

鸡 *NRG4* 基因组及转录本结构分析

高智慧^{1,2,3}, 黄佳新^{1,2,3}, 罗昊玉^{1,2,3}, 徐海冬^{1,2,3}, 娄明^{1,2,3}, 宁博林^{1,2,3},
邢晓旭^{1,2,3}, 牟芳^{1,2,3}, 李辉^{1,2,3}, 王宁^{1,2,3}

1. 农业农村部鸡遗传育种重点实验室, 哈尔滨 150030
2. 黑龙江省普通高等学校动物遗传育种与繁殖重点实验室, 哈尔滨 150030
3. 东北农业大学动物科学技术学院, 哈尔滨 150030

摘要: 神经调节蛋白4(neuregulin 4, *NRG4*)是一个重要的脂肪细胞因子, 在维持哺乳动物能量平衡、调节糖脂代谢和预防非酒精性脂肪性肝病中起着非常重要的作用。目前, 人 *NRG4* 基因的基因组结构、转录异构体和蛋白异构体等已有深入的研究。本实验室前期研究显示, 鸡脂肪组织也表达 *NRG4*, 但是目前有关鸡 *NRG4*(chicken *NRG4*, *cNRG4*)基因的基因组结构、转录异构体和蛋白异构体尚不清楚。为此, 本研究采用 RACE(rapid-amplification of cDNA ends)和 RT-PCR(reverse transcription-PCR)等技术, 系统开展了 *cNRG4* 基因的基因组和转录本的结构分析。结果发现, *cNRG4* 基因编码区很小, 但该基因却非常复杂, 存在选择性转录起始位点、选择性拼接、内含子滞留、隐匿外显子和选择性多聚腺苷酸化, 这导致 *cNRG4* 基因产生4种不同的5'UTR异构体(*cNRG4 A*、*cNRG4 B*、*cNRG4 C*和*cNRG4 D*)和6种不同的3'UTR异构体(*cNRG4 a*、*cNRG4 b*、*cNRG4 c*、*cNRG4 d*、*cNRG4 e*和*cNRG4 f*)。基因组结构分析发现, *cNRG4* 基因跨越基因组21,969 bp (Chr.10: 3,490,314~3,512,282), 由11个外显子和10个内含子构成。与NCBI数据库中的*cNRG4* 基因mRNA序列(NM_001030544.4)相比, 本研究发现了*cNRG4* 基因的2个新外显子和1个隐匿外显子。生物信息学分析、RT-PCR、克隆和测序分析发现, *cNRG4* 基因能编码3种蛋白异构体(*cNRG4-1*、*cNRG4-2*和*cNRG4-3*)。本研究为进一步开展*cNRG4* 基因的功能和调控研究奠定了基础。

关键词: 鸡; *NRG4*; 基因组结构; 转录异构体; 蛋白异构体

收稿日期: 2023-01-03; 修回日期: 2023-04-04; 网络发布日期: 2023-04-19

基金项目: 国家自然科学基金项目(编号: 31872346)和国家肉鸡产业技术体系(编号: CARS-41)资助[Supported by the National Natural Science Foundation of China (No. 31872346) and the China Agriculture Research System (No. CARS-41)]

作者简介: 高智慧, 在读硕士研究生, 专业方向: 动物遗传育种与繁殖。E-mail: gaozhahui@neau.edu.cn

通讯作者: 王宁, 博士, 教授, 博士生导师, 研究方向: 动物遗传育种与繁殖。E-mail: wangning@neau.edu.cn

DOI: 10.16288/j.ycz.23-001

Characterization of the genomic and transcriptional structure of chicken *NRG4* gene

Zhihui Gao^{1,2,3}, Jiixin Huang^{1,2,3}, Haoyu Luo^{1,2,3}, Haidong Xu^{1,2,3}, Ming Lou^{1,2,3}, Bolin Ning^{1,2,3}, Xiaoxu Xing^{1,2,3}, Fang Mu^{1,2,3}, Hui Li^{1,2,3}, Ning Wang^{1,2,3}

1. Key Laboratory of Chicken Genetics and Breeding, Ministry of Agriculture and Rural Affairs, Harbin 150030, China

2. Key Laboratory of Animal Genetics, Breeding and Reproduction, Education Department of Heilongjiang Province, Harbin 150030, China

3. College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China

Abstract: Neuregulin 4 (NRG4) is an important adipocytokine, which plays crucial roles in maintaining energy balance, regulating glucose and lipid metabolism, and preventing non-alcoholic fatty liver disease in mammals. At present, the genomic organization, transcript and protein isoforms of human *NRG4* gene have been fully explored. Previous studies in our laboratory have shown that the *NRG4* gene is expressed in chicken adipose tissue, but the chicken *NRG4* (*cNRG4*) genomic structure, transcript and protein isoforms are still unknown. To this end, in this study, the genomic and transcriptional structure of the *cNRG4* gene were systematically investigated using rapid amplification of cDNA ends (RACE) and reverse transcription-polymerase chain reaction (RT-PCR). The results showed that the coding region (CDS) of the *cNRG4* gene was small, but it had a very complex transcriptional structure characterized by multiple transcription start sites, alternative splicing, intron retention, cryptic exons, and alternative polyadenylation, thus leading to production of four 5'UTR isoforms (*cNRG4 A*, *cNRG4 B*, *cNRG4 C*, and *cNRG4 D*) and six 3'UTR isoforms (*cNRG4 a*, *cNRG4 b*, *cNRG4 c*, *cNRG4 d*, *cNRG4 e*, and *cNRG4 f*) of the *cNRG4* gene. The *cNRG4* gene spanned 21,969 bp of genomic DNA (Chr.10:3,490,314~3,512,282) and consisted of 11 exons and 10 introns. Compared with the *cNRG4* gene mRNA sequence (NM_001030544.4), two novel exons and one cryptic exon of the *cNRG4* gene were identified in this study. Bioinformatics analysis, RT-PCR, cloning and sequencing analysis showed that the *cNRG4* gene could encode three protein isoforms (*cNRG4-1*, *cNRG4-2* and *cNRG4-3*). This study lays a foundation for further research on the function and regulation of the *cNRG4* gene.

Keywords: chicken; NRG4; genomic structure; transcript isoform; protein isoform

神经调节蛋白 4(neuregulin4, NRG4)是一个重要的脂肪细胞因子,属于 EGF(epidermal growth factor)家族。哺乳动物的研究发现, *NRG4* 基因在棕色脂肪组织中高表达,此外,在白色脂肪组织、胰腺、肺、心脏等多种器官和组织中也有不同程度的表达^[1]。NRG4 通过其 EGF 结构域与受体 ErbB4(v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4)的胞外区域相结合,使 ErbB4 和 ErbB3 发生二聚化,进而激活下游磷脂酰肌醇(phosphatidyl inositol 3 kinase, PI3K)和丝裂原活化

蛋白激酶(mitogen-activated protein kinase, MAPK)等信号通路^[2],从而促进细胞增殖,抑制细胞凋亡以及改善细胞能量代谢等。临床研究显示,成人非酒精性脂肪性肝病(nonalcoholic fatty liver disease, NAFLD)的发生率与血清 NRG4 水平呈负相关^[3]。对转基因小鼠(*Mus musculus*)的研究发现, NRG4 可降低肝脏的脂肪合成,改善 NAFLD 和胰岛素抵抗^[4]。流行病学调查显示,糖尿病患者血清中的 NRG4 含量明显高于正常对照组^[5]。这些研究提示, NRG4 有望成为多种疾病的潜在治疗靶点。

