科学通报 2009 年 第 54 卷 第 10 期: 1429 ~ 1435
论 文 www.scichina.com csb.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

次要等位基因频率筛选阈值对人类单体型图精度的影响: 基于中国汉族与藏族人群测序数据的比较

熊诗诣[①], 郝元涛[②], 饶绍奇[②], 黄玮俊[①], 胡彬[①], 拉布[③], 普布卓玛[③], 格桑卓嘎[③], 王一鸣[①*]

① 中山大学中山医学院医学遗传教研室, 广州 510089;
② 中山大学公共卫生学院医学统计与流行病学系, 广州 510089;
③ 西藏自治区第二医院妇产科, 拉萨 850002
* 联系人, E-mail: ywzhong@hotmail.com

2009-01-12 收稿, 2009-03-27 接受
教育部"985 工程"二期、广东省自然科学基金(批准号: 031673)、广州市科技计划基金(批准号: 2002Z3-C7191, 2004Z3-C7501)资助项目

摘要 基因组变异是个体间疾病易感性和药物反应等表型多样性的遗传基础. 国际人类单体型图(International HapMap)旨在为复杂疾病相关遗传变异的研究提供路线图. 单核苷酸多态性(SNPs)是 HapMap 的基本要素. SNPs 等位基因频率影响连锁不平衡结构、单体型的构建、标签 SNPs 的筛选, 是影响 HapMap 精度的主要因素之一. 因此, 次要等位基因频率筛选阈值的选择对图谱精度有深远影响. 迄今大多数研究者选用自定的阈值, 且鲜有针对次要等位基因频率筛选阈值对 HapMap 精度影响的研究. 为探讨次要等位基因频率筛选阈值对相应 HapMap 精度的影响, 本研究用中国汉、藏族人群 15 号染色体中心粒区域基因的测序结果按不同次要等位基因频率筛选阈值(≥0.01, ≥0.05, ≥0.10)将以往的数据分成了 3 组, 即 0.01 组、0.05 组以及 0.10 组, 分别构建了 3 组数据的 HapMap, 并比较了各组 HapMap 精度、关联分析的研究效能及节约/总成本比值. 结果显示, 0.01 组有最高的关联分析研究效能(相比 0.05 组: 汉族, $P = 0.019$; 藏族, $P = 0.029$), 并捕获了最多的人群特异性单体型(相比 0.05 组, $P = 0.012$). 在所检区域内, 与 0.10 阈值相比, 0.05 阈值并没有显著提高关联分析的研究效能(汉族, $P = 0.191$; 藏族, $P = 1.000$)及人群特异性单体型的捕获($P = 0.592$). 同时, 在藏族人群中, 0.05 与 0.10 组产生了相同数据的标签 SNPs 效率、连锁不平衡结构域的数目和平均长度、关联分析研究效能及节约/总成本比值. 结果提示, 较低的次要等位基因频率筛选阈值更适合着重于人群特异性单体型的研究; 不同人群最佳次要等位基因频率筛选阈值可能不尽相同. 由于本研究检测基因数目有限, 这一重要议题仍需更多深入的探讨.

关键词
单体型图
次要等位基因频率筛选阈值
单体型图精度

基因组变异(genomic variation)在决定诸如疾病易感性、进程及对药物反应等个体表型差异中起重要作用. 单核苷酸多态性(SNPs)间存在连锁不平衡(linkage disequilibrium, LD), 一个连锁不平衡结构域(LD block)中只存在数量有限的几个常见单体型(common haplotype), 且其大部分遗传信息能被少数几个标签SNPs(tagSNPs)所捕获的事实是国际人类单体型图(International HapMap)的基本原理[1]. HapMap

引用格式: 熊诗诣, 郝元涛, 饶绍奇, 等. 次要等位基因频率筛选阈值对人类单体型图精度的影响: 基于中国汉族与藏族人群测序数据的比较. 科学通报, 2009, 54: 1429~1435
Xiong S Y, Hao Y T, Rao S Q, et al. Effects of cutoff thresholds for minor allele frequencies on HapMap resolution: A real dataset-based evaluation of the Chinese Han and Tibetan populations. Chinese Sci Bull, 2009, 54: 2069−2075, doi: 10.1007/s11434-009-0302-4

旨在揭示全基因组LD结构、常见单体型组成、筛选相应tagSNPs, 为鉴定不同人群复杂疾病基因的研究提供强有力的工具.

