

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 577.21

Ускоренная публикация

НОВЫЙ ДВУНАПРАВЛЕННЫЙ ПРОМОТОР ИЗ ГЕНОМА ЧЕЛОВЕКА

© 2011 г. А. С. Орехова, П. С. Свердлова, П. В. Спиринов, О. Г. Леонова,  
В. И. Попенко, В. С. Прасолов, П. М. Рубцов\*

Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991

Поступила в редакцию 01.12.2010 г.

Принята к печати 21.12.2010 г.

В геномах человека и других млекопитающих выявлено большое число близкорасположенных пар генов, которые транскрибируются в противоположных направлениях. По данным биоинформатического анализа в геноме человека до 10% генов имеют подобную организацию. В настоящей работе клонирован фрагмент генома человека, разделяющий гены гипотетических белков с неизвестной функцией – CCDC142 (Coiled-Coil Domain Containing) и TTC31 (TetraTricopeptide repeat Containing), локализованные в области 2p13.1 и ориентированные “голова к голове”. Межгенная область CCDC142-TTC31 перекрывается с CpG-островком и содержит множество потенциальных сайтов связывания факторов транскрипции. Показано, что в системе временной экспрессии репортерного гена люциферазы в клетках эмбриональной почки человека (HEK293) фрагмент ДНК, включающий эту область, действует как двунаправленный промотор. Сконструированы векторы, содержащие противоположно ориентированные гены двух флуоресцентных белков – зеленого (EGFP) и красного (DsRed2), и встроенный между ними фрагмент межгенной области CCDC142-TTC31. При трансфекции клеток HEK293 этим вектором наблюдается одновременный синтез двух флуоресцентных белков. Получены укороченные версии межгенной области и определена их промоторная активность. Минимальный промоторный фрагмент содержит ряд элементов – Inr, BRE, DPE, характерных для лишенных TATA-боксов промоторов. Таким образом, нами клонирован и охарактеризован новый двунаправленный промотор, который может использоваться для одновременной конститутивной экспрессии двух разных генов в клетках человека.

**Ключевые слова:** двунаправленные промоторы, геном человека, ген CCDC142 (Coiled Coil Domain Containing), ген TTC31 (TetraTricopeptide repeat Containing), межгенная область, векторы экспрессии.

NOVEL BIDIRECTIONAL PROMOTER FROM HUMAN GENOME, by A. S. Orekhova, P. S. Syerdlova, P. V. Spirin, O. G. Leonova, V. I. Popenko, V. S. Prassolov, P. M. Rubtsov\* (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia; \*e-mail: rubtsov@eimb.ru). In human and other mammalian genomes a number of closely linked gene pairs transcribed in opposite directions are found. According to bioinformatic analysis up to 10% of human genes are arranged in this way. In present work the fragment of human genome was cloned that separates genes localized at 2p13.1 and oriented “head-to-head”, coding for hypothetical proteins with unknown functions – CCDC (Coiled Coil Domain Containing) 142 and TTC (TetraTricopeptide repeat Containing) 31. Intergenic CCDC142-TTC31 region overlaps with CpG-island and contains a number of potential binding sites for transcription factors. This fragment functions as bidirectional promoter in the system of luciferase reporter gene expression upon transfection of human embryonic kidney (HEK293) cells. The vectors containing genes of two fluorescent proteins – green (EGFP) and red (DsRed2) in opposite orientations separated by the fragment of CCDC142-TTC31 intergenic region were constructed. In HEK293 cells transfected with these vectors simultaneous expression of two fluorescent proteins is observed. Truncated versions of intergenic region were obtained and their promoter activity measured. Minimal promoter fragment contains elements Inr, BRE, DPE characteristic for TATA-less promoters. Thus, from the human genome the novel bidirectional promoter was cloned that can be used for simultaneous constitutive expression of two genes in human cells.

**Keywords:** bidirectional promoters, human genome, gene CCDC142 (Coiled Coil Domain Containing), gene TTC31 (TetraTricopeptide repeat Containing), intergenic region, expression vectors.

Принятые сокращения: Inr – инициаторный элемент; BRE – элемент, узнаваемый фактором транскрипции TFIIIB (TFIIIB Recognition Element); DPE (Downstream Promoter Element) – нижележащий промоторный элемент; BGH – гормон роста крупного рогатого скота.

\* Эл. почта: rubtsov@eimb.ru

В геномах человека и других млекопитающих выявлено большое число сложных локусов, в которых расположены гены, транскрибируемые в противоположных направлениях с общих регуляторных областей, так называемых двунаправленных промоторов. Проведен полногеномный анализ таких генов у мыши и человека [1]. Компьютерный анализ 36606 транскрипционных единиц генома мыши и 42887 транскрипционных единиц генома человека обнаружил 1638 и 2113 двунаправленных промоторов соответственно. Проверка особенностей экспрессии генов мыши, транскрибируемых с двунаправленных промоторов, с помощью гибридизации на микрочипах показала, что очень часто гены, транскрибируемые в противоположных направлениях с одного двунаправленного промотора, регулируются координировано [1]. По оценкам Collins et al. [2] в геноме человека примерно 10% генов располагаются на расстоянии менее 1 т.п.н. друг от друга и ориентированы “голова к голове”, т.е. контролируются двунаправленными промоторами. Показано также [1], что двунаправленные промоторы обогащены сайтами связывания определенных факторов транскрипции. Наряду с сайтами связывания GA-связывающего белка (GABP), входящего в семейство факторов транскрипции ETS, это сайты связывания факторов Nrf1, CCAAT, YY1 и SP1. Установлено [3], что двунаправленные промоторы составляют более 25% всех промоторов, расположенных в обогащенных CpG-динуклеотидами областях генома (CpG-островках) и гиперметилированных в опухолевых клетках.

