

# 线粒体融合基因 2 启动子的生物信息学研究

王佐广, 牛秋丽, 刘雅, 刘洁琳, 顾伟, 刘阔, 楼煜清, 文杰, 吴兆苏, 温绍君

(首都医科大学附属北京安贞医院 北京市心肺血管疾病研究所高血压研究室, 北京市 100029)

[关键词] 线粒体融合基因 2; 启动子; 转录调控; 转录因子; 生物信息学

[摘要] **目的** 线粒体融合基因 2 是一个新发现的负调控血管平滑肌细胞增殖的基因, 在原发性高血压患者血管平滑肌细胞中低表达, 但其调控机制还不明确。本研究通过生物信息学方法研究线粒体融合基因 2 的转录和调控机制。**方法** 通过使用在线软件, 对线粒体融合基因 2 上游 5' 端的启动子分布、转录因子及其结合位点、CpG 岛、TATA box 分析、保守区域等进行分析。同时对该区 SNP 的分布及其功能进行研究。**结果** 线粒体融合基因 2 上游 5' 端存在 5 个启动子, 但是起核心作用的可能只有 2 个; 有 2 个保守区域, 存在 69 个转录因子结合位点, 后者主要与炎症反应、氧化应激、细胞增殖与分化和 MAPK、STAT 信号传导通路相关; 一个 CpG 岛; 15 个 TATA box, 有 8 个处于 -2000 bp 以内; 32 个 SNP 位点, 其中 9 个处于核心启动子区, 5 个可能会影响线粒体融合基因 2 的转录。**结论** 线粒体融合基因 2 上游 5' 端 -400 bp 以内可能存在着核心启动子, 该启动子区存在着一些重要的与线粒体融合基因 2 功能相关的转录因子。同时, 启动子区存在的大量 SNP 可能会影响基因的转录调控。

[中图分类号] Q81

[文献标识码] A

## Bioinformatic Study of Mitofusion 2 Gene Promoter Region

WANG Zuo-Guang, NIU Qiu-Li, LIU Ya, LIU Jie-Ling, GU Wei, LIU Kuo, LOU Yu-Qing, WEN Jie, WU Zhao-Su, and WEN Shao-Jun

(Department of Hypertension, Beijing Institute of Heart, Lung & Blood Vessel Diseases, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, China)

[KEY WORDS] Mitofusion 2; Promoter; Transcription Regulatory; Transcription Factor; Bioinformatics

[ABSTRACT] **Aim** Mitofusion 2 (Mfn2) is a gene which can negatively regulate the hyperplasia of vascular smooth muscle cells (VSMC). The gene is down-regulated in the VSMC of essential hypertensive patients. However, the potential mechanism is undiscovered. The object of this study is to investigate the mechanism of Mfn2 gene transcription and regulation by bioinformatics. **Method** By using online software, the promoter distribution, transcription factors and their binding sites, CpG island, TATA box, and conservative region were analyzed in the 5' uncoding region of Mfn2 gene.

Simultaneously, the distribution and function change of SNP were studied. **Results** There were five promoters in the 5' uncoding region of Mfn2 gene, which two of them played a key role in the transcription of Mfn2. Sixty nine transcription factors' binding sites were found in the two conservative districts. These transcription factors were mainly associated with inflammation, oxidative stress, hyperplasia and differentiation of cells, MAPK and STAT signal pathway. One CpG island was found. Eight of 15 TATA boxes were found in the region of -2000 bp upstream of the transcription origin site. Thirty two SNP sites were found, nine were in the core promoter region, and five may affect the transcription of Mfn2. **Conclusion** The core promoter distributes in the -400 bp region of 5' uncoding region of Mfn2. There were a lot of transcription factors which were closely related with the function of Mfn2 existed. In addition, some SNPs which may affect the transcription regulation of Mfn2 were found in the promoter region.

[收稿日期] 2012-03-13

[基金项目] 国家自然科学基金(81270216)和国家 863 高技术项目基金资助(2008AA02Z441), 北京市自然科学基金重大项目(7120001)及北京市自然科学基金项目(7102045)

[作者简介] 王佐广, 博士, 助理研究员, 医师, 主要从事高血压的基础与临床研究。吴兆苏, 硕士, 研究员, 博士研究生导师, 主要从事心血管疾病的流行病学研究。通讯作者温绍君, 硕士, 研究员, 医师, 博士研究生导师, 主要从事高血压的基础与临床研究, E-mail 为 wzg70@hotmail.com。

