 test of goodness of fit among the random population sample.

The potential confounding effect of population stratification was evaluated with the genomic control method.[15] Briefly, 96 unlinked single nucleotide polymorphisms spread throughout the genome were selected and 72 successfully typed in 100 randomly selected cases and 100 from the general population sample. Differences in allele frequencies of the 72 single nucleotide polymorphisms between the two groups were tested with a χ^2 statistic. In the absence of stratification, the expected mean χ^2 value for all single nucleotide polymorphisms typed is 1.

Linkage disequilibrium between single nucleotide polymorphisms within the same gene or for genes that cluster together on a chromosome was assessed with the normalised disequilibrium parameter, D',[16] using the EM algorithm[17] among the combined cases and controls.

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 enrolment 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.

Table 1 Characteristics of the study population

Characteristic	Coronary heart disease cases	Random population
Sample size	352	418
% male	70%	44%
Mean current age (SD)	48.1 (7.4)	43.2 (14.3)
(range)	(29–74)	(20–70)
Mean age at diagnosis (SD)	39.3 (5.0)	Not applicable
(range)	(22–51)	
% diabetes	14%	4%
% hypertension	44%	13%
Mean body mass index (SD)	29.3 (5.7)	26.7 (6.2)
(range), kg/m²	(16–61)	(15–58)
% current smokers	30%	25%
Qualifying events:		Not applicable
Angiogram >70% stenosis	15%	
Coronary artery bypass graft	15%	
Myocardial infarction	54%	
Angioplasty	12%	
Other	4%	

All variables differed significantly (p<0.0001) between cases and controls, except current smoking status.

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.bmjjournals.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.bmjjournals.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 >0.008).

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 χ^2 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, LRP1, 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 LRP1_3 single nucleotide polymorphism associated in our study was in strong linkage disequilibrium with the LRP1_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.[10] 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

Table 2 Results of the analysis of polymorphisms associated with coronary heart disease or myocardial infarction in prior studies

Gene	Polymorphism	Common name	P value coronary heart disease unadjusted/adjusted	P value myocardial infarction unadjusted/adjusted	Representative reference
ABCA1	ABCA1_1	R219K	0.97/0.74	0.83/0.25	18
ADRA2B	ADRA2B_1	I/D	0.84/0.66	0.96/0.82	19
AGT	AGT_1	M235T	0.50/0.93	0.45/0.70	20
AGTR1	AGTR1_1	1166 A/C	0.89/0.78	0.92/0.95	21
APOA1	APOA1_1	−75 G/A	0.05/0.17	0.10/0.37	22
APOB	APOB_1	R3611Q	0.21/0.24	0.37/0.28	23
APOB	APOB_2	XbaI	0.23/0.16	0.51/0.24	23
APOC3	APOC3_1	Sst1	0.34/0.98	0.12/0.47	24
APOE	APOE_1	*	0.02/0.03	0.02/0.04	25
CD14	CD14_1	−159 T/C	0.73/0.54	0.94/0.84	26
CETP	CETP_4	Taq1B	0.28/0.28	0.60/0.69	27
CYBA	CYBA_1	H72Y	0.10/0.54	0.08/0.33	28
EDN1	EDN1_2	5665 G/T	0.16/0.58	0.43/0.58	9
F2	F2_2	20210 G/A	0.98/0.99	0.84/0.61	29
F5	F5_1	Leiden	0.51/0.29	0.17/0.11	30
F7	F7_2	R353Q	0.03/0.06	0.05/0.04	31
FGB	FGB_3	−455 G/A	0.002/0.04	0.08/0.23	32
GP1BA	GP1BA_1	T145M	0.05/0.04	0.05/0.05	33
HFE	HFE_1	C282Y	0.15/0.37	0.35/0.39	34
IL6	IL6_1	−174 G/C	0.85/0.91	0.83/0.78	35
ITGB3	ITGB3_1	PL-A1/A2	0.27/0.07	0.21/0.15	36
LPL	LPL_1	HindIII	0.25/0.50	0.47/0.71	37
LRP1	LRP1_5	4012 C/T	0.09/0.06	0.14/0.10	38
MTHFR	MTHFR_1	677 C/T	0.19/0.10	0.07/0.01	39
NOS3	NOS3_1	D298E	0.29/0.27	0.25/0.29	40
PAI1	PAI1_1	I/D	0.71/0.78	0.77/0.82	41
PON1	PON1_1	M55L	0.69/0.53	0.56/0.33	42
PPARG	PPARG_2	161 C/T	0.22/0.35	0.25/0.55	43
SCARB1	SRB1_2	HaeIII	0.87/0.52	0.82/0.48	44
SELE	SELE_1	S128R	0.12/0.17	0.06/0.09	45
SELP	SELP_1	T715P	0.27/0.29	0.28/0.46	46

*one of two single nucleotide polymorphisms that define the E2/E3/E4 haplotypes

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 2q12-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_3 and coronary heart disease may reflect haplotypes previously associated with coronary heart disease.[47]

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 all 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,[10] Yamada and colleagues assessed 112 candidate gene polymorphisms in Japanese individuals with myocardial infarction[9] and Ozaki et al examined over 90 000 gene based single nucleotide polymorphisms in Japanese patients who had had myocardial infarction.[8] 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 thrombospondin 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,[48] a population of myocardial infarction cases and controls from the Cleveland Clinic[49] and in the Atherosclerosis Risk in Communities study.[50] The THBS2 association has also been replicated in the European study of premature coronary heart disease[48] and the Atherosclerosis Risk in Communities study.[50] 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,[51] which may account for predisposition to myocardial infarction. Now that we have expanded our assessment of candidate vascular biology genes from 62 to

Table 3 Significant associations between coronary heart disease or myocardial infarction and single nucleotide polymorphisms in replication genes

Gene (polymorphism)	Genotype	Random population	Coronary heart disease cases	Myocardial infarction cases	Coronary heart disease adjusted odds ratio (95% confidence interval)	p coronary heart disease	myocardial infarction adjusted odds ratio (95% confidence interval)	p myocardial infarction	D'*
APOE (APOE_1)	CC	5	12	9	2.92 (0.90, 9.45)	0.03	4.63 (1.38, 15.6)	0.04	N/A
	CT	96	91	42	1.54 (1.05, 2.28)		1.23 (0.76, 1.99)		
	TT	310	220	125	1.00		1.00		
ACE (DCP1_1)	GG	78	78	43	0.87 (0.55, 1.38)	0.01	0.91 (0.53, 1.56)	0.02	unknown
	AG	185	124	64	0.58 (0.39, 0.85)		0.53 (0.33, 0.85)		
	AA	137	130	72	1.00		1.00		
F7 (F7_2)	AA	4	13	7	3.88 (1.19, 12.6)	0.06	5.05 (1.29, 19.8)	0.04	N/A
	AG	82	68	40	1.20 (0.79, 1.82)		1.34 (0.82, 2.20)		
	GG	312	243	130	1.00		1.00		
FGB (FGB_1)	TT	19	5	4	0.43 (0.14, 1.27)	0.04	0.65 (0.20, 2.14)	0.25	0.99
	CT	133	83	47	0.66 (0.46, 0.97)		0.70 (0.45, 1.09)		
	CC	254	240	125	1.00		1.00		
FGB (FGB_2)	AA	19	5	4	0.43 (0.12, 1.28)	0.02	0.65 (0.20, 2.15)	0.20	1.00
	AG	137	78	44	0.63 (0.43, 0.92)		0.67 (0.43, 1.06)		
	GG	255	239	124	1.00		1.00		
FGB (FGB_3)	CC	20	6	5	0.49 (0.18, 1.36)	0.04	0.76 (0.25, 2.31)	0.23	N/A
	CT	137	82	46	0.64 (0.44, 0.93)		0.68 (0.43, 1.06)		
	TT	253	239	125	1.00		1.00		
GP1BA (GP1BA_1)	TT	1	4	4	5.31 (0.56, 50.6)	0.04	10.2 (1.08, 96.1)	0.05	N/A
	CT	44	53	24	1.69 (1.03, 2.77)		1.61 (0.87, 2.96)		
	CC	337	268	150	1.00		1.00		
IL1RN (IL1RN_3)	CC	31	22	13	1.45 (0.75, 2.80)	0.03	1.51 (0.70, 3.27)	0.37	unknown
	CT	148	139	69	1.61 (1.12, 2.33)		1.31 (0.84, 2.03)		
	TT	201	133	77	1.00		1.00		
LRP1 (LRP1_3)	TT	40	21	11	0.89 (0.47, 1.69)	0.03	1.06 (0.48, 2.34)	0.03	0.99
	CT	172	157	85	1.57 (1.09, 2.26)		1.82 (1.16, 2.87)		
	CC	189	115	57	1.00		1.00		
MTHFR (MTHFR_1)	TT	46	47	31	1.69 (1.01, 2.83)	0.10	2.41 (1.33, 4.38)	0.01	N/A
	CT	172	115	65	0.97 (0.67, 1.40)		1.07 (0.68, 1.67)		
	CC	191	151	73	1.00		1.00		
SELP (SELP_3)	AA	8	14	11	3.82 (1.43, 10.2)	0.01	5.31 (1.87, 15.1)	0.005	−0.43
	AG	90	100	54	1.38 (0.93, 2.04)		1.40 (0.87, 2.25)		
	GG	270	207	108	1.00		1.00		
THPO (THPO_1)	GG	79	55	33	0.52 (0.31, 0.89)	0.01	0.64 (0.34, 1.22)	0.19	unknown
	AG	176	146	84	0.55 (0.36, 0.85)		0.62 (0.37, 1.05)		
	AA	75	87	42	1.00		1.00		

