

## · 现场调查 ·

# 多因子降维法在人群散发性结直肠癌交互作用分析中的应用

金明娟 刘冰 张爽爽 张勇晶 徐玫 马新源 姚开颜 陈坤

**【摘要】** 目的 探讨 DNA 修复基因多态性与中国南方汉族人群散发性结直肠癌发病的相关性, 验证多因子降维法(MDR)应用于多因子疾病基因-基因、基因-环境交互作用分析的可行性。方法 采用自然人群为基础的病例对照研究设计, 运用 PCR-RFLP 方法对 206 例结直肠癌病例和 845 例正常对照开展 OGG1 Ser326Cys, XRCC1 Arg194Trp、Arg280His 和 Arg399Gln, XPD Lys751Gln 和 XRCC3 Thr241Met 等 DNA 修复体系常见单核苷酸多态性(SNP)的检测分型。结果 个体特征与结直肠癌的关联分析表明, 年龄与结直肠癌的发病正相关, 高年龄组( $\geq 61$ 岁)与低年龄组( $\geq 42$ 岁至 $< 61$ 岁)相比, 结直肠癌患病风险增高有统计学意义(校正 OR = 2.04, 95% CI: 1.49~2.80); 家族肿瘤史同样与结直肠癌的发病存在有统计学意义的正相关关系(校正 OR = 1.51, 95% CI: 1.05~2.17)。前述各 SNP 的等位基因和基因型分布频率在正常对照组和病例组间差异均无统计学意义( $P > 0.05$ )。采用 MDR 对基因-基因、基因-环境交互作用模型的筛选分析表明, 最佳的交互作用模型包含了年龄分布、饮酒史、XRCC1 Arg194Trp 和 OGG1 Ser326Cys 等 4 个因子(平均检验准确度 = 0.616, 交叉验证一致性 = 10/10,  $P = 0.011$ ); 进一步以筛选出的低风险组合为参照, logistic 拟合分析发现高风险组合可以使结直肠癌的患病风险增高并有统计学意义(OR = 2.72, 95% CI: 1.66~4.47)。结论 DNA 修复基因多态性对中国人散发性结直肠癌风险的遗传影响符合低外显性特征, 并与环境因子可能存在复杂的联合作用。

**【关键词】** 结直肠肿瘤; 基因多态性, 单核苷酸多态性; DNA 修复基因; 多因子降维

**Application of multifactor dimensionality reduction on the interactions between gene-gene, gene-environment and the risk sporadic colorectal cancer in Chinese population** JIN Ming-juan\*, LIU Bing, ZHANG Shuang-shuang, ZHANG Yong-jing, XU Mei, MA Xin-yuan, YAO Kai-yan, CHEN Kun. \*Department of Epidemiology and Health Statistics, Zhejiang University School of Medicine, Hangzhou 310058, China

Corresponding author: CHEN Kun, Email: CK@zju.edu.cn

**【Abstract】 Objective** To identify the association between risk of sporadic colorectal cancer and the common single nucleotide polymorphisms (SNPs) in DNA repairs genes, gene to gene interactions among them and their gene to environment interactions with common environmental factors. **Methods** In this population-based case-control study, 206 primary colorectal cancer cases and 845 cancer-free healthy controls were enrolled. Genotyping was carried out using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) technique, with the status of subjects case or controls unknown. Multifactor dimensionality reduction (MDR) and logistic analysis were both used for association analysis. **Results** As compared to the younger age group ( $\geq 42$ ,  $< 61$  years), the risk of colorectal cancer in older age group ( $\geq 61$  years) increased significantly (OR = 2.04, 95% CI: 1.49-2.80). Similar result was observed in the family cancer history (OR = 1.51, 95% CI: 1.05-2.17). However, no significant association between any single DNA repair gene SNP and colorectal cancer risk was discovered. Results from MDR analysis only showed a significant interaction among the four following factors: age, alcohol drinking, XRCC1 Arg194Trp and OGG1 Ser326Cys (the cross-validation consistency = 10/10, the average testing accuracy = 0.616,  $P = 0.011$ ). Using a logistic regression model, the "high-risk" individuals had a significantly elevated risk of colorectal cancer compared to those "low-risk" individuals

