遗传 HEREDITAS (Beijing) 2012 年 12 月, 34(12): 1577－1582
ISSN 0253-9772 www.chinagene.cn

研究报告

DOI: 10.3724/SP.J.1005.2012.01577

人与小鼠核糖体蛋白基因内含子中的转录调控位点分析

李慧敏[1, 2], 陈丹[2, 3]

1. 云南民族大学数学与计算机科学学院, 昆明 650031;
2. 云南大学数学与统计学院, 昆明 650091;
3. 云南大学生物资源保护和利用重点实验室, 昆明 650091

摘要: 前期对酵母和果蝇核糖体蛋白(Ribosomal protein, RP)基因内含子序列中的寡核苷酸分析表明, 内含子中含有潜在的转录因子结合位点。为进一步发掘核糖体蛋白基因内含子参与转录调控的证据, 文章首先基于频率分析方法抽提出人和小鼠核糖体蛋白基因第一内含子中高频(Over-represented)出现的寡核苷酸片段 (亦称模体, Motif), 这些寡核苷酸中超过 85%与已知的转录因子结合位点吻合, 是潜在的转录调控元件。对抽提出的寡核苷酸进行碱基组成分析, 发现 95%以上的寡核苷酸富含碱基 C 和 G, 而较少富含 A 和 T。从寡核苷酸在内含子中的分布情况看, 它们相对靠近第一内含子的 5′端, 即距离基因转录起始位点和上游区域较近。推测这些特征可能与基因转录调控有关。

关键词: 人; 小鼠; RP 基因; 内含子; 转录调控

Analysis of transcriptional regulatory sites in introns of human and mouse ribosomal protein genes

LI Hui-Min[1, 2], CHEN Dan[2, 3]

1. *School of Mathematics and Computer Science*, *Yunnan University of Nationalities*, *Kunming* 650031, *China*;
2. *School of Mathematics and Statistics*, *Yunnan University*, *Kunming* 650091, *China*;
3. *Laboratory for Conservation and Utilization of Bio-resources*, *Yunnan University*, *Kunming* 650091, *China*

Abstract: Previous studies from oligonucleotides in the ribosomal protein (RP) genes of the yeast and fruitfly indicated that the potential transcriptional regulatory sites are located in the introns of the genes. The transcriptional regulatory sites in introns are still poorly understood. To explore the functional significance of transcriptional regulation of introns, we extracted over-represented oligonucleotides (also known as motifs) in the first introns of the human and mouse ribosomal protein genes by statistical comparative analysis, and found that over 85% of these oligonucleotides were consistent with the known transcriptional factor binding sites, which might be potential transcriptional regulatory elements. By analyzing

收稿日期: 2012–04–19; 修回日期: 2012–06–10
基金项目: 国家自然科学基金项目 (编号：31160181), 云南省应用基础研究基金项目(编号：2007A023M) 和云南民族大学引进人才科研项目资助
作者简介: 李慧敏, 博士, 副教授，研究方向：生物统计学与生物信息学。E-mail: lihuimin_1980@126.com
通讯作者: 陈丹, 博士研究生, 讲师, 研究方向：生物数学与生物信息学。E-mail: danchen@ynu.edu.cn
李慧敏和陈丹为共同通讯作者。
网络出版时间: 2012/10/24 17:06:36
URL: http://www.cnki.net/kcms/detail/11.1913.R.20121024.1706.002.html

the base compositions of these elements, we found that a majority (>95%) of the detected motifs were rich in C and G and only a few of them were rich in A and T. Moreover, the oligonucleotides were close to the 5'-ends of the first introns (the distances between the motifs and the transcriptional start sites or upstream regions of genes are short). We speculated that the properties of over-represented motifs in the first introns might be associated with the transcriptional control.