神经调节蛋白(neuregulins, NRGs)家族包括4个成员,分别为NRG1、NRG2、NRG3和NRG4,这4个NRG家族成员均具有多个转录异构体和蛋白异构体^[6-9]。人*NRG4*基因包含9个外显子和8个内含子,跨越基因组71,509 bp^[8]。人*NRG4*基因能产生*NRG4A1*、*NRG4A2*、*NRG4B1*、*NRG4B2*和*NRG4B3*共5种转录异构体,这些转录异构体拥有相同的起始密码子,由于选择性剪接它们的终止密码子位置不同,导致人*NRG4*基因的5种转录异构体编码5种不同的NRG4蛋白异构体(NRG4A1、NRG4A2、NRG4B1、NRG4B2和NRG4B3)^[10]。其中NRG4A1和NRG4A2具有完整的EGF结构域和跨膜结构域,定位于细胞膜,经过ADAM/TACE型蛋白酶加工后,两者位于膜外的EGF结构域被释放到细胞外,作为分泌因子参与机体的调控^[8]。与NRG4A1和NRG4A2相比,其余3种NRG4异构体(NRG4B1~3)缺失跨膜结构域,EGF结构域也不完整,均定位在细胞质^[11]。

与哺乳动物不同,鸡(*Gallus gallus*)缺乏棕色脂肪组织。本实验室前期研究显示,鸡腹部脂肪组织表达*NRG4*,并且发现东北农业大学高脂系肉鸡脂肪组织*NRG4*表达量是低脂系肉鸡脂肪组织表达量的2.55倍^[12]。目前人们对哺乳动物*NRG4*基因的基因组结构、转录异构体和蛋白异构体都已有了深入研究,但对鸡*NRG4*(chicken *NRG4*, *cNRG4*)基因的基因组结构、转录异构体和蛋白异构体还不甚了解。本文采用RACE(rapid-amplification of cDNA ends)和RT-PCR(reverse transcription-PCR)等技术,对*cNRG4*基因的基因组和转录本结构进行了分析。

1 材料与方法

1.1 实验材料

克隆载体 pEASY-Blunt Simple 和 pEASY-T1 Simple、菌株 Trans1-T1 Phage Resistant Chemically Competent Cell、Trans2K DNA Marker 和 Trans5K DNA Marker 均购自北京全式金生物技术有限公司; AxyPrep DNA 凝胶回收试剂盒购自美国 Axygen 公司; Phanta Max super-fidelity DNA polymerase 购自南京 Vazyme 公司; RNAiso Plus、SMARTer[®] RACE 5'/3'Kit 和 PrimeScript[™] RT reagent Kit with

gDNA Eraser 购自宝生物工程(大连)有限公司; FirstChoice[™] RLM-RACE Kit 购自美国 Thermo Fisher 公司; 5'/3'RACE Kit, 2nd Generation 购自德国 Roche 公司。腹部脂肪组织采自4周龄 AA 白羽肉鸡(东北农业大学阿城实习基地)。

1.2 引物设计与合成

根据 NCBI 数据库中 *cNRG4* 基因 (NM_001030544.4) 序列, 利用 Oligo 7 软件设计 *cNRG4* 基因的 RACE 引物和 RT-PCR 引物(表 1), 由北京睿德兴科生物技术有限公司合成。

1.3 RNA 提取

取 AA 白羽肉鸡腹部脂肪组织样, 置于研钵中, 加入液氮充分研磨成粉末, 采用 RNAiso Plus 试剂提取 RNA。使用紫外分光光度计测定样品总 RNA 在 260 nm 与 280 nm 的吸光度, 根据两者的比值 ($OD_{260/280}$) 评价总 RNA 的质量, 并计算总 RNA 的浓度, 分装冻存于 -80°C 。

表 1 本研究所用引物信息

Table 1 Primers used in this study

引物类型	引物名称	引物序列(5'→3')
RACE	GSP1-R1	AGAAGGAGAAGTGAAGCCAAGAGA ACTG
	GSP2-F1	GTTTGACAGTCTCTTGGCTTCACTTCTC
RT-PCR	E76-F/E4-R2	GCCGAGCCAATGGGCGACG GATCCATAACTGGTGCCACAGAGT
	E85-F/E4-R1	GGTGAGGCGCGAGAG GACAAAAAGATCCATAACTGGTGC
	E56-F/E4-R2	CTCCTTCACAGGCACAGTTCAG GATCCATAACTGGTGCCACAGAGT
	I1-F/E4-R1	CTCTGTAAACGTTCCGCTG GACAAAAAGATCCATAACTGGTGC
	E4-F /I5-R	TTCCTACTGTACCCAGTCCATTC CTACTTGATTAATAGCTTCAGCAC
	E5-F /I7-R	CTTCTCGGAGTCTTTTAAATTGGA TCTCTAAGATAAAAGTAGGCCTC
	CDS1-F/R	ATGCGAACAGATCATGAAG TTACTCGCAAAGGAAGTAG
	CDS2-F/R	ATGCGAACAGATCATGAAG TTACTCCGTTATTATTTTGTTCAG
	CDS3-F/R	ATGCGAACAGATCATGAAG TAGTACATGATTACCAGAAGCAA

1.4 cDNA 末端快速扩增(RACE)

按照 SMARTer[®] RACE 5'/3'Kit、FirstChoice[™] RLM-RACE Kit 和 5'/3'RACE Kit, 2nd Generation 说明书分别扩增 *cNRG4* 基因 mRNA 的 5'和 3'端。5'和 3'RACE 引物序列如表 1 所示。以鸡腹部脂肪组织提取的混合 RNA 为模板, 分别反转录获得 5'RACE-Ready cDNA 和 3'RACE-Ready cDNA。利用获得的 cDNA 分别进行 PCR 扩增, 反应体系均为: PCR-Grade H₂O 15.5 μL, 2×SeqAmp Buffer 25.0 μL, SeqAmp DNA Polymerase 1.0 μL, 5'或者 3'RACE-Ready cDNA 2.5 μL, 10×Universal Primer A Mix(UPM)(由试剂盒提供)5 μL, 5'或者 3'GSP 1 μL, 总体积为 50 μL。

cNRG4 基因 5'和 3'RACE 的 PCR 反应条件均为: 94°C 5 min; 94°C 30 s, 72°C 3 min, 5 个循环; 94°C 30 s, 70°C 30 s, 72°C 3 min, 5 个循环; 94°C 30 s, 68°C 30 s, 72°C 3 min, 25 个循环; 72°C 5 min。5'和 3'RACE 的 PCR 产物经过琼脂糖凝胶电泳之后, 进行切胶回收, 使用 AxyPrep DNA 凝胶回收试剂盒对 RACE-PCR 产物进行凝胶纯化。将纯化的 5'和 3'RACE 的 PCR 产物分别与 pEASY-T1 Simple 克隆载体连接, 将连接产物转化 Trans1-T1 感受态细胞后, 涂布氨苄青霉素(Ampicillin, Amp)抗性 LB 平板, 37°C 过夜培养, 然后随机挑选阳性菌落, 由北京睿博兴科生物技术有限公司进行测序。

1.5 RT-PCR 检测

RT-PCR 用于新发现的外显子和内含子滞留的验证以及 CDS 区克隆。取鸡腹部脂肪组织总 RNA (1 μg), 按照 PrimeScript[™] RT reagent Kit with gDNA Eraser 试剂盒说明书, 去除基因组 DNA 并反转录合成 cDNA。以获得的 cDNA 为模板进行 PCR 扩增, RT-PCR 引物序列如表 1 所示。RT-PCR 反应体系为: 2×Phanta Max Buffer 25 μL, dNTP Mix(10 mmol/L) 1 μL, 上、下游引物(10 μmol/L)各 2 μL, Phanta Max Super-Fidelity DNA Polymerase 1 μL, 模板 1~5 μL, 加 ddH₂O 至总体积为 50 μL。反应条件: 95°C 预变性 3 min; 95°C 变性 15 s, 60°C 退火 15 s, 72°C 延伸 30 s, 共 35 个循环; 72°C 延伸 5 min。RT-PCR 产物经过琼脂糖凝胶电泳之后, 进行切胶回收, 使用 AxyPrep DNA 凝胶回收试剂盒对产物进行凝胶纯

化。将 RT-PCR 产物和 pEASY-Blunt Simple 克隆载体连接, 将连接产物转化 Trans1-T1 感受态细胞后, 涂布 Amp 抗性 LB 平板, 37°C 过夜培养, 随机挑选阳性菌落, 由北京睿博兴科生物技术有限公司进行测序。

