恒河猴 PPARgamma 基因的生物信息学分析
焦传珍
(德州学院生物系，山东德州 253023)

摘要：恒河猴 Macaca mulatta 是猕猴 Pan troglodytes 后第 2 个完成基因组测序的非人类灵长类动物，在生命科学领域中具有重要意义。本文利用生物信息学方法，寻找与人类肥胖密切相关的 PPARg 基因在恒河猴中的同源基因，对基因的序列、外显子信息、编码蛋白及其理化性质进行分析，并预测其编码蛋白的结构与功能，构建其同源基因的系统进化树，旨在为今后人类肥胖的相关研究提供一定的依据。

关键词：恒河猴；PPARgamma；生物信息学
中图分类号：Q959.848 文献标识码：A 文章编号：1000 – 7083(2009)04 – 0545 – 03

Bioinformatic Analysis of the PPARgamma Gene from Macaca mulatta
JIAO Chuan-zhen
(Dezhou University, Dezhou, Shandong Province 253023, China)

Abstract: The rhesus monkey is the second animal only the chimpanzee in its importance to life sciences, in particular in regards to determining their complete genome sequence. In this article, the PPARgamma gene, which was just discovered and is linked with human obesity, was examined. Its exons, coding proteins, and physical and chemical characteristics were analyzed. The structure and function of the proteins were forecasted, and a phylogenetic tree of the homologous gene was constructed. All of these will provide baseline information on human obesity for the future research.

Key words: Macaca mulatta; PPARgamma; bioinformatics

恒河猴 Macaca mulatta 是人类的近亲，在形态结构、生理机能和生化代谢方面同人类非常相似，成为生物医学和药物研究不可缺少的实验动物。测序结果显示恒河猴基因组与人类的基因组序列相似度为 92%—95%，而与黑猩猩 Pan troglodytes 基因组相似度在 98% 以上。因此恒河猴为这 3 种亲缘关系相近的灵长类动物（人类、黑猩猩和恒河猴）的对比提供了一个理想的参考点（Heidi et al., 2002）。

现代医学已证明，肥胖与许多慢性非传染性疾病的发生和发展密切相关，对肥胖的产生原因和预防的研究始终是医学和生物学的热点。PPARg 基因是使人体细胞转化为脂肪细胞的关键的一个环节，PPARg 蛋白是一种细胞核激素受体，负责调节细胞内信号的基因表达。PPARg 基因在前脂肪细胞转化或脂肪细胞的增生过程中起作用（金谷雷等, 2005），高表达于脂肪组织，而 PPARg 基因的激活伴随噬噬烧二酮类改变脂肪局部分布，脂肪细胞表型，并且上调多种与脂肪酸代谢和甘油三酯蓄积基因的表达。因此，PPARg 基因激活受体改善脂肪组织功能，对阻止糖尿病胰岛素抵抗的进展和动脉粥样硬化的内皮功能紊乱的进展具有作用。本文利用生物信息学方法，分析与人类肥胖密切相关的 PPARg 基因在恒河猴中的同源基因 PPARgamma，并预测其结构和功能，旨在通过对恒河猴 PPARgamma 基因的研究为人类肥胖的相关研究提供一定的依据。

1 分析方法

1.1 恒河猴 PPARgamma 基因结构分析

在 GenBank 数据库(http://www. ncbi. nlm. nih. goV/)中，搜索人类 PPARg 基因在恒河猴中的同源序列，分析其内含子和外显子信息。

1.2 恒河猴 PPARgamma 基因编码蛋白质的理化性质分析

利用互联网上提供的 ExPaSy (http://us. ExpaSy. org/tools/) 软件包中的 Computer pI/MW 和 BioEdit 软件进行蛋白质的氨基酸组成、分子质量、等电点以及疏水性分析。

1.3 恒河猴 PPARgamma 基因编码蛋白质的结构分析

利用互联网 ExPaSy 数据库，分析该蛋白质跨膜
结构域。利用PUMA2服务器的SOPM软件进行蛋白序列的二级结构分析，利用ExPaSy的3djigsaw工具（http://www.bmm.icnet.uk/servers/3djigsaw/）向蛋白质立体结构数据库PDB（Protein Data Bank）提交该蛋白质序列，利用RasMol软件显示该蛋白的三维分子结构。

### 1.4 恒河猴PPARgamma基因编码蛋白质的功能分析

利用Scansite（http://scansite.mit.edu/）网站工具，分析功能结构域及磷酸化位点。

### 1.5 构建系统进化树

选取不同进化层次的典型物种相应的同源基因，构建系统进化树。

#### 表1 恒河猴PPARgamma基因外显子信息

<table>
<thead>
<tr>
<th>序列号</th>
<th>Exon 1</th>
<th>Exon 2</th>
<th>Exon 3</th>
<th>Exon 4</th>
<th>Exon 5</th>
<th>Exon 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_007859</td>
<td>1 - 106</td>
<td>30592 - 30821</td>
<td>32190 - 32358</td>
<td>43661 - 43798</td>
<td>57565 - 57764</td>
<td>68339 - 68789</td>
</tr>
</tbody>
</table>

### 2.2 恒河猴PPARgamma基因编码蛋白质的理化性质分析

利用BioEdit软件及互联网上工具进行氨基酸组成、分子量、等点电荷分析结果，蛋白质分子式为C_{2275}H_{4641}N_{767}O_{776}S_{127}，分子数均为8091，蛋白质质量为57590.1，理论等点电为5.61。在氨基酸组成上，Leu有52个，占10.3%，含量最高，其次，Lys和Ser各有39个，分别占7.7%；Trp含量只有1个，占0.2%（表2）。