Keywords: human; mouse; RP gene; intron; transcriptional regulation

转录调控是真核基因表达过程中重要的环节，目前对真核基因转录调控的研究主要集中在基因的上游区域[1~3]。内含子(Intron)作为非编码序列的重要成员，近年来，越来越多的实验表明基因的内含子(特别是第一内含子)中也含有转录调控元件，这些元件可以作为启动子(Promoter)或增强子(Enhancer)发挥作用，或两者兼有[4~6]。因此，有必要对内含子中的转录调控元件进行研究。

一个生物体含有成千上万的基因，若对所有基因内含子中的调控元件进行分析，其难度不言而喻。核糖体蛋白(Ribosomal protein, RP)基因是生物体内蛋白质合成的重要器件，一般具有高表达、进化上保守和共调控的性质[7]，用计算分析易于找出这些基因中非随机出现的序列片段。因此可利用它初步考察基因内含子参与转录调控的可能性，以及分析内含子中调控元件的性质。前期我们对两种模式生物(酵母和果蝇)RP基因内含子进行过分析，发现它们中确实含有潜在的正调控元件[8~11]。酵母(*Saccharomyces cerevisiae*)和果蝇(*Drosophila melanogaster*)均为低等真核生物，它们内含子参与转录调控的事实对研究这些生物内含子的功能提供了有价值的参考。

哺乳动物(Mammalian)属于高等真核生物，也有一些关于其内含子具有转录调控功能的实验报道。如，研究发现小鼠*c-fos*基因的第一个内含子中包含一个具有转录终止功能的长序列[12]；若敲除小鼠*RPL32* 基因第一个外显子和第一个内含子中的两个YY1 的结合位点，其转录速率则降低 90%[13]；在人*RPS6* 基因的第一个内含子中探测到转录因子YY1 的结合位点[14]。然而，由于哺乳动物基因序列及其转录调控的复杂性，对其内含子转录调控功能的研究还处于模糊的探索阶段。本文将在前期研究的基础上，以人(*Homo sapiens*)和小鼠(*Mus musculus*)的RP基因为例，挖掘哺乳动物RP基因内含子中转录调控元件的特征，期望为内含子参与转录调控提供更多的理论证据。

1 样本和方法

1.1 样本

由于实验分析结果主要表明基因的第一内含子中含有转录调控元件，因此本文针对第一内含子序列进行分析。人和小鼠 RP 基因第一内含子序列取自核糖体蛋白基因数据库(RPG: http://ribosome.med.miyazaki-u.ac.jp/)，分别获得 80 和 79 条第一内含子序列。作为对照集，从 NCBI GenBank 数据库(http://www.ncbi.nlm.nih.gov)中分别取出人和小鼠所有基因的非编码序列。

1.2 提取潜在转录调控模体

利用频率分析方法[15]提取RP基因启动子中潜在的转录调控模体。设S_1和S_2分别为RP基因第一内含子序列和对照序列。某长度为l的寡核苷酸w在S_1和S_2中出现的次数分别记为$n_1(w)$和$n_2(w)$，n_1和n_2分别表示所有形式的l核苷酸在S_1和S_2中出现的总次数。w在序列集S_1和S_2中出现的频率分别为$n_1(w)/n_1$和$n_2(w)/n_2$，记为$f_1(w)$和$f_2(w)$。由于绝大多数转录因子在两条链上都具有活性[16]，寡核苷酸的统计在DNA双链上进行（方向为 5′→3′)。采用单边假设检验法进行检验,即零假设H_0：$f_1(w)\leqslant f_2(w)$，备择假设H_a：$f_1(w)> f_2(w)$。$f_1(w)$和$f_2(w)$的差异$U(w)$按以下公式计算：

$$U(w)=\frac{f_1(w)-f_2(w)-\delta_1\times 0.5/n_1-\delta_2\times 0.5/n_2}{\sqrt{\frac{n_1(w)+n_2(w)}{n_1+n_2}\left(1-\frac{n_1(w)+n_2(w)}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

其中

$$\delta_i=\begin{cases}0, & \text{当} n_i(b)\geqslant 30\text{时}\\ 1, & \text{当} n_i(b)<30\text{时}\end{cases},\qquad i=1,2$$

即当$n_1(b)$或$n_2(b)$小于 30 时, 对$U(w)$值进行连续性矫正。当$n_1(w)$和$n_2(b)$都小于 5 时, 直接用二项分布计算概率$P(f_1(w)\leqslant f_2(w))$。

