

人 NPCEDRG 基因启动子的克隆及 CCAAT/NFY 结合位点初步分析*

侯德富^{1, 2, 3)**} 关勇军^{2)**} 关瑞^{1, 2)} 欧阳咏梅²⁾ 余艳辉²⁾ 陈主初^{1, 2)***}

⁽¹⁾中南大学湘雅医院卫生部肿瘤蛋白质组学重点实验室, 长沙 410008; ⁽²⁾中南大学肿瘤研究所, 长沙 410008;

⁽³⁾湖南师范大学医学院生物化学与分子生物学教研室, 长沙 410013)

摘要 NPCEDRG 基因是采用基因定位候选克隆策略获得的一个鼻咽癌候选抑瘤基因。NPCEDRG 在鼻咽癌细胞和组织中表达下调, 重新恢复 NPCEDRG 基因在 CNE2 细胞系的表达, 可部分逆转 CNE2 的恶性表型。为揭示 NPCEDRG 基因在鼻咽癌细胞和组织中表达下调的分子机制, 联合应用生物信息学和报告基因载体系统分析方法对 NPCEDRG 基因启动子区进行克隆及功能分析, 系统发育进化足迹分析结果表明, NPCEDRG 基因 5'端调控区-180~+235 bp 区间在脊椎动物中高度保守, 该保守区域中存在包括 CCAAT/NFY、STAT1 和 SP1 等转录因子结合位点。构建 Luc 和 / 或 EGFP 报告基因表达载体并检测其启动子活性, -146~-8 bp 区域有较强的启动子活性, 电泳迁移阻滞分析实验(EMSA)提示, CCAAT/NFY 转录因子结合位点是 NPCEDRG 基因的转录调控元件。因此, 研究确定-146~-8 bp 区域是 NPCEDRG 基因核心启动子区域且启动子核心元件 CCAAT/NFY 可能参与 NPCEDRG 基因的转录调控。

关键词 NPCEDRG 基因, 核心启动子, 转录调控, CCAAT/NFY 结合位点

学科分类号 Q71, R73

DOI: 10.3724/SP.J.1206.2010.00673

鼻咽癌(nasopharyngeal carcinoma, NPC)是一种具有明显种族和地区分布特征的上皮源性头颈部恶性肿瘤^[1]。其病因涉及到遗传因素、EB 病毒的感染、饮食和某些环境中的化学致癌物的作用^[2-4]等。鼻咽癌的发生是一个多因素多步骤的病理过程, 已发现数十个与鼻咽癌相关的基因, 其中抑瘤基因的失活更为普遍, 且发挥更重要的作用。研究发现位于染色体 3p21.3 区的 RASSF1A^[5]、BLU^[6]等候选抑瘤基因与鼻咽癌发生密切相关。中南大学肿瘤研究所对湖南鼻咽癌高发家系进行连锁分析, 发现染色体 3p21 部分区域与湖南家族性鼻咽癌发病紧密连锁^[2, 7]。有研究证实在我国南方, 正常鼻咽上皮和鼻咽癌前病变以及鼻咽癌组织存在 3p 染色体(包括 3p21.3)丢失^[8], 有力地证明染色体 3p21.3 区抑瘤基因的功能失活为鼻咽癌发生发展中的早期事件。

NPCEDRG 基因(GenBank No. AF538150)是中南大学卫生部肿瘤蛋白质组学重点实验室采用基因

定位候选克隆策略获得的一个在正常鼻咽和鼻咽癌组织表达差异具有显著性的基因, 定位于染色体 3p21.3, 该点存在于湖南家族性鼻咽癌的遗传易感区 3p21.31~21.2 区域内^[9-10]。我们前期研究结果表明, NPCEDRG 在部分鼻咽癌细胞系、鼻咽癌组织中表达下调, 过表达 NPCEDRG 基因可抑制鼻咽癌细胞增殖和细胞周期进程, 并部分逆转 CNE2 的恶性表型, 提示 NPCEDRG 基因可能为一个鼻咽癌候选抑瘤基因^[10-11]。为揭示 NPCEDRG 基因在鼻咽癌细胞和组织中表达下调的分子机制, 更深入地阐明 NPCEDRG 基因的生物学功能, 本研究应用

* 国家自然科学基金(30772401)和湖南省卫生厅科研基金(B2007006)资助项目。

** 共同第一作者。

*** 通讯联系人。

Tel: 0731-82355116, Fax: 0731-84327321, E-mail: tcbl@xysm.net

收稿日期: 2010-12-22, 接受日期: 2011-04-12

生物信息学分析 NPCEDRG 基因核心启动子区域及转录因子结合位点, 并利用报告基因、转染技术对该基因核心启动子区域进行初步鉴定, 应用电泳迁移阻滞分析实验 (electrophoretic mobility shift assay, EMSA) 分析其转录调控元件 CAAT 盒, 以期明确 NPCEDRG 基因在鼻咽癌发生机制中的功能提供一些线索.

1 材料和方法

1.1 材料

1.1.1 细胞. CNE1 为鼻咽高分化鳞状细胞癌细胞系, CNE2、HNE2、HNE3 为鼻咽低分化鳞状细胞癌细胞系, 人乳腺癌细胞系 MCF7、人肺腺癌细胞系 A549 和人宫颈癌细胞系 HeLa 均为中南大学肿瘤研究所保存, 于 RPMI 1640+10% 新生小牛血清 (newborn calf serum, NCS) 的培养基中, 37°C、5% CO₂、饱和湿度条件下培养.

1.1.2 载体. pGEM-T Easy、pGL3-Enhancer、pGL3-Control 购自 Promega 公司; pEGFP-c2 载体购自 BD Biosciences 公司.

1.1.3 试剂. Lipofectamine™ 2000 脂质体购自 Invitrogen 公司; M-MLV 逆转录试剂盒、核酸分子质量标准 DL2000 Marker、质粒抽提试剂盒、胶回收试剂盒购自大连宝生物公司; Wizard® Genomic DNA Purification Kit、高保真 DNA 聚合酶 (Advantage® 2 PCR Enzyme System)、双荧光素酶检测试剂盒 (Dual-Luciferase® Reporter Assay System)、*Nco* I、*Xba* I、*Mlu* I 和 *Bgl* II 购自 Promega 公司; DNA-free™ Kit 购自 Ambion 公司; 鼠抗人 GFP 抗体、鼠抗人 β -Actin 抗体、辣根过氧化物酶标记羊抗鼠抗体购自 SantaCruz 公司; BCA Protein Assay Kit、ECL 试剂盒、核蛋白抽提试剂盒 (NE-PER® Nuclear and Cytoplasmic Extraction Reagents)、Halt™ Protease Inhibitor Cocktail Kit、EMSA 试剂盒 (LightShift® Chemiluminescent EMSA Kit) 购自 Pierce 公司.

