

多因子降维法分析基因-基因交互作用的应用进展

唐迅 李娜 陈大方 胡永华

【关键词】 多因子降维法; 病例对照研究; 基因-基因交互作用

Recent advances in applications of multifactor dimensionality reduction for detecting gene-gene interactions TANG Xun, LI Na, CHEN Da-fang, HU Yong-hua. Key Laboratory of Epidemiology, Ministry of Education, Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing 100083, China
Corresponding author: HU Yong-hua, Email: yhhu@bjmu.edu.cn

【Key words】 Multifactor dimensionality reduction; Case-control study; Gene-gene interactions

多因子降维法 (multifactor dimensionality reduction, MDR) 是近年来发展的一种分析交互作用的新方法^[1], “因子”即交互作用研究中的变量 (如基因型或环境因素), “维”是指研究的多因子组合中的因子数 (如基因型数目); 该方法以疾病易感性分类 (高危、低危) 的方式建模, 将研究中的多个因子看作一个多因子组合 (基因型组合), 这样就把高维的结构降低到单维两水平 (即高危或低危), 即为“降维”, 最后, 再通过交叉验证和置换检验来评估该单维的多因子组合识别和预测疾病的能力。作为一种非参数的分析方法, 它无需指定遗传模式 (显性或隐性遗传) 和交互作用模型 (线性或非线性模型, 加法或乘法模型), 因而既克服了传统的 logistic 回归分析的线性模型对小样本参数估计所导致的 I 类和 II 类错误增大的问题, 又解决了研究多个因子之间交互作用时高

维空间中数据分布稀疏的“维度困扰” (curse of dimensionality) 的问题^[2]; 并且, 相对于其他的数据挖掘方法而言, 其使用条件宽松、过程简单易懂、结果解释合理, 并可通过免费软件实现, 具体过程及方法参见参考文献 [3]。自从 2001 年 Ritchie 等^[1]首次提出该方法以来, 已经得到了广泛应用和关注, 现将有关文献进展综述如下, 本文将结合实际应用的研究实例, 重点强调该方法的适用条件和使用策略, 并讨论其发展前景。

1. MDR 方法应用广泛: 2001 年 Ritchie 等^[1]在研究散发性乳腺癌时首次提出 MDR 的方法, 但可能是由于当时缺少相应的分析软件, 此类研究的应用实例直至近 2 年才逐渐增多, 所研究的人群及疾病也更加广泛, 基本涵盖了高血压、糖尿病、心血管疾病和恶性肿瘤等常见的复杂疾病, 主要的研究如表 1 所示。通常在遗传流行病学研究中, 样本量不足是一个普遍的问题, MDR 方法可以在相对较小的样本量条件下进行复杂疾病的高阶交互作用分析, 正如表 1 中成功应用该方法的文献所列出的, 总样本量范围介于 177~2163 之间。

起初, 研究者只是对单个位点研究未发现阳性关联的数据, 尝试采用 MDR 方法分析交互作用, 例如, 2004 年 Cho 等^[4]采用 MDR 方法研究 2 型糖尿病时发现, UCP2 和 PPAR γ 基因多态性之间存在两位点的基因-基因交互作用 (OR = 0.51, 95% CI: 0.34~0.77, P = 0.0016)。其后, Coffey 等^[5]对心肌梗死发病中的 ACE、PAI-1 和 t-PA 基因