В ходе работы, направленной на поиск потенциальных белковых партнеров ядерного рецептора DAX1 с использованием дрожжевой двухгибридной системы, нами был клонирован фрагмент кДНК, кодирующей гипотетический белок CCDC142 с неизвестной функцией. Ген, кодирующий этот белок, локализован в области 2p13.1. Согласно предсказаниям, основанным на сравнении геномной последовательности и EST, он состоит из 9 экзонов. Общая длина гена – 8895 п.н. Со стороны теломеры ген *CCDC142* соседствует с геном *MRPL53*, кодирующим митохондриальный рибосомный белок, который находится на расстоянии 1306 п.н. и транскрибируется в том же направлении. Со стороны центромеры рядом с геном *CCDC142* располагается ген *TTC31* в ориентации “голова к голове”. Межгенная область имеет размер всего 93 п.н. (расстояние между предполагаемыми точками инициации транскрипции, соответствующими 5'-концам наиболее длинных EST). Анализ межгенной области с помощью программы MatInspector выявил сайты связывания факторов транскрипции, характерные для двунаправленных промоторов [2]. Целью настоящей работы было клонирование и характеристика промоторной области гена *CCDC142*.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В работе использовали штамм *Escherichia coli* JM109, рестрикционные эндонуклеазы BamHI, EcoRI, NcoI, SalGI, XhoI фирмы “Promega”, BglII, KpnI, NheI, NotI, SmaI, Eco32I (EcoRV), ДНК-лигазу фага T4, щелочную фосфатазу из кишечника телят (CIAP) и Taq-ДНК-полимеразу фирмы “MBI Fermentas” (Литва), олигонуклеотиды (табл. 1) фирмы “Синтол” (Россия). Все генноинженерные процедуры проводили согласно стандартным протоколам [3]. Плазмиды выделяли из клеток *E. coli* с помощью наборов Plasmid Mini Kit фирмы “Qiagen”.

**Аmplификация и клонирование фрагментов геномной ДНК человека.** Суммарную ДНК выделяли из периферической крови с помощью набора Wizard<sup>R</sup> Genomic DNA Purification Kit (“Promega”). Фрагменты геномной ДНК, содержащие межгенную область генов *CCDC142* и *TTC31*, амплифицировали, используя ПЦР со следующими парами праймеров (табл. 1): 1) ccd-ttc PR-Kp и ccd-ttc PR-Bg (продукт размером 646 п.н.); 2) ccd-ttc PR-Bg-n и ccd-ttc PR-Kp-n (646 п.н.); 3) Nh-ccd\_Tis-L и ccd-ttc PR-Bg (423 п.н.); 4) Nh-ccd\_Tis-L и ccd-ttc PR-Kp-n (423 п.н.); 5) Nh-ccd\_Tis-R и ccd-ttc PR-Bg (254 п.н.); 6) Nh-ccd\_Tis-R и ccd-ttc PR-Kp (254 п.н.); 7) Nh-ttc\_Tis-L и ccd-ttc PR-Bg (283 п.н.); 8) Nh-ttc\_Tis-L и ccd-ttc PR-Kp-n (283 п.н.); 9) Nh-ttc\_Tis-R и ccd-ttc PR-Kp (390 п.н.); 10) Nh-ttc\_Tis-R и ccd-ttc PR-Bg-n (390 п.н.). Продукты ПЦР разделяли электрофоретически в 1%-ных агарозных гелях и выделяли из геля, используя компоненты набора QIAquick<sup>R</sup> Gel Extraction Kit (“Qiagen”). Выделенные фрагменты клонировали в векторе pGEM<sup>R</sup>-T Easy (“Promega”). Рекомбинантные клоны отбирали на чашках с агаризованной средой, содержащей ампициллин, изопропилтио-β-D-галактозид (ИПТГ) и (X-Gal). Присутствие вставок подтверждали с помощью ПЦР на колониях с праймерами M13-F и M13-R (табл. 1). Плазмиды из положительных клонов выделяли и секвенировали с теми же праймерами.

**Выделение РНК и проведение обратной транскрипции–ПЦР.** Суммарную РНК из тканей человека выделяли гуанидинтиоцианатным методом [5], из культивируемых клеток – с использованием реагента TRIzol<sup>R</sup> (“Invitrogen<sup>TM</sup>”). Целостность РНК проверяли с помощью электрофореза в 1%-ном агарозном геле, концентрацию РНК определяли на спектрофотометре NanoDrop 1000 (“Termo Fisher Scientific Inc.”). Обратную транскрипцию проводили с реагентами ImProm-II<sup>TM</sup> Reverse Transcription (“Promega”). Синтез кДНК с 1 мкг суммарной РНК праймировали олиго(dT)<sub>15</sub>. Реакцию останавливали прогреванием реакционной смеси при 70°C в течение 15 мин. Для последующей ПЦР брали 1 мкл реакционной смеси.

**Таблица 1.** Синтетические олигонуклеотиды, использованные в работе

Олигонуклеотид	Нуклеотидная последовательность (5' → 3')*	Число звеньев
ccd-ttc PR-Kp	<u>GGTACC</u> CAGTGAACCTCCCAGCGAAG	25
ccd-ttc PR-Bg	<u>AGATCT</u> GGGCCGTAGAGAGCAGTCTAC	27
ccd-ttc PR-Bg-n	<u>AGATCT</u> AGTGAACCTCCCAGCGAAG	25
ccd-ttc PR-Kp-n	<u>GGTACC</u> GGGCCGTAGAGAGCAGTCTAC	27
Nh-ccd Tis-L	<u>GCTAGC</u> GCCGTTTTCTCCAGTCCGGGA	27
Nh-ttc Tis-L	<u>GCTAGC</u> CTTTTCATTTCCGGGTCAGTGT	27
Nh-ccd Tis-R	<u>GCTAGC</u> CCGGACTGGAGAAAACGGC	25
Nh-ttc Tis-R	<u>GCTAGC</u> CACAGTGACCCGGAAATGAAAG	25
ccd142-F1	GCCCTCCGT <u>GAATTC</u> CAGGCGTCTCGCTCAGGTA	34
ccd142-R1	CAGCCCTCC <u>GTCGAC</u> AGAGGAGGTAAGTCCTTTTG	35
ttc31_F1	<u>AGATCT</u> CATGGCGCCGATTCCAAAGACTG	29
ttc31_R2	TTGAGCAAAGCTGGTACCCAAC	22
M13-F	GCCAGGGTTTTCCCAGTCACGA	22
M13-R	GAGCGGATAACAATTTACACAGG	24
RV-Pr3	CTAGCAAAATAGGCTGTCCC	20
GL-Pr2	CTTTATGTTTTTGGCGTCTTCCA	22
BGH	AACTAGAAGGCACAGTCG	18
EGFP-N	CGTCGCCGTCCAGCTCGACCAG	22
Red-N	GTACTGGAAGTGGGGGGACAG	21

\*Подчеркнуты нуклеотидные последовательности участков узнавания рестриктаз BglII (AGATCT), EcoRI (GAATTC), KpnI (GGTACC), NheI (GCTAGC) и SalGI (GTCGAC).