取显著性水平$\alpha=10^{-4}$, 即当$U(w)>3.70$ ($P<10^{-4}$)时, 拒绝零假设H_0接受备择假设H_a, 此时认为w在序列集S_1中的出现频率显著高于在S_2中的出现频率, 是S_1中潜在的转录因子结合位点(Transcriptional factor binding site, TFBS)。$U(w)$值越大, 差异越显著。

2　结果与分析

2.1　内含子中的 6 核苷酸

由于真核基因转录因子结合位点的核心序列长度一般为 6~10 bp[17], 本文主要考察 6 核苷酸。6 核苷酸共有 4 096 种形式, 按照$u>3.70$, 同时为了使抽提出的模体具有一定的普遍性, 要求 6 核苷酸在RP序列中出现条数大于总RP基因条数的 1/4。结果在人和小鼠RP基因第一内含子中分别抽提出 363 和 373 个高频出现的模体 (限于篇幅, 表 1 只列出一部分), 其中有 269 个模体在两种RP基因内含子中共有。

按照Tompa建立的序列片段与转录因子 (Transcriptional factor, TF)进行匹配的准则[18], 将探测的模体与TRANSFAC中相应物种(人或小鼠)的调控位点进行对照, 发现分别有 91.7%和 86.1%的模体与已知结合位点吻合。如, CTCCGA($U_h(w)$ =11.93, P_h=4.29e-33, 其中$U_h(w)$和P_h分别表示模体 w 在小鼠RP基因中的$U(w)$值和P值)、CTCCGT($U_m(w)$ =8.69, P_m=1.81e-18, 其中$U_m(w)$和P_m分别表示模体w在小鼠RP基因中的$U(w)$值和P值)和ACTCAC($U_h(w)$ = 6.05, P_h=7.36e-10; $U_m(w)$=4.47, P_m=3.98e-6)是AP-1 潜在的结合位点 (AP-1 结合位点RSTGACTNMNW,其中R：A或G; S：C或G; N：A,C,G或T; M：C或A; W：A或T), CGCGCC($U_h(w)$=17.92, P_h=3.93e–72; $U_m(w)$=35.36, P_m=3.2e–274)是E2F结合位点TTTSG CGCG MNR中的序列片段。实验已经证实在人和小鼠的某些基因内含子中存在YY1 的结合位点CCATNTWN NNW或NCGGCCATCTTGNCTSNW, 而两物种RP基因中共有高频模体GCGGCC($U_h(w)$= 30.41, P_h= 2.27e–203; $U_m(w)$=58.09, P_m<1e–275)和 CGGCCA ($U_h(w)$=11.6, P_h=2.05e–31; $U_m(w)$=12.06, P_m=8.71e-34) 恰好是YY1 结合位点的一部分。在脊椎动物中出现最频繁的转录因子为 SP1, 其结合位点序列之一为GGGCGGGG。我们探测到大量SP1 以及SP1-related 结合位点, 并且它们都具有较高的$U(w)$值, 这表明本文抽提的绝大部分 6 核苷酸是潜在的转录因子结合位点。

2.2　高频 6 核苷酸碱基组成分析

如果 6 核苷酸中有 4 个或 4 个以上碱基为 C 或 G, 称该 6 核苷酸富含 CG; 若有 4 个或 4 个以上碱基为 A 或 T, 则称该 6 核苷酸富含 AT。结果发现, 在人和小鼠 RP 基因第一内含子中探测到的 6 核苷酸中, 分别有 96.68%和 99.46%富含 CG。特别地, 所有 64 种形式为 SSSSSS(S：C 或 G)的 6 核苷酸中, 分别有 59 和 58 个在人和小鼠中被抽提出来, 并且具有较高的$U(w)$值。进一步将 6 核苷酸分成 4 类：(1)6 核苷酸全部由碱基 C、G 构成; (2)6 核苷酸中有 5 个位置由 C、G 构成, 只有 1 个位置为 A 或 T; (3)6 核苷酸中有 4 个位置是 C、G, 2 个位置是 A 或 T 和(4)其他。分析表明, 在人和小鼠中, 分别有 16.3%和 15.59%第一类模体; 56.35%和 67.2%第 2 类模体; 24.03%和 16.67%第 3 类模体; 只有不到 4%的模体为第 4 类。这表明 CG-rich 元件在 RP 基因转录调控过程中可能发挥着重要作用。同时, 各类模体在人和小鼠 RP 基因中的碱基组成差别不大, 表明两种哺乳 RP 基因内含子中的调控元件具有保守性。

