人核受体 *nr5a2(hb1f)* 基因组序列的生物信息学分析

张成康^{①*} 林伟^{②*} 蔡彦宁^① 董辉^② 李娟^① 付刚^② 谢幼华^①
黄谷杨^② 汪垣^{①†}

(^①中国科学院上海生命科学研究院, 中国科学院上海生物化学和细胞生物学研究所, 上海 200031; ^②国家人类基因组南方研究中心, 上海 201203. * 同等贡献. † 联系人, E-mail: wangyuan@server.shnc.ac.cn)

摘要 在克隆了人核受体基因 *nr5a2(hb1f)* 的 cDNA 的基础上, 测定了该基因的全基因组序列. 采用生物信息学手段对 *nr5a2(hb1f)* 基因组序列进行了深入分析. 序列同源性比较及编码序列预测结果提示, 在测定的 210 kb 基因组序列中, 特别是人 *nr5a2(hb1f)* 基因的内含子中可能还存在其他的编码信息. 通过比较基因组方法分析不同生物 *nr5a* 基因的结构, 发现各种生物 *nr5a* 基因的结构具有保守性, 并且在进化过程中有明显的基因重复现象. 对斑马鱼、小鼠和人 *nr5a2* 基因上游序列的比较, 显示这些生物 *nr5a2* 基因的启动子区域相当保守, 提示不同生物 *nr5a2* 基因可能受到同一类转录因子的调控.

关键词 核受体 *nr5a2(hb1f)* 基因结构 生物信息学 比较基因组学 启动子

核受体超家族是目前已知的最大的一类真核转录因子家族, 包含众多在基因调控、细胞分化和生长发育中起重要作用的真核转录因子^[1~3]. 人核受体 NR5A2(hB1F^[4], 又名 hFTF^[5], CPF^[6])是核受体超家族的一个新成员, 属于其中的 NR5A 亚家族^[7]. NR5A2 是一个重要的真核转录因子, 对多种肝脏特异基因, 如 α -甲胎蛋白(*afp*), 胆固醇 7 α -羟化酶(*cyp7a1*), 甾醇 12 α -羟化酶(*cyp8b1*)和乙肝病毒增强子 II (HBV EN II)的转录调控起着重要的作用. 最近的研究还发现, NR5A2 在胆酸合成途径及维持胆固醇的生理平衡中起着重要的调控作用^[8~10]. 除此以外, NR5A2 还可能调控几种重要的肝细胞转录因子基因, 如 *hnf1 α* , *hnf3 β* 和 *hnf4 α* 的转录^[11,12]. 这些结果都说明, NR5A2 在生物基因表达调控及其生长发育中具有重要的作用. 获得人 *nr5a2(hb1f)* 基因的全部基因组序列并分析其结构, 对于研究揭示其表达规律和生理功能具有重要的意义.

我们实验室在国际上首先克隆了人 *nr5a2(hb1f)* 基因的 cDNA, 并研究了 NR5A2(hB1F)对乙肝病毒增强子 II (HBV EN II)的转录调控^[4,13]. 以这个 cDNA 为探针, 我们从人的基因组文库中筛选到包含人 *nr5a2(hb1f)* 全基因的 BAC 克隆和包含人 *nr5a2(hb1f)* 基因 5' 上游序列的噬菌体克隆, 利用鸟枪法测定了人 *nr5a2(hb1f)* 基因的全部序列, 并对该基因的启动子做了初步分析^[14]. 本文采用生物信息学手段对人 *nr5a2(hb1f)* 的基因组序列进行了深入分析. 序列同

源性比较及编码序列预测结果提示人 *nr5a2(hb1f)* 基因的内含子中可能存在其他编码信息. 对线虫、果蝇、斑马鱼、小鼠和人的 *nr5a* 基因结构的比较发现, 各种生物的 *nr5a* 基因具有保守的内含子/外显子结构. 各种生物的 *nr5a2* 基因及与其同源的 *nr5a1* 基因可能来源于线虫中的同一个祖先基因. 从低等生物到高等生物, *nr5a* 基因在进化过程中具有明显的基因重复现象, 并表现出显著的增大趋势. 对几种生物 *nr5a2* 基因启动子的比较, 为深入了解 *nr5a2* 基因的调控机制提供了有益的信息.