第Ⅰ, Ⅱ期HapMap数据的发表及其在某些疾病易感基因鉴定上的成功应用, 使其研究策略的效能和优势已获肯定, 但图谱构建中仍有些问题并未解决. 遗传标记物(genetic markers)的等位基因频率(allele frequency)是影响图谱效能的主要因素之一, 它影响LD结构[2,3]、单体型组成和tagSNPs筛选[3]. 等位基因频率不仅在不同人群中存在差异, 同时还有人群特异性等位基因(population-specific allele)的存在[4]. 因此, 次要等位基因频率筛选阈值(minor allele frequency cutoff thresholds, MAF cutoff值)的选择对HapMap结构及精度有极大影响. 相对于较多的筛选tagSNPs的方法、构建LD结构的方法及其阈值的比较性研究, 迄今鲜有MAF cutoff值对HapMap影响的研究. 尽管多数人参照了国际HapMap计划的阈值. 大部分研究者自行选定自己的MAF cutoff值. 而国际HapMap计划以 0.05 为其MAF cutoff值[1,5]; 千人基因组计划(1000 genome project) Ⅰ期以 0.05, Ⅱ期则以 0.01 为其筛选阈值[6]; Framingham心脏研究计划以 0.10[7]; 而Iida以 0.20 为筛选阈值[8].

为评估不同MAF cutoff值对HapMap精度在关联分析中的效能及所需费用的影响, 本文根据中国汉族和藏族人群 15 号染色体中心粒区域基因的测序数据[9,10]针对MAF cutoff值对HapMap精度的影响进行了系统的对比分析, 并将结果报告如下.

1 材料与方法

(ⅰ) 数据的来源与分组. 所有SNPs数据来自本课题组以往对汉、藏人群各 50 例样本的测序结果[9,10], 并将所有数据按MAF≥0.01 (0.01 组), MAF≥0.05 (0.05 组), MAF≥0.10 (0.10 组)分为 3 组.

(ⅱ) 3 组单体型图(HapMap)的构建. 我们基于这 3 组数据分别构建了 3 个相应的单体型图. 构图方法见文献[9,10].

(ⅲ) 各组HapMap精度的评估. (1) tagSNPs所捕获的遗传信息. 参照Ke等人的方法[11]比较了 3 个HapMap中获取相同数目SNPs所需最少tagSNPs数目. 参照国际HapMap计划方法[1]比较了 3 个HapMap中每个tagSNP标记的平均基因组长度(kb). (2) 对单体型的影响. 参照Carlson等人的方法[12]比较了 3 个HapMap中单体型的组成及其频率. 参照Crawford等人的方法[13]比较了各图所捕获的人群特异性单体型(population-specific haplotypes)的数目. (3) 对LD 结构的影响. 分别比较了 3 个HapMap中LD block的平均长度以及每个基因内的平均LD blocks数目. (4) 3 个HapMap的关联分析研究效能(study power)和节约/总成本比值(cost-savings). 由于决定复杂疾病、性状的常见遗传变异对表型的影响往往不是很强, 研究效能就成为设计与评价关联分析的重要因素. 高精度遗传变异图谱可提高研究效能、提高信息的获取量, 但也增加了研究成本. 我们用CaTS power calculator 比较了 3 个HapMap的关联分析研究效能和节约/总成本比值[14], 并采用了这一软件设定的样本量(sample size)、显著水平(significant level)、患病率(disease prevalence)、基因型相对危险度(genotype relative risk)和遗传模型(genetic model)[14]. 节约/总成本比值(cost-savings)则用 3 个HapMap中检测tagSNPs相对于检测同一候选节段内所有SNPs所节约的成本占总成本的百分比表示[14].

(ⅳ) 统计学分析. 所有数值变量以 $\bar{x} \pm$ SD描述. 3 组(0.01 组、0.05 组以及 0.10 组)间的差异用随机区组设计的线性模型方差分析(ANOVA)[15]进行比较. 组间比较用最小显著差异法(least significant difference, LSD)[16]. 显著性水平α为 0.05. 所有统计分析用软件SPSS 13.0[17]进行.

2 结果

2.1 tagSNPs 捕获的遗传信息

(ⅰ) 获取相同数量SNPs所需最少tagSNPs数目. 当 MAF cutoff 值从 0.01 升至 0.05 时, 汉、藏两人群中获取相同数量 SNPs 所需最少 tagSNPs 数目显著降低(汉族人群: (6 ± 1)~(5 ± 2), P = 0.0014; 藏族人群: (5 ± 3)~(4 ± 3), P = 0.021). 但 0.05 与 0.10 组之间的差异无统计学意义(汉族人群: (5 ± 2)~(4 ± 2), P = 0.156; 藏族人群: (4 ± 3)~(4 ± 3), P = 1.000)(表 1~3).

