

文章编号: 1674-8085(2016)01-0051-06

人线粒体转录终止因子 2 基因启动子区的 生物信息学分析

*熊 伟^{1,2}, 张晓娟³, 孙美涛¹, 张海洋¹, 徐 彤¹

(1. 大理大学基础医学院, 云南, 大理 671000; 2. 云南省昆虫生物医药研发重点实验室, 云南, 大理 671000;

3. 大理市第一人民医院, 云南, 大理 671000)

摘 要: 人线粒体转录终止因子 2 基因 (mitochondrial transcription termination factor 2, *MTERF2*) 在线粒体基因表达、细胞能量代谢及细胞周期调控过程中发挥重要作用, 但 *MTERF2* 基因自身的表达调控机制仍不明确。通过生物信息学方法预测和分析人线粒体转录终止因子 2 基因启动子区的结构与功能, 结果显示: 人 *MTERF2* 基因全长 9876 bp, 含有 3 个外显子和 2 个内含子。基因上游 5'端 2500 bp 的核苷酸序列存在 3 个启动子, 启动子区序列中存在长为 874 bp 的 CpG 岛, 启动子区共发现 244 个转录因子结合位点。人 *MTERF2* 基因启动子区的生物信息学分析, 提高了针对该基因启动子的研究效率, 并为分析基因启动子的功能奠定了重要的理论基础。

关键词: 人线粒体转录终止因子 2; 启动子; 转录因子; 生物信息学

中图分类号: Q811.4

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2016.01.011

BIOINFORMATICS ANALYSIS OF HUMAN *MTERF2* GENE PROMOTOR REGION

*XIONG Wei^{1,2}, ZHANG Xiao-juan³, SUN Mei-tao¹, ZHANG Hai-yang¹, XU Tong¹

(1. College of Basical Medical Sciences, Dali University, Dali, Yunnan 671000, China;

2. Yunnan Provincial Key Laboratory of Entomological Biopharmaceutical R&D, Dali, Yunnan 671000, China;

3. Dali First People's Hospital, Dali, Yunnan 671000, China)

Abstract: Human mitochondrial transcription termination factor 2 (*MTERF2*) plays an important role in mitochondrial gene expression, energy metabolism and cell cycle regulation. However, the potential regulation mechanism of *MTERF2* gene expression was still undiscovered. Online bioinformatics tools were used to analyze human *MTERF2* gene's promoter. The results showed that the full length of human *MTERF2* gene was 9876 bp, which consist of three exons and two introns. There were three promoters in the 5' unconding region of human *MTERF2* gene, two of which played a key role in the transcription of *MTERF2*. A 874 bp CpG island was found in human *MTERF2* gene promoter region. In addition, total 244 transcription factor binding sites were found by AliBaba 2.1 protocol. Gene promoter related bioinformatics can improve the efficiency of human *MTERF2* gene promoter research, and provide significant information for the prediction of gene promoter function.

Key words: human *MTERF2*; promoter; transcription factor; bioinformatics

收稿日期: 2015-10-12; 修改日期: 2015-12-02

基金项目: 国家自然科学基金项目(31260276); 云南省教育厅科学研究基金重点项目(2014Z126); 大理大学博士科研启动基金项目(BSKY2012018); 大理大学大学生创新创业训练计划项目(201402)

作者简介: *熊 伟(1982-), 男, 湖南株洲人, 副教授, 博士, 硕士生导师, 主要从事细胞分子生物学与生物信息学研究(E-mail: xwailp@163.com); 张晓娟(1984-), 女, 云南大理人, 护师, 主要从事儿科护理研究(E-mail: 791372366@qq.com); 孙美涛(1988-), 女, 山东日照人, 硕士生, 主要从事细胞分子生物学研究(E-mail: smtshandong@126.com); 张海洋(1989-), 男, 安徽合肥人, 硕士生, 主要从事细胞分子生物学研究(E-mail: hanyangzhang88@163.com); 徐 彤(1989-), 女, 云南昭通人, 硕士生, 主要从事生物信息学研究(E-mail: xutong1989@163.com).