1.6 生物信息学分析

利用 UCSC(<http://genome.ucsc.edu/>)和 NCBI(<http://www.ncbi.nlm.nih.gov>)数据库查询 *NRG4* 基因序列; 采用 Oligo 7 软件进行引物设计; 采用 DNAMAN 6.0 进行序列比对; 采用 Chromas 查看测序峰图; 采用 ORF Finder(<https://www.ncbi.nlm.nih.gov/orffinder/>)预测编码蛋白; 采用 Signal IP-5.0(<http://www.cbs.dtu.dk/services/SignalP/>)预测分泌信号肽; 采用 TMHMM2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)预测跨膜结构; 采用 SMART(<http://smart.embl-heidelberg.de/>)分析蛋白序列; 使用 ITB tools (<http://itbtools.ba.itb.cnr.it/utrscan>)分析多聚腺苷酸化信号(polyadenylation signal, PAS)。

2 结果与分析

2.1 鸡 *NRG4* 基因 5'末端克隆和序列分析

为了确定 *cNRG4* mRNA 转录本的 5'端序列, 取 5 只 AA 白羽肉鸡腹部脂肪组织的混合 RNA 进行 5'RACE 分析。琼脂糖电泳分析显示, *cNRG4* 基因 5'RACE 扩增产物主要有 3 条带, 大小分别约为 600 bp、500 bp 和 400 bp(图 1A)。将 5'RACE 扩增产物进行凝胶回收和克隆, 共随机挑选 187 个重组质粒进行测序分析。测序结果与 *cNRG4* 基因 mRNA 序列(NM_001030544.4)和基因组序列(NC_052541.1) 比对发现, 这些 5'RACE 扩增产物序列与 *cNRG4* 基因的外显子 3(E3)、E4 和 E5 序列完全相同, 但 E3 上游序列差异较大。根据 E3 上游序列的不同, 将这些 *cNRG4* 基因 5'RACE 扩增产物命名为 *cNRG4-A* (19 个克隆)、*cNRG4-B* (147 个克隆)、*cNRG4-C* (17 个克隆) 和 *cNRG4-D* (4 个克隆)(图 1B)。与 NM_001030544.4 相比, 本研究发现 *cNRG4-A*、*cNRG4-B* 和 *cNRG4-D* 分别存在 1 个新的外显子, 大小分别是 76 bp、85 bp 和 56 bp, 按照 NM_001030544.4 的外显子命名, 本研究将新发现的这

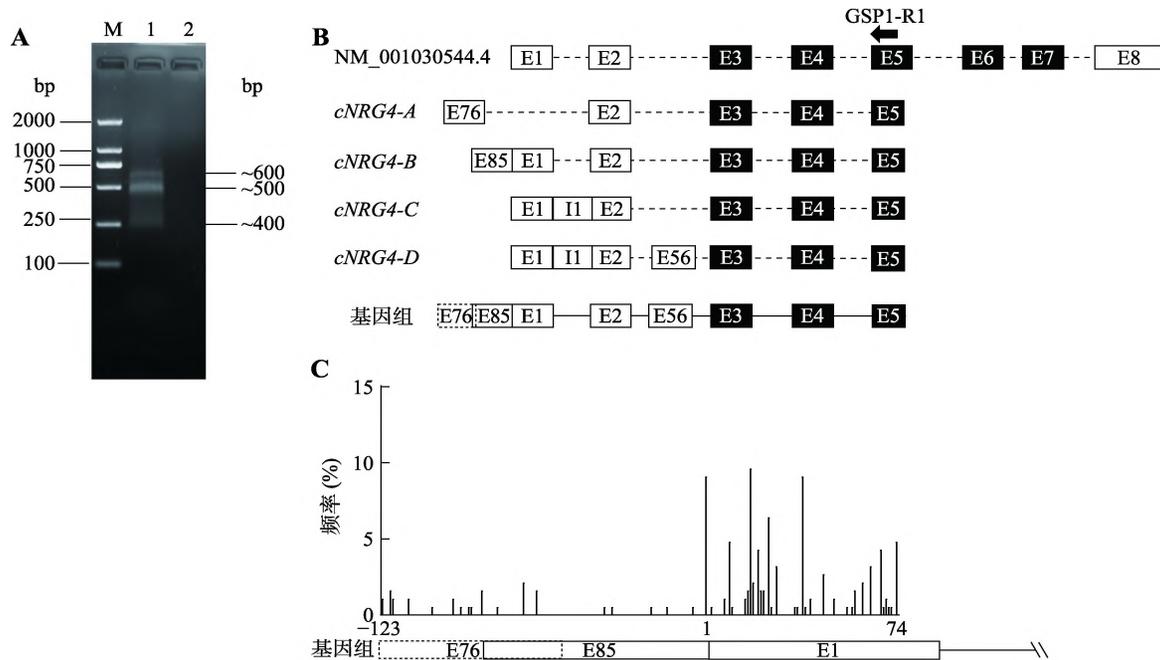


图 1 鸡 *NRG4* 基因 5'RACE 分析

Fig. 1 5'RACE analysis of the chicken *NRG4* gene

A: *cNRG4* 基因 5'RACE PCR 扩增产物的琼脂糖凝胶电泳分析。M: DL2000 DNA marker; 泳道 1: *cNRG4* 基因 5'RACE PCR 扩增片段; 泳道 2: 阴性对照。B: *cNRG4* 基因 5'RACE 产物的外显子分析。上面一排为 *cNRG4* 基因 NM_001030544.4 外显子分析结果; 中间 4 排分别为 4 种不同 5'RACE 扩增产物的外显子分析结果; 最下面一排为 5'RACE 扩增产物序列在基因组上的位置; GSP1-R1 为 5'RACE 的特异性引物。C: *cNRG4* 基因的转录起始位点分析。*cNRG4* 基因 TSS 在-123~+74 范围内的分布频率(%)。

3 个 *cNRG4* 基因外显子分别命名为 E76、E85 和 E56。其中 E85 和 E76 位于 E1 的上游, E76 的 3'端与 E85 的 5'端有 38 bp 的序列重叠, 两者为互斥外显子 (mutually exclusive exons, MEE)。E56 是一个隐匿外显子 (cryptic exon), 位于 E2 与 E3 之间的内含子区。此外, 本研究发现在 *cNRG4-C* 和 *cNRG4-D* 都存在内含子 1 (I1) 滞留, 大小为 194 bp。序列分析显示, *cNRG4-A*、*cNRG4-B*、*cNRG4-C* 和 *cNRG4-D* 的剪接都遵循 GT-AG 法则。

测序分析发现 *cNRG4* 具有 54 个转录起始位点 (transcription start site, TSS), 如图 1C 所示, 以 NM_001030544.4 的第一个碱基为 +1, *cNRG4* 的 TSS 离散地分布在 -123~+74 (-123/+74) 的范围内。其中 E76 有 12 个 TSS, 分布在 (-123/-64) 范围内, 导致 E76 大小在 17~76 bp 范围内变化; E85 有 9 个 TSS, 分布在 (-85/0) 范围内, 导致 E85 大小在 4~85 bp 范围内变化; E1 有 33 个 TSS, 分布在 (+1/+74) 范围内, 导致 E1 大小在 2~75 bp 范围内变化。

2.2 鸡 *NRG4* 基因 3'末端克隆和序列分析

为了获得完整 *cNRG4* mRNA 转录本序列, 取 5 只 AA 白羽肉鸡腹部脂肪组织的混合 RNA 进行 3'RACE 分析。琼脂糖电泳分析显示, *cNRG4* 基因 3'RACE 的 PCR 扩增产物主要有 3 条带, 大小分别约为 2000 bp、1500 bp 和 600 bp (图 2A)。将 3'RACE 扩增产物进行凝胶回收和克隆, 共随机挑选 36 个重组质粒进行测序分析。测序分析结果显示, *cNRG4* 基因有 5 种不同的 3'RACE 扩增产物, 大小分别是 2140 bp、553 bp、1458 bp、1449 bp 和 2422 bp, 命名为 *cNRG4-a* (1 个克隆)、*cNRG4-b* (15 个克隆)、*cNRG4-c* (13 个克隆)、*cNRG4-d* (4 个克隆) 和 *cNRG4-e* (3 个克隆)。

