

ORIGINAL ARTICLE

ISSN 1738-5520

Korean Circulation J 2005;35:759-765

© 2005, The Korean Society of Circulation

단계적 회귀분석을 이용한 혈중 지질농도에서 유전자의 상대적 효과 분석

연세대학교 의과대학 의학통계학과,¹ 연세대학교 심혈관계질환 유전체 연구센터²송기준¹ · 고민진¹ · 박찬미¹ · 임길섭¹ · 장양수² · 김동기¹

Analysis of the Relative Effects of SNPs within a Gene to Serum Lipid Profiles Using Stepwise Linear Regression

Kijun Song, PhD¹, Min Jin Go, BS¹, Chan Mi Park, MS¹,Kil Seob Lim, PhD¹, Yang Soo Jang, MD, PhD² and Dong Kee Kim, PhD¹¹Department of Biostatistics and ²Cardiovascular Genome Center, Yonsei University College of Medicine, Seoul, Korea

ABSTRACT

Background and Objectives : It is very important to distinguish between the primary and secondary genetic effects at different sites within a small genetic region. Therefore, we evaluated the relative effects of single nucleotide polymorphisms (SNPs) within a gene on the serum lipid profiles by using individual data. **Subjects and Methods :** To evaluate the contributions of SNPs in a region to the serum lipid profiles (total cholesterol, triglyceride, low density lipoprotein, high density lipoprotein), we used data that consisted of 808 individuals (327 males and 481 females) who did not have cardiovascular disease. In this study, we used a stepwise regression procedure to analyze the relative effects of four single nucleotide polymorphisms (ACE6, ACE7, ACE8, ACE10) in a gene region on the development of the serum lipid profiles in each gender group. **Results :** In the males, there were epistatic interaction effects between two loci (ACE6 × ACE7, ACE6 × ACE8, ACE6 × ACE10, ACE8 × ACE10 and ACE7 × ACE8) and among three loci (ACE6 × ACE7 × ACE8, ACE6 × ACE7 × ACE10 and ACE6 × ACE8 × ACE10). Also, there are interaction effects between two loci (ACE6 × ACE7, ACE6 × ACE8, ACE6 × ACE10, ACE7 × ACE10 and ACE8 × ACE10) and among three loci (ACE6 × ACE7 × ACE8, ACE6 × ACE7 × ACE10, ACE6 × ACE8 × ACE10 and ACE7 × ACE8 × ACE10) in the females. **Conclusion :** The results suggested that each of these loci is important in causing a relative change of the serum lipid profiles, even with simultaneously accounting for the effects at the other loci. In the results of the analysis, there existed the effects of individual loci and significant interaction between the loci on the serum lipid profiles in each gender group. It was confirmed that this stepwise regression method can be suitable for evaluating the relative effects of SNPs and it is easily performed. (Korean Circulation J 2005;35:759-765)

KEY WORDS : angiotensin converting enzyme ; Lipids ; Regression ; Polymorphism genetic.

서 론

질병과 관련 있는 유전자를 발견하는데 있어서 중요한 방법론적 논점의 하나는 질병발생과 직접적으로 관련된 대립

유전자(alleles)와 질병에 직접적으로 관여하지는 않으나 연관 불균형(linkage disequilibrium) 상태로 그 유전자의 근처에 위치하고 있는 유전자를 구별해내는 일이다. 이러한 작업은 유전자간의 연관 불균형의 강도나 단일 유전자 내에서