**Конструирование плазмид с репортерным геном люциферазы для определения промоторной активности фрагментов ДНК.** Фрагменты межгенной области генов *CCDC142* и *TTC31*, клонированные в векторе pGEM<sup>R</sup>-T Easy, вырезали рестриктазами KpnI и BglII, KpnI и NheI или NheI и BglII, сайты узнавания которых были введены в праймеры для амплификации (табл. 1). Полученные фрагменты встраивали в вектор pGL3-Basic<sup>R</sup> (“Promega”), расщепленный теми же рестриктазами. Присутствие вставок подтверждали с помощью ПЦР на колониях с праймерами RV-Pr3 и GL-Pr2 (табл. 1).

**Конструирование векторов для одновременной экспрессии генов двух флуоресцентных белков в клет-**

**ках млекопитающих.** Для конструирования векторов экспрессии, содержащих межгенную область *CCDC142-TTC31*, предварительно получили плазмиду с двумя сигналами полиаденилирования, фланкируемыми полилинкер. С этой целью из вектора pcDNA3.1(+) (“Invitrogen”) удаляли BglII–BamHI-фрагмент, содержащий промотор и энхансер цитомегаловируса. Вместо него встраивали BamHI-фрагмент размером 262 п.н., содержащий сигнал полиаденилирования вируса SV40. Полученный вектор pcDNA-Stop содержит полилинкер, с одной стороны которого помещен сигнал полиаденилирования SV40, а с другой – сигнал полиаденилирования гена гормона роста крупного рогатого скота (BGH).

На следующем этапе в этот вектор встраивали межгенную область *CCDC142-TTC31*. Вектор *pcDNA-Stop* расщепляли рестриктазой *Eco321*, обрабатывали ДНК-полимеразой *Taq* в присутствии *dTTP* в течение 1 ч при 70°C. Таким образом получили линейную форму вектора с 3'-выступающими остатками *dT*. Эту форму смешивали с продуктом ПЦР-амплификации межгенной области *CCDC142-TTC31* с праймерами *ccd-ttc PR-Bg-n* и *ccd-ttc PR-Kp-n* (фрагмент размером 646 п.н.) и обрабатывали ДНК-лигазой *фага T4*. После трансформации лигазной смесью клеток штамма *JM109 E. coli* клоны со вставками отбирали с помощью ПЦР на колониях, используя две комбинации пар праймеров: *ccd-ttc PR-Bg-n/BGH* и *ccd-ttc PR-Kp-n/BGH*. В результате получили два варианта плазмид, отличающихся ориентацией фрагмента межгенной области *CCDC142-TTC31* — *pcDNA-ccdc-ttc-Pr-6* и *pcDNA-ccdc-ttc-Pr-18*.

На следующей стадии конструирования в полученные векторы встроили ген зеленого флуоресцентного белка *EGFP* из плазмиды *rEGFP-N1* ("Clontech"). Для этого *rEGFP-N1* сначала расщепили рестриктазами *BamHI* и *NotI*. Продукты расщепления обрабатывали ДНК-полимеразой *Taq* в присутствии четырех *dNTP* в течение 1 ч при 70°C. Фрагмент, содержащий ген *EGFP*, с 3'-выступающими остатками *dA* клонировали в векторе *rGEM-Teasy* ("Promega"). Рекомбинантные клоны отбирали на среде с *X-Gal/ИПТГ*, наличие вставок подтверждали с помощью ПЦР на колониях, используя праймеры *M13F* и *M13R*. В полученных клонах ген *EGFP* фланкирован участками узнавания рестриктазы *EcoRI*, что позволило вырезать его этой рестриктазой и встроить в векторы *pcDNA-ccdc-ttc-Pr-6* и *pcDNA-ccdc-ttc-Pr-18*, расщепленные *EcoRI* и обработанные *CIAP*. Клоны с правильной ориентацией гена *EGFP* по отношению к промоторной области отбирали с помощью ПЦР на колониях с праймерами *GFP-N* и *BGH* (табл. 1). В результате получили векторы — *pcDNA-ccdc-ttc-Pr-6-EGFP* и *pcDNA-ccdc-ttc-Pr-18-EGFP*.

На заключительном этапе в векторы встраивали фрагменты с геном красного флуоресцентного белка *DsRed2* из вектора *pDsRed2-1* ("Clontech"). Плазмиду *pcDNA-ccdc-ttc-Pr-6-EGFP* гидролизывали рестриктазами *BglII* и *NotI* и лигировали с *BglII-NotI*-фрагментом, вырезанным из *pDsRed2-1*. Плазмиду *pcDNA-ccdc-ttc-Pr-18-EGFP* расщепляли рестриктазами *KpnI* и *NotI* и лигировали с *KpnI-NotI*-фрагментом из *pDsRed2-1*. В обоих случаях рекомбинантные клоны отбирали с помощью ПЦР на колониях с праймерами *GFP-N* и *DsRed-N*. В результате получили два вектора, содержащих гены флуоресцентных белков *EGFP* и *DsRed2*, с расположенным между ними фрагментом межгенной области *CCDC142-TTC31* — *pcDNA-*

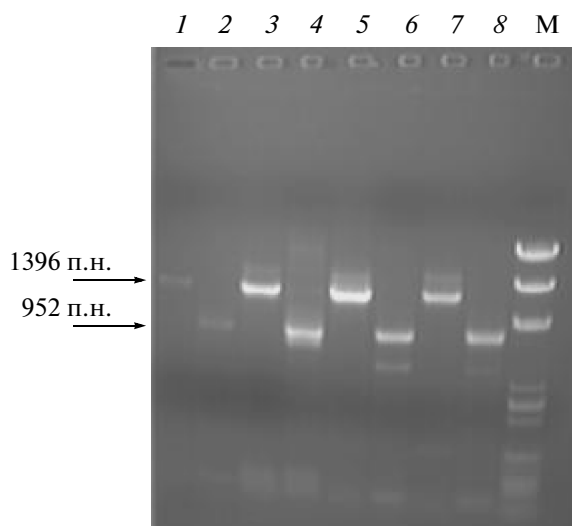
*ccdc-ttc-Pr-6-EGFP-DsRed2* и *pcDNA-ccdc-ttc-Pr-18-EGFP-DsRed2*.