血管平滑肌细胞(vascular smooth muscle cells, VSMC)的异常增殖是原发性高血压(essential hypertension, EH)动脉粥样硬化常见的原因,同时也是EH的常见结果,但是其潜在的分子机制还不清楚。1997年陈光慧等<sup>[1]</sup>从自发性高血压大鼠(spontaneously hypertensive rat, SHR)血管中分离出来一种可以抑制VSMC增殖的基因,即增殖抑制基因,后来被统一命名为线粒体融合基因2(mitofusin 2, Mfn2)。研究发现, Mfn2基因在SHR和球囊损伤后增殖的WKY大鼠动脉VSMC中转录明显降低。而过表达Mfn2可以明显抑制培养的VSMC增殖,并能阻断球囊损伤所诱导的大鼠颈动脉VSMC增殖和再狭窄<sup>[2]</sup>。近来的其它几项研究发现Mfn2的抗增殖作用是由ERK/MAPK信号通路介导并与随后的细胞周期顿抑有关<sup>[3,4]</sup>。因而, Mfn2是一个具有细胞增殖负调控功能的基因。另外, Mfn2在线粒体融合过程中发挥着非常重要的作用。由于线粒体分裂和融合是确保细胞代谢、线粒体DNA混合及线粒体形态、数目和生物学功能的重要事件,作为两种线粒体融合蛋白中的一种, Mfn2属于大GTP结合蛋白家族<sup>[5,6]</sup>。如果线粒体分裂和融合之间的平衡被打破,则会诱导出现细胞内能量代谢、氧化磷酸化、钙信号转导异常以及凋亡等。例如,线粒体产生的过氧化物可以引起线粒体氧化应激,并导致线粒体功能下降,而这将会引发如高血压和动脉粥样硬化等一系列的病理生理学过程<sup>[7]</sup>。由此可见, Mfn2在VSMC增殖的负调控中非常重要,同时也是线粒体融合的一种关键基因,因而对于EH发病及其并发症具有重要的意义。但是Mfn2基因作为一种负调控基因,为什么在EH患者和SHR的VSMC中低表达?其潜在的分子机制是什么?明确这些问题对于防止EH患者VSMC增殖、动脉粥样硬化的形成以及心血管系统并发症方面,具有非常重要的意义。我们的一些前期研究已经发现Mfn2上游调控区存在着一些单核苷酸多态性(single nucleotide polymorphisms, SNP)与EH明显相关<sup>[8]</sup>。但是该基因详细的转录调控机制还不明确。因此,本研究将从生物信息学的角度去探讨该基因转录调控机制,为进一步从分子遗传学角度揭示该基因转录调控机制提供参考。

## 1 材料和方法

### 1.1 人线粒体融合基因2的基因序列

Mfn2基因位于1p36.22,该基因也被称为CMT2A、CMT2A2或CPRP1。本研究从<http://www.ncbi.nlm.nih.gov/gene/bank>数据库中获得Mfn2基因序列。可通过NCBI Reference Sequence: NC\_007945.1检索得到。

### 1.2 程序和数据库<sup>[9,10]</sup>

启动子在线分析软件为Promoter 2.0: <http://www.cbs.dtu.dk/services/promoter>; Neural Network Promoter Prediction: [www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html); PromoterSCAN: <http://www.bimas.cit.nih.gov/cgi-bin/molbio/proscan>。该软件是使用PROSCAN Version 1.7 suite,由明尼苏达大学高级生物科学计算中心的Dan Prestidge博士编写。FirstEF: <http://rulai.cshl.org/tools/FirstEF/>,由Ramana Davuluri, Ivo Grosse和Michael Zhang博士编写的在线人类DNA第一个外显子和启动子分析软件。

甲基化CpG岛预测在线软件为CpG Island Searcher: <http://www.uscnorris.com/cpgislands2/cpg.aspx>及欧洲分子生物学实验室(The European Molecular Biology Laboratory, EMBL)提供的CpG island在线预测软件<http://www.ebi.ac.uk/Tools/emboss/cpgplot/>。

转录因子预测有转录因子结合位点: <http://mbs.cbrc.jp/research/db/TFSEARCH.html>。TATA盒预测软件为TATA signal prediction: [http://zeus2.itb.cnr.it/cgi-bin/wwwHC\\_TATA.pl](http://zeus2.itb.cnr.it/cgi-bin/wwwHC_TATA.pl)。

两序列间保守区域转录因子结合部位在线预测软件为: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>。PMatch 1.0: [www.gene-regulation.com/cgi-bin/pub/programs/p-match/bin/p-match.cgi](http://www.gene-regulation.com/cgi-bin/pub/programs/p-match/bin/p-match.cgi)。Consite在线分析软件: <http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/>。

SNP的筛选及其潜在功能预测采用SNP潜在功能进行预测的在线软件。FastSNP: [http://fastsnp.ibms.sinica.edu.tw/pages/input\\_CandidateGeneSearch.jsp](http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp); F-SNP: <http://compbio.cs.queensu.ca/F-SNP>。