表 1 获取相同数目 SNPs 所需最少 tagSNPs 数目

基因	汉族			藏族		
	0.01 组	0.05 组	0.10 组	0.01 组	0.05 组	0.10 组
NT_026446.32	6	4	4	6	4	4
NT_026446.33	5	4	4	1	1	1
NT_026446.34	6	4	4	5	2	2
NT_026446.37	8	8	7	8	6	6
NT_026446.72	6	5	3	7	7	7
均数	6	5	4	5	4	4
标准差	1	2	2	3	3	3

表 2　汉族人群变量统计学分析结果

	P 值			
	3 组比较 a)	0.01 组 *vs*.0.05 组 b)	0.01 组 *vs*.0.10 组 b)	0.05 组 *vs*. 0.10 组 b)
获取相同数目 SNPs 所需最少 tagSNPs 数目	0.004 c)	0.014 d)	0.002 c)	0.156
每个 tagSNP 标记的平均基因组长度	0.305	0.732	0.152	0.254
单体型总数	0.152	0.276	0.06	0.336
常见单体型数目	0.280	0.475	0.123	0.358
LD block 平均长度	0.140	0.791	0.072	0.110
各基因内平均 LD block 数目	0.066	0.077	0.028 d)	0.535
关联分析研究效能	0.007 c)	0.019 d)	0.002 c)	0.191
节约/总成本比值	0.053	0.260	0.019 d)	0.124

a) 随机区组设计的线性模型方差分析(ANOVA); b) 组间比较采用最小显著差异法(LSD); c) $0.001<P\leq0.01$; d) $0.01<P\leq0.05$

表 3　藏族人群变量统计学分析结果

	P 值			
	3 组比较 a)	0.01 组 vs.0.05 组 b)	0.01 组 vs.0.10 组 b)	0.05 组 vs. 0.10 组 b)
获取相同数目 SNPs 所需最少 tagSNPs 数目	0.032 c)	0.021 c)	0.021 c)	1.000
每个 tagSNP 标记的平均基因组长度	0.124	0.077	0.077	1.000
单体型总数	0.015 c)	0.024 c)	0.006 d)	0.382
常见单体型数目	0.190	0.203	0.084	0.573
LD block 平均长度	0.410	0.256	0.256	1.000
各基因内平均 LD block 数目	0.410	0.256	0.256	1.000
关联分析研究效能	0.045 c)	0.029	0.029	1.000
节约/总成本比值	0.028 c)	0.019 c)	0.019 c)	1.000

a) 随机区组设计的线性模型方差分析(ANOVA); b) 组间比较采用最小显著差异法(LSD); c) $0.01<P\leq0.05$; d) $0.001<P\leq0.01$

(ⅱ) 每个 tagSNP 标记的平均基因组长度. 每个 tagSNP 标记的平均基因组长度随 MAF cutoff 值升高而增加. 汉族人群: 0.01 组, (14.22 ± 17.82) kb; 0.05 组, (16.28 ± 20.45) kb; 0.10 组, (23.42 ± 33.76) kb (图 1(a)). 藏族人群: 0.01 组, (14.13 ± 14.96) kb; 0.05 组和 0.10 组均为(16.95 ± 16.07) kb (图 1(b)). 但 3 组间的差异在两人群中均无统计学意义(表 2 和 3).

2.2　对单体型的影响

(ⅰ) 单体型总数(total haplotypes). 3 个 HapMap 的单体型总数随 MAF cutoff 值升高而下降:

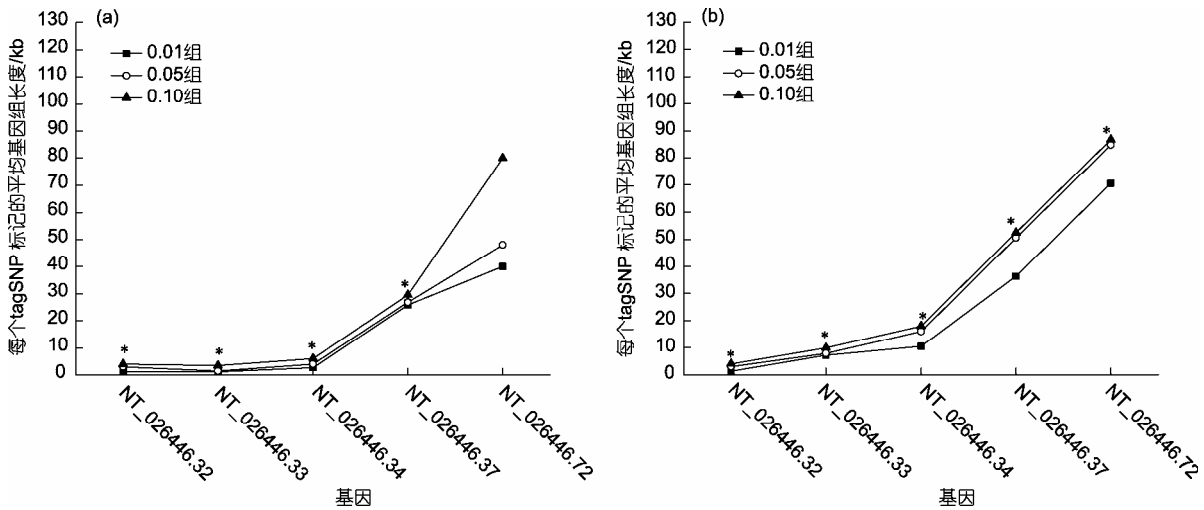

图 1　各 MAF cutoff 组所构建的 HapMap 中每个 tagSNP 标记的平均基因组长度(kb)

(a) 汉族人群; (b) 藏族人群. *表示 MAF cutoff 0.05 与 0.10 组数值相同

汉族人群, 0.01 组为 14 ± 7, 0.05 组 12 ± 7, 0.10 组 11 ± 6(图 2(a)), 但 3 组间差异无统计学意义(P = 0.152)(表 2); 藏族人群中变化趋势相同, 0.01 组 14 ± 10, 0.05 组 12 ± 10, 0.10 组 12 ± 9, 但其差异有统计学意义(P = 0.015)(图 2(b), 表 3).

(ⅱ) 常见单体型. 常见单体型(频率≥5%的单体型)的数目随 MAF cutoff 值升高缓慢下降: 汉族人群, 0.01 组、0.05 组、0.10 组常见单体型的数目分别为 4.5 ± 2.6, 4.3 ± 2.5 和 4.0 ± 2.4(图 2(c)); 藏族人群变化趋势相同, 分别为 4.5 ± 3.7, 4.0 ± 3.1 和 3.8 ± 2.9(图 2(d)). 但两人群 3 组间差异均无统计学意义(表 2, 3).

(ⅲ) 人群特异性单体型. 汉藏两组人群特异性单体型共同比较结果显示, MAF cutoff 值由 0.01 升至 0.05 时, 不同 HapMap 所捕获的人群特异性单体型的数目下降, 有显著性意义(汉族人群: (8.6 ± 5.6)~(4.2 ± 4.3); 藏族人群: (7.8 ± 3.6)~(5.2 ± 3.7), P = 0.012). 但 MAF cutoff 值由 0.05 升至 0.10 时, 差异无统计学意义(汉族人群: (4.2 ± 4.3)~(3.8 ± 3.9); 藏族人群: (5.2 ± 3.7)~(4.2 ± 3.4), P = 0.592)(图 3, 表 4).

2.3 对 LD 结构的影响

(ⅰ) LD block 平均长度. MAF cutoff 值由 0.01 升至 0.05 时, HapMap 所显示的两人群 LD block 平均长度均增加(汉族人群: (21.2 ± 21.2)~(22.4 ± 21.2) kb; 藏族人群: (23.2 ± 24.0)~(24.5 ± 23.2) kb). MAF cutoff 值由 0.05 升至 0.10 时, 汉族人群 LD block 平均长度增加((22.4 ± 21.2)~(30.5 ± 33.1) kb), 但藏族人群却没有变化(表 5). 两人群的变化均无统计学意义(表 2 和 3).