**Трансфекцию клеток НЕК293** (линия эпителио-подобных клеток почки эмбриона человека) проводили с использованием липофектамина ("Invitrogen™"). К образцам ДНК (0.5 мкг/мкл в *TE*-буфере) добавляли 120 мкл среды *DMEM* и 150 мкл липофектамина. Смесь перемешивали на вортексе и инкубировали в течение 20 мин при 20°C, добавляли 1.2 мл среды *DMEM*, снова перемешивали на вортексе, центрифугировали в течение 1 мин при 6000 об/мин и инкубировали в течение 4 ч. Затем среду *DMEM* заменяли обогащенной факторами роста средой *DMEM*. Клетки высевали в 24-луночные планшеты и выращивали в *CO<sub>2</sub>*-инкубаторе при 37°C в атмосфере 5% *CO<sub>2</sub>* в течение 24–48 ч.

**Определение активности люциферазы.** Плазмиды с репортерным геном люциферазы светлячка вводили в клетки НЕК293, как описано выше. Каждый образец, кроме тестируемой репортерной конструкции, содержал также плазмиду *pRL-tk* ("Promega") с геном люциферазы *Renilla* под контролем промотора гена тимидинкиназы вируса простого герпеса. Результаты определения активности люциферазы светлячка нормировали по активности люциферазы *Renilla*, что позволяло учесть различия в эффективности трансфекции клеток в разных образцах. В качестве контроля использовали плазмиду *pGL3-Basic* (не содержит промотора и служит отрицательным контролем) и *pGL3-Control* (содержит промотор и энхансер вируса *SV40* и служит положительным контролем).

Из лунок планшетов с трансфицированными клетками удаляли культуральную среду, отмывали лунки 1 мл однократного фосфатно-солевого буфера (*PBS*), клетки лизировали, добавляя 100 мкл *Passive Lysis Buffer* из набора *Dual-Glo™ Luciferase Assay System* ("Promega") и выдерживая в течение 10 мин при комнатной температуре. Лизаты переносили в пробирки на 1.5 мл и центрифугировали при 13200 *g* в течение 10 мин. По 20 мкл лизатов вносили в лунки 96-луночного планшета и определяли последовательно активность люциферазы светлячка и люциферазы *Renilla*, используя реагенты набора *Dual-Glo™ Luciferase Assay System*. Люминесценцию измеряли на мультиридере *Plate SHAMELEON™ V* ("Hydrex Oy", Финляндия). Статистическую обработку результатов проводили с помощью программы *Microsoft Excel*.

**Секвенирование ДНК.** Нуклеотидные последовательности определяли в Межинститутском Центре коллективного пользования "ГЕНОМ", ИМБ РАН (<http://www.genome-centre.narod.ru/>), организованном при поддержке РФФИ (грант № 00-04-55000). Образцы секвенировали на автоматическом секваторе *ABI PRISM Model 3100* ("Applied Biosys-



**Рис. 1.** Анализ экспрессии генов *CCDC142* (1, 3, 5, 7) и *TTC31* (2, 4, 6, 8) в клетках НЕК293 (1, 2), гипофизе (3, 4), взрослых (5, 6) и фетальных (7, 8) надпочечниках человека методом обратной транскрипции-ПЦР. М – маркеры длины (2415, 1651, 1151+1101, 580, 564, 514, 487, 420, 396, 255 п.н.). Стрелками показано положение продуктов ПЦР.

tems”), результаты секвенирования анализировали с использованием программы Chromas (<http://www.techneylsium.com.au/chromas.htm>).

**Флуоресцентная микроскопия.** Флуоресценцию белков EGFP и DsRed2 в трансфицированных клетках выявляли на микроскопе Leica DMI 4000 (США).

**Базы данных и компьютерные программы.** Гомологии вставок клонов и участков генома человека (база данных GenBank, NCBI, США) анализировали с использованием следующих программ:

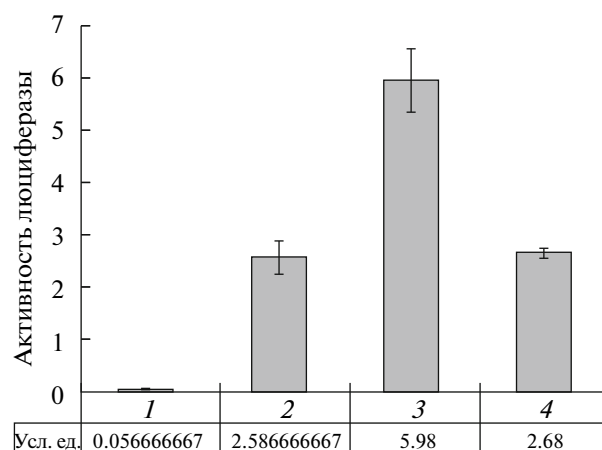
BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) и BLAT (<http://www.genome.ucsc.edu/cgi-bin/hgBlat?command=start>).

Поиск проводили с помощью ресурсов отделения NR (non redundant) базы данных GenBank NCBI и htgs – раздел базы данных, где помещены незавершенные последовательности генома человека (<http://www.ncbi.nlm.nih.gov/>) и сервера UCSC Human Genome Browser (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>). Поиск потенциальных сайтов связывания факторов транскрипции проводили, используя программу MatInspector (<http://www.genomatix.de/cgi-bin/./eldorado/main.pl>).

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### Анализ экспрессии генов *CCDC142* и *TTC31*

Ген гипотетического белка *CCDC142* – потенциального партнера ядерного рецептора *DAX1* – локализован в области 2p13.1 рядом с геном другого ги-

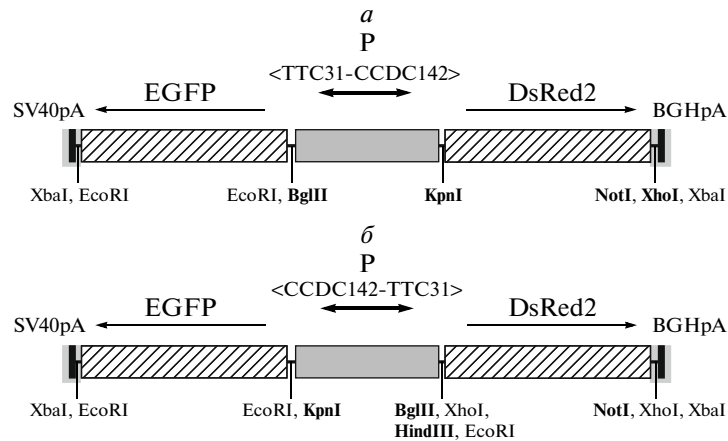


**Рис. 2.** Анализ промоторной активности межгенной области *CCDC142-TTC31* в системе временной экспрессии репортерного гена люциферазы в клетках НЕК293. Представлены результаты трех независимых опытов. Приведены средние значения и стандартные отклонения. 1 – pGL3-Basic; 2 – pGL3-Control; 3 – pGL3-Basic-TTC31-CCDC142-Pr-full; 4 – pGL3-Basic-CCDC142-TTC31-Pr-full.