논문접수일 : 2005년 6월 20일

수정논문접수일 : 2005년 8월 5일

심사완료일 : 2005년 9월 20일

교신저자 : 김동기, 120-752 서울 서대문구 신촌동 134 연세대학교 의과대학 의학통계학과

전화 : (02) 2228-2491 · 전송 : (02) 364-8037 · E-mail : dkkimbios@yumc.yonsei.ac.kr

질병에 관여하는 단일 염기 다형성(single nucleotide polymorphism; SNP)이 하나 이상 존재할 수 있다는 점 때문에 매우 복잡한 일이 될 수 있다. 만약, 염색체상에서 특정한 질병의 발생에 관여하는 국부(region)의 위치가 알려질 경우, 그 국부에는 수많은 단일 염기 다형성이 존재할 수 있는데, 이 때 어떤 단일 염기 다형성이 혹은 어떤 조합(combination)이 질병발생에 중요한 역할을 하는 지 규명하는 것은 매우 의미 있는 일이다. 또한, 질병발생과의 관련성(association)이 단순히 연관 불균형 상태에 있기 때문에 간접적으로 나타나는 것인지 아니면 직접적으로 질병을 야기하는 효과 때문인지 규명하는 것 역시 중요한 일이라고 할 수 있다. 이를 검증하기 위해 흔히 접근되는 분석 방법은 반수체(haplotype) 단위를 이용하는 것인데, 이는 환자군과 정상군에서 관심 있는 유전소(locus)에 있는 대립유전자들의 상대 빈도(relative frequency)가 그 주변의 유전소에 있는 대립유전자들의 상대 빈도와 같은지를 비교하는 것이다.¹⁻⁴⁾ 이러한 방법은 관심 있는 유전소에는 동일한 대립유전자들을 갖고 있고 그렇지 않은 유전소에는 다른 대립유전자들을 갖고 있는 반수체들에 대하여 환자군과 정상군에서의 비차비(odds ratio)를 계산하여 그 효과를 평가하는 방법으로 변형되기도 하였다.⁵⁻⁷⁾ 그런데, 이러한 반수체를 이용한 분석 방법은 몇 가지의 단점을 내포하고 있는데, 첫째 유전소에 있는 복수의 대립유전자들을 동시에 비교하기 때문에 어느 대립유전자가 질병 발생에 직접적으로 관련이 있는 지 밝혀내는 것이 명확하지 않을 수 있고, 둘째 반수체를 구성하고 있는 유전소의 수가 다소 많은 경우에는 통계학적 비교를 반복적으로 수행해야 하므로 다중 검정(multiple testing)에 따른 유의 수준(significance level)의 문제가 대두될 수 있다. 또한, 반수체를 구성하고 있는 유전소의 순서 등에 의한 조합(combination)에 의해 독특하게 나타날 수 있는 추가적인 정보들을 무시할 가능성이 존재하고 있다.^{8,9)} 이러한 기존의 방법에 대한 한계를 효율적으로 극복하기 위해 회귀 모형(regression model)에 기초한 방법이 제시되었는데, 질병의 유무(예를 들어, 환자군, 정상군으로 구분된 경우)처럼 두 개의 범주로 구분되어 있는 질적 형질(qualitative trait)인 경우, 반수체를 구성하고 있는 단일 염기 다형성들의 상대적인 효과를 평가하기 위해 단계적 로지스틱 회귀분석(stepwise logistic regression analysis) 방법이 적용되었다.¹⁰⁻¹³⁾ 이에 본 연구에서는 양적 형질(quantitative trait)에 대하여 반수체를 구성하고 있는 개별 단일 염기 다형성 및 그 조합들의 상대적 효과를 평가하기 위한 방법으로 단계적 선형회귀분석(stepwise linear regression analysis)을 제안하고자 한다. 또한 이를 이용하여 양적 형질로서 심혈관계질환(cardiovascular disease) 발생과 밀접하게 관련이 있다고 알려진 혈중지질농도(serum lipid profile)에 대한 특정한 다형성들의 상대적 효과를 실제로 평가하고자 한다.

대상 및 방법

본 연구에서는 연세대학교 심혈관계질환 유전체 연구센터에서 실시하는 건강검진에 참여한 일반인들 중에서, 질병 이환에 따른 영향을 배제하기 위해 심혈관계질환이 없는 남자 327(40.5%)명, 여자 481(59.5%)명, 총 808명을 그 대상으로 하였다. 연구 대상들은 심혈관계질환에 대한 위험인자가 없고 당뇨, 고혈압, 뇌혈관질환, 말초혈관질환 등이 없으며 또한 지질강하제나 호르몬제 등을 복용하지 않는 정상 성인들로 구성되어 있다.

단일 염기 다형성으로서 ACE(Angiotensin Converting Enzyme)에서 4개의 유전소 즉, -6(Alu I/D), -7(G14480C), -8(A14519G), -10(A22982G)들의 집합(collection)을 이용했는데, 이들에 관한 유전형을 분석하는 절차는 다음과 같다. 먼저, DNA는 PUREGENE DNA Purification kit(GENTRA, Minneapolis, MA)를 이용하여 5 mL 혈액 내의 백혈구 세포로부터 추출하였다. DNA 추출 후, ACE Alu I/D의 다형성 분석은 ACE 유전자 인트론 16번에 primer를 이용한 중합효소 연쇄반응(PCR)을 수행하였는데, 이 때 PCR thermal cycler(PTC-200, MJ Research, Waltham, MA)를 이용하여 실시하였다. G14480C, A14519G 및 A22982G의 다형성 분석은 SNP stream 25K system(Orchid Biosciences, Princeton, NJ)을 이용하여 SNP-IT(SNP-Identification Technology) assay 방법으로 단일염기확장 반응을 수행하였고, primer와 각각의 PCR 생성물을 재생시킨 후 biotin 또는 fluorescein isothiocyanate(FITC)으로 표지된 각각의 dideoxynucleotide와 반응시켜 단일염기를 확장하였다. 반응이 끝난 후 결과분석은 ELISA 판독기를 이용하여 분석하였으며 최종 유전형은 QC Review™ program을 이용하여 확인하였다.

양적 형질은 심혈관계질환과 관련 있다고 알려진 혈중 지질농도로서 총 콜레스테롤, 중성지방, LDL-콜레스테롤, HDL-콜레스테롤의 수치를 이용하였다.

본 연구에서는 개별적인 단일 염기 다형성 및 그것들로 구성된 조합(combination)의 상대적 효과를 평가하기 위해 단계적 회귀분석 방법을 적용하는데 먼저, 유전소의 집합에서 각각의 유전형(genotype)에 따라 가변수(dummy variable)를 정의하고 다음과 같이 선형회귀모형(linear regression model)을 설정한다. 즉, 2개의 대립유전자로 구성된 첫 번째 유전소에 있는 단일 염기 다형성의 유전형을 AA, Aa, aa로 표현하기로 한다면, 그것이 혈중지질농도에 유의한 영향을 주는 지 분석하기 위해 다음의 모형 (1)과 같은 회귀식(regression equation)을 설정한다.

$$\text{혈중지질농도} = \beta_0 + \beta_1 I_{AA} + \beta_2 I_{Aa} \quad (1)$$

여기에서, I_{AA} 는 유전형이 AA이면 1, 아니면 0의 값을 갖는

가변수이고 β_1 는 그에 따른 회귀계수(regression coefficient)이다. 마찬가지로 I_{Aa} 는 유전형이 Aa이면 1, 아니면 0이 되는 가변수이고 β_2 는 그에 따른 회귀계수이다. 위의 회귀 모형 (1)에서 회귀계수의 추정치(estimate)가 통계학적으로 유의하다면 유전적 영향이 존재한다고 평가한다. 만약, 두 번째의 유전소에 있는 단일 염기 다형성 또한 혈중지질농도에 영향을 미치는지 분석하고자 한다면 두 번째 유전소에 있는 유전형을 포함시킨 다음과 같은 모형 (2)를 설정할 수 있다.