将 3'RACE 扩增产物序列与 *cNRG4* 基因 mRNA 序列 (NM_001030544.4) 和基因组序列 (NC_052541.1) 进行比对, 发现 3'RACE 扩增产物均包含 *cNRG4* 基因 E6 和 E7, 且 *cNRG4-c* 序列与 NM_001030544.4

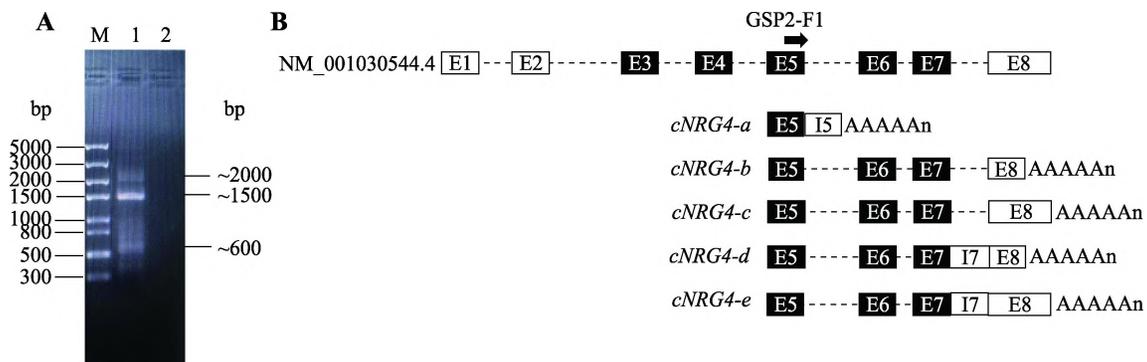


图 2 鸡 *NRG4* 基因 3'RACE 分析

Fig. 2 3'RACE analysis of the chicken *NRG4* gene

A: *cNRG4* 基因 3'RACE PCR 扩增产物的琼脂糖凝胶电泳分析。M: DL5000 DNA marker; 泳道 1: *cNRG4* 基因 3'RACE PCR 扩增片段; 泳道 2: 阴性对照。B: *cNRG4* 基因 3'RACE 产物的外显子分析。最上面一排为 *cNRG4* 基因 mRNA 序列(NM_001030544.4)的外显子分析结果, 下面为 5 种不同 3'RACE 扩增产物序列的外显子分析结果。GSP2-F1 为 3'RACE 特异性引物。

的 3' 端序列完全相同, 提示本研究已成功获得 *cNRG4* 基因 mRNA 3' 端序列, 但是这些扩增产物的 3' 末端序列和长度差异较大, 说明 *cNRG4* 基因存在选择多聚腺苷酸化(alternative polyadenylation, APA)。序列分析发现 *cNRG4-a*、*cNRG4-d* 和 *cNRG4-e* 存在内含子滞留, 内含子分别为 I5 和 I7, 大小分别是 2068 bp 和 896 bp; *cNRG4-a* 的 poly(A) 加尾位点位于 I5 上, *cNRG4-b* 和 *cNRG4-d* 的 poly(A) 加尾位点位于 E8 中部, *cNRG4-c* 和 *cNRG4-e* 的 poly(A) 加尾位点位于 E8 末端(图 2B)。

2.3 鸡 *NRG4* 基因新外显子和内含子滞留验证

5'RACE 和 3'RACE 测序分析结果显示, *cNRG4* 基因的一些转录本存在新外显子(E76、E85 和 E56)和内含子滞留(I1、I5 和 I7)。为验证它们的存在, 本研究跨内含子设计了这些新外显子和内含子滞留的 6 对特异性检测引物, 引物位置如图 3A 所示。以鸡腹部脂肪组织总 RNA 为材料进行 RT-PCR 验证。RT-PCR 扩增、克隆及测序分析显示: E76、E85、E56、I1、I5 和 I7 经 RT-PCR 扩增, 条带大小与预期

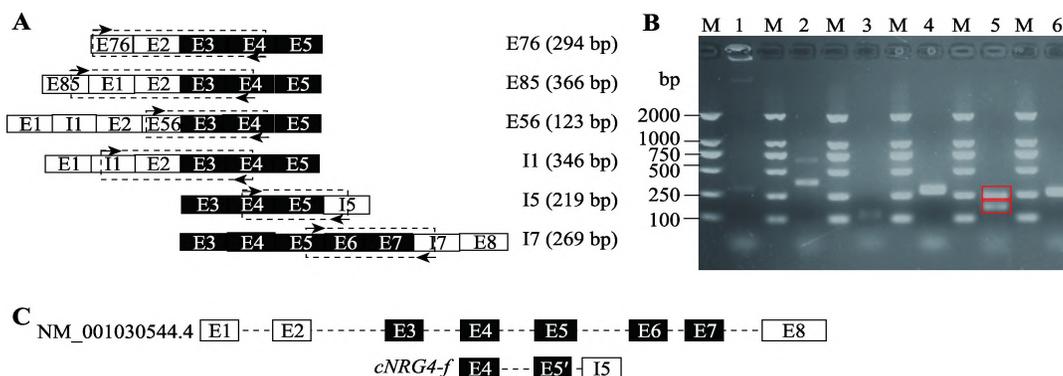


图 3 鸡 *NRG4* 基因新外显子和内含子滞留的 RT-PCR 鉴定

Fig. 3 RT-PCR identification of novel exons and intron retention in chicken *NRG4* gene

A: RT-PCR 鉴定新外显子和内含子滞留的引物位置(箭头所示)示意图。由上到下分别鉴定 E76、E85、E56、I1、I5 和 I7; 虚线框代表引物预计扩增出的片段。B: RT-PCR 鉴定新外显子和内含子滞留的琼脂糖凝胶电泳分析。M: DL2000 DNA marker; 泳道 1~6 依次为 RT-PCR 扩增 E76、E85、E56、I1、I5 和 I7 的琼脂糖电泳图, 在鉴定 I5 的电泳图中分子量大的条带为 *cNRG4-a*, 分子量小的条带为 *cNRG4-f*, 两者均用红色方框标记。C: *cNRG4-f* 和 *cNRG4* 基因 mRNA 序列(NM_001030544.4)外显子分析结果。E5 序列长度为 150 bp, E5' 序列长度 71 bp。

相符,分别为 294 bp、366 bp、123 bp、346 bp、219 bp 和 269 bp(图 3B),测序结果与 RACE 结果一致,说明 *cNRG4* 基因确实存在外显子 E76、E85、E56 及 3 个内含子(I1、I5 和 I7)的滞留。在 E85 和 I5 的 RT-PCR 验证中,除了扩增出预期大小的条带外,都扩增出了一条额外条带(图 3B)。测序分析显示,在 E85 的 RT-PCR 验证中扩增出的额外条带为非特异性条带,其序列与 *cNRG4* 基因无关,出现非特异扩增带的主要原因是 E85 比较小,只有 85 bp,难以设计高特异性的上游检测引物。但值得注意的是,在 I5 的 RT-PCR 验证中扩增出的额外条带确实为 *cNRG4* 基因转录本的一部分,该扩增产物包含 I5,但是其 E5 外显子的 3'端缺失了 79 bp 序列,本研究将其命名为 E5'外显子,将该序列命名为 *cNRG4-f*(图 3C)。生物信息学分析发现,在 EST 数据库中有 1 个 *cNRG4* 基因序列(BU348396),其 3'端与 *cNRG4-f* 完全相同,说明 *cNRG4* 基因确实存在这样一种 3'末端序列。