基金项目: 国家自然科学基金资助项目(30471492); 浙江省自然科学基金资助项目(R205319)

作者单位: 310058 杭州, 浙江大学医学院公共卫生学院流行病学与卫生统计学系(金明娟、刘冰、张爽爽、张勇晶、陈坤); 义乌出入境检验检疫局(徐玫); 浙江省嘉善县肿瘤防治所(马新源、姚开颜)

通讯作者: 陈坤, Email: CK@zju.edu.cn

classified by the above MDR model ( $OR = 2.72$ ,  $95\% CI: 1.66-4.47$ ). **Conclusion** The impact of polymorphisms in DNA repair genes on the risk of sporadic colorectal cancer exhibited a low-penetrance characteristics while the intricate interactions existing among them and with environmental factors.

**【Key words】** Colorectal neoplasm; Genetic polymorphism, Single nucleotide polymorphism; DNA repair gene; Multifactor dimensionality reduction

近年来结直肠癌在我国的发病迅速上升,据估计 2005 年结直肠癌年龄标化发病率男性和女性将分别达到 15.0/10 万和 9.7/10 万,居常见恶性肿瘤发病率第 5 位<sup>[1]</sup>。结直肠癌同样是多因子疾病:一方面,生活方式、饮食习惯等外环境因素在其发生发展过程中起重要作用;另一方面,并非所有接触环境危险因素的个体都伴随着结直肠癌的发生,个体对环境危险因素的作用存在着易感性,不同个体间遗传背景的差异决定着这种疾病易感性的差异。美国国家环境卫生科学研究所于 1997 年 10 月提出了环境基因组计划,纳入的 10 个大类候选基因中 DNA 修复基因居首位<sup>[2,3]</sup>。多因子疾病病因学研究中,有关交互作用的识别和定量分析,logistic 回归是最为经典的方法,然而其应用往往受到统计学检验效能和样本量的限制。因此,近年来,国内外学者们都在不断尝试提出新的分析方法和思路。多因子降维法 (multifactor dimensionality reduction, MDR) 是由 Ritchie 等<sup>[4]</sup> 在研究散发性乳腺癌时首先提出,目前已成为一种颇有应用前景的分析多个因子间交互作用的方法。本研究的主要目的是通过实际的应用研究,明确 DNA 修复基因多态性与国人结直肠癌发病相关性的同时,明确 MDR 用于多因子疾病基因-基因、基因-环境交互作用分析的实际可行性和应用前景。

### 材料与方法

1. 研究对象:以全国结直肠癌调整死亡率最高县——浙江省嘉善县为研究现场,以该县 1990 年 4 月建立的包括 10 个乡镇的当地居民 (64 693 人) 随访队列为研究人群。选择队列内 1990 年 5 月至 2005 年 5 月发生的原发性结直肠癌存活病例,排除直肠类癌、继发性结直肠癌、1989-1990 年筛查时检出的及不属于队列人群的原发性结直肠癌病例,最终纳入有效病例 206 例 (结肠癌 93 例,直肠癌 113 例)。对照组 (队列内正常人群的代表性样本) 通过 2002 年和 2005 年分别开展的两次抽样调查获得:其中,2002 年通过单纯随机抽样方法抽取个体 400 例,排除死亡 16 例和结直肠癌病例 1 例,共需调查 383 例,实际共调查 343 例;2005 年在基线资

料库中,剔除死亡和恶性肿瘤发病个体后,以性别和年龄为分层因素,采用分层随机抽样抽取调查对象 550 例,最终有效调查个体 502 例;经比较分析,两组对照人群在年龄和性别等基本因素的分布上均衡可比,代表性样本合计 845 人构成了本次研究的对照人群。

