

HPCL2 基因表达的选择性剪切分析*

牛宇欣^{1,2)} 胡松年³⁾ 杨焕明^{1,2)}**

(¹⁾中国科学院遗传与发育生物学研究所人类基因组中心, 北京 100101;

²⁾中国科学院北京基因组研究所, 北京 101300; ³⁾杭州华大基因研究中心, 杭州 380001)

摘要 2-羟基植烷酸辅酶 A 裂解酶 (2-hydroxypanthyl-CoA lyase, HPCL2) 是 3-甲基脂肪酸 α -氧化途径中的关键酶。采用鸟枪法测序技术, 得到了 HPCL2 基因的基因组序列。结果显示, HPCL2 基因组大小为 40 829 bp, 共有 17 个外显子, 16 个内含子。外显子平均大小为 116 bp, 内含子平均大小为 2 429 bp, 属于结构紧凑基因。从 dbEST 数据库中筛选与 HPCL2 基因的 mRNA (GenBank Acc.: AJ131753) 相互重叠表达序列标签 (EST) 共有 213 个, 分别来自 29 个不同的组织。将 EST 与基因组序列进行 Blast 分析, 共检测 17 个 EST 具有选择性剪切, 其中 14 例属外显子遗漏 (exon skipping), 2 例外显子增加 (exon inclusion) 和 1 例剪切位点改变 (splicing site shift)。外显子遗漏主要发生在 5'端的第 3 到第 8 个外显子。结果表明, 外显子遗漏可能是 HPCL2 基因选择性剪切的主要形式。

关键词 HPCL2, 基因结构, 选择性剪切, 外显子遗漏

学科分类号 Q78

由于 β 位上存在一个甲基基团, 一些 3-甲基脂肪酸, 如植烷酸 (phytanic acid, 3, 7, 11, 15-tetramethylhexadecanoic acid), 不能通过正常的 β -氧化途径降解^[1]。因此 3-甲基脂肪酸需经过一个 α -氧化反应, 转化为 2-甲基脂肪酸后, 才能进一步通过 β -氧化途径被降解。 α -氧化途径细胞定位于过氧化物酶体^[2]。已经发现几种先天性代谢遗传病是由于过氧化物酶体内 α -氧化途径中酶的缺失而引起, 如植烷酸 α -羟化酶缺陷引起的 RD 型雷姆氏病等^[3]。在 α -氧化途径的第三步反应中, 一个 2-羟基三甲基乳酰辅酶 A (2-hydroxy-3-methylacyl-CoA) 裂解, 生成一个甲酰辅酶 A 和一个 2-甲基脂肪醛^[2,4]。Foulon 等^[5]首先在 1999 年从大鼠肝脏的过氧化物酶体里分离得到了这种裂解酶, 克隆编码了该酶的 cDNA, 并命名该基因为 2-羟基植烷酸辅酶 A 裂解酶基因 (2-hydroxyphytanoyl-CoA lyase, HPCL2)。但进一步关于该基因的基因结构和表达调控研究一直未见报道。

1 实验方法

1.1 BAC 基因组测序和序列组装

方法见参考文献 [6]。

1.2 序列分析

基因结构的分析用 Blast 程序在本地计算机上完成, 重复序列的标记和统计用 RepeatMasker 程序, 用 TNNW 程序进行启动子序列分析和 TATA

box 定位。

1.3 选择性剪切分析

采用电子选择性剪切分析方法 (*in silico* alternative splicing) 研究 HPCL2 的表达调控。从 UniGene 数据库找出所有与 HPCL2 基因 mRNA (GenBank Acc.: AJ131753) 相互重叠的表达序列标签 (EST), 并将所有的 EST 序列与基因组序列进行 Blast 分析, 以检查各 EST 的表达特征。