1.2 方法

1.2.1 RT-PCR 检测 NPCEDRG 基因表达. 按 TRIZOL™ 试剂操作程序进行总 RNA 的抽提, rDNase I 消化 Total RNA 中残留 gDNA, 于 1.0% 琼脂糖凝胶电泳确定 RNA 的质量和完整性, 用紫外分光光度仪检测 RNA 的浓度和纯度. 采用 TAKARA 公司 M-MLV 逆转录酶进行 cDNA 第一链的合成. 取 2 μ l cDNA 模板于 50 μ l 反应体系进行

PCR 扩增, 扩增条件: 预变性, 94°C 5 min; 94°C 30 s, 55°C 30 s, 72°C 45 s, 37(NPCEDRG) 或 28 个(GAPDH)循环; 72°C 10min. 取 9 μ l 扩增产物用 2% 琼脂糖凝胶电泳检测. 回收 PCR 产物克隆至 pMD20-T 载体并测序. NPCEDRG 引物序列 F367(5' ATATCGGAGCTACGCCTGAT 3') 和 R663 (5' AGATGCCTCAAATCCCACAG 3'), 目的片段长度 297 bp; 内对照 GAPDH 引物 F-GAPDH (5' TCGGAGTCAACGGATTTGGT 3') 和 R-GAPDH (5' TGGAATTTGCCATGGGTGGA 3'), 目的片段长度 105 bp.

1.2.2 NPCEDRG 基因 5' 端调控区生物信息学分析. 以 NPCEDRG 基因翻译起始密码子 ATG 的第一个碱基 A 为 +1, 选取 5' 端上游 3 500 bp 至下游 500 bp 的 gDNA 序列, 以此 4 000 bp 片段作为 CpG 岛分析、转录起始位点分析、启动子分析及转录因子结合位点分析用序列. 利用在线软件 DGSF (http://sdmc.lit.org.sg/promoter/dragonGSF1_0/genestart.htm)、DGCPF (http://sdmc.lit.org.sg/promoter/CGrich1_0/CGRICH.htm)^[12] 及 Genomatix (<http://www.genomatix.de/cgi-bin/eldorado/main>) 中 PromoterInspector 和 Gene2Promoter 两个程序等在线软件预测 NPCEDRG 基因候选启动子区域及转录起始位点^[13]; 应用 cpgplot (<http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html>) 分析 NPCEDRG 基因的 CpG 岛^[14]. 进一步查询 DBTSS (<http://dbtss.hgc.jp/>) 获取 NPCEDRG 基因转录起始位点数据信息^[15]. 应用 ECR browser (<http://ecrbrowser.dcode.org/>) 等在线程序分析与预测 NPCEDRG 基因调控区转录因子结合位点^[16].

1.2.3 NPCEDRG 启动子区 Luc 及 pEGFP 报告基因载体构建及鉴定.

按 Promega 基因组 DNA 抽提试剂盒提取正常人全血 gDNA, 应用 Primer 5.0 针对 NPCEDRG 基因 5' 端调控区序列设计 2 条正向引物及 1 条共有反向引物, 正向引物的 5' 端引入 *Mlu* I 酶切位点 (下划线部分) 和保护性碱基, 反向引物的 5' 端引入 *Bgl* II 酶切位点 (下划线部分) 和保护性碱基. 其引物序列分别为: NPCEDRG-For(-625), 5' cgagcgt-ACAGACTACCGCCCTAAGCA 3', NPCEDRG-For(-146) 5' cgagcgtATCCTGAGCTTCGGTGATTG 3', NPCEDRG-Rev(-8), 5' gaagatctCAGCCGCAACT-AAGTGCAAC 3', 扩增片段大小分别为 617 bp (-625 bp 至 -8 bp 区间) 和 138 bp (-146 bp 至 -8 bp

区间)。利用高保真 DNA 聚合酶 Advantage[®] 2 Polymerase 对上述 2 个片段进行 PCR 扩增, 扩增产物加“A”尾后, 重组至 pGEM-T 载体上, 经酶切和测序证实后用 *Mlu* I 和 *Bgl* II 双酶切, 胶回收目的酶切片段, 连接到同样经 *Mlu* I 和 *Bgl* II 双酶切制备好的 pGL3-Enhancer 线性化载体上, 构建含 NPCEDRG 基因 5'端调控序列和萤火虫荧光素酶报告基因的重组质粒, 重组质粒经酶切鉴定并测序。

以 EGFP cDNA 为模板设计引物对, 在下游引物 5'端加上 *Xba* I 酶切位点(下划线部分)和保护性碱基, 序列如下: EGFP-For, 5' GACTTTCCAAA-ATGTCGTAACAACCTCC 3', EGFP-Rev, 5' tgc-tctagaTTACTTGTACAGCTCGTC 3'。利用 EGFP 基因编码区引物, 以载体 pEGFP-C2 为模板, PCR 扩增 EGFP 基因全长编码区序列, 纯化 PCR 产物, 并用 *Nco* I 和 *Xba* I 双酶切 PCR 产物, 胶回收目的酶切片段, 连接至以 *Nco* I 和 *Xba* I 双酶切切除其下游荧光素酶 (luciferase) 基因而线性化的 NPCEDRG 启动子介导的报告载体 pGL3-Enhancer、pGL3-en617 和 pGL3-en138。转化至 JM109 感受态细胞中, 获得的重组质粒分别命名为 pEGFP-Enhancer、pEGFP-en617 和 pEGFP-en138, 测序鉴定重组质粒。

1.2.4 NPCEDRG 基因 5'端调控区域的功能分析。 选用 NPCEDRG 基因表达相对较高的 MCF-7、HeLa、CNE2 细胞作为靶细胞进行后续研究。MCF-7、HeLa、CNE2 细胞接种于 24 孔培养板中, 当细胞生长至 80%~90% 融合度时, 按 Lipofectamine 2000 脂质体说明书将 800 ng pGL3/Basic 重组质粒和 10 ng pHRL-SV40 共转染, 同时设立 pGL3-Enhancer 阴性对照组和 pGL3-Control 阳性对照组。转染 30 h 后, 按双荧光素酶检测试剂盒操作步骤进行荧光素酶报告基因活性检测, 弃培养液, 用预冷的 PBS 洗涤细胞 2 次, 每孔加入 100 μ l 1 \times PLB, 放置于摇床室温缓慢摇 15~20 min, 再反复冻融后将细胞裂解物转入 0.5 ml 离心管中, 短暂离心约 10 s, 取 20 μ l 细胞裂解物加入 100 μ l 荧光素酶底物(LAR II), 混匀后用单光子检测仪测定荧光素酶发光值, 再加入 100 μ l 反应终止液(Stop & Glo[®] Reagent), 混匀后用单光子检测仪测定内标海肾荧光素酶发光值, 两者比值即为荧光素酶的相对活性 RLA(relative luciferase activity)。

1.2.5 NPCEDRG 基因启动子区调控 EGFP 表达水平检测。 将 MCF7 细胞接种于 6 孔培养板, 待细

胞汇合 80%~90% 时, 按 Lipofectamine[™] 2000 脂质体说明书分别将报告质粒 pEGFP-Enhancer、pEGFP-en617 和 pEGFP-en138 转染至 MCF7 细胞, 其中以 pEGFP-Enhancer 作为阴性对照, 转染后 24 h, 荧光显微镜下直接观察和流式细胞仪分析转染细胞绿色荧光蛋白的活体表达水平, 并应用 Western blot 检测 EGFP 表达。