2.3　6 核苷酸在内含子序列中的分布

尽管本文只考察 6 核苷酸, 但实际的调控元件往往是由较长的序列组成。为了解调控元件在序列中的分布情况, 将探测到的 6 核苷酸在 RP 基因第一内含子序列和非编码序列中进行比较。发现由它们重叠或连接可以获得一些长序列片段, 并且这些长序列片段中有一些与已知转录因子结合位点相似。如, 在小鼠基因中探测到 CGGCGC、GCGGCC 和 GCCATC, 它们形成的长片段 CGGCGGCCATC 是 YY1 潜在的结合位点。其次, 高频寡核苷酸在第一内含子序列中形成长序列片段的密度较大：它们在人和小鼠 RP 基因内含子中的平均长度分别为 12 bp 和 13 bp, 平均密度为 37.8%和 45.5%; 而在非编码序列中的平均长度和平均密度分别为 8 bp 和 7 bp 以

表 **1** 人和小鼠 **RP** 基因中部分高频出现的模体信息

高频模体(人)	*U*(*w*)值(人)	出现基因数(人)	*P* 值(人)	高频模体(小鼠)	*U*(*w*)值(小鼠)	出现基因数(小鼠)	*P* 值(小鼠)
CGCGGC	39.25	38	<1e-274	CGCGGC	75.78	43	<1e-274
CGGCGC	34.60	34	9.75e-263	CCGCGG	70.90	29	<1e-274
GCGGCC	30.40	42	2.27e-203	CCCGCG	66.81	39	<1e-274
CCGCGG	29.76	21	6.30e-195	CGCCGC	65.22	43	<1e-274
CGGGCC	29.17	50	1.87e-187	CCGCCG	58.85	43	<1e-274
CCGGCG	29.03	32	1.24e-185	GCGGCC	58.09	50	<1e-274
CGCCGC	28.65	41	1.06e-180	CGGCCG	55.63	24	<1e-274
CCGCCG	26.36	38	1.82e-153	GCCGCC	52.84	47	<1e-274
CGGCCC	25.95	51	7.76e-149	CCGCGC	52.68	37	<1e-274
CCCCGG	23.76	43	3.83e-125	CGCCCG	51.35	41	<1e-274
CCGCGC	23.76	37	3.98e-125	GCCCGC	48.38	43	<1e-274
GCCGCC	22.61	45	1.71e-113	CCGGCG	44.35	34	<1e-274
CCCGCA	22.03	44	6.09e-108	CGGCCC	44.19	43	<1e-274
CCCGCG	21.85	30	3.57e-106	CGCGAC	43.78	24	<1e-274
CGGGGC	21.59	48	1.05e-103	CGGGCC	42.31	40	<1e-274
CCCGGG	20.45	38	2.61e-93	CCGGGC	40.52	47	<1e-274
GGCGCC	20.24	26	2.01e-91	CCCGGC	40.41	49	<1e-274
CCGGGC	19.80	46	1.38e-87	CGGCGC	40.07	28	<1e-274
CCGCAG	19.10	41	1.11e-81	CCGCGA	39.57	25	<1e-274
GCCCGC	18.97	43	1.42e-80	CCGGCC	38.07	42	<1e-274
CCGGAG	18.89	36	6.35e-80	CGCGCC	35.36	27	3.20e-274
GCCGCA	18.43	31	3.42e-76	ACGCGG	34.63	22	4.67e-263
CCGGCC	18.25	43	8.75e-75	CGGGGC	34.55	39	6.27e-262
CGCGCC	17.92	34	3.93e-72	GCCGGC	33.27	26	4.47e-243
CCGAAC	17.72	25	1.58e-70	CCCGGG	33.00	29	4.34e-239
AGCCCG	17.56	33	2.63e-69	CCGCCC	32.32	45	1.60e-229
CCGCAC	17.28	35	3.40e-67	CCCCGC	32.18	48	1.55e-227
CCCGGA	16.87	29	3.72e-64	CCCCGG	31.31	40	1.89e-215
CCCGGC	16.87	46	3.76e-64	AGGCCG	31.11	36	9.10e-213
CGGAGC	16.19	34	3.18e-59	CGGCGA	30.62	24	3.35e-206
CGCCCG	16.10	38	1.33e-58	CCCGCC	30.52	42	7.74e-205

