

基因多态性与疾病关联研究中 Hardy-Weinberg 平衡的影响因素分析

陈阿群 臧静 冯云霞 王凯娟

【摘要】 目的 根据炎性基因多态性与胃癌易感性家系比较研究的数据,分析基因多态性与疾病关联研究中 Hardy-Weinberg(H-W)平衡的影响因素,揭示群体中等位基因频率在逐代传递中平衡偏离的原因。方法 对胃癌风险因素的各种基因型的状态进行 H-W 平衡检验、连锁不平衡分析、Cochran-Armitage 趋势检验和分析。结果 (1)单位点 IL-1B-31、IL-1B-511、IL-1RN 和 TNF-A-308 的病例组和对照组均显著背离 H-W 平衡($P < 0.01$),MIF-173 趋于平衡。(2)双位点统计分析,各位点组合的 H-W χ^2 检验均偏离表型组合概率的期望值($P < 0.01$)。(3)Cochran-Armitage 趋势检验不同组 IL-1B-511 和 IL-1RN 位点分布的差异有统计学意义($P < 0.05$),表明群体中可能存在群体分层。结论 (1)受连锁不平衡相结合的频率依赖性选择、突变及其共同作用的影响,IL-1B-31、IL-1B-511、IL-1RN 和 TNF-A-308 位点的基因型频率在群体中分布偏离 H-W 平衡。(2)双位点遗传平衡检验模式更能反映表型组合适合度的差异。(3)群体分层是 H-W 平衡偏离的另一影响因素。

【关键词】 基因多态性; Hardy-Weinberg 平衡; 影响因素

Influential factors of Hardy-Weinberg equilibrium on the study of association between gene polymorphism and disease CHEN A-qun, ZANG Jing, FENG Yun-xia, WANG Kai-juan. Department of Epidemiology, College of Public Health, Proteomics Research Center, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: WANG Kai-juan, Email: kjwang@163.com

【Abstract】 **Objective** According to the data collected from gastric cancer families comparative study, the influence factors of Hardy-Weinberg (H-W) equilibrium in the association studies of gene polymorphism and disease were analyzed to reveal the reasons that affecting the equilibrium deviation in the group. **Methods** Varieties of risk genotype for gastric cancer were analyzed and detected with H-W equilibrium, genetic linkage disequilibrium analysis and Cochran-Armitage trend test. **Results** (1) Significant deviations from H-W equilibrium were observed in IL-1B-31, IL-1B-511, IL-1RN and TNF-A-308 of the cases and controls ($P < 0.01$). MIF-173 tended to be equilibrium in the population. (2) Deviations from expectations of phenotypes combination probability were observed in two-site H-W χ^2 tests ($P < 0.01$). (3) The Cochran-Armitage trend test showed that distribution of IL-1B-511 and IL-1RN were significantly different ($P < 0.05$), suggesting that population stratification might have existed in the group. **Conclusions** (1) Affected by frequency-dependent selection, under the combination of linkage disequilibrium, mutations and interaction, genotype frequency of IL-1B-31, IL-1B-511, IL-1RN and TNF-A-308 showed deviation from H-W equilibrium in population. (2) Two-site genetic equilibrium test model seemed better to reflect the differences of phenotypic combination fitness. (3) Population stratification was another factor to express the deviation from H-W equilibrium.

【Key words】 Gene polymorphism; Hardy-Weinberg equilibrium; Influence factor

在病例对照研究中, Hardy-Weinberg (H-W) 平

衡检验被认为是数据研究的基础。但 H-W 定律只是群体遗传学规律中的一条特殊规律,并不适应于许多类型自然群体的遗传平衡^[1]。本研究从基因多态性与疾病关联研究中,采用 H-W 平衡检验、双位点分析和 Cochran-Armitage 趋势检验等方法,分析 H-W 平衡的影响因素,揭示群体中等位基因频率在逐代传递中平衡偏离的原因。

DOI: 10.3760/cma.j.issn.0254-6450.2009.11.027

基金项目:国家自然科学基金(30972547);河南省高等学校创新人才培养工程(2005)