1 材料与方法

(i) 人 *nr5a2(hb1f)* 基因组序列的生物信息学分析. 将人 *nr5a2(hb1f)* 基因的 cDNA 序列^[4]与基因组序列^[14]比较, 得到人 *nr5a2(hb1f)* 基因的结构. 以每 500 bp 为单位, 计算测定的基因组序列的碱基组成和 CpG 二核苷酸的分布. 利用 RepeatMasker 搜寻基因组序列中的重复序列, 并将发现的重复序列用连续的“N”替换, 得到“过滤序列”(filtered sequence). 利用 BLASTN 将“过滤序列”与 NCBI 的非冗余库(NR)、人的表达序列标签库(EST)和人的 UniGene 库相比较, 寻找基因组序列中可能存在的编码信息. 将这些信息与利用 GENSCAN, MZEF 和 FGENESH 等基因识别软件预测的可能编码信息相比较, 寻找基因组序列中可能包含的其他编码信息.

(ii) 不同生物 *nr5a* 基因结构的比较. 利用

TBLASTN, 将 NR5A 亚家族各成员中高度保守的 Ftz-F1 盒结构(KLEAVRADMRGGRNKFGPMYKR-DRA)与果蝇和线虫的全基因组序列及已公布的其他物种(斑马鱼、小鼠、大鼠和人)的基因组序列进行比较, 每种生物中与 Ftz-F1 盒结构匹配最佳的基因组序列被认为包含了该生物的 *nr5a* 基因. 利用 BLASTN 和 SIM4 将每种生物的 *nr5a* 基因组序列与已经报道的该生物 *nr5a* 基因的 cDNA 序列相比较, 得到每种生物 *nr5a* 基因的结构, 对不同生物 *nr5a* 基因的外显子、内含子的数目及大小进行比较分析.

(iii) 不同生物 *nr5a2* 基因 5' 上游序列的比较. 从斑马鱼、小鼠和人的 *nr5a2* 基因组序列中取出 1 kb 的 5' 上游序列, 利用 ClustalW 对多个序列进行联配分析(alignment). 并利用 MatInspector 寻找保守序列中可能存在的转录因子结合位点.

2 结果与讨论

2.1 人 *nr5a2(hb1f)* 基因组序列的生物信息学分析

人 *nr5a2(hb1f)* 基因长约 150 kb, 由 8 个外显子和 7 个内含子组成^[14](图版 I, 附本刊后, 下同), 是哺乳动物中第一个完成全基因组序列测定的 *nr5a* 亚家族成员. 在全部测定的 210 kb 基因组序列中, 利用 RepeatMasker 发现了 149 个属于 SINE 家族的重复序列(124 个 Alu, 25 个 MIR), 42 个属于 LINE 家族的重复序列(14 个 LINE1, 28 个 LINE2), 10 个 LTR 型的重复序列(7 个 MaLR, 3 个逆转录病毒类似序列)以及 18 个 DNA 元件(9 个 MER1 类型, 7 个 MER2 类型和 2 个 Mariners). 除此以外, 还发现了 46 个简单重复元件和 28 个低复杂度元件. 其中, 以 2 个最大的内含子(分别为 62 和 53 kb)中所包含的重复序列的数目和种类最多(图版 I).

利用 BLASTN 将“过滤”的 210 kb 基因组序列和 NCBI 的非冗余库(NR)比较, 除了 *nr5a2(hb1f)* 基因的转录本以外, 未发现任何其他 cDNA. 但与人的序列表达标签库(EST)比较, 除了发现 *nr5a2(hb1f)* 基因的 EST 以外, 还发现在 2 个大内含子中存在许多 EST 的同源片段(图版 I). 用“过滤序列”搜寻人的 UniGene 库, 发现了 12 个 UniGene. 其中的 Hs.183123 就是 *nr5a2(hb1f)* 基因本身, 其他 11 个 UniGene 都仅由序列表达标签组成, 表 1 列出了 12 个 UniGene 的信息. 在 *nr5a2(hb1f)* 基因的 2 个大内含子中发现了 8 个