2.4 鸡 *NRG4* 基因转录本结构分析

整合 5'RACE、3'RACE 以及新外显子和内含子滞留的验证结果,按照随机组合排列,*cNRG4* 基因可能产生多达 24 种转录异构体(图 4),本研究进一步利用 ORF Finder 对这 24 种潜在转录本的 CDS 进行了分析。结果提示,这些 *cNRG4* 基因转录异构体可能存在 3 种不同的 CDS(CDS1、CDS2 和 CDS3),大小分别为 258 bp、351 bp 和 198 bp。将这 3 个 CDS 分别编码的蛋白命名为 *cNRG4-1*、*cNRG4-2* 和 *cNRG4-3*。在 24 种潜在 *cNRG4* 基因转录异构体中,4 种潜在异构体拥有 CDS1,16 种潜在异构体拥有 CDS2,4 种潜在异构体拥有 CDS3。CDS1、CDS2 和 CDS3 的起始密码子 ATG 均位于 E3; CDS1 和 CDS3 的终止密码子位于 I5,而 CDS2 的终止密码子位于 E7(图 4)。

cNRG4 基因的起始密码子均位于 E3,根据 5'RACE 扩增产物序列,*cNRG4* 基因存在 4 种 5'UTR 异构体。其中 *cNRG4-A* 所包含的 5'UTR 由 E76、E2 和部分 E3 组成,其 5'UTR 命名为 *cNRG4 A*; *cNRG4-B* 所包含的 5'UTR 由 E85、E1、E2 和部分 E3 组成,其 5'UTR 命名为 *cNRG4 B*; *cNRG4-C* 所包含的 5'UTR 由 E1、I1、E2 和部分 E3 组成,其 5'UTR 命名为 *cNRG4 C*; *cNRG4-D* 所包含

的 5'UTR 由 E1、I1、E2、E56 和部分 E3 组成,其 5'UTR 命名为 *cNRG4 D*(图 1B)。这 4 种 *cNRG4* 的 5'UTR 的最长和最短序列已提交 GenBank,序列号分别为 OP893938、OP893939、OP893940、OP893941、OP893942、OP893943、OP893944 和 OP893945。

根据 3'RACE 克隆测序结果和终止密码子的位置,*cNRG4* 基因存在 6 种 3'UTR 异构体,大小分别是 2064 bp、384 bp、1289 bp、1280 bp、2184 bp 和 2045 bp。3'RACE 扩增产物 *cNRG4-a* 所包含的 3'UTR 由部分 I5 组成,其 3'UTR 命名为 *cNRG4 a*; *cNRG4-b* 所包含的 3'UTR 由部分 E7 和部分 E8 组成,其 3'UTR 命名为 *cNRG4 b*; *cNRG4-c* 所包含的 3'UTR 由部分 E7 和 E8 组成,其 3'UTR 命名为 *cNRG4 c*, *cNRG4 c* 与 NM_001030544.4 的 3'UTR 序列相同; *cNRG4-d* 所包含的 3'UTR 由部分 E7、I7 和部分 E8 组成,其 3'UTR 命名为 *cNRG4 d*; *cNRG4-e* 所包含的 3'UTR 由部分 E7、I7 和 E8 组成,其 3'UTR 命名为 *cNRG4 e*; *cNRG4-f* 所包含的 3'UTR 由部分 I5 组成,其 3'UTR 命名为 *cNRG4 f*(图 2B,图 3C)。这 6 个 *cNRG4* 转录异构体的 3'UTR 序列已提交 GenBank,序列号分别为 OP893946、OP893947、OP893948、OP893949、OP893950 和 OP893951。

cNRG4 基因转录本具有不同长度和序列的 3'UTR,表明 *cNRG4* 基因存在 APA。与此相一致,ITB tools 分析发现,*cNRG4* 基因具有 3 个 PAS,分别位于 I5 和 E8,其中 E8 有两个 PAS(图 4)。根据 PAS 的位置,PAS 分为编码区 PAS(coding region PAS, CR-PAS)和非编码区 PAS(untranslated region PAS, UTR-PAS)。*cNRG4* 基因的 3 个 PAS 分别为 1 个 CR-PAS 和 2 个 UTR-PAS。*cNRG4 a* 和 *cNRG4 f* 的 PAS 位于 I5,导致 *cNRG4* 编码蛋白变小;另外 4 种 3'UTR 异构体的 PAS 都位于 E8 上,均不影响 *cNRG4* 基因的蛋白编码。此外,*cNRG4* 基因的 3 个 PAS 都位于 poly(A)加尾位点上游 10~30 核苷酸处,都具有六聚体 AAUAAA 或其相近变异体、下游的 U/GU-rich 元件和临近剪接位点的 CA(或 UA)序列等。

2.5 鸡 *NRG4* 蛋白异构体分析

由于 *cNRG4* 基因转录异构体数量众多(24 种),且转录本间序列差异较小,难以利用常规方法逐一加以鉴定。因此,本研究设计引物,以鸡腹部脂肪

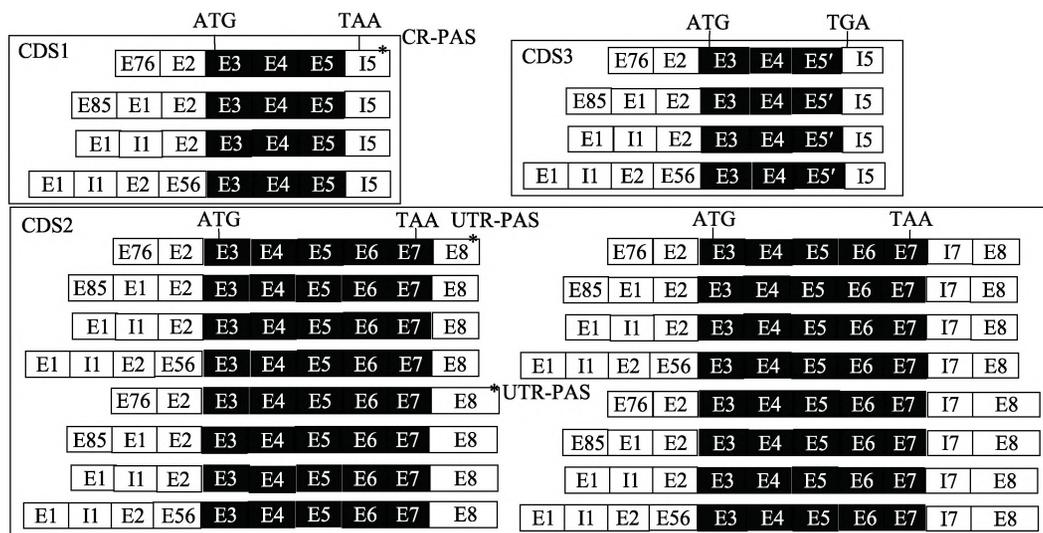


图 4 鸡 *NRG4* 基因转录本结构分析

Fig. 4 Transcript analysis of chicken *NRG4* gene

cNRG4 基因 24 种潜在转录异构体的 CDS 分析。这 24 种 *cNRG4* 基因潜在转录异构体共存在 3 种不同的 CDS: CDS1、CDS2 和 CDS3。这 3 种 CDS 的起始密码子均位于 E3, CDS1 和 CDS3 编码的蛋白终止密码子均位于 I5, CDS2 编码的蛋白终止密码子位于 E7。3 个星号(*)为本研究发现的 3 个 PAS, 其中 1 个位于 I5, 为 CR-PAS, 另外 2 个位于 E8, 均为 UTR-PAS。

