

胃癌易感基因筛选及多基因危险度分析

李鹏飞, 冯靖宇, 严滢滢, 符刚, 沈孝兵

摘要: [目的] 探讨我国汉族人群胃癌患者遗传易感基因, 并进行多基因危险度分析。[方法] 采用 1:1 配对病例对照研究方法, 收集南京市汉族人群原发性胃癌患者 585 例和相应非肿瘤及非消化道疾病患者为研究对象, 应用限制性片段长度多态性聚合酶链反应 (PCR-RFLP) 和基因特异性聚合酶链反应 (AS-PCR) 技术分析 *CYP2E1*、*GSTM1*、*GSTT1*、*NAT2*、*ALDH2*、*MTHFR*、*XRCC1*、*IL-1 β* 、*VDR*、*TNF* 等基因型别; 以条件 logistic 回归模型对基因及基因间的交互作用进行分析, 选出易感基因, 多基因联合作用危险度分析统计模型对选出的危险因素进行多基因危险度评价。[结果] 原发性胃癌遗传易感因素有 8 项, 分别是 *CYP2E1* (*c1/c1*)、*NAT2* 表型 (慢乙酰化型)、*MTHFR*1298C (*A/C*)、*IL-1 β* (*C/T*)、*NAT2M2* (*A/A*)、*XRCC1194* (*T/T*)、*NAT2M1* (*T/T*)、*VDR TaqI* (*T/T*)。利用多基因联合作用危险度分析模型对多基因危险度评价分析, 可以更直观地发现多基因组合的 *OR* 值与其基因频率存在高度相关性, 即随易感基因的增加, 易感基因组合危险度分布曲线会向更加危险的方向移动。并呈现一定的量化关系。[结论] 通过对筛选出的易感基因进行多基因危险度分析, 可更有效地推进对汉族人群中胃癌的高危人群识别及对其采取预防和干预措施。

关键词: 胃癌; 基因多态性; 多基因; 危险度分析

Screening of Susceptibility Genes and Multi-gene Risk Analysis in Gastric Cancer LI Peng-fei, FENG Jing-yu, YAN Ying-ying, FU Gang, SHEN Xiao-bing (Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing 210009, China). Address correspondence to SHEN Xiao-bing, E-mail: xb.shen@seu.edu.cn

Abstract: [Objective] To explore the relations between genetic polymorphism and susceptibility to gastric cancer in Chinese population, and to study the action of genetic factors in development of gastric carcinoma. [Methods] A case-control study was performed in 585 controls without tumor matched to 585 original gastric cancer patients in Nanjing city. The genotypes of *CYP2E1*, *GSTM1*, *GSTT1*, *NAT2*, *ALDH2*, *MTHFR*, *XRCC1*, *IL-1 β* , *VDR* and *TNF* were detected by molecular biological techniques (PCR-RFLP and AS-PCR). Sole gene and gene-gene interactions were analyzed by Logistic regression model, and the multi-gene risk of susceptibility genes were analyzed by a multi-gene risk analysis statistic model proposed by Demchuk E. [Results] The genetic risk factors associated with primary gastric cancer susceptibility were *CYP2E1*(*c1/c1*), *NAT2* phenotype (slow acetylase), *MTHFR*1298C (*A/C*), *IL-1 β* (*C/T*), *NAT2M2* (*A/A*), *XRCC1194* (*T/T*), *NAT2M1* (*T/T*), *VDR TaqI* (*T/T*). The multi-gene risk analysis showed there was close relation between the *OR* of polygene combination and the genotype frequency. The *OR* increased along with the susceptibility gene. [Conclusion] This new analyze model is helpful to screen and protect the high-risk group of gastric cancer.

Key Words: gastric cancer; polymorphism; Multi-Genes; risk analysis

胃癌是严重危害人类健康的疾病。尽管近年来胃癌的发病率在北美和欧洲呈现下降趋势, 但在全球范围内胃癌依然是第 4 大常见肿瘤和第 2 大常见肿瘤相关死因, 仍是一个严重的公共健康问题。据报道全球的胃癌病例将持续增加到 2050 年^[1-2], 而中国是世界上胃癌发病率和死亡率高的国家之一^[3]。大量流行病学研究证实胃癌的发生是一个多基因、多步骤、多