UniGene(Hs.319829, Hs.289242, Hs.256691, Hs.330657, Hs.114135, Hs.303971, Hs.190377, Hs.220981 和 Hs.158615), 其中的 5 个(Hs.289242, Hs.256691, Hs.330657, Hs.114135, Hs.158615)分布在互补链上. 这些 UniGene 的“表达特征”与人 *nr5a2(hb1f)* 基因一致. 另一个 UniGene Hs.13580 则位于 *nr5a2(hb1f)* 基因的下游. 在测定的 210 kb 基因组序列, 尤其是 *nr5a2(hb1f)* 基因的 2 个大内含子中发现如此众多的 UniGene 和 EST 同源区域, 提示其中很可能存在人 *nr5a2(hb1f)* 基因以外的其他编码信息. GENSCAN, MZEF 和 FGENESH 等基因识别软件的预测结果同样也暗示这段序列中可能包含其他编码信息(图版 I). 根据最新公布的人类基因组测序结果, 我们测定的 BAC 克隆位于 NT_004680 中, 其下游 190 kb 是一个 G 蛋白偶联的受体基因 *gpr-25*, 在 *nr5a2* 和 *gpr-25* 之间的区域内是否存在其他的基因, 也将随着研究的深入而得到解决.

2.2 不同生物 *nr5a* 基因结构的比较

对不同生物 *nr5a* 基因结构的比较发现, *nr5a* 基因在进化过程中有明显的基因重复现象(表 2). 线虫只有一个 *nr5a* 基因位点, 在果蝇中则有 2 个 *nr5a* 基因位点, 但其中的 *nr5a3* 基因在处于进化下游的生物中未再出现. 斑马鱼、小鼠和人 *nr5a* 基因很明显的经重复而成 *nr5a1* 和 *nr5a2* 两个基因. 另外, 从低等生物到高等生物, *nr5a* 基因表现出明显的增大趋势. 线虫的 *nr5a* 基因只有 4.7 kb, 而人的 *nr5a2(hb1f)* 基因大约有 150 kb. *nr5a* 基因在进化过程中的这种增大很可能是通过在内含子中不断插入重复序列而形成的.

比较各种生物的 *nr5a* 基因, 可以发现它们的基因结构具有非常保守的特性. NR5A 的 N 端可变区一般由一个或两个外显子编码; 负责结合靶位点的两个锌指(Zinc finger)各由一个外显子编码; Ftz-F1 盒, 绞链区和配体结合/转录激活域的一部分由一个大外显子编码; 最后几个外显子编码 NR5A 配体结合/转录激活域的大部分. 对 NR5A 转录激活功能很重要的 AF-2 区一般由最后一个外显子编码(图 1).

nr5a 基因结构的复杂性还表现在各种生物的 *nr5a* 基因大都具有选择性剪接(图 1). *nr5a* 基因的选择性剪接主要发生在 N 端的可变区和 C 端配体结合/转录激活域. 发生在 *nr5a* 的 N 端可变区的选择性剪接产生的不同异构体, 在转录激活功能上没有明显的差

表 1 210 kb 基因组序列中所包含的 UniGene 和部分 EST^{a)}

UniGene	长度 /bp	起点	终点	相似度/%	链	基因	成员	定位	表达特征
Hs.183123	3908	1	1217	100	F	nr5a2	AF190464, AB019246, XM_001441, AF146343, NM_003822, U80251, U93553, AF124247, AF228413, AF049102, H72424, H72864, R10138, AI240772, AI979325, N59515, N59115, AV732004, AU098581, AA627793	1q32.11	肾上腺, 生殖细胞, 肾, 子宫, 胚胎, 睾丸, 肝脏, 胰脏, 结肠
		12961	13099	100					
		17086	17197	100					
		18744	18887	100					
		21481	22131	100					
		84513	84637	100					
		94126	94273	100					
		147195	149033	100					
Hs.277964	369	6076	6444	100	F	ESTs	AW197405, BE046003		生殖细胞
Hs.319829	448	24226	24637	100	F	ESTs	H79446, H79331, AI292200		肺
Hs.289242	705	33730	34134	95	R	ESTs	AI078465, AV699722		
Hs.256691	453	48751	49196	100	R	EST	AI990476		生殖细胞
Hs.330657	231	48967	49196	95	R	EST	BE552190		生殖细胞
Hs.114135	524	65492	66015	100	R	ESTs	H78052, H78053, AA777855, AA702675	Chr1	肝脏, 脾脏
Hs.303971	392	85004	85188	100	F	EST	AI281443		淋巴
Hs.190377	954	102810	103354	99	F	ESTs	AI381423, AA704845, AI861938, AA203218		结肠, 胚胎
Hs.220981	478	121134	121612	85	F	ESTs	AI055813, AI992022, W38786, N92527		脑, 甲状旁腺
Hs.158615	313	134821	135120	100	R	EST	AI370040		胚胎
Hs.13580	1264	186870	187966	92	F	ESTs	AI796492, AI559684, AI864516, AI148734, AI784337, AI564254, AI685335, AW009759, AI368788, AI831840, AA910981, AI825679, AI811256, AI380138, AI337279, AI696023, AA454598, AI695982, T82322, AW079353, AI971714, F31492, AW362111, AA458517, N50335, AA307936, C00192		B-细胞, 血液, 中枢神经系统, 结肠, 眼, 生殖细胞, 肾, 肺, 淋巴, 肌肉, 卵巢, 前列腺, 子宫, 胚胎, 脑, 胰脏, 甲状腺