人线粒体转录终止因子 2 (mitochondrial transcription termination factor 2, *MTERF2*) 又称为线粒体转录终止因子结构域 3 (mitochondrial transcription termination domain containing 3, *MTERFD3*) 或线粒体转录终止样因子 (mitochondrial transcription termination like factor, *MTERFL*), 该基因是通过比较血清饥饿和血清培养的人成纤维细胞基因表达谱差异, 发现的 1 个细胞增殖抑制基因^[1]。前期研究发现, *MTERF2* 定位于细胞线粒体内, 过表达 *MTERF2* 基因对细胞增殖有较强的抑制作用, 但并不导致细胞凋亡^[1]。人 *MTERF2* 属于线粒体转录终止因子蛋白家族, 对人类细胞线粒体基因的复制和表达具有调控作用, 同时调控细胞的氧化磷酸化过程^[2-3]。也有研究显示 *MTERF2* 是哺乳动物线粒体的类核蛋白质, *MTERF2* 基因敲除小鼠出现明显的呼吸功能障碍和能量代谢异常, 且较正常小鼠体重显著减轻^[4-5]。目前, 针对人 *MTERF2* 蛋白的体内和体外生理功能已有比较深入的研究, 但是对于人 *MTERF2* 基因表达调控的机制仍然缺乏了解。

本研究利用生物信息学技术对人 *MTERF2* 基因序列、启动子区进行预测和分析, 获得该基因 CpG 岛及启动子区转录因子结合部位, 以期为人 *MTERF2* 基因调控区的分析与鉴定、基因表达调控研究提供重要的信息^[6]。

1 材料和方法

1.1 人 *MTERF2* 基因组 DNA 序列

人 *MTERF2* 基因定位于人类第 12 号染色体长臂 (12q24.1), 该基因也被称为 *MTERFL* 或 *MTERFD3*, 基因跨越 9876 bp (Chromosome 12: 106977291~106987166), 含有 3 个外显子和 2 个内含子。转录 mRNA 的 ID 为 XM_011538767.1, 编码蛋白质产物由 385 个氨基酸组成, 该基因在 GenBank 尚未记录启动子区的序列。

1.2 程序和数据库

在线核酸序列数据库: <http://www.ncbi.nlm.nih.gov/genbank/>。启动子在线分析软件: Promoter 2.0 (<http://www.cbs.dtu.dk/services/promoter/>); Neural Network Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html) ; Promoter SCAN (<http://www-bimas.cit.nih.gov/molbio/proscan>) ; First EF (<http://rulai.cshl.org/cgi-bin/tools/FirstEF>)。转录因子预测软件: AliBaba 2.1 (<http://www.gene-regulation.com/pub/programs/alibaba2/index.html>)。CpG island 预测软件: (<http://www.uscnorris.com/cpgislands2/cpg.aspx>); CpGPlot (<http://www.ebi.ac.uk/emboss/cpgplot/index.html>)。

1.3 方法

1.3.1 获得人 *MTERF2* 基因序列

在 <http://www.ncbi.nlm.nih.gov/genbank> 数据库中检索 *MTERF2* 获得人 *MTERF2* 基因 ID 为 80298, 采用 Genbank 格式获得基因信息。

1.3.2 人 *MTERF2* 基因启动子区序列的确定

在 NCBI 数据库的 GenBank 中查找, 获得人 *MTERF2* 基因的转录起始位点在基因序列图上的定位, 并获取转录起始位点上游-2000 bp 至转录起始点下游的+500 bp 序列 (包括第一外显子和第一内含子)。该序列包含潜在的基因启动子序列^[6]。

1.3.3 人 *MTERF2* 基因启动子序列的预测分析

采用在线启动子预测软件在默认条件下对取得的人 *MTERF2* 基因序列进行预测分析。

1.3.4 人 *MTERF2* 基因启动子 CpG 岛的分析

输入人 *MTERF2* 基因上游至 5'侧翼 2500 bp 的序列, 按照预测软件默认条件进行 CpG 岛的预测分析。

1.3.5 人 *MTERF2* 基因启动子区转录因子结合位点分析

利用 AliBaba 2.1 程序, 输入人 *MTERF2* 基因上游至 5'侧翼 2500 bp 的序列, 搜索 TRANSFAC 4.0 数据库, 获得该基因启动子区转录因子结合位点^[6]。