потетического белка *TTC31*. Гены ориентированы “голова к голове” и транскрибируются в противоположных направлениях. Прежде чем клонировать межгенную область, мы проверили, происходит ли координированная транскрипция генов *CCDC142* и *TTC31* в тканях, экспрессирующих *DAX1* – надпочечниках и гипофизе, а также в клетках НЕК293, которые в дальнейшем планировали использовать для функционального анализа промоторной активности. С этой целью применили метод обратной транскрипции-ПЦР с праймерами, специфичными к экзонам этих генов – экзонам 1 и 4 гена *CCDC142* (ccd142-F1/ccd142-R1) и экзонам 1 и 10 гена *TTC31* (ttc31-F1/ttc31-R2) соответственно (табл. 1). Во всех случаях выявлены продукты ПЦР ожидаемой длины – 1396 п.н. для гена *CCDC142* и 952 п.н. для гена *TTC31* (рис. 1), что подтвердило одновременную экспрессию двух генов в изученных образцах.

### Клонирование и анализ промоторной активности межгенной области *CCDC142-TTC31* человека

Для более детальной характеристики промоторов фрагмент межгенной области *CCDC142-TTC31* длиной 637 п.н. амплифицировали и встроили в вектор pGL3-Basic (“Promega”) перед репортерным геном люциферазы светлячка в двух противоположных ориентациях. Этот фрагмент включает 293 п.н. начала транскрибируемой области гена *CCDC142*, 93 п.н. межгенной области и 252 п.н. начала транскрибируемой области гена *TTC31*. На



**Рис. 3.** Схемы кассет экспрессии в векторах pcDNA-ccdc-ttc-Pr-6-EGFP-DsRed2 (а) и pcDNA-ccdc-ttc-Pr-18-EGFP-DsRed2 (б). P – двунаправленный промотор межгенной области *CCDC142-TTC31*; *EGFP* и *DsRed2* – гены зеленого и красного флуоресцентных белков; SV40pA и BGHpA – сайты полиаденилирования вируса SV40 и гена гормона роста крупного рогатого скота. Стрелками показаны направления транскрипции. Внизу отмечены положения участков узнавания рестриктаз, уникальные сайты выделены полужирным.

следующем этапе сконструированными векторами трансфицировали клетки HEK293 и измерили активность люциферазы. Каждый образец, кроме тестируемой репортерной конструкции, содержал также плазмиду pRL-tk с геном люциферазы Renilla под контролем промотора гена тимидинкиназы вируса простого герпеса. По активности люциферазы Renilla нормировали результаты определения активности люциферазы светлячка, что позволяло учесть различия в эффективности трансфекции клеток в разных образцах. В качестве контроля использовали также плазмиду pGL3-Basic (не содержит промотора и служат отрицательным контролем) и плазмиду pGL3-Control (содержит промотор и энхансер вируса SV40 и служит положительным контролем). Установлено, что активность промотора гена *CCDC142* (конструкция pGL3-Basic-*CCDC142-TTC31-Pr-full*) примерно равна, а активность промотора гена *TTC31* (конструкция pGL3-Basic-*TTC31-CCDC142-Pr-full*) примерно вдвое выше активности промотора SV40 в плазмиде pGL3-Control (рис. 2). Таким образом, межгенная область *CCDC142-TTC31* действует в клетках HEK293 как относительно сильный двунаправленный промотор, что позволяет использовать ее для создания векторов для одновременной конститутивной экспрессии двух разных генов в клетках млекопитающих.

**Использование межгенной области *CCDC142-TTC31* для одновременной экспрессии двух генов в клетках млекопитающих на примере генов флуоресцентных белков**

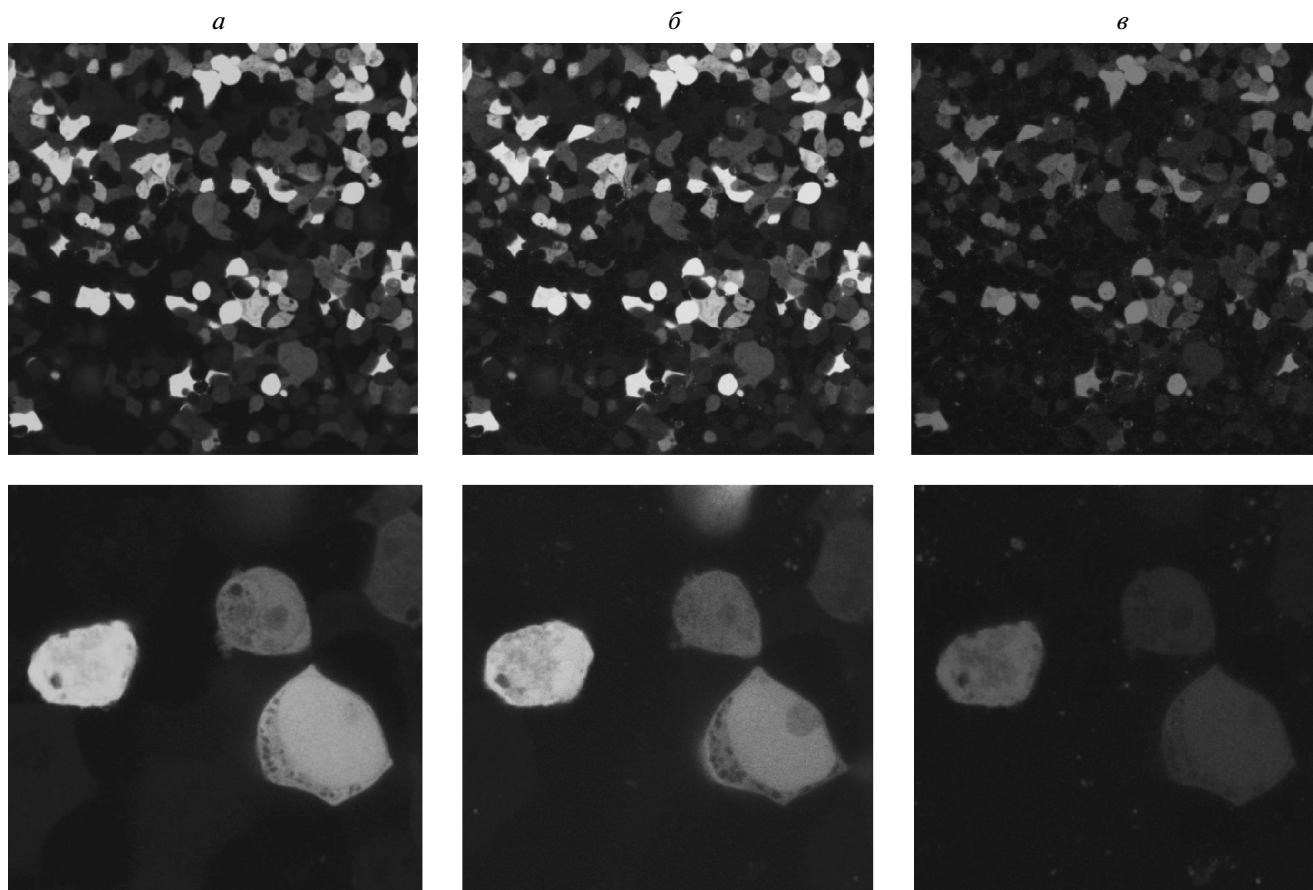
Возможность одновременной экспрессии двух генов мы проверяли с использованием векторов,