a) 表中列出了利用 BLASTN 将“过滤”的 210 kb 基因组序列与人 UniGene 数据库相比较所找到的 UniGene, 关于每个 UniGene 的详细信息可以在 <http://www.ncbi.nlm.nih.gov/UniGene> 上找到. 表中的起点和终点分别表示每个 UniGene 在 210 kb 基因组序列上的位置. F 和 R 分别代表正链和负链(以 *nr5a2* 基因所在的链为正链)

表 2 不同生物 *nr5a* 基因结构的比较

种名	基因	基因组序列	基因大小	cDNA	外显子	备注
线虫	<i>nr5a</i>	z54270	4.7 kb	<i>nhr-25</i> (AF179215)	9	基因组序列已经测定, 外显子的数目、大小确定
果蝇	<i>nr5a</i>	AE003519	约 48 kb	<i>ftz-fla</i> (M63711)	10	基因组序列已经测定, 外显子的数目、大小确定. <i>ftz-fla</i> 和 <i>ftz-flb</i> 源自不同启动子和剪接形式 ^[15]
				<i>ftz-flb</i> (M98397)	9	
	<i>nr5a3</i>	AE003669	约 33 kb	<i>Dmftz-f1B</i> (L06423)	9	DmFtz-F1B 的 Ftz-F1 盒与其他成员只有 57%的相同, 根据文献[16]报道, 数据库中尚无记录
斑马鱼	<i>nr5a1</i>	未发表		未发表		
	<i>nr5a2</i>	未发表	未确定	<i>zff1</i> (AF014926)	8	基因组序列尚未测定, 但外显子的数目、大小已经确定 ^[16]
小鼠	<i>nr5a1</i>	AC067926	未确定	<i>msf-1</i> (AB000490)	9	基因组序列尚未测定, 外显子的数目、大小尚不能确定. 文献[18]报道 <i>msf-1</i> 和 <i>melp3</i> 源自不同的启动子 ^[17]
	<i>nr5a2</i>	AC036121	未确定	<i>mlrh-1</i> (M81385)	(9?)	
						基因组序列尚不完全, 外显子的数目、大小尚未完全确定 ^[12]
人	<i>nr5a1</i>	NT_008491	约 20 kb	<i>hsf-1</i> (NM_004959)	9(6?)	基因组序列尚不完全, SIM4 确定的外显子的数目与文献[18]有差别
	<i>nr5a2</i>	NT_004680	约 150 kb	<i>hb1f</i> (NM_003822)	8	

<https://engine.scichina.com/doi/10.1360/csb2002-47-1-44>

图 1 *nr5a* 基因的典型结构特征

(a) NR5A 蛋白质的典型结构. NR5A 蛋白质从 N 端到 C 端可以分为可变区(A/B), DNA 结合区(C), 绞链区(D)和配体结合/激活区(E), II, III 代表在 NR5A 亚家族各成员中保守的两个区域. (b) *nr5a* 基因组的典型结构. *nr5a* 基因一般有 8~10 个外显子, 并且存在选择性剪接^[4,15,16]、不同启动子的选择性利用^[16,17]和不同 3' 加尾信号的选择性利用^[14]等多种变化方式. NR5A 的结构域和外显子之间具有明显的对应关系. 图中的箭头表示对不同启动子的选择性利用

别^[4,17]. 而 C 端配体结合/转录激活域内的选择性剪接通常表现为对最后一个外显子的选择性利用, 由此产生的不同异构体由于包含/不包含 AF-2 而表现出转录激活/抑制的活性^[16]. 人 *nr5a2* 基因还具有一种与不同的选择性剪接方式, 该剪接发生在编码绞链区的外显子中, 由此产生的异构体由于不含绞链区, 不能形成正常的空间结构, 因而不具有转录激活功能^[6].