2 结果与分析

2.1 人 *MTERF2* 基因特征

人 *MTERF2* 基因 GenBank 登录号为 NC_000012.12, 位于第 12 号染色体长臂 (12q24.1), 该基因也被称为 *MTERFD3* 或 *MTERFL*, 基因跨越 9876 bp, 含有 3 个外显子和 2 个内含子 (图 1)。第一个外显子位于 106987166~106986969, 长度为 198 bp; 第 2 个外显子位于 106985225~106985115,

长度为 111 bp; 第 3 个外显子位于 106978771~106977291, 长度为 1481 bp, 该区域含有翻译起始密码子 ATG 和所有编码蛋白质的序列。该基因转

录 mRNA 全长 1763 bp, ID 号为 XM_011538767.1, 编码蛋白质产物含有 385 个氨基酸, N 端 1~35 个氨基酸为导肽, 其后 350 个氨基酸为成熟蛋白质的序列。

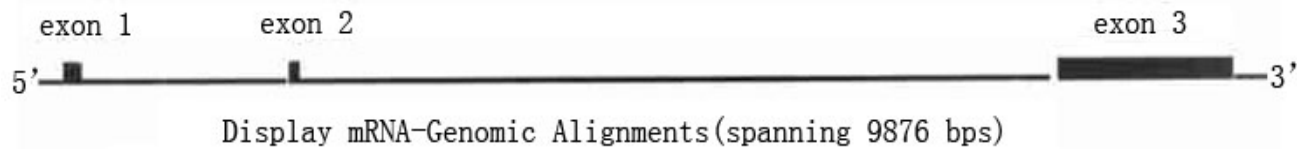


图 1 人 *MTERF2* 基因的序列结构特征

Fig.1 Stucture characteristics of human *MTERF2* gene sequence

2.2 人 *MTERF2* 基因启动子预测结果

2.2.1 Promotor2.0 预测结果

Promotor 2.0 是根据对启动子区各种转录信号 (如 CCAAT box、GC box、TATA box) 及信号间保守的空间排列顺序的识别, 用神经网络的方法进行预测^[7]。Promotor 2.0 预测结果显示, 人 *MTERF2* 基因上游存在着 3 个启动子。其中 500 bp 处为高度可能的启动子, 其它则为临界性预测 (表 1)。

2.2.2 Neural Network 预测结果

Neural Network Promotor Prediction 综合利用

了 Promotor 2.0 的神经网络算法和 PROSCAN 的一些分析方法, 特异性比较高^[8]。Neural Network 预测结果显示, 人 *MTERF2* 基因上游同样有 3 个启动子。除具体位置分布略有差异外, 与 Promotor 2.0 预测的结果基本一致 (表 2)。

表 1 Promotor 2.0 软件预测的人线粒体转录终止因子 2 基因的启动子

Table 1 Promoters of human *MTERF2* predicted by the Promotor 2.0

启动子	位置	得分	可信度
1	500 bp	1.805	高度可能预测
2	1000 bp	0.572	临界性预测
3	1700 bp	0.632	临界性预测

表 2 Neural Network 软件预测的人线粒体转录终止因子 2 基因的启动子

Table 2 Promoters of human *MTERF2* predicted by the Neural Network promotor prediction

启动子	起始位点	结束位点	得分	启动子序列
1	379 bp	429 bp	0.98	atgcactgtttaaatatttggcgctggaaaaactttgagtagcatttatt
2	1808 bp	1858 bp	0.95	aaggaaagggtgtgaaaatgccagatgccccgccagcacaggctgggct
3	2389 bp	2439 bp	0.94	tttccttgcttataaacgttggctcaatttgaactggcagcagggcataat