作者单位:450001 郑州大学公共卫生学院流行病学教研室、郑州大学蛋白质组学应用研究中心

通信作者:王凯娟, Email: kjwang@163.com

对象与方法

1. 研究对象:2003 年 8 月对河南省新乡和新乡县约 4 万人的消化道肿瘤调查确定的随访队中的 4 个乡镇常住居民队列,采用整群抽样方法进行入户面访,调查胃癌病例及其家族一级亲属及二级亲属胃癌患病情况,病例包括调查前 5 年内已患胃癌的现患病例及调查期间新发现的病例,病例均按县级以上医院病理诊断确诊;对照家系选自和病例同一村民组、年龄±5 岁、同性别、无上消化道相关疾病,与病例家系无亲缘关系,1:1 比例选取,同样调查其家庭中一级亲属及二级亲属资料。调查获得胃癌病例家系一级亲属与二级亲属成员 296 人,对照家系一级与二级亲属成员 319 人。提取研究对象外周静脉血基因组 DNA,用 PCR-RFLP 和 PCR 方法,对胃癌候选基因 IL-1B-31、IL-1B-511、TNF-A-308 和 MIF-173 的单核苷酸多态性和 IL-1RN 的数目可变的串联序列重复多态性进行基因频率的比较²¹。

2. 单位点 H-W 平衡检验³¹: $\chi^2 = \sum (\text{期望值} - \text{观察值})^2 / \text{期望值}$ 。

3. 双位点组合遗传平衡检验:各组双位点表型组合 H-W 平衡期望值 χ^2_{τ} 检验⁴¹。

4. 单体型分析和连锁不平衡分析:运用 SNPAP 软件对 IL-1B-511、IL-1B-31 及 IL-1RN 进行 Haplotype 单体型分析;用 Haploview 4.0 软件进行连锁不平衡分析,确定连锁不平衡指数 D'。

5. Cochran-Armitage 趋势检验:趋势检验针对的是风险因素有多种状态,各状态对患病与正常的影响程度⁵¹,每一列认为是风险因素状态的某种次序,即前一种状态对疾病的影响要大于后一种状态对疾病的影响(表 1)。

计算公式

$$\chi^2_{\sigma} = \frac{n[n(x_1 \times D_2 + D_1) - n_1(x_1 \times D_2 + D_1 + x_1 \times C_2 + C_1)]^2}{n_1 \times n_2 \times [n(x_1^2 D_2 + D_1 + x_1^2 C_2 + C_1) - (x_1 \times D_2 + D_1 + x_1 \times C_2 + C_1)^2]}$$

表 1 对各位点基因多态性的状态进行 Cochran-Armitage 趋势检验

疾病状态	基因型			合计
	AA	Aa	aa	
病例组	$D_2(d_2)$	$D_1(d_1)$	$D_0(d_0)$	n_1
对照组	$C_2(c_2)$	$C_1(c_1)$	$C_0(c_0)$	n_2
合计	$D_2 + C_2(r_2)$	$D_1 + C_1(r_1)$	$D_0 + C_0(r_0)$	n
影响值	χ	1	0	

结果

1. 炎性基因单位点 H-W 平衡检验:表 2 显示 IL-1B-31, IL-1B-511, IL-1RN 和 TNF-A-308 各组 χ^2 (df=1)值均大于 6.63(P<0.01),吻合度不佳,研究人群处于遗传不平衡状态。MIF-173 位点病例组 $\chi^2 = 4.07$;对照组 P>0.05;可认为趋向于遗传平衡状态。

2. 双位点分析与遗传平衡:为观察不同基因位点之间相互作用的结果对遗传平衡的影响,进行双位点分析⁴¹,结果见表 3。

病例组和对照组的双位点组合 H-W 期望值 χ^2_{τ} 检验,对遗传平衡的偏离程度较高(P<0.01),且 MIF-173 与其他 4 个位点联合后均显著偏离平衡,提示双位点遗传平衡检验模式对于选择压力的反应更敏感。

3. 单体型分析和连锁不平衡分析:以 IL-1B-511、IL-1B-31 及 IL-1RN 的单体型和连锁不平衡分析结果为例说明连锁不平衡和群体遗传平衡状态偏离的关系。

表 4 以 CC*L 为参照单体型,CT*L 单体型是胃癌的保护因素(Δ),TT*L、TT*2、CT*2、TC*2 和 CC*2 单体型是胃癌的危险因素;各组内不同位点等位基因间连锁不平衡系数 D' 在 0.03 和 0.43 之间,IL-1B-511、IL-1B-31、IL-1RN 有连锁关系。