### 1.3 线粒体融合基因2启动子的确定

使用在线启动子分析软件Promoter 2.0,取其默认设置,进行启动子预测。Neural Network Promoter Prediction在线软件时对积分大于0.85者进行分析。PROSCAN对启动子的预测其Cutoff = 53.000。FirstEF对启动子的分析采用默认条件,对正义链和反义链均进行分析。

### 1.4 转录因子的预测

对Mfn2基因5'端上游的序列共5000个碱基对进行分析。CpG岛分析使用P-Match 1.0软件和CpG island预测软件(所用标准为CpG岛的大小>100, GC% >50.0%, 观察值/期望值>0.6)。TATA box使用在线软件分析。转录因子使用PROMO程序和TRANSFAC(8.3版本)数据。

1.5 两序列间保守区域转录因子结合部位分析

经 P-Match1.0 程序搜索 TRANSFAC 6.0 数据库,获得人 Mfn2 基因 5'端上游的序列共 5000 bp 后,先进行正负链转录因子结合部位分析,再用进化足迹法分析,最后利用 Consite 程序确定人和小鼠同源基因保守区域。

1.6 SNP 的筛选及其潜在功能预测

根据 dbSNP、Ensembl、TFSearch、PolyPhen、ESEfinder、RescueESE、Genebank 和 Swiss-Prot 数据库分析的结果,采用决策树(decision tree)方法对启动子区的 32 种 SNP 引起的功能改变进行定量评价和危险度分级,结合对启动子序列的分析及其 SNP 所在的位置,预测其可能的功能变化。所使用的参数均为默认值。

2 结 果

2.1 线粒体融合基因 2 的启动子序列

Mfn2 基因总长为 40335 bp,有 19 个外显子,18 个内含子,757 个氨基酸,cDNA 含 4685 个碱基。其转录起始位点自 5001 bp,5'端未编码区全长为 5000

bp。该基因的 1 ~ 362 bp 为氨酰羟化酶( procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1, PLOD1) 基因的编码序列( GeneID:5351 )。

2.2 启动子分析

2.2.1 Promoter 2.0 预测 Mfn2 基因上游存在着 5 个启动子。其中 4100 bp 处为高度可能的启动子,其它则为临界性预测(表 1)。

表 1. Promoter 2.0 软件预测的线粒体融合基因 2 的启动子  
Table 1. Promoters of Mfn2 predicted by the Promoter 2.0

启动子	位置	得分	可信度
1	1300 bp	0.730	临界性预测
2	2500 bp	0.596	临界性预测
3	3200 bp	0.623	临界性预测
4	4100 bp	1.113	高度可能预测
5	4800 bp	0.601	临界性预测

2.2.2 Neural Network 预测 Mfn2 基因上游同样有 5 个启动子。除具体位置分布略有差异外,与 Promoter 2.0 预测的结果基本一致(表 2)。

表 2. Neural Network 软件对线粒体融合基因 2 启动子的预测

Table 2. Promoters of Mfn2 predicted by the Neural Network Promoter Prediction

启动子	起始位点	结束位点	得分	启动子序列
1	1115 bp	1165 bp	0.95	TAATAATATGAAAAAATCAGCCGTTGTGTTGGCATGCGCCTGTAATCCC
2	2307 bp	2357 bp	0.85	ATTAAAAATACAAAAATAAGGTGGGCTTTATGGCACATGCCTGTAAACCC
3	3750 bp	3800 bp	0.89	TGTAGTCAGCATATAAAATAGTAATGGGCTAATAGCCCTGCACTGTGGCAG
4	4296 bp	4346 bp	0.89	GCCTCGCCCTCCTAAAAATGCTGGAATTACAGGCTTGAGCCACCGCGCCCG
5	4628 bp	4678 bp	0.91	TGCTGTGGACGTTAAATACAGCGGTGGATGTTAGAGACGCAGTTCAGTCA

2.2.3 PROSCAN 预测 PROSCAN 对 Mfn2 基因启动子的预测是依据该基因序列与真核生物 Pol II 启动子序列比较后的同源性得分进行排序。从表中可以看到与启动子结合的转录因子主要存在于 4950 ~ 5000 bp 之间(表 3)。

2.2.4 FirstEF 预测 应用 FirstEF 软件对启动子预测时发现启动子主要位于 4600 bp 之后,即 -400 bp 以内(表 4),这与上述三种软件所得到的结果基本一致。

2.3 转录因子结合位点分析

利用 BLAST 程序对小鼠和人的 Mfn2 基因转录起始点上游 5'端保守区域进行分析时发现,在 300 ~ 1100 bp 间和 3000 ~ 5000 bp 间存在着保守区域(图