1.2.6 EMSA 初步鉴定 NPCEDRG 基因启动子区 CAAT/NFY 结合位点。

取对数生长期 MCF7 细胞, 用预冷 PBS 洗涤细胞 2 次, 离心收集细胞, 按照 NR-PER Nuclea and Cytoplasmic Extraction Reagents 试剂盒使用说明书抽提细胞核蛋白。在抽提试剂 CER I 和 NER 中分别加入蛋白酶抑制剂 Halt[™] Protease Inhibitor Cocktail, BCA 法测量核蛋白浓度, 于 -80 $^{\circ}$ C 保存备用。

系统发育进化足迹分析结果提示, NPCEDRG 启动子区 -141 bp 至 -116 bp 区域只有 CAAT/NFY 转录因子的结合位点, 截取该区域 26 bp 寡核苷酸序列作为探针, 并根据 rVista 软件采用置换突变的方法设计突变探针。探针序列为: Wt-CAAT/NFY, 5' GCCAGCCCAGCCAATCACCGAAGCTC 3', Mt-CAAT/NFY, 5' GCCAGCCCAGCGTATTCCGAAGCTC 3', 并对其 Wt-CCAAT/NFY 结合的 DNA 双链探针的 5'端进行生物素标记。生物素标记双链探针、野生型双链探针及突变型探针均由大连宝生物生物技术公司合成。按照 Pierce 公司提供的 EMSA 试剂盒操作程序进行, 20 μ l 总反应体系加入 20 μ g 核蛋白、20 fmol 生物素标记双链探针和 10 pmol 的野生型双链探针或突变型探针。为克服非特异结合反应, 加入冷探针后, 在室温反应 5 min 后再加入标记探针, 反应液放于室温孵育 20 min。上样于预电泳好的 6% 非变性聚丙烯酰胺凝胶中, 0.5 \times TBE 缓冲液, 采用 100V 电压、冰浴中电泳, 电泳结束后转膜及紫外交联, 化学发光检测生物素标记的 DNA。

2 结 果

2.1 不同细胞株中 NPCEDRG 基因的 mRNA 表达水平不同

RT-PCR 是一种半定量的 PCR, 采用以下措施保证其有效性: a. 消化干净总 RNA 中的痕量 DNA; b. 对 RNA 精确定量, 设定内对照 GAPDH; c. 在 GAPDH 与目的基因的平台期前结束 PCR。预实

验提示 NPCEDRG 基因以及 GAPDH 分别在 40 个及 30 个循环进入平台期. 本实验针对 NPCEDRG 基因和 GAPDH 分别选择 37 个和 28 个循环对其进行 PCR 扩增. 如图 1a 所示, 与 GAPDH 比较, NPCEDRG 基因在所检测的 7 个不同肿瘤细胞中均

为低表达, 且在 A549 和 HNE2 等细胞系中表达非常低, 产物片段为 297 bp. 将 RT-PCR 产物克隆至 pMD20-T 载体上, 测序结果表明, PCR 扩增获得的产物即为基因的目的序列, 见图 1b.

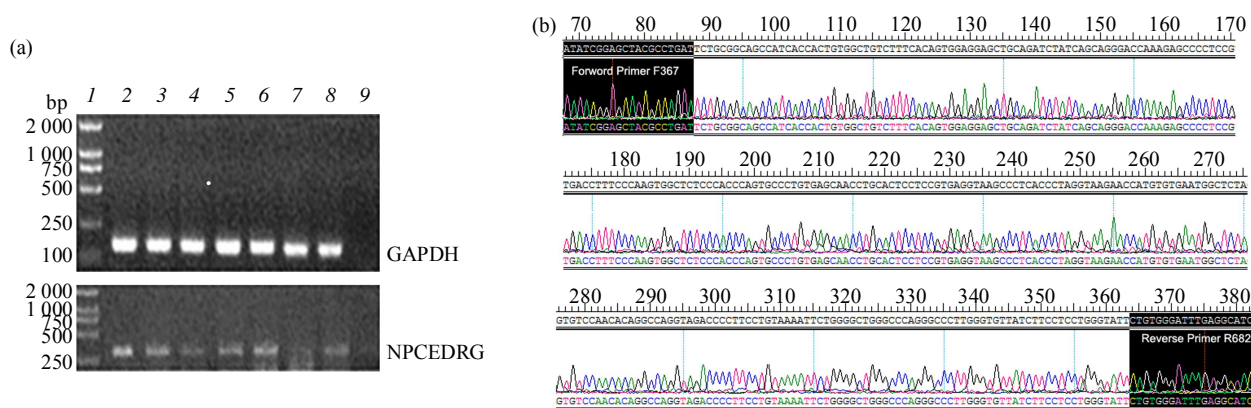


Fig. 1 Analysis of NPCEDRG expression in cultured cells

(a) RT-PCR analysis of NPCEDRG mRNA purified from cultured cell lines. GAPDH was also analyzed as a loading control. 1: DL2000; 2: MCF7; 3: HeLa; 4: A549; 5: CNE1; 6: CNE2; 7: HNE2; 8: HNE3; 9: Water. (b) Sequencing results of RT-PCR product of NPCEDRG gene.

2.2 NPCEDRG 基因 5'端上游调控区序列生物信息学分析结果

以 NPCEDRG 起始密码子 ATG 的第一个碱基 A 为+1, 提取上游 3 500 bp 及下游 500 bp 共 4 000 bp 的 gDNA 序列(其包含 NPCEDRG 的第一个外显子、5'端上游约 3 500 bp 的 gDNA 序列以及 283 bp 的第一个内含子序列)进行 CpG 岛、TSS 及启动子区域分析. 如图 2a 所示: Gene2promoter 软件分析 TSS 位于-73 bp 和-49 bp 处, 启动子介于-625~+74 bp; PromoterInspector未预测到任何启动子区; Dragon Gene Start Finder 预测 TSS 位于-7 bp 处, 启动子介于-764~+155 bp 区间; Dragon GC+Promoter Finder 预测 TSS 位于-62 bp 和-12 bp 处; CpGplot 和 Methprimer 分析该区域存在 2 个 CpG 岛, CpG 岛 1 位于-1 573~-960 bp, 长度为 614 bp; CpG 岛 2 位于-624~+265 bp, 长度为 889 bp 的 2 个 CpG 岛. DBTSS 数据库中 NPCEDRG 基因 TSSs 查询结果表明, 在多细胞混合文库中 NPCEDRG 基因包含多个 TSSs 主要位于-85~-9 bp 区域, 个别 EST 起始于 ATG 下游 +107 bp 处, 见图 2b. 应用 ECR Browser 软件系统对 NPCEDRG 基因-750~+250 bp 区域进行系统发育进化足迹分析, TBA 多序列比对分析结果表明

NPCEDRG 基因 5'端有约 415 bp 序列(-180~+235 bp 区间)在脊椎动物中高度保守, 该区域与黑猩猩、恒河猴、狗、牛及小鼠等高度匹配, 见图 2c; rVista 分析结果显示在保守区域中存在包括 CAAT/NFY、STAT1 和 SP1 等转录因子结合位点, 见图 2d. 综合各生物信息学分析结果, 提示 -180~+74 bp 区域可能包含 NPCEDRG 基因的核心启动子区域且启动子核心元件 CAAT/NFY 可能参与 NPCEDRG 基因的转录调控.

