

DOI: 10.3724/SP.J.1005.2011.00239

TNFSF4 基因与冠心病的关联研究

陈民志¹, 程广辉¹, 马龙¹, 王慧¹, 邱熔芳¹, 薛付忠², 刘奇迹¹

1. 山东大学医学院医学遗传学研究所, 山东大学实验畸形学教育部重点实验室, 济南 250012;
2. 山东大学公共卫生学院流行病学与卫生统计学研究所, 济南 250012

摘要: *TNFSF4* (Tumor necrosis factor superfamily, number 4) 基因是动脉粥样硬化的易感基因。但在瑞典、德国人群中进行的病例对照关联分析却得到了相反的结果。为探讨中国汉族人群中该基因与冠心病的关联性, 从山东大学齐鲁医院选取了 498 例病例及 509 例对照, 分析了 *TNFSF4* 基因上 5 个 SNP 位点 (rs1234314、rs45454293、rs3850641、rs1234313、rs3861950) 与冠心病之间的关联性。在采用传统的以单个 SNP 位点为单位以及以单体型为单位的统计分析方法的基础上, 引进基于主成分的 logistic 回归分析方法进行处理。结果显示: Armitage 趋势检验中只有 rs3861950 位点 ($P=0.0324$) 具有统计学意义, 经 Bonferroni 多重检验校正后, 5 个 SNP 位点均无统计学意义; 调整混杂因素的 logistic 回归分析中, 5 个 SNP 位点均无统计学意义; 单体型分析中, CTAGT ($P=0.0006$)、CTAAC ($P=0.0123$)、CCAGT ($P=0.0004$)、GTGGT ($P=0.0329$)、GCGAC ($P<0.0001$) 以及 GCAAC ($P=0.0173$) 这 6 个单体型在病例组和对照组中的频率差异具有统计学意义; 基于主成分的 logistic 回归分析中, 第一主成分具有统计学意义 ($P=0.0236$)。结果表明, *TNFSF4* 基因与冠心病之间存在关联性。

关键词: *TNFSF4* 基因; 冠心病; 主成分分析; logistic 回归分析; 关联分析

Association study between *TNFSF4* and coronary heart disease

CHEN Min-Zhi¹, CHENG Guang-Hui¹, MA Long¹, WANG Hui¹, QIU Rong-Fang¹,
XUE Fu-Zhong², LIU Qi-Ji¹

1. Institute of Medical Genetics, School of Medicine, Shandong University, Jinan 250012, China;
2. Institute of Epidemiology and Health Statistics, School of Public Health, Shandong University, Jinan 250012, China

Abstract: Previous studies suggest that *TNFSF4* is a susceptibility gene of atherosclerosis. But case-control association analysis in Swedish population and German population provided inconsistent, even opposite results. In order to explore the relationship between this gene and coronary heart disease (CHD) in Chinese Han population, we collected 498 cases and 509 controls from Qilu hospital of Shandong University and analyzed the association between five single-nucleotide polymorphisms (SNPs) (rs1234314, rs45454293, rs3850641, rs1234313, and rs3861950) of *TNFSF4* and CHD. On the basis of using traditional statistical analysis methods based on single SNP and haplotypes, we introduced principal component score-based logistic regression analysis to deal with the data. The results suggested that in Armitage trend test, only rs3861950 was significant, when used the Bonferroni correction, and all of the five SNPs were not statistically significant. In the logistic regression analysis which adjusts the confounding factors, all of the five SNPs were not statistically significant. In haplotype analysis, the frequencies of six haplotypes were significantly different in cases and controls (CTAGT

收稿日期: 2010-07-14; 修回日期: 2010-10-04

基金项目: 国家 863 项目 (编号: 2006AA02A406) 和国家 973 项目 (编号: 2007CB512001) 资助

作者简介: 陈民志, 硕士研究生, 专业方向: 群体遗传学。Tel: 0531-88393283; E-mail: chenminzhi11205@163.com

通讯作者: 刘奇迹, 教授, 博士生导师, 研究方向: 单基因病致病基因的定位与克隆, 复杂疾病的易感基因鉴定及功能分析。

E-mail: liuqiji@sdu.edu.cn

($P=0.0006$), CTAAC ($P=0.0123$), CCAAGT ($P=0.0004$), GTGGT ($P=0.0329$), GCGAC ($P<0.0001$), and GCAAC ($P=0.0173$). In principal component score-based logistic regression analysis, the first principal component has statistical significance ($P=0.0236$). These results indicate that *TNFSF4* is a susceptibility gene of CHD in Chinese Han population.