$$\text{혈중지질농도} = \beta_0 + \beta_1 I_{AA} + \beta_2 I_{Aa} + \beta_3 I_{BB} + \beta_4 I_{Bb} \quad (2)$$

여기에서, I_{BB} 와 I_{Bb} 는 모형 (1)에서 정의한 것과 같은 방식으로 설정된 가변수들이고, β_3 과 β_4 는 각각 그에 따른 회귀계수이다. 두 번째 유전소에서도 혈중지질농도에 유의한 효과가 존재하는지는 모형 (1)과 모형 (2)의 비교를 통하여 분석하는데, 이는 모형 (2)가 혈중지질농도를 설명하는데 모형 (1)보다 더 적절하다면 두 번째 단일 염기 다형성의 효과 또한 유의하다고 판단하는 것이다. 비슷한 경우로 두 단일 염기 다형성간에 유전적 교호작용(epistatic interaction)의 효과가 존재하는지 분석할 경우에는 다음의 모형 (3)과 위의 모형 (2)를 비교하여 분석한다.

$$\begin{aligned} \text{혈중지질농도} = & \beta_0 + \beta_1 I_{AA} + \beta_2 I_{Aa} + \beta_3 I_{BB} + \beta_4 I_{Bb} \\ & + \beta_5 I_{AABB} + \beta_6 I_{AaBB} + \beta_7 I_{AABb} + \beta_8 I_{AaBb} \quad (3) \end{aligned}$$

위의 모형 (3)에서, I_{AABB} , I_{AaBB} , I_{AaBb} 는 두 개의 유전소에 유전적 교호작용을 나타내는 가변수들인데, 예를 들어, I_{AABB} 의 경우 두 개의 단일 염기 다형성의 유전형이 AA와 BB이면 1, 그 외에는 모두 0의 값을 갖는 가변수를 말한다. 또한 $\beta_0 \dots \beta_8$ 은 각 가변수에 해당되는 회귀계수들이다.

지금까지 설명한 방법을 이용하여 관심 있는 유전소의 개수만큼 각 단일 염기 다형성의 개별적인 주효과(main effect)와 그것들의 교호작용의 효과에 대한 유의성을 평가할 수 있다. 그런데, 이러한 과정에서 집합을 구성하고 있는 단일 염기 다형성의 개수가 많아지게 되면 그 주효과와 교호작용에 대한 조합의 수 역시 많아지게 된다. 아울러 그에 따른 모형의 비교 또한 다양한 각 경우에 대하여 복잡하게 이루어져야 하므로, 유의성 평가를 효율적으로 수행할 필요성이 대두된다. 이러한 효율적인 분석을 수행하기 위해, 본 연구에서는 적절한 회귀모형의 구축을 위해 일반적으로 사용되는 단계적 변수 선택(stepwise variable selection) 방법을 적용하였는데 그 절차는 다음과 같다.¹⁴⁻¹⁷⁾ 먼저, 종속변수에 영향력이 큰 독립변수를 하나씩 선형회귀모형에 추가해 가는데, 이렇게 독립변수들이 추가된 선형모형에서 특정한 독립변수가 종속변수에 유의한 영향을 미치지 못하면 이 독립변수를 선형회귀모형에서 제외시킨다. 이러한 방법을 최적

의 모형을 구성할 때까지 계속 반복하는 것이다.

SAS ver. 8.1(SAS Institute, Cary, NC)을 이용하여 지금까지 기술한 단계적 회귀분석 방법을 실제 자료에 적용하여 분석하였다.

결 과

연구 대상은 분석하고자 하는 단일 염기 다형성들의 정보를 모두 알고 있는 개인들로서, 평균 연령은 48.6±10.4세이고 남자의 경우 평균 연령은 48.4±10.7세, 여자의 평균 연령은 48.8±10.1세로 분포하고 있었다. 분석하고자 하는 혈중지질농도에서 남자의 경우, 평균±표준편차는 총 콜레스테롤에서 206.37±39.82, 중성지방은 168.17±132.53, LDL-콜레스테롤은 129.18±35.10, HDL-콜레스테롤은 45.82±10.22의 분포를 보여주고 있었다. 여자에서는 총 콜레스테롤의 경우 209.30±38.08, 중성지방은 127.84±100.05, LDL-콜레스테롤은 131.95±32.74, HDL-콜레스테롤은 51.51±11.98의 분포를 보여주고 있었다(Table 1). ACE 단일 염기 다형성에서 유전형의 빈도는 다음과 같다(Table 2). 또한, 유전형에 따라 구성 가능한 효과를 나타내는 가변수를 생성했는데, 그 개수는 개별적인 주효과에서 8개, 두 개의 다형성들간의 교호작용에서 24개, 세 개의 다형성들간의 교호작용에서 32개, 네 개의 경우에서 16개로 총 80개의 가변수를 만들어 분석에 이용했다.