表 3 PROSCAN 软件预测的人线粒体转录终止因子 2 基因的启动子

Table 3 Promoters of human *MTERF2* predicted by the PROSCAN

转录因子	正负链	位置 (bp)	权重
AP-2	-	1947	1.091000
KROX24	+	1950	2.151000
T-Ag	+	1953	1.086000
EGR-1	-	1958	2.294000
KROX24	-	1958	1.912000
(Sp1)	-	1959	2.233000
(Sp1)	+	1967	4.589000
Sp1	+	1968	6.023000
Sp1	+	1968	3.129000
Sp1	+	1969	3.013000
Sp1	+	1970	3.191000
Sp1	-	1974	3.061000
AP-2	-	1974	1.672000
Sp1	-	1975	3.119000
(Sp1)	-	1976	6.819000
EARLY-SEQ1	-	1976	5.795000
Sp1	-	1977	3.129000
Y	+	1999	9.680000
CTF	+	2001	1.704000
NFI	-	2007	4.221000
junB-US2	-	2008	1.510000
UCE2	+	2125	1.278000
T-Ag	+	2147	1.086000
SIF	-	2149	1.161000
AP-2	+	2156	1.108000
NF-kB	+	2161	1.080000
NF-kB	-	2169	1.008000

2.2.3 PROSCAN 预测结果

PROSCAN 是根据转录因子结合部位在基因组中分布的不平衡性,将转录因子结合部位分布密度和 TATA box 的权重矩阵结合起来,从基因组 DNA 序列中识别出启动子区^[9]。PROSCAN 预测结果显示,与人 *MTERF2* 基因启动子结合的转录因子主要存在于 1933~2183 bp 之间,启动子得分为 76.07,启动子 Cutoff=53.00 (表 3)。

2.2.4 First EF 预测结果

FirstEF 软件是通过搜索 5'非转录区定位技术构建的第一外显子数据库,识别第一剪切点,结合 CpG 岛信息,确定基因的启动子区^[10]。FirstEF 预

测结果显示,人 *MTERF2* 基因启动子位于 1347~1916 bp 之间, *P* 值为 0.9991。同时,FirstEF 软件预测的 CpG 岛位置为 1828~2029 bp。

2.3 人 *MTERF2* 基因 CpG 岛预测和分析

2.3.1 CpG Island Search 预测结果

使用 CpG Island Search 软件对人 *MTERF2* 基因 CpG 岛进行分析时发现,人 *MTERF2* 基因 CpG 岛位于启动子区 1627~2500 bp 之间[(G+C)% = 56.6%; ObsCpG/Exp = 0.749, length = 874 bp],检索标准为 [(G+C)% = 55%; ObsCpG/Exp = 0.65, length > 200 bp]。且横跨第一启动子和转录起始位点,这符合 CpG 岛的分布特点(图 2)。