Keywords: *TNFSF4* gene; coronary heart disease; principal component analysis; logistic regression analysis; association analysis

TNFSF4(Tumor necrosis factor superfamily, number 4)基因长约 1 kb, 定位于 1q25.1^[1]。其编码的 OX40 配体(OX40L/CD134L)属于TNF家族成员, 为型跨膜糖蛋白^[2]。该基因是一个通过动物模型发现的动脉粥样硬化的易感基因, 该基因敲除小鼠在高脂饮食时表现出比对照小鼠具有抵抗动脉粥样硬化的表型, 而*TNFSF4* 基因过表达的小鼠则易患动脉粥样硬化^[1]。这表明该基因在动脉粥样硬化的发生中起着重要的作用。已有多篇文献报导了该基因与冠心病之间的关联分析结果^[1,3-6]。但在不同人群中进行的病例对照关联分析却得到了不一致的结果。2005年, Wang等^[1]利用瑞典人群进行的病例对照关联分析发现, *TNFSF4* 基因多态与心肌梗死和冠脉狭窄程度具有相关性; 但Koch等^[3]在德国人群中利用 1 211 例对照与 3 657 例病例进行的关联分析发现该基因与心肌梗死没有相关性。不同群体的遗传结构差异可能是造成关联分析结果不一致的原因。因此, *TNFSF4* 基因是否是中国汉族人群冠心病的易感基因值得深入研究。

本文在采用以单个 SNP 位点为单位和以单体型为单位的统计分析方法进行处理的基础上, 引进基于主成分的 logistic 回归分析方法来探讨中国汉族人群中 *TNFSF4* 基因与冠心病之间的关联性。

1 材料和方法

1.1 研究对象

本研究中的 1 007 个样本均来自于汉族人群。其中病例组 498 人, 是山东大学齐鲁医院从 2006 年 9 月到 2009 年 1 月收集的临床确诊患有冠心病的病人, 所有病人都进行了详细的体格检查, 行冠状动脉造影并且至少有一支冠状动脉狭窄大于 50%, 并排除甲亢、继发性高血压、风湿病、严重的肝肾功能不全等。对照组 509 人, 也是该院在同一时间段收集的通过临床及实验室检查无冠心病发病史的正

常个体。对所有样本均收集或测量了他们的收缩压、舒张压、总胆固醇、甘油三酯和葡萄糖水平, 并计算了他们的体重指数。

1.2 SNP 等位基因分型

查询 HapMap 数据库, 从 *TNFSF4* 基因中选取的 5 个 SNP 位点中, 有 3 个(rs3850641、rs1234313、rs3861950)来自在 2005 年对瑞典人群进行的病例对照关联分析研究^[1], 另外两个(rs1234314、rs45454293)位于 *TNFSF4* 基因 5 调控区域。对所有 SNP 位点采用 PCR-RFLP(Polymerase chain reaction and restriction fragment length polymorphism)方法分型。引物利用 primer5.0 设计, 由北京博尚生物技术有限责任公司合成, 上述 5 个位点的 PCR 产物分别用 *Mfe*、*HpyCHIV*、*Hinc*、*ScrF* 和 *Sac* 酶切, 琼脂糖凝胶电泳检测, 读取并记录基因型。

1.3 统计分析方法

1.3.1 单点检验和混杂因素的调整

采用 Armitage 趋势检验进行单点分析, 同时, 采用 logistic 回归模型调整混杂因素。

1.3.2 整体基因关联分析

本研究中, 用单体型分析来比较病例组和对照组的单体型频率的差异, 进而揭示 *TNFSF4* 基因上特定单体型与冠心病之间的关联性。统计方法上采用的是卡方检验。

另外, 一些研究小组提出了将主成分分析和 logistic 回归分析两者结合起来的检验方法—基于主成分的 logistic 回归分析方法^[7-10]。研究结果表明该方法的检验效能比基于单个 SNP 位点或者基于单体型分析方法更高^[7-10]。因此, 本研究引进该方法来分析 *TNFSF4* 基因与冠心病之间的关联性。其步骤是主成分分析、主成分的提取及主成分的 logistic 回归分析。