2. 资料收集与实验检测:研究通过一对一的上门问卷调查收集个体特征和生活方式等常见环境暴露因素的信息,其中,吸烟史定义为每天吸烟 1 支及以上且持续 1 年及以上,或者在短时间内累计量达到此水平者;饮酒史定义为每天至少一次且持续 3 个月以上,或者在短时间内累计次数达到此水平者;家族肿瘤史指一、二级亲属中有无个体患恶性肿瘤。每名对象同时采集静脉血 5 ml,采用改良盐析法抽提外周血白细胞基因组 DNA 后,  $-60^{\circ}\text{C}$  冰箱储存待检。OGG1 Ser326Cys, XRCC1 Arg194Trp, Arg280His 和 Arg399Gln, XPD Lys751Gln 和 XRCC3 Thr241Met 各单核苷酸多态性 (SNP) 的检测分型均采用 PCR-RFLP 分析技术和盲法检测。引物序列参考相关文献 [5-9] 方法,其中 XRCC1 Arg280His 纠正了文献报道中 2 个错误碱基,所有引物均由上海生工生物工程技术有限公司 (Sangon) 合成。为提高分型结果的正确性和可信性,显像模糊的样品均予以复检,而每个批次的 PCR 扩增中均平行设置阴性对照,并以 10%~20% 的总样本量进行抽样复检。最终,各基因型检测的成功率均在 95% 以上。问卷调查和血样采集都经相关伦理学委员会同意,并经每位对象签署知情同意后书后进行。

3. MDR 基本原理和分析步骤:其基本原理是将  $n$  个因子在  $n$  维空间形成的多因子组合,单纯根据疾病易感性 (高风险和低风险) 归类建模,将高维结构降到一维两水平,而后通过交叉验证和置换检验来评价模型是否有统计学意义及其预测疾病易感性的有效性。MDR 的基本分析步骤包括:①判断研究设计类型是否符合要求,应为病例对照研究或不一致同胞对设计;②判断纳入分析的指标是否符合要求,应是离散型变量,同时结局变量应为二分类变

量;③对数据存在的错分和缺失情况进行评估,若是多态基因型需进行对照组的 Hardy-Weinberg(H-W)检验;④对纳入存在于同一基因的多个多态位点进行连锁不平衡检验,以明确是否需要转换成单体型后纳入分析;⑤分析是否存在遗传异质性、位点异质性问题;⑥以记事本格式(.txt)建立符合 MDR 分析要求的数据集,若存在缺失值,可先采用 MDR-Data Tool Software 程序包对缺失值进行填补;⑦采用 MDR Software 程序包进行 MDR 分析,如果是非配对病例对照研究设计可以在程序默认的参数设置下运行;⑧对分析结果予以评估,有统计学意义的模型组合可将高、低风险型分别合并后,采用 logistic 回归对具体的风险效应予以测算,以进一步验证结论的正确性。MDR-Data Tool Software 和 MDR Software 两个程序均是基于 Java 程序编写,可在 <http://www.epistasis.org/mdr.html> 免费下载。

4. 统计学分析:整个统计分析过程结合应用了 Excel 2003、SPSS 13.0、EH Linkage Software 1.2、MDR-Data Tool Software 0.4.3 和 MDR Software 1.0.0 等软件。单因素均衡性检验中病例组和对照组个体具体年龄的差异比较采用成组 *t* 检验,而性别和等位基因等分类变量的分布差异采用 Pearson  $\chi^2$  分布检验;H-W 遗传平衡检验采用  $\chi^2$  拟合优度;连锁不平衡检验采用 EH Linkage Software 1.2 进行  $\chi^2$  似然比检验;通过运行 MDR-Data Tool Software 0.4.3 对缺失值进行填补后,采用 MDR Software 1.0.0 软件进行基因-基因、基因-环境交互作用的模型筛选分析,两程序的参数设置均参照系统默认。 $P < 0.05$  为差异或风险估计有统计学意义。