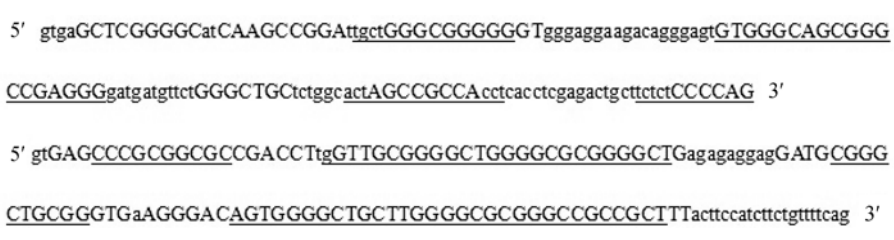

图 **1** 高频模体在人和小鼠 **RP** 基因第一内含子中的定位(以 *RPS13* 基因序列为例)

大写字母代表探测到的高频出现模体; 下划线部分表示该段模体与已知的转录因子结合位点相似。

及 14.2%和 12.8%。这表明, 潜在调控模体在 RP 基因第一内含子中的分布更集中。例如, 图 1 显示了人和小鼠 *RPS13* 基因内含子的情况。

为进一步了解调控元件在内含子中的分布情况,

将每条内含子序列分成三部分。沿着 5′→3′的方向, 各部分序列的长度分别为l_1、l_2和l_3。其中l_1、l_2和l_3按如下定义: 若第i条序列的长度为L_i, $l_1=l_2=\begin{cases} l & L_i=3l \text{ 或 } 3l+1 \\ l+1 & L_i=3l+2 \end{cases}$, $l_3=L_i-l_1-l_2$。我们分别称各条子序列为第一内含子的 5′端(5′-end)、5′/3′区域 (5′/3′region) 和 3′端 (3′-end) 。计算发现, 6 核苷酸重叠或连接形成的长片段在 5′端具有最长的长度以及最大的密度, 而在 3′-端的平均长度和平均密度均相对最小(表 2), 这提示潜在调控元件更靠近第一内含子的 5′端。由于人和小鼠第一外显子普遍较短[19], 因此我们认为调控元件距离基因转录起始位点或上游区域较近。

表 2 由高频出现的 6 核苷酸形成的长片段在基因序列中的平均长度和密度

		5′-端	5′/3′区域	3′-端
人	长度(bp)	12	11	9
	密度(%)	53	35	25
小鼠	长度(bp)	14	12	9
	密度(%)	64	43	28

2.4 与 MEME 算法结果的比较分析

要对转录调控元件的信息有比较准确而全面地认识, 还需要综合多种方法的结果。MEME[20] (Multiple EM for Motif Elicitation) 算法是在生物序列模式识别时应用最广泛的工具之一, 其原理是反复应用期望最大(Expectation maximization, EM)算法搜索模体, 输入模体长度和期望出现的模体数目, MEME将输出所有符合条件的模体, 直到它认为没有统计上显著的模体为止。

为了保持研究的一致, 输入模体长度 6, 模体个数为 300, 在人和小鼠 RP 基因第一个内含子中, 分别得到 40 和 56 个显著的 6 核苷酸 (表 3)。分析发现, 这些 6 核苷酸虽然与频率分析方法抽提到的不完全相同, 但是一般相差一个碱基。也就是说, 它们在碱基构成上极为相似。将这些 6 核苷酸与 TRANSFAC 数据库中已知转录因子结合位点进行对比发现, 它们绝大部分与已知的转录因子结合位点吻合。