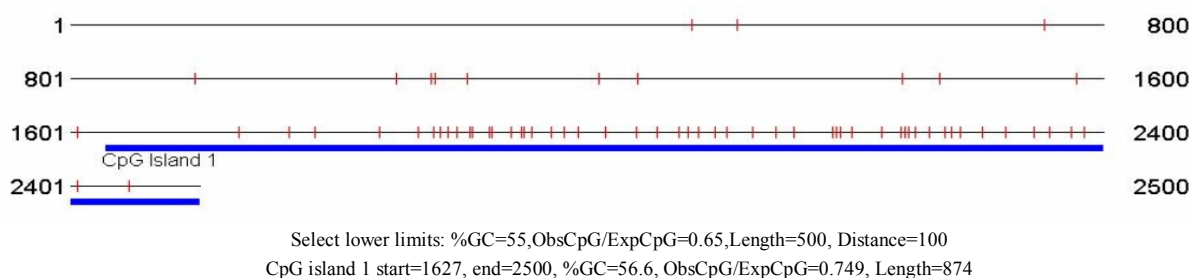


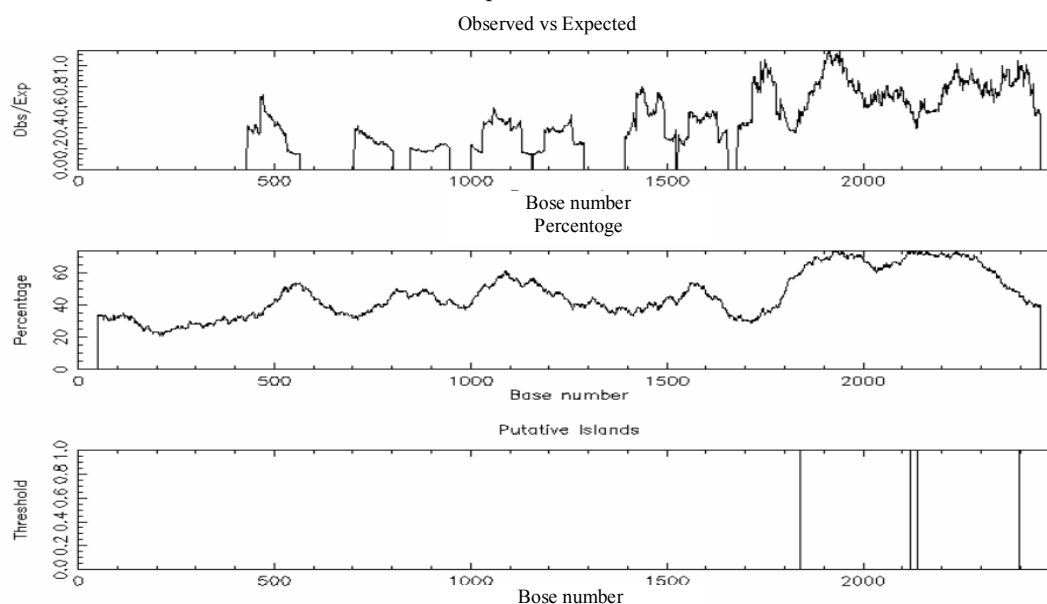
图 2 人 *MTERF2* 基因上游 5'端启动子区 CpG 岛分析 (图中粗线蓝色区域为 CpG 岛所在的位置)

Fig. 2 CpG island prediction of 5' conservative region of human *MTERF2* gene

2.3.2 CpG Plot 预测结果

使用 EMBL-EBI 提供的 CpG Plot 软件[标准为 (G+C)% > 50.00%; ObsCpG/Exp > 0.60, length > 200 bp]在人 *MTERF2* 基因 5'端上游启动子区发现 2 个 CpG

岛。第 1 个 CpG 岛位于 1842~2120 bp 之间,长度为 279 bp; 第 2 个 CpG 岛位于 2140~2398 bp 之间,长度为 259 bp。CpG Plot 预测的 CpG 岛位置与 CpG Island Search 软件预测的覆盖位置基本吻合(图 3)。



(柱状图为 CpG 岛所在位置)

图 3 人 *MTERF2* 基因上游 5'端启动子区 CpG 岛预测

Fig. 3 CpG island prediction of 5' conservative region of human *MTERF2* gene

2.4 人 *MTERF2* 基因转录因子结合位点分析

使用 AliBaba 2.1 程序搜索 TRANSFAC 4.0 数据库, 获得人 *MTERF2* 基因启动子区 2500 bp 正负链转录因子结合位点共计 244 个, 主要包括 NF-1、Oct-1、HNF-1、HNF-3、C/EBP、Sp1、TBP、GATA-1、NF-kB、AP-1、c-Myc、Id3、Pit-1a、WT1、c-Jun、GR、Adf-1、Egr-1、Antp、USF、Krox-20、SRF、MEB-1、T-Ag、CACCC-binding factor 等。AliBaba 2.1 程序预测到转录因子的结合位点数目远多于 PROSCAN 软件预测的结果, 但其预测结果有一部分是一致的。

3 讨论

人 *MTERF2* 属于线粒体转录终止因子蛋白家族, 位于人类第 12 号染色体长臂 (12q24.1), 含有 3 个外显子, 2 个内含子^[11]。人 *MTERF2* 基因编码的蛋白质产物由 385 个氨基酸组成, N 端 1-35 个氨基酸为进入线粒体的导肽, 其后 350 个氨基酸为成熟的蛋白质。人 *MTERF2* 蛋白能调节线粒体基因的复制和转录过程, 并参与细胞能量代谢及细胞周期调控, 对神经退行性疾病、线粒体糖尿病、肿瘤等疾病的诊断与治疗具有一定的应用前景^[12]。

