

中国人群 MEGSIN 基因单核苷酸多态性的检测

黄 霜¹, 李幼姬², 黄玮俊¹, 陈素琴¹, 古宏标^{1,3}, 胡 彬¹, 王一鸣¹

(1. 中山大学基础医学院遗传学教研室, 广东 广州 510080; 2. 中山大学附属第一医院肾内科, 广东 广州 510080;
3. 广东药学院, 广东 广州 510240)

摘 要:【目的】研究中国人群 MEGSIN 基因多态性及其分布特征,并与国外数据库进行比较。【方法】随机选取 208 例广东地区汉族个体,提取基因组 DNA,对包含外显子、外显子-内含子交界区及 5' UTR 区、3' UTR 区的 PCR 产物进行直接测序,综合正反向结果识别和鉴定基因内 SNPs。所得结果在美国国立生物技术信息中心(NCBI)的 SNP 数据库(dbSNP)中进行查询和比较。【结果】在所有研究对象中共发现 24 个 SNPs,主要位于非编码区,22 个为替换型 SNP,1 个为插入型 SNP,另一个为缺失型 SNP;其中有 6 个 SNPs 在数据库中未报道,有 4 个数据库已报道的 SNPs,在本次研究未能证实在中国人群中存在多态性。【结论】中国汉族人群 MEGSIN 基因的多态性分布与美国数据库中基于高加索人群的资料存在差异。本研究不仅有利于了解 MEGSIN 基因结构,且为在中国人群中研究 MEGSIN 基因相关疾病并最终找到致病位点提供可靠数据。

关键词: MEGSIN; 多态性; 单核苷酸; 中国人

中图分类号: R394.5

文献标识码: A

文章编号: 1672-3554(2006)02-0140-04

Investigation of Single Nucleotide Polymorphisms in MEGSIN Gene among Chinese Population

HUANG Shuang¹, LI You-ji², HUANG Wei-jun¹, CHEN Su-qin¹, GU Hong-biao¹, HU Bin¹, WANG Yi-ming¹

(1. Department of Medical Genetics, 2. Department of Nephrology, The first Affiliated Hospital, SUN Yat-sen University, Guangzhou 510080, China; 3. Guangdong College of Pharmacy, Guangzhou 510240, China)

Abstract:【Objective】To investigate the single nucleotide polymorphisms of the MEGSIN gene in Chinese population and to compare them with the data in the aboard database. 【Methods】Genomic DNA was extracted randomly from 208 Chinese Han individuals in Guangdong Province. PCR products spanning the exons, intron- exon boundaries, 5' and 3' UTR of the gene were sequenced in both directions. The results were compared against the NCBI (National Center for Biotechnology Information U.S.) databases. 【Results】Twenty-four SNPs were identified in this study. They were located mostly in the non-coding intronic regions. Among them 22 SNPs were transition, 1 SNP was insertion and another was deletion. There were 6 novel SNPs in our results, and 4 already in the database were not identified in our subjects. 【Conclusion】MEGSIN gene polymorphisms vary between Chinese and the Caucasian populations. Results obtained in this study not only help the understanding of the gene's genomic structures, but also provide useful information for association studies on MEGSIN gene-related diseases with the aim of the eventual identification of the causal variants.

Key word: MEGSIN; polymorphism, single nucleotide; Chinese

[J SUN Yat-sen Univ(Med Sci), 2006(2):140-143]

IgA 肾病是我国最常见的肾小球疾病,临床转归不一,在发病 10 年后 10%~20% 病人发展为终末期肾衰竭(ESRF),对病人的健康有极大的危害。到目前为止,IgA 肾病的病因和发病机理不清。近

年来,国内外大量的研究进展表明本病与遗传因素、环境因素相关,属多基因疾病^[1]。MEGSIN 基因是在肾小球系膜细胞优势表达的基因,基因碱基对数目约为 30 kb,含 8 个外显子,转录产物 2 249

收稿日期 2005-09-19

基金项目 美国中华医学基金资助项目(98-677);国家自然科学基金资助项目(10206);广东省自然科学基金资助项目(013140)