2.3 启动子片段 Luc 报告基因载体的构建及其活性检测

以正常人外周血细胞 gDNA 为模板, 利用高保真 DNA 聚合酶 Advantage[®] 2 Polymerase 分别扩增 617 (-625~-4 bp)和 138 (-146~-8 bp)等片段, 加“A”尾后, 重组至 pGEM-T 载体上, *Mul* I/*Bgl* II 双酶切及测序确证序列无突变后, 重组至 Luc 报告基因载体 pGL3-Enhancer 上, 获得 pGL3-en617 和 pGL3-en138 2 个 NPCEDRG 基因启动子报告基因载体, 重组报告基因载体经 *Mul* I/*Bgl* II 双酶切后, 1%琼脂糖凝胶电泳结果显示获得预期大小的片段(图 3a), 测序证实序列无突变(图 3b), 成功构建启动子片段 Luc 报告基因载体.

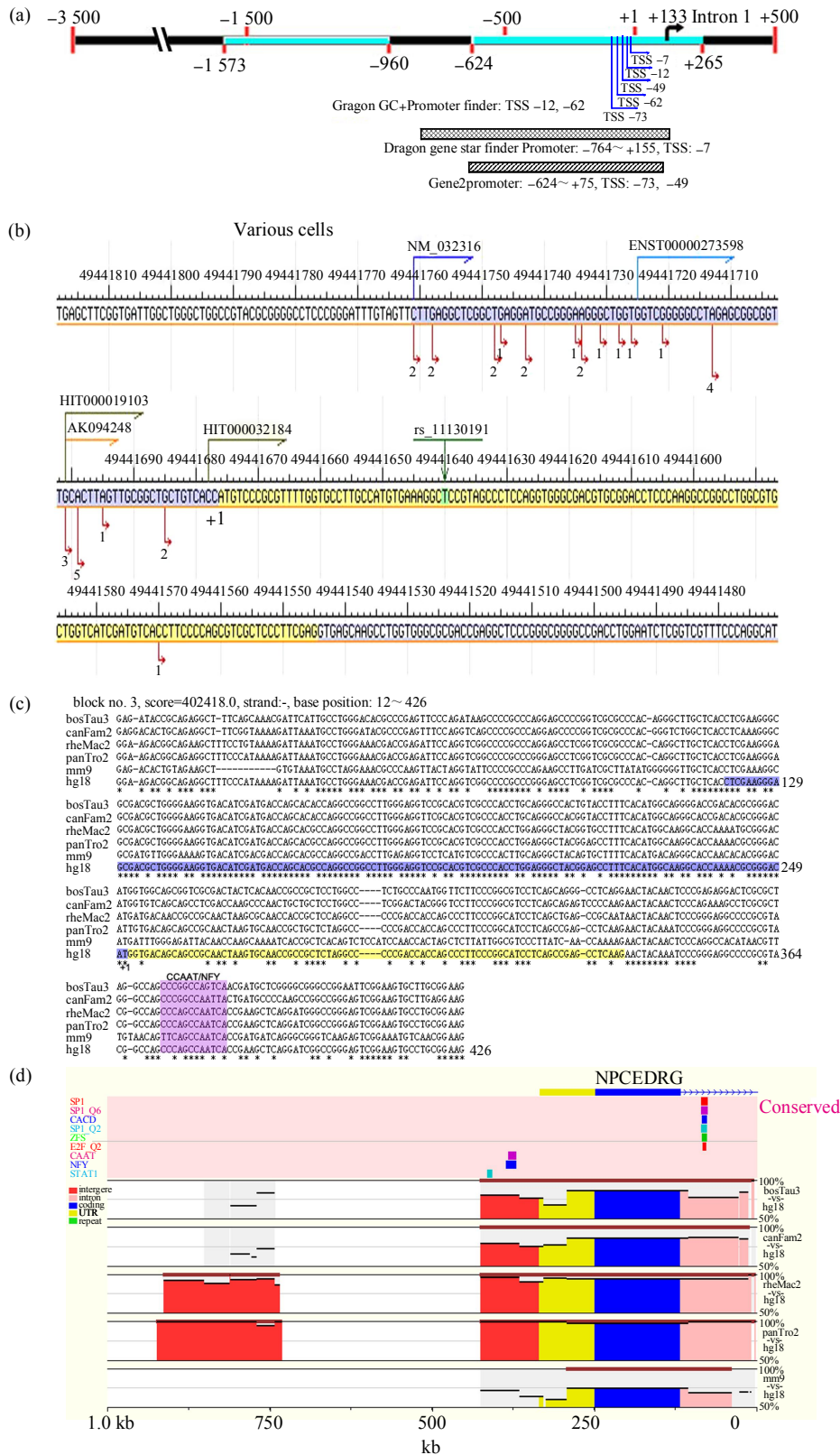


Fig. 2 Analysis of NPCEDRG promoter region by a combination of bioinformatics

(a) Schematic representation of the putative NPCEDRG promoter region, TSSs and CpG Island. The putative promoter regions by DGSF and Gene2promoter are shown as a hatched box, and a dot box, respectively. The putative TSSs are indicated with bent arrows. Two islands by cpplot are shown as bright green boxes. The translation start site will be referred to as +1 throughout the remainder of this report. (b) The TSSs of NPCEDRG from various cells by DBTSS, the number below the bent arrows represent the number of clones. (c) TBA BLAST result of the conserved region of NPCEDRG among human, chimpanzee, rhesus, dog, cow and mouse (yellow: 5' UTR, blue: code exon, of human NPCEDRG). (d) Schematic representation of the putative transcription factor binding sites (TFBS) by rVista. Dynamically overlay TFBS prediction with the conservation profile and perform clustering and 9 multi-conserved TFBSs were identified, such as CAAT/NFY, STAT1 and SP1.

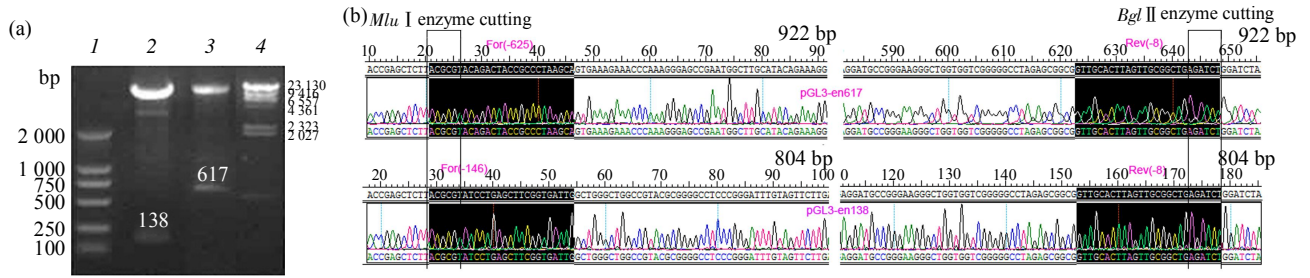


Fig. 3 Verification of luciferase reporter constructs by enzyme cutting and DNA sequencing

The NPCEDRG gene promoter regions were inserted into the luciferase reporter vector and verified by (a) enzyme cutting and (b) DNA sequencing. 1: DL2000; 2: pGL3-en138; 3: pGL3-en617; 4: λ -HindIII.