성별에 따른 유전적인 영향을 구분하여 파악하기 위해 남

Table 1. Distribution of serum lipid profiles

Trait	Male (n=327)	Female (n=481)
Total cholesterol	206.94±40.34	208.03±37.85
Triglyceride	168.97±135.46	126.28±102.18
LDL-C	129.89±35.88	131.06±33.02
HDL-C	45.95±10.72	51.57±12.29

Values are Mean±SD. LDL-C: low density lipoprotein-cholesterol, HDL-C: high density lipoprotein-cholesterol

Table 2. Frequency of ACE-SNP genotypes

SNP	Genotype	Frequency (%)	
		in male	in female
ACE6	DD	57 (17.43)	64 (13.31)
	ID	159 (48.62)	243 (50.52)
	II	111 (33.94)	174 (36.17)
ACE7	CC	111 (33.94)	173 (35.97)
	CG	160 (48.93)	241 (50.10)
	GG	56 (17.13)	67 (13.93)
ACE8	AA	109 (33.33)	176 (36.59)
	AG	163 (49.85)	241 (50.10)
	GG	55 (16.82)	64 (13.31)
ACE10	AA	100 (30.58)	162 (33.68)
	AG	166 (50.76)	235 (48.86)
	GG	61 (18.65)	84 (17.46)

SNP: single nucleotide polymorphism, ACE: angiotensin converting enzyme, 6: Alu I/D, 7: G14480C, 8: A14519G, 10: A22982G

자와 여자 집단을 각각 나누어서 혈중 지질농도에 대한 ACE 단일 염기 다형성들의 상대적 효과를 평가하였는데, 남자의 경우에서 그 결과는 다음과 같다(Table 3). 먼저, 총 콜레스테롤의 경우 ACE6, ACE7, ACE8의 주효과에 대한 R^2 의 합은 4.5%이었고, 두 개의 유전자간의 교호작용, 즉 ACE6과 ACE7의 교호작용(ACE6×ACE7), ACE6과 ACE8(ACE6×ACE8) 및 ACE6과 ACE10의 교호작용(ACE6×ACE10)들에 대한 효과가 상대적으로 존재하고 있었는데, 그 때 R^2 의 합은 2.7%이었다. 또한 세 개의 유전자, ACE6, ACE7 및 ACE8 간의 교호작용(ACE6×ACE7×ACE8)에 대한 R^2 은 0.2%이었다. 모형에 포함된 효과들에 대한 총 R^2 의 합은 7.4%이었는데, 이 때 교호작용들에 대한 R^2 의 합은 2.9%로서 전체에서 39.2%의 상대적인 효과를 설명하고 있는 것으로 나타났다. 중성지방의 경우에는 주효과는 존재하지 않았고 교호작용의 효과만 존재하고 있었는데, ACE6과 ACE10(ACE6×ACE10), ACE8과 ACE10(ACE8×ACE10)의 교호작용에 대한 R^2 의 합은 2.3%이었고, ACE6, ACE8 및 ACE10간의 교호작용(ACE6×ACE8×ACE10)에 대한 R^2 은 1.1%로 나타났다. LDL-콜레스테롤에서는 ACE8에서 주효과에 대한 R^2 이 0.9%이었고, ACE6과 ACE8들과, ACE7과 ACE8들간의 교호

작용(ACE7×ACE8)에 대한 R^2 의 합이 4.7%이었으며, ACE6, ACE7 및 ACE8간의 교호작용과 ACE6, ACE8 및 ACE10간의 교호작용의 효과에 대한 R^2 의 합은 1.2%이었다. 이 때 총 R^2 의 합이 6.8%이었는데, 교호작용들에 대한 R^2 의 합은 5.9%로서 전체 R^2 의 86.8%를 차지하고 있었고, 이것은 주효과에 비해 6.6배나 높은 수치이다. 즉, 다형성들의 조합에 의한 상대적인 효과가 크게 존재하고 있다는 사실을 암시하는 것이라고 볼 수 있다. HDL-콜레스테롤의 경우에는 ACE6, ACE8, ACE10에서 주효과에 대한 R^2 의 합은 9.9%이고, ACE6과 ACE7의 교호작용과 ACE6과 ACE10의 교호작용에 대한 R^2 의 합은 2.0%, ACE6, ACE7 및 ACE10간의 교호작용(ACE6×ACE7×ACE10)의 R^2 은 1.2%로 나타났다.

여자에서 혈중 지질농도에 대한 ACE 단일 염기 다형성들의 상대적 효과를 분석한 결과는 다음과 같다(Table 4). 총 콜레스테롤에서는 ACE6과 ACE10에서 주효과에 대한 R^2 의 합이 0.7%로 나타났고, 두 개의 단일 염기 다형성간의 교호작용에서 ACE6과 ACE7, ACE6과 ACE8의 교호작용에 대한 R^2 의 합은 0.3%이었다. 또한 ACE6, ACE7 및 ACE10간의 교호작용에 대한 R^2 은 0.8%로서 전체 R^2 에서 교호작용들로 설명되는 R^2 은 61.1%를 차지하고 있었다. 중성지방의 경우에

Table 3. Relative effects of ACE-SNPs to serum lipid profiles using stepwise regression analysis in male

