

中国北京汉族人群和日本东京人群 ROR2 基因单核苷酸多态性比较

王红 赵凯平

【摘要】 目的 比较中国北京汉族人群(CHB)和日本东京人群(JPT)受体酪氨酸激酶样孤立受体2(ROR2)基因单核苷酸多态性(SNP)的异同。方法 收集国际人类基因组单体型图计划(HapMap)公布的CHB及JPT的ROR2基因SNP数据,利用Haploview和SPSS 13.0软件区分纯合与非纯合基因型SNP,以基因型测定率 $>80\%$ 、H-W平衡($P>0.01$)、MAF >0.01 且无性别差异($P>0.05$)为标准,在非纯合基因型SNP中确定合格SNP;标签SNP根据 $r^2\geq 0.8$ 或对数优势比(LOD) ≥ 3 进行确定;并比较合格SNP及标签SNP的MAF、单体域及单体型频率。结果 CHB和JPT的ROR2基因SNP在两人群中名称相同且均为404个,其中纯合基因型SNP为101个。两人群共有257个名称相同的合格SNP,其中224个(87.2%)SNP在两人群中的微效等位基因亦相同。257个合格SNP在CHB和JPT中分别形成18和27个单体域,除2个为JPT独有外,其余单体域在两人群完全或部分重叠。两人群中名称相同的标签SNP为50个,分别占CHB和JPT各自标签SNP的64.9%和70.4%。结论 通过分析HapMap数据可避免将纯合基因型SNP纳入遗传易感性研究,CHB和JPT的ROR2基因SNP及单体域特征以共性为主,与人群的地域和种族属性一致。

【关键词】 受体酪氨酸激酶样孤立受体2;单核苷酸多态性

Comparison of minor allele frequency and haplotype frequencies for single nucleotide polymorphisms in receptor tyrosine kinase-like orphan receptor 2 gene using HapMap data from Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) WANG Hong, ZHAO Kai-ping. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
Corresponding author: WANG Hong, Email: hwang2010@yahoo.cn

This work was supported by grants from the US National Institute of Dental and Craniofacial Research (No. R21-DE-013707, R01-DE-014581) and Fogarty Institution (No. D43-TW006176).

【Abstract】 **Objective** Single nucleotide polymorphisms (SNPs) in receptor tyrosine kinase-like orphan receptor 2 (ROR2) gene were analyzed and compared between Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) using the HapMap data, to provide basis for SNP determination. ROR2 gene related etiologic studies were conducted in the above mentioned two populations. **Methods** Monotonic and un-monotonic SNPs of ROR2 gene were distinguished by Haploview program. Minor allele frequency (MAF), haplotype blocks and haplotype frequencies were analyzed in eligible SNPs and tag SNPs respectively with genotyping call rate $>80\%$, MAF $>1\%$, H-W equilibrium ($P>0.01$) and no gender difference ($P>0.05$). Tag SNPs were determined under the criteria of $r^2\geq 0.8$ or logarithm of the odd score (LOD) ≥ 3 for pairwise eligible SNPs in CHB and JPT. Common tag SNPs for CHB and JPT were directly reported by Haploview program or being identified from those which were highly related to tag SNPs reported by haploview program under SPSS 13.0 software. **Results** A total of 404 common SNPs were provided for both CHB and JPT samples by HapMap, where 101 common monotonic SNPs between CHB and JPT had the common minor alleles. The common SNPs between CHB and JPT were 257. In the 257 common eligible SNPs, 224 (87.2%) had common minor alleles. Among the 18 and 27 haplotype blocks identified in 257 common eligible SNPs between CHB and JPT, except for 2 independent haplotype blocks identified only in JPT. Other haplotype blocks between CHB and JPT were overlapped partly or completely. A number of 50 common tag SNPs between CHB and JPT were determined and the proportions in CHB and JPT were 64.9% and 70.4% respectively. **Conclusion** Analysis of HapMap data provided an

DOI: 10.3760/cma.j.issn.0254-6450.2011.11.024

基金项目:美国国立卫生研究院国家牙科与颅面研究院基金(R21-DE-013707, R01-DE-014581); Fogarty基金(D43-TW006176)

作者单位:100191 北京大学公共卫生学院流行病学与卫生统计学系

通信作者:王红, Email: hwang2010@yahoo.cn

opportunity to avoid monotonic SNPs that had been included in ROR2 gene related etiologic studies. SNPs in ROR2 gene had common features in alleles, MAF, haplotype blocks and haplotype frequencies between CHB and JPT populations, which were consistent with the geographic and ethnic origins of the two populations.