2.3 不同生物 *nr5a2* 基因 5' 上游序列的比较

比较 1 kb 的斑马鱼、小鼠和人 *nr5a2* 基因的启动子区域^[12,14,16], 发现人和小鼠 *nr5a2* 基因的启动子序列相似性较高, 它们和斑马鱼 *nr5a2* 基因启动子的相似性较低(图 2). 3 种生物 *nr5a2* 基因的转录起始点附近有一段约 150 bp 的高保守区, 人和小鼠 *nr5a2*

图 2 人、小鼠和斑马鱼 *nr5a2* 基因的上游序列比较

利用 ClustalW 对人、小鼠和斑马鱼 *nr5a2* 基因的上游序列进行联配比较, 并利用 MatInspector 寻找保守序列中可能的转录因子结合位点. hB1F, 人 *nr5a2* 基因; mFTF, 小鼠 *nr5a2* 基因; zFF1, 斑马鱼 *nr5a2* 基因. 在 3 种生物中都保守的碱基用黑框标出, 在两种生物中保守的碱基则用灰框标出. 各种生物 *nr5a2* 基因的转录起始位点和翻译起始位点分别用箭头标出. 在 3 种生物中高度保守的一段 150 bp 序列用星号★标出, GATA 表示可能的 GATA 结合位点. 人和小鼠 *nr5a2* 基因上游序列中的保守区域用黑线框出. TSS 示转录起始位点, ATG 示翻译起始位点

基因启动子的这段序列相似性高达 86%。这段保守区域包含一些转录因子(如 GATA)的结合位点,说明不同生物的 *nr5a2* 基因可能受到同一类转录因子的调控。另一方面,不同生物 *nr5a2* 基因的启动子区域也具有一定的差异。例如,细胞转染实验证明小鼠 *nr5a2* 基因的启动子能够被 HNF4 和 NR5A2 自身激活,但不能被 HNF1 α 激活^[12]。相反,人 *nr5a2(hb1f)* 基因的启动子不能被 HNF4 和 NR5A2 自身激活,却可以被 HNF1 α 激活^[14]。这些结果说明,各种生物中来源于同一祖先的 *nr5a2* 基因在进化中不仅保持了一定的保守性,同时随着生物进化的进程,又发展出各自特有的调控机制,以适应不同生物生长发育的要求。

目前已经发现,果蝇的 *nr5a* 基因、斑马鱼的 *nr5a2* 基因和小鼠的 *nr5a1* 基因都具有 2 个启动子,并且分别在发育的不同阶段和/或不同的组织中指导基因的转录^[16,17]。对人 *nr5a2(hb1f)* 基因的碱基组成和 CpG 二核苷酸分布的计算,发现包含第 1 和第 2 内含子的一段 16 kb 的序列是典型的 GC 富含区(图版 I)。根据文献报道,真核基因组序列中的 GC 富含区很可能包含基因的启动子^[19]。利用真核基因启动子预测软件 PromoterInspector^[20]对人 *nr5a2(hb1f)* 基因序列进行分析,除了预测出我们已经鉴定的一个启动子以外,还提示这段 GC 富含区可能包含一个潜在的启动子。人 *nr5a2(hb1f)* 基因是否具有 2 个启动子,这 2 个启动子在生长发育中具有什么样的作用,都值得进行深入研究。

致谢 本工作作为国家自然科学基金(批准号:39893320-4)、国家人类基因组南方中心(批准号:CNCS-98-M-06)和国家重点基础研究发展规划(批准号:G1999054105)资助项目。

参 考 文 献

1 Beato M, Herrlich P, Schutz G. Steroid hormone receptors: many actors in search of a plot. *Cell*, 1995, 83: 851~857

2 Mangelsdorf D J, Evans R M. The RXR heterodimers and orphan receptors. *Cell*, 1995, 83: 841~850

3 Mangelsdorf D J, Thummel C, Beato M, et al. The nuclear receptor superfamily: the second decade. *Cell*, 1995, 83: 835~839