2 结果与分析

2.1 HPCL2 的结构特征

HPCL2 基因大小为 40 829 bp, 平均 GC 含量为 38.4%, 比人类基因组 GC 含量的平均值 41.0% 低。HPCL2 基因共有 17 个外显子, 最大外显子长度为 187 bp, 最小的外显子只有 41 bp, 平均大小为 116 bp。内含子大小变化幅度较大, 最大的第 10 内含子为 4 827 bp, 而最小的为第 1 内含子, 仅 188 bp, 所有内含子平均长度为 2 429 bp。HPCL2 基因的外显子、内含子分布特点表明, 该基因属于结构紧凑基因, 所有的外显子-内含子接头顺序符合 “GT-AG” 原则。

* 国家高技术 “863” 计划资助项目 (863-J19) 和中国科学院创新工程资助项目 (KSCX1-D4)。

** 通讯联系人。

Tel: 010-80494969, E-mail: hmyang@genetics.ac.cn

收稿日期: 2003-04-10, 接受日期: 2003-05-28

HPCL2 翻译起始位点位于第 1 个外显子 98 bp, 在此区域存在一个典型的 Kozak 结构. 终止编码为 TAA, 3'UTR 大小为 133 bp. 用 TNNW 程序分析翻译起始位点的 5' 端序列结构, 在第 1 个外显子上游 479 bp 处发现 TATA box. 对该基因编码的多肽序列分析表明, 焦磷酸硫胺素的结合位点位于第 14 个外显子.

2.2 HPCL2 基因的选择性剪切分析

2.2.1 EST 组织来源说明: Unigene 数据库中, 共搜索到 213 个与 AJ131753 相互重叠的 EST, 来源于 29 种不同的器官/组织, 其中有 16 个 EST 来源于两种或两种以上的器官/组织的混合样品. 其中 EST 数目最多的器官是脑, 共有 25 个 EST, 占总数的 11.7%. 在所有的 EST 中, 有 48 个 EST 来源于肿瘤/癌组织, 占总数的 22.5%.

2.2.2 外显子遗漏: 将所有 EST 序列与基因组序列进行 Blast 分析, 以检测该基因在不同组织中的选择性剪切. 从 213 个相互重叠的 EST 中, 共发现有 14 个 EST 发生外显子遗漏, 其中来自于脑组织的 EST 有 5 个, 分别来自 4 个不同的 EST 文库.

来自肠组织的 EST 有 6 个, 6 个 EST 来自同一癌变细胞系 EST 文库. 另外发生外显子遗漏的 3 个 EST 分别来自淋巴组织、睾丸组织和癌变的皮肤组织.

在脑组织来源的 5 个 EST 中, 有 5 种外显子遗漏类型, 遗漏外显子数目变化较大, 其中有 2 个 EST 遗漏 3 个外显子, 1 个 EST 遗漏 2 个外显子, 2 个 EST 遗漏 1 个外显子. 遗漏的外显子包括从第 4 个外显子到第 8 个外显子. 而来自癌变细胞系的 6 个 EST 遗漏的外显子相同, 均为第 5 个外显子遗漏. 来自淋巴组织的 AA306411 和来自癌变皮肤组织的 BG761200 均为第 6 个外显子遗漏, 来自睾丸组织的 BI562289 为第 3 个外显子遗漏.

2.2.3 其他剪切类型: 除外显子遗漏外, 还发现另外两种类型的剪切形式. 来自树突状细胞组织的 EST AV713962, 由于第 6 个内含子的剪切位点发生后移, 导致第 7 个外显子部分序列 (49 bp) 被剪切, 属于剪切位点改变. AA046732 和 AA046602 是来自子宫组织的 2 个 EST, Blast 分析发现, 这 2 个 EST 都包含有 175 bp 长的 16 号内含子序列 (图 1), 属于外显子增加.