因素参与的复杂过程^[4], 早期筛检胃癌的危险因素是预防胃癌及降低其发生率的重要手段。

本研究试图通过查阅国内外最新文献, 选取当前研究比较多, 对胃癌发生发展影响相对较大的基因作为我们研究的目的基因, 主要涉及的基因为代谢酶基因, 包括 *GSTM1*、*GSTT1*、*CYP2E1*、*NAT*、*MTHFR*、*ALDH2*^[5-8]等; DNA 修复功能相关的基因如 *XRCC1*^[9]; 与炎症反应相关的细胞因子基因及一些与细胞周期调控功能相关的基因如 *IL-1 β* 、*TNF*^[10-11]等。首先, 对这些基因结合一些环境因素 (如吸烟、饮酒等) 进行单因素和多因素分析选出易感基因; 然后, 对其交互作用进行分析; 最后, 拟采用多基因联合作用的分析模型评估胃癌中多基因的组合作用, 以进一步推进易感人群的识别。

[基金项目] 江苏省自然科学基金项目 (编号: BK2009283)

[作者简介] 李鹏飞 (1985—), 男, 硕士生; 研究方向: 环境流行病学;

E-mail: pengfeili1985@sina.com

[通信作者] 沈孝兵教授, E-mail: xb.shen@seu.edu.cn

[作者单位] 东南大学公共卫生学院环境医学工程教育部重点实验室, 江苏 南京 210009

1 对象与方法

1.1 研究对象

病例组:收集于东南大学附属中大医院、江苏省肿瘤医院、南京鼓楼医院、南京市第二医院等自 2005 至 2010 年均经过组织病理学检查确诊为原发性胃癌的患者 585 例。取样前未进行过化疗或放疗。年龄为 18~90 岁[(61.15±12.61)岁];汉族,男性 429 例,女性 156 例。对照组:按 1:1 配比,在医院内随机选取与患者同性别、年龄差别±5 岁、同民族、同居住地和居住年限相近的非肿瘤及非消化道患者为对照。

1.2 外周血基因组 DNA 提取及基因多态性检测

抽取研究对象静脉血 3 mL 加 乙二胺四乙酸二钠盐(EDTA)抗凝-20℃低温保存。采用蛋白酶 K 消化法,异丙醇、乙醇抽提 DNA,紫外分光光度计定量,4℃保存备用。应用限制性片段长度多态性聚合酶链反应(PCR-RFLP)和基因持异性聚合酶链反应(AS-PCR)技术分析基因的多态性^[12-13]。

1.3 数据分析

用 Epidata 3.0 建立数据库,逻辑纠错后应用 SPSS 13.0 统计软件进行分析。病例组和对照组的基因型频率差异以 χ^2 检验确定,检验水准 $\alpha=0.05$ 。以比值比(odds ratio, OR)以及 95% 可信区间(95% confidence interval, 95%CI)表示各种基因型与胃癌风险的相关性。OR 值以非条件的 logistic 回归模型计算,并对年龄、性别等混杂因子进行校正。最后参考 DEMCHUK 建立的多基因联合作用危险度分析模型对筛选出的基因作多基因危险度分析^[14]。

2 结果

2.1 胃癌遗传因素的单因素条件 logistic 回归分析

运用分层的 logistic 回归模型所作胃癌遗传因素的单因素分析,表 1 显示,在 $\alpha=0.05$ 的水平上,筛选出的危险因素有:CYP2E1、NAT2M1、NAT2M2、NAT2 表型、XRCC1194、MTHFR A1298C、IL-1 β 、VDR TaqI 等 8 项,提示这些危险因素可能与胃癌的易感性有关。