组织的总 RNA 为材料, 采用 RT-PCR 验证了 *cNRG4* 基因 3 种不同 CDS。RT-PCR 扩增结果显示: 利用 CDS1 特异扩增引物(CDS1-F/R)和 CDS2 特异扩增引物(CDS2-F/R)分别扩增出一条预期大小的特异条带 (258 bp 和 351 bp); 利用 CDS3 扩增引物(CDS3-F/R) 扩增出预期大小的两条带, 一条包含 CDS1 的片段 (321 bp) 和一条包含 CDS3 的扩增片段 (242 bp) (图 5A)。这些扩增产物的测序分析与预期完全一致, CDS1、CDS2 和 CDS3 全长分别为 258 bp、351 bp 和 198 bp, 表明 *cNRG4* 基因转录异构体确实存在 3 种不同 CDS(CDS1、CDS2 和 CDS3)。将这 3 个 CDS 分别编码的蛋白命名为 cNRG4-1、cNRG4-2 和 cNRG4-3。其中 cNRG4-1 蛋白异构体包含 85 个氨基酸, cNRG4-2 蛋白异构体包含 116 个氨基酸, cNRG4-3 蛋白异构体包含 65 个氨基酸(图 5B)。cNRG4-2 与 NCBI 数据库中的 cNRG4 氨基酸序列 (NP_001025715.2) 完全相同, cNRG4-1 和 cNRG4-3 是两个新发现的 cNRG4 蛋白异构体。cNRG4-1 与人的 6 个 NRG4 蛋白异构体 (NP_612640.1、CAL35830.1、EAW99228.1、CAL35831.1、CAL35829.1 和 XP_047288140.1) 同源性分别为 68%、76%、77%、72%、65% 和 65%; 与小鼠的 3

个 NRG4 蛋白异构体 (NP_114391.1、AAH34839.1 和 EDL25847.1) 同源性分别为 61%、74% 和 46%。cNRG4-3 与人的 6 个 NRG4 蛋白异构体同源性分别为 73%、76%、77%、72%、73% 和 73%; 与小鼠的 3 个 NRG4 蛋白异构体同源性分别为 81%、74% 和 78%。cNRG4 基因 3 种 CDS 全长序列已提交 GenBank, 序列号分别为 OP288945、OP288946 和 OP288947。

NRG4 是作为前体合成的, 前体 NRG4 通过蛋白酶水解后成为有活性的形式^[4]。利用在线软件 Signal IP-5.0 预测分析显示, 这 3 个 cNRG4 蛋白异构体缺乏分泌蛋白的经典分泌信号, 这与人、小鼠 NRG4 蛋白分泌信号肽的分析结果一致。利用 TMHMM2.0 进行跨膜结构分析显示, cNRG4-1 和 cNRG4-2 都存在一个跨膜螺旋结构, cNRG4-3 没有跨膜结构。利用 SMART 进行结构域分析, 结果发现 cNRG4-1 和 cNRG4-2 都包含 1 个 EGF 结构域、1 个跨膜结构域和 1 个与跨膜结构域重叠的低复杂性区域 (low-complexity regions, LCRs), 其二者的区别在于跨膜结构域后的蛋白长度不同, 而 cNRG4-3 只包含一个 EGF 结构域, 没有跨膜结构域和低复杂性区域。

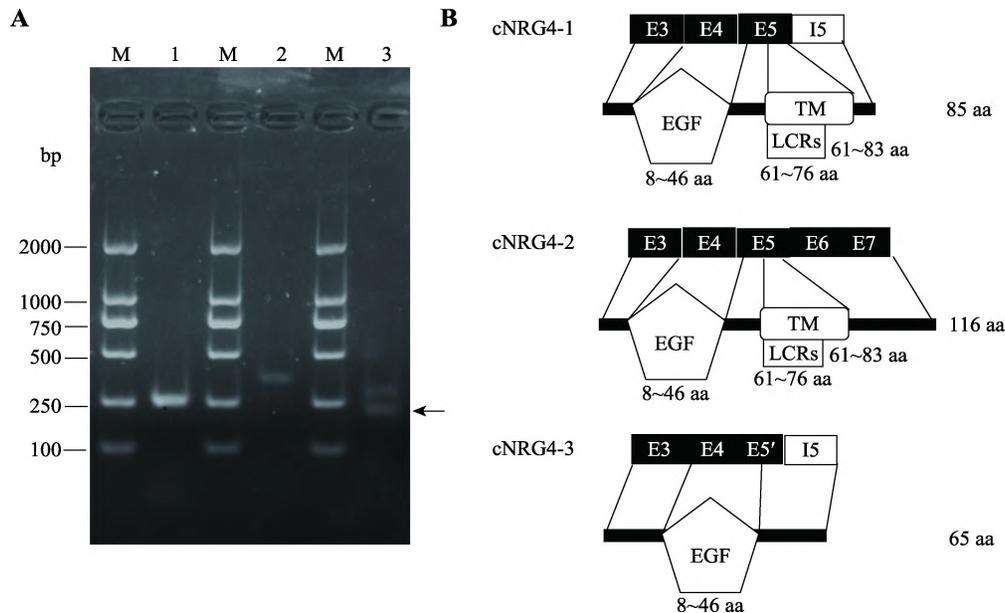


图 5 鸡 *NRG4* 蛋白质异构体的鉴定与分析

Fig. 5 Identification and analysis of chicken *NRG4* protein isoforms

A: *cNRG4* 基因 3 种不同 CDS 的 RT-PCR 鉴定分析。M: DL2000 DNA marker; 泳道 1: CDS1 的 RT-PCR 扩增片段; 泳道 2: CDS2 的 RT-PCR 扩增片段; 泳道 3: CDS3 的 RT-PCR 扩增片段。箭头所指的条带为目标条带。B: *cNRG4* 蛋白异构体的结构域分析。LCRs: low-complexity regions, *cNRG4-1* 全长有 85 个氨基酸(aa), 其中 8~46 aa 为 EGF 结构域, 61~83 aa 为跨膜结构域(TM), 61~76 aa 为低复杂性区域(LCRs)。cNRG4-2 全长 116 aa, 其中 8~46 aa 为 EGF 结构域, 61~83 aa 为跨膜结构域(TM), 61~76 aa 为低复杂性区域(LCRs)。cNRG4-3 全长 65 aa, 其中 8~46 aa 为 EGF 结构域。

3 讨论

本研究发现 *cNRG4* 基因存在多个转录异构体, 这些转录异构体编码 3 个 *cNRG4* 蛋白异构体。与本研究结果类似, 人 *NRG4* 基因能产生 5 种转录异构体和相对应的 5 种蛋白异构体(NRG4 A1、NRG4 A2、NRG4 B1、NRG4 B2 和 NRG4 B3)^[11]。NRGs 家族其他成员也普遍存在多转录异构体和蛋白异构体。例如, 人 *NRG1* 基因由于选择性转录起始位点和选择性剪接能产生 33 种转录异构体和 6 种蛋白异构体(NRG1 I-VI)^[6]; 人 *NRG2* 基因由于选择性剪接能产生至少 10 种转录异构体和 2 种蛋白异构体(NRG2- α 和 NRG2- β)^[7]; 人 *NRG3* 基因由于选择性剪接能产生 15 种转录异构体^[9]。

在 5'RACE 测序分析中, 本研究发现 *cNRG4* 基因存在选择性转录起始位点、互斥外显子、隐匿外显子和内含子滞留现象, 其转录本的 5'端非常复杂, 为了保证结果的准确性和可靠性, 本研究对 187 个 5'RACE 的 PCR 产物克隆质粒进行了测序。而在

3'RACE 测序分析中发现, *cNRG4* 基因 mRNA 的 3'端比较简单, 只存在内含子滞留和选择性多聚腺苷酸化, 因此本研究只对 36 个 3'RACE 的扩增产物的克隆质粒进行了测序。本研究发现, *cNRG4* 基因具有选择性转录起始位点、选择性剪接、选择性多聚腺苷酸化以及内含子滞留和隐匿外显子, 这导致 *cNRG4* 基因存在多个转录异构体和蛋白异构体。人 *NRG4* 基因也具有多个转录异构体和蛋白异构体, 但其造成的原因是选择性剪接^[11]。目前尚无报道其他动物 *NRG4* 基因存在选择性转录起始位点、隐匿外显子、内含子滞留以及选择性多聚腺苷酸化。此外, 生物信息学分析显示, *cNRG4* 基因 5'UTR 与人和小鼠 *NRG4* 基因 5'UTR 的序列同源性均低于 45%, *cNRG4* 基因 3'UTR 与人和小鼠 *NRG4* 基因的 3'UTR 序列的同源性均低于 40%。这些结果提示, 鸡与人和鼠等其他动物 *NRG4* 基因的表达调控机制不同。