作者简介 黄 霜(1977-),女,海南儋县人,硕士生;王一鸣,教授,博士生导师,通讯作者。E-mail: ywzhong@hotmail.com

bp, 开放性读码框 1 143 bp, 编码含 380 个氨基酸的蛋白质, 表达产物属丝氨酸蛋白酶抑制剂超家族的成员, 参与介导系膜细胞的增生和细胞外基质的沉积, 免疫组化和原位杂交技术已经证明了此基因在 IgA 肾病人中表达升高, 同时伴有系膜细胞增生及细胞外基质的扩增^[2-5]。以往研究表明, MEGSN 基因 3' UTR 区两个 SNPs 2093C 和 2180T 与 IgA 肾病相关, 是 IgA 肾病的易感基因^[6]。本研究为首次对中国汉族人群 MEGSN 基因进行单核苷酸多态性分析, 并确定其分布特征, 为进行 MEGSN 基因相关疾病的研究并最终找到致病位点奠定基础。

1 材料和方法

1.1 研究对象

随机选取 208 例彼此无血缘关系的广东地区汉族人, 在获得参与者的知情同意后, 采集外周血, 所有参与者既往无肾脏病及其它系统性疾病史。

1.2 DNA 提取

外周血 10 mL, 用 Qiagen Maxi QiaAmp DNA kit (Qiagen, 德国) 抽提 DNA, 严格按试剂盒说明进行操作。

表 1 MEGSN 基因测序区域及相应引物

Table 1 Sequencing region and primers of MEGSN gene

Region	Primer sequence	Size of production (bp)	Temperature of annealing (°C)
Exon1	Forward 5' TGCCCTTAGTTGATTTGATGTCTA 3'	936	60
	Reverse 5' CAGTGAGGTAAGAATAGGATGC 3'		
Exon 2	Forward 5' CCAACCAACGAGAAGAGATAGTTT 3'	524	59
	Reverse 5' TGGAGATGTGTGAAGATGTGAC 3'		
Exon 3	Forward 5' GGACTGACTACCAACAAGCAT 3'	627	63
	Reverse 5' CCACATTTCTCACTTCCATCTC 3'		
Exon 4	Forward 5' TGGGCAAGAAAGGATGAAGTTA 3'	570	57
	Reverse 5' AAGTAGCAATGAGAAACAAAACCA 3'		
Exon 5	Forward 5' CTTCCAGTCCATTTCCATA 3'	454	57
	Reverse 5' ATTATTCAAAACCAATTGC 3'		
Intron 5	Forward 5' ACTCTTCTGTGTGCTGGACTGG 3'	578	60
	Reverse 5' CTTATGCCAOCCTTCAACCAAT 3'		
Exon 6	Forward 5' GGGCCTAGAAAGTCTGGACACA 3'	521	58
	Reverse 5' CAAGACACGTTTGGTGGTGTTC A 3'		
Exon 7	Forward 5' TGGCTTTACCACTAAGATTTGC 3'	578	57
	Reverse 5' TGACATTTGAGAGATTTCCAA 3'		
Exon 8-1	Forward 5' ATGGGTGAGCAGAGAGTGAAT 3'	923	63
	Reverse 5' AGAAATGAGCGTGGGAAGAAAG 3'		
Exon 8-2	Forward 5' TGACCAACCAATTCATTTGTTA 3'	1195	62
	Reverse 5' CAGTCTCTATTCAGGTAAGTGTTT 3'		

1.3 SNPs 的检测

对 208 例个体 MEGSN 基因的 5' UTR 区、外显子、外显子-内含子交界区约 200 bp、部分内含子区以及 3' UTR 区进行直接测序。根据 NCBI 中发表的 MEGSN 基因序列 (GenBank ID AF027866) 以在线软件 Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) 进行引物设计 (表 1), 由上海英骏公司合成。PCR 反应体系为 30 μ L, 含 1 \times PCR Buffer、4 种 dNTP 各 20 mmol/L、MgCl₂ 1.6 mmol/L, 上游和下游引物各 0.17 μ mol/L、Taq DNA 聚合酶 (MBI) 1 U, 反应在 Geneamp 9700PCR 仪 (Applied Biosystems) 上完成。扩增程序为: 94 预变性 3 min; 94 变性 45 s, 引物特异性退火温度 (见表 1) 45 s, 72 延伸 45 s, 共 35 个循环; 最后 72 延伸 10 min。PCR 产物经大肠杆菌核酸外切酶和虾碱性磷酸酶 (usb 公司) 纯化后以 BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) 按标准程序进行双向测序, 产物在 ABI 3100 测序仪 (Applied Biosystems) 上分析, 综合正反向测序结果识别 SNPs 及对所有 SNP 位点的部分样本进行再次扩增测序最终确认 SNPs 的存在。