值得注意的是, MEME 算法抽提到了一些富含 A、T 的元件, 而由于它们的 $U(w)<3.70$ 而未被频率分析方法探测到。这或许是因为 MEME 算法的对照序列是由原序列随机打乱(Shuffled)后得到, 而本文频率分析方法的对照集则是基因的非编码序列, 它本身也包含内含子序列。事实上, 用频率分析方法抽提出的 6 核苷酸在 RP 基因内含子中的出现频率显著高于非编码序列, 属于 RP 基因第一内含子特有的。

3 讨 论

本文以人和小鼠 RP 基因为例, 分析了哺乳动物 RP 基因第一内含子中的转录调控元件特征。首先, 利用频率分析方法探测到一批高频出现的模体, 这些模体绝大部分与实验上已知的转录因子结合位点吻合, 是潜在的调控元件。据此我们推测, 在人和小鼠 RP 基因第一内含子中确实存在转录调控元件。换句话说, 第一内含子具有转录调控的功能, 而不只是转录后被剪切, 这与一些实验的结果基本一致。

内含子中转录调控元件的特征可以从另一个角度说明其在基因转录调控过程中的重要作用。我们对抽提出的模体的碱基组成进行分析表明: 绝大多

表 3 利用 MEME 算法探测到的模体

物种	模体
人	AAAAAA,AAAAAG,AAAAAT,AAAACA,AAAATA,AAGAAA,AGAAAA,AGAAAG,AGAAAT,AGAACA,AGAGAG,AGAGGA,AGGAAA,AGGAGA,CCCGCC,CCTGCC,**CTCTCC**,CTTTCC,GCCGCC,GCCTCC,**GCTCCC**,GCTTCC,GGAAAG,GGAAGA,GGAAGC,GGAGAG,GGAGGA,GGAGGC,GGCGCC,GGCTCC,GGGAAA,GGGAAC,GGGAAG,GGGAGA,GGGAGG,GGTCGG,GTGAGT,TGCTCG,TGCTGG,TTCCCA
小鼠	AAAAAA,AAAACA,AAACAG,**AAAGAA**,AAAGCC,AAGAAA,ACCAAC,AGAAAA,AGAAAG,AGAGAA,AGAGAG,AGGAAG,AGGGAG,CAGAGC,CAGGCC,CCGGGC,CGAGCC,**CTCTCC**,CTGAAA,CTGAAT,CTGCGG,CTGTGT,GAAAAA,GAGAGG,GCCGGC,GCGGGG,GGAAAG,GGAAGG,GGAGAA,GGAGAC,GGAGAG,GGAGCC,GGATGG,GGCCGC,GGCTGG,GGGAGG,GGGAGT,GGGCGG,GGGGAA,GGGGCC,GTGAGA,GTGAGC,GTGAGT,GTTCTC,GTTCTG,TAACTT,TCCCTT,TCTCTG,TCTCTT,TTAAAA,TTAGAA,TTTAAA,TTTCAG,TTTTCT,TTTTTC,GGAGCC

注: 方框内模体表示在 TRANSFAC 数据库中未出现的片段。

数模体富含碱基C或G, 较少富含A或T。研究表明, 多数情况下, 许多含C、G的元件都是转录调控的关键因素, 是许多重要转录因子 (如SP1) 的结合位点, 因此, 我们抽提的 6 核苷酸富含C和G这一特征有利于某些转录因子的结合。对高频模体在内含子中的分布分析表明, 它们相对倾向于内含子的 5′-端。结合人和小鼠RP基因第一个外显子普遍较短的事实, 也可以认为它们比较靠近基因上游区域。已经知道, 基因上游区域是转录调控的核心, 其中含有较多的转录调控元件。此外, 基因的转录通常需要多个转录因子的协同作用, 协同作用的转录因子结合位点距离一般不能太远[21]。从这个意义上讲, 内含子中潜在调控元件靠近基因上游的现象有利于上游和内含子中位点之间的协同作用。

本文结果从理论上分析了第一内含子中转录调控元件的特征, 进一步支持了内含子参与转录调控的推测, 为理解内含子的功能和哺乳动物 RP 基因的转录调控机制提供了有价值的参考。基因转录调控的一个重要特征是组合调控, 接下来的工作是揭示蕴含在内含子中的组合调控规律。