本研究发现 *cNRG4* 基因可以编码 3 种蛋白异构体(*cNRG4-1*、*cNRG4-2* 和 *cNRG4-3*)。研究显示,

人 NRG4 具有 5 种蛋白异构体(NRG4 A1、NRG4 A2、NRG4 B1、NRG4 B2 和 NRG4 B3)^[11]。其中 cNRG4-1 和 cNRG4-2 与人的 NRG4 A 型相似, 具有完整的 EGF 结构域和跨膜结构域。人 NRG4 A 蛋白异构体是目前研究最广泛的, 其定位于细胞膜, 通过蛋白酶加工, 释放膜外 EGF 结构域, 分泌入血液, 作为经典配体发挥作用^[5], 本研究推测 cNRG4-1 和 cNRG4-2 与人 NRG4 A 类似, 均定位于细胞膜, 经过酶切加工后释放 EGF 结构域, 进入血液, 通过结合靶细胞受体 ErbB4, 触发下游信号通路。cNRG4-3 蛋白异构体是鸡独有的, 其具有完整的 EGF 样结构域, 但缺失跨膜结构域。人 NRG4 B 蛋白异构体具有 2/3 的 EGF 结构域, 缺失跨膜结构域, 定位于细胞内^[11]。由于 cNRG4-3 没有跨膜结构域, 推测其可能定位于细胞质或细胞核而发挥作用^[4,13]。

本研究发现 *cNRG4* 基因的基因组(NC_052541.1) 大小为 21,969 bp(Chr.10: 3,490,314~3,512,282), 由 11 个外显子和 10 个内含子构成。与 *cNRG4* 基因 mRNA 序列(NM_001030544.4)相比, 本研究发现了 *cNRG4* 基因存在 2 个新外显子和 1 个隐匿外显子; *cNRG4* 基因 5'端在基因组的位置向前延伸了 123 bp(图 6)。

最近的研究表明, 人和鼠的大多数基因并非只有 1 个 TSS^[14,15], 而是具有一系列分布紧密的多个 TSS, 即选择性转录起始位点, 这一区域构成一个 TSS 簇^[16]。有研究将 TSS 簇根据 TSS 数目与分布特征分为 2 种类型: 尖峰型, 分布集中且具有单一、丰度较高的 TSS, 通常从一个固定位置开始转录, 大多与 TATA-box 相关^[17]; 宽峰型, 具有多个丰度相似的邻近 TSS, 离散分布于宽泛区域^[14,16~18], 可从多个位置开始转录, 绝大多数与 CpG 岛相关^[17]。

本研究发现 *cNRG4* 基因具有多个 TSS, 离散地分布在基因组(-123/+74)。为了排除由于 RNA 降解和试剂盒差异等原因导致 *cNRG4* 基因出现多个 TSS 的可能性, 本研究尝试了多个批次的新鲜实验材料, 试用了 3 种不同原理的 RACE 分析试剂盒, 结果都证实 *cNRG4* 基因确实存在选择性 TSS, *cNRG4* 基因 TSS 簇呈宽峰型。与此相一致, *cNRG4* 基因启动子区存在一个 1206 bp 的 CpG 岛(-1019/+186)^[12]。人 *NRG4* 基因的结构信息是通过生物信息学和 RT-PCR 扩增获得的^[11], 人类 *NRG4* 基因是否存在选择性转录起始位点还不清楚。选择性转录起始点会导致基因转录本的 5'UTR 长度及其所包含调控元件的不同, 这些调控元件如上游开放阅读框(upstream open reading fragment, uORF)会影响 mRNA 稳定性、翻译效率和定位等^[18]。推测选择性 TSS 可能是 *cNRG4* 基因表达的一个重要调控方式。

内含子滞留最初在植物和病毒中被描述, 后来发现哺乳动物的基因也存在内含子滞留^[19~22]。内含子滞留在基因表达调控中发挥多种作用, 不仅可以导致无义介导的 mRNA 降解来降低基因表达, 还可以导致产生新的转录异构体和蛋白异构体^[21,22]。CDS 区的内含子滞留会导致基因产生新的蛋白异构体^[23], 5'UTR 区的内含子滞留会影响 mRNA 稳定性和翻译效率等^[24,25]; 3'UTR 区的内含子滞留也能影响 mRNA 的稳定性、翻译效率和定位等^[26,27]。本研究首次发现 *cNRG4* 基因存在内含子滞留(I1、I5 和 I7), 其中 I1 位于 5'UTR 区, 产生新的 5'UTR 异构体(*cNRG4 C* 和 *cNRG4 D*); I5 位于主 CDS 区, 导致产生了新的蛋白异构体 cNRG4-1 和 cNRG4-3; I7 位于 3'UTR 区域, 产生新的 3'UTR(*cNRG4 d* 和 *cNRG4 e*)。I1 和 I7 滞留并未影响蛋白编码, 推测它们会引入新的调控元件, 从而实现 *cNRG4* 基因表达的精细调控。

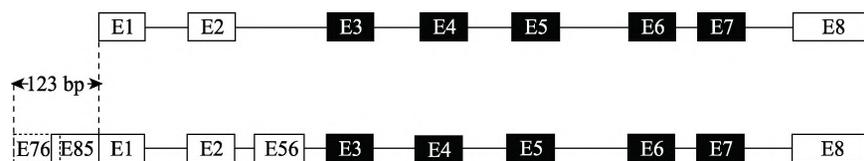


图 6 鸡 *NRG4* 基因组结构分析

Fig. 6 Genomic structure analysis of chicken *NRG4* gene

图上方为 *cNRG4* 基因 mRNA 序列(NM_001030544.4)的基因组结构。下方为本研究所确定的 *cNRG4* 基因组结构, E76、E85 和 E56 为本研究新发现的 *cNRG4* 基因的 3 个新外显子。

真核生物基因普遍具有多个 PAS, APA 是一个重要的转录后调控机制^[28]。PAS 分为 CR-PAS 和 UTR-PAS, 其中 UTR-PAS 最为常见^[29], UTR-PAS 产生具有不同 3'UTR 长度的转录异构体^[30]; CR-PAS 会导致基因编码不同 C 端的蛋白异构体, 其中一些异构体可能缺乏某些重要的功能域^[31]。本研究首次发现 *cNRG4* 基因存在 APA, 并同时具有 UTR-PAS 和 CR-PAS。有研究报道, 鸡生长激素受体(*GHR*)基因和鸡转化生长因子 β 受体 II(*TGFBR2*)基因同时具有 CR-PAS 和 UTR-PAS^[32,33]。*NRG4* 在动物体内发挥着多种重要的作用, 考虑到 APA 在转录后调控中的重要性, 未来有必要探讨 APA 在 *cNRG4* 基因表达调控中的作用和机制。

下游开放阅读框(downstream open reading fragment, dORFs)是真核生物基因中广泛存在的一个转录后调节元件, 它可以提高 mRNA 的翻译效率^[34-36]。本研究分析发现 *cNRG4* 基因的 3'UTR(*cNRG4 c*)有一个 dORF, 可以编码一个由 233 个氨基酸组成的蛋白, 其编码区大于 *cNRG4* 基因的编码区。将该 dORF 的核苷酸序列进行 BLAST 分析, 发现 *cNRG4* 的 dORF 序列与兔(*Lagopus muta*)、白斑兔(*Lagopus leucura*)、鹌鹑(*Coturnix japonica*)、鸿雁(*Anser cygnoides*)、黑天鹅(*Cygnus atratus*)和红鸭(*Oxyura jamaicensis*)等物种 *NRG4* 基因的 mRNA 同源性在 78%以上, 提示该 dORF 可能在 *NRG4* 基因的转录后调控中发挥重要作用。下一步有必要探究该 dORF 对 *cNRG4* 基因表达的调控作用。

参考文献(References):

