

УДК 577.133.3:57.087.1

## SNP\_TATA\_COMPARATOR: WEB-СЕРВИС ПРИМЕНЕНИЯ УРАВНЕНИЯ РАВНОВЕСИЯ ТВР/ТАТА-КОМПЛЕКСА В СРАВНИТЕЛЬНОЙ ОЦЕНКЕ SNPs ПРОМОТОРОВ ГЕНОВ, СВЯЗАННЫХ С БОЛЕЗНЯМИ ЧЕЛОВЕКА

© 2013 г. Д.А. Рассказов<sup>1</sup>, К.В. Гунбин<sup>1</sup>, П.М. Пономаренко<sup>1</sup>,  
О.В. Вишневский<sup>1,2</sup>, М.П. Пономаренко<sup>1</sup>, Д.А. Афонников<sup>1,2</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики  
Сибирского отделения Российской академии наук, Новосибирск, Россия,  
e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Новосибирский национальный исследовательский государственный университет,  
Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Исследование проявления полиморфизма регуляторных районов генов на уровне их экспрессии имеет важное фундаментальное и прикладное значение. Создан Web-сервис SNP\_TATA\_Comparator для применения оценки *in silico* на основе уравнения равновесия ТВР/ТАТА-комплекса (ТВР, ТАТА-связывающий белок), экспериментально доказанного для биохимических проявлений *in vitro*, к связанным с болезнями SNP (Single nucleotide polymorphism) кор-промоторов генов человека. В результате обеспечен свободный доступ в режиме «реального времени» к анализу имеющейся персонафицированной информации об отклонениях индивидуальных геномов от так называемого референсного генома человека (Вып. 68, Flisek *et al.*, 2011), т. е. варианта генома, который общепринят в качестве стандарта для сравнительного анализа). При этом автоматически учитываются все доказанные транскрипты гена (база данных GENCODE, Вып. 17). Web-сервис SNP\_TATA\_Comparator предназначен для диагностики, мониторинга, профилактики и лечения заболеваний с учетом индивидуальных геномов пациентов в рамках персонафицированной медицины.

**Ключевые слова:** SNP, ТАТА-бокс, ТВР, экспрессия генов, болезни человека.

### ВВЕДЕНИЕ

Голдберг и Хогнесс (Lifton *et al.*, 1978) открыли (A+T)-богатый район длиной 8 п.о. в промоторах генов гистонов дрозофилы, который назвали «ТАТА бокс» по виду его консенсуса ТАТА(t/a)A(t/a)g (синонимы: Goldberg-Hogness box и Hogness box). Затем на ТАТА-боксе промотора гена кональбумина был открыт (Davison *et al.*, 1983) стабильный комплекс ДНК/белок, формирование которого предшествовало связыванию РНК полимеразы II (RNAPII). Год спустя из всего ДНК/белкового комплекса-мишени для связывания RNAPII был выделен (Parker, Topol, 1984) иницирующий транскрипционный фак-

тор TFIID, который связывал ТАТА-бокс