#### 表2 基因编码蛋白质的氨基酸组成

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Ala</th>
<th>Cys</th>
<th>Asp</th>
<th>Asn</th>
<th>Gln</th>
<th>Arg</th>
<th>Ser</th>
<th>Thr</th>
<th>Val</th>
<th>Trp</th>
<th>Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>28</td>
<td>10</td>
<td>37</td>
<td>34</td>
<td>27</td>
<td>21</td>
<td>14</td>
<td>36</td>
<td>39</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>mol%</td>
<td>5.5</td>
<td>2.0</td>
<td>7.3</td>
<td>6.7</td>
<td>5.3</td>
<td>4.2</td>
<td>2.8</td>
<td>7.1</td>
<td>7.7</td>
<td>10.3</td>
<td>3.4</td>
</tr>
</tbody>
</table>

对恒河猴PPARg蛋白进行亲疏水性分析得知该蛋白有1个亲水性区域，4个疏水性区域。

### 2.3 恒河猴PPARgamma基因编码蛋白质结构分析

利用Tmpred服务器（http://www.ch.embnet.org/software/TMPRED_form.html），通过分析疏水区域预测蛋白质的跨膜区域以及方向，输出结果包括可能的跨膜螺旋区、相关性列表，建议的跨膜拓扑模型。结果表明该蛋白质从内向外，由外向内分别含有两个可能的跨膜螺旋区，其中从内向外螺旋的位置在47～67, 477～494；由外向内螺旋的位置在47～66, 413～435, 472～494。为了确认以上结果，再利用ExPaSy的TMHMM工具预测该蛋白质的跨膜区，此结果说明该蛋白质不是膜蛋白，无跨膜区（http://www.ebi.ac.uk/services/TMHMM-2.0/）。

利用SOPM（http://compbio.mes.anl.gov/puma2/）软件工具，预测基因编码蛋白质的二级结构，α螺旋（Alpha helix）有196个，占38.81%，β折叠片（Beta turn）39个，占7.72%，延伸链（Extended strand）有105个占20.79%，无规则卷曲（Random coil）有165个，占32.67%（图1）。

利用ExPaSy服务器的3djigsaw（http://www.bmm.icnet.uk/servers/3djigsaw/）工具预测蛋白质的三维结构（图2），二者存在细微差别。

#### 2.4 恒河猴PPARgamma基因编码蛋白质的功能分析

利用Scansite分析motif（http://scansite.mit.edu/），采用默认参数，选择high Stringency，寻找蛋白质序列中易于被特定的蛋白质磷酸化的模块并确定结构域。结果发现，此蛋白不含同源结构域。其磷酸化位点分析发现SH2结构域；再Y78位置发现Fyn SH2和She SH2序列，在Y355位置发现SHIP SH2序列；Pro_ST_kin结构域；在S112位置发现Erk1 Kinase序列；Kin_bind结构域；在D69位置...
发现 PDK1 Binding 序列（图 3）。

2.5 构建系统进化树

利用 ClustalX 1.83 软件的 Bootstrap N-J Tree 方法获得系统进化树数据，然后用 Treeview 软件处理数据生成系统进化树（图 4）。选择在进化上具有代表性的 8 个物种（人类、黑猩猩、恒河猴、家犬、牛、小

鼠、大鼠、原鸡）进行比对，采取默认参数进行构建，结果显示，黑猩猩和人的这一基因相似性最高，它们和恒河猴的相似性也很高，并与家犬和牛在进化上有较近的遗传距离，说明基因编码蛋白质的变异不大。

3 讨论

本文运用生物信息学方法首次对恒河猴的 PPARgamma 基因进行分析，从其基因序列、编码蛋白的理化性质、结构及功能做简要分析，结果发现此基因与人类肥胖相关基因 PPARg 具有 99.2% 以上的同源性，通过三级结构观察，恒河猴 PPARgamma 蛋白和人的 PPARg 蛋白的立体结构有很高的相似性，这是由于两个蛋白的序列同源所决定的。对恒河猴 PPARgamma 基因的研究可能为今后进一步研究人类肥胖相关问题提供依据（Yang et al., 2008）。

在生命科学和医学的研究和应用中，利用生物信息学分析生物分子数据，可以大大提高研究开发的科学性及效率（瞿飞等，2008），如：根据基因功能分析结果来检测与疾病相关的基因，根据蛋白质分析结果研究某些疾病的发病机理并进行新药设计。随着分子生物学技术和计算机科学的迅速发展，以及互联网在全球普及，生物信息学以其大规模信息量、快速等优势已成为推动基因组学和后基因组学研究的一种重要技术，必将加速生命科学研究（杨美香等，2005）。

4 参考文献

金谷雷，任旭生，宋军。2005。水稻 14_3_3 蛋白家族的生物信息学分析[J]。遗传学报，32(7)；726 ~ 732。

瞿飞，赵梅，焦传珍，等。2008。紫色球海胆 Tel/lef 基因的生物信息学分析[J]。水产科学，27(1)；32 ~ 34。

杨美香，曲迟，韩克军，等。2005。应用生物信息学方法分析人 HICA56 基因[J]。基础医学与临床，25(2)；169 ~ 172。

Heidi，Camp，Bruce Spiegelman。2002。Scientists Discover Gene For Fat Cell Development [N]。Science Daily，2002-01 ~ 07

Yang L，Zhou ZG，Zheng XL。2008。RNA interference against peroxisome proliferator-activated receptor delta gene promotes proliferation of human colorectal cancer cells[J]。Dis Colon Rectum，6(24)；846 ~ 851。