содержащих гены двух флуоресцентных белков – зеленого (*EGFP*) и красного (*DsRed2*), и встроенную между ними межгенную область *CCDC142-TTC31*. Схемы кассет экспрессии векторов представлены на рис. 3. Межгенная область должна направлять транскрипцию гена *EGFP* в одном направлении и гена *DsRed2* – в противоположном. Кассеты экспрессии содержат также сигналы полиаденилирования SV40 и гена *BGH* после генов *EGFP* и *DsRed2* соответственно.

После трансфекции клеток HEK293 синтезируемые белки *EGFP* и *DsRed2* визуализировали с помощью флуоресцентной микроскопии. На рис. 4 показаны результаты, полученные при трансфекции клеток вектором pcDNA-ccdc-ttc-Pr-18-EGFP-DsRed2. Видно, что в трансфицированных клетках одновременно экспрессируются оба гена. При наложении изображений зеленой (рис. 4а) и красной (рис. 4б) флуоресценции они полностью совпадают (рис. 4в). Сходные результаты получены с использованием вектора pcDNA-ccdc-ttc-Pr-6-EGFP-DsRed2 с противоположной по отношению к генам *EGFP* и *DsRed2* ориентацией промоторного фрагмента. Таким образом, двунаправленный промотор межгенной области *CCDC142-TTC31* может быть использован для одновременной конститутивной экспрессии двух генов в клетках человека.

**Делеционное картирование минимального двунаправленного промотора межгенной области *CCDC142-TTC31***

Следующей задачей было картирование границ двунаправленного промотора. С этой целью получены укороченные фрагменты межгенной области *CCDC142-TTC31*, которые встраивали в вектор



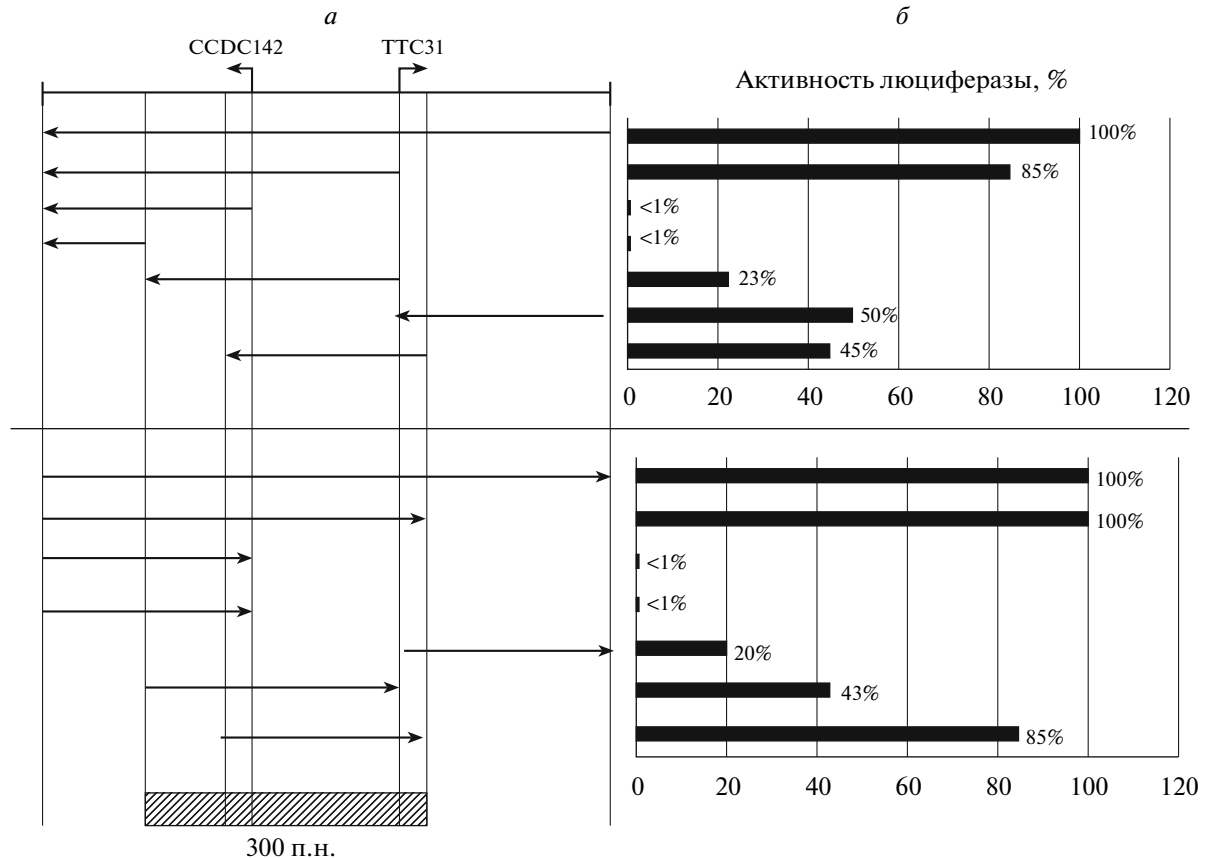
**Рис. 4.** Анализ одновременной экспрессии генов флуоресцентных белков EGFP и DsRed2 под контролем двунаправленного промотора межгенной области *CCDC142-TTC31* в клетках HEK293. *a* – Флуоресценция EGFP; *б* – флуоресценция DsRed2; *б* – совмещение изображений *a* и *в*.