2 结 果

对所有样本进行 MEGSN 基因的直接测序 (位点 797 G > A 测序结果, 见图 1), 测序区段共 6.3 kb, 包括 1 140 bp 编码区以及 5 200 bp 非编码区, 共发现 24 个 SNPs (表 2)。其中 22 个为替换型 SNPs, 1 个为插入型 SNPs, 一个为缺失型 SNP; 在 22 个替换型 SNPs 中, 15 个为转换型 (68.2%), 7 个为颠换型 (31.8%), 转换发生的频率是颠换的 2.1 倍。所有 24 个 SNPs 中, 1 个在 5' UTR 区, 2 个在编码区, 16 个在内含子区, 5 个在 3' UTR 区。位于编码区的 2 个 SNPs 中, 一个为错义突变, 另一个为同义突变。

多态位点等位基因的频率在 0.010-0.498 之间。分布密度为: 编码区 1 SNP/570 bp, 而非编码区则为 1 SNP/236 bp。所得结果与 NCBI (截至 2005 年 8 月) 的 SNP 数据库中已登记的 SNPs 进行比较, 其中有 6 个 SNPs 数据库中未报道, 占 25%; 另 4 个 SNPs (rs12964116、rs17071824、rs11873045、rs1055900) 为数据库中已有登记, 但本次研究在中国人群中未发现多态性。

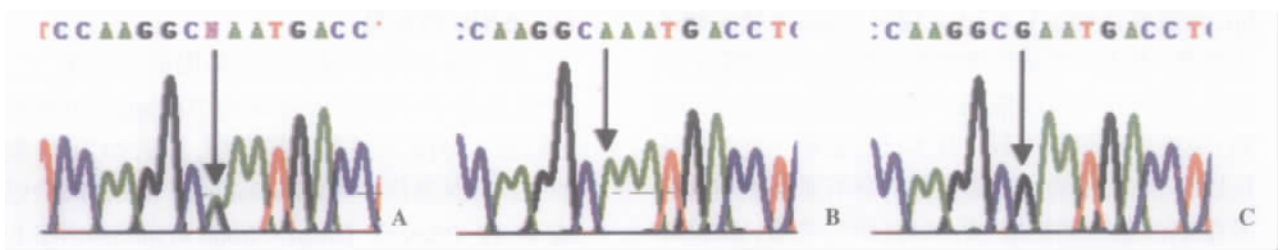


图 1 SNP 797 G > A 的测序图(箭头所指)

Fig. 1 The sequence of 797 G > A (indicated by arrow)

A: A/G B: A/A C: G/G

表 2 中国人群 MEGSIN 基因单核苷酸多态性位点的位置、类型及等位基因频率

Table 2 Characterization of 26 SNPs in the MEGSIN gene

No.	Position ¹⁾	Location in gene ²⁾	Allele freq	Identity to dbSNP
1	-6697 G>A	E1-5'UTR	0.855:0.145	rs17781440
2	-6468 T>G	I1	0.858:0.142	rs1548320
3	169-261 A>G	I2	0.748:0.252	rs1720844
4	169-211 A>G	I2	0.517:0.483	rs1720845
5	167-117 T>C	I2	0.976:0.024	rs8093326
6	220-109 A>T	I3	0.522:0.478	rs1701631
7	337-49 C>T	I4	0.981:0.019	rs8091304
8	455-300 C>A	I5	0.966:0.034	rs8083983
9	455-85del A	I5	0.966:0.034	novel
10	455-28 G>A	I5	0.750:0.250	rs1720858
11	455-7_455-6ins T	I5	0.536:0.464	novel
12	597+44 C>T	I6	0.529:0.471	rs1720857
13	597+78 T>C	I6	0.529:0.471	rs1720856
14	598-15 T>A	I6	0.519:0.481	rs2658459
15	744+60 C>G	I7	0.750:0.250	rs2689399
16	745-117 A>G	I7	0.976:0.024	novel
17	745-19 C>G	I7	0.988:0.012	novel
18	797 G>A	E8-CDS	0.890:0.110	rs17782413
19	936 G>T	E8-CDS	0.976:0.024	novel
20	* +18 A>G	E8-3'UTR	0.502:0.498	rs165353
21	* +485 T>C	E8-3'UTR	0.976:0.024	novel
22	* +586 C>T	E8-3'UTR	0.750:0.250	rs1055901
23	* +660 T>C	E8-3'UTR	0.526:0.474	rs2658460
24	* +674 T>C	E8-3'UTR	0.750:0.250	rs1055902