【Key words】 Receptor tyrosine kinase-like orphan receptor 2; Single nucleotide polymorphism

确定研究标记是进行疾病遗传易感性研究的一个重要步骤,国际人类基因组单体型图计划(HapMap)发布的单核苷酸多态性(SNP)数据为此提供了参考依据^[1]。通过对SNP微效等位基因频率(minor allele frequency, MAF)的分析可以避免将纯合基因型SNP纳入研究,对不同人群SNP进行比较有助于发现人群共有和特有疾病易感等位基因^[2,3],标签SNP是人类基因组中具有代表性和特征性的SNP,由于可以代表与其高度相关的一组SNP,以标签SNP进行病因学研究可达到经济有效的效果。受体酪氨酸激酶样孤立受体2(ROR2)基因在软骨形成和生长发育过程中起重要作用^[4],在人骨肉瘤细胞系SaOS-2和U2OS细胞中的过度表达与肿瘤恶性程度有关,是相关疾病的重要候选基因^[5]。本研究利用HapMap公布的资料,对中国北京汉族人群(Han Chinese in Beijing, China, CHB)和日本东京人群(Japanese in Tokyo, Japan, JPT)ROR2基因SNP的MAF、标签SNP、单体域(Haplotype block)和单体型频率进行研究。

对象与方法

1. 研究对象:利用HapMap计划II期所公布的CHB(45人)和JPT(45人)的ROR2基因的SNP数据进行分析。合格SNP的判定标准为:基因型测定率>80%、H-W平衡($P>0.01$)、MAF>0.01且无性别差异($P>0.05$)。

2. 统计学分析:MAF估计、标签SNP确定、单体域构建及单体型频率估计均由Haploview软件完成,两人群中名称相同的标签SNP确定由SPSS 13.0软件辅助完成。MAF的计算参考文献^[6]。两人群名称相同且合格SNP的MAF差值相对比= $\frac{|MAF_a - MAF_b|}{MAF_a} \times 100\%$,其中MAF_a和MAF_b分别代表CHB和JPT的MAF。标签SNP根据各人群中成对SNP间 $r^2 \geq 0.8$ 或对数优势比(LOD) ≥ 3 的标准确定。 r^2 代表SNP间的关联程度;LOD代表SNP间的连锁程度,LOD的计算公式为: $\text{lod}(\theta) = \text{lod}_{10}[L(\hat{\theta})/L(0.5)]$,其中 θ 为重组分数, $\hat{\theta}$ 为重组分数估计值。Haploview软件多次运行确定的标签SNP可能不同^[7],该软件确定标签SNP的方法为:与其他所有

SNP间 $r^2 < 0.8$ 或LOD < 3 的SNP均作为标签SNP,SNP间 $r^2 \geq 0.8$ 或LOD ≥ 3 的每组SNP中,随机确定一个作为标签SNP。本研究两人群中名称相同的标签SNP一部分通过比较Haploview软件分析提供的标签SNP获得,另一部分通过比较Haploview软件所提供标签SNP的强关联SNP获得,当某个SNP在两人群中存在多个共同强关联SNP时,依据 r^2 最大、LOD最大、物理距离最短的顺序确定。单体域的构建采用Gabriel等^[8]提出的连锁不平衡系数(D')置信区间(CI)进行, D' 值95%CI在0.70~0.98范围的相邻SNP归入同一单体域,单体域图根据两点间LOD和 D' 值的水平用不同的颜色表示,单体型频率的最大似然估计参照文献^[9]。

结 果

1. 基本情况:HapMap公布的ROR2基因SNP在CHB和JPT两人群中名称相同且均为404个,其中纯合基因型SNP分别为103和102个,约占测定SNP总数的25%。纯合基因型SNP中101个在两人群中名称相同,其中97个(96%)在两人群中有相同的微效等位基因。此外,在CHB为纯合基因型的rs11787772和rs12378533,其微效等位基因在JPT的频率均为0.989;在JPT为纯合基因型的rs12685556,其微效等位基因在CHB的频率为0.967。