表 1 应用 MDR 的研究实例

文献	疾病	研究人群	病例/对照人数	基因/位点个数	MDR 最佳模型	预测误差 (%)	交叉验证一致性
Ritchie 等 ^[1]	散发性乳腺癌	美国白人妇女	200/200	5/10	3 基因 4 位点交互模型 (P < 0.001)	46.73	9.8/10
Cho 等 ^[4]	2 型糖尿病	韩国人	504/133	15/23	2 基因 2 位点交互模型 (P = 0.0016)	20.43	8.6/10
Williams 等 ^[6]	原发性高血压	加纳黑人	126/51	8/13	2 基因 2 位点交互模型 (P = 0.013)	29.50	9/10
Tsai 等 ^[7]	房颤	中国台湾人	250/250	3/8	2 基因 3 位点交互模型 (P < 0.001)	37.26	10/10
Coffey 等 ^[5]	心肌梗死	美国医生	343/343	3/3	2 基因 2 位点交互模型 (P = 0.09) ^a	46.00	10/10
Bastone 等 ^[8]	动脉粥样硬化	美国白人	300/300	10/11	3 基因 3 位点交互模型 (P = 0.106) ^a	40.00	7.5/10
Soares 等 ^[9]	家族性淀粉样多神经病	葡萄牙人	92/85	6/10	2 基因 3 位点交互模型 (P = 0.006)	21.70	10/10
Qin 等 ^[10]	精神分裂症	中国汉族人	253/140	2/16	2 基因 3 位点交互模型 (P < 0.001)	30.75	10/10
Xu 等 ^[11]	前列腺癌	瑞典白人	1383/780	20/57	4 基因 4 位点交互模型 (P = 0.019)	43.28	5.6/10
Chan 等 ^[12]	哮喘	中国香港儿童	240/140	8/12	2 基因 2 位点交互模型 (P = 0.014)	33.70	10/10
Mannila 等 ^[13]	心肌梗死	瑞典白人	387/387	5/5	2 基因 2 位点交互模型 (P = 0.03)	44.30	9.9/10
Brassat 等 ^[14]	多发性硬化症	美国黑人	442/293	34/49	3 基因 3 位点交互模型 (P < 0.01)	23.90	5/10
Shifman 等 ^[15]	精神分裂症	北欧犹太人	507/450	2/6	2 基因 4 位点交互模型 (P < 0.0001)	40.00	10/10
Hsieh 等 ^[16]	糖尿病肾病	中国台湾人	144/120	28/231	2 基因 2 位点交互模型 (P = 0.0032)	37.30	11/12

注: ^aP > 0.05

基金项目: 国家自然科学基金资助项目 (30671807); 高等学校博士学科点专项科研基金资助项目 (200600061111)
作者单位: 100083 北京大学公共卫生学院流行病学与卫生统计学系 流行病学教育部重点实验室
通讯作者: 胡永华, Email: yhhu@bjmu.edu.cn

进行研究,首先采用条件 logistic 回归分析,发现 ACE 和 PAI-1 基因多态性之间的交互作用有统计学意义 ($P=0.02$);后又采用 MDR 分析,虽然发现相同的结果,但模型的显著性水平没有达到 0.05 ($P=0.09$),并且,其预测误差的估计值也接近 50%,趋于没有阳性关联的无效假设。这一结果强调了在进行交互作用分析时模型验证过程的重要意义,这也可以从某种程度上部分解释此类单基因关联研究的阳性结果为何通常难以重复的原因。

除了结合传统的 logistic 回归分析,后续的研究还综合采用了其他的一些新方法。Bastone 等^[8]通过对冠状动脉粥样硬化的遗传风险研究(SIRCA)的数据,分别进行 MDR 以及模式形成和递归划分法(patterning and recursive partitioning, PRP)分析,发现其结果高度一致,同时找出了三个位点交互的模型,预测误差分别为 40.0% 和 40.17%,虽然 $P>0.05$,模型的检验不具有统计学意义,但最终 Bastone 等仍然通过理论推导发现,MDR 实际上就是递归划分法的一个特例。

近年来,越来越多的大样本研究开始应用 MDR 方法,并且随着 DNA 微阵列等基因芯片技术广泛应用,研究中的交互作用位点越来越多^[11-16]。在这种情况下,优化 MDR 分析的策略显得尤为重要。Williams 等^[6]研究原发性高血压时,首先将 MDR 用于全部 13 个多态性位点的分析,然后根据 4 条不同的生理通路对不同通路的多态性位点进行单独的 MDR 分析。这种分析策略首先强调通路的关系,然后才是通路的集合。该研究表明,根据基因功能或生物学通路选择基因,然后进行 MDR 分析,可能更有效,同时这也有利于进行生物学交互作用的解释。后续的一些研究也采用了这种基于通路选择基因位点的研究策略,例如, Xu 等^[11]研究了炎症通路上的基因位点之间的交互作用与瑞典白人前列腺癌发病的关系。