**结 果**

1. 研究对象的一般特征与生活方式因子:如表 1 所示,病例组和对照组仅平均年龄和年龄分布差异有统计学意义( $P < 0.01$ )。而个体特征与结直肠癌风险的关联分析表

明,年龄与其存在正相关关系,与低年龄组( $\geq 42$  至  $< 61$  岁)个体相比,高年龄组( $\geq 61$  岁)个体结直肠癌患病风险增高有统计学意义(校正 OR = 2.04, 95% CI: 1.49~2.80);而家族肿瘤史同样与结直肠癌的发病存在有统计学意义的正相关关系,校正 OR 值(95% CI)为 1.51(1.05~2.17)。

2. DNA 修复基因多态性分布及单位点多态性与结直肠癌发病的相关性:如表 2 所示,所有等位基因分布对照组与病例组间差异无统计学意义( $P > 0.05$ )。相应各位点基因型分布在两组间差异也无统计学意义( $P > 0.05$ )。DNA 修复基因单基因多态性与结直肠癌患病风险的关联分析表明,所有变异杂合型和纯合型都可使结直肠癌的患病风险增高(OR 值  $> 1$ ),但效应无统计学意义( $P > 0.05$ )。

3. MDR 基因-基因交互作用模型分析:经  $\chi^2$  拟

表1 病例组和对照组的基本因素分布

因素	对照		病例		P 值 <sup>a</sup>	OR 值(95% CI) <sup>c</sup>
	例数	构成 (%)	例数	构成 (%)		
合计	845	100.00	206	100.00		
年龄(岁)						
$\bar{x} \pm s$	61.84 ± 10.83		65.25 ± 9.47		0.000	
42~	441	52.19	72	34.95	0.000	1.00
61~	404	47.81	134	65.05		2.04(1.49~2.80)
性别						
女	452	53.49	101	49.03	0.250	1.00
男	393	46.51	105	50.97		1.21(0.89~1.65)
BMI <sup>b</sup>						
$< 22$	432	51.99	116	57.43	0.165	1.00
$\geq 22$	399	48.01	86	42.57		0.93(0.68~1.29)
吸烟史						
无	527	62.37	123	59.71	0.481	1.00
有	318	37.63	83	40.29		0.95(0.60~1.51)
饮酒史						
无	601	71.12	146	70.87	0.943	1.00
有	244	28.88	60	29.13		0.95(0.63~1.43)
家族肿瘤史						
无	664	78.58	152	73.79	0.139	1.00
有	181	21.42	54	26.21		1.51(1.05~2.17)

注:<sup>a</sup> Pearson  $\chi^2$  分布检验;<sup>b</sup> 病例组有 4 例缺失值,对照组 14 例缺失值;<sup>c</sup> 校正因素为年龄分布和性别构成

表2 DNA 修复基因多态等位基因在病例组和对照组的分布

基因及位点	病例组		对照组		P 值 <sup>a</sup>
	W 等位基因	M 等位基因	W 等位基因	M 等位基因	
OGG1 Ser326Cys	179(45.43)	215(54.57)	753(45.09)	917(54.91)	0.197
XRCC1 Arg194Trp	283(70.05)	121(29.95)	1102(65.75)	574(34.25)	0.570
XRCC1 Arg280His	364(88.78)	46(11.22)	1482(88.32)	196(11.68)	0.249
XRCC1 Arg399Gln	290(71.78)	114(28.22)	1237(73.81)	439(26.19)	0.010
XPD Lys751Gln	363(89.85)	41(10.15)	1541(91.84)	137(8.16)	0.232
XRCC3 Thr241Met	388(95.10)	20(4.90)	1604(95.59)	74(4.41)	0.052