排除MAF<0.01、基因型测定频率 $\leq 80\%$ 、H-W不平衡以及有性别差异的SNP,CHB和JPT的合格SNP分别为263和293个,其中71.9%(189)和82.9%(243)SNP的MAF>10.0%。合格SNP中有257个在两人群中名称相同即共同合格SNP,分别占各自合格SNP的97.7%(CHB)和87.7%(JPT)。上述257个共同合格SNP中,224个(87.2%)SNP在两人群中的微效等位基因亦相同,其中在两人群中MAF相同的有4个,MAF差值相对比<20%的占53.1%,相对比 $\geq 50\%$ 的占20.5%。

2. 单体域:CHB和JPT中257个共同合格SNP中分别构建了18和27个单体域。其中6个单体域完全相同,2个单体域为JPT独有,其余单体域在两人群相互重叠。

3. 标签SNP:CHB及JPT的合格SNP中,由Haploview软件确定的标签SNP分别为77个和71

个,其中在两人群中名称相同的标签 SNP 即共同标签 SNP 为 38 个,加上根据 SNP 间相关性所确定的 12 个标签 SNP(表 1),两人群中共同标签 SNP 共计 50 个,分别占各自标签 SNP 的 64.9%(CHB)和 70.4%(JPT)。

4. 共同标签 SNP 单体域及单体型频率:50 个共同标签 SNP 在 CHB 和 JPT 分别构建 3 个和 2 个单体

域(图 1)。其中 2 号单体域所包含的 SNP 名称及 4 种单体型在两人群均完全相同,单体型均以前两种为主,累计频率在 CHB 和 JPT 分别为 84.0% 和 77.6%;CHB 的 2 个和 JPT 的 1 个独立单体域中,单体型均为 3 种且以前两种单体型为主,累计频率分别为 92.2%、85.3% 和 92.3%(表 2)。

表 1 CHB 和 JPT 人群 ROR2 基因 50 个共同标签 SNP 的 MAF 和单体域

SNP	CHB	JPT	SNP	CHB	JPT	SNP	CHB	JPT
rs1135150	0.110(A)	0.114(T)	rs10992097	0.100(C)	0.136(C)	rs4744107	0.478(C)	0.500(C)
rs10761129	0.071(C)	0.047(C)	rs4743851	0.233(T)	0.256(T)	rs1881389	0.349(G)	0.442(G)
rs10992065	0.324(T)	0.284(T)	rs3935382	0.267(T)	0.284(T)	rs11795340	0.070(C)	0.119(C)
rs10992066	0.037(T)	0.028(T)	rs9409658	0.367(A)	0.411(A)	rs7870049	0.467(C, 2)	0.489(C, 2)
rs7869073	0.078(G)	0.056(G)	rs10119449	0.200(A)	0.144(A)	rs10992149	0.364(A, 2)	0.298(A, 2)
rs9409458	0.488(G)	0.500(A)	rs16907793	0.167(T)	0.261(T)	rs4744114	0.466(A, 2)	0.352(A, 2)
rs10992075	0.311(G)	0.318(G)	rs10992118	0.476(A, 1)	0.393(A)	rs7873326	0.091(A)	0.116(A)
rs16907725	0.278(A)	0.300(A, 1)	rs10992121	0.375(G, 1)	0.375(G)	rs7035117	0.111(A)	0.067(A)
rs16907728	0.200(G)	0.222(G, 1)	rs17514885	0.044(A)	0.022(A)	rs2001917	0.378(T)	0.318(A)
rs7849691	0.012(T)	0.011(T)	rs7022147	0.200(G)	0.159(G)	rs10820921	0.144(T)	0.136(T)
rs12683181	0.500(C)	0.465(T)	rs7853430	0.183(T)	0.231(T)	rs10992170	0.261(G, 3)	0.190(G)
rs10992085	0.156(A)	0.178(A)	rs4743855	0.398(T)	0.420(T)	rs6479386	0.333(T, 3)	0.322(T)
rs9409461	0.222(A)	0.205(A)	rs7867707	0.344(A)	0.311(A)	rs7034006	0.067(A)	0.100(A)
rs7026175	0.300(C)	0.311(C)	rs17587223	0.111(A)	0.070(A)	rs7037255	0.355(A)	0.342(A)
rs17586213	0.122(C)	0.167(C)	rs7874163	0.167(G)	0.193(G)	rs10992178	0.170(A)	0.189(A)
rs10992090	0.178(A)	0.144(A)	rs16907877	0.067(T)	0.056(T)	rs3750440	0.278(G)	0.330(C)
rs7041425	0.078(A)	0.034(A)	rs4744105	0.159(A)	0.141(A)			