4. 群体分层与 Cochran-Armitage 趋势检验:表 5 中 IL-1B-511 (P<0.01) 和 IL-1RN (P<0.05) 位点趋势检验的差异有统计学意义,提示两位点 SNP 的 2 个等位基因并不独立,可能存在正相关,群体中可能

表 2 炎性基因多态性各位点的基因型频率和单位点 H-W 平衡检验

基因多态性 位点	病例家系			χ^2 值	对照家系			χ^2 值
	例数(基因型频率%)				例数(基因型频率%)			
	野生型	杂合子	纯合子		野生型	杂合子	纯合子	
IL-1B-31	117(39.5)	53(17.9)	126(42.6)	121.83*	107(33.5)	60(18.8)	152(47.6)	121.12*
IL-1B-511	76(25.7)	94(31.8)	126(42.6)	35.48*	158(49.5)	54(16.9)	107(33.5)	135.84*
IL-1RN	185(62.5)	41(13.9)	70(23.6)	134.36*	267(83.7)	33(10.3)	19(6.0)	72.58*
TNF-A-308	142(48.0)	66(22.3)	88(29.7)	85.90*	237(74.3)	50(15.7)	32(10.0)	69.27*
MIF-173	168(56.8)	101(34.1)	27(9.1)	4.07 ^b	193(60.5)	114(35.7)	12(3.8)	0.93

注: *P<0.01; ^bP<0.05

表 3 双位点表型组合概率的 H-W 期望值 χ^2 检验

分组	双位点组合	χ^2 值	P 值
病例组	IL-1B-31/MIF-173	138.04	0.00
	IL-1B-511/MIF-173	44.49	0.00
	IL-1RN/MIF-173	190.21	0.00
	TNF-308/MIF-173	97.97	0.00
对照组	IL-1B-31/MIF-173	134.67	0.00
	IL-1B-511/MIF-173	152.27	0.00
	IL-1RN/MIF-173	83.44	0.00
	TNF-308/MIF-173	109.45	0.00

表 4 炎性基因 IL-1B-31、IL-1B-511 和 IL-1RN 位点单体型分析和连锁不平衡分析

单体型	病例家系 (%)	对照家系 (%)	χ^2 值	P 值	基因组合	连锁不平衡系数 (D')
CC*L	63(10.6)	87(13.6)	参照	-	病例组	
TC*L	155(26.2)	154(24.1)	2.70	0.10	IL-1B-31-511	0.21
CT*L Δ	108(18.2)	253(39.7)	6.95	0.01	IL-1B-511-1RN	0.04
TT*L	85(14.4)	73(11.4)	4.29	0.04	IL-1B-31-1RN	0.18
CC*2	27(4.6)	10(1.6)	11.41	0.00*	对照组	
CT*2	50(8.4)	22(3.4)	11.66	0.00*	IL-1B-31-511	0.43
TC*2	43(7.3)	23(3.6)	9.83	0.00*	IL-1B-511-1RN	0.03
TT*2	61(10.3)	16(2.5)	28.44	0.00*	IL-1B-511-1RN	0.26
合计	592(100.0)	638(100.0)	-	-	-	-

注:同表 2; D' 为双等位基因配对检验方法得出的连锁不平衡系数

表 5 炎性基因多态性各基因型数据的 Cochran-Armitage 趋势检验

位点	炎性因子基因型(影响值)			χ^2 值	P 值
IL-1B-31	TT(0.89)	CT(1)	CC(0)	0.10	0.75
IL-1B-511	TT(0.70)	CT(1)	CC(0)	12.58	0.00*
IL-1RN	22(2.97)	L2(1)	LL(0)	7.11	0.01 ^a
TNF-A-308	AA(2.09)	GA(1)	GG(0)	2.08	0.15
MIF-173	CC(2.54)	GC(1)	GG(0)	2.87	0.09

注: *P<0.01; ^aP<0.05

存在群体分层或近亲婚配。

讨 论

评估标记基因型频率间的 H-W 平衡是群体遗传学数据研究的基础^[6], 根据炎性基因多态性与胃癌易感性家系比较研究的数据结果分析基因多态性与疾病关联研究中 H-W 平衡的影响因素, 揭示群体中等位基因频率在逐代传递中平衡偏离的原因。