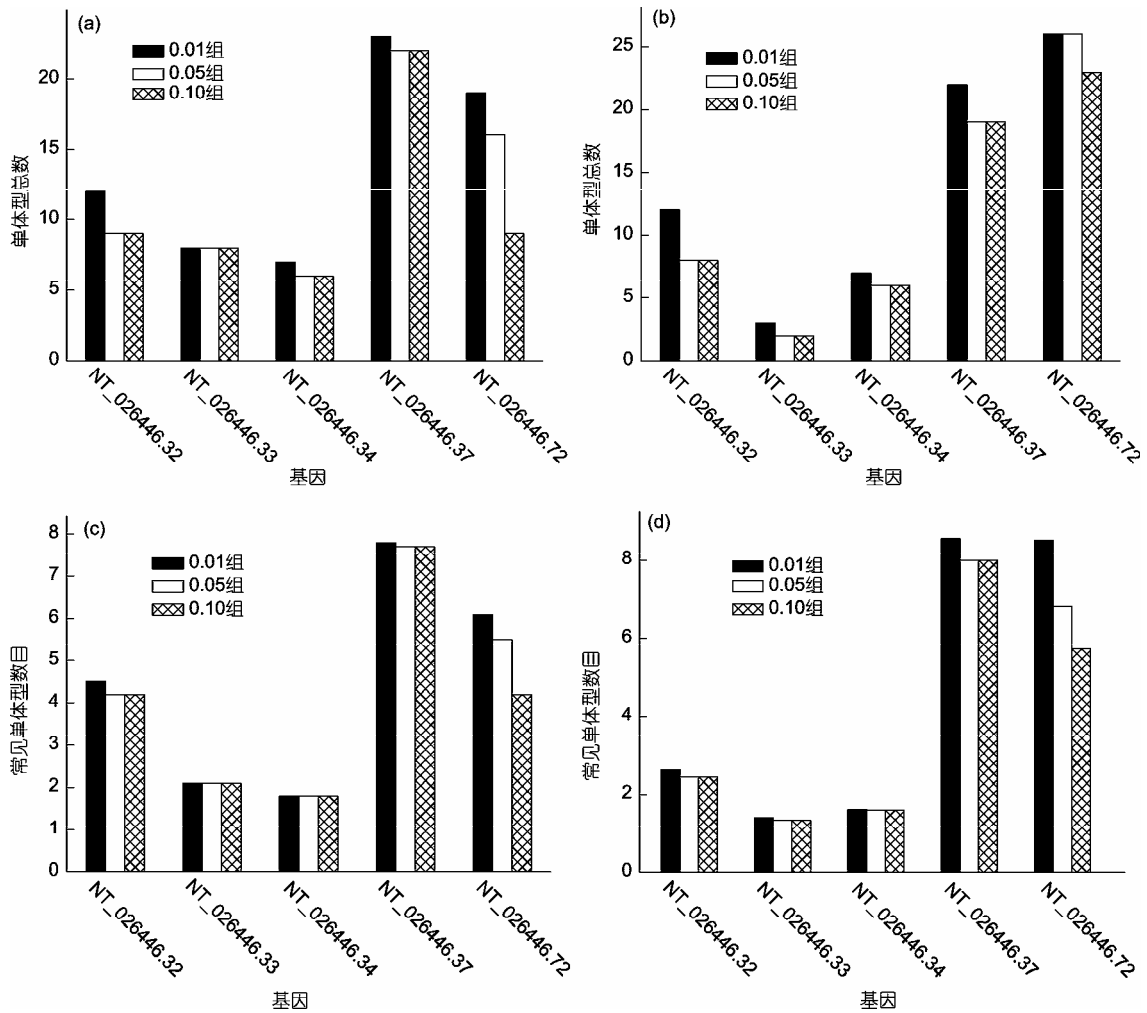

图 2 各 MAF cutoff 组所构建的 HapMap 中单体型总数和常见单体型数目

(a) 汉族人群单体型总数; (b) 藏族人群单体型总数; (c) 汉族人群常见单体型数目; (d) 藏族人群常见单体型数目

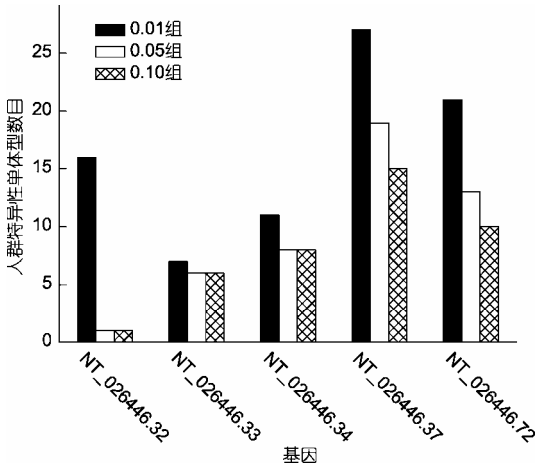

图 3　各 MAF cutoff 组构建的 HapMap 中人群特异性单体型数目

3 组间比较 $P = 0.008$. 两组比较: 0.01 *vs*. 0.05 组: $P = 0.012$; 0.01 *vs*. 0.10 组: $P = 0.003$; 0.05 *vs*. 0.10 组: $P = 0.592$