SNP Effect (genotype)	Total cholesterol		Triglyceride		LDL-C		HDL-C	
	$\hat{\beta}$ (SE)	R^2 *	$\hat{\beta}$ (SE)	R^2	$\hat{\beta}$ (SE)	R^2	$\hat{\beta}$ (SE)	R^2
ACE6 (ID)	-138.0 (39.5)	1.4					18.3 (10.2)	6.9
ACE6 (II)	-139.2 (39.7)	1.7						
ACE7 (GG)	-139.5 (48.3)	1.1						
ACE8 (AG)					116.2 (35.2)	0.9		
ACE8 (GG)	-7.5 (28.4)	0.3					-3.8 (5.4)	0.2
ACE10 (AG)							15.3 (10.2)	0.8
ACE10 (GG)							-1.2 (5.3)	2.0
Sum of R^2		4.5				0.9		9.9
ACE6 (ID)×ACE7 (CG)	-19.0 (48.3)	0.3					-24.9 (12.0)	1.0
ACE6 (ID)×ACE8 (AG)	67.0 (55.8)	0.3			-61.0 (49.6)	0.5		
ACE6 (ID)×ACE10 (AG)			-75.2 (67.6)	0.4			16.7 (17.6)	0.3
ACE6 (ID)×ACE10 (GG)	36.1 (19.9)	0.9						
ACE6 (II)×ACE8 (AG)					-107.0 (42.9)	2.9		
ACE6 (II)×ACE8 (GG)	-66.3 (48.8)	1.0			-52.9 (35.2)	0.7		
ACE6 (II)×ACE10 (AG)	-9.0 (11.7)	0.2					-19.9 (10.6)	0.7
ACE7 (GG)×ACE8 (AG)					-87.0 (49.6)	0.6		
ACE8 (AG)×ACE10 (AG)			95.5 (67.6)	0.7				
ACE8 (AG)×ACE10 (GG)			126.4 (60.9)	1.2				
Sum of R^2		2.7		2.3		4.7		2.0
ACE6 (ID)×ACE7 (CG)×ACE8 (AG)	-53.3 (62.5)	0.2			-60.1 (35.2)	0.7		
ACE6 (ID)×ACE7 (CG)×ACE10 (AG)							-27.4 (15.8)	1.2
ACE6 (II)×ACE8 (AG)×ACE10 (AG)			88.6 (116.7)	1.1	-40.0 (35.1)	0.5		
Sum of R^2		0.2		1.1		1.2		1.2
Total R^2		7.4		3.4		6.8		13.1

ACE: angiotensin converting enzyme, SNP: single nucleotide polymorphism, LDL-C: low density lipoprotein-cholesterol, HDL-C: high density lipoprotein-cholesterol, $\hat{\beta}$: estimate of regression coefficient, SE: standard error, 6: Alu I/D, 7: G14480C, 8: A14519G, 10: A22982G. *: R^2 s are percent

Table 4. Relative effects of ACE-SNPs to serum lipid profiles using stepwise regression analysis in female

SNP Effect (genotype)	Total cholesterol		Triglyceride		LDL-C		HDL-C	
	$\hat{\beta}$ (SE)	R^2 *	$\hat{\beta}$ (SE)	R^2	$\hat{\beta}$ (SE)	R^2	$\hat{\beta}$ (SE)	R^2
ACE6 (ID)	-5.8 (6.7)	0.2						
ACE6 (II)			23.3 (14.8)	0.5			-1.6 (1.8)	0.2
ACE7 (CG)					-18.7 (19.1)	0.2		
ACE8 (AG)			-41.3 (51.8)	0.2	20.9 (19.1)	0.1	12.8 (8.5)	0.5
ACE10 (AG)	10.0 (6.7)	0.5					6.1 (3.5)	0.8
Sum of R^2		0.7		0.7		0.3		1.5
ACE6 (ID)×ACE8 (GG)	28.0 (37.8)	0.1						
ACE6 (ID)×ACE10 (AG)			-40.7 (60.1)	0.1				
ACE6 (II)×ACE7 (CG)					46.4 (38.1)	0.2	-10.9 (12.1)	0.2
ACE6 (II)×ACE7 (GG)	37.2 (37.8)	0.2					18.0 (12.2)	0.5
ACE6 (II)×ACE10 (AG)					8.6 (9.8)	0.2		
ACE7 (GG)×ACE10 (AG)			243.3 (128.4)	0.5				
ACE8 (AG)×ACE10 (GG)					6.3 (9.1)	0.2	-16.6 (9.1)	0.2
Sum of R^2		0.3		0.6		0.6		0.9
ACE6 (ID)×ACE7 (CG)×ACE10 (AG)	73.0 (37.8)	0.8					17.7 (7.4)	1.5
ACE6 (ID)×ACE7 (GG)×ACE8 (AG)			98.1 (56.3)	0.6	-56.0 (46.6)	0.3	24.3 (12.4)	0.8
ACE6 (II)×ACE7 (GG)×ACE10 (GG)					23.7 (33.1)	0.2		
ACE6 (ID)×ACE8 (AG)×ACE10 (AG)					99.8 (60.3)	0.9	-35.7 (11.1)	0.9
ACE7 (CG)×ACE8 (AG)×ACE10 (AG)			93.6 (78.6)	0.3				
ACE7 (GG)×ACE8 (AG)×ACE10 (GG)					6.6 (39.2)	0.2		
Sum of R^2		0.8		0.9		1.6		3.2
Total R^2		1.8		2.2		2.5		5.6

ACE: angiotensin converting enzyme, SNP: single nucleotide polymorphism, LDL-C: low density lipoprotein-cholesterol, HDL-C: high density lipoprotein-cholesterol, $\hat{\beta}$: estimate of regression coefficient, SE: standard error, 6: Alu I/D, 7: G14480C, 8: A14519G, 10: A22982G. *: R^2 's are percent