*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.

111 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 our approach as an effective means of hypothesis generation.

The current report represents a significant extension of our interim report, providing distinctive insight into the potential significance of additional vascular biology genes and particular single nucleotide polymorphisms. Using a database of single nucleotide polymorphisms identified through systematic screening of the coding region of candidate genes, we have identified novel genetic associations between single nucleotide polymorphisms in the endothelial converting enzyme (ECE1), histidine rich glycoprotein (HRG), phospholipase C, gamma 1 (PLCG1), syndecan (SDC4) and plasminogen activator inhibitor-2 (PAI2) genes and coronary heart disease or myocardial infarction. The validity and relevance of these associations to coronary heart disease or myocardial infarction requires further validation in other study populations.

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 4 Significant associations between single nucleotide polymorphisms and coronary heart disease or myocardial infarction uncovered in this population for the first time

| Gene (polymorphism)	Genotype	Random population	Coronary heart disease cases	Myocardial infarction cases	Coronary heart disease adjusted odds ratio (95% confidence interval)	p coronary heart disease	Myocardial infarction adjusted odds ratio (95% confidence interval)	p myocardial infarction
THBS4 (THBS4_1)	CC	19	20	11	1.77 (0.85, 3.68)	0.08	2.22 (0.93, 5.28)	0.002
	CG	121	126	79	1.42 (0.99, 2.03)		2.07 (1.35, 3.17)	
	GG	266	196	94	1.00		1.00	
THBS2 (THBS2_2)	GG	38	15	6	0.51 (0.25, 1.05)	0.03	0.38 (0.14, 1.01)	0.02
	GT	147	147	83	1.29 (0.90, 1.85)		1.40 (0.92, 2.15)	
	TT	199	155	80	1.00		1.00	
PAI2 (PAI2_1)	GG	15	15	9	1.56 (0.67, 3.65)	0.21	2.35 (0.90, 6.12)	0.03
	GC	146	139	85	1.32 (0.93, 1.88)		1.64 (1.08, 2.49)	
	CC	247	181	88	1.00		1.00	
PAI2 (PAI2_2)*	GG	17	15	9	1.34 (0.59, 3.06)	0.24	1.99 (0.78, 5.07)	0.03
	GC	141	135	83	1.34 (0.94, 1.91)		1.69 (1.11, 2.56)	
	CC	244	181	89	1.00		1.00	

PAI2_1 and PAI2_2 are in complete linkage disequilibrium (D′ = 1.00)

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[52] 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.[11] In addition, disease heterogeneity and linkage disequilibrium with nearby loci, as illustrated in our work must be considered. Recently, Colhoun et al[53] 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,[54] the genome wide
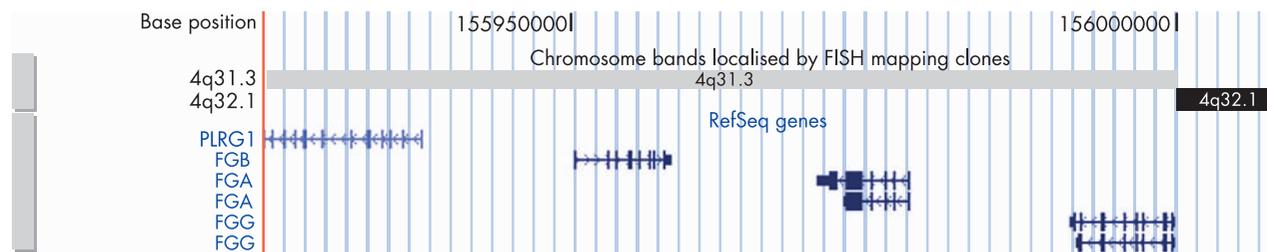
Figure 1 Relative position of fibrinogen genes clustered on chromosome 4q28 (http://genome.ucsc.edu/). FISH, fluorescence in situ hybridisation; PLRG1, pleiotropic regulator 1.
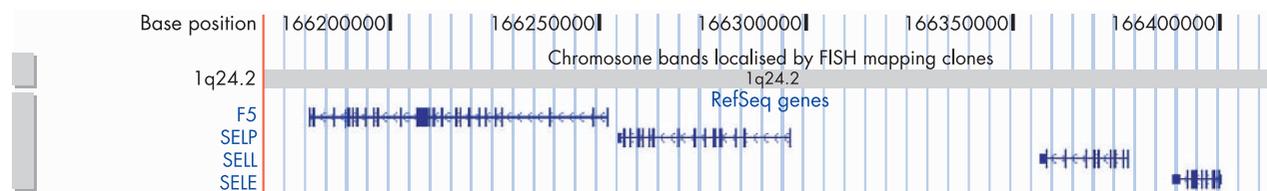
Figure 2 Relative position of selectin and factor V genes clustered on chromosome 1q24 (http://genome.ucsc.edu/). FISH, fluorescence in situ hybridisation.

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 reports 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, confounding and interaction may 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, San Diego State University, San Diego, CA 92182, USA

A Parker, K Glatt, Millennium Pharmaceuticals, Inc., Cambridge, MA 02139, USA

Q Wang, E F Plow, S Rao, G Q Shen, R Cannata, E J Topol, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

W J Rogers, University of Alabama, Birmingham, AL 35203, USA

L K Newby, Duke University, Durham, NC 27708, USA

D J Moliterno, University of Kentucky, Lexington, KY 40536, USA

Conflicts of interest: none declared.

*The names of investigators and their institutions are presented in reference 10.