启动子是 RNA 聚合酶识别、结合并开始基因转录所必需的一段 DNA 序列。目前, 对基因启动子区的预测算法有多种, 不同的算法有不同的预测结果, 且存在一定的假阳性, 需综合预测和分析^[13]。本研究采用 4 种不同的启动子分析软件对人 *MTERF2* 基因启动子区进行分析, 结果发现该基因可能存在着 3 个启动子, 尤其是-450 bp 左右和-2000 bp 之内, 是两个非常重要的启动子区。所采用的 4 种分析软件结构都相似地显示了这一结果。这提示人 *MTERF2* 基因的核心启动子就在此范围内。与文献报道的基因启动子位置一般在转录起始位点-2000 bp 以内是一致的^[14]。

CpG 岛是哺乳类动物基因组中长度为 0.5-4.0 Kb 的一段富含 CpG 二核苷酸成分的 DNA 序列, 几乎都位于基因的启动子区和第一个外显子区^[12]。有研究表明, 对于一个基因而言, 局部序列的 GC 含量存在不均一性, 这为研究该基因的进化、结构和功能提供了线索^[15]。由于 CpG 岛对于基因的转录有

重要的调控作用, 所以对于转录调控区的 CpG 岛研究也是非常有必要的。现阶段使用实验方法检测基因启动子区 CpG 岛的难度相对较大。基于生物信息学方法, 通过预测软件以理论预测的方法可以快速有效的获得基因的结构和状态。本研究使用 CpG Island Search 软件和 CpG Plot 软件对基因启动子区 CpG 岛进行预测时, CpG Island Search 软件预测人 *MTERF2* 基因启动子区有 1 个 CpG 岛, 位于启动子区 1627~2500 bp 之间, 长度为 874 bp。而 CpG Plot 软件则预测人 *MTERF2* 基因启动子区有 2 个 CpG 岛, 长度分别为 259 bp 和 279 bp, 但这两个序列所在的位置与 CpG Island Search 软件分析的位置是基本吻合的。

转录因子通常也称为反式作用因子, 它们通过与顺式作用元件和 RNA 聚合酶的相互作用来调节不同基因的转录活性。不同的转录因子对不同发育阶段的各种信号或外界环境刺激作出反应, 激活或抑制基因的转录, 从而控制不同基因的表达^[16]。我们首先在 TRANSFAC 4.0 数据库进行注册, 再利用 AliBaba 2.1 程序搜索, 获得人 *MTERF2* 基因启动子区 2500 bp 正负链转录因子结合部位共计 244 个。由于在转录因子结合位点的分析中, 程序只能针对已知转录因子的结合位点进行预测, 对未知的或新的转录因子结合位点则无法进行分析, 同时由于真核生物之间存在特异性和预测软件原理局限性, 该预测方法仍存在着一些不足, 这些生物信息学软件分析的结果有待于今后的实验研究来进一步验证。

本研究通过在 NCBI 的 GenBank 数据库获得人 *MTERF2* 基因序列, 并获取人 *MTERF2* 基因上游至 5'侧翼 2500 bp 序列, 采用不同的在线分析软件对人 *MTERF2* 基因序列特征、启动子序列、CpG 岛以及转录因子结合位点进行了预测和分析, 为下一步构建人 *MTERF2* 基因启动子表达载体、启动子活性的检测奠定了坚实的理论基础。

参考文献:

- [1] Chen Y, Zhou G, Yu M, et al. Cloning and functional analysis of human mTERFL encoding a novel mitochondrial transcription termination factor-like protein[J]. Biochem Biophys Res Commun, 2005, 337(4): 1112-1118.