注:括号外数据为例数,括号内数据为构成比(%);<sup>a</sup> 对照组 H-W 遗传平衡检验

合优度检验,在正常对照组,仅 XRCC1 Arg399Gln 基因型频率分布不符合 H-W 遗传平衡定律 ( $P = 0.010$ ),因此,后期的数据分析中将该位点剔除。而采用 EH Linkage Software 1.2 软件分析表明, XRCC1 Arg194Trp 和 Arg280His 两位点在病例组和对照组都不存在有统计学意义的连锁不平衡现象(病例组  $\chi^2 = 5.23$ , 对照组  $\chi^2 = 6.85$ ,  $df = 3$ ,  $P$  值均  $> 0.05$ )。因此,后续的 MDR 分析中,纳入了 OGG1 Ser326Cys, XRCC1 Arg194Trp 和 Arg280His, XPD Lys751Gln 和 XRCC3 Thr241Met 等 5 个多态位点,分析结果见表 3。各因子模型组合均无统计学意义 ( $P > 0.05$ ),但以仅纳入 XRCC1 Arg194Trp 的单因子模型最佳(平均检验准确度 = 0.533,交叉验证一致性 = 10/10)。

4. MDR 基因-环境交互作用模型分析:研究进一步纳入了个体特征和生活方式等环境因子,进行基因-环境交互作用的 MDR 分析。分析发现,最佳模型是包含了年龄分布、饮酒史、XRCC1 Arg194Trp 和 OGG1 Ser326Cys 的 4 个因子模型(平均检验准确度 = 0.616,交叉验证一致性 = 10/10,  $P = 0.011$ )。研究对象经以上 4 个因子又生组合后,在风险度上区分为“高风险”和“低风险”两个组合。进一步将高、低风险组合分别合并,采用 logistic 分析表明,与筛选出的低风险组合相比,高风险组合可以使结直肠癌的患病风险增高,并有统计学意义 ( $OR = 2.72, 95\% CI: 1.66 \sim 4.47$ )。

### 讨 论

人群散发性结直肠癌属于典型的多因子疾病,环境因素虽可解释 70% 结直肠癌的发生,但机体存在着遗传易感倾向。而这种个体倾向性的差异主要表现为基因多态性,其中 SNP 是最重要的物质基础<sup>[10,11]</sup>。因此,对于多因子疾病病因学研究,基因-基因、基因-环境交互作用的识别和定量分析尤为重要。

MDR 分析通过交叉验证和置换检验集属性选择、属性构建和分类于一体,是一种综合的、效能较高的数据分析策略,特别适合于离散型遗传和环境因素复杂高阶交互作用的识别和分析,是对 logistic 回归分析的一种有效补充。作为一种非参数检验方法,它的应用不受遗传模式(显性或隐性)和交互作用模型(线性或非线性,加法或乘法)的限制<sup>[12,13]</sup>。然而,MDR 作为一种新生的数据挖掘技术,也存在不足之处:一方面,普通的 MDR 分析仅适用于结局变量为二分类的(有病或无病,死亡或存活)病例对照或不一致同胞对研究;另一方面,遗传异质性、位点异质性及拟表型等会大大降低 MDR 的检验效能。另外,尽管理论推算和实例应用都证明与传统方法相比,对于复杂疾病的高阶交互作用分析和识别 MDR 分析有着更为优异的把握度,但是对于 MDR 样本量和把握度(效能)的估算至今仍未有明确的估算公式。Ritchie 等<sup>[14]</sup>仅通过数据模拟研究证明当病例和对照样本量均  $< 50$  个时,检验效能会明显下降。有关 MDR 分析的小样本研究应用成功案例,有 Williams 等<sup>[15]</sup>以加纳黑人为研究人群,纳入病例 126 例、对照 51 人,开展原发性高血压多基因(ACE 等 8 基因 13 个位点)交互作用的筛选研究;另外,Soares 等<sup>[16]</sup>也在葡萄牙人群中纳入病例 92 例、对照 85 人开展了常染色体显性遗传病家族性淀粉样多发性神经病多基因(APCS 等 6 基因 10 个位点)影响与发病年龄的相关性研究。目前,MDR 分析在心血管疾病、高血压、糖尿病、恶性肿瘤等多基因疾病交互作用分析中都得到了一定的应用<sup>[15,17-19]</sup>。其中,Coffey 等<sup>[20]</sup>于 2004 年将其应用于心肌梗死基因-基因交互作用分析时,采用 MDR 分析将基因型组合区分为高、低风险后,以“低风险型”组合为参照,经 logistic 回归拟合,分析发现“高风险型”罹患心肌梗死的危险度增高有统计学意义 ( $OR = 1.44, 95\% CI: 1.06 \sim 1.95$ )。而在我国尚缺乏类似的应用实例。