MDR 方法并非只限于多基因病的研究,2005 年 Soares 等^[9]通过对葡萄牙人群中家族性淀粉样多神经病这一常染色体显性疾病的研究,提供了很好的实例。事实上,由于多个基因位点和环境因素都可能对病程有影响,而使研究更为复杂。Soares 等首先根据发病年龄是否超过 40 岁,将病例分为普通型和早发型,分层进行研究,通过 MDR 分析发现,普通型与 APCS 基因有关,而晚发型与 RBP4v1、RBP4v2 和 APCSv1 基因有关,提示 RBP4 基因很可能对该病的发病年龄有影响。这种分层研究的策略对避免遗传异质性的影响,将有重要的价值。

2. MDR 方法适用性与应用策略:目前,虽然 MDR 方法得到了广泛应用,但也需要注意其适用条件,主要包括样本量及效能、数据结构(如病例和对照数是否平衡、同胞对等家系资料的结构)、遗传异质性等方面,这些问题必须在进行 MDR 分析前考虑清楚。

(1)样本量及效能:目前进行 MDR 分析所需的样本量大小仍不明确,但 Ritchie 等^[17]通过模拟数据研究显示,病例、对照数各为 200 例时,检测两位点交互作用,MDR 的效能达

到 80% 以上。即使在 5% 的基因型错分和 5% 缺失数据的情况下,对于大多数模型,该结果仍然真实。当然进行高阶交互作用需要更多的样本量,目前尚无样本量的理论计算公式,但模拟研究证明 <50 个病例和 50 个对照的数据集,其检测效能将会明显降低^[18]。

(2)数据结构:MDR 可用于病例对照或不一致同胞对研究设计,所检测的因子应该是分类资料的独立变量,如单核苷酸多态性(SNP)或其他序列变异(插入、缺失等);环境危险因素也能作为分类变量,进行基因-基因或基因-环境交互作用的研究。结局变量作为因变量必须是二分变量,如病例/对照,有效/无效等。对于病例、对照数相等的平衡数据,MDR 方法与对于每个个体的基因型数据转化为单一的多位点分布的贝叶斯分类(naive Bayes classifier)一致^[19]。MDR 也可以用于病例、对照数不相等的非平衡数据研究。此时,可考虑扩大抽样或减少抽样。扩大抽样是指对样本数较低的个体组进行数据集内的随机重复抽样,直至病例和对照组样本量相等为止;而减少抽样则是从样本数较多的个体组的数据集中随机删除个体,直至平衡为止。但采用这两种方法进行调整也存在着一些问题:扩大抽样可能会引起虚假关联,这是由于一些特殊样本被重复抽样,另外,还可能造成研究效能虚高的假象;而减少抽样所采用数据去除的原理,同样也会由于某些特殊样本被去除而引起虚假关联,同时,由于样本量变小还可能導致效能降低。因此,这两者孰优孰劣,目前尚无定论,对于此两种方法影响的模拟数据的研究仍在进行中^[18]。另一种更保守的分析非平衡数据的方法就是调整 MDR 的阈值,阈值是决定分配给特定多位点基因型组合的患病风险状态(高危、低危)的比值,通常为 1,但如果 50 个病例与 200 个对照,阈值就为 0.25,而不是 1。当阈值调整后,计算分类误差时也需要进行调整以适应非平衡数据,例如,采用平衡的准确度统计量(即: $1 - \text{平衡的分类误差} = [1 - (\text{灵敏度} + \text{特异度})/2]$)。解决非平衡数据的问题,也有其他一些方法,如建立新的适合度函数等,这些方法还有待进一步的研究。