1)。

使用 PROMO 程序和 TRANSFAC(8.3 版本)数据库,共发现了 979 个高得分的转录因子结合位点。最大得分为 100,最小为 85,Cutoff 值为 85。再使用 Consite 程序分析人和小鼠两序列间保守区域转录因子结合部位,结果发现人类在 380 ~ 400 bp、650 ~ 750 bp、1500 ~ 1700 bp 和 2900 ~ 3100 bp 处为保守区域(图 2)。由于 380 ~ 400 bp 和 650 ~ 750 bp 两处保守区距离 Mfn2 基因的转录起始点较远,而该序列在 1 ~ 362 bp 处为 PLOD1 的编码区,这两个保守区可能与 PLOD1 有关。因此,仅对 1500 ~ 1700 bp 和 2900 ~ 3100 bp 两处的转录因子进行分析。结果发现这两处分别有 30 个和 39 个转录因子结合位点

(表 5)。

表 3. PROSCAN 对线粒体融合基因 2 启动子的预测

Table 3. Promoters of Mfn2 predicted by the PROSCAN

转录因子	正负链	位置(bp)	权重
T-Ag	－	4999	1.086000
AP-2	＋	4980	1.091000
(Sp1)	＋	4977	2.233000
UCE. 2	－	4975	1.278000
UCE. 2	＋	4972	1.216000
(Sp1)	－	4963	4.589000
Sp1	－	4962	12.906000
JCV_repeated_sequenc	－	4961	1.427000
Sp1	－	4961	3.013000
Sp1	－	4960	3.191000
GCF	－	4959	2.361000
Sp1	＋	4956	3.061000
AP-2	＋	4956	1.672000
Sp1	＋	4955	3.119000
(Sp1)	＋	4954	6.819000
EARLY-SEQ1	＋	4954	5.795000
APRT-mouse_US	＋	4953	7.604000

表 4. FirstEF 软件对线粒体融合基因 2 启动子的预测

Table 4. Promoters of Mfn2 predicted by the FirstEF

正负链	启动子	<i>P</i> 值	CpG 岛位置	
正义链：1	4649..5218	0.9996	4877..5078	
	2	4645..5214	0.9824	4877..5078
反义链：1	5271..4702	0.8108	5159..4958	
	2	5478..4909	0.8032	5159..4958
	3	5551..4982	0.7702	5159..4958

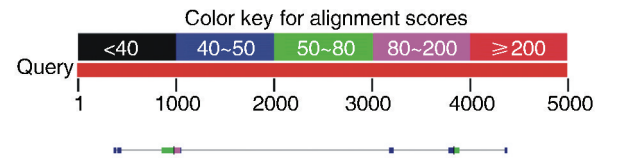


图 1. BLAST 程序对人线粒体融合基因 2 转录起始点上游 5' 端保守区域分析

Figure 1. BLAST analysis of 5' conservative region of human Mfn2

在对 Mfn2 基因进行 TATA box 分析时发现,该基因的启动子区存在着 15 个 TATA box。其中 8 个位于 -2000 bp 以内(表 6)。

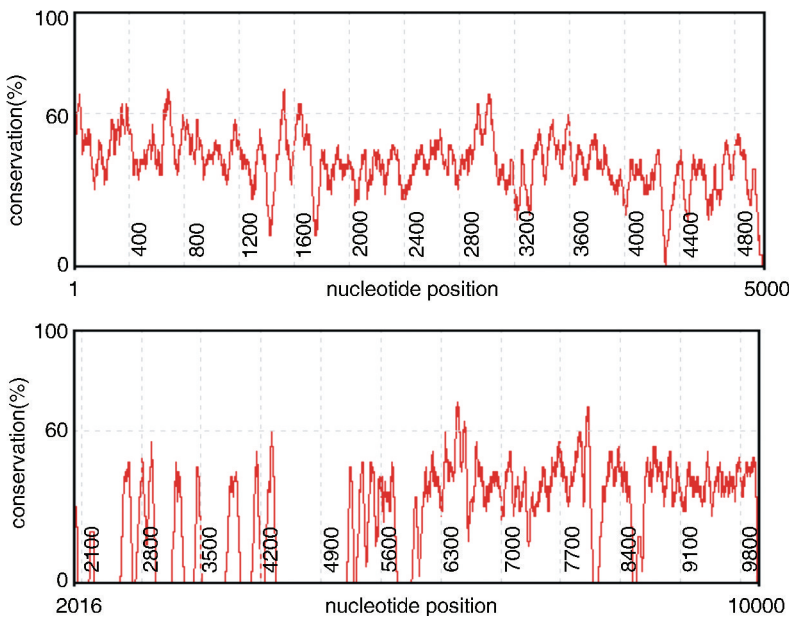


图 2. 人(上图)和小鼠(下图)线粒体融合基因 2 转录起始点上游 5'端保守区域分析

Figure 2. Analysis of 5' conservative region of human (above) and mouse (below) Mfn2 by the Consite