1. 遗传流行病学病因推断中孟德尔群体遗传定律的影响因素: 以家系为单位比较不同亲属关系的病例和对照的炎性基因多态性的基因频率, 表 2 单位点 H-W 平衡检验显示 IL-1B-31、IL-1B-511、IL-1RN 和 TNF-A-308 各位点均显著背离 H-W 平衡, 而 MIF-173 位点可认为趋向于遗传平衡状态。研究对象来自于同一个人群, 而 H-W 平衡检验结果相

驳, 显示研究群体从基因位点发生原始突变时的原始群体到调查人群所经历的世代数不同, 致使不同位点的等位基因在人群中的频率不同, 对抗 H-W 平衡的能力也不同, 世代数每增加一次就向平衡理论值逼近一步。单体型分析以 CC*L 作为参照, 5 种单体型与胃癌相关且在病例组人群中常见, 提示等位基因的遗传异质性是胃癌发病的基础, 说明胃癌在群体遗传学史上至少发生了 5 次独立的突变, 显示了不同位点基因型的差异源于原始突变的不同, 发生多次独立的突变至少产生了 3 种不同的基因型, 等位基因遗传异质性使不同位点的等位基因频率在群体中分布不同, 各位点等位基因频率在病例组和对照组中差异均有统计学意义亦使 H-W 平衡复杂化, 产生偏离甚至相驳的现象。

2. 双位点表型组合分析与遗传平衡: 单位点 MIF-173 的 H-W 平衡检验趋于遗传平衡状态, 对其与另 4 个位点进行表型组合分析, 表 3 各组双位点表型组合的 H-W 期望值 χ^2 检验对遗传平衡的偏离程度不同但均有显著差异。偏离程度的不同提示不同基因双位点表型组合适合度的差异, 也表明选择作用的多向性; 平衡选择作用将有利于基因系统中长期保持 2 个基因位点多态现象, 也提示双位点遗传平衡检验模式对于选择压力的反应更敏感, 更能反映基因表型组合适合度的差异, 表明自然选择对炎性因子基因多态性演变的影响是通过复杂的基因关系网络而起作用, 作用的最终结果可能使群体遗传结构处于十分微弱的演变之中, 以至在一个世代中用单位点 H-W 平衡检验方法不能察觉。

3. 突变、选择、与连锁不平衡相结合的频率依赖性选择与 H-W 平衡: 对于随机交配中独立分配的基因座, 不连锁的基因有 50% 的重组机会, 而连锁基因趋于平衡的速率较慢, 因为它们重组频率比不连锁的基因低, 连锁越紧密, 达到平衡需要的时间越长^[7]。表 4 单体型分析显示: 特定的单体型与单位点突变相关性病例组比对照组人群更常见, IL-1B-511、IL-1B-31、IL-1RN 同处于一条染色体上, 距离很近^[8], 不同位点等位基因间的非随机性组合分析显示 IL-1B-511、IL-1B-31、IL-1RN 标记基因和疾病基因之间存在连锁不平衡, 可能是在携带特定单型型的染色体上, 某次发生了致病的突变, 但在原始突变事件后时间很短, 不足以经过反复互换引起随机化或连锁平衡, 紧密连锁基因座的三种等位基因组成的 CC*L, TC*L 和 CT*L 单体型具有选择优势而在群体中长久保存, 作用的结果使群体遗传结

构处于不断的演变之中,产生连锁不平衡而引起基因型频率的变化,导致对遗传平衡的偏离。

对同一人群的前期研究显示:与 GG 野生型比较, TNF-A-308 位点突变型纯合子对胃癌的发病风险要高于杂合子(A/A:OR=4.59;G/A:OR=2.20)^[2], A/A 与 G/G 比较,在一级亲属中的风险高于二级亲属(一级亲属:OR=5.51;二级亲属:OR=2.09),突变导致的选择机制不利于等位基因的携带者(TNF-A-308A),尤其不利于纯合子个体(TNF-A-308A/A),使其基因型的适合度降低。决定劣势性状的 A 等位基因的拷贝频率在未来世代的群体中降低,经多代淘汰,有害的 A 基因逐渐减少,随亲属血缘关系的疏远其优势比也降低。很大程度上影响了人群中的基因组成,引起各等位基因频率在人群中变化,提示突变是变异的主要动力,使选择作用发生效应,是一种潜在的改变群体中等位基因频率的力量,为影响群体的基因平衡的重要因素。