(3)遗传异质性:Ritchie 等^[17]通过模拟数据的研究表明,遗传异质性或位点异质性以及拟表型会大大降低 MDR 的效能。遗传异质性之所以会导致效能较低,是因为不同的因子组合导致疾病,将会降低预测的准确度和模型交叉验证的一致性。拟表型是因为环境因素变化的影响所产生的表型变化,并不会遗传到下一代,所以基因本身的结构不变,只是表型改变。其降低效能可能是由于某些随机事件导致半数的病例受影响。因此,同时对环境因素进行 MDR 分析,或者先以环境因素分层,或许能够提高其效能。事实上对于所有统计学方法,异质性都是一个困扰的问题。另外,可以考虑通过采用诸如聚类分析、基于递归分类树的方法(tree-based recursive partitioning methods)^[20]、有序子集分析(ordered subset analysis)等手段^[21],构造一些相对同质的亚组,对于不同亚组再进行 MDR 分析,或将亚组状态作为协变

量进行分析,以便较好地解决这一问题。

(4)模型验证和过度拟合:MDR 算法的核心就是属性选择、属性构造以及分类,并通过交叉验证和置换检验进行模型选择和评估。正是由于组合运用了交叉验证和置换检验的方法,从而降低了多重检验所导致的 I 类错误出现的机会。需要注意的是,从数学角度来讲,计算预测误差和分类错误是一样的,但在数据集中,用来作为计算统计量的数据部分是不同的。分类误差是用训练样本进行计算的,而预测误差则是用检验样本进行计算的。两者都是衡量 MDR 模型错分临床结局的个体数。但在模型选择中,一般采用预测误差而非非分类误差作为决定性指标,这是因为采用分类误差时会出现过度拟合的情况。分类模型过度拟合是此类高维数据分析时的一个严重的问题,特别是当样本量较小时,随着评估位点数目的增加,分类误差将始终降低。因此,虽然高阶模型过度拟合数据,但其预测能力变差。当交叉验证一致性最大的模型与预测误差最小的模型不一致时,应采用统计学简约性(statistical parsimony)的原则,选择包含因子数目较少的模型作为最佳模型^[16]。

(5)交叉验证和穷举算法:交叉验证过程正是避免过度拟合的较好办法。最早在进行 MDR 分析时,数据集必须正好被交叉验证等份数目(通常为 10)所整除,这意味着不能整除时有少量个体数据将被弃除,而现在进行 MDR 分析时已无此限制,数据集将尽可能被平均划分而无数据丢失。但是,通常 MDR 进行 10 重交叉验证,作为一种穷举算法,进行大量数据分析时对计算机的要求相当高,因此 MDR 不适用于识别 5 个位点以上的交互作用模型。最近的研究表明,把交叉验证的次数从 10 重减为 5 重,将会极大地降低计算时间,但并不会导致效能的降低^[22]。这样不仅降低了分析原始数据的时间,还减少了置换检验的运算时间,置换检验也是极其繁琐的计算过程,因此整个过程的运算时间将大大降低,这意味着可以分析更大样本的数据集。但需要注意的是交叉验证次数减为 3 重,甚至取消时将会导致效能降低。

3. MDR 方法学的扩展:虽然目前大多数研究都是考虑两位点 SNP 基因型,但 MDR 的应用并不限于此,环境危险因素、其他多位点基因型,如微卫星标记或单体型均可应用,因而也可用于基因-环境交互作用的研究,以及遗传药理学研究^[23],例如,用于识别药物代谢状态的乙酰化水平^[24],不同的设计所关注的交互作用的研究变量不同。

MDR 方法本身也是在不断发展的,如扩展的 MDR (EMDR)^[25]、MDR-PDT 等方法的提出^[26],将其应用范围扩展到了家系资料。Ma 等^[25]首次将 MDR 方法扩展到自闭症的家系研究,在 470 个美国白人家系中发现 GABRA4 和 GABRB1 基因存在两位点的交互作用,同时采用的多位点基因型家系不平衡检验(genotype-pedigree disequilibrium test, geno-PDT)和存在连锁情况下的关联检验(association in the presence of linkage test, APL)也证实了该结果。值得一提的是,该研究扩展了一些新的特性到 MDR 方法中,包括:①允