注: 括号外数据为 MAF, 括号内数据为微效等位基因

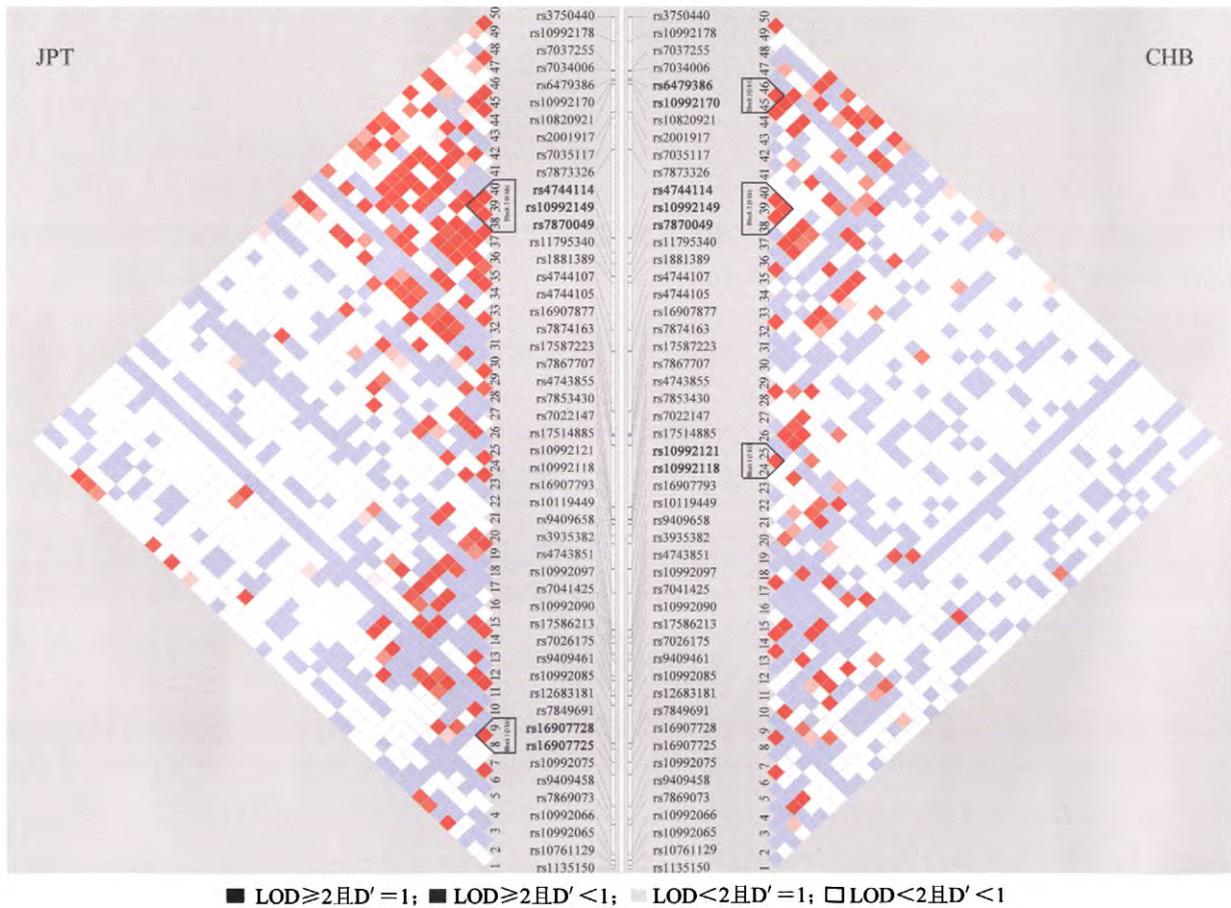


图 1 CHB 和 JPT 人群 50 个共同标签 SNP 的单体域

表2 CHB和JPT人群50个共同标签SNP的单体型频率

Block	CHB		JPT	
	单体型	频率	单体型	频率
1	AA	0.478	GA	0.700
	TG	0.375	AG	0.222
	TA	0.147	AA	0.078
2	CGG	0.467	CGG	0.489
	TAA	0.373	TAA	0.287
	TGA	0.101	TGG	0.160
	TGG	0.060	TGA	0.064
3	TC	0.667	-	-
	GT	0.256	-	-
	TT	0.078	-	-

注: - 无此单体型

讨 论

MAF是筛选SNP的一个重要指标。本研究结果显示,HapMap计划公布的CHB和JPT的ROR2基因的404个SNP中,101个纯合基因型SNP在两人群中名称相同,提示在上述两人群中研究时,应避免选择这些SNP作为研究标记。

