Clin Chem Lab Med 2003; 41(4):529–534 © 2003 by Walter de Gruyter · Berlin · New York

Analysis of Multiple Single Nucleotide Polymorphisms of Candidate Genes Related to Coronary Heart Disease Susceptibility by Using Support Vector Machines

Yeomin Yoon[1], Junghan Song[2], Seung Ho Hong[3] and Jin Q. Kim[2]*

[1] Department of Laboratory Medicine, Cheju National University College of Medicine, Jeju, South Korea
[2] Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul, South Korea
[3] Jeju National University of Education, Jeju, South Korea

Coronary heart disease (CHD) is a complex genetic disease involving gene-environment interaction. Many association studies between single nucleotide polymorphisms (SNPs) of candidate genes and CHD have been reported. We have applied a new method to analyze such relationships using support vector machines (SVMs), which is one of the methods for artificial neuronal network. We assumed that common haplotype implicit in genotypes will differ between cases and controls, and that this will allow SVM-derived patterns to be classifiable according to subject genotypes. Fourteen SNPs of ten candidate genes in 86 CHD patients and 119 controls were investigated. Genotypes were transformed to a numerical vector by giving scores based on difference between the genotypes of each subject and the reference genotypes, which represent the healthy normal population. Overall classification accuracy by SVMs was 64.4% with a receiver operating characteristic (ROC) area of 0.639. By conventional analysis using the χ^2 test, the association between CHD and the SNP of the scavenger receptor B1 gene was most significant in terms of allele frequencies in cases *vs.* controls (p = 0.0001). In conclusion, we suggest that the application of SVMs for association studies of SNPs in candidate genes shows considerable promise and that further work could be usefully performed upon the estimation of CHD susceptibility in individuals of high risk. Clin Chem Lab Med 2003; 41(4): 529–534

Key words: Support vector machines; Coronary heart disease; Single nucleotide polymorphisms.

Abbreviations: apo, apolipoprotein; BMI, body mass index; CHD, coronary heart disease; HDL-C, high density lipoprotein-cholesterol; LDL-C, low density lipoprotein-cholesterol; Lp(a), lipoprotein(a); ROC, receiver operating characteristic; SNP, single nucleotide polymorphism; SRB1, scavenger receptor B1; SVMs, support vector machines.

Introduction

Coronary heart disease (CHD) is the leading cause of death in developed countries (1). CHD is the complex genetic disease involving many genes, environmental influences, and important gene-environment interactions (2). CHD patients show varying clinical and angiographic features and that the importance of the different pathogenetic components may be different in different patients (3). Prevention of CHD is made difficult not only by the multiple predisposing causes of CHD but also by the individual's susceptibility to these causes. Emerging evidence suggests that common variations in genes, called single nucleotide polymorphisms (SNPs), are associated with CHD and can modulate the effect of environmental risk factors on the development of CHD (4). However, the identification of SNPs in genes associated with complex disease including CHD is often very difficult because their individual contributions are likely to be small (5). Therefore, we not only analyzed the association between a single SNP and CHD using the χ^2 test for allele frequency in cases and controls but also estimated the combinational effects of multiple SNPs in attempt to detect CHD using support vector machines (SVMs).

SVMs (6, 7) have been successfully applied to a wide range of pattern recognition problems, including microarray gene expression data. SVMs can classify genes into some functional categories based on expression data obtained from a microarray and have allowed predictions to be made concerning the functions of unannotated yeast genes (8). SVMs, which are based on a solid mathematical foundation, attempt to solve a universal problem of classification that we need to know which belongs to which group. Applied to multiple SNP data, the process of constructing SVMs begins with the transformation of genotypes of multiple SNPs of each individual into a numerical vector. Vectors are labeled positively if the individuals are in the CHD group and are labeled negatively if they are in the control group. Using this training set of SNP vectors, SVMs would learn to discriminate between the CHD and control group (9). Having learned the vector features of the class, SVMs can recognize a new individual as a member of the CHD group or of the control group based on their SNP data. Moreover, the SVMs could also be retrained to identify outliers that may have previously been assigned to the incorrect class in the training set. Then, SVMs would use the information in the training set to determine what SNP features are characteristic of a given CHD or control group, and use this information to decide whether any

*E-mail of the corresponding author: jqkim@plaza.snu.ac.kr

given SNP data are likely belong to a CHD or control group.