许个体资料有部分基因型数据的缺失;②衡量预测误差时加入了 χ^2 统计量;③加入了非固定的置换检验以得到模型的统计学意义,对于多位点模型,该检验将考虑所有多个位点的全部组合。并且,该研究表明,这种 EMDR 比经典的 MDR 方法的假阳性和假阴性率更低。在家系研究中,真实的单体型可知,这为进行单体型研究的应用提供了条件。但总体而言,目前此类家系研究甚少,对于单体型资料进行 MDR 分析也有待进一步研究。

此外,Motsinger 和 Ritchie^[16]正在试图扩展 MDR,以分析超过两分类的多结局变量,如“未患病”、“轻度患病”和“重度患病”,这对于分析高血压和糖尿病等疾病而言更有意义。由于 MDR 采用的是列联表的方法,这种方法用于等级资料的分析并非难事。最近,Bush 等^[27]又将其机器学习的算法进行了改进,通过递归框并法(recursive binning method)提出了可用于更大规模的交互作用分析的并联 MDR 方法(Parallel MDR),据称可以研究任何离散变量,包括等位基因、基因型、单体型等,并且研究样本及变量数目的上限也将不受限制。

对于 MDR 分析所发现的交互作用结果进行解释,仍然是目前面临的巨大挑战。针对此问题,Moore 等^[28]结合图形归纳算法(constructive induction algorithms)提出了一种更为灵活的 MDR 分析框架:首先,采用熵统计量选择感兴趣的多态性位点的组合,一旦所感兴趣的 SNP 被选择,可对其采用 MDR 构造新的属性,新的属性或属性的集合构造完毕后,可以用任何类型的机器学习算法进行评估。最后,通过熵交互曲线和交互聚类图的方法作为直观的工具进行交互作用模型的解释。

综上所述,应用 MDR 进行交互作用研究的过程,一般包括但不限于以下步骤:①根据研究设计判断是否具备应用 MDR 的条件,如病例对照研究或不一致同胞对设计可应用 MDR,而病例父母研究的家系资料则应考虑 MDR-PDT;②收集样本和资料,需满足基本的样本量条件,并参考生物学机制的致病通路及国际人类基因组单体型图计划(HapMap)的数据^[29],选择需要研究的基因型位点;③完成基因分型,评估缺失值,并检验 Hardy-Weinberg 平衡;④通过连锁不平衡检验,确定各位点之间的相关性,较高的相关性将导致多重共线性,因此应该将存在较高连锁不平衡的 SNP 变成单体型进行分析;⑤检验数据的人群分层(如种族、年龄、性别等的差异),并考虑遗传异质性的问题,可先采用聚类分层分析;⑥按照 MDR 分析的数据格式构造数据集,运用软件进行 MDR 分析;⑦通过置换检验对模型进行评价;⑧采用其他分析方法,如 logistic 回归分析、数据集关联法等验证所得结论;⑨采用交互聚类图等方法作为直观的工具,尝试解释可能的生物学机理。研究基因型与表型的关系,最终还是为了提高预防、诊断和治疗疾病的能力。虽然 MDR 可以为研究数个 SNP 的交互作用提供工具,但对于数以百万计的 SNP 而言,这一方法所提供的解释还甚少。组合应用多种方法,包括

MDR、数据集关联法以及随机森林法,对于发现复杂疾病发病中的交互作用可能更为有效。需要注意的是,没有任何一种方法是“放之四海而皆准”的万能工具,而综合应用各种方法的组合策略更值得推荐。