表 4　3 组 MAF cutoff 值构建的 HapMap 所捕获的人群特异性单体型统计学分析结果

	P 值			
	3 组比较 [a)]	0.01 组 *vs*. 0.05 组 [b)]	0.01 组 *vs*. 0.10 组 [b)]	0.05 组 *vs*. 0.10 组 [b)]
人群特异性单体型数目	0.008 [c)]	0.012 [d)]	0.003 [c)]	0.592

a) 随机区组设计的线性模型方差分析(ANOVA); b)组间比较采用最小显著差异法(LSD); c) $0.001 < P \leq 0.01$; d) $0.01 < P \leq 0.05$

(ⅱ) 各基因内 LD blocks 的均数. 相应于 LD block 平均长度变化, 各基因内平均 LD blocks 数目在两人群中随 MAF cutoff 值升高而减少(汉族人群: 0.01 组, 2.8 ± 1.6; 0.05 组, 2.0 ± 1.0; 0.10 组, 1.8 ± 0.8. 藏族人群: 0.01 组, 2.4 ± 1.1; 0.05 组和 0.10 组均为 2.2 ± 1.3. 表 5). 但 3 组间的差异在汉藏两人群中均无统计学意义(表 2 和 3).

表 5　各 MAF cutoff 值所构建的 HapMap 中 LD block 数目与 LD block 平均长度

基因	MAF cutoff	LD block 数目		LD block 平均长度/kb	
		汉族	藏族	汉族	藏族
NT_026446.32	0.01	4	2	1.6	3.9
	0.05	2	2	3.5	3.9
	0.10	2	2	3.5	3.9
NT_026446.33	0.01	1	1	6.0	6.0
	0.05	1	1	6.0	6.0
	0.10	1	1	6.0	6.0
NT_026446.34	0.01	1	2	16.0	9.5
	0.05	1	1	16.0	16.0
	0.10	1	1	16.0	16.0
NT_026446.37	0.01	4	3	54.5	38.6
	0.05	3	3	54.5	38.6
	0.10	3	3	81.8	38.6
NT_026446.72	0.01	4	4	27.6	57.9
	0.05	3	4	31.9	57.9
	0.10	2	4	45.3	57.9

2.4　关联分析的研究效能和节约/总成本比值.

(ⅰ) 研究效能. MAF cutoff 值由 0.01 升至 0.05 时, 两人群关联分析的研究效能均显著下降(汉族人群: (84.1 ± 14.6)%降至(75.5 ± 14.1)%, $P = 0.019$; 藏族人群: (75.3 ± 13.7)%降至(64.8 ± 2.7)%, $P = 0.029$). 而 MAF cutoff 值由 0.05 升至 0.10 时, 基于我们所检测基因而构建的 HapMap 虽然显示了研究效能的降低, 但这一降低并没有统计学意义(汉族人群: 0.10 组, (71.3 ± 13.9)%, $P = 0.191$; 藏族人群 0.10 与 0.05 组研究效能相同, $P = 1.000$)(图 4, 表 2 和 3).

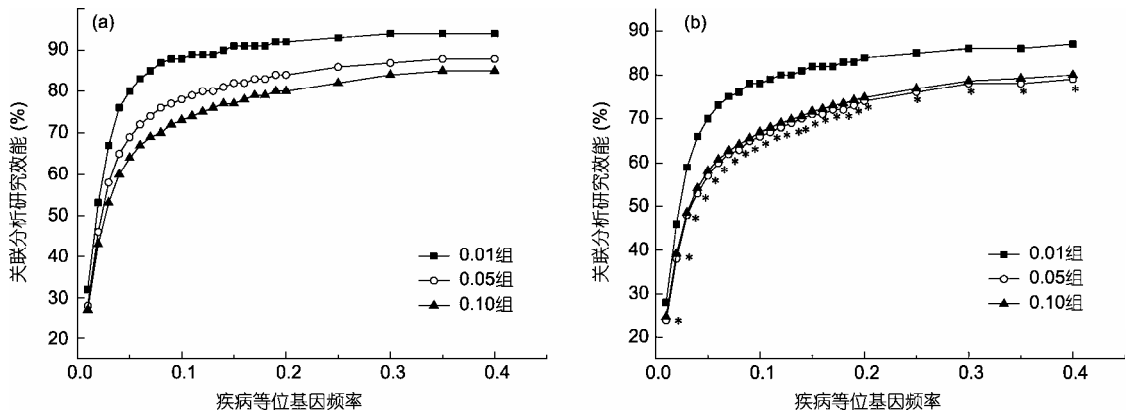

图 4　各 MAF cutoff 组所构建的 HapMap 中在关联分析中的研究效能(%)