REFERENCES

1 **National Cholesterol Education Program Expert Panel**. *Detection, evaluation and treatment of high blood cholesterol in adults (adult treatment panel III). Final report* 2002:II–19.
2 **Pohjola-Sintonen S**, Rissanen A, Liskola P, Luomanmaki K. Family history as a risk factor of coronary heart disease in patients under 60 years of age. *Eur Heart J* 1998;**19**:235–9.
3 **Rissanen AM**, Nikkila EA. Coronary artery disease and its risk factors in families of young men with angina pectoris and in controls. *Br Heart J* 1977;**39**:875–83.
4 **Pajukanta P**, Cargill M, Viitanen L, Nuotio I, Kareinen A, Perola M, Terwilliger JD, Kempas E, Daly M, Lilja H, Rioux JD, Brettin T, Viikari JS, Ronnemaa T, Laakso M, Lander ES, Peltonen L. Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland. *Am J Hum Genet* 2000;**67**:1481–93.
5 **Francke S**, Manraj M, Lacquemant C, Lecoeur C, Lepretre F, Passa P, Hebe A, Corset L, Yan SL, Lahmidi S, Jankee S, Gunness TK, Ramjuttun US, Balgobin V, Dina C, Froguel P. A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 2001;**10**:2751–65.
6 **Broeckel U**, Hengstenberg C, Mayer B, Holmer S, Martin LJ, Comuzzie AG, Blangero J, Nurnberg P, Reis A, Riegger GA, Jacob HJ, Schunkert H. A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 2002;**30**:210–4.
7 **Harrap SB**, Zammit KS, Wong ZY, Williams FM, Bahlo M, Tonkin AM, Anderson ST. Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler Thromb Vasc Biol* 2002;**22**:874–8.
8 **Ozaki K**, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002;**32**:650–4.
9 **Yamada Y**, Izawa H, Ichihara S, Takatsu F, Ishihara H, Hirayama H, Sone T, Tanaka M, Yokota M. Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. *N Engl J Med* 2002;**347**:1916–23.
10 **Topol EJ**, McCarthy J, Gabriel S, Moliterno DJ, Rogers WJ, Kristin Newby L, Freedman M, Metivier J, Cannata R, O'Donnell CJ, Kottke-Marchant K, Murugesan G, Plow EF, Stenina O, Daley GQ. Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation* 2001;**104**:2641–4.
11 **Lohmueller KE**, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177–82.
12 **Ioannidis JP**, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis, DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003;**361**:567–71.
13 **Cargill M**, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of

14 single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999;**22**:231–8.
14 **Livak KJ**. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 1999;**14**:143–9.
15 **Reich DE**, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;**20**:4–16.
16 **Lewontin R**. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964;**49**:49–67.
17 **Excoffier L**, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;**12**:921–7.
18 **Clee SM**, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, Brooks-Wilson A, Genest Jr J, Kastelein JJ, Hayden MR. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation* 2001;**103**:1198–205.
19 **Snapir A**, Heinonen P, Tuomainen TP, Alhopuro P, Karvonen MK, Lakka TA, Nyyssonen K, Salonen R, Kauhanen J, Valkonen VP, Pesonen U, Koulu M, Scheinin M, Salonen JT. An insertion/deletion polymorphism in the α_{2B}-adrenergic receptor gene is a novel genetic risk factor for acute coronary events. *J Am Coll Cardiol* 2001;**37**:1516–22.
20 **Wenzel K**, Blackburn A, Ernst M, Affeldt M, Hanke R, Baumann G, Felix SB, Kleber FX, Rohde K, Glaser C, Speer A. Relationship of polymorphisms in the renin-angiotensin system and in E-selectin of patients with early severe coronary heart disease. *J Mol Med* 1997;**75**:57–61.
21 **Xiang K**, Zheng T, Sun D, Li J. [The relationship between angiotensin II type 1 receptor gene and coronary heart disease, hypertension and diabetes mellitus.] *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 1998;**15**:9–12. (In Chinese.)
22 **Reguero JR**, Cubero GI, Batalla A, Alvarez V, Hevia S, Cortina A, Coto E. Apolipoprotein A1 gene polymorphisms and risk of early coronary disease. *Cardiology* 1998;**90**:231–5.
23 DNA polymorphisms of the apolipoprotein B gene in patients with premature coronary artery disease. *Atherosclerosis* 1990;**82**:7–17.
24 **Rees A**, Stocks J, Williams LG, Caplin JL, Jowett NI, Camm AJ, Galton DJ. DNA polymorphisms in the apolipoprotein C-III and insulin genes and atherosclerosis. *Atherosclerosis* 1985;**58**:269–75.
25 **Luc G**, Bard JM, Arveiler D, Evans A, Cambou JP, Bingham A, Amouyel P, Schaffer P, Ruidavets JB, Cambien F. Impact of apolipoprotein E polymorphism on lipoproteins and risk of myocardial infarction. The ECTIM Study. *Arterioscler Thromb Vasc Biol* 1994;**14**:1412–9.
26 **Unkelbach K**, Gardemann A, Kostrzewa M, Philipp M, Tillmanns H, Haberbosch W. A new promoter polymorphism in the gene of lipopolysaccharide receptor CD14 is associated with expired myocardial infarction in patients with low atherosclerotic risk profile. *Arterioscler Thromb Vasc Biol* 1999;**19**:932–8.
27 **Ordovas JM**, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler Thromb Vasc Biol* 2000;**20**:1323–9.
28 **Inoue N**, Kawashima S, Kanazawa K, Yamada S, Akita H, Yokoyama M. Polymorphism of the NADH/NADPH oxidase p22 phox gene in patients with coronary artery disease. *Circulation* 1998;**97**:135–7.
29 **Franco RF**, Reitsma PH. Genetic risk factors of venous thrombosis. *Hum Genet* 2001;**109**:369–84.
30 **Rosendaal FR**, Siscovick DS, Schwartz SM, Psaty BM, Raghunathan TE, Vos HL. A common prothrombin variant (20210 G to A) increases the risk of myocardial infarction in young women. *Blood* 1997;**90**:1747–50.
31 **Iacoviello L**, Di Castelnuovo A, De Knijff P, D'Orazio A, Amore C, Arboretti R, Kluft C, Benedetta Donati M. Polymorphisms in the coagulation factor VII gene and the risk of myocardial infarction. *N Engl J Med* 1998;**338**:79–85.
32 **Behague I**, Poirier O, Nicaud V, Evans A, Arveiler D, Luc G, Cambou JP, Scarabin PY, Bara L, Green F, Cambien F. β fibrinogen gene polymorphisms are associated with plasma fibrinogen and coronary artery disease in patients with myocardial infarction. The ECTIM Study. Etude Cas-Temoins sur l'Infarctus du Myocarde. *Circulation* 1996;**93**:440–9.
33 **Gonzalez-Conejero R**, Lozano ML, Rivera J, Corral J, Iniesta JA, Moraleda JM, Vicente V. Polymorphisms of platelet membrane glycoprotein Ib associated with arterial thrombotic disease. *Blood* 1998;**92**:2771–6.
34 **Waalen J**, Felitti V, Gelbart T, Ho NJ, Beutler E. Prevalence of coronary heart disease associated with HFE mutations in adults attending a health appraisal center. *Am J Med* 2002;**113**:472–9.
35 **Humphries SE**, Luong LA, Ogg MS, Hawe E, Miller GJ. The interleukin-6 -174 G/C promoter polymorphism is associated with risk of coronary heart disease and systolic blood pressure in healthy men. *Eur Heart J* 2001;**22**:2243–52.
36 **Carter AM**, Ossei-Gerning N, Wilson IJ, Grant PJ. Association of the platelet Pl^A polymorphism of glycoprotein IIb/IIIa and the fibrinogen Bβ 448 polymorphism with myocardial infarction and extent of coronary artery disease. *Circulation* 1997;**96**:1424–31.
37 **Peacock RE**, Hamsten A, Nilsson-Ehle P, Humphries SE. Associations between lipoprotein lipase gene polymorphisms and plasma correlations of lipids, lipoproteins and lipase activities in young myocardial infarction survivors and age-matched healthy individuals from Sweden. *Atherosclerosis* 1992;**97**:171–85.
38 **Schulz S**, Schagdarsurengin U, Greiser P, Birkenmeier G, Muller-Werdan U, Hagemann M, Riemann D, Werdan K, Glaser C. The LDL receptor-related protein (LRP1/A2MR) and coronary atherosclerosis—novel genomic variants and functional consequences. *Hum Mutat* 2002;**20**:404.
39 **Morita H**, Taguchi J, Kurihara H, Kitaoka M, Kaneda H, Kurihara Y, Maemura K, Shindo T, Minamino T, Ohno M, Yamaoki K, Ogasawara K, Aizawa T, Suzuki S, Yazaki Y. Genetic polymorphism of

5,10-methylenetetrahydrofolate reductase (MTHFR) as a risk factor for coronary artery disease. *Circulation* 1997;**95**:2032–6.