- [2] Xiong W, Huang W, Jiao Y, et al. Production, purification and characterization of mouse monoclonal antibodies against human mitochondrial transcription termination factor 2 (*MTERF2*)[J]. *Protein Expr Purif*, 2012, 82(3): 11-19.
- [3] Huang W, Yu M, Jiao Y, et al. Mitochondrial transcription termination factor 2 binds to entire mitochondrial DNA and negatively regulates mitochondrial gene expression[J]. *Acta Biochim Biophys Sin*, 2011, 43(6): 472-479.
- [4] Pellegrinia M, Asin-Cayuela J, Erdjument-Bromage H, et al. *MTERF2* is a nucleoid component in mammalian mitochondria[J]. *Biochim Biophys Acta-Bioenergs*, 2009, 1787(5): 296-302.
- [5] Wenz T, Luca C, Torraco A, et al. *MTERF2* regulates oxidative phosphorylation by modulating mtDNA transcription[J]. *Cell Metab*, 2009, 9(6): 499-511.
- [6] 孙勇,王良喜,孙曙光,等. 人三叶因子 3 基因启动子区的生物信息学分析[J]. *医学研究生学报*, 2013, 26(4): 340-342.
- [7] 张万里,李伟,王川清,等. 小鼠 EDNRB 基因启动子生物信息学分析[J]. *生物信息学*, 2011, 9(1): 72-75.
- [8] 杨树德,曹奇志,冯小英,等. 人 LAIR-1/CD305 基因启动子的生物信息学分析[J]. *细胞与分子免疫学杂志*, 2010, 26(8): 826-828.
- [9] 唐亮,王燕,陈碧峰. 人 Toll 样受体 9 基因启动子区别的生物信息学分析[J]. *生物技术通讯*, 2014, 25(2): 210-212.
- [10] 陈瑶,何云刚,余敏,等. 一个人类线粒体转录终止样基因的克隆[J]. *云南大学学报:自然科学版*, 2003, 25(12): 205-209.
- [11] 熊伟,余敏,左绍远. 线粒体转录终止因子蛋白家族在线粒体基因表达中的调节作用[J]. *中国生物化学与分子生物学报*, 2015, 31(3): 223-231.
- [12] 王佐广,牛秋丽,刘雅,等. 线粒体融合基因 2 启动子的生物信息学研究[J]. *中国动脉硬化杂志*, 2013, 21(3): 97-103.
- [13] 张强,卜凡莉,徐金会,等. 黑线仓鼠 KiSS-1 基因的生物信息学与系统进化研究[J]. *井冈山大学学报:自然科学版*, 2012, 33(5): 99-106.
- [14] 陈保锋,曾梅,申跃武,等. 人 Boule 基因启动子区结合蛋白的生物信息学分析[J]. *现代生物医学进展*, 2012, 13(31): 6184-6188.
- [15] 管晓翔,陈巍巍,陈龙邦,等. P27^{kip1} 基因启动子区的生物信息学分析[J]. *医学研究生学报*, 2010, 23(10): 27-30.

(上接第 50 页)

- [6] Raymundo-Pinero E, Khomenko V, Frackowiak E, et al. Performance of manganese oxide/CNTs composites as electrode materials for electrochemical capacitors[J]. *J. Electrochem. Soc*, 2005, 152(1): 229-235.
- [7] Zhang Y, Feng H, Wu X, et al. Progress of electrochemical capacitor electrode materials[J]. In: *J. Hydrogen Energ.*, 2009, 34(11): 4889-4899.
- [8] Yan J, Wu T, Cheng J, et al. Preparation and electrochemical properties of lamellar MnO₂ for supercapacitors[J]. *Mater. Res. Bull.*, 2010, 45(2): 210-215.
- [9] Wei W, Cui X, Chen W, et al. Electrochemical cyclability mechanism for MnO₂ electrodes utilized as electrochemical supercapacitors[J]. *J. Power Sources*, 2009, 186(2): 543-550.
- [10] 张宗杰,解勤兴,张宇峰. 金属氧化物/碳超级电容器复合材料研究进展[J]. *绵阳师范学院学报*, 2014, 33(2): 50-55.
- [11] 李俊,王先友,黄庆华,等. 超级电容器 MnO₂/CRF 复合电极材料的制备及性能的研究[J]. *功能材料*, 2006, 12(37): 1938-1941.
- [12] 朱修锋,景晓燕. 金属氧化物超级电容器及其应用进展[J]. *功能材料与器件学报*, 2002, 8(3): 325-330.
- [13] 袁磊,王朝阳,付志兵,等. MnO₂/碳气凝胶复合电极材料制备及性能的研究[J]. *原子能科学技术*, 2010, 44(7): 864-868.