表 1 胃癌遗传因素单因素条件 logistic 回归分析结果

基因型(Gene)	b	S _b	χ^2	P	OR(95%CI)
GSTM1	0.2294	0.1184	3.7517	0.0528	1.258 (0.997~1.568)
GSTT1	0.0149	0.1186	0.0591	0.4417	1.096 (0.868~1.382)
CYP2E1	0.2742	0.0932	8.6542	0.0033	1.315 (1.096~1.579)
NAT2M1	0.6862	0.1440	22.7033	<0.0001	1.986 (1.498~2.634)
NAT2M2	0.4114	0.1023	16.1823	<0.0001	1.509 (1.235~1.844)
NAT2M3	0.0149	0.0996	0.0223	0.8812	1.015 (0.835~1.234)
NAT2	0.7810	0.1959	15.9009	<0.0001	2.184 (1.488~3.206)
ALDH2	0.0479	0.0632	0.5748	0.4484	1.049 (0.927~1.187)
XRCC1194	0.5570	0.0994	31.3855	<0.0001	1.745 (1.436~2.121)
XRCC1399	0.0040	0.0872	0.6120	0.4340	1.071 (0.902~1.270)
MTHFR677	0.0591	0.0342	1.2569	0.1625	1.013 (0.889~1.195)
MTHFRA1298C	0.3207	0.1204	7.0994	0.0077	1.378 (1.088~1.745)
IL-1 β	0.1819	0.0851	4.5694	0.0325	1.199 (1.015~1.417)
VDR Apa I	0.0533	0.0944	0.3195	0.5719	1.055 (0.877~1.269)
VDR Taq I	0.4780	0.1892	6.3794	0.0115	1.613 (1.113~2.337)
TNF- β	0.0535	0.0845	0.4004	0.5269	1.055 (0.894~1.245)

2.2 原发性胃癌遗传危险因素的多因素条件 logistic 回归模型分析

对第一部分筛选出的危险项进行多因素条件 logistic 回归模型作进一步分析,表 2 显示,在 $\alpha=0.05$ 的水平上原发性胃癌的危险因素是:CYP2E1(c1/c1)、NAT2M1(T/T)、NAT2M2(A/A)、XRCC1194(T/T)、MTHFRA1298C(A/C)等。

表 2 多因素条件 logistic 回归分析结果

基因型(Gene)	b	S _b	χ^2	P	OR(95%CI)
CYP2E1(c1/c1)	0.3076	0.1011	9.2663	0.0023	1.360 (1.116~1.658)
NAT2M1(T/T)	0.7727	0.1596	23.4467	<0.0001	2.166 (1.584~2.961)
NAT2M2(A/A)	0.5133	0.1133	20.5347	<0.0001	1.671 (1.388~2.086)
XRCC1194(T/T)	0.5502	0.1051	27.4005	<0.0001	1.734 (1.411~2.310)
MTHFRA1298C(A/C)	0.3518	0.1304	7.2810	0.0070	1.422 (1.101~1.835)

2.3 原发性胃癌的交互作用分析

对原发性胃癌的遗传度危险因素交互作用分析,表 3 显示,以上 5 种基因除 NAT2M1 与 MTHFR1298(A/C)之间不存在交互作用外($P>0.05$),其他基因间均存在交互作用($P<0.05$),且表现为协同作用($OR>1$),提示同时携带这些有交互作用的基因可使患胃癌的风险性增加^[15-16]。

表 3 胃癌易感基因交互作用的 OR 值

基因型(Gene)	CYP2E1(c1/c1)	NAT2M1(T/T)	NAT2M2(A/A)	XRCC1194(T/T)	MTHFRA1298C(A/C)
CYP2E1(c1/c1)	—	—	—	—	—
NAT2M1(T/T)	6.873*	—	—	—	—
NAT2M2(A/A)	2.915*	11.592*	—	—	—
XRCC1194(T/T)	7.596*	12.163*	12.160*	—	—
MTHFR1298C(A/C)	3.702*	4.036	4.460*	6.581*	—

[注]*: $P<0.05$ 。

2.4 多基因联合危险度分析

尽管有关肿瘤与易感基因的关系研究已有很多,但如何评价多种易感基因的联合作用与肿瘤的患病风险一直是比较难的一个问题。目前还没有大家比较认可的方法,在本研究中参考 DEMCHUK 构建的多基因联合分析模型用于评价“哮喘的基因多态性与易感性关系”^[14],试图对单因素筛选出的 8 种易感基因的组合作用危险度做出初步分析。

在该研究选用的这个模型中,每一个遗传因素都被划分为含有易感基因或没有两种情况,如果人体存在某个遗传易感基因型用 X 表示,不存在则为 0,那么此 8 种基因型的组合共有 256 种组合。例如,(00000000)则表示未含有任何一种上述遗传易感基因,其 OR 值为 1;(XXXXXXXX)表示含有全部上述 8 种遗传易感基因型,其 OR 值为各单个基因型 OR 值的乘积,其基因频率为各单基因的基因频率的乘积,其中单基因频率是指收集的研究对象中某一基因类型在该总体中的分布。易感基因型的 OR 值和在对照人群中的基因频率见表 4。