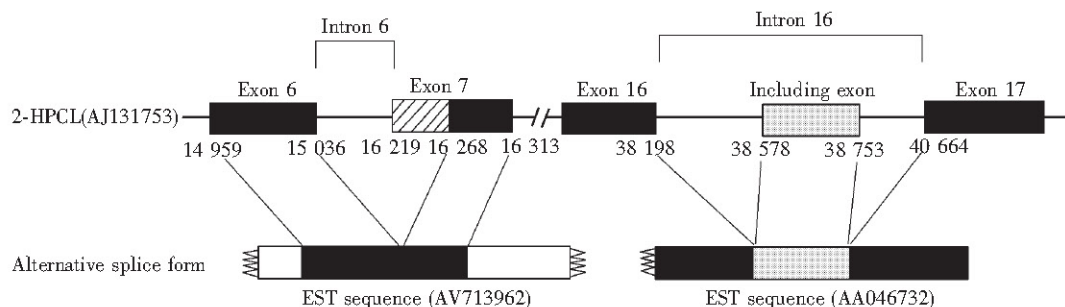


Fig. 1 The sketch map of two splice variances of HPCL2 gene

3 讨 论

可变剪接是真核细胞一种重要的 mRNA 转录后加工机制, 由此而产生相应的多种 mRNA 异构体 (isoform), 使一个基因可以翻译为多种多肽序列^[7], 是基因表达多样性的重要表现形式. 通过对人类基因组的分析表明, 有 60% 的基因在表达过程中可通过选择性剪切, 产生多种形式的 mRNA^[8,9], 有的基因能够产生几万种不同形式的 mRNA^[10]. 选择性剪切的不同变化同细胞生理、发育调节以及肿瘤的发生、转移有关. 因此, 研究基因的选择性剪切对了解基因功能有重要意义.

人类的 EST 数据库包括了 600 多种不同组织

的 cDNA 文库, 其组织供体的年龄从胚胎发育的各个时期一直到最大年龄 75 岁, 基本能够涵盖人体不同发育时期和生命过程. 因此 EST 数据库含有丰富的选择性剪切类型, 通过对 EST 数据库的分析, 系统地了解基因的选择性剪切, 能够较为全面了解基因的表达调节^[11]. 随着高通量测序技术的发展, EST 数据库收录的 EST 数目将会继续增加, 而且覆盖的组织类型更为丰富, 利用 dbEST 进行基因的数字化表达分析将成为基因功能分析的重要手段之一^[12].

本研究利用人类 dbEST 数据库, 调查了 HPCL2 基因选择性剪切的基本类型. 从 213 条 HPCL2 基因的 EST 中, 共检测到不同类型选择性

剪切形式的 EST 有 17 个。其中, 外显子遗漏有 14 例, 占全部剪切类型的 82%。表明外显子遗漏可能是 HPCL2 基因选择性剪切的主要形式。真核细胞的 mRNA 成熟加工过程中, 许多基因中的内含子并不是一次全部切去, 而是在不同的细胞或不同的发育阶段选择性剪切其中部分内含子, 生成不同的 mRNA 及蛋白质分子^[13]。HPCL2 基因外显子遗漏主要发生在靠近 5' 端的几个外显子, 可发生单个外显子遗漏, 也可同时遗漏 2 或 3 个外显子。

除外显子遗漏的变异类型外, 同时还发现了另外两种类型的变异体, 一种是由于 3' 端剪切位点的改变, 从而使第 7 个外显子较正常 mRNA 相比缺失了 45 bp, 另一种类型的变异体是第 16 个内含子 175 bp 的部分序列未被切除, 依然保持在 mRNA 中。通过对 HPCL2 基因的选择性剪切研究表明, 在不同的发育时期, 在不同的组织类型中, HPCL2 基因有着复杂的转录表达调控机制, 所有这些变异类型的生物学功能与可能的疾病发生相关性尚待进一步的研究。

致谢 本文承中国科学院生物物理研究所陈润生研究员审阅, 并给予极大的帮助, 特此感谢。

参 考 文 献

- 1 Verhoeven N M, Jakobs C. Human metabolism of phytanic acid and pristanic acid. *Prog Lipid Res*, 2001, **40** (6): 453 ~ 466
- 2 Mihalik S J, Rainville A M, Watkins P A. Phytanic acid alpha-oxidation in rat liver perox-isomes. Production of alpha-hydroxyphytanoyl-CoA and formate is enhanced by dioxxygenase cofactors. *Eur J Biochem*, 1995, **232** (2): 545 ~ 551