将人 NPCEDRG 基因启动子区 Luc 表达载体 pGL3-en617 和 pGL3-en138、空载体 pGL3-Enhancer、阳性对照 pGL3-Contral 质粒与内对照载体 phRL-SV40 分别共转染至 MCF7、HeLa 和 CNE2 细胞中, 用双荧光素酶检测试剂盒进行荧光素酶报告基因活性检测, 测定荧光素酶发光值和内标海洋腔肠荧光素酶发光值, 计算并分析各重组质粒在各细胞中荧光素酶的相对活性 RLA。如图 4 所示, 在 MCF7、HeLa 和 CNE2 细胞中, pGL3-en617 与 pGL3-en138 均有较强的启动子活性, 其中 pGL3-en617 启动子活性最高, 在 3 种细胞中分别为 SV40 启动子活性的 52.34%、41.42%及 33.46%, pGL3-en138(即-146~-8 bp)约为 SV40 启动子活性的 15.80%~21.12%, 提示-146~-8 bp 区间为 NPCEDRG 基因的核心启动子区域。

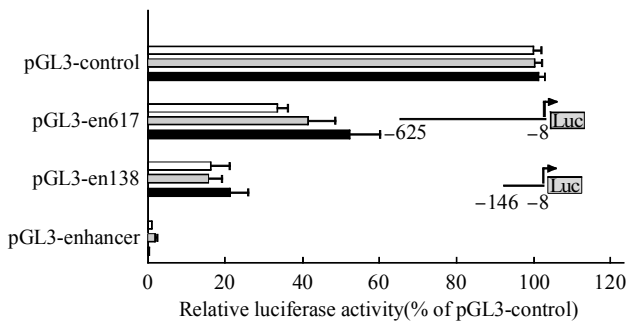


Fig. 4 Luciferase activity of human NPCEDRG gene promoter

Constructs of NPCEDRG promoter region were inserted into the luciferase reporter vector, followed by transfection into MCF7, HeLa and CNE2 cells; the phRL-SV40 vector, coding for Renilla luciferase, was co-transfected as an internal reference. The luciferase values are represented as luciferase activity relative to that of the pGL3-control vector in each cell line. The results are represented as $\bar{x} \pm s$ of at least three independent experiments. □: MCF7; ▤: HeLa; ■: CNE2.

2.4 启动子区 EGFP 报告基因载体构建及其活性检测

为证实 Luc 报告基因质粒所得到的结果的可靠性, 我们进一步构建 EGFP 报告基因载体. 以 pEGFP-c2 载体为模板, 利用高保真 DNA 聚合酶 Advantage[®] 2 Polymerase 扩增 EGFP 基因编码区 cDNA 片段, PCR 产物胶回收, Nco I 和 Xba I 双酶切, 胶回收后与 Nco I 和 Xba I 双酶切的 pGL3-en617、pGL3-en138 及 pGL3-Enhancer 等线性化载体连接、转化, 分别获得 pEGFP-en617、pEGFP-en138 及 pEGFP-Enhancer 等 EGFP 报告基因载体. 重组 EGFP 报告基因载体经 Nco I 和 Xba I 双酶切后 1%琼脂糖凝胶电泳, 均获得一个约 850 bp 大小的 EGFP 片段及分别约为 4.0 kb、3.5 kb 及 3.4 kb 的载体片段, 见图 5, 说明成功构建了启动子片段 EGFP 报告基因载体。

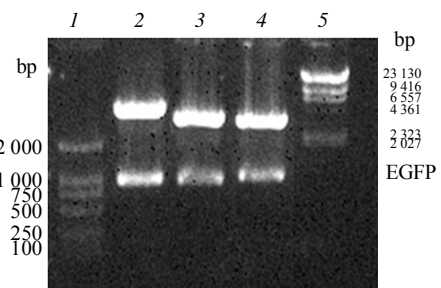


Fig. 5 Verification of EGFP constructs by enzyme cutting 1: DL2000; 2: pEGFP-en617; 3: pEGFP-en138; 4: pEGFP-enhancer; 5: λ -HindIII.

将 pEGFP-en617、pEGFP-en138 及 pEGFP-Enhancer 等 EGFP 报告基因载体转染至 Luc 报告基

因载体活性较高的 MCF7 细胞中, 荧光显微镜下直接观察和流式细胞仪分析绿色荧光蛋白 EGFP 的表达, 如图 6a、b 所示, pEGFP-en617 和 pEGFP-en138 均检测得到一定程度的 EGFP 蛋白表达, 其中 pEGFP-en617 荧光强度较强, 而 pEGFP-en138 次之, 无启动子的 pEGFP-Enhancer 无 EGFP 表达或非常弱. 常规方法抽提细胞总蛋白并用 BCA 法

测定蛋白质浓度, 各取 50 μg 细胞总蛋白进行 Western blot 分析, 如图 6c 所示, 与荧光显微镜直接观察及流式细胞仪计数结果基本一致. 该结果与上述 NPCEDRG 基因启动子区 Luc 报告基因载体荧光素酶活性测定结果基本吻合, 再次确证 NPCEDRG 基因 5' 端调控区 -146 ~ -8 bp 区间为其核心启动子区.