## 2.4 CpG 岛分析

使用 CpG Island Searcher 软件对 CpG 岛进行预测时发现,Mfn2 基因有一个 CpG 岛位于启动子区

即 -4551 ~ +871 bp 间[(C + G)% = 62.9% ;Obs/Exp =0.727,大小为 1321 bp][标准为(C + G)% = 55% ;Obs/Exp = 0.65 , Length > 200 bp)].且横跨



转录起始位点和第一启动子,这符合启动子和 CpG 的分布特点(图 3)。

表 5. 线粒体融合基因 2 5'端保守区域转录因子结合位点

Table 5. Transcript factor binding site analysis of 5' conservative region of Mfn2

编号	准入号	转录因子	得分	编号	准入号	转录因子	得分	编号	准入号	转录因子	得分
1	M00148	SRY	100	24	M00142	NIT2	100	47	M00184	MyoD	85. 1
2	M00148	SRY	100	25	M00008	Sp1	89	48	M00141	Lyf-1	100
3	M00048	ADR1	93. 8	26	M00271	AML-1a	88. 7	49	M00253	cap	96. 2
4	M00048	ADR1	93. 8	27	M00109	C/EBPb	86. 2	50	M00073	deltaE	92. 6
5	M00253	cap	92. 6	28	M00253	cap	86	51	M00220	SREBP-	91
6	M00160	SRY	87. 9	29	M00101	CdxA	85. 7	52	M00048	ADR1	89. 2
7	M00077	GATA-3	87. 2	30	M00148	SRY	85. 5	53	M00271	AML-1 a	88. 7
8	M00100	CdxA	87. 2	31	M00253	cap	99. 5	54	M00111	CF1 /	88. 2
9	M00075	GATA-1	86. 1	32	M00048	ADR1	93. 8	55	M00041	CRE-BP	87. 5
10	M00160	SRY	85. 9	33	M00140	Bcd	93. 7	56	M00039	CREB	86. 6
11	M00091	BR-C Z	85. 7	34	M00019	Dfd	89. 4	57	M00041	CRE-BP	86. 2
12	M00029	HSF	95. 4	35	M00087	Ik-2	89	58	M00253	cap	86. 2
13	M00028	HSF	94. 3	36	M00141	Lyf-1	100	59	M00048	ADR1	86. 2
14	M00029	HSF	86. 3	37	M00140	Bcd	93. 7	60	M00039	CREB	85. 6
15	M00048	ADR1	86. 2	38	M00253	cap	90. 4	61	M00033	p300	85. 1
16	M00029	HSF	95. 4	39	M00019	Dfd	89. 3	62	M00048	ADR1	100
17	M00028	HSF	94. 3	40	M00008	Sp1	89	63	M00077	GATA-3	92. 2
18	M00140	Bcd	93. 7	41	M00073	deltaE	89	64	M00075	GATA-1	90. 6
19	M00253	cap	91. 9	42	M00253	cap	88. 4	65	M00271	AML-1 a	87. 4
20	M00019	Dfd	89. 4	43	M00087	Ik-2	87. 3	66	M00100	CdxA	87. 2
21	M00008	Sp1	89	44	M00060	Sn	87. 1	67	M00076	GATA-2	86. 2
22	M00087	Ik-2	89	45	M00029	HSF	86. 9	68	M00101	CdxA	85. 7
23	M00037	NF-E2	87. 7	46	M00028	HSF	85. 9	69	M00083	MZF1	100

注:1~30 为 1500~1700 bp 保守区的转录因子,31~69 为 2900~3100 bp 保守区的转录因子。

表 6. 线粒体融合基因 2 转录起始点上游 5'端的 TATA box 分析

Table 6. TATA box analysis of 5' conservative region of Mfn2

编号	位置	结合序列	编号	位置	结合序列	编号	位置	结合序列
1	337	TCTATTAAAG	6	2815	AATTTAGGGA	11	3561	GCAATAAATT
2	1681	GATATGAGGT	7	2897	GCTAAAAAGG	12	3588	GTTATAGAGC
3	2682	AGTATGCAGA	8	3232	TTTATAGTTC	13	3711	TGTATTTAAC
4	2694	TGTTTCAGAG	9	3520	TCTATATCAG	14	3761	CATATAAATA
5	2701	GAGATAAAGC	10	3522	TATATCAGAG	15	4364	CCTAAAAATT

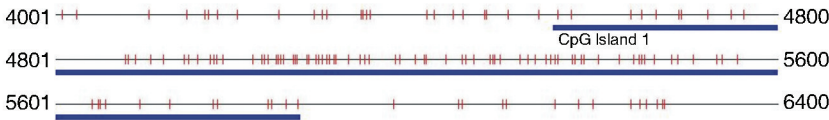


图 3. 线粒体融合基因 2 在转录起始点上游 5'端 CpG 岛预测 图中蓝色区域为 CpG 岛所在的位置。

Figure 3. CpG island prediction of 5'conservative region of Mfn2

使用 EMBL 提供的软件[ 标准为 Obs/Exp ratio >0. 60, ( C + G ) % > 50. 00% , Length > 200 bp ] 在 Mfn2 基因 5'端转录起始点上游发现一个 CpG 岛。位于 - 4865 ~ 493 bp 之间,大小为 629 bp, ( C +