40 **Hingorani AD**, Liang CF, Fatibene J, Lyon A, Monteith S, Parsons A, Haydock S, Hopper RV, Stephens NG, O'Shaughnessy KM, Brown MJ. A common variant of the endothelial nitric oxide synthase (Glu298→Asp) is a major risk factor for coronary artery disease in the UK. *Circulation* 1999;**100**:1515–20.

41 **Eriksson P**, Kallin B, van't Hooft FM, Bavenholm P, Hamsten A. Allele-specific increase in basal transcription of the plasminogen-activator inhibitor 1 gene is associated with myocardial infarction. *Proc Natl Acad Sci U S A* 1995;**92**:1851–5.

42 **Sen-Banerjee S**, Siles X, Campos H. Tobacco smoking modifies association between Gln-Arg192 polymorphism of human paraoxonase gene and risk of myocardial infarction. *Arterioscler Thromb Vasc Biol* 2000;**20**:2120–6.

43 **Wang XL**, Oosterhof J, Duarte N. Peroxisome proliferator-activated receptor gamma C161→T polymorphism and coronary artery disease. *Cardiovasc Res* 1999;**44**:588–94.

44 **Hong SH**, Kim YR, Yoon YM, Min WK, Chun SI, Kim JQ. Association between HaeIII polymorphism of scavenger receptor class B type I gene and plasma HDL-cholesterol concentration. *Ann Clin Biochem* 2002;**39**:478–81.

45 **Wenzel K**, Felix S, Kleber FX, Brachold R, Menke T, Schattke S, Schulte KL, Glaser C, Rohde K, Baumann G. E-selectin polymorphism and atherosclerosis: an association study. *Hum Mol Genet* 1994;**3**:1935–7.

46 **Herrmann SM**, Ricard S, Nicaud V, Mallet C, Evans A, Ruidavets JB, Arveiler D, Luc G, Cambien F. The P-selectin gene is highly polymorphic: reduced frequency of the Pro715 allele carriers in patients with myocardial infarction. *Hum Mol Genet* 1998;**7**:1277–84.

47 **Francis SE**, Camp NJ, Dewberry RM, Gunn J, Syrris P, Carter ND, Jeffery S, Kaski JC, Cumberland DC, Duff GW, Crossman DC. Interleukin-1 receptor antagonist gene polymorphism and coronary artery disease. *Circulation* 1999;**99**:861–6.

48 **Boekholdt SM**, Trip MD, Peters RJ, Engelen M, Boer JM, Feskens EJ, Zwinderman AH, Kastelein JJ, Reitsma PH. Thrombospondin-2 polymorphism is associated with a reduced risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* 2002;**22**:e24–7.

49 **Wessel J**, Topol EJ, Ji M, *et al*. Replication of the association between the thrombospondin-4 A387P polymorphism and myocardial infarction. *Am Heart J* 2004 *(in press)*.

50 **Walton B**, Wilerson J, J C, *et al*. Thrombospondin-2 and Lymphotoxin-a gene variations predict coronary heart disease in a large prospective study. *Circulation* 2003;**108**(suppl IV):771 (abstract).

51 **Stenina OI**, Desai SY, Krukovets I, Kight K, Janigro D, Topol EJ, Plow EF. Thrombospondin-4 and its variants: expression and differential effects on endothelial cells. *Circulation* 2003;**108**:1514–9.

52 **Hirschhorn JN**, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;**4**:45–61.

53 **Colhoun HM**, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;**361**:865–72.

54 **Wang L**, Fan C, Topol SE, Topol EJ, Wang Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science* 2003;**302**:1578–81.

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

Please visit the *Journal of Medical Genetics* website [www.jmedgenet.com] for a link to the full text of this article.

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.

▲ *Gut* 2004;**53**:381–386.

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
A2M	α-2-macroglobulin	A2M_1	mis (V/I)	0.32	CTCCAGAG[G/A]TCAAGTCC
ABCA1	ATP-binding cassette, subfamily A, member 1	ABCA1_1	mis (R/K)	0.26	TTTGTGGCCTACCAA[G/A]GGAGAAACTGGCTGC
ACADM	acyl coenzyme A dehydrogenase, C4-C12 straight chain; MCAD	ACADM_1	silent	0.26	TACAGAATATCCTGT[A/G]GAAAAACTAATGAGG
ACAT2	acetyl coenzyme A acetyltransferase 2	ACAT2_1	mis (K/R)	0.41	TGGTGTCAACTAGAA[A/G]AGGTGAGTATATCAT
ACAT2	acetyl coenzyme A acetyltransferase 2	ACAT2_2	silent	0.18	TGGTGCTGCAGCTGT[T/C]GTTCTTATGAAGAAG
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	DCP1_1	silent	0.42	GAATGTGATGGCCAC[A/G]TCCCGGAAATATGAA
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	DCP1_2	mis (P/L)	0.43	CGATGCTGGAGAAGC[C/T]GGCCGACGGGCGGGA
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	DCP1_3	silent	0.38	GTACAAGGATCTGCC[C/T]GTCTCCCTGCGTCGG
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	DCP1_4	silent	0.43	CTGCTCCAGGTACTT[T/C]GTCAGCTTCATCATC
ADRA2B	adrenergic, α-2B-, receptor	ADRA2B_1	I/D	0.41	GCATCTCCAGAGGATGAAGCT[-/GAAGAGGAG]
ADRB2	adrenergic receptor, β 2	ADRB2_1	mis (E/Q)	0.44	CACGACGTCACGCAG[G/C]AAAGGGACGAGGTGT
AGT	angiotensinogen	AGT_1	mis (M/T)	0.42	CTGGCTGCTCCCTGA[T/C]GGGAGCCAGTGTGGA
AGTR1	angiotensin receptor 1	AGTR1_1	3' UTR	0.30	ACTACCAAATGAGC[A/C]TTAGCTACTTTCAGA
ANXA4	annexin IV; lipocortin IV	ANXA4_1	mis (T/M)	0.16	TTGTGGGGATGATGA[C/T]GCCCACGGTGCTGTA
APOA1	apolipoprotein AI	APOA1_1	promoter	0.17	AGCAGCAACAGGGCC[G/A]GGGCTGGGCTTATCA
APOB	apolipoprotein B	APOB_1	mis (R/Q)	0.09	GGAAAAATGAAGTCC[G/A]GATTCATTCTGGGTC
APOB	apolipoprotein B	APOB_2	silent	0.50	CAAATTCCGAGAGAC[C/T]CTAGAAGATACACGA
APOC3	apolipoprotein CIII	APOC3_1	3' UTR	0.08	GCCTATCCATCCTGC[C/G]AGCTCCTTGGGTCCT
APOE	apolipoprotein E	APOE_1	mis (C/R)	0.13	GACATGGAGGACGTG[T/C]GCGGCCGCCT GGTGC
BDKRB1	bradykinin receptor B1	BDKRB1_1	silent	0.16	GCACAGAGTGCTGCC[G/A]ACATTTATCATCTCC
BDKRB2	bradykinin receptor B2	BDKRB2_1	silent	0.16	CAAGGAGATCCAGAC[G/A]GAGAGGAGGGCCACG
CCKBR	cholecystokinin B receptor	CCKBR_1	mis (V/I)	0.05	TTCATCTTTGGCACC[G/A]TCATCTGCAAGGCGG
CCL2	chemokine (C-C motif) ligand 2	SCYA2_1	silent	0.39	CCCAGTCACCTGCTG[T/C]TATAACTTCACCAAT
CD14	CD14 antigen	CD14_1	promoter	0.49	TCCTTCCTGTTACGG[C/T]CCCCCTCCCTGAAAC
CD36	CD36 antigen; thrombospondin receptor	CD36_2	promoter	0.42	GTGACTCATCAGTTC[A/C]TTCCTGTAAAATTC
CD36	CD36 antigen; thrombospondin receptor	CD36_3	intronic	0.42	ATAATTTTTTTGCT[A/G]TCATAATCGCCTCAT
CD36	CD36 antigen; thrombospondin receptor	CD36_4	silent	0.01	TTTGAGTTTGGTTCC[G/T]TACCCTGTTACTACC
CETP	cholesterol ester transfer protein	CETP_1	mis (A/P)	0.07	GTGACTACCGTCCAG[G/C]CCTCCTATTCTAAGA

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
CETP	cholesterol ester transfer protein	CETP_2	mis (I/V)	0.33	AGCAGCTCCGAGTCC[A/G]TCCAGAGCTTCCTGC
CETP	cholesterol ester transfer protein	CETP_3	silent	0.06	CTCTGAGCGAGTCTT[C/T]CACTCGCTGGCCAAG
CETP	cholesterol ester transfer protein	CETP_4	intronic	0.43	GAATCACTGGGGGTTC[G/A]AGTTAGGGTTCAGAT
COL5A2	collagen, type V, alpha 2	COL5A2_1	silent	0.18	CCCAACGGGCTCTCC[G/A]GGTACCTCTGGTCCT
COMP	cartilage oligomeric matrix protein (TSP-5)	COMP_1	mis (N/D)	0.05	CGCAACCAGGCCGAC[A/G]ACTGCCCTAGGGTAC
COMP	cartilage oligomeric matrix protein (TSP-5)	COMP_2	3' UTR	0.16	GCACACAGGCCCTGC[C/T]GGCCTTCTCAGCCCT
CYBA	p22 phox	CYBA_1	mis (H/Y)	0.36	CCCAGGGGACAGAAG[T/C]ACATGACCGCCGTGG
ECE1	endothelin converting enzyme 1	ECE1_1	silent	0.28	TCTCATCAACACCAC[C/T]GACAGATGCCTGCTC
ECE1	endothelin converting enzyme 1	ECE1_2	silent	0.06	GAGGGTCACTGCCGA[T/C]CAGCTCAGGAAAGCC
EDN1	endothelin 1	EDN1_1	intronic	0.23	TTAAAGACTATTAAT[T/C]ACACTAATATAGTTT
EDN1	endothelin 1	EDN1_2	mis (K/N)	0.20	CAAGCTGAAAGGCAA[G/T]CCCTCCAGAGAGCGT
EDN1	endothelin 1	EDN1_3	silent	0.14	GGCAACAGACCGTGA[G/A]AATAGATGCCAATGT
EDNRA	endothelin A receptor	EDNRA_1	intronic	0.39	ACAAGGCAACTGTGA[C/G]TCCGGGAATCTCTTC
EDNRB	endothelin B receptor	EDNRB_1	silent	0.40	AAAAGATTGGTGGCT[G/A]TTCAGTTTCTATTTC
ESR1	estrogen receptor 1	ESR1_1	silent	0.18	GGGTTTCCCTGCCAC[A/G]GTCTGAGAGCTCCCT
F10	coagulation factor X	F10_1	silent	0.13	TTTCTGTGGTGGAAC[C/T]ATTTGAGCGAGT TC
F11	coagulation factor XI	F11_1	silent	0.12	TGCGTCTGTTCGTGG[C/T]GAGTGGCCGTGGCAG
F11	coagulation factor XI	F11_2	silent	0.03	AAGATGCACGGATGA[C/T]GTCCACTGCCACTTT
F11	coagulation factor XI	F11_3	silent	0.13	TGGATTGCCCAGTAC[G/A]CGCATTAAAAAGAGC
F13A1	coagulation factor XIII A-1	F13A1_1	mis (Q/E)	0.21	ATGACTGTGACAGTT[C/G]AGTTTACCAATCCTT
F13A1	coagulation factor XIII A-1	F13A1_2	mis (L/P)	0.21	TCTACACCGGGGTCC[T/C]GAAGGCAGAATTCAA
F13A1	coagulation factor XIII A-1	F13A1_3	intronic	0.20	GGGCAGGCAGTGCAG[A/G]GCTGTGGGAAGCTTT
F13A1	coagulation factor XIII A-1	F13A1_4	mis (M/K)	0.24	ATGCCAATTTGCAAA[A/T]GGACATCTT CCTGGA
F2	coagulation factor II; prothrombin	F2_1	mis (T/M)	0.12	CCGACAGCAGCACCA[C/T]GGGACCCTGGTGCTA
F2	coagulation factor II; prothrombin	F2_2	3' UTR	0.01	AAAGTGACTCTCAGC[G/A]AGCCTCAATGCTCCC
F2	coagulation factor II; prothrombin	F2_3	intronic	0.06	CTGCCTCCTGTACCC[G/A]CCCTGGGAC AAGAAC
F5	coagulation factor V	F5_1	mis (R/Q)	0.02	GATCCCTGGACAGGC[G/A]AGGAATACAGGTATT
F5	coagulation factor V	F5_2	mis (N/T)	0.04	ACCTCATTGGCAAGA[A/C]CTCAGTTCT CAATTC
F5	coagulation factor V	F5_3	intronic	0.26	AGAATGAGGCCCTGG[A/G]AGGACCCTC CTAGTG
F5	coagulation factor V	F5_4	mis (V/M)	0.31	GACAGCAACATGCCT[G/A]TGGACATGAGAGAAT
F7	coagulation factor VII	F7_1	silent	0.14	CTCCTGTCGGTGCCA[C/T]GAGGGGTACTCTCTG
F7	coagulation factor VII	F7_2	mis (R/Q)	0.11	ATGCCACCCACTACC[A/G]GGGCACGTG GTACCT
F7	coagulation factor VII	F7_3	intronic	0.11	GCTGCAGGTGCGTCC[G/A]GGGAGGTTTTCTCCA

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
F9	coagulation factor IX	F9_1	mis (A/T)	0.30	CTCACCCGTGCTGAG[G/A]CTGTTTTTCCTGATG
FABP3	fatty acid binding protein 3	FABP3_1	mis (K/R)	0.04	ACATTCTCACCCTAA[A/G]AACACACAGCACCTT
FASN	fatty acid synthase	FASN_1	mis (V/I)	0.03	AGCAGCACCTCCCAC[G/A]TCCCGGAGGTGGACC
FGA	fibrinogen, α	FGA_1	mis (T/A)	0.24	TCTGGACCTGGAAGT[A/G]CTGGAAGCTGGAACT
FGB	fibrinogen, β	FGB_1	silent	0.21	TGAGACTGTGAATAG[C/T]AATATCCCAACTAAC
FGB	fibrinogen, β	FGB_2	mis (R/K)	0.21	CATGGTACTCAATGA[G/A]GAAGATGAGTATGAA
FGB	fibrinogen, β	FGB_3	promoter	0.22	TGACAAGTAAATAAG[C/T]TTTGCTGGGAAGATG
FGB	fibrinogen, β	FGB_4	promoter	0.24	CTATTGATTTTAATG[A/G]CCCCTTTTGAAATAG
FGG	fibrinogen, γ	FGG_1	mis (G/R)	0.04	GGAGCTAAACAGAGC[G/A]GGCTTTACTTTATTA
GBE1	glucan (1,4-α)-branching enzyme 1 (glycogen branching enzyme)	GBE1_1	mis (T/A)	0.03	AACATGAGTGTCCTG[A/G]CTCCTTTTACTCCAG
GP1BA	glycoprotein 1B (platelet), α	GP1BA_1	mis (T/M)	0.06	CCCCAGGGCTCCTGA[C/T]GCCCACACCCAAGCT
HFE	haemochromatosis	HFE_1	mis (C/Y)	0.06	ATATACGT[G/A]CCAGGTGG
HRG	histidine-rich glycoprotein	HRG_1	mis (H/R)	0.20	ATGGGGCCCAAAGAC[A/G]TTCTCATAATAATAA
HRG	histidine-rich glycoprotein	HRG_2	mis (N/I)	0.39	CTCTAAAGCCAGACA[A/T]TCAGCCCTTTCCTCA
HSD17B4	hydroxysteroid (17 β) dehydrogenase 4	HSD17B4_1	mis (W/R)	0.10	CGCCTCAGTGGAGAC[T/C]GGAATCCCTTACACA
IL1A	interleukin 1, αa	IL1A_1	mis (A/S)	0.29	ATCAAGCCTAGGTCA[G/T]CACCTTTTAGCTTCC
IL1B	interleukin 1, β	IL1B_1	silent	0.21	AGAACCTATCTTCTT[C/T]GACACATGG GATAAC
IL1R1	interleukin 1 receptor, type 1	IL1R1_1	mis (A/G)	0.03	ACTTATGTTATAATG[C/G]ACAAGCCATATTTAA
IL1R2	interleukin 1 receptor, type II	IL1R2_1	intronic	0.02	CTCTGTCAGCCCCCG[C/A]ATCAACCTGACATGG
IL1RN	interleukin 1 receptor antagonist	IL1RN_1	mis (S/G)	0.15	AGTGGCCCCACCACC[A/G]GTTTTGAGTCTGCCG
IL1RN	interleukin 1 receptor antagonist	IL1RN_2	silent	0.28	CATCCGCTCAGACAG[T/C]GGCCCCACCACCAGT
IL1RN	interleukin 1 receptor antagonist	IL1RN_3	intronic	0.28	CAACCAACTAGTTGC[T/C]GGATACTTGCAAGGA
IL6	interleukin 6	IL6_1	promoter	0.41	CTAGTTGTGTCTTGC[C/G]ATGCTAAAGGACGTC
INS	insulin	INS_1	intronic	0.29	CTCAGCCCTGCCTGTC[T/A]
IRS2	insulin receptor substrate 2	IRS2_1	intronic	0.13	GTCACATGTCAGAGA[A/G]TATCATTAAAAGAAG
ITGA2B	integrin, α 2B	ITGA2B_1	silent	0.36	CGTGTACCCCCAGGT[C/T]GGCTTCTTCAAGCGG
ITGA2B	integrin, α 2B	ITGA2B_2	mis (I/S)	0.36	ACTGGGGGCTGCCCA[T/G]CCCCAGCCCCTCCCC
ITGB1	integrin, β 1	ITGB1_1	silent	0.09	TATGGACCTGTCTTA[T/C]TCAATGAAAGACGAT
ITGB1	integrin, β 1	ITGB1_2	silent	0.08	AGGAACATTATCTGC[A/C]AATTCTAGCAATGTA
ITGB2	integrin, β 2; CD18	ITGB2_1	silent	0.24	TTTCGCGGGCGACGG[A/G]AAGCTGGGCGCCATC
ITGB2	integrin, β 2; CD18	ITGB2_2	silent	0.24	ACTCTCCTCCAGGGT[C/A]TTCCTGGATCACAAC

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
ITGB3	integrin, β 3	ITGB3_1	mis (L/P)	0.16	TACAGGCCCTGCCTC[T/C]GGGCTCACCTCGCTG
ITGB3	integrin, β 3	ITGB3_2	silent	0.29	TTCCCAGCAGGACGA[G/A]TGCAGCCCCCGAGAG
ITGB3	integrin, β 3	ITGB3_3	silent	0.08	AGGCATTGTCCAGCC[T/C]AATGACGGGCAGTGT
ITGB3	integrin, β 3	ITGB3_4	silent	0.30	CGAGTGCAGCCCCCG[A/G]GAGGGTCAGCCCGTC
ITPKB	inositol 1,4,5-trisphosphate 3-kinase B	ITPKB_1	silent	0.28	GGACCCCAACTCAGC[T/C]TTCCTGCATACCCTG
LDLR	low density lipoprotein receptor	LDLR_1	silent	0.45	AAGCATCGATGTCAA[T/C]GGGGGCAACCGGAAG
LDLR	low density lipoprotein receptor	LDLR_2	silent	0.46	CCCAGAGGATATGGT[T/C]CTCTTCCACAACCTC
LDLR	low density lipoprotein receptor	LDLR_3	silent	0.11	AGTGGGCGACAGATG[C/T]GAAAGAAACGAGTTC
LDLR	low density lipoprotein receptor	LDLR_4	silent	0.26	GCACACAACCACCCG[G/A]CCTGTTCCCGACACC
LIPA	lysosomal acid lipase A; acid cholesteryl ester hydrolase	LIPA_1	mis (P/T)	0.32	TGTTTGGTTCTCTGG[C/A]CCCTGCATTCTGAGG
LIPA	lysosomal acid lipase A; acid cholesteryl ester hydrolase	LIPA_2	mis (G/R)	0.11	AGGGGTCT[G/A]GAGGGAAA
LIPC	hepatic lipase	LIPC_1	mis (V/M)	0.04	CCTCAGGTGGACGGC[G/A]TGCTAGAAAACTGGA
LIPC	hepatic lipase	LIPC_2	silent	0.41	CTTCCAGGAATCTGT[T/G]CAACTCTCTCGAAGC
LIPC	hepatic lipase	LIPC_3	mis (N/S)	0.37	CTCCAGATGATGCCA[A/G]TTTTGTGGATGCCAT
LIPC	hepatic lipase	LIPC_4	silent	0.49	CATTCATACCTTTAC[C/G]CGGGAGCACATGGGC
LIPC	hepatic lipase	LIPC_5	promoter	0.22	ACACAGTAGCTTTAA[G/A]TTGATTAATTTGGAA
LPL	lipoprotein lipase	LPL_1	intronic	0.29	TATAGGATTTAAAGC[T/G]TTTATACTAA ATGTG
LRP1	low density lipoprotein-related protein 1 (α 2 macroglobulin receptor)	LRP1_2	silent	0.27	TGAGTACCCGACCTG[C/T]GGCCCCAGTGAGTTC
LRP1	low density lipoprotein-related protein 1 (α 2 macroglobulin receptor)	LRP1_3	silent	0.31	CTTCCGGCTGAAGGA[C/T]GACGGCCGGACGTGT
LRP1	low density lipoprotein-related protein 1 (α 2 macroglobulin receptor)	LRP1_4	silent	0.31	CCCCTTCGGGCTGGC[C/T]GTGTATGGGGAGCAC
LRP1	low density lipoprotein-related protein 1 (α 2 macroglobulin receptor)	LRP1_5	silent	0.31	TGAGGGCGAGCTCTG[C/T]GGTGAGGCCTGGTCC
LRP2	lipoprotein-related protein 2; glycoprotein 330 megalin low density	LRP2_1	mis (A/T)	0.28	TCCAATGC[G/A]CATCTGGG
LRP2	lipoprotein-related protein 2; glycoprotein 330 megalin low density	LRP2_2	mis (N/S)	0.31	GACAATGCATCCCCA[A/G]CTCCTGGGTGTGTGA
LRPAP1	low density lipoprotein-related protein-associated protein 1; RAP; MRAP	LRPAP1_1	silent	0.10	CTACAGCACTGAGGC[T/C]GGTGAGTGCCCGCCC
LRPAP1	low density lipoprotein-related protein-associated protein 1; RAP; MRAP	LRPAP1_2	3' non-coding	0.27	CGAACTCTGAAGGCA[C/T]TGGGGAGCCCAGCCC

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
MMP9	matrix metalloproteinase 9; gelatinase B	MMP9_1	mis (R/Q)	0.33	CAGGACTCTACACCC[G/A]GGACGGCAATGCTGA
MMP9	matrix metalloproteinase 9; gelatinase B	MMP9_2	mis (Q/R)	0.14	ACGTCTTCCAGTACC[A/G]AGGTGAGGGCTGAGG
MMP9	matrix metalloproteinase 9; gelatinase B	MMP9_3	mis (A/V)	0.02	GCTGCTGCTTTGCTG[C/T]CCCCAGACAGCGCCA
MOG	myelin oligodendrocyte glycoprotein	MOG_1	mis (L/V)	0.23	AGATCACT[C/G]TTGGCCTC
MTHFR	methylenetetrahydrofolate reductase	MTHFR_1	mis (A/V)	0.32	AGGTGTCTGCGGGAG[T/C]CGATTTCATCATCAC
NOS2A	nitric oxide synthase 2 (inducible, hepatocytes)	NOS2A_1	silent	0.26	CCGGGACTTCTGTGA[C/T]GTCCAGCGCTACAAC
NOS2A	nitric oxide synthase 2 (inducible, hepatocytes)	NOS2A_2	silent	0.20	GGCCCTGGTCCAAGG[C/T]ATCCTGGAGCGAGTG
NOS2A	nitric oxide synthase 2 (inducible, hepatocytes)	NOS2A_3	mis (S/L)	0.20	AGAAACTGAAGAAAT[C/T]GCTCTTCATGCTGAA
NOS3	nitric oxide synthase 3 (endothelial); ENOS	NOS3_1	mis (D/E)	0.31	GCAGGCCCCAGATGA[G/T]CCCCCAGAACTCTTC
NOS3	nitric oxide synthase 3 (endothelial); ENOS	NOS3_2	I/D	0.17	GGCTGGAGGAGGGGAAA[GAAGTCTAGACCTGCTGCA GGGG TGAG/-]
P2RY1	purinergic receptor P2Y, G-protein coupled, 1 (ADP receptor)	P2RY1_1	silent	0.03	CGCTGCCTTCCTGGC[C/T]GGTCCGGGTTCGTCC
P2RY1	purinergic receptor P2Y, G-protein coupled, 1 (ADP receptor)	P2RY1_2	silent	0.15	ATCGATTTACCTGGT[A/G]ATCATTGTACTGACT
PAI1	serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	PAI1_1	I/D	0.47	TCTGGACACGTGGGG[G/del]AGTCAGCCGTGTATC
PAI2	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	PAI2_1	mis (S/C)	0.22	TTTTCGGCAGATTTT[C/G]CTCACCCTAAAACTA
PAI2	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	PAI2_2	mis (N/K)	0.22	GCATAAGATAACCAA[C/G]TGCATTTTATTTTTC
PAI2	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	PAI2_3	mis (N/D)	0.22	AATGCATCCACAGGG[A/G]ATTATTTACTGGAAA
PAI2	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	PAI2_4	intronic	0.21	TGTTTTTTTCTTCCT[C/G]TCTTTGCTTCTAGAT
PCSK1	proprotein convertase subtilisin/kexin type 1	PCSK1_1	intronic	0.19	GAGCCAGGAGTGGTC[T/C]AGAGCCCGAGGGTGG
PDGFA	platelet derived growth factor, A chain	PDGFA_1	silent	0.24	AGCTCACGGGGTCCA[C/T]GCCACTAAGCATGTG
PDGFRA	platelet derived growth factor receptor, α	PDGFRA_1	silent	0.13	AGTGAGCTGGCAGTA[C/T]CCCATGTCTGAAGAA
PECAM1	platelet/endothelial cell adhesion molecule (CD31 antigen); ELAM1	PECAM1_1	mis (N/S)	0.35	CCAAGCAGAAGGCTA[A/G]CAAGGAACAGGAGGG
PECAM1	platelet/endothelial cell adhesion molecule (CD31 antigen); ELAM1	PECAM1_2	intronic	0.46	CCGAGAAGAACAGAT[G/A]ATCCCTGTATTTCAA
PECAM1	platelet/endothelial cell adhesion molecule (CD31	PECAM1_3	silent	0.49	ACTGCAGAGTACCAG[C/G]TGTTGGTGGAAGGAG

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
	antigen); ELAM1				
PIK3R1	phosphatidylinositol 3-kinase, regulatory, 1	PIK3R1_1	mis (M/I)	0.17	TATGAATAACAATAT[G/A]TCCTTACAAAATGCT
PLAU	plasminogen activator, urokinase	PLAU_1	mis (P/L)	0.22	AGGTGGGCCTAAAGC[C/T]GCTTGTCCAAGAGTG
PLAU	plasminogen activator, urokinase	PLAU_2	silent	0.10	GCTTGCTCACCACAA[C/T]GACATTGCCTTGCTG
PLCG1	phospholipase C, gamma 1	PLCG1_1	mis (I/T)	0.46	ACGAGCTGACCTTCA[T/C]CAAGAGCGCCATCAT
PLCG1	phospholipase C, gamma 1	PLCG1_2	mis (S/G)	0.17	CAGGAGTTCATGCTC[A/G]GCTTCCTCCGAGACC
PLOD2	lysine hydrolase 2	PLOD2_1	intronic	0.51	TGAAATTTATTTTGG[G/A]TGAATGACTGGCATG
PON1	paraoxonase 1, serum, alt transcript 1	PON1_1	mis (M/L)	0.29	CTGGCTCTGAAGAC[A/T]TGGAGATACTGCCTAA
PON2	paraoxonase 2, serum	PON2_1	silent	0.22	TGAAGAAG[C/G]AGAAAATT
PPARG	peroxisome proliferative activated receptor, gamma	PPARG_1	mis (P/A)	0.12	GATTCTCCTATTGAC[G/C]CAGAAAGCGATTCCT
PPARG	peroxisome proliferative activated receptor, gamma	PPARG_2	silent	0.14	GATTGTCACGGAACA[C/T]GTGCAGCTACTGCAG
PRCP	prolylcarboxypeptidase (angiotensinase C)	PRCP_1	mis (T/S)	0.03	AGGATATCACAGACA[C/G]TCTGGTTGCAGTCAC
PRCP	prolylcarboxypeptidase (angiotensinase C)	PRCP_2	mis (E/D)	0.18	GGATGTGGCTGAGGA[A/C]CTGAAAGCTATGTTG
PROC	protein C	PROC_1	silent	0.29	CTTCCTCAATTGCTC[T/G]CTGGACAACGGCGGC
PROC	protein C	PROC_2	silent	0.36	GGCCCACTGCATGGA[T/C]GAGTCCAAGAAGCTC
PROS1	protein S (α)	PROS1_1	silent	0.39	AGCTCACTCATGTCC[A/G]TCAGTTTGG AAAAAG
PTGS2	cyclooxygenase (COX2)	PTGS2_1	silent	0.17	AATTATGAGTTATGT[C/G]TTGACATCCAGATCA
SAA1	serum amyloid A1, β	SAA1_1	mis (H/R)	0.12	AGAGATTCTTTGGCC[A/G]TGGTGCGGAGGACTC
SAA1	serum amyloid A1, β	SAA1_2	mis (K/R)	0.01	GGGGCAGGAGTGGCA[A/G]AGACCCCAATCACTT
SCARB1	scavenger receptor class B, member 1	SRB1_2	silent	0.49	TCACTTCCTCAACGC[C/T]GACCCGGTT CTGGCA
SCARB1	scavenger receptor class B, member 1	SRB1_3	silent	0.01	CCCCACCTATCGCTT[C/T]GTGGCTCCCAAAACC
SCARB1	scavenger receptor class B, member 1	SRB1_4	mis (V/I)	0.01	GAGAGCGACTACATC[G/A]TCATGCCCAACATCC
SCARB1	scavenger receptor class B, member 1	SRB1_5	intronic	0.05	AGC[C/G]TGCGGCCCCAGCTC
SCARB1	scavenger receptor class B, member 1	SRB1_6	mis (G/S)	0.10	AGGCGCGCAGACATG[G/A]GCTGCTCCGCCAAAG
SCARB1	scavenger receptor class B, member 1	SRB1_7	intronic	0.10	AGGGGAGGGTGGGCC[C/T]GGCCATGGCTGCTCG
SDC4	syndecan 4 (amphiglycan, ryudocan	SDC4_1	mis (F/L)	0.38	TTCGCGCTGCTGCTG[T/C]TCTTCGTAGGCGGAG
SELE	E-selectin	SELE_1	mis (S/R)	0.09	CCTAAACCTTTGGGT[G/T]AAAAGAACTCTTGAA
SELE	E-selectin	SELE_2	mis (S/R)	0.08	ACCAATACATCCTGC[A/C]GTGGCCACGGTGAAT
SELL	L-selectin	SELL_1	mis (P/S)	0.12	ATGGACTGTACTCAC[C/T]CTTTGGGAAACTTCA
SELL	L-selectin	SELL_2	mis (F/L)	0.27	GGGCCCCAGTGTCAG[T/C]TTGGTAAGTCTCTTT
SELL	L-selectin	SELL_3	silent	0.20	AACCAACAAATCTCT[C/T]ACTGAAGAA GCAGAG
SELP	P-selectin	SELP_1	mis (T/P)	0.11	AATGGCCACTGGTCA[A/C]CTACCGTGCCAACCT
SELP	P-selectin	SELP_2	mis (V/L)	0.10	CTTCCTACTCCAGGG[G/T]TGCAATGTCCAGCCC

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
SELP	P-selectin	SELP_3	mis (S/N)	0.14	ACCTGGAAGCCCCCA[G/A]TGAAGGAACCATGGA
SELP	P-selectin	SELP_4	mis (N/D)	0.06	TGTCATTTCTCTTGT[A/G]ACAATGGCTTTAAGC
SELPLG	P-selectin ligand	SELPLG_1	mis (M/I)	0.06	CCAGAAAT[G/A]CTGAGGA
SERPINA5	protein C inhibitor; PCI, PLANH3; PAI3	SERPINA5_1	mis (A/V)	0.10	TTGACCTCTACAGGG[C/T]CTTGGCTTCCGCTGC
SERPINA5	protein C inhibitor; PCI, PLANH3; PAI3	SERPINA5_2	mis (K/E)	0.09	CAGAAAAGCTCAGAG[A/G]AGGAGCTGCACAGAG
SERPINC1	serine (or cysteine) proteinase inhibitor, clade C (antithrombin), member 1	AT3_1	silent	0.36	GAGCCTGGCCAAGGT[G/A]GAGAAGGAACTCACC
SERPIND1	serine (or cysteine) proteinase inhibitor, clade D (heparin cofactor), member 1	HCF2_1	silent	0.45	CCTCATCTACGAGCA[C/T]CGCACCAGCTGCCTG
SREBF2	sterol regulatory element binding protein 2	SREBF2_1	mis (R/S)	0.04	CCCCCCACTCTCCAG[G/C]AGCTCCGTGCTCAAG
SREBF2	sterol regulatory element binding protein 2	SREBF2_2	intronic	0.07	TGAGAGTGGTGGGGA[A/G]GAGCCTTGT CTTCTT
SREBF2	sterol regulatory element binding protein 2	SREBF2_3	mis (A/G)	0.29	TTGCAGCTGCTGCCG[C/G]CAACCTACAAACCTG
TBXA2R	thromboxane A2 receptor	TBXA2R_1	silent	0.36	GGACCCCTGGGTGTA[T/C]ATCCTGTTCCGCCGC
TGFBI	TGFBi	TGFBI_1	silent	0.27	TTCTGTGCAGAGCCT[C/T]TGCATTGAGAACAGC
TGFBI	TGFBi	TGFBI_2	silent	0.48	GAACTGTGCCCGGCT[C/G]CTGAAAGCCGACCAC
THBD	thrombomodulin	THBD_1	intronic	0.37	TGAGATGTAAAAGGT[A/G]TTAAATTGATGTTGC
THBD	thrombomodulin	THBD_2	mis (A/V)	0.19	CCGACTCGGCCCTTG[C/T]CCGCCACATTGGCAC
THBS1	thrombospondin I	THBS1_1	mis (N/S)	0.10	CTGGCCCA[A/G]TGAGAACC
THBS1	thrombospondin I	THBS1_2	silent	0.10	CAGCCCCCAGATGAA[T/C]GGGAAACCCTGTGAA
THBS1	thrombospondin I	THBS1_3	intronic	0.15	ATGGGCTTGAGAAAA[C/T]CCCCAGGATCACTTC
THBS1	thrombospondin I	THBS1_5	silent	0.15	CATCAGTGAGACCGA[T/C]TTCCGCCGATTCCAG
THBS1	thrombospondin I	THBS1_6	silent	0.11	TCAGGAGTGTGACAA[A/G]AGATGTAAGCATCTT
THBS1	thrombospondin I	THBS1_7	intronic	0.11	CCTCTGCACCCAGCC[C/T]GTTAGCATGAACCCT
THBS1	thrombospondin I	THBS1_8	silent	0.03	TCGTCTCTGCAACAA[C/T]CCCNCACCCCAGTTT
THBS1	thrombospondin I	THBS1_9	mis (A/T)	0.12	CTCTGCAACAACCCC[G/A]CACCCCAGTTTGGAG
THBS2	thrombospondin II	THBS2_1	3' UTR	0.13	TGCAAATGGGTGTGA[T/C]GCGGTTCCAGATGTG
THBS2	thrombospondin II	THBS2_2	3' UTR	0.29	AAATATCAATGGAAC[T/G]CAGAGATGAACATCT
THBS2	thrombospondin II	THBS2_3	intronic	0.01	CTTCATTAAATGTAG[C/T]GACTGTCAACAGCGT
THBS2	thrombospondin II	THBS2_4	intronic	0.13	ATGCCAAACAGGAAC[G/A]ATCCATAACTTTAGT
THBS4	thrombospondin IV	THBS4_1	mis (A/P)	0.20	GAGTGTCGAAATGGA[G/C]CGTGCGTTCCCAACT
THBS4	thrombospondin IV	THBS4_3	mis (G/S)	0.02	GCTTACCAGACCGTG[G/A]GCCTGGATCCTGAAG
THBS4	thrombospondin IV	THBS4_4	intronic	0.50	AAGGCTGTCAGTCAG[C/T]GGGTGCTCGGATCTC
THPO	thrombopoietin	THPO_1	intronic	0.51	CATCAGCATTGTCTC[A/G]TGTACAGCTCCCTTC

Gene symbol	Gene name	SNP	Mutation type	Variant frequency	Flanking sequence
TIMP1	tissue inhibitor of metalloproteinase 1	TIMP1_1	silent	0.48	CACTACCTGCAGTTT[T/C]GTGGCTCCCTGGAAC
VWF	von Willebrand factor	VWF_1	silent	0.11	CCCCTTCAGCGAGGC[A/C]CAGTCCAAAGGGGAC
VWF	von Willebrand factor	VWF_2	silent	0.12	GTACTTGGCCTCAAC[C/T]GCCACCAATGACTGT
VWF	von Willebrand factor	VWF_3	mis (A/T)	0.38	GAAGGGCTCGAGTGT[A/G]CCAAAACGTGCCAGA
VWF	von Willebrand factor	VWF_4	mis (V/L)	0.09	GACATCCTGCAGCGG[G/T]TGCGAGAGATCCGCT
VWF	von Willebrand factor	VWF_5	silent	0.37	GCTGTCCCCCGTCTA[C/T]GCCGGGAAGACCTGC
ZNF239	zinc finger protein 239	ZNF239_1	silent	0.21	CGCCTTTGCTCGTGC[C/G]GTGGACACACGGCTG