- [1] Yang F, Li XN. Research progress of neuregulin 4 biological function. *Acta Physiol Sin*, 2017, 69(3): 351-356.
杨帆, 李晓南. 神经调节蛋白 4 生物学功能的研究进展. *生理学报*, 2017, 69(3): 351-356.
- [2] Ledonne A, Mercuri NB. On the modulatory roles of neuregulins/ErbB signaling on synaptic plasticity. *Int J Mol Sci*, 2019, 21(1): 275.
- [3] Dai YN, Zhu JZ, Fang ZY, Zhao DJ, Wan XY, Zhu HT, Yu CH, Li YM. A case-control study: association between serum neuregulin 4 level and non-alcoholic fatty liver disease. *Metabolism*, 2015, 64(12): 1667-1673.
- [4] Wang GX, Zhao XY, Meng ZX, Kern M, Dietrich A, Chen ZM, Cozacov Z, Zhou DQ, Okunade AL, Su X, Li SM, Blüher M, Lin JD. The brown fat-enriched secreted factor Nrg4 preserves metabolic homeostasis through attenuation of hepatic lipogenesis. *Nat Med*, 2014, 20(12): 1436-1443.
- [5] Chen LL, Peng MM, Zhang JY, Hu X, Min J, Huang QL, Wan LM. Elevated circulating Neuregulin4 level in patients with diabetes. *Diabetes Metab Res Rev*, 2017, 33(4): e2870.
- [6] Steinthorsdottir V, Stefansson H, Ghosh S, Birgisdottir B, Bjornsdottir S, Fasquel AC, Olafsson O, Stefansson K, Gulcher JR. Multiple novel transcription initiation sites for NRG1. *Gene*, 2004, 342(1): 97-105.
- [7] Rimer M, Prieto AL, Weber JL, Colasante C, Ponomareva O, Fromm L, Schwab MH, Lai C, Burden SJ. Neuregulin-2 is synthesized by motor neurons and terminal Schwann cells and activates acetylcholine receptor transcription in muscle cells expressing ErbB4. *Mol Cell Neurosci*, 2004, 26(2): 271-281.
- [8] Hayes NVL, Gullick WJ. The neuregulin family of genes and their multiple splice variants in breast cancer. *J Mammary Gland Biol Neoplasia*, 2008, 13(2): 205-214.
- [9] Carteron C, Ferrer-Montiel A, Cabedo H. Characterization of a neural-specific splicing form of the human neuregulin 3 gene involved in oligodendrocyte survival. *J Cell Sci*, 2006, 119(Pt 5): 898-909.
- [10] Hayes NVL, Newsam RJ, Baines AJ, Gullick WJ. Characterization of the cell membrane-associated products of the neuregulin 4 gene. *Oncogene*, 2008, 27(5): 715-720.
- [11] Hayes NVL, Blackburn E, Smart LV, Boyle MM, Russell GA, Frost TM, Morgan BJT, Baines AJ, Gullick WJ. Identification and characterization of novel spliced variants of neuregulin 4 in prostate cancer. *Clin Cancer Res*, 2007, 13(11): 3147-3155.
- [12] Guo YQ, Wang WJ, Gao ZH, Mu F, Xu HD, Li H, Wang N. Cloning, expression and promoter analysis of adipokine *NRG4* gene in chicken. *Chin J Agric Biotechnol*, 2021, 29(11): 2129-2138.
郭亚琦, 王伟佳, 高智慧, 牟芳, 徐海冬, 李辉, 王宁. 鸡脂肪细胞因子 *NRG4* 基因的克隆、表达及启动子分析. *农业生物技术学报*, 2021, 29(11): 2129-2138.
- [13] Pfeifer A. *NRG4*: an endocrine link between brown adipose tissue and liver. *Cell Metab*, 2015, 21(1): 13-14.
- [14] Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, 2018, 19(10): 621-637.
- [15] Forutan M, Ross E, Chamberlain AJ, Nguyen L, Mason B, Moore S, Garner JB, Xiang RD, Hayes BJ. Evolution of tissue and developmental specificity of transcription start sites in *Bos taurus indicus*. *Commun Biol*, 2021, 4(1): 829.
- [16] Mejía-Guerra MK, Li W, Galeano NF, Vidal M, Gray J,

- Doseff AI, Grotewold E. Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *Plant Cell*, 2015, 27(12): 3309–3320.
- [17] Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engström PG, Frith MC, Forrest ARR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 2006, 38(6): 626–635.
- [18] Thieffry A, López-Márquez D, Bornholdt J, Malekroudi MG, Bressendorff S, Barghetti A, Sandelin A, Brodersen P. PAMP-triggered genetic reprogramming involves widespread alternative transcription initiation and an immediate transcription factor wave. *Plant Cell*, 2022, 34(7): 2615–2637.
- [19] Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gontopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*, 2014, 24(11): 1774–1786.
- [20] Hammarskjöld ML. Regulation of retroviral RNA export. *Semin Cell Dev Biol*, 1997, 8(1): 83–90.
- [21] Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R. Intron retention is a major phenomenon in alternative splicing in arabidopsis. *Plant J*, 2004, 39(6): 877–885.
- [22] Rekosh D, Hammarskjöld ML. Intron retention in viruses and cellular genes: detention, border controls and passports. *Wiley Interdiscip Rev RNA*, 2018, 9(3): e1470.
- [23] Marquez Y, Höpfler M, Ayatollahi Z, Barta A, Kalyna M. Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res*, 2015, 25(7): 995–1007.
- [24] Tahmasebi S, Jafarnejad SM, Tam IS, Gontopoulos-Pournatzis T, Matta-Camacho E, Tsukumo Y, Yanagiya A, Li WC, Atlasi Y, Caron M, Braunschweig U, Pearl D, Khoutorsky A, Gkogkas CG, Nadon R, Bourque G, Yang XJ, Tian B, Stunnenberg HG, Yamanaka Y, Blencowe BJ, Giguère V, Sonenberg N. Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc Natl Acad Sci USA*, 2016, 113(44): 12360–12367.
- [25] Weatheritt RJ, Sterne-Weiler T, Blencowe BJ. The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol*, 2016, 23(12): 1117–1123.
- [26] Sun SY, Zhang Z, Sinha R, Karni R, Krainer AR. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol*, 2010, 17(3): 306–312.
- [27] Thiele A, Nagamine Y, Hauschildt S, Clevers H. AU-rich elements and alternative splicing in the beta-catenin 3'UTR can influence the human beta-catenin mRNA stability. *Exp Cell Res*, 2006, 312(12): 2367–2378.
- [28] Nourse J, Spada S, Danckwardt S. Emerging roles of RNA 3'-end cleavage and polyadenylation in pathogenesis, diagnosis and therapy of human disorders. *Biomolecules*, 2020, 10(6): 915.
- [29] Chen W, Jia Q, Song YF, Fu HH, Wei G, Ni T. Alternative polyadenylation: methods, findings, and impacts. *Genomics Proteomics Bioinformatics*, 2017, 15(5): 287–300.
- [30] Jambhekar A, Derisi JL. Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA*, 2007, 13(5): 625–642.
- [31] Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*, 2017, 18(1): 18–30.
- [32] Lau JS, Yip CW, Law KM, Leung FC. Cloning and characterization of chicken growth hormone binding protein (cGHBP). *Domest Anim Endocrinol*, 2007, 33(1): 107–121.
- [33] Ning BL, Huang JX, Xu HD, Lou YQ, Wang WS, Mu F, Yan XH, Li H, Wang N. Genomic organization, intragenic tandem duplication, and expression analysis of chicken TGFBR2 gene. *Poult Sci*, 2022, 101(12): 102169.
- [34] Wu QS, Wright M, Gogol MM, Bradford WD, Zhang N, Bazzini AA. Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J*, 2020, 39(17): e104763.
- [35] Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*, 2014, 33(9): 981–993.
- [36] Dodbele S, Wilusz JE. Ending on a high note: downstream ORFs enhance mRNA translational output. *EMBO J*, 2020, 39(17): e105959.

(责任编辑: 赵要风)