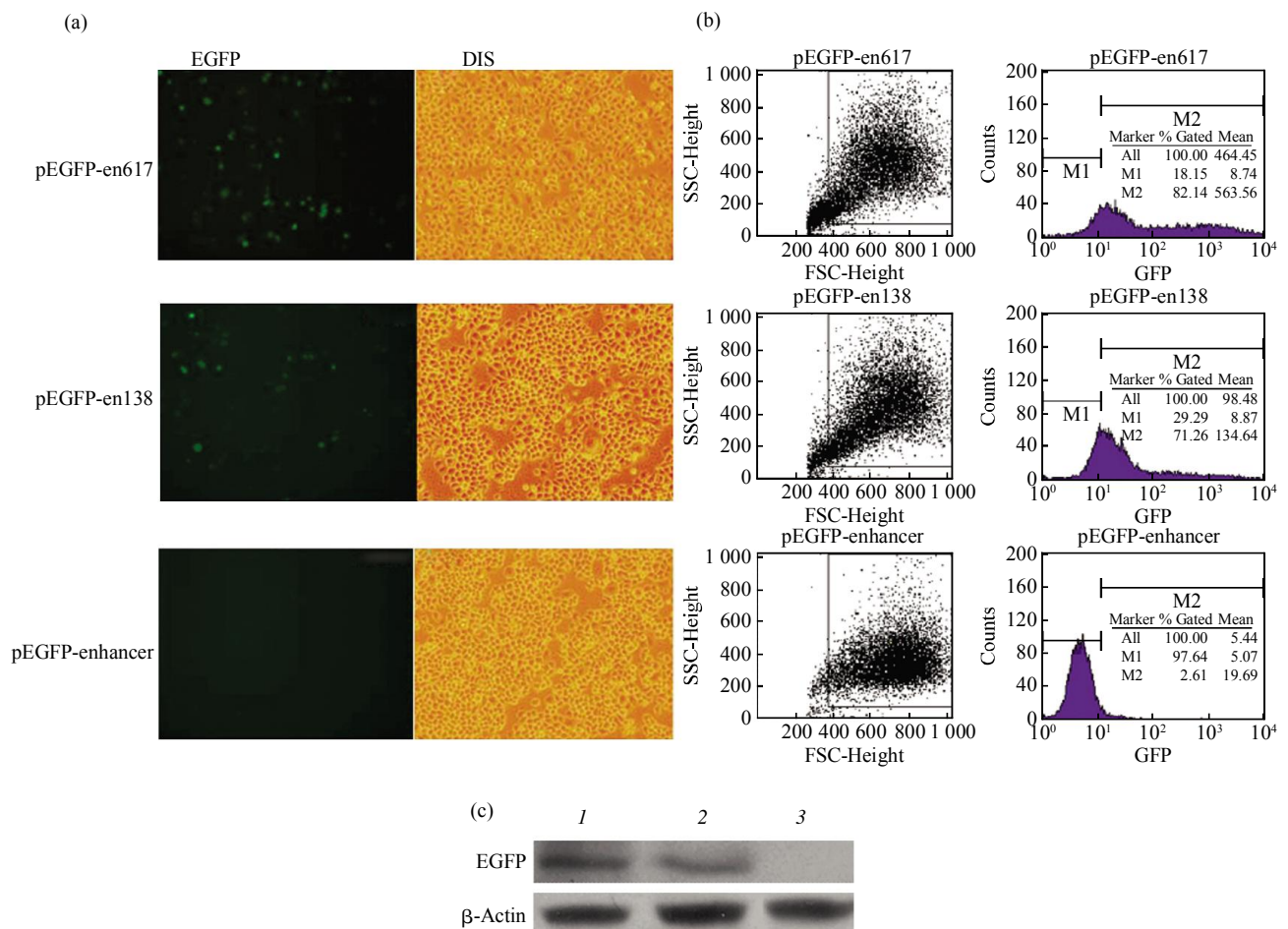


Fig. 6 Detected the EGFP expression in transfected cell lines by using fluorescence microscopy (10 \times 10), FACS and Western blot

The constructs of pEGFP-en617, pEGFP-en138, and negative control pEGFP-enhancer were transiently transfected into MCF7 cells, and the expression of EGFP was detected by fluorescence microscopy (a), FACS (b) and Western blot (c). The MCF7 cells transfected with pEGFP-en617 construct displayed high green fluorescence; pEGFP-en138 displayed moderate green fluorescence; pEGFP-enhancer has weak green fluorescence. 1: pEGFP-en617; 2: pEGFP-en138; 3: pEGFP-enhancer.

2.5 EMSA 鉴定 NPCEDRG 基因启动子区 CAAT/NFY 结合位点

系统发育进化足迹分析发现, 在多物种保守的 NPCEDRG 启动子区存在 CCAAT/NFY、STAT1 及 SP1 等转录因子结合位点, 人工合成 NPCEDRG 启动子区 CCAAT/NFY (-141 ~ -116) 26 bp 寡核苷酸

序列作为探针(图 7a), 分别以 5' 端进行生物素标记的探针(biotin-CAAT/NFY)、未标记的野生型探针(Wt CAAT/NFY)及突变型探针(Mt CAAT/NFY)与 MCF7 细胞核提取物进行结合反应(图 7b). 转录因子 CAAT/NFY 的结合鉴定反应中, 探针可以与核蛋白结合形成阻滞条带, 并可被特异性竞争抑制,

冷探针的突变体却不能抑制二者结合(图 7c). 说明在 MCF7 细胞中 NPCEDRG 启动子区 CCAAT/NFY

结合位点是特异的, 初步确定 CCAAT/NFY 转录因子结合位点参与了 NPCEDRG 基因的表达调控.

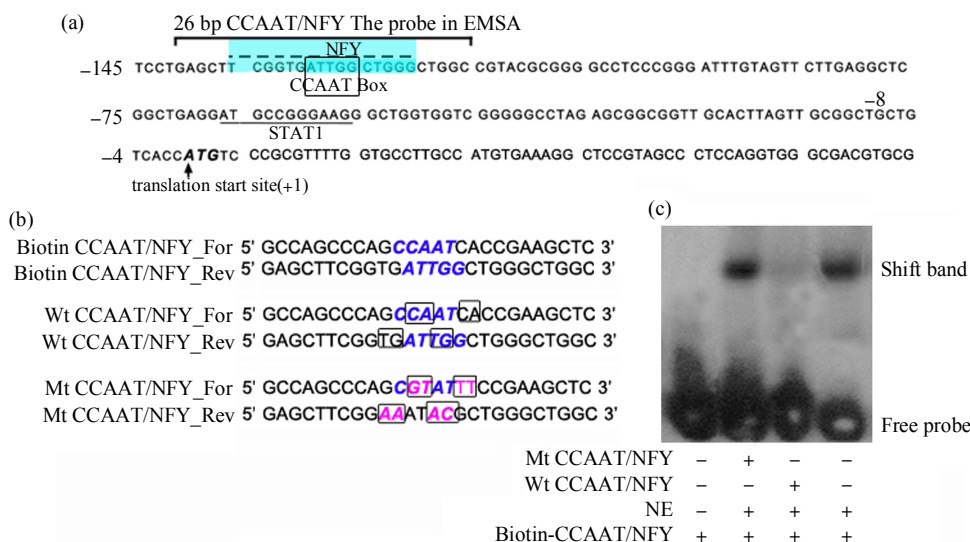


Fig. 7 Preliminary identification of CCAAT/NFY binding site in NPCEDRG gene promoter

(a) Schematic representation of the putative TFBSs of the NPCEDRG gene (-146~-8) by rVista and the probe located at the region -141~-116 bp. The start codon (ATG) is in italic bold. Nucleotide numbers are indicated on the right side of the sequence. (b) Schematic representation of the double-strand (ds) probes for EMSA. (c) Characterization of CCAAT/NFY binding site of NPCEDRG promoter. Binding reactions were performed with 20 fmol of ds probe marked with biotin, 20 μg of nuclear extracts (NE) from exponentially growing MCF7 cells, Wt-ds probe as the specific competitor and Mt-ds probe as the nonspecific competitor.