4. 群体分层与 H-W 平衡:Wigginton 等^[9]指出, H-W 检验会增大犯第 I 类错误的概率。有学者认为通过测试 H-W 平衡,基因分型的错误是不太可能被显示出来的^[10]。Relton 等^[11]建议对等位基因使用 Cochran-Armitage 趋势检验基因型计数的趋势,其理论基础是连锁不平衡^[12]。表 5 炎性基因多态性各基因型数据的 Cochran-Armitage 趋势检验显示 IL-1B-511 和 IL-1RN 的 2 个位点与 SNP 的 2 个等位基因并不独立,可能存在正相关,各位点的单核苷酸多态性之间由于遗传背景的差异导致群体分层,或者存在近亲婚配现象,加上地理、文化等原因,使群体内的 H-W 平衡定律偏离。

与连锁不平衡相结合的频率依赖性选择、突变及其共同作用的影响;群体分层或者近亲婚配现象的存在;造成基因频率在群体中的增加或减少,影响基因多态性与疾病关联研究中的 H-W 平衡。研究抽取的样本量不够大,效应估计值较小,统计效能低;基因-环境交互作用等混杂因素的影响也会使 H-W 平衡复杂化;致使孟德尔随机化原则不能起到作用,而产生偏性结果。基因多态性与疾病关联研究中,双位点分析和 Cochran-Armitage 趋势检验能

更准确的解释对遗传平衡的偏离程度和背离平衡的原因。当然,疾病关联研究中 H-W 平衡的影响因素分析是基于对前期研究的基因多态性的结果而进行的,随着遗传流行病学和群体遗传学的进一步深入研究,将能更深刻的阐明肿瘤遗传平衡复杂化的原因。

参 考 文 献

- [1] 罗庆. 自交率对 Hardy-Weinberg 平衡的效应. 遗传, 1985, 7(1): 46-47.
- [2] Feng YX, Zhang JZ, Dai LP, et al. Inflammatory cytokine gene polymorphisms in gastric cancer cases and controls family members from Chinese areas at high cancer prevalence. *Cancer Letters*, 2008, 270(2): 250-259.
- [3] 陈莉雅, 陈平燕. 应正确把握遗传研究中有关病例对照研究的前提条件. 中华流行病学杂志, 2007, 28(11): 1140.
- [4] 罗佳滨, 李月秋, 王庆元, 等. Hp, Gc, ABO 三个非连锁基因系统间双位点遗传平衡检验. 遗传, 1994, 16(2): 10-14.
- [5] 李赵海, 覃红, 张红. 基于群体数据的关联分析. 遗传学中的统计方法. 北京: 科学出版社, 2006: 85-87.
- [6] Daniel JS, Anthony JB, Gregory DJ, et al. Exact tests of Hardy-Weinberg equilibrium and homogeneity of disequilibrium across strata. *Am J Hum Genet*, 2006, 12(79): 1071-1079.
- [7] Susan Elrod, William Stansfield. *Population Genetics and Development. Genetics*. 田清涛, 等译. 4 版. 北京: 科学出版社, 2004: 227-230.
- [8] Perri F, Piepoli A, Bonvicini C, et al. Cytokine gene polymorphisms in gastric cancer patients from two Italian areas at high and low cancer prevalence. *Cytokine*, 2005, 30(5): 293-302.
- [9] Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet*, 2005, 76: 887-893.
- [10] Zou GY, Donner A. The merits of testing Hardy-Weinberg equilibrium in the analysis of unmatched case-control data: a cautionary note. *Ann Hum Genet*, 2006, 70: 923-933.
- [11] Relton CL, Wilding CS, Pearce MS, et al. Gene-gene interaction in folate-related genes and risk of neural tube defects in a UK population. *J Med Genet*, 2004, 41: 256-260.
- [12] Jonathan KP, Noah AR. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*, 1999, 65: 220-228.

(收稿日期: 2009-04-06)

(本文编辑: 张林东)