表 4 易感基因在人群中的分布频率和 OR 值

Table 4 Susceptibility gene variants' frequencies and ORs

基因型 (Gene)	基因频率 (Frequency)	OR
CYP2E1	0.5348	1.315
NAT2M1	0.0068	1.986
NAT2M2	0.0359	1.509
NAT2	0.1572	2.184
XRCC1194	0.0667	1.745
MTHFRA1298C	0.3521	1.378
IL-1β	0.5641	1.199
VDR TaqI	0.9200	1.613

运用该模型初步评估多种易感基因对胃癌的组合作用, 结果图 1 列出了由上述 8 个易感基因全部组合的 OR 值与基因频率, 每一个点都是一个唯一的多基因型组合。由图 1 可以直观地观察到, 多基因联合的 OR 值与其基因频率是高度相关的, 大部分组合都集中在中间部位, OR 值较高的多基因组合虽然危险度较高, 但是其在人群的出现频率低, 因此对于筛选易感人群的意义不大。由图 2 可以直观地观察到, 随着易感基因的数目或其他危险因素的增加, OR 的回归曲线沿着危险度更大的方向移动。综合图 2 则可以参考实际情况, 选出一些基因作

为筛选易感人群的依据, 从而实现以较低的成本对易感人群保护的目的。另外随着样本数和筛选出的基因数的增加, 则可以拟合基因的累积频率和 OR 的关系曲线图, 这样就可以利用此模型对多基因作出定量的评价。于是就可以根据这一模型来分析哪些基因的组合危险度更大, 以做到对胃癌的预防及干预。

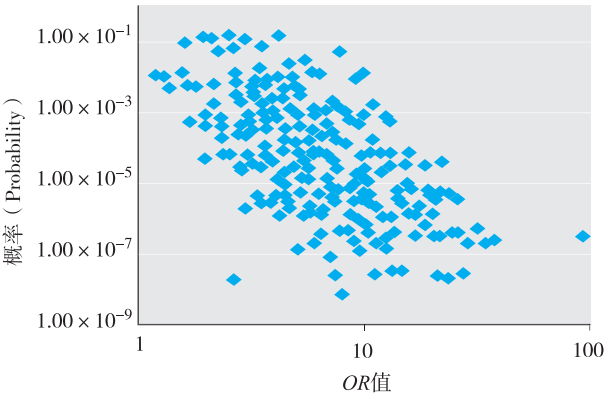
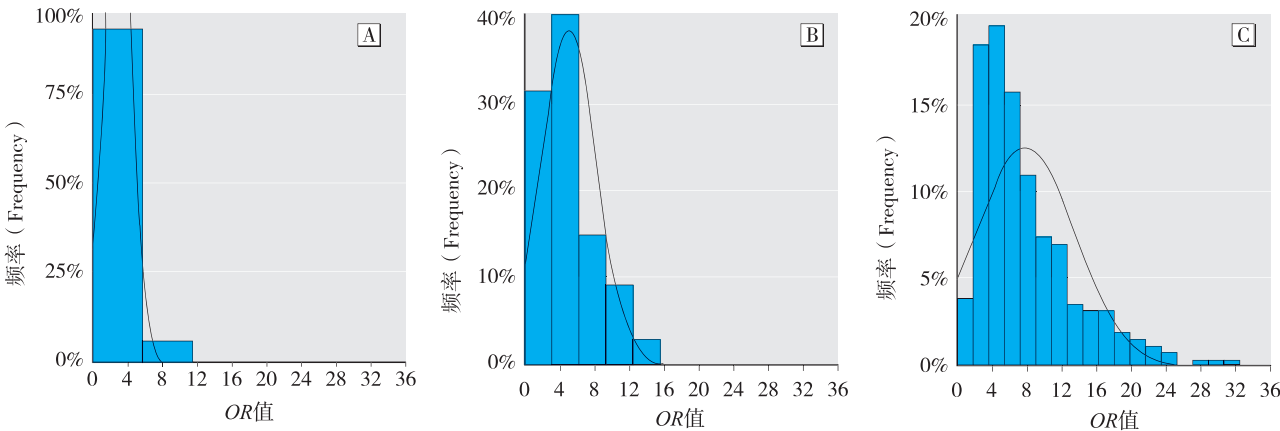


图 1 8 种易感基因任意组合的频率与 OR 值的分布图

Figure 1 Frequencies and ORs of genotypes calculated using 8 selected gene variants