G)% = 71.86% , Obs/Exp = 0.85。与 CpG Island Searcher 所得到的结果是一致的(图4)。

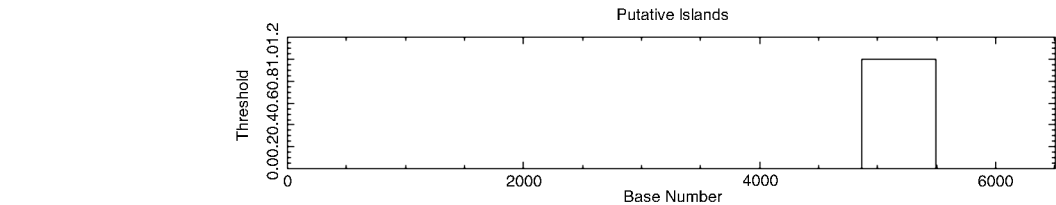


图4. 线粒体融合基因2在5'端转录起始点上游 CpG 岛预测 柱状图为预测的 CpG 岛所在位置。

Figure 4. CpG island prediction of 5'conservative region of Mfn2

2.5 SNP 及其功能预测

Mfn2 基因在5'端转录起始点上游存在着32个多态性位点。其中12个SNP位于启动子区,5个处于-2000 bp 以内,可能对基因转录具有调节功能。另外还有20个SNP位于5'端,但是其功能还不清楚。表7和表8分别为FastSNP和F-SNP的分析结果。

表7. FastSNP对线粒体融合基因2 5'端转录起始点上游 SNP 及其功能预测

Table 7. SNPs and function prediction of 5' conservative region of Mfn2 by FastSNP

SNP ID( rs )	功能	Lower Risk	Upper Risk	区域	SNP 位置
rs1049519	Promoter/regulatory region	1	3	5upstream	-4905
rs12088157	Promoter/regulatory region	1	3	5upstream	-4677
rs12061195	Promoter/regulatory region	1	3	5upstream	-4676
rs12024161	Promoter/regulatory region	1	3	5upstream	-4617
rs6698008	Promoter/regulatory region	1	3	5upstream	-3451
rs11588109	Promoter/regulatory region	1	3	5upstream	-3126
rs34061017	Promoter/regulatory region	1	3	5upstream	-2647
rs61776499	Promoter/regulatory region	1	3	5upstream	-1944
rs3753579	Promoter/regulatory region	1	3	5upstream	-1778
rs3820189	Promoter/regulatory region	1	3	5upstream	-1379
rs12758442	Promoter/regulatory region	1	3	5upstream	-1040
rs12743901	Promoter/regulatory region	1	3	5upstream	-1039
rs1063393	Upstream with no known function	0	0	5upstream	-4957
rs15431	Upstream with no known function	0	0	5upstream	-4890
rs1804187	Upstream with no known function	0	0	5upstream	-4834
rs34657615	Upstream with no known function	0	0	5upstream	-4687
rs11587948	Upstream with no known function	0	0	5upstream	-4348
rs61776496	Upstream with no known function	0	0	5upstream	-4089
rs61776497	Upstream with no known function	0	0	5upstream	-3561
rs4846081	Upstream with no known function	0	0	5upstream	-3528
rs34030040	Upstream with no known function	0	0	5upstream	-2767
rs2336382	Upstream with no known function	0	0	5upstream	-2736
rs5772482	Upstream with no known function	0	0	5upstream	-2659
rs35156669	Upstream with no known function	0	0	5upstream	-2656
rs3047174	Upstream with no known function	0	0	5upstream	-2648
rs7551289	Upstream with no known function	0	0	5upstream	-2449
rs7516628	Upstream with no known function	0	0	5upstream	-2313
rs61776498	Upstream with no known function	0	0	5upstream	-2310
rs34712694	Upstream with no known function	0	0	5upstream	-1050
rs57696082	Upstream with no known function	0	0	5upstream	-831
rs2236053	Upstream with no known function	0	0	5upstream	-166
rs2236054	Upstream with no known function	0	0	5upstream	-45

表 8. F-SNP 对线粒体融合基因 2 5'端转录起始点上游 SNP 及其功能预测

Table 8. SNP and function prediction of 5' conservative region of Mfn2 by F-SNP

SNP	功能	位置
rs3753579	+	5upstream
rs3820189	+	5upstream
rs3047222	?	5upstream
rs12758422	+	5upstream
rs12743901	+	5upstream
rs2236053	-	5upstream
rs2236054	-	5upstream