4 Li M, Xie Y H, Kong Y Y, et al. Cloning and characterization of a novel human hepatocyte transcription factor, hB1F, which binds and activates enhancer II of hepatitis B virus. *J Biol Chem*, 1998, 273: 29022~29031

5 Galarneau L, Drouin R, Belanger L. Assignment of the fetoprotein transcription factor gene (FTF) to human chromosome band 1q32.11 by in situ hybridization. *Cytogenet Cell Genet*, 1998, 82:

269~270

6 Nitta M, Ku S, Brown C, et al. CPF: an orphan nuclear receptor that regulates liver-specific expression of the human cholesterol 7 α -hydroxylase gene. *Proc Natl Acad Sci USA*, 1999, 96: 6660~6665

7 Nuclear receptors nomenclature committee. A unified nomenclature system for the nuclear receptor superfamily. *Cell*, 1999, 97: 161~163

8 Lu T T, Makishima M, Repa J J, et al. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Mol Cell*, 2000, 6: 507~515

9 Goodwin B, Jones S A, Price R R, et al. A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Mol Cell*, 2000, 6: 517~526

10 Castillo-Olivares A, Gil G. Role of FXR and FTF in bile acid-mediated suppression of cholesterol 7 α -hydroxylase transcription. *Nucleic Acids Res*, 2000, 28: 3587~3593

11 Rausa F M, Galarneau L, Belanger L, et al. The nuclear receptor fetoprotein transcription factor is coexpressed with its target gene HNF-3 β in the developing murine liver, intestine and pancreas. *Mech Dev*, 1999, 89: 185~188

12 Pare J F, Roy S, Galarneau L, et al. The mouse fetoprotein transcription factor (ftf) gene promoter is regulated by three gata elements with tandem e box and nkx motifs, and ftf in turn activates the hnf3 β , hnf4 α , and hnf1 α gene promoters. *J Biol Chem*, 2001, 276: 13136~13144

13 蔡彦宁, 李岷, 周青, 等. hB1F 与 HNF1 协同调控 HBV 增强子 II 的功能. *科学通报*, 2000, 45(15): 1635~1639

14 Zhang C K, Lin W, Cai Y N, et al. Characterization of genomic structure and tissue-specific promoter of the human nuclear receptor NR5A2 (hB1F) gene. *Gene*, 2001, 273: 239~249

15 Parker K L, Schimmer B P. Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocr Rev*, 1997, 18: 361~377

16 Lin W, Wang H W, Sum C, et al. Zebrafish ftz-f1 gene has two promoters, is alternatively spliced, and is expressed in digestive organs. *Biochem J*, 2000, 348: 439~446

17 Wong M, Ramayya M S, Chrousos G P, et al. Cloning and sequence analysis of the human gene encoding steroidogenic factor 1. *J Mol Endocrinol*, 1996, 17: 139~147

18 White K P, Hurban P, Watanabe T, et al. Coordination of drosophila metamorphosis by two ecdysone-induced nuclear receptors. *Science*, 1997, 276: 114~117

19 Cross S H, Bird A P. CpG islands and genes. *Curr Opin Genet Dev*, 1995, 5: 309~314

20 Scherf M, Klingenhoff A, Werner T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J Mol Biol*, 2000, 297: 599~606

人 *nr5a2(hb1f)* 基因组序列的生物信息学分析。从上至下分别为：人 *nr5a2(hb1f)* 基因的结构（长箭头表示人 *nr5a2(hb1f)* 基因，红、蓝、黑 3 种颜色的折线分别代表人 *nr5a2(hb1f)* 基因的 3 种不同剪接形式）；基因序列的碱基组成和 CpG 二核苷酸分布；基因序列中重复序列的分布情况；基因序列中分布的人 EST 和 UniGene；MZEF、FGENESH 和 GENSCAN 预测的可能编码信息

人 *nr5a2(hb1f)* 基因 人 *nr5a2(hb1f)* 基因 *nr5a2(hb1f)* *nr5a2(hb1f)*

人 *nr5a2(hb1f)*
基因组序列

人 *nr5a2(hb1f)*
基因组序列