We describe here not only the use of χ^2 statistics to determine the association between a single SNP and CHD but also the use of SVMs to classify subjects as members of the CHD or of the control group based on a set of multiple SNP data. We genotyped 14 SNPs on 10 candidate genes related to CHD risk in 86 CHD patients and 119 age-matched healthy controls.

Materials and Methods

Study subjects and samples

Eighty-six patients (54 males and 32 females) with CHD, as documented by coronary angiography because of recent myocardial infarction or angina, were selected at Seoul National University Hospital. The normal control group consisted of 119 age-matched individuals (63 males and 56 females) who were selected by health-screening at the same hospital in order to exclude those with a history of chest pain, diabetes, hypertension, and general illnesses. Blood samples were placed into EDTA tubes and stored at –70 °C until assay.

Lipid and apolipoprotein analysis

The concentrations of plasma cholesterol and triglycerides were determined using enzymatic methods (Roche Diagnostics, Mannheim, Germany). High density lipoprotein-cholesterol (HDL-C) was measured directly with HDL-C diagnostic kits (Kyowa Medex, Tokyo, Japan) using a Hitachi 747 automatic chemistry analyzer. The level of low density lipoprotein-cholesterol (LDL-C) was calculated using the formula of Friedwald *et al.* (10), and the levels of apolipoprotein (apo)A-I and apoB were measured by immunonephelometric assay (Bering Nephelometer, Beringwerke AG, Germany). Lipoprotein(a) (Lp(a)) was measured using commercially available enzyme-linked immunosorbent assay kit (IMMUNO GmbH, Heidelberg, Germany). Body mass index (BMI) was calculated by dividing weight by (height)^2.

Selection of candidate genes

We selected 10 genes probably related to CHD, based on encoded molecules that have roles in thrombosis, thrombolysis, vasodilator tone, and lipid metabolism. The ten genes were: *apoCIII*, *apoE*, lipoprotein lipase, scavenger receptor B1 (*SRB1*), lipoprotein receptor-related protein, factor VII, plasminogen activator inhibitor 1 (*PAI-1*), glycoprotein 1b α-polypeptide (*GP1BA*), superoxide dismutase (*SOD*), and the endothelial nitric oxide synthase (*eNOS*) genes. The accession number of appropriate the GeneBank reference sequences, the location of the sequences, and the bases potentially substituted in the 14 SNPs and the dbSNP numbers are summarized in Table 1.

Genotyping SNPs

DNA samples were extracted from peripheral blood by standard methods. Genomic DNA was subjected to PCR and the identity of the PCR products were confirmed by digestion with a restriction enzyme and subsequent agarose electrophoresis. Fourteen pairs of oligomers were chosen to serve as PCR primers to amplify regions containing each of the SNPs in the 10 candidate genes. The nucleotide sequence of these primers, restriction enzymes, and the expected sizes of the PCR products are indicated in Table 2.

Statistical analysis

Statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS, SPSS Inc., Chicago, IL, USA), version 9.01 for Windows. Variables in two or three groups were compared using the Mann-Whitney U-test or the Kruskal-Wallis test. The χ^2 test and Fisher's exact test were used to test for independent relationships between variables. The difference of the allele frequencies of CHD patients and controls were evaluated using χ^2 test. The Hardy-Weinberg equilibrium of alleles at individual loci was assessed using χ^2 statistics.

Table 1 Genotyped SNPs.