3 讨 论

NPCEDRG 基因是中南大学卫生部肿瘤蛋白质组学重点实验室采用基因定位候选克隆策略获得的一个在正常鼻咽和鼻咽癌组织表达差异具有显著性的基因^[9-10]. 我们前期研究结果表明, NPCEDRG 基因在正常人多种组织中表达, 而在鼻咽癌组织、鼻咽癌细胞系及其他肿瘤细胞系中表达下调, 功能研究提示 NPCEDRG 基因可能为一个鼻咽癌候选抑瘤基因^[10-11]. 为明确 NPCEDRG 基因在鼻咽癌组织、鼻咽癌细胞系及其他肿瘤细胞系中表达下调机制, 我们对 NPCEDRG 基因转录调控机制进行了研究. RT-PCR 结果显示, NPCEDRG 基因在所检测的 7 种不同肿瘤细胞中均为低表达, 且在 A549 和 HNE2 等细胞系中表达非常低, 而在 MCF7、HeLa 及 CNE2 等细胞中相对较高, 因此, 我们选用 MCF7、HeLa 和 CNE2 细胞作为后续启动子研究的靶细胞.

真核生物转录调控大多是通过顺式作用元件和反式作用因子相互作用而实现的. 真核基因启动子是 RNA 聚合酶结合位点周围的一组转录控制组件, 每一组件含 7~20 bp 的 DNA 序列, 其核心元件包括 TATA 盒、BRE、Inr、MTE、DPE、DCE 及 XCPE1 等, 主要位于 TSS 附近的 50~100 bp, 为控制转录起始, 并决定着基因表达强度的序列, 在转录调控中扮演着非常重要的角色^[17-19]. 研究启动子的常用策略是联合应用计算机软件分析方法和报告基因载体分析系统, 作为分析 NPCEDRG 基因表达调控机制的第一步, 本研究首先对其起始密码子 ATG 上游启动子区域进行了初步分析. 真核生物基因启动子的转录控制组件序列特征是利用信息学方法识别启动子所必须考虑的主要信息来源, 研究表明, 约只有 10%~20% 的蛋白编码基因含有 TATA 盒和特定的 TSS, 而绝大部分蛋白编码基因启动子区不含 TATA 盒(TATA-less promoter), 但常有多个选择性启动子区(alterlative promoter, AP)

和 TSSs, 其中有约 72% 的基因启动子区含有 CpG 岛^[19-21]. 由于不同基因的启动子区结构可能不同, 因此, 在进行启动子预测时, 选择用于分析的 DNA 区段很重要. 通常, 仅使用转录起始位点的上游区段在很多程序里都会导致假阴性, 而如果所用区段包含转录起始位点、翻译起始位点和部分编码区等信息, 阳性率则会显著提高. 因此, 我们选取 NPCEDRG 基因 5' 端上游共 4 000 bp 区段(包含转录起始位点、翻译起始位点、部分编码区及部分内含子等), 利用多个生物信息学软件预测该基因其启动子区域、TSSs 及 CpG 岛, 结果表明预测启动子区与 CpG 岛重叠区域为 -624 ~ +75 bp, 且有多数 TSSs 主要介于 -85 ~ -9 bp 区间. 系统发育进化足迹分析表明, NPCEDRG 基因 5' 端有约 415 bp 序列(-180 ~ +235 bp 区间)与黑猩猩、猕猴、狗、牛及小鼠等高度匹配, rVista 分析结果显示在保守区域中存在包括 CCAAT/NFY、STAT1 和 SP1 等重要转录因子结合位点. 综合各生物信息学分析结果提示, NPCEDRG 基因启动子为 TATA-less 启动子, 且其核心启动子位于 -180 ~ +74 bp 区, 可能包含启动子核心元件 CCAAT/NFY、STAT1 和 SP1 等参与 NPCEDRG 基因的转录调控. 在此基础上, 我们构建 Luc 报告基因表达载体和 EGFP 报告基因表达载体对 NPCEDRG 基因 5' 端上游调控区进行了克隆和功能性分析. 通过对 Luc 报告基因表达载体在 CNE2、MCF7 及 HeLa 细胞中表达活性分析发现, 证实 -146 ~ -8 bp 区间为 NPCEDRG 基因的核心启动子区域, 且 -625 ~ -8 bp 区间启动子活性达到 -146 ~ -8 bp 区间的 2 倍以上, 提示 -625 ~ -216 bp 区间可能存在增强 NPCEDRG 基因表达的重要调控元件, 但有待进一步研究证实.

凝胶电泳迁移率阻滞分析(EMSA)是研究序列特异性 DNA 结合蛋白的最常用方法, 常用作基因转录调控因子的初步筛选, 采用此方法可以鉴定特定细胞核中是否存在特定调控序列的结合蛋白, 也可用来鉴定特定基因序列中是否存在特定 DNA 结合蛋白的结合位点. 利用系统发育足迹分析法对其转录因子结合位点分析结果提示, NPCEDRG 基因基本启动子区正好处于人类和小鼠的高度保守区域, 该区域含有 CCAAT 盒等重要的启动子组件. 研究表明 NFY 是一个普遍存在的转录因子, 其与基因的核心启动子区 CCAAT 盒结合参与基因的转录起始^[22]. NFY 为一种异源三聚体, 由 NFYA、NFYB 和 NFYC 组成, NFYA 和 NFYC 先组成异

源二聚体, 然后吸引 NFYB 形成三聚体才能结合 CCAAT 序列. NFYA 和 NFYC 均有一个疏水性的富含谷氨酰胺(Q-rich)结构域, 该结构域为 NFY 转录激活作用所必需^[22-24]. NFY 与 CCAAT 盒的结合为许多哺乳动物基因转录起始所必需, 尤其是 TATA-Less 启动子基因的转录, NFY 结合 CCAAT 盒后可与其他基础转录因子及相关因子(如 RNAP II、TFIID、TBP、PC4 和 HATs 等)相互作用和 / 或募集转录起始复合物^[25-28], 从而提高基因转录活性和效率. 本实验中采用 EMSA 方法鉴定被测细胞核中是否存在 CCAAT 区域的特异结合蛋白, 结果在 MCF7 细胞核提取液中检测到 CCAAT 序列的结合蛋白, 此结合可被同源序列竞争抑制, 而不被突变序列竞争抑制, 说明结合蛋白与 CCAAT 序列的结合是特异性的. 因此, 我们推测 CCAAT/NFY 参与 NPCEDRG 基因的转录调控, 但有待进一步证实. 总之, 本研究结果为深入探讨 NPCEDRG 基因表达调控机制奠定了基础.