1.3.2.1 主成分分析

主成分分析是基于多个变量的线性组合提取数据综合信息的一种多元统计分析方法。主成分分析可基于基因分型相关系数、连锁不平衡及其他一些矩阵^[8], 本文基于基因分型相关系数矩阵计算主成分。候选基因 p 个SNP位点按加性模型(低频率等位基因纯合子赋2, 杂合子赋1, 高频率等位基因纯合子赋0)进行赋值, 得到一个数据矩阵 (X_1, X_2, \dots, X_p) 。应用主成分分析后, 我们得到 p 个主成分 (F_1, F_2, \dots, F_p) , 其中 $F_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p, i = 1, 2, \dots, p$ 。这些主成分间, $\text{cov}(F_i, F_j) = 0, i \neq j$ 并且 $\text{Var}(F_1) \geq \text{Var}(F_2) \geq \dots \geq \text{Var}(F_p)$ 。

1.3.2.2 主成分的提取

进行主成分分析后, 提取的主成分的个数依赖于特征值及主成分的累计贡献率。通常采用累计贡献率达到70%~85%作为准则, 如果用特征值来确定主成分的个数, 一般取特征值大于或等于1为准则。目前大都将两者结合应用确定主成分的提取个数。

对于基于病例对照研究关联分析, 有文献提出了4种不同的主成分提取方法^[11]: 一是分别从病例组和对照组中提取主成分, 然后利用相应主成分表达式计算各个体的主成分得分; 二是从病例组中提取主成分, 然后用该主成分表达式计算病例组和对照组个体的主成分得分; 三是从对照组中提取主成分, 然后用该表达式计算病例组和对照组个体的主成分得分; 四是将病例组和对照组合并为一组, 基于该合并组的主成分表达式计算病例组和对照组个体的主成分得分。其中的第三种既合理, 又具有较高的检验效能^[11], 本文将采用此种策略提取主

成分。

1.3.2.3 主成分的 logistic 回归分析

采用上述主成分提取策略并计算病例组和对照组个体的主成分, 进而以主成分为自变量, 以年龄、性别等指标为协变量, 建立 logistic 回归模型:

$$\ln \frac{P}{1-P} = \alpha + \beta F' + \gamma C',$$

其中, p 为患病的概率, α 为常数项, $F = (F_1, F_2, \dots, F_m)$ 是由提取的 m 个主成分组成的矩阵, $C = (C_1, C_2, \dots, C_k)$ 是由 k 个协变量组成的矩阵, $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ 及 $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$ 是相应的系数。

本文所有分析采用 SAS 9.1.3 完成。

2 结果与分析

2.1 协变量的描述统计

表1是性别(SEX)、年龄(AGE)、体重指数(BMI)、收缩压(SBP)、舒张压(DBP)、总胆固醇(TC)、甘油三酯(TG)和葡萄糖(GLU)等协变量在病例组与对照组之间的分布特征, 可见除性别(0.9006)和年龄($P=0.0500$)外, 其他协变量在病例组和对照组之间的差异均具有统计学意义($P < 0.05$)。

2.2 单位点 Armitage 趋势检验

表2是 *TNFSF4* 基因5个SNP位点的 Armitage 趋势检验的结果, 可见在单点检验水平上, 除 rs3861950 位点($P=0.0324$)具有统计学意义外, 其余4个SNP位点(rs1234314、rs45454293、rs3850641及rs1234313)均无统计学意义。经 Bonferroni 多重检验校正后, 5个SNP位点均无统计学意义。

表1 协变量的统计描述结果

协变量	对照组	病例组	假设检验 P 值
SEX(男/女)	360/149	354/144	0.9006 ^a
AGE(岁)	60.37±11.30	61.26±10.71	0.0500 ^b
BMI(kg/m ²)	24.43±3.33	25.94±7.55	<0.0001 ^b
SBP(mmHg)	120.86±10.68	130.87±19.11	<0.0001 ^b
DBP(mmHg)	73.47±8.27	76.33±11.52	<0.0001 ^b
TC(mmol/L)	4.65±1.13	4.76±0.68	0.0020 ^b
TG(mmol/L)	1.05±0.33	2.00±1.32	<0.0001 ^b
GLU(mmol/L)	4.90±0.48	6.01±1.85	<0.0001 ^b