参考文献(References):

[1] Hu Z, Gallo SM. Identification of interacting transcription factors regulating tissue gene expression in human. *BMC Genomics*, 2010, 11(1): 49. DOI

[2] Hu ZH, Hu BY, Collins JF. Prediction of synergistic transcription factors by function conservation. *Genome Biol*, 2007, 8(12): R257.1–R257.20. DOI

[3] Yu XP, Lin J, Masuda1 T, Esumi N, Zack DJ, Qian J. Genome-wide prediction and characterization of interactions between transcription factors in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 2006, 34(3): 917–927. DOI

[4] Curi GC, Chan RL, Gonzalez DH. The leader intron of *Arabidopsis thaliana* genes encoding cytochrome *c* oxidase subunit 5c promotes high-level expression by increasing transcript abundance and translation efficiency. *J Exp Bot*, 2005, 56(419): 2563–2571. DOI

[5] Choi J, Newman AP. A two-promoter system of gene expression in *C. elegans*. *Dev Biol*, 2006, 296(2): 537–544. DOI

[6] Gazave E, Marqués-Bonet T, Fernando O, Charlesworth B, Navarro A. Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol*, 2007, 8(2): R21.1–R21.13. DOI

[7] Hu HY, Li XM. Transcriptional regulation in eukaryotic ribosomal protein genes. *Genomics*, 2007, 90(4): 421–423. DOI

[8] 张昆林, 张静, 罗静初. 酵母基因上游与内含子可能存在的转录协同作用. 生物化学与生物物理进展, 2005, 32(1): 46–52. DOI

[9] Hu J, Li HM, Zhang J. Analysis of transcriptional synergy between upstream regions and introns in ribosomal protein genes of yeast. *Comput Biol Chem*, 2010, 34(2): 106–114. DOI

[10] 胡俊, 张静. 酵母基因内含子中二聚体寡核苷酸转录调控位点的统计分析. 生物化学与生物物理进展, 2004, 31(5): 449–454. DOI

[11] 李慧敏, 胡俊, 张静. 果蝇核糖体蛋白基因中潜在转录协同作用模体的统计分析. 云南大学学报 (自然科学版), 2010, 32(3): 338–345. DOI

[12] Mechti N, Piechaczyk M, Blanchard JM, Jeanteur P, Lebleu B. Sequence requirements for premature transcription arrest within the first intron of the mouse c-fos gene. *Mol Cell Biol*, 1991, 11(5): 2832–2841. DOI

[13] Chung S, Perry RP. The importance of downstream δ-factor binding elements for the activity of the rpL32 promoter. *Nucleic Acids Res*, 1993, 21(14): 3301–3308. DOI

[14] Antoine M, Kiefer P. Functional characterization of transcriptional regulatory elements in the upstream region and intron 1 of the human S6 ribosomal protein gene. *Biochem J*, 1998, 336 (2): 327–335. DOI

[15] Zhang J, Hu J, Shi XF, Cao H, Liu WB. Detection of potential positive regulatory motifs of transcription in yeast introns by comparative analysis of oligonucleotide frequencies. *Comput Biol Chem*, 2003, 27(4-5), 497–506. DOI

[16] van Helden J, André B, Collado-Vides J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol*, 1998, 281(5): 827–842. DOI

[17] Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res*, 1996, 24(1): 238–241. DOI

[18] Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu YT, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Régnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*, 2005, 23(1): 137–144. DOI

[19] Li HM, Zhang J. Comparison of promoter sequences in the eukaryotic ribosomal protein genes. In: 2008 International Conference on Bioinformatics and Biomedical Engineering. Shanghai: IEEE, 2008: 180–183. DOI

[20] Bailey TL, Williarns N, Misleh C, Li WW. MEME: discover-

ing and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*, 2006, 34(Web Server issue): W369-W373. DOI

[21] Griffith J, Hochschild A, Ptashne M. DNA loops induced by cooperative binding of λ repressor. *Nature*, 1986, 322(6081): 750–752. DOI