(a) 汉族; (b) 藏族. * 表示 0.05 组与 0.10 组数值相同. 3 组比较: 汉族人群, $P = 0.007$; 藏族人群, $P = 0.045$. 两组比较: 0.01 较 0.05 组研究效能显著增高(汉族人群, $P = 0.019$; 藏族人群, $P = 0.029$). 但 0.05 与 0.10 组的差异无统计学意义(汉族人群, $P = 0.191$; 藏族人群, $P = 1.000$)

(ii) 节约/总成本比值. 3 个 HapMap 的节约/总成本比值随 MAF cutoff 值升高而提高. 汉族人群, 3 组(0.01 组、0.05 组和 0.10 组)节约/总成本比值分别为(18.1 ± 16.5)%, (34.4 ± 20.2)%和(55.2 ± 21.7)% (图 5). 3 组比较产生了边缘性 P 值 0.053 (表 2). 藏族人群的变化趋势相同(0.01 组: (28.9 ± 34.2)%, 0.05 组及 0.10 组均为(46.4 ± 29.2)%), P = 0.028(图 5, 表 3).

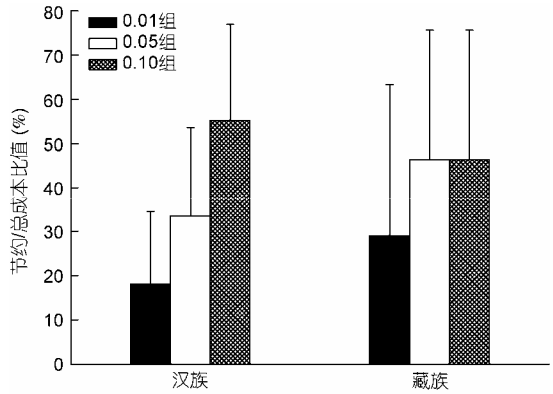

图 5 节约/总成本比值

各 MAF cutoff 组所构建的 HapMap 中检测 tagSNPs 相对于检测同一候选节段内所有 SNPs 所节约的成本占总成本的百分比

4 讨论

人类HapMap项目的主要目的在于为复杂疾病的相关基因鉴定提供有效的工具[18,19]. 在该计划获得成功的同时, 也面临一些未解决的问题, 如HapMap捕获或丢失了多少遗传信息, 如何在成本与效益之间取得平衡, 以及在“重测序”(re-sequencing)和“深度测序”(deep sequencing)研究[20,21]中如何解决人群差异性等. 与此相关的正是构建HapMap时MAF cutoff值的选择. 正如预期的一样, 本研究表明, 在 3 个HapMap中以MAF cutoff值 0.01 所构建的图谱捕获了最多的遗传信息, 但成本也最高. 在所检测的基因中, MAF cutoff值从 0.05 到 0.10 的改变并没有显著地提高两人群的HapMap研究效能(汉族人群, P = 0.191; 藏族人群, P = 1.000), 也没有捕获更多人群特异性单体型(P = 0.592). 这一 0.05 HapMap与 0.10 HapMap的相似性在藏族人群中更为明显, 其产生了相同数据的标签效率(tagging efficiency)、LD blocks数目、LD block平均长度、关联分析研究效能和节约/总成本比值. 因此, 在本次研究的区域内, 藏族人群的结果显示, 相比 0.10 MAF cutoff值, 0.05 并没有提高图谱精度. 这些结果提示, 最佳MAF cutoff值应根据不同的研究目的而确定. 低MAF cutoff值更适于以人群特异性单体型为主要目的的研究. 不同人群的最佳MAF cutoff值可能不尽相同. 在大样本量的研究实施前, 进行预实验探索最佳cutoff值是可取的方法. 显然, 如果经费充足, cutoff值 0.01 所构建的图谱可以捕获最多的遗传信息. 由于本次实验所研究的基因数目有限, 对我们的结果, 尤其是阴性结果的解释应慎重. 国际科学界对这一重要课题仍应进行进一步的研究.