는 ACE6, ACE8의 주효과가 존재하고 있었는데, 그 때 R^2 의 합은 0.7%이었고 ACE6과 ACE10의 교호작용 및 ACE7과 ACE10의 교호작용(ACE7×ACE10)의 효과에 대한 R^2 은 0.6%이었다. 세 개의 다형성에서 ACE6, ACE7 및 ACE8간의 교호작용과 ACE7, ACE8 및 ACE10간의 교호작용(ACE7×ACE8×ACE10)에 대한 효과가 나타났는데, 그 때 R^2 의 합은 0.9%이었다. 전체 R^2 의 합 2.2%에 대하여 교호작용들의 R^2 은 1.5%로서 약 68.2%를 차지하고 있었다. LDL-콜레스테롤에서는 ACE7과 ACE8의 주효과에서 R^2 이 0.3%이었고 ACE6과 ACE7, ACE6과 ACE10, ACE8과 ACE10의 두 개의 다형성들간의 교호작용에 대한 R^2 은 0.6%, 아울러 세 개의 경우에서 ACE6, ACE7 및 ACE8간, ACE6, ACE7 및 ACE10간, ACE7, ACE8 및 ACE10간, 그리고 ACE7, ACE8 및 ACE10간의 교호작용에 대한 R^2 은 1.6%이었다. 이 때 교호작용들에 의한 R^2 은 전체 R^2 의 88%를 차지하는 것으로 나타났다. HDL-콜레스테롤의 경우, ACE6, ACE8 및 ACE10의 개별적인 주효과에서 R^2 이 1.5%이었고, 두 개의 다형성들에서 ACE6과 ACE7, ACE8과 ACE10간의 교호작용에 대한 R^2 은 0.9%이었으며, 또한 세 개의 다형성들에서 ACE6, ACE7 및 ACE10간, ACE6, ACE7 및 ACE8간, ACE6, ACE8 및 ACE10간의 교호작용에 대한 R^2 은 3.2%이었다. 여기에서도 교호작용에 의한 상대적인 효과가 전체 R^2 =5.6%에 대

하여 73.2%를 점유하고 있었다.

고 찰

단일 염기 다형성과 같이 고밀도(high density)의 유전 지도(genetic map)에 대한 접근이 가능해짐으로 인하여, 이것들을 개별적으로 혹은 반수체와 같이 몇 개의 집합 단위로 이용하여 질병과 유전자간의 관련성 (association)을 분석하는 일은 복잡한 질병에 대한 유전적 영향을 연구하는데 있어서 중요한 위치를 점하고 있다.¹⁸⁻²⁰⁾

본 연구에서는 염색체상의 아주 작은 범위의 특정한 국부에서 강한 연관 불균형상태로 존재하는 단일 염기 다형성들의 양적 형질에 대한 상대적인 기여도(contribution)를 평가하기 위해 단계적 회귀분석의 절차를 적용하였다. 특히, 양적 형질로서 심혈관계질환과 관련 있다고 알려진 혈중 지질 농도들에 대하여 4개의 유전소로 구성된 ACE 단일 염기 다형성들의 상대적인 효과를 실제로 평가하였다. ACE 다형성들을 분석에 이용한 이유는 각 개별 다형성들이 각종 심혈관계질환 즉, 본태성 고혈압, 관상동맥질환, 허혈성 심질환, 심근 경색, 비후성 심근병증, 급성 심장사 등의 발생과 관련 있다고 국내·외의 연구들을 통해 보고되고 있고 집중적으로 연구되고 있는 것들로서 그 임상적 중요성을 지니고 있기 때

문이다.²¹⁻³⁰⁾ 구체적으로 기존의 한국인을 대상으로 한 심혈관계질환과 ACE 단일 염기 다형성들과의 관련성에 관한 연구에서 특히, 김 등²¹⁾의 연구에 의하면 남자의 경우 총콜레스테롤 수치와 ACE6(Alu I/D), ACE7(G14480C), ACE10(A22982G)간의 개별적 관련성이 존재하는 것으로 보고되었다. 또한 이 등²³⁾의 연구에서는 ACE6(Alu I/D)과 총콜레스테롤 수치 및 허혈성 심질환과의 관련성이 존재하는 것으로 보고되었으며, 박 등²⁴⁾의 연구에서도 허혈성 심질환 등과의 관련성이 보고되었다. 국외의 경우, Ribichini 등²⁹⁾의 연구에서는 ACE6(Alu I/D)이 관동맥 재협착 발생의 위험인자로 분석되었으며 Nakai 등³⁰⁾의 연구에서는 관상동맥질환 발생의 위험 인자로 보고되었다.

그러나 기존의 보고된 결과들은 개별적인 다형성들 각각과 질병과의 관련성을 분석한 것이며 그것들의 조합에 의한 교호작용의 영향 등 유전자의 상대적인 효과에 관한 고려가 이루어지지 않은 결과들이다. 즉, 하나의 다형성에서 동일한 유전형을 갖고 있지만 이웃하는 몇 개의 다형성들의 유전형이 어떻게 구성되어지는가에 따라 질병 발생의 양상은 개인별로 다르게 나올 수 있다는 가능성을 간과한 것이라고 할 수 있다. 이러한 사실은 본 연구의 결과를 통해서도 확인할 수 있었는데, 예를 들어 남자의 경우에서 중성지방이나 LDL-콜레스테롤에서는 오히려 개별적인 다형성들의 효과보다는 교호작용에 의한 상대적인 효과가 더 크게 나타났다는 것이다. 또한 여자의 경우에서도 모든 혈중 지질농도들에서 교호작용의 효과가 더 크게 존재하고 있었다는 점은 유전자의 구성에 따른 상대적인 효과의 중요성을 간과할 수 없다는 것을 나타내고 있는 것이다.