注: SEX: 性别; AGE: 年龄; BMI: 体重指数; SBP: 收缩压; DBP: 舒张压; TC: 总胆固醇; TG: 甘油三酯; GLU: 葡萄糖; 下表同。^a: 表示进行的是卡方检验; ^b: 表示由于数据不能同时达到正态性及方差齐要求, 进行的是 Wilcoxon 秩和检验。

表 2 单位点 Armitage 趋势检验结果

SNP	基因型	病例组		对照组		Armitage 趋势检验		
		频数	频率	频数	频率	卡方值	P 值	Bonferroni 校正后显著性
rs1234314	CC	96	0.1928	100	0.1965	1.073	0.3003	不显著
	CG	217	0.4357	242	0.4754			
	GG	185	0.3715	167	0.3281			
rs45454293	TT	12	0.0241	11	0.0216	0.936	0.3332	不显著
	CT	101	0.2028	121	0.2377			
	CC	385	0.7731	377	0.7407			
rs3850641	GG	16	0.0321	13	0.0255	0.563	0.4533	不显著
	AG	121	0.2430	143	0.2809			
	AA	361	0.7249	353	0.6935			
rs1234313	GG	56	0.1124	68	0.1336	0.456	0.4996	不显著
	AG	213	0.4277	211	0.4145			
	AA	229	0.4598	230	0.4519			
rs3861950	TT	3	0.0060	8	0.0157	4.579	0.0324	不显著
	CT	63	0.1265	82	0.1611			
	CC	432	0.8675	419	0.8232			

表 3 调整混杂因素的 logistic 回归分析结果

	参数估计	标准误	Wald 卡方值	P 值	OR 值	95% 置信区间
常数项	-10.9110	1.2816	72.4791	<0.0001		
SEX	0.0315	0.2126	0.0219	0.8823	1.032	(0.680, 1.565)
AGE	-0.0022	0.0091	0.0570	0.8112	0.998	(0.980, 1.016)
BMI	-0.0275	0.0165	2.7759	0.0957	0.973	(0.942, 1.005)
SBP	0.0592	0.0088	45.6402	<0.0001	1.061	(1.043, 1.079)
DBP	-0.0201	0.0123	2.6854	0.1013	0.980	(0.957, 1.004)
TC	-0.7533	0.1103	46.6408	<0.0001	0.471	(0.379, 0.584)
TG	3.1624	0.2612	146.6198	<0.0001	23.628	(14.162, 39.421)
GLU	1.0092	0.1298	60.4918	<0.0001	2.744	(2.127, 3.538)
SNP1	-0.1364	0.1753	0.6056	0.4364	0.872	(0.619, 1.230)
SNP2	-0.0647	0.2066	0.0982	0.7540	0.937	(0.625, 1.405)
SNP3	0.0308	0.2149	0.0205	0.8862	1.031	(0.677, 1.571)
SNP4	0.0250	0.1814	0.0190	0.8903	1.025	(0.719, 1.463)
SNP5	-0.5108	0.2899	3.1052	0.0780	0.600	(0.340, 1.059)

注: SNP1~5 依次表示的是 SNP 位点 rs1234314、rs45454293、rs3850641、rs1234313 和 rs3861950, 下表同。

2.3 调整混杂因素的 logistic 回归分析

将混杂因素(协变量)及 5 个 SNP 位点引入 logistic 模型中, 结果显示, 在调整混杂因素的影响后, 5 个 SNP 位点均无统计学意义(表 3)。

2.4 单体型分析

单体型分析结果显示, CTAGT($P=0.0006$)、CTAAC($P=0.0123$)、CCAGT($P=0.0004$)、GTGGT($P=$

0.0329)、GCGAC($P<0.0001$)以及 GCAAC($P=0.0173$)这 6 个单体型在病例组和对照组中的频率差异具有统计学意义(表 4)。

2.5 主成分的 logistic 回归分析

主成分 logistic 分析结果见表 5。根据主成分提取准则, 提取第一主成分和第二主成分用于构建 logistic 回归模型(第一、第二主成分的特征值均大于 1, 累积