[注]A.CYP2E1(c1/c1), NAT2 表型(慢乙酰化型), MTHFRA1298C(A/C), IL-1β(C/T)4 种基因组合后 ORs 的分布; B.CYP2E1(c1/c1), NAT2 表型(慢乙酰化型), MTHFRA1298C(A/C), IL-1β(C/T)NAT2M2(A/A), XRCC1194(T/T)6 种基因组合后 ORs 的分布; C.CYP2E1(c1/c1), NAT2 表型(慢乙酰化型), MTHFRA1298C(A/C), IL-1β(C/T)NAT2M2(A/A), XRCC1194(T/T), NAT2M1(T/T), VDR TaqI(T/T)8 种基因组合 [Distribution of relative disease risk calculated using gastric cancer associated gene variants.(1)contain four genes: CYP2E1(c1/c1), NAT2 phenotype(slow acetylator), MTHFRA1298C(A/C), and IL-1β(C/T); (2)NAT2M2(A/A), XRCC1194(T/T)with A; (3)NAT2M1(T/T), VDR TaqI(T/T)with B.]。

图 2 易感基因与危险因素 OR 回归曲线

Figure 2 Regression curve of OR in susceptibility genes with risk factors

3 讨论

该研究通过病例对照研究对 585 例南京市原发性胃癌及对照进行基因多态性分析, 筛选出的易感基因, 跟目前绝大多数研究报道的较为一致。为胃癌易感标志物的构建及高危人群的筛选提供一定的依据。但研究发现肿瘤有很多易感基因, 且有些易感因素的 OR 值较小, 如果直接用该研究我们筛选出的基因作为胃癌的筛选标志物, 筛选高危人群并对其做出干预, 实际上存在一定的困难。因此, 如何从整体上对筛选出的易感基因作出分析并应用到胃癌预防中去, 是目前研究的一个重点同

时也是一个难点。

以往利用 logistic 回归分析方法来研究单核苷酸多态性 (SNPs) 之间的基因 - 基因交互作用, 因为 OR 值较低对模型的参数的结果很难准确下结论。更重要的是在研究多位点基因 - 基因间的交互作用时, SNPs 的位点每增加一个, 样本量将呈指数级的上升。即便如此, 分布在高维空间中的数据仍显得相对稀疏, 有可能出现某些基因型组合没有观察值, 这就是近来提到的维度困扰^[17]。该研究采用 DEMCHUK 建立的多基因联合作用危险度分析模型来定性和定量相结合地评价多基因的

危险度。从结果可以直观的观察到每增加一种易感基因或保护基因对其总体危险度的影响情况。另外还可以观察到任意基因组合的危险性大小,因此可以通过这一模型分析结果合理选出一些基因或者其组合作为筛选胃癌高危人群的遗传标志物。

近年来有关基因筛选和评价模型的报道出现很多。如最近提出的关联规则筛选疾病相关的 SNP 位点及其组合的分析方法。邹莉玲等^[18]利用关联规则筛选肺癌有关的 SNPs 多基因位点及组合效果较好。另外研究者还提出一种新的评价多基因的危险度方法——“广义多因子降维分析法”。2001 年 RITCHIE 等^[19]首次提出了多因子降维法(multifactor dimensionality reduction, MDR)。但由于该法不能用于数量性状的研究,为此 LOU 等^[20]在 2007 年提出了一种基于 MDR 基本原理的扩展方法——广义多因子降维法(generalized multifactor dimensionality reduction, GMDR),又称基于计分的多因子降维法(score-based MDR)。该法可以通过将广义线性模型的概念引入到 MDR 中,使其不但能够分析连续变量,且能够纳入协变量,从而控制协变量引起的干扰,提高预测的准确度。GMDR 能够处理多种样本类型和结局变量类型,与其他连续变量交互作用分析方法相比具有一定优势。但目前这些方法还处于实验阶段,还有待进一步完善。

参考文献:

- [1] NEUGUT AI, HAYEK M, HOWE G. Epidemiology of gastric cancer [J]. *World J Gastroenterol*, 2006, 12(3): 354-362.
- [2] FORMAN D, BURLEY VJ. Gastric cancer: global pattern of the disease and an overview of environmental risk factors [J]. *Best Pract Res Clin Gastroenterol*, 2006, 20(4): 633-649.
- [3] HOHENBERGER P, GRETSCHER S. Gastric cancer [J]. *Lancet*, 2003, 362(9380): 305-315.
- [4] CHEN B, CAO L, ZHOU Y, et al. Glutathione S-transferase T1 (GSTT1) gene polymorphism and gastric cancer susceptibility: a meta-analysis of epidemiologic studies [J]. *Dig Dis Sci*, 2010, 55(7): 1831-1838.
- [5] COLOMBO J, ROSSIT AR, CAETANO A, et al. GSTT1, GSTM1 and CYP2E1 polymorphisms in gastric cancer and chronic gastritis in a Brazilian population [J]. *World J Gastroenterol*, 2004, 10(9): 1240-1245.
- [6] AL-MOUNDHRI M S, AL-KINDI M, AL-NABHANI M, et al. NAT2 polymorphism in Omani gastric cancer patients-risk predisposition and clinic pathological associations [J]. *World J Gastroenterol*, 2007, 13(19): 2697-2702.
- [7] DE RE V, CANNIZZARO R, CANZONIERI V, et al. MTHFR polymorphisms in gastric cancer and in first-degree relatives of patients with gastric cancer [J]. *Tumour Biol*, 2010, 31(1): 23-32.
- [8] YOU WC, HONG JY, ZHANG L, et al. Genetic polymorphisms of CYP2E1, GSTT1, GSTP1, GSTM1, ALDH2, and ODC and the risk of advanced precancerous gastric lesions in a Chinese population [J]. *Cancer Epidemiol Biomarkers Prev*, 2005, 14(2): 451-458.
- [9] 符刚, 沈孝兵. DNA 修复酶基因多态性与胃癌易感性关系的研究进展 [J]. *环境与职业医学*, 2007, 24(6): 628-632.
- [10] MURPHY G, THORNTON J, MCMANUS R, et al. Association of gastric disease with polymorphisms in the inflammatory related genes IL-1B, IL-1RN, IL-10, TNF and TLR4 [J]. *Eur J Gastroenterol Hepatol*, 2009, 21(6): 630-635.
- [11] SUGIMOTO M, YAMAOKA Y, FURUTA T. Influence of interleukin polymorphisms on development of gastric cancer and peptic ulcer [J]. *World J Gastroenterol*, 2010, 16(10): 1188-1200.
- [12] SHEN X, ZHANG J, YAN Y, et al. Analysis and estimates of the attributable risk for environmental and genetic risk factors in gastric cancer in a Chinese population [J]. *J Toxicol Environ Health A*, 2009, 72(11-12): 759-766.
- [13] YAN L, YANAN D, DONGLAN S, et al. Polymorphisms of XRCC1 gene and risk of gastric cardiac adenocarcinoma [J]. *Dis Esophagus*, 2009, 22(5): 396-401.
- [14] DEMCHUK E, YUCESAY B, JOHNSON VJ, et al. A statistical model for assessing genetic susceptibility as a risk factor in multifactorial diseases: lessons from occupational asthma [J]. *Environ Health Perspect*, 2007, 115(2): 231-234.
- [15] 梁戈玉, 浦跃朴, 尹立红. 基因多态性在肺癌发生中交互作用 [J]. *中国公共卫生*, 2007, 23(8): 902-903.
- [16] LI H, CHEN XL, LI HQ. Polymorphism of CYP1A1 and GSTM1 genes associated with susceptibility of gastric cancer in Shandong province China [J]. *World J Gastroenterol*, 2005, 11(37): 5757-5762.
- [17] MOORE JH, RITCHIE MD. The challenges of whole genome approaches to common diseases [J]. *JAMA*, 2004, 291(13): 1642-1643.
- [18] 邹莉玲, 赵耐青, 秦国友, 等. 应用关联规则筛选疾病相关的 SNP 位点及其组合的分析方法 [J]. *中国卫生统计*, 2009, 26(3): 226-228, 233.
- [19] RITCHIE MD, HAHN LW, ROODI N, et al. Multifactor-dimensionality reduction reveals high order interactions among estrogen metabolism genes in sporadic breast cancer [J]. *Am J Hum Genet*, 2001, 69(1): 138-147.
- [20] LOU XY, CHEN GB, YAN L, et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence [J]. *Am J Hum Genet*, 2007, 80(6): 1125-1137.

(收稿日期: 2011-03-01)

(英文编审: 金克峙; 编辑: 张晶; 校对: 王晓宇)