Genes	Gene Bank Reference Sequence			Base	
	Accession number	Location of SNP	dbSNP number	Major allele	Minor allele
SRB1	NM_005505.2 (gi21361199)	1158	rs5888	*C*	*T*
ApoE	K00396.1 (gi 178850)	586	rs7412	*C*	*T*
	K00396.1 (gi 178850)	448	rs429358	*T*	*C*
eNOS	D26607.1 (gi558523)	7002	rs1799983	*G*	*T*
	D26607.1 (gi558523)	20454	rs1799985	*G*	*T*
SOD	U10116.1 (gi529149)	5256	rs2536512	*T*	*G*
ApoCIII	J00098.1 (gi 178765)	5163	rs5128	*G*	*C*
LPL	AF050163.1 (gi3293304)	4509	rs285	*T*	*C*
	AF050163.1 (gi3293304)	8393	rs320	*T*	*G*
	AF050163.1 (gi3293304)	9040	rs328	*C*	*G*
Factor VII	NM_019616.1 (gi10518502)	1223	rs6046	*G*	*A*
PAI-1	AF386492.2 (gi14488407)	837	rs1799889	*G*	–
LRP1	AF058399.1 (gi3493546)	516	rs1799986	*C*	*T*
GP1BA	AF395009.1 (gi14600281)	2217	rs6065	*C*	*T*

SRB1, scavenger receptor B1; apoE, apolipoprotein E; eNOS, endothelial nitric oxide synthase; SOD, superoxide dismutase; ApoCIII, apolipoprotein CIII; LPL, lipoprotein lipase, PAI-1, plasminogen activator inhibitor 1; LRP1, lipoprotein receptor-related protein 1; GP1BA, glycoprotein 1b α-polypeptide (GP1BA).

Table 2 PCR primers and restriction enzyme digestion to detect SNPs.

Genes	SNPs (dbSNP)	Primers	Digest	Pruduct sizes, bp
SRB1	rs5888	ccttgtttcttcccatcctcacttcctcaaggc	*Hae*III	CC: 154 ,33, 31
	(C/T)	caccaccccagcccacagcagc		TT: 154, 64
ApoE	rs7412	tccaaggagctgcaggcggcgca	*Afl*III	CC: 218
	(C/T)	gccccggcctggtacactgcca		TT: 168, 50
	rs429358	tccaaggagctgcaggcggcgca	*Hae*II	TT: 218
	(T/C)	gccccggcctggtacactgcca		CC: 195, 23
eNOS	rs1799983	tccctgaggagggcatgaggct	*Ban*II	GG: 320, 137
	(G/T)	tgagggtcacacaggttcct		TT: 457
	rs1799985	cccctgagtcatctaagtattc	*Hind*II	GG: 577, 99
	(G/T)	agctctggcacagtcaag		TT: 374, 203, 99
SOD	rs2536512	gagacatgtacgccaaggtc	*Bst*UI	AA: 114, 39
	(T/G)	gctgccggaagaggac		GG: 66, 48, 39
ApoCIII	rs5128	ggagggtgattcctacctta	*Sst*I	GG: 710
	(G/C)	tttgacttgtgctggggttc		CC: 377, 333
LPL	rs285	atcaggcaatgcgtatgaggtaa	*Pvu*II	TT: 431
	(T/C)	gagacacagatctcttaagac		CC: 222, 209
	rs320	gatgctacctggataatcaaag	*Hind*III	TT: 354
	(T/G)	cagctagacattgctagtgt		GG: 141, 213
	rs328	catccattttcttccacaggg	*Hinf*I	CC: 140
	(C/G)	tagcccagaatgctcaccagact		GG: 118, 22
Factor VII	rs6046	gggagactccccaaatatcac	*Msp*I	GG: 206, 67, 39
	(G/A)	acgcagccttggctttctctc		AA: 273, 39
PAI-1	rs1799889	5G: gtctggacacgtggggg	Allele-specific PCR	5G: 139
	(G/–)	4G: gtctggacacgtgggga		4G: 138
		tgcagccagccacgtgattgtctag		
LRP1	rs1799986	ggggtccaggactgcatgta	*Rsa*I	CC: 32, 19, 8
	(C/T)	aagtccgtacctcggcagtg		TT: 51, 8
GP1BA	rs6065	cactactgaaccaaccccaag	*Bbi*II	CC: 271, 201, 119
	(C/T)	ttgtggcagacaccaggatgg		TT: 390, 201

For abbreviations: see Table 1.