表4 单体型频率估计及显著性水平

单体型	频率			卡方值	P 值 ^a
	病例组	对照组	合并组		
CTGGT	0.03367	0.02510	0.02931	1.3005	0.2541
CTGGC	0.00298	0.00771	0.00455	2.8017	0.0942
CTGAT	0.00189	0.00380	0.00286	0.6500	0.4201
CTGAC	0.04139	0.03551	0.03900	0.4826	0.4873
CTAGT	0.00000	0.01166	0.00577	11.9370	0.0006
CTAGC	0.00466	0.00590	0.00569	0.1977	0.6566
CTAAT	0.00000	0.00000	0.00000	0.0023	0.9618
CTAAC	0.00313	0.01258	0.00734	6.2652	0.0123
CCGGT	0.02262	0.02199	0.02228	0.0093	0.9232
CCGGC	0.01085	0.01747	0.01463	1.5572	0.2121
CCGAT	0.00000	0.00000	0.00000	0.00001	0.9942
CCGAC	0.02887	0.01624	0.02305	3.5929	0.0580
CCAGT	0.00338	0.02055	0.01206	12.4507	0.0004
CCAGC	0.20399	0.17930	0.19157	1.9813	0.1592
CCAAT	0.00016	0.00277	0.00158	2.2089	0.1372
CCAAC	0.05308	0.07360	0.06287	3.6110	0.0574
GTGGT	0.00004	0.00497	0.00272	4.5510	0.0329
GTGGC	0.00254	0.00017	0.00120	2.3789	0.1230
GTGAT	0.00000	0.00000	0.00000	0.0010	0.9750
GTGAC	0.00131	0.00149	0.00181	0.2012	0.6537
GTAGT	0.00000	0.00124	0.00039	2.2859	0.1306
GTAGC	0.00087	0.00120	0.00097	0.0655	0.7980
GTAAT	0.00000	0.00004	0.00002	0.0371	0.8472
GTAAC	0.03305	0.02910	0.03144	0.2674	0.6051
GCGGT	0.00217	0.00118	0.00141	0.4445	0.5050
GCGGC	0.00158	0.00194	0.00231	0.2885	0.5912
GCGAT	0.00000	0.00000	0.00000	0.0003	0.9853
GCGAC	0.00373	0.02844	0.01475	21.4562	<0.0001
GCAAT	0.00536	0.00152	0.00384	2.0303	0.1542
GCAGC	0.03161	0.03897	0.03497	0.8150	0.3667
GCAAT	0.00000	0.00145	0.00069	1.5313	0.2159
GCAAC	0.50711	0.45411	0.48093	5.6681	0.0173

注: ^a P 值是根据 100 次置换检验确定的。

贡献率达 68.66%)。将第一、第二主成分(PC1、PC2)及协变量(SEX、AGE、BMI、SBP、DBP、TC、TG 和 GLU)作为自变量引入 logistic 回归模型,结果显示,在调整了其他协变量的情况下,第一主成分(PC1)与冠心病具有统计学关联性($P=0.0236$)(表 5)。因此,在中国汉族群体中, *TNFSF4* 基因与冠心病存在关联性。

第一主成分和第二主成分对应 SNP 位点的线性

组合系数见表 6。其中,第一主成分对应 5 个 SNP 位点的系数在 0.36 到 0.49 之间,系数比较平均且都比较大。

3 讨论

在病例对照关联分析中,造成结论不一致的原因有群体遗传背景不同、群体亚结构、遗传异质性、样本含量不足、所使用的统计分析方法的检验效能

表 5 主成分的 logistic 回归分析结果

	参数估计	标准误	Wald 卡方值	P 值	OR 值	95% 置信区间
常数项	-10.8890	1.2789	74.7431	<0.0001	.	.
SEX	-0.00180	0.0089	0.0219	0.8824	1.032	(0.682, 1.560)
AGE	0.0375	0.2106	0.0520	0.8196	0.998	(0.981, 1.016)
BMI	-0.0254	0.0173	2.4025	0.1211	0.974	(0.943, 1.007)
SBP	0.0579	0.0087	45.4354	<0.0001	1.061	(1.043, 1.079)
DBP	-0.0200	0.0122	2.7080	0.0998	0.980	(0.957, 1.004)
TC	-0.7527	0.1099	47.0194	<0.0001	0.470	(0.379, 0.584)
TG	3.1637	0.2594	147.1570	<0.0001	23.545	(14.133, 39.224)
GLU	0.9828	0.1296	60.4567	<0.0001	2.743	(2.127, 3.537)
PC1	-0.0052	0.0415	5.1253	0.0236	0.870	(0.771, 0.981)
PC2	-0.0175	0.0676	0.0034	0.9535	0.995	(0.845, 1.172)