表3 MDR 分析基因-基因、基因-环境交互作用模型因子数

因子数	基于 DNA 多态基因				基于 DNA 多态基因和环境因子			
	因子组合	交叉验证一致性	平均检验准确度	P 值	因子组合	交叉验证一致性	平均检验准确度	P 值
1	XRCC1 Arg194Trp	10/10	0.533	0.184	年龄分布	10/10	0.584	0.184
2	XRCC1 Arg194Trp/OGG1 Ser326Cys	10/10	0.494	0.816	年龄分布/家族肿瘤史	10/10	0.574	0.184
3	XRCC1 Arg194Trp/XRCC1 Arg280His/XPD Lys751Gln	6.4/10	0.445	0.956	年龄分布/性别分布/XRCC1 Arg194Trp	8.9/10	0.568	0.999
4	XRCC1 Arg194Trp/XRCC1 Arg280His/OGG1 Ser326Cys/XPD Lys751Gln	9.9/10	0.470	0.691	年龄分布/饮酒史/XRCC1 Arg194Trp/OGG1 Ser326Cys	10/10	0.616	0.011

本研究属非配对病例对照研究,纳入的分析因子均为离散型变量,而结局变量为结直肠癌发病或未发病,缺失值 $<5\%$ 且采取包括复检等多项措施尽量避免错分,初步判定符合 MDR 的应用条件。分析过程中,首先采用  $\chi^2$  拟合优度检验对正常对照组 6 个 SNP 多态位点分布是否符合 H-W 遗传平衡定律进行检验,发现仅 XRCC1 Arg399Gln 基因频率分布不符合 H-W 遗传平衡,分析原因可能是由随机抽样误差引起的偏性,予以剔除;尔后采用 EH Linkage Software 1.2 对 XRCC1 Arg194Trp 和 Arg280His 同一基因两位点在病例组和对照组是否存在连锁不平衡现象进行检验分析,结果表明两位点不存在连锁不平衡现象,因此进行交互作用分析时将其作为两个独立的因子来考虑;最后,对 OGG1 Ser326Cys, XRCC1 Arg194Trp 和 Arg280His, XPD Lys751Gln 和 XRCC3 Thr241Met 5 个 SNP 的基因-基因交互作用 MDR 筛选分析无统计学意义的发现。对基因-环境交互作用的分析,研究纳入了年龄、性别、BMI 分布、吸烟史、饮酒史和家族肿瘤史等 6 个环境因子以及上述的 5 个 SNP,MDR 分析表明,最佳模型是包含年龄分布、饮酒史、XRCC1 Arg194Trp 和 OGG1 Ser326Cys 的 4 个因子模型,且经 logistic 拟合分析,与筛选出的低风险组合相比,高风险组合可以使结直肠癌的患病风险增高并有统计学意义 ( $OR = 2.72, 95\% CI: 1.66 \sim 4.47$ )。

本研究结果表明,DNA 修复基因多态性对结直肠癌发病的影响符合多基因疾病的低外显性特征,单个位点的多态性与结直肠癌的发病不存在相关性,且与环境因子可能存在着复杂的联合作用。在后续的研究中,应纳入更多的有功能意义的多态位点,在开展基于整个 DNA 修复通路 MDR 分析的同时,根据不同的具体修复通路,开展更为深入细致地分析。另外,研究纳入的病例为当地现患存活病例,可能存在一定的存活偏倚即选择偏倚从而影响结果真实性,研究结论有待后续的研究予以进一步证实。