dbSNP: database of single nucleotide polymorphisms; 1) Nucleotide numbering is according to nomenclature for the description of sequence variations by human genome variation society (<http://www.genomic.unimelb.edu.au/mdi/mutnomen/>); the A of the ATG of the initiator Methionine codon is denoted nucleotide +1; the nucleotide 3 of the translation stop codon is * 1, the next * 2, etc; ins, insertion polymorphism; del, deletion polymorphism; 2) E: Exon; I: Intron; UTR: Untranslated region; CDS: Coding region. Amino acid substitution: Arg266Gln for No.18; Ser312 Ser for No.19

3 讨论

人类基因组计划的研究成果表明,多态性是人类基因组组成的显著特征之一。基因多态性在不同的人群中有不同的表现,部分多态性可能与疾病的易感性相关。SNPs是人类可遗传的变异中最常见的一种,占所有已知多态性的90%以上,可满足多基因病高精度基因定位的要求,可直接用于指导易感基因的克隆^[7,8]。人们已寄希望于用它定位肿瘤、精神性疾病、心血管疾病、糖尿病等一系列多基因疾病的主基因^[9]。根据SNPs分布部位的不同可分为编码区SNP(cSNP)和非编码区SNP(非cSNP);cSNP可能导致氨基酸的替换,从而影响蛋白质的功能和某些生物学性状^[10];位于基因调控区(启动子区、5'UTR区、3'UTR区)的SNP(rSNP)则可能影响基因表达的量,从而引起人体某些性状的改变和疾病的发生;另外,基因中外显子-内含子交界区的SNPs均可能影响蛋白质的表达、功能和稳定性,同样十分重要^[11,12]。基于以上原因,本研究主要对Megsin基因的编码区、调控区和外显子-内含子交界区进行测序,检测这些区段内的多态位点的分布。

MEGSIN基因的高表达已证明和IgA肾病相关^[2-5],并已确定在中国人群MEGSIN基因的3'UTR区2093C和2180T与IgA肾病相关^[6]。本研究对广东地区汉族人群MEGSIN基因进行单核苷酸多态性分析,并确定其分布特征,为进一步研究IgA肾病及MEGSIN基因相关疾病奠定基础。

在MEGSIN基因6.3 kb的测序区域内,包括1140 bp编码区及5200 bp非编码区,共发现24个SNPs均为二态性。其中22个为替换型SNPs,1个为插入型SNPs,一个为缺失型SNP;在22个替换型SNPs中,15个为转换型(68.2%),7个为颠换

型 (31.8%), 转换发生的频率是颠换的 2.1 倍, 确切原因不明。所有 24 个多态位点中, 1 个在 5' UTR 区, 2 个在编码区, 16 个在内含子区, 5 个在 3' UTR 区。位于编码区的 2 个 SNPs, 1 个为错义突变, 另一个为同义突变。24 个多态位点等位基因的频率在 0.010-0.498 之间。分布密度为: 编码区 1 SNP/570 bp, 非编码区为 1 SNP/236 bp; SNPs 在非编码区的密度大, 这可能是因为在非编码序列的 SNPs 大多不影响蛋白质编码, 承受的选择压力较小, 所以得以保存; 处于编码序列的 SNPs 承受的选择压力大, 易于被选择。与 NCBI (截至 2005 年 8 月) 中已登记的 SNPs 进行比较, 有 6 个 SNPs 数据库中未报道, 占 25%, 包括了位于编码区的一个同义突变、位于 3' UTR 区的一个多态位点、内含子区的 4 个多态位点; 4 个数据库中已登记的 SNPs (rs12964116、rs17071824、rs11873045、rs1055900)。本次研究在中国人群中未能证实存在多态性, 可能是因为数据库中的数据主要来源于高加索人群, 与中国人群在遗传背景上有不同; 其余的多态位点与数据库报导的一致。