表 6 第一、第二主成分对应 SNP 位点线性组合系数

SNP 位点	PC1	PC2
SNP1	0.485581	-0.382195
SNP2	0.367654	0.574241
SNP3	0.446797	0.428990
SNP4	0.444312	-0.577462
SNP5	0.481665	0.081729

不足等多种原因。众所周知,在关联分析中经常出现的一个统计学问题是假阴性,这主要是由两方面的原因造成:一方面是选取的样本含量不够。在复杂疾病的发生过程中,单个SNP位点的作用是微效的,如果选取的样本含量不够,就无法检验出关联性,造成假阴性结果。另一方面是统计方法的选择。目前,在基于群体的病例对照关联分析中有两种分析策略。一类是以单个SNP为单位的统计分析方法。常用的是单个位点卡方检验^[12],其中,用的较多的是基于加性模型的Armitage趋势检验^[13],它不要求H-W平衡。随着SNP分型技术的发展,研究中所涉及到的SNP位点数目越来越多,该方法所面临的一个巨大挑战就是多重检验,用Bonferroni校正又过于保守^[7],logistic回归分析^[14]能调整混杂因素的影响,在一定程度上回避了多重校正的问题,但是SNP之间的多重共线性以及所分析的SNP数目较大时带来的高纬度均可导致logistic回归分析的检验效能不高^[15]。另一类是以单体型为单位的分析方法。此类分析方法通过比较病例组和对照组之间的单体型频率的差异揭示特定单体型与疾病的关联性。其统计方法仍然主要采用卡方检验方法和logistic回归分析方

法。然而这类方法的弊端在于单体型推断的复杂性和当群体不满足哈迪温伯格平衡(HWE)时的不确定性。

为提高基于群体的病例对照研究设计的关联分析的有效性,一些研究小组提出了将主成分分析和logistic回归分析两者结合起来的检验方法——基于主成分的logistic回归分析方法^[7-10]。研究结果表明,该方法比基于单个SNP位点或者基于单体型的方法更有效^[7-10]。它分为两步:第一步是提取主成分,基于某候选基因区域内SNP位点的连锁不平衡信息,采用主成分分析提取其综合信息;第二步是将主成分得分(SNP位点的线性组合)作为自变量引入logistic回归模型,借助logistic回归模型研究基因与疾病之间的关联性。这种基于主成分的logistic回归分析方法有许多优点。一方面由于提取的主成分之间是正交的,这就避免了多重共线性问题;另一方面,选取少数几个有代表性的主成分能达到降维的目的,同时也能减少因多重检验所致的假阳性结果。当位点之间的连锁不平衡较弱时,多个位点提供的信息要比单个位点多,并且一个基因上的多个位点间通常存在很强的交互作用^[16],基于主成分的logistic回归分析方法正是考虑多个位点间连锁不平衡信息的多位点分析方法,因而在关联分析中备受青睐^[7-10]。

目前国内外对将该方法应用于病例对照关联分析中的报导较少。我们在采用传统统计分析方法基础上,引进基于主成分的logistic回归分析方法来研究 *TNFSF4* 基因与冠心病之间的关联性。本文基于主成分的logistic回归分析结果显示,在中国汉族人

群中, *TNFSF4* 基因与冠心病之间是存在关联性的。单体型分析的结果也支持该结论。

本文结果与瑞典人群病例对照关联分析结果相一致, 进一步扩充了国内外对 *TNFSF4* 基因与冠心病之间关联性的认识。瑞典人群中存在相关性而德国人群中不存在相关性提示不同人群的遗传结构对关联分析的结果可能有影响, 这还有待于在更多人群中得到验证。