#### 参 考 文 献

- [1] Yang L, Parkin DM, Ferlay J, et al. Estimates of cancer incidence in China for 2000 and projections for 2005. *Cancer Epidemiol Biomarkers Prev*, 2005, 14: 243-250.
- [2] 吴德生. 环境与健康//陈学敏. 环境卫生学. 1 版. 北京: 人民卫生出版社, 2004: 11-45.
- [3] 朱守民, 夏昭林. DNA 损伤修复基因与遗传易感性. *环境与职业医学*, 2003, 20(1): 50-52.

- [4] Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*, 2001, 69: 138-147.
- [5] Abdel-Rahman SZ, El-Zein RA. The 399Gln polymorphism in the DNA repair gene XRCC1 modulates the genotoxic response induced in human lymphocytes by the tobacco-specific nitrosamine NNK. *Cancer Lett*, 2000, 159: 63-71.
- [6] Butkiewicz D, Rusin M, Enewold L, et al. Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis*, 2001, 22: 593-597.
- [7] David-Beabes GL, Lunn RM, London SJ. No association between the XPD (Lys751Gln) polymorphism or the XRCC3 (Thr241Met) polymorphism and lung cancer risk. *Cancer Epidemiol Biomarkers Prev*, 2001, 10: 911-912.
- [8] Le Marchand L, Donlon T, Lum-Jones A, et al. Association of the hOGG1 Ser326Cys polymorphism with lung cancer risk. *Cancer Epidemiol Biomarkers Prev*, 2002, 11: 409-412.
- [9] Sturgis EM, Zheng R, Li L, et al. XPD/ERCC2 polymorphisms and risk of head and neck cancer: a case-control analysis. *Carcinogenesis*, 2000, 21: 2219-2223.
- [10] Baglioni S, Genuardi M. Simple and complex genetics of colorectal cancer susceptibility. *Am J Med Genet C Semin Med Genet*, 2004, 129: 35-43.
- [11] Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 2001, 409: 928-933.
- [12] Moore JH, Gilbert JC, Tsai CT, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol*, 2006, 241: 252-261.
- [13] 唐迅, 李娜, 胡永华. 应用多因子降维法分析基因-基因交互作用. *中华流行病学杂志*, 2006, 27: 437-441.
- [14] Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol*, 2003, 24: 150-157.
- [15] Williams SM, Ritchie MD, Phillips JA 3rd, et al. Multilocus analysis of hypertension: a hierarchical approach. *Hum Hered*, 2004, 57: 28-38.
- [16] Soares ML, Coelho T, Sousa A, et al. Susceptibility and modifier genes in Portuguese transthyretin V30M amyloid polyneuropathy: complexity in a single-gene disease. *Hum Mol Genet*, 2005, 14: 543-553.
- [17] Agirbasli D, Agirbasli M, Williams SM, et al. Interaction among 5, 10 methylenetetrahydrofolate reductase, plasminogen activator inhibitor and endothelial nitric oxide synthase gene polymorphisms predicts the severity of coronary artery disease in Turkish patients. *Coron Artery Dis*, 2006, 17: 413-417.
- [18] Hsieh CH, Liang KH, Hung YJ, et al. Analysis of epistasis for diabetic nephropathy among type 2 diabetic patients. *Hum Mol Genet*, 2006, 15: 2701-2708.
- [19] Oestergaard MZ, Tyrer J, Cebrian A, et al. Interactions between genes involved in the antioxidant defence system and breast cancer risk. *Br J Cancer*, 2006, 95: 525-531.
- [20] Coffey CS, Hebert PR, Ritchie MD, et al. An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC Bioinformatics*, 2004, 5: 49.

(收稿日期: 2007-12-13)

(本文编辑: 张林东)