본 연구의 결과를 통해 볼 때, 혈중지질농도에서 ACE 다형성들에 의해 설명되는 전체의 합에서 그 조합에 의한 효과의 비율(proportion)은 개별 다형성의 효과에 비해 상대적으로 크다는 것을 알 수 있는데, 이것은 인구집단에서 각 개인들이 특정한 하나의 다형성에서는 동일한 유전형을 갖고 있음에도 불구하고 이웃하는 몇 개의 다형성들의 유전형이 어떻게 구성되어지는가에 따라 혈중지질농도의 변화에 미치는 영향이 상대적으로 다를 수 있으므로 질병 발생의 양상은 그에 따라 개인별로 다르게 나올 수 있다는 점을 의미하는 것이다.

본 연구에서 적용된 분석 방법은 새롭게 개발된 통계학적 분석 방법이 아님에도 불구하고 다중의 유전소로 구성된 반수체나 복수의 유전자들과 질병간의 관련성을 분석하는데 거의 적용되지 않은 것이 사실이다. 본 연구의 방법은 기존의 회귀분석 방법의 틀 내에서 상용화된 통계분석 프로그램들을 이용하여 손쉽게 적용할 수 있고, 환경적인 영향 등에 따른 공변량(covariates)들을 고려하여 분석할 수도 있다는 장점을 가지고 있다. 본 연구에서 적용한 방법은 각각의 유전소에 위치한 단일 염기 다형성들과 양적 형질간의 관련성을 차례대로 분석할 수 있는데, 그 때마다 다른 유전소에 존재하

는 유전형의 구성에 따라 조건부적으로 각 단일 염기 다형성의 효과를 평가할 수 있다는 장점을 가지고 있다. 또한, 각 단일 염기 다형성의 주효과와 교호작용의 효과를 포함하는 여러 가지 회귀모형을 설정함에 의해 다양한 상대적인 유전적 효과들에 대한 분석을 탄력적으로 수행할 수 있다. 그런데, 특정한 국부에 존재하는 단일 염기 다형성의 수가 과도하게 많아지면 그에 따른 각 효과들을 평가하기 위해 추정해야할 회귀계수의 수가 또한 많아지게 되므로, 표본 크기가 충분히 크지 않다면 분석을 수행하는데 어려움이 발생할 수 있다. 따라서 분석에 이용할 유전소의 집합을 구성하는 단일 염기 다형성의 수를 어느 정도로 결정할 것인가에 관한 문제가 생길 수 있다. 본 연구의 분석 방법은 독립적인 개인으로 구성된 자료에 적용하는 것이 적절한데, 만약 가족으로 구성된 자료(family data)인 경우에는 가족 구성원들간의 상관성(correlation)을 고려하여 분석할 수 있는 방법으로서의 확장이 필요할 것으로 여겨진다.

요 약

배경 및 목적 :

염색체상에서 특정한 질병의 발생에 관여하는 유전자들의 국부의 위치가 알려질 경우, 그 국부에는 수많은 단일 염기 다형성이 존재할 수 있는데, 이 때 어떤 단일 염기 다형성이 혹은 어떤 조합이 질병발생에 직접적인 영향을 주는 지 규명하는 것은 매우 의미 있는 일이다. 또한, 질병발생과의 관련성이 단순히 연관 불균형 상태에 있기 때문에 간접적으로 나타나는 것인지 아니면 직접적으로 질병을 야기하는 효과 때문인지 규명하는 것 역시 중요한 일이라고 할 수 있다. 이를 위해 본 연구에서는 양적 형질로서 심혈관계질환과 관련 있다고 알려진 혈중 지질농도에 대한 유전자들의 상대적 영향을 평가하기 위해 단계적 회귀 분석 방법을 적용하여 분석하였다.

방 법 :

본 연구에서는 각 단일 염기 다형성들의 집합에서 각각의 유전형에 대한 가변수를 정의한 후, 그 가변수들이 포함된 회귀 모형의 유의성에 대한 검정을 통해 유전자의 상대적인 영향을 평가하기 위하여 단계적 회귀 분석 방법을 적용하였다. 단일 염기 다형성으로서는 ACE6, ACE7, ACE8, ACE10들로 이루어진 집합을 이용했다. 양적 형질은 혈중 지질농도들로서 총 콜레스테롤, 중성지방, LDL-콜레스테롤, HDL-콜레스테롤의 수치를 이용하였고, 단일 염기 다형성의 정보를 모두 알고 있으며 심혈관계질환을 가지고 있지 않은 남자 327(40.5%)명, 여자 481(59.5%)명, 총 808명을 대상으로 분석하였다.

결 과 :

남자의 경우 혈중 지질농도들 별로 ACE6, ACE7, ACE8, ACE10에서 주효과가 각각 서로 다른 양상으로 존재하고 있

었고, 두 개의 다형성들간의 교호작용과 세 개의 다형성들간의 교호작용이 존재하고 있었다. 중성지방과 HDL-콜레스테롤에서는 개별적인 주효과에 비해 교호작용의 효과가 더 크게 나타났다. 여자에서도 주효과와 두 개 및 세 개의 다형성들간의 교호작용에 의한 효과가 존재하고 있었는데, 모든 혈중 지질농도들에서 교호작용에 의한 상대적인 효과가 개별적인 주효과보다 더 크게 나타났다.

결론 :

본 연구에서는 손쉽게 적용할 수 있는 단계적 회귀분석 방법을 이용하여 혈중 지질농도에 대한 단일 염기 다형성들의 개별적인 효과와 그것들로 구성된 조합의 상대적인 효과를 평가하였는데, 각 성별에 따라 개별적인 다형성의 효과뿐만 아니라 두 개 혹은 세 개의 다형성들간의 교호작용에 대한 효과 또한 존재한다는 사실을 알 수 있었다.