此外, 由第一主成分和第二主成分对应 SNP 位点的系数表(表 6)可知, 位点 rs1234314(SNP1)在第一主成分中的系数为 0.485581, 是 5 个系数中最大的, 因而该位点对第一主成分的贡献最大。位点 rs3861950(SNP5)在第一主成分中的系数为 0.481665, 也比较大, 并且在单点 Armitage 检验中, 该位点具有统计学意义($P=0.3237$), 这可能是第一主成分在多因素分析中显著的主要原因。但在本研究中, 我们还无法确定哪个位点是起主要作用的功能位点, 这有待于进行进一步深入研究。

参考文献(References):

- [1] Wang XS, Ria M, Kelmenson PM, Eriksson P, Higgins DC, Samnegård A, Petros C, Rollins J, Bennet AM, Wiman B, de Faire U, Wennberg C, Olsson PG, Ishii N, Sugamura K, Hamsten A, Forsman-Semb K, Lagercrantz J, Paigen B. Positional identification of *TNFSF4*, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nat Genet*, 2005, 37(4): 365–372. [\[DOI\]](#)
- [2] Lazta U, Dürkop H, Schnittger S, Ringeling J, Eitelbach F, Hummel M, Fonatsch C, Stein H. The human OX40 homolog: cDNA structure, expression and chromosomal assignment of the ACT35 antigen. *Eur J Immunol*, 1994, 24(3): 677–683. [\[DOI\]](#)
- [3] Koch W, Hoppmann P, Mueller JC, Schömig A, Kastrati A. Lack of support for association between common variation in *TNFSF4* and myocardial infarction in a German population. *Nat Genet*, 2008, 40(12): 1386–1387. [\[DOI\]](#)
- [4] Olofsson PS, Söderström LÅ, Jern C, Sirsjö A, Ria M, Sundler E, de Faire U, Wiklund PG, Öhrvik J, Hedin U, Paulsson-Berne G, Hamsten A, Eriksson P, Hansson GK. Genetic variants of *TNFSF4* and risk for carotid artery disease and stroke. *J Mol Med*, 2009, 87(4): 337–346. [\[DOI\]](#)
- [5] M larstig A, Eriksson P, Rose L, Diehl KA, Hamsten A, Ridker PM, Zee RY. Genetic variants of tumor necrosis factor superfamily, member 4 (*TNFSF4*), and risk of incident atherothrombosis and venous thromboembolism. *Clin Chem*, 2008, 54(5): 833–840. [\[DOI\]](#)
- [6] Ria M, Eriksson P, Boquist S, Ericsson CG, Hamsten A, Lagercrantz J. Human genetic evidence that OX40 is implicated in myocardial infarction. *Biochem Biophys Res Commun*, 2006, 339(3): 1001–1006. [\[DOI\]](#)
- [7] Zhang FY, Wagener D. An approach to incorporate linkage disequilibrium structure into genomic association analysis. *J Genet Genomics*, 2008, 35(6): 381–385. [\[DOI\]](#)
- [8] Oh S, Park T. Association tests based on the principal-component analysis. *BMC proceedings*, 2007, 1(Suppl. 1): S130. [\[DOI\]](#)
- [9] Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol*, 2007, 31(5): 383–395. [\[DOI\]](#)
- [10] Wang K, Abbott D. A principal components regression approach to multilocus genetic association studies. *Genet Epidemiol*, 2008, 32(2): 108–118. [\[DOI\]](#)
- [11] Peng QQ, Zhao JH, Xue FZ. PCA-based bootstrap confidence interval tests for gene-disease association involving multiple SNPs. *BMC Genet*, 2010, 11: 6. [\[DOI\]](#)
- [12] Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics*, 1997, 53(4): 1253–1261. [\[DOI\]](#)
- [13] Slager SL, Schaid DJ. Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Human Hered*, 2001, 52(3): 149–153. [\[DOI\]](#)
- [14] Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev*, 2006, 7(10): 781–791. [\[DOI\]](#)
- [15] Schaid DJ, McDonnell SK, Hebring SJ, Cunningham JM, Thibodeau SN. Nonparametric tests of association of multiple genes with human disease. *Am J Human Genet*, 2005, 76(5): 780–793. [\[DOI\]](#)
- [16] Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Ame J Human Genet*, 2002, 70(2): 425–434. [\[DOI\]](#)