- 3 Jansen G A, Wanders R J, Watkins P A, *et al.* Phytanoyl-coenzyme A hydroxylase deficiency-the enzyme defect in Refsum's disease. *N Engl J Med*, 1997, **337** (2): 133 ~ 134
- 4 Croes K, Casteels M, de Hoffmann E, *et al.* alpha-oxidation of 3-methyl-substituted fatty acids in rat liver. Production of formic acid instead of CO₂, cofactor requirements, subcellular localization and formation of a 2-hydroxy-3-methylacyl-CoA intermediate. *Eur J Biochem*, 1996, **240** (3): 674 ~ 683
- 5 Foulon V, Antonenkov V D, Croes K, *et al.* Purification, molecular cloning, and expression of 2-hydroxyphytanoyl-CoA lyase, a peroxisomal thiamine pyrophosphate-dependent enzyme that catalyzes the carbon-carbon bond cleavage during α -oxidation of 3-methyl-branched fatty acids. *Proc Natl Acad Sci USA*, 1999, **96**: 10039 ~ 10044
- 6 吴清发, 董 伟, 齐晓华, 等. 代谢型谷氨酸受体 mGluR7 的基因组结构及其家族蛋白胞外区的基因组结构比较. *科学通报*, 2002, **47** (7): 508 ~ 515
Wu Q F, Dong W, Qi X H, *et al.* *Chin Sci Bull*, 2002, **47** (16): 1330 ~ 1336
- 7 Sunyaev S, Hanke J, Brett D, *et al.* Individual variation in protein-coding sequences of human genome. *Adv Protein Chem*, 2000, **54**: 409 ~ 37
- 8 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 2001, **409**: 860 ~ 921
- 9 Venter J C, Adams M D, Myers E W, *et al.* The sequence of the human genome. *Science*, 2001, **291**: 1304 ~ 1351
- 10 Graveley B R. Alternative splicing increasing diversity in the proteomic world. *Trends Genet*, 2001, **17**: 100 ~ 107
- 11 Hanke J, Brett D, Zastow I, *et al.* Alternative splicing of human genes more the rule than the exception? *Trends Genet*, 1999, **15** (10): 389 ~ 390
- 12 钱 骏, 董 利, 张必成, 等. 表达序列标签数据库搜索鉴定小鼠 UBAP1 基因及其数字化表达分析. *生物化学与生物物理进展*, 2002, **29** (2): 323 ~ 327
Qian J, Dong L, Zhang B C, *et al.* *Prog Biochem Biophys*, 2002, **29** (2): 323 ~ 327
- 13 Tollervey D, Caceres J F. RNA processing marches on. *Cell*, 2000, **103** (5): 703 ~ 709

Expression Regulation Analysis of HPCL2 Gene by *In silico* Alternative Splicing*

NIU Yu-Xin^{1,2)}, HU Song-Nian³⁾, YANG Huan-Ming^{1,2)} **

(¹⁾ Human Genome Center, Institute of Genetics and Developmental Biology, The Chinese Academy of Sciences, Beijing 100101, China;

²⁾ Beijing Genomic Institute, The Chinese Academy of Sciences, Beijing 101300, China; ³⁾ Hangzhou Genomics Institute, Hangzhou 310008, China)

Abstract In the third step of the α -oxidation of 3-methyl-branched fatty acids such as phytanic acid, a 2-hydroxy-3-methylacyl-CoA is cleaved into formyl-CoA and a 2-methyl-branched fatty aldehyde, which is regulated by 2-hydroxypanthyl-CoA lyase HPCL2. Using shot-gun sequencing technique, the genome sequence of the gene (HPCL2) code for the lyase was obtained. The gene is 40 829 bp in genome size including 17 exons and 16 introns, with an average 116 bp in exon size and 2 429 bp in intron size. The different transcriptional isoforms were also investigated by using method of alternative splicing *in silico*. In order to investigate the level of alternative

splicing and to search for novel splice variants, total 213 ESTs derived from 29 different tissues were collected and aligned against the genome sequence of the gene. 17 ESTs were found having different splicing type that could be classified into three types of alternative splicing. 14 of the ESTs were detected to have exon skipping, 2 intron not spliced and 1 splice site shift. The data suggested that exon skipping was possibly one of the major mechanisms of transcription regulation of HPCL2 gene.