参 考 文 献

- [1] Parkin D M, Whelan S L, Ferlay J, *et al.* Cancer incidence In Five Continents. Vol. 7. Lyon: IARC, 1997: 334-337
- [2] Zeng Z Y, Zhou Y, Zhang W, *et al.* Family-based association analysis validates chromosome 3p21 as a putative nasopharyngeal carcinoma susceptibility locus. *Genet Med*, 2006, **8**(3): 156-160
- [3] Chien Y C, Chen J Y, Liu M Y, *et al.* Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med*, 2001, **345**(26): 1877-1882
- [4] Ward M H, Pan W H, Cheng Y J, *et al.* Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. *Int J Cancer*, 2000, **86**(5): 603-609
- [5] Xiong W, Zeng Z Y, Xia J H, *et al.* A susceptibility locus at cChromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res*, 2004, **64**(6): 1972-1974
- [6] Chan A S, To K F, Lo K W, *et al.* High frequency of chromosome 3p deletion in histologically normal nasopharyngeal epithelia from southern Chinese. *Cancer Res*, 2000, **60**(19): 5365-5370
- [7] Chow L S, Lam C W, Chan S Y, *et al.* Identification of RASSF1A modulated genes in nasopharyngeal carcinoma. *Oncogene*, 2006, **25**(2): 310-316
- [8] Yau W L, Lung H L, Zabarovsky E R, *et al.* Functional studies of the chromosome 3p21.3 candidate tumor suppressor gene BLU/ZMYND10 in nasopharyngeal carcinoma. *Int J Cancer*, 2006, **119**(12): 2821-2826
- [9] 贺修胜, 陈主初, 田 芳, 等. 鼻咽癌中染色体 3p21 区域一个表达下调的 EST 的鉴定. *癌症*, 2003, **22**(1): 1-5
He X S, Chen Z C, Tian F, *et al.* *Chin J Cancer*, 2003, **22**(1): 1-5
- [10] 关勇军, 贺修胜, 侯德富, 等. 一个鼻咽癌相关 EST 的鉴定及其全长 cDNA 序列分析. *生命科学研究*, 2006, **10**(2): 173-177

- Guan Y J, He X S, Hou D F, *et al.* Life Sci Res, 2006, **10**(2):173-177
- [11] 阳 帅, 胡 华, 邓 敏, 等. 鼻咽癌相关新基因 NPCEDRG 表达对 CNE2 细胞生长的影响. 生物化学与生物物理进展, 2010, **37**(2): 167-174
Yang S, Hu H, Deng M, *et al.* Prog Biochem Biophys, 2010, **37**(2): 167-174
- [12] Bajic V B, Seah S H, Chong A, *et al.* Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters. Bioinformatics, 2002, **18**(1): 198-199
- [13] Werner T. Computer-assisted analysis of transcription control regions. MatInspector and other programs. Meth Mol Biol, 2006, **132**: 337-349
- [14] Larsen F, Gundersen G, Lopez R, *et al.* CpG islands as gene markers in the human genome. Genomics, 1992, **13**(4): 1095-1107
- [15] Wakaguri H, Yamashita R, Suzuki Y, *et al.* DBTSS: database of transcription start sites, progress report 2008. Nucleic Acids Res, 2008, **36**(Database issue): D97-101
- [16] Ovcharenko I, Nobrega M A, Loots G G, *et al.* ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Research, 2004, **32**(Web server issue): W280-W286
- [17] Martinez E, Chiang C M, Ge H, *et al.* TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. EMBO J, 1994, **13**(13): 3115-3126
- [18] Juven-Gershon T, Hsu J Y, Theisen J W, *et al.* The RNA polymerase II core promoter-the gateway to transcription. Curr Opin Cell Biol, 2008, **20**(3): 253-259
- [19] Sandelin A, Carninci P, Lenhard B, *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet, 2007, **8**(6): 424-436
- [20] Bajic V B, Tan S L, Christoffels A, *et al.* Mice and men: their promoter properties. PLoS Genet, 2006, **2**(4): e54
- [21] FitzGerald P C, Shlyakhtenko A, Mir A A, *et al.* Clustering of DNA sequences in human promoters. Genome Res, 2004, **14**(8): 1562-1574
- [22] Mantovani R. The molecular biology of the CCAAT-binding factor NF-Y. Gene, 1999, **239**(1): 15-27
- [23] de Silvio A, Imbriano C, Mantovani R. Dissection of the NF-Y transcriptional activation potential. Nucleic Acids Res, 1999, **27**(13): 2578-2584
- [24] Ceribelli M, Benatti P, Imbriano C, *et al.* NF-YC complexity is generated by dual promoters and alternative splicing. J Biol Chem, 2009, **284**(49): 34189-34200
- [25] Bellorini M, Lee D K, Dantonel J C, *et al.* CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues. Nucleic Acids Res, 1997, **25**(11): 2174-2181
- [26] Kabe Y, Yamada J, Uga H, *et al.* NF-Y is essential for the recruitment of RNA polymerase II and inducible transcription of several CCAAT box-containing genes. Mol Cell Biol, 2005, **25**(1): 512-522
- [27] Tapias A, Monasterio P, Ciudad C J, *et al.* Characterization of the 5'-flanking region of the human transcription factor Sp3 gene. Biochim Biophys Acta, 2005, **1730**(2): 126-136
- [28] Morachis J M, Murawsky C M, Emerson B M. Regulation of the p53 transcriptional response by structurally diverse core promoters. Genes Dev, 2010, **24**(2): 135-147

Cloning of Human NPCEDRG Core Promoter and Preliminary Identification of Its CCAAT/NFY Binding Site*

HOU De-Fu^{1,2,3**}, GUAN Yong-Jun^{2**}, GUAN Rui^{1,2)}, OUYANG Yong-Mei²⁾,
YU Yan-Hui²⁾, CHEN Zhu-Chu^{1,2)***}

¹⁾ Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha 410008, China;

²⁾ Cancer Research Institute, Xiangya School of Medicine, Central South University, Changsha 410008, China;

³⁾ Department of Biochemistry and Molecular Biology, Medical College of Hunan Normal University, Changsha 410013, China)

Abstract NPCEDRG is an NPC associated suppressive gene cloned by positional candidate cloning strategy. Its transcriptional down-expression has been shown in the cell lines and primary tumor tissues of NPC. Reintroduction of NPCEDRG into CNE2, a cell line derived from NPC, was effective to induce cell differentiation, control cell growth, and regulate the cell cycle. To uncover the molecular mechanisms underlying down-expression of NPCEDRG in NPC cells, bioinformatics approaches and functional assays in different tumor cell lines were used to identify and characterize the NPCEDRG core promoter and *cis*-acting elements. The conserved region from -180 to +235 bp was found in the potential promoter among 6 vertebrate species by the ECR browser, and there have several potential binding sites for transcription factors, such as CCAAT/NFY, STAT1 and SP1. To characterize the NPCEDRG core promoter, transient luciferase and/or EGFP reporter assay were carried out with the construct pGL3-en138. The results demonstrated that the core promoter is located at the conserved region from -146 to -8 nucleotides. Gel shift assay revealed the specific binding of some nuclear proteins to probes containing a putative CCAAT/NFY site, suggesting that the CCAAT/NFY site contributes to the regulation of NPCEDRG gene expression.

Key words NPCEDRG gene, core promoter, transcription regulation, CCAAT/NFY binding site

DOI: 10.3724/SP.J.1206.2010.00673

*This work was supported by grants from The National Natural Science Foundation of China (30772401) and Scientific Research Fund of Hunan Provincial Health Department (B2007006).

**These authors contributed equally to this work.

***Corresponding author.

Tel: 86-731-82355116, Fax: 86-731-84327321, E-mail: tcb1@xysm.net

Received: December 22, 2010 Accepted: April 12, 2011