중심 단어 : 혈중 지질농도 ; 단계적 회귀분석.

본 논문은 보건복지부 보건의료바이오기술개발사업 중 특정센터연구지원사업(00-PJ3-PG6-GN01-0001)의 연구비 지원을 받아 이루어졌다.

REFERENCES

- 1) Thomson G, Robinson WP, Kuhner MK, et al. *Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. Am J Hum Genet* 1988;43:799-816.
- 2) Valdes AM, Thomson G. *Detecting disease-predisposing variants: the haplotype method. Am J Hum Genet* 1997;60:703-16.
- 3) Li H. *A permutation procedure for the haplotype method for identification of disease-predisposing variants. Ann Hum Genet* 2001;65:189-96.
- 4) Falk CT, Rubenstein P. *Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet* 1987;51:227-33.
- 5) Dudbridge F, Koeleman BP, Todd JA, Clayton DG. *Unbiased application of the transmission/disequilibrium test to multilocus haplotypes. Am J Hum Genet* 2000;66:2009-12.
- 6) Clayton D. *A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet* 1999;65:1170-7.
- 7) Mignot E, Lin L, Rogers W, et al. *Complex HLA-DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. Am J Hum Genet* 2001;68:686-99.
- 8) Cucca F, Dudbridge F, Loddo M, et al. *The HLADPB1: associated component of the IDDM1 and its relationship to the major loci HLA-DQB1, -DQA1, and -DRB1. Diabetes* 2001;50:1200-5.
- 9) Cucca F, Lampis R, Congia M, et al. *A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. Hum Mol Genet* 2001;10:2025-37.
- 10) Cordell HJ, Clayton DG. *A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet* 2002;70:124-41.
- 11) Kleinbaum DG, Klein M. *Logistic Regression: a self-learning text. 2nd ed. New York: Springer; 2002.*
- 12) Hosmer DW, Lemeshow S. *Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons; 2000.*
- 13) Jaccard J. *Interaction Effects in Logistic Regression. Thousand Oaks, Calif.: Sage Publications; 2001.*
- 14) Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models. 4th ed. New York: McGraw-Hill; 1996.*
- 15) Bowerman BL, O'Connell RT. *Linear Statistical Models: an applied approach. 2nd ed. Boston: PWS-KENT Pub. Co.; 1990.*
- 16) Fahrmeir L, Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models. 2nd ed. New York: Springer; 2001.*
- 17) Vittinghoff E. *Regression Methods in Biostatistics: linear, logistic, survival, and repeated measures models. New York: Springer; 2005.*
- 18) Clark AG. *The role of haplotypes in candidate gene studies. Genet Epidemiol* 2004;27:321-33.
- 19) Schaid DJ. *Genetic epidemiology and haplotypes. Genet Epidemiol* 2004;27:317-20.
- 20) Schaid DJ. *Evaluating associations of haplotypes with traits. Genet Epidemiol* 2004;27:348-64.
- 21) Kim J, Shin DJ, Bae Y, et al. *Association between I/D, G14480C, A22982G polymorphisms of angiotensin I-converting enzyme gene and essential hypertension in the Korean population. Korean Circ J* 2004;34:1137-47.
- 22) Kwon SU, Kim DK, Kim JH, et al. *Association of angiotensin-converting enzyme gene polymorphism and angiotensin-converting enzyme activity in Korean patients with coronary artery disease. Korean Circ J* 2000;30:1350-6.
- 23) Lee MM, Kim HS, Song JM, et al. *Insertion/deletion polymorphism of angiotensin I-converting enzyme gene in Korean patients with ischemic heart disease. Korean Circ J* 1996;26:5-13.
- 24) Park HY, Kwon HM, Kim HS, Song KS, Kim CH. *An I/D polymorphism in angiotensin-converting enzyme gene in myocardial infarction. Korean Circ J* 1996;26:465-72.
- 25) Lindpaintner K, Pfeffer MA, Kreutz R, et al. *A prospective evaluation of an angiotensin-converting enzyme gene polymorphism and the risk of ischemic heart disease. N Engl J Med* 1995;332:706-11.
- 26) Winkelmann BR, Nauck M, Klein B, et al. *Deletion polymorphism of the angiotensin I-converting enzyme gene is associated with increased plasma angiotensin-converting enzyme activity but not with increased risk for myocardial infarction and coronary artery disease. Ann Intern Med* 1996;125:19-25.
- 27) Agerholm-Larsen B, Nordestgaard BG, Steffensen R, Sorensen TI, Jensen G, Tybjaerg-Hasen A. *ACE gene polymorphism-ischemic heart disease and longevity in 10,150 individuals. Circulation* 1997;95:2358-67.
- 28) Marian AJ, Yu QT, Workman R, Greve G, Roberts R. *Angiotensin-converting enzyme polymorphism in hypertrophic cardiomyopathy and sudden cardiac death. Lancet* 1993;342:1085-6.
- 29) Ribichini F, Steffenino G, Dellavalle A, et al. *Plasma activity and insertion/deletion polymorphism of angiotensin I-converting enzyme: a major risk factor and a marker of risk for coronary stent restenosis. Circulation* 1998;97:147-54.
- 30) Nakai K, Itoh C, Miura Y, et al. *Deletion polymorphism of the angiotensin I-converting enzyme gene is associated with serum ACE concentration and increased risk for CAD in the Japanese. Circulation* 1994;90:2199-202.