Key words HPCL2, gene structure, alternative splicing, exon skipping

* This work was supported by grants from State 863 High Technology R&D Project of China (863-J19) and by CAS Innovation Program (KSCX1-D4).

** Corresponding author. Tel: 86-10-80494969, E-mail: hmyang@genetics.ac.cn

Received: April 10, 2003 Accepted: May 28, 2003

知识与动态

血清蛋白质组分析方法*

李 峰¹⁾ 肖志强²⁾ 陈主初^{1,2)}**

(¹⁾中南大学湘雅医学院肿瘤研究所, 长沙 410078; ²⁾中南大学湘雅医院医学实验中心, 长沙 410008)

蛋白质组学一经出现, 人们便开始尝试将其用于疾病特定分子标志物的研究当中, 比如通过比较疾病与正常生理状况下蛋白质表达谱的差异来寻找与疾病密切相关的蛋白质。但采用该方法有一明显的缺点, 即需要分析大量同一肿瘤组织的二维凝胶电泳 (two-dimensional gel electrophoresis, 2-DE) 图谱, 否则难以得到有统计学意义的差异蛋白质。另外, 由于细胞内的蛋白质表达是不均匀的, 这就导致了高表达和易溶解的蛋白质远较其他低表达、难溶解的蛋白质容易出现在 2-DE 图谱中, 因而明显地降低了计算机匹配分析发现有生物学意义的差异蛋白质的灵敏性。

血清蛋白质组分析方法 (serologic proteome analysis, SERPA) 是采用 2-DE 技术和免疫印迹技术相结合的一种新的蛋白质组分析方法, 其最早应用出现在肿瘤蛋白质组学的研究当中。从肿瘤免疫学的观点来看, 肿瘤细胞是“非己细胞”, 肿瘤细胞总是或多或少地表达区别于正常细胞的肿瘤抗原, 因而在患者血清中极有可能存在其相应的抗体, 所以 SERPA 技术可以结合蛋白质组学能分离肿瘤细胞内成千上万个蛋白质的优势, 和抗原抗体反应高度特异性的特点, 从而能够快速筛选出与肿瘤密切相关的蛋白质, 进而识别出肿瘤的标志物。其实验过程如下: 首先采用 2-DE 分离肿瘤组织、癌旁正常对照组织的蛋白质, 转至硝酸纤维素膜后, 与患者或正常人血清做蛋白质印迹分析来

建立蛋白质印迹反应图谱, 通过计算机分析确定差异反应的蛋白质斑点, 然后采用质谱鉴定和生物信息学联用的方法对肿瘤组织平行胶 (replica gel) 中相应的差异蛋白质点进行鉴定, 筛选出肿瘤分子标志物 (肿瘤相关抗原), 最后用 ELISA、免疫组化等方法对该分子标志物进行验证。

目前, 国外已有不少研究小组开始采用 SERPA 技术对包括肺癌、乳腺癌、神经母细胞瘤 (neuroblastoma)、肾细胞癌、肝癌以及卵巢癌等多种肿瘤在内的人类肿瘤进行了研究, 并寻找到了部分潜在的肿瘤分子标志物。总而言之, 血清蛋白质组分析方法是功能蛋白质组学研究思路的一大突破, 它能够敏感同时又很特异地识别出疾病状态和生理情况下细胞中的差异表达蛋白质, 因此, 我们有理由相信, 在不远的将来, 随着血清蛋白质组学研究的深入开展, 人们能够更容易获得疾病相关的特异蛋白质, 开发出诊断试剂和疫苗, 从而真正实现疾病的早期诊断和早期治疗。

* 国家“973”计划项目 (2001CB510)、国家自然科学基金项目 (30000028, 30240056) 和教育部跨世纪优秀人才培养计划基金 (教育部科技函 [2002] 48 号) 资助。

** 通讯联系人。

Tel: 0731-4805447, Fax: 0731-4485482

E-mail: tcb1@xymu.net