pGL3-Basic. Полученными конструкциями трансфицировали клетки HEK293 и определяли в них люциферазную активность. Схематическое изображение делеционных вариантов межгенной области *CCDC142-TTC31* и результаты определения люциферазной активности после трансфекции клеток HEK293 приведены на рис. 5. Укорачивание исходного фрагмента межгенной области с обеих сторон и приводит к снижению промоторной активности, однако центральная часть фрагмента 300 п.н. сохраняет относительно высокую активность (примерно 45–85% активности исходного фрагмента). Нуклеотидные последовательности обеих цепей минимального промоторного фрагмента представлены на рис. 6. В этих последовательностях мы провели поиск регуляторных элементов кор-промоторов – инициаторного элемента (Inr) с консенсусом YYANWYY [6], элемента, узнаваемого фактором транскрипции ПВ (BRE), с консенсусом SSRCGCC [7] и нижележащего элемента (DPE) с консенсусом DSWYVY [8], где R – G/C, Y – T/C, W – A/T, S –

G/C, D – A/G/T, V – A/C/G. Анализ нуклеотидной последовательности минимального фрагмента выявил такие элементы (рис. 6), причем их положение относительно условных точек инициации транскрипции генов *CCDC142* и *TTC31* согласуется с типичной архитектурой кор-промоторов [9–11].

#### Ортологи генов *CCDC142* и *TTC31* человека

Ортологи гена *CCDC142* обнаружены у других видов млекопитающих. Результаты выравнивания нуклеотидных последовательностей кДНК этих генов и выведенных аминокислотных последовательностей их белковых продуктов свидетельствует об их консервативности (табл. 2). Уровень идентичности аминокислотных последовательностей *CCDC142* человека и шимпанзе составляет более 99%, человека и собаки – 85%, человека и крупного рогатого скота – 83%, человека и грызунов (мышь, крыса) – более 75%. Ортологи гена *TTC31* обнаружены у приматов, собаки, полосатого данио (табл. 3), но не выяв-



**Рис. 5.** Делеционное картирование двунаправленного промотора межгенной области *CCDC142-TTC31* человека. *а* – Схемы делеционных вариантов межгенной области (стрелки показывают направление транскрипции репортерного гена люциферазы, внизу выделено положение минимального промоторного фрагмента); *б* – относительная активность люциферазы. Приведены средние значения трех независимых опытов.

ляются у грызунов. В то же время организация других генов, примыкающих к гену *CCDC142*, в синтенных участках геномов человека и грызунов идентична. Таким образом, ген *TTC31* утрачен у грызунов.

Функции белковых продуктов генов *CCDC142* и *TTC31* не известны. Белок *CCDC142*, как следует из его названия, содержит структурный домен “спиральная спираль” [12], но лишен каких-либо функциональных доменов. Нами установлено, что слитый белок EGFP-*CCDC142*, синтезируемый в клетках млекопитающих, локализуется в цитоплазме (данные не представлены). В аминокислотной последовательности белка *CCDC142* человека присутствуют пять LXXLL-подобных мотивов, которые могут участвовать в белок-белковых взаимодействиях, как это показано на примере некоторых ядерных рецепторов [13, 14]. Эти мотивы возможно обуславливают взаимодействие белка *CCDC142* с ядерным рецептором *DAX1 (NR0B1)*, выявленное нами в дрожжевой двухгибридной системе.

Функции белка *TTC31*, предсказанного методами биоинформатики на основе анализа нуклеотид-

ных последовательностей геномов, также неизвестны. Этот белок содержит домен TTR (TetraTricopeptide Repeat). Белки с TTR-доменом обнаружены у разных организмов – от бактерий до человека. Они участвуют в таких разнообразных процессах, связанных с белок-белковыми взаимодействиями, как функционирование шаперонов, регуляция клеточного цикла, транскрипция, транспорт белков [15, 16].

#### **Широкое распространение двунаправленных промоторов в геномах млекопитающих**

Как уже отмечалось, биоинформатический анализ генома человека и других млекопитающих показал, что они содержат большое число двунаправленных промоторов, разделяющих близко расположенные гены, ориентированные “голова к голове” [1, 2, 17]. Доля таких генов в геноме человека, согласно [18], составляет около 10%. Лишь малая часть двунаправленных промоторов содержит ТАТА-боксы – 9% по сравнению с 29% однонаправленных промоторов [19]. В то же время, двунаправленные промоторы значительно чаще,



а

CHR2: 74, 709, 961–74, 710, 261 (–цепь) ген *CCDC142*

AGCTTGATCCGCCCCACAGTCTTTGGAATCGGCGCCATCGCATCTACAGT  
 GACCCGAAATGAAAGGTAGTCTAGTATTTGATACCAGCCAAAGGGCGGG  
 GTGGGACAGAGTGGGCGG**CCATGTT**TGTTAGGGGCAGAAGCCTC**CCATTA**  
**CGGAGCACGAGAGAGTCCATGAAGGTCCC**CGCGACTCCC**GGACTGGAGAA**  
**AACGGCTCTT**GCGATGGGGCGAAGTCCGAGCTGCGGGCGGGCGTTGGTCCG  
 TGCAGGGAAGTGGGAATCGTTAGGTTTCGTTCTGGACCCGCCGCCCATGG

б

CHR2: 74, 709, 961–74, 710, 260 (+цепь) ген *TTC31*

CCATGGGGCGGGCGGGTCCAGAACGAACCTAACGATTCCCACCTCCCTGCA  
 CGGACCAACGCCCGCCGAGCTCGGACTTCGCCCATCGCAAGAGCCGTT  
 TTCTCCAGTCCGGGAGTCCGGGGACCTTCATGGACTCTCTCGTGCTCCG  
 TAATGGGAGGCTTCTGCCCTAACAAACATGGCCGCCCACTCTGTCCAC  
**CCCGCCCTTGGCTGGTATCAAATACTAGACT**ACCTTTCATTTCCGGGTC  
 ACTGTAGATGCGATGGCGCCGATCCAAAGACTGTGGGGCGGATCAAGCT