SNPs 是人类在漫长进化过程中基因组与内外环境交互作用的累积结果。虽然 SNPs 的数目极其繁多、密度很高, 但值得指出的是, 不同的 SNPs 位点由于其在进化过程中产生的顺序有先有后、在不同人群中分布频率不完全一样、杂合程度亦不尽相同, 在利用它们进行相关分析或者连锁不平衡分析时所能提供的信息量也可能存在较大差别^[7]。目前国内外对于 MEGSN 基因多态性的研究很少, 我们通过检测, 发现中国人群的有关数据与国外权威数据库中基于高加索人群的资料存在较大的差异。这提示我们, 某些 SNPs 存在种族差异, 所以在研究中应该通过实验来检测中国人群的数据。本研究为首次对中国人群 MEGSN 基因进行 SNPs 分析。本次研究样本数为 208 例, 属较大样本, 有利于发现罕见的多态位点和获得可靠可信的等位基因的分布情况, 为建立中国人群的多态信息库提供资料。本研究不仅有利于了解 MEGSN 基因结构, 且为在中国人群中研究 MEGSN 基因相关疾病并最终找到致病位点提供可靠基础。我们的下一步研究, 将继续检测该基因的其他 SNPs 位点, 特别是基因调控区, 以得到更全面的结果。

参考文献:

- [1] HSU S I, RAMIREZ S B, WINN M P, et al. Evidence for genetic factors in the development and progression of IgA nephropathy [J]. *Kidney Int*, 2000, 57(5): 1818-1835.
- [2] INAGI R, MIYATA T, SUZUKI D, et al. Specific tissue distribution of megsin, a novel serpin, in the glomerulus and its up-regulation in IgA nephropathy [J]. *Biochem Biophys Res Commun*, 2001, 286(5): 1098-1106.
- [3] MIYATA T, INAGI R, NANGAKU M, et al. Overexpression of the serpin megsin induces progressive mesangial cell proliferation and expansion [J]. *J Clin Invest*, 2002, 109(5): 585-593.
- [4] INAGI R, NANGAKU M, MIYATA T, et al. Mesangial cell-predominant functional gene, megsin [J]. *Clin Exp Nephrol*, 2003, 7(2): 87-92.
- [5] NANGAKU M, MIYATA T, SUZUKI D, et al. Cloning of rodent megsin revealed its up-regulation in mesangioproliferative nephritis [J]. *Kidney Int*. 2001, 60(2): 641-652.
- [6] LI Y J, DU Y, LI C X, et al. Family-based association study showing that immunoglobulin A nephropathy is associated with the polymorphisms 2093C and 2180T in the 3' untranslated region of the Megsin gene [J]. *J Am Soc Nephrol*, 2004, 15(7): 1739-43.
- [7] BROOKES A J. The essence of SNPs [J]. *Gene*, 1999, 234(2): 177-1786.
- [8] OHNISHI Y, TANAKA T, YAMADA R, et al. Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population [J]. *Hum Genet*, 2000, 106(3): 288-292.
- [9] LEONID K. Prospects for whole genome linkage disequilibrium mapping of common disease genes [J]. *Nat Genet*, 1999, 22(2): 139-144.
- [10] CARTEGNI L, CHEW S L, KRAINER A R. Listening to silence and understanding nonsense: exonic mutations that affect splicing [J]. *Nat Rev Genet*, 2002, 3(4): 285-298.
- [11] TABOR H K, RISCH N J, MYERS R M, et al. Candidate - gene approaches for studying complex genetic traits: practical considerations [J]. *Nat Rev Genet*, 2002, 3(5): 391-397.
- [12] CARDON L R, BELL J I. Association study designs for complex disease [J]. *Nat Rev Genet*, 2001, 2(2): 91-99.

(编辑 张敏瑞)