+ 表示有功能改变, - 表示没有功能改变, ? 表示无预测结果。

3 讨 论

EH 发病时出现的 VSMC 增殖是动脉粥样硬化的基本和主要的病理变化,因而如何防止 VSMC 增殖是临床治疗高血压时必须要考虑的问题。长期以来,虽然在防治增殖性心血管疾病方面取得了一些进展,但是还没有完全解决这个问题。Mfn2 基因是一个近来发现的功能比较强大的基因,对 VSMC 增殖的负调控功能使得它从发现之初就受到高度重视。我们的前期研究发现在 EH 患者 VSMC 中,该基因低表达<sup>[11]</sup>,但是其低表达的机制还不清楚。因此,对于该基因的启动子研究是基因表达调控机制研究的前提。

在本研究所用的四种启动子分析软件中, Promoter 2.0 是根据对启动子区各种转录信号(如 TA-TA box、CCAAT box 等)及信号间保守的空间排列顺序的识别,用神经网络方法进行预测。Promoter SCAN 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与 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。Neural Network Promoter Prediction 综合了以上二种软件的一些分析方法,特异性较高。但上述程序预测的假阳性率较高,如 Promoter 2.0 每 23 kb 出现一个假阳性, Promoter SCAN 平均每 19 kb 出现一个假阳性,而 Neural Network Promoter Prediction 的假阳性率约为 61%。FirstEF 是搜索通过 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,确定启动子区。这种方法使预测的敏感性和特异性都明显提

高(均高于 90%)。但是对于不含 CpG 岛的启动子,其预测的精确性相对较低。本研究通过对启动子的分析发现该基因可能存在着 5 个启动子,尤其 -350 bp 左右和 -2000 bp 以内,是两个非常重要的启动子区。所应用的四种软件的分析结果都非常相似地显示了这种结果。这提示核心启动子就在此范围内。与公认的启动子的位置一般在 2000 bp 以内是一致的<sup>[12]</sup>。

在对人和小鼠的 Mfn2 基因上游 5'端的保守区域进行比较分析时我们发现,在人 Mfn2 基因 -2000 bp 处存在着高度保守的区域。对人和小鼠 Mfn2 基因上游 5'端进行同源性检索时发现 -1000 bp 左右和 -2000 bp 两个区域相似性较大,同源性最多可达到 93%。因此,综合结果表明 -2000 bp 以内这个区域可能是基因表达调控的关键所在。同时, -4000 bp 以外存在着保守区域,人和鼠的相似性也比较高,但是由于该序列还编码 PLOD1 基因,因此,对于 Mfn2 基因而言可能并不是很重要。

对 1500 ~1700 bp 和 2900 ~3100 bp 两处比较保守的区域进行转录因子结合位点分析时发现,这两处有大量的转录因子结合位点,分别有 30 个和 39 个转录因子结合位点。尤其是一些与已知的 VSMC 增殖的发病机制,如炎症反应、氧化应激、VSMC 增殖与分化和 MAPK、STAT 信号传导通路等相关的转录因子(如 Gata-3、HSF、SP-1、SN、MYOD-1 以及 CREB 等)等的发现<sup>[13-16]</sup>,更提示了这两个区域的重要性和生物信息学预测与实验结果的一致性。TATA box 的分析预测结果显示 15 个 TATA box 中,有 10 个在 2900 ~3100 bp 保守区以内,而 1500 ~1700 bp 区内只有 2 个。由于 TATA box 对转录起始具有重要作用,这表明 2900 ~3100 bp 保守区的重要性。

由于 CpG 岛对于基因的转录有重要的调控作用。因此,对于转录调控区的 CpG 岛研究也是非常必要的。本研究在使用 First EF 软件对 CpG 岛预测时,发现了一个 CpG 岛,位于 Mfn2 基因上游 5'端非编码区序列 -400 bp 以内,并且处于启动子所在的范围内,体现了其功能与空间位置的一致性。这与使用 CpG Island Searcher 和 EMBL 提供的软件分析得到的结果基本一致。

SNP 是第三代遗传标记,是现代分子遗传学的重要内容。我们在前期研究中发现了一些重要的 SNP 与 EH 相关<sup>[8]</sup>。而 Mfn2 基因在人类 EH 患者 VSMC 中的表达特点是转录下调。因此,研究 SNP 与转录调控的关系也是我们研究 Mfn2 基因低表达