**Рис. 6.** Нуклеотидные последовательности минимального двунаправленного промотора межгенной области *CCDC142-TTC31* человека и локализация потенциальных элементов кор-промоторов – Inr (консенсус – YYANWYY), BRE (консенсус SSRCGCC) и DPE (консенсус DSWYVY), где R – G/C, Y – T/C, W – A/T, S – G/C, D – A/G/T, V – A/C/G. Приведены (–) и (+)-цепи (а и б) хромосомы 2 человека с сервера UCSC Human Genome Browser (референсная последовательность GRCh37/hg19, февраль 2009 г.). Подчеркнуты транскрибируемые области, соответствующие наиболее протяженным в 5'-направлении EST, курсивом показаны 5'-фланкирующие последовательности. Потенциальные элементы кор-промоторов затенены, потенциальные Inr-элементы дополнительно выделены полужирным.

чем однонаправленные связаны с CpG-островками (90 против 45%) [3, 19]. Практически половина генов, разделенных двунаправленными промоторами, коэкспрессируется или регулируются координировано [18, 19]. Для двунаправленных промоторов характерно также присутствие элементов кор-промоторов (Inr, BRE, DPE) [11, 19]. Особый интерес представляет сравнительное картирование двунаправленных промоторов в геномах разных видов млекопитающих. Предпринятая попытка картирования ортологов двунаправленных промоторов

человека в геноме крупного рогатого скота [20] показала, что по меньшей мере у 70% пар генов человека, транскрибирующихся с двунаправленных промоторов, имеются ортологичные пары в геноме крупного рогатого скота.

Клонированный и охарактеризованный в настоящей работе двунаправленный промотор межгенной области *CCDC142-TTC31* из генома человека является типичным представителем этого класса промоторов. Он лишен ТАТА-бокса, тесно связан с

**Таблица 2.** Консервативность белка и гена *CCDC142*

Сравниваемые виды	Уровень идентичности, %	
	белок	ген
<i>Homo sapiens/Pan troglodytes</i>	99.2	99.3
<i>Homo sapiens/Canis lupus familiaris</i>	79.6	85.0
<i>Homo sapiens/Bos taurus</i>	77.3	83.0
<i>Homo sapiens/Mus musculus</i>	74.1	76.7
<i>Homo sapiens/Rattus norvegicus</i>	74.2	77.2

**Таблица 3.** Консервативность белка и гена *TTC31*

Сравниваемые виды	Уровень идентичности, %	
	белок	ген
<i>Homo sapiens/Pan troglodytes</i>	98.1	99.0
<i>Homo sapiens/Canis lupus familiaris</i>	83.3	86.1
<i>Homo sapiens/Danio rerio</i>	40.5	51.2

СрG-островком, содержит элементы Inr, BRE и DPE кор-промоторов.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (09-04-01497а).

### СПИСОК ЛИТЕРАТУРЫ

- Engstrom P.G., Suzuki H., Ninomiya N., Akalin A., Sessa L., Lavorgna G., Brozzi A., Luzzi L., Tan S.L., Yang L., Kunarso G., Ng E.L., Batalov S., Wählestedt C., Kai C., Kawai J., Carninci P., Hayashizaki Y., Wells C., Bajic V.B., Orlando V., Reid J.F., Lenhard B., Lipovich L. 2006. Complex loci in human and mouse genomes. *PLoS Genet.* **2(4)**, e47.
- Collins P.J., Kobayashi Y., Nguyen L., Trinklein N.D., Myers R.M. 2007. The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet.* **3(11)**, e208.
- Shu J., Jelinek J., Chang H., Shen L., Qin T., Chung W., Oki Y., Issa J.P. 2006. Silencing of bidirectional promoters by DNA methylation in tumorigenesis. *Cancer Res.* **66**, 5077–5084.
- Maniatis T., Fritsch E.F., Sambrook J. 1982. *Molecular cloning. A laboratory manual*, 2<sup>nd</sup> Ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor. Laboratory Press.
- Chomczynski P., Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.
- Smale S.T., Baltimore D. 1989. The “initiator” as a transcription control element. *Cell.* **57**, 103–113.
- Lagrange T., Kapanidis A.N., Tang H., Reinberg D., Ebright R.H. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific binding by transcription factor IIB. *Genes Dev.* **12**, 34–44.
- Kutach A.K., Kadonaga J.T. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20**, 4754–4764.
- Butler J.E.F., Kadonaga J.T. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16**, 2583–2592.
- Smale S.T. 2001. Core promoters: active contributors to combinatorial gene regulation. *Genes Dev.* **15**, 2503–2508.
- Свердлов Е.Д., Виноградова Т.В. 2010. Эволюция взглядов на молекулярные механизмы жизнедеятельности в свете полногеномной информации на примере представлений о кор-промоторах. *Молекуляр. биология.* **44**, 773–785.
- Burkhard P., Stetefeld J., Strelkov S.V. 2001. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88.
- Zhang H., Thomsen J.S., Johansson L., Gustafsson J.A., Treuter E. 2000. DAX-1 functions as an LXXLL-containing corepressor for activated estrogen receptors. *J. Biol. Chem.* **275**, 39855–39859.
- Suzuki T., Kasahara M., Yoshioka H., Morohashi K.-I., Umesono K., 2003. LXXLL-related motifs in DAX-1 have target specificity for the orphan nuclear receptors Ad4BP/SF-1 and LRH-1. *Mol. Cell. Biol.* **23**, 238–249.
- Das A.K., Cohen P.W., Barford D. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192–1199.
- D’andrea L.D., Regan L. 2003. TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655–662.
- Adachi N., Lieber M.R. 2002. Bidirectional gene organization: a common architectural feature of the human genome. *Cell.* **109**, 807–809.
- Trinklein N.D., Aldred S.F., Hartman S.J., Schroeder D.I., Ottillar R.P., Myers R.M. 2004. An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–66.
- Yang M.Q., Elnitski L.L. 2008. Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics.* **9** (Suppl 2), S3.
- Piontkivska H., Yang M.Q., Larkin D.M., Lewin H.A., Reecy J., Elnitski L. 2009. Cross-species mapping of bidirectional promoters enables prediction of unannotated 5'UTRs and identification of species-specific transcripts. *BMC Genomics.* **10**, 189.