致谢 感谢所有血样的捐献者, 感谢李彩霞博士、张晋昕教授对本研究在统计学分析中的支持.

参考文献

1 The International HapMap Consortium. A haplotype map of the human genome. Nature, 2005, 437: 1299—1320[DOI]
2 Payseur B A, Place M, Weber J L, et al. Linkage disequilibrium between STRPs and SNPs across the human genome. Am J Hum Genet, 2008, 82: 1039—1050[DOI]
3 Gabriel S B, Schaffner S F, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science, 2002, 296: 2225—2229[DOI]
4 Guthery S L, Salisbury B A, Pungliya M S, et al. The structure of common genetic variation in United States populations. Am J Hum Genet, 2007, 81: 1221—1231[DOI]
5 The International Hapmap consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature, 2007, 449: 851—861[DOI]
6 Siva N. 1000 Genomes project. Nat Biotechnol, 2008, 26: 256
7 Vasan R S, Larson M G, Aragam J, et al. Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. BMC Med Genet, 2007, 8(Suppl 1): S2[DOI]
8 Iida A, Ozaki K, Tanaka T, et al. Fine-scale SNP map of an 11-kb genomic region at 22q13.1 containing the galectin-1 gene. J Hum

Genet, 2005, 50: 41—45
9 黄玮俊, 李彩霞, 陈素琴, 等. 中国汉族人群 15 号染色体中心粒区域 5 个基因的高精度单倍型及单倍型域构建. 科学通报, 2004, 49: 649—656
10 黄玮俊, 李彩霞, 拉布, 等. 藏族人群 15 号染色体中心粒区域基因的高精度连锁不平衡和单体型图谱及其与汉族人群的比较. 科学通报, 2006, 5: 283—291
11 Ke X, Durrant C, Morris A P, et al. Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples. Hum Mol Genet, 2004, 13: 2557—2565[DOI]
12 Carlson C S, Eberle M A, Rieder M J, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet, 2004, 74: 106—120[DOI]
13 Crawford D C, Carlson C S, Rieder M J, et al. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am J Hum Genet, 2004, 74: 610—622[DOI]
14 Skol A D, Scott L J, Abecasis G R, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet, 2006, 38: 209—213[DOI]
15 Lindman H R. Analysis of variance in complex experimental designs. San Francisco: W. H. Freeman & Co, 1974
16 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc Ser B, 1995, 57: 289—300
17 Nie N H, Hull C H, Jenkins J G, et al. SPSS: Statistical Package for the Social Sciences, Berkshire: McGraw-Hill Education, 1975
18 Wang W Y, Barratt B J, Clayton D G, et al. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet, 2005, 6: 109—118[DOI]
19 McCarthy M I, Abecasis G R, Cardon L R, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet, 2008, 9: 356—369[DOI]
20 Anderson C A, Pettersson F H, Barrett J C, et al. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. Am J Hum Genet, 2008, 83: 112—119[DOI]
21 Bhangale T R, Rieder M J, Nickerson D A, et al. Estimating coverage and power for genetic association studies using near-complete variation data. Nat Genet, 2008, 40: 841—843[DOI]

·书 讯·

《基因组研究手册》

——基因组学、蛋白质组学、代谢组学、生物信息学、伦理和法律问题

(翻译版)

著者：〔加拿大〕**C. W.** 森森
译者：谢东
出版：科学出版社　**2009** 年 **3** 月
定价：**95.00**



本书旨在为读者引入“组学”的各种基本概念，了解其基本内容，介绍各种常用的“组学”技术，并结合对实例的引用和阐述，展示“组学”的应用及前景. 同时，就“组学”的迅猛发展所带来的社会伦理问题，进行了开放性的讨论. 基于上述目的，本书分为 4 个部分：第一部分，关键的生物，对“组学”进行了概括介绍；第二部分，基因组和蛋白质组技术，结合应用实例，详细地介绍了基因组学和蛋白质组学的技术发展；第三部分，生物信息学，对这一在组学研究和应用中不可或缺的工具进行了细致全面的阐述；第四部分，伦理、法律和社会问题，归纳和展示了由“组学”的迅速发展所带来的不可避免的社会问题，并开展了多方面、多视角的讨论.

本书可供分子生物学、生物技术、医药卫生，以及生命科学领域相关科研、技术人员，研究生和高年级本科生参考使用.