机制的一个重要内容。本研究通过对上游 5'端未编码区 32 个 SNP 的分析发现,一些 SNP 位于启动子区, -2000 bp 以内的有 5 个处于转录调控区,因而可能会影响启动子的功能,其机制可能是顺式作用元件碱基的改变影响了转录因子与顺式作用元件之间的亲和力,从而影响转录因子发挥正常的转录调节功能<sup>[17]</sup>。另外还有 4 个位于 -1050 bp 以内,但是是否会影响启动子的功能还不清楚。其它可能会影响启动子功能与功能未知的 23 个 SNP 则分布于其它区域。由于 EH 是一种遗传率相当高的疾病,结合启动子与转录因子等的分析结果,SNP 的分析和功能预测对于下一步转录调控和分子遗传机制研究具有重要的指导作用。

综上所述,Mfn2 基因上游 5'端非编码区序列 -2000 bp,尤其是 -400 bp 左右可能存在着核心启动子。同时,启动子区存在着一些重要的与 Mfn2 功能密切相关的转录因子。因此,对 Mfn2 基因转录调控的研究应该以此为切入点。另外,启动子区存在的 SNP 可能会影响转录因子正常功能的发挥,这提示在涉及基因表达调控异常的遗传相关疾病的研究中,对启动子区 SNP 研究具有颇为重要的意义。

#### [参考文献]

- [1] Chen KH, Guo X, Ma D, et al. Dysregulation of HSG triggers vascular proliferative disorders [J]. *Nat Cell Biol*, 2004, 6: 872-883.
- [2] Xia Y, Wu Y, He X, et al. Effects of mitofusin-2 gene on cell proliferation and chemotherapy sensitivity of MCF-7 [J]. *J Huazhong Univ Sci Technolog Med Sci*, 2008, 28: 185-189.
- [3] Guo X, Chen KH, Guo Y, et al. Mitofusin 2 triggers vascular smooth muscle cell apoptosis via mitochondrial death pathway[J]. *Circ Res*, 2007, 101: 113-122.
- [4] de Brito OM, Scorrano L. Mitofusin 2: a mitochondria-shaping protein with signaling roles beyond fusion [J]. *Antioxid Redox Signal*, 2008, 10: 621-633.
- [5] Herzig S, Martinou JC. Mitochondrial dynamics: to be in good shape to survive [J]. *Curr Mol Med*, 2008, 8: 131-137.
- [6] Liesa M, Palacín M, Zorzano A. Mitochondrial dynamics in mammalian health and disease [J]. *Physiol Rev*, 2009,

89: 799-845.

- [7] 王佐广, 温绍君, 吴兆苏. 氧化应激与高血压的相关性研究进展[J]. *心血管病学进展*, 2009, 30: 581-584.
- [8] Wang Z, Liu Y, Liu J, et al. HSG/Mfn2 gene polymorphism and essential hypertension: a case-control association study in Chinese [J]. *J Atheroscler Thromb*, 2011, 18: 24-31.
- [9] 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: 826-828.
- [10] 张万里, 李伟, 吴川清, 等. 小鼠 EDNRB 基因启动子生物信息学分析[J]. *生物信息学*, 2011, 9: 72-76.
- [11] 刘亚萍, 温绍君, 刘雅, 等. 增殖抑制基因在原发性高血压患者血管平滑肌细胞增殖中的作用 [J]. *中华心血管病杂志*, 2007, 35: 914-918.
- [12] Sanecka A, Ansems M, van Hout-Kuijer MA, et al. Analysis of genes regulated by the transcription factor LUMAN identifies ApoA4 as a target gene in dendritic cells [J]. *Mol Immunol*, 2012, 50: 66-73.
- [13] Raymond M, Van VQ, Wakahara K, et al. Lung dendritic cells induce T(H)17 cells that produce T(H)2 cytokines, express GATA-3, and promote airway inflammation [J]. *J Allergy Clin Immunol*, 2011, 128: 192-201.
- [14] Chan J, Prado-Lourenco L, Khachigian LM, et al. TRAIL promotes VSMC proliferation and neointima formation in a FGF-2-, Sp1 phosphorylation-, and NF-kappaB-dependent manner [J]. *Circ Res*, 2010, 106: 1061-071.
- [15] Jo C, Cho SJ, Jo SA. Mitogen-activated protein kinase kinase 1 (MEK1) stabilizes MyoD through direct phosphorylation at tyrosine 156 during myogenic differentiation [J]. *J Biol Chem*, 2011, 286: 18903-913.
- [16] Zhou B, Liu Y, Kahn M, et al. Interactions between  $\beta$ -Catenin and transforming growth factor- $\beta$  signaling pathways mediate epithelial-mesenchymal transition and are dependent on the transcriptional co-activator cAMP-response element-binding protein (CREB)-binding protein (CBP) [J]. *J Biol Chem*, 2012, 287: 7026-038.
- [17] Gan-Or Z, Bar-Shira A, Dahary D, et al. Association of sequence alterations in the putative promoter of RAB7L1 with a reduced parkinson disease risk [J]. *Arch Neurol*, 2012, 69: 105-110.

(此文编辑 许雪梅)