参 考 文 献

- [1] 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(1):138-147.
- [2] Moore JH. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn*, 2004, 4(6):795-803.
- [3] 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(5):437-441.
- [4] Cho YM, Ritchie MD, Moore JH, et al. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia*, 2004, 47(3):549-554.
- [5] 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(1):49.
- [6] Williams SM, Ritchie MD, Phillips JA 3rd, et al. Multilocus analysis of hypertension: a hierarchical approach. *Hum Hered*, 2004, 57(1):28-38.
- [7] Tsai CT, Lai LP, Lin JL, et al. Renin-angiotensin system gene polymorphisms and atrial fibrillation. *Circulation*, 2004, 109(13):1640-1646.
- [8] Bastone L, Reilly M, Rader DJ, et al. MDR and PRP: a comparison of methods for high-order genotype-phenotype associations. *Hum Hered*, 2004, 58(2):82-92.
- [9] 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(4):543-553.
- [10] Qin S, Zhao X, Pan Y, et al. An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. *Eur J Hum Genet*, 2005, 13(7):807-814.
- [11] Xu J, Lowey J, Wiklund F, et al. The interaction of four genes in the inflammation pathway significantly predicts prostate cancer risk. *Cancer Epidemiol Biomarkers Prev*, 2005, 14(11 Pt 1):2563-2568.
- [12] Chan IH, Leung TF, Tang NL, et al. Gene-gene interactions for asthma and plasma total IgE concentration in Chinese children. *J Allergy Clin Immunol*, 2006, 117(1):127-133.
- [13] Mannila MN, Eriksson P, Ericsson CG, et al. Epistatic and pleiotropic effects of polymorphisms in the fibrinogen and coagulation factor XIII genes on plasma fibrinogen concentration, fibrin gel structure and risk of myocardial infarction. *Thromb Haemost*, 2006, 95(3):420-427.
- [14] Brassat D, Motsinger AA, Caillier SJ, et al. Multifactor dimensionality reduction reveals gene-gene interactions associated with multiple sclerosis susceptibility in African Americans. *Genes Immun*, 2006, 7(4):310-315.
- [15] Shifman S, Levit A, Chen ML, et al. A complete genetic association scan of the 22q11 deletion region and functional evidence reveal an association between DGCR2 and schizophrenia. *Hum Genet*, 2006, 120(2):160-170.
- [16] Hsieh CH, Liang KH, Hung YJ, et al. Analysis of epistasis for diabetic nephropathy among type 2 diabetic patients. *Hum Mol Genet*, 2006, 15(18):2701-2708.
- [17] 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(2):150-157.
- [18] Motsinger AA, Ritchie MD. Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics*, 2006, 2(5):318-328.
- [19] Hahn LW, Moore JH. Ideal discrimination of discrete clinical endpoints using multilocus genotypes. *In Silico Biol*, 2004, 4(2):183-194.
- [20] Shannon WD, Province MA, Rao DC. Tree-based recursive partitioning methods for subdividing sibpairs into relatively more homogeneous subgroups. *Genet Epidemiol*, 2001, 20(3):293-306.
- [21] Hauser ER, Watanabe RM, Duren WL, et al. Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol*, 2004, 27(1):53-63.
- [22] Motsinger AA, Ritchie MD. The effect of reduction in cross-validation intervals on the performance of multifactor dimensionality reduction. *Genet Epidemiol*, 2006, 30(6):546-555.
- [23] Ritchie MD, Motsinger AA. Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies. *Pharmacogenomics*, 2005, 6(8):823-834.
- [24] Sabbagh A, Darlu P. SNP selection at the NAT2 locus for an accurate prediction of the acetylation phenotype. *Genet Med*, 2006, 8(2):76-85.
- [25] Ma DQ, Whitehead PL, Menold MM, et al. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. *Am J Hum Genet*, 2005, 77(3):377-388.
- [26] Martin ER, Ritchie MD, Hahn L, et al. A novel method to identify gene-gene effects in nuclear families; the MDR-PDT. *Genet Epidemiol*, 2006, 30(2):111-123.
- [27] Bush WS, Dudek SM, Ritchie MD. Parallel multifactor dimensionality reduction: a tool for the large-scale analysis of gene-gene interactions. *Bioinformatics*, 2006, 22(17):2173-2174.
- [28] 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(2):252-261.
- [29] The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437(7063):1299-1320.

(收稿日期:2006-12-21)

(本文编辑:尹廉)