MAF水平与研究效能有关,在疾病遗传易感性研究中,一般选择MAF>0.1的SNP作为遗传标记^[10]。本研究中CHB和JPT的ROR2基因SNP中MAF>10%的分别占71.9%和82.9%,提供足够的可供选择遗传标记。

当研究涉及≥2个人群时,对人群间的共性位点进行合并分析有助于通过扩大样本量提高研究效能。本研究CHB和JPT的ROR2基因SNP在多方面表现出相似性:①HapMap测定的全部SNP在两人群中名称完全相同;②纯合基因型SNP名称及微效等位基因在两人群中基本相同;③两人群共同合格SNP中,微效等位基因在两人群中相同者接近90.0%,其中MAF差值相对比<20.0%的超过半数;④共同合格SNP中,CHB形成的全部单体域与JPT形成的92.6%的单体域全部或部分重叠;⑤50个共同标签SNP所形成的单体域中,有1个单体域的SNP组成及单体型在两人群中均完全相同,累计频率相近。

本研究结果表明,HapMap所公布CHB和JPT的ROR2基因SNP遗传标记在两人群中以共性为主,伴随较低程度的人群间差异。如少数SNP仅在一个人群为纯合基因型SNP,由于样本和基因型测定的原因使某些SNP仅在一个人群符合研究条件等,这些SNP形成人群各自的特有位点。研究涉及多个人群时,发现人群间特有和共性位点同等重要,

对人群特有位点进行单独分析,有助于人群特异性易感位点的发现^[11,12]。

CHB和JPT同属亚洲人群,本文利用Haploview软件对HapMap计划公布的ROR2基因SNP数据进行综合分析和比较研究,为两人群ROR2基因与相关疾病的病因学研究提供基础,也为其他基因的初步分析提供方法。

参 考 文 献

- [1] Chen HF, Lin LN, Li W, et al. Genetic variants of haplotypes in the 5' -flanking region of protein phosphatase 2A-A α gene in Cantonese, China. *Chin Prev Med*, 2011 (3): 219-222. (in Chinese)
陈慧峰,林丽娜,李文,等. 广东人群PP2A-A α 亚基基因5'-侧翼区单体型分析. *中国预防医学杂志*, 2011(3):219-222.
- [2] Guthery SL, Salisbury BA, Pungliya MS, et al. The structure of common genetic variation in United States populations. *Am J Hum Genet*, 2007, 81(6): 1221-1231.
- [3] Zhao ML, Qi SW, Liu L, et al. Discrimination of SNP ancestry informative markers for Japanese, European, African and Chinese Han populations. *J Zhengzhou Univ (Med Edition)*, 2007 (3): 540-542. (in Chinese)
赵美乐,齐守文,刘良,等. 中国汉族、日本、欧裔和非裔人群SNP始祖多态位点的鉴定. *郑州大学学报(医学版)*, 2007(3): 540-542.
- [4] Dechiara TM, Kimble RB, Poueymirou WT, et al. ROR2, encoding a receptor-like tyrosine kinase, is required for cartilage and growth plate development. *Nat Genet*, 2000, 24(3): 271-274.
- [5] Minami Y, Oishi I, Endo M, et al. Ror-family receptor tyrosine kinases in noncanonical Wnt signaling: their implications in developmental morphogenesis and human diseases. *Dev Dyn*, 2010, 239(1): 1-15.
- [6] Jia XF, Wang H. Application of FBAT and PLINK programs on allele frequency calculation and transmission disequilibrium test. *Chin J Birth Health Hered*, 2009(10): 20-22. (in Chinese)
贾小芳,王红. FBAT和PLINK软件计算等位基因频率和传递不平衡分析的应用研究. *中国优生与遗传杂志*, 2009(10): 20-22.
- [7] Carlson CS, Eberle MA, Rieder MJ, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*, 2004, 74(1): 106-120.
- [8] Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science*, 2002, 296(5576): 2225-2229.
- [9] Qin ZS, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet*, 2002, 71(5): 1242-1247.
- [10] Xiong SY, Hao YT, Rao SQ, 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. *Chin Sci Bull*, 2009(12): 2069-2075.
- [11] Risch N, Burchard E, Ziv E, et al. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol*, 2002, 3(7): t2007.
- [12] Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*, 2003, 348(12): 1170-1175.

(收稿日期:2011-05-20)

(本文编辑:万玉立)