Support vector machines

When SVMs classify, they separate a given set of binary-labeled training data with a hyper-plane that is maximally distant from the point of each set. For cases in which no linear separation is possible, SVMs can work using kernel functions, which automatically realize a non-linear mapping to a feature space. The hyper-plane found by the SVM in feature space corresponds to a non-linear decision boundary in the input space. For a set of 14 SNPs in each individual, their representation as a vector in a Euclidean space is found as follows: The reference genotype at each SNP site can be selected by choosing the most common genotypes in the normal controls. By giving scores based on difference between the genotypes of the SNPs in each subject (CHD patients and controls) and the reference genotype, a vector of 14 dimensions was assigned to each individual. A set of *n* SNPs sites from *m* individuals was represented as a difference score matrix, in which each of the *m* rows consists of an *n*-element difference score vectors. In our experiments, the number of individuals *m* was 203 and the number of difference scores *n* was 14. At each SNP location, we awarded difference score uniformly such as *diff(w/w, w/w)* = 1, *diff(w/w, w/m)* = 2.5, *diff(w/m, m/m)* = 7.5, and *diff(w/w, m/m)* = 10. Here, *w* and *m* represent wild and mutated genotypes, respectively. For best performance, the difference scores were adjusted at each SNP location by giving weights of χ values, which were obtained from the χ^2 test of allele frequency of a single SNP between CHD and controls (9).

Let the *j*th individual input difference score point $x^j = (x^j_1, ..., x^j_n)$ be the realization of the random vector X^j. Let this input point be labeled by CHD (+1) or controls (–1), $Y^j \in \{+1,-1\}$. Let $\Phi : I \subseteq \Re^n \rightarrow I \subseteq \Re^N$ be a mapping from the input space $I \subseteq \Re^n$ to a feature space *F*. A data set S of *m* labeled data point: $S = \{(X^1,y^1), ..., (X^{\mathrm{m}},y^{\mathrm{m}})\}$. The SVM learning algorithm finds a hyper-plane (w,*b*) such that the quantity $\gamma = \min_i y^i\{\langle W,\phi(X^i)\rangle - b\}$ is maximized, where $\langle . \rangle$ denotes an inner product, the vector *W* has the same dimensionality as *F*, *b* is a real number, and γ is called the margin. The quantity $\{\langle W,\phi(X^i)\rangle - b\}$ corresponds to the distance between the point X^i and the decision boundary. The corresponding decision function is $f(X) = sign\{\langle W,\phi(X)\rangle - b\}$. It is easy to prove that, for the maximal margin hyper-plane, $w = \sum_{i=1}^{m} \alpha_i y_i \phi(X^i)$, where α_i are positive real numbers that maximize $\sum_i \alpha_i - \sum_{ij} \alpha_i \alpha_j y_i y_j \langle \phi(X_i), \phi(X_j)\rangle$ subject to $\sum_{i=1}^{m} \alpha_i y_i = 0, \alpha_i > 0$, the decision function can equivalently be expressed as $f(X) = sign\left(\sum_{i=1}^{m} \alpha_i y_i \langle \phi(X^i), \phi(X)\rangle - b\right)$. From the equation it is possible to see that the α_i associated with the training point X^i expresses the strength with which that point is embedded in the final decision function. A remarkable property of this alternative representation is that often only a subset of the points will be associated with non-zero α_i. These points are called support vectors and are the points that lie closest to the separating hyper-plane. The matrix $K_{ij} = \langle \phi(X^i), \phi(X^j)\rangle$ is called the kernel matrix. In the case when the data are not linearly separable, one can use more general functions, $K_{ij} = K(X^i,X^j)$, that provide non-linear decision boundaries. Two classical choices are polynomial kernels $K(X^i,X^j) = (\langle X^i,X^j\rangle + 1)^d$ and Gaussian kernels $K(X^i,X^j) = e^{-\frac{\|X^i - X^j\|}{\sigma^2}}$, where *d* and σ are kernel parameters. The experiments presented in this paper were performed using a freely available implementation of the SVM classifier which can be obtained at http://www.cs.columbia.edu/~bgrundy/svm (8).

Validation of SVM methods

The goal of SVM is that a learning classifier should become trained well enough in terms of its teaching examples to be able to generalize new examples (actual error rate). The actual error rate of small data numbers was determined by hold-one-out cross validation (11). The SVM is trained using data from all but one of the subjects. The vector of subject not used in training is then assigned a class by the SVM. A single SVM experiment consists of a series of hold-one-out experiments, the vector of each subject being held out and tested exactly once (11). A receiver operating characteristic (ROC) curve was constructed for each of the SVM methods, and the area under the ROC curve quantified the diagnostic accuracy of a test as a single number, with 1 indicating perfect discrimination and 0.5 signifying discrimination no better than random assignment (12).

Results

Conventional analysis for the association between single SNP and CHD

Demographic features and risk factor status among subjects are summarized in Table 3. No significant difference was found between the age and sex distribution of the CHD patients and controls. The levels of BMI, cholesterol, HDL-C, LDL-C, and apoAI showed significant differences between the CHD patients and controls. Allele and genotype frequencies of SNPs in candidate genes are shown in Table 4. An association between CHD and the SNP at *SRB1* gene was statistically significant in terms of allele frequency comparisons in CHD patients *vs*. the controls (p = 0.0001). However, no significant difference was found between the allele frequencies of the other SNPs in the CHD patients and controls.

Table 3 Characterization of CHD patients and controls.

	CHD patients (n = 86)	Controls (n = 119)	p[a]
Age	60.1 ± 7.7	58.0 ± 7.5	0.082
Sex			0.181
Male	54	63	
Female	32	56	
Smoking			0.053
Current	25	17	
Former	15	24	
Never	46	78	
Hypertension			
(+)	46	0	
(–)	40	119	
Diabetes			
(+)	21	0	
(–)	65	119	
BMI (kg/m^2)[a]	24.70 ± 2.80	23.40 ± 2.80	0.002
Chol (mmol/l)	5.06 ± 1.10	5.30 ± 0.93	0.129
TG (mmol/l)	1.43 ± 0.83	1.29 ± 0.55	0.337
HDL-C (mmol/l)[a]	1.06 ± 0.30	1.51 ± 0.38	0.000
LDL-C (mmol/l)	3.34 ± 1.01	3.20 ± 0.84	0.242
ApoAI (g/l)[a]	1.06 ± 0.25	1.40 ± 0.26	0.000
ApoB (g/l)	1.03 ± 0.28	1.09 ± 0.26	0.095
Lp(a) (g/l)	0.31 ± 0.23	0.26 ± 0.18	0.184

Values are mean ± SD. BMI, body mass index; Chol, total cholesterol; TG, triglyceride; HDL-C, high denstiy lipoprotein-cholesterol; LDL-C, low density lipoprotein-cholesterol; Apo, apolipoprotein; Lp(a), lipoprotein(a). [a]Significant differences between CHD patients and controls (p-value by Kruskal-Wallis test and χ^2 test).

SVM analysis for the association between multiple SNPs and CHD

Table 5 summarizes the results of a hold-one-out cross-validation experiment using all four SMV methods. Performance was evaluated for each method and was involved allocating a positive or negative classification label for each member of the test set based only on what it has learned from the training set. The first four columns are the categories false positive (FP), false negative (FN), true positive (TP), and true negative (TN), and the fifth is a measure of overall accuracy. No significant difference in accuracy and in ROC areas was observed for the different SVM methods.

Discussion

It is estimated that 90% of naturally occurring sequence variations are SNPs, which can be small enough to manifest detectable linkage disequilibrium in human population (13). Mutations, which have recently been introduced into a population, tend to demonstrate linkage disequilibrium with nearby polymorphisms. Detecting association between these SNPs and disease may provide useful evidence for the existence of a susceptibility locus within such a region and can lead to the identification of the gene and of pathogenic polymorphisms. However, as SNPs are biallelic they have relatively little power in association studies compared with the information that could be obtained by using haplotypes. With regard to complex disease, such as CHD, each related gene contributes disease susceptibility to some extent, but such genes may also interact with each other. If different mutational events have occurred in a gene related to disease at different points in some population history, then the frequencies of both these haplotypes of mutations will be increased among cases. However, it is very difficult to detect these using conventional methods, when only the multilocus genotypes are available for study. Studying individual SNPs might be fail to detect association with disease because each allele might be associated with a different gene, which have opposing effects on disease development, the two thus tend to cancel each other and the result is little difference in the overall allele frequencies in comparisons between cases and controls (14).

Some association has been maintained through link-

Table 4 Genotypic and allele frequencies of SNPs in CHD patients and controls.

Genes	SNPs		Alleles				Genotype		
	SNPs (dbSNP)	Groups	Major	Minor	χ value	p[a]	Major homo	Hetero-zygotes	Minor homo-zygotes
SRB1	rs5888	CHD	0.85	0.15	4.109	0.000	61	25	0
	(C/T)	Controls	0.67	0.33			61	37	21
ApoE	rs7412	CHD	0.94	0.06	1.734	0.087	76	10	0
	(C/T)	Controls	0.98	0.02			114	5	0
	rs429358	CHD	0.92	0.08	1.362	0.178	73	12	1
	(T/C)	Controls	0.87	0.13			89	29	1
eNOS	rs1799983	CHD	0.91	0.09	0.311	0.761	71	15	0
	(G/T)	Controls	0.92	0.08			101	18	0
	rs1799985	CHD	0.98	0.02	1.266	0.207	83	3	0
	(G/T)	Controls	0.95	0.05			108	11	0
SOD	rs2536512	CHD	0.62	0.38	1.183	0.243	31	46	9
	(T/G)	Controls	0.69	0.31			53	57	9
ApoCIII	rs5128	CHD	0.73	0.27	0.998	0.321	45	35	6
	(G/C)	Controls	0.67	0.33			57	46	16
LPL	rs285	CHD	0.68	0.32	0.285	0.780	39	39	8
	(T/C)	Controls	0.66	0.34			52	54	13
	rs320	CHD	0.75	0.25	1.001	0.320	48	33	5
	(T/G)	Controls	0.79	0.21			73	43	3
	rs328	CHD	0.88	0.12	0.155	0.883	66	19	1
	(C/G)	Controls	0.89	0.11			93	25	1
Factor VII	rs6046	CHD	0.94	0.06	0.537	0.593	75	11	0
	(G/A)	Controls	0.92	0.08			100	18	1
PAI-1	rs1799889	CHD	0.57	0.43	0.387	0.698	32	35	19
	(G/–)	Controls	0.55	0.45			39	52	28
LRP1	rs1799986	CHD	0.92	0.08	0.164	0.878	74	11	1
	(C/T)	Controls	0.93	0.07			104	14	1
GP1BA	rs6065	CHD	0.89	0.11	0.138	0.893	70	14	2
	(C/T)	Controls	0.90	0.10			96	23	0

[a]p-Value of χ^2 test, allele frequency between CHD and controls in various SNPs. For abbeviations: see Table 1.

Table 5 Comparison of accuracies and ROC areas for various SVM methods.

Methods	Kernel function	FP	FN	TP	TN	Accuracy[a]	ROC area	p[b]
Polynomial 1	$(\vec{X}\cdot\vec{Y}+1)$, C = 1	50	33	53	69	59.4%	0.603 ± 0.039	0.012
Polynomial 2	$(\vec{X}\cdot\vec{Y}+1)^2$, C = 1	47	36	50	72	59.5%	0.613 ± 0.039	0.006
Polynomial 3	$(\vec{X}\cdot\vec{Y}+1)^3$, C = 1	46	35	51	73	60.5%	0.623 ± 0.039	0.003
Radial basis	$\exp(-7 \,\|\, \vec{X}-\vec{Y} \,\|^2)$	43	30	56	76	64.4%	0.639 ± 0.039	0.001

[a] Accuracy, percentage of (TP+TN)/total subjects; [b] p-value of ROC areas, under the nonparametric assumption and null hypothesis true area = 0.5.

age disequilibrium, and particular haplotypes should be more commonly found on chromosomes bearing pathogenic mutations. Such haplotypes should act in a similar way to multiallelic markers, and should be better able to produce detectable associations, when there are multiple mutation events. Then, in association studies between multiple SNPs and complex disease, it is more efficient to use a marker haplotype rather than a single biallelic marker (5, 13).

Our results show that SVMs can provide a simple and practical method for dealing with a set of multiple SNP data as a marker haplotype, such is produced using marker haplotypes, which are obtained from standard case-control studies. SVMs can produce a discriminant function (hyper-plane function), which has the ability to classify sets of input values (difference score vectors) according to their output values (CHD patients and controls), so that a given set of input values will produce a set of outputs close to the observed values (6). Standard methods are to adjust the SVM parameters (hyper-plane function) in order to produce output values, which approximate the target values. Sets of input data and target outputs are repeatedly presented to the SVMs and changes are made to modify parameters of the hyper-plane function in a way which improves the overall performance of the SVMs.

In terms of the association study between multiple SNPs and CHD, SVMs allow multiple SNP data to be analyzed simultaneously, even when haplotypes are unavailable. Such analyses complement conventional analyses based on single markers.

References

1. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. Lancet 1997; 349:1498–504.
2. Lusis A, Weireb A, Drake TA. Genetics of atherosclerosis. In: Topol EJ, editor. Textbook of cardiovascular medicine. Philadelphia: Lippincott-Raven, 1998:2389–413.
3. Betriu A, Castaner A, Sanz GA, Pare JC, Roig E, Coll S, *et al.* Angiographic findings 1 month after myocardial infarction: a prospective study of 259 survivors. Circulation 1982; 1099–105.
4. Niccoli G, Iacoviello L, Cianflone D, Crea F. Coronary risk factors: new perspectives. Int J Epidemiol 2001; 30 Suppl 1:S41–7.
5. Sham PC, Zhao JH, Curtis D. The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations. Ann Hum Genet 2000; 64:161–9.
6. Vapnik V. Statistical learning theory. New York: Wiley, 1998.
7. Scholkopt C, Burges JC, Smola AJ. Advances in kernel methods. Cambridge, MA: MIT Press, 1999.
8. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, *et al.* Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 2000; 97:262–7.
9. Kim G, Kim M. Application of support vector machine to detect an association between a disease or trait and multiple SNP variation, (http://xxx.lanl.gov/abs/cs.CC/0104015).
10. Friedwald WT, Levy RI, Fredrikson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972; 18:499–502.
11. Burden FR, Brereton RG, Walsh PT. Cross-validatory selection of test and validation sets in multivariate calibration and neural networks as applied to spectroscopy. Analyst 1997; 122:1015–22.
12. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39:561–77.
13. Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 1999; 96:15173–7.
14. Curtis D, North BV, Sham PC. Use of an artificial neural network to detect association between a disease and multiple marker genotypes. Ann Hum Genet 2001; 65:95–107.

Received 29 November 2002, accepted 13 January 2003

Corresponding author: Dr. Jin Q. Kim, Department of
Laboratory Medicine, Seoul National University College of
Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-799,
Korea
Fax: +82-2-745-6653, E-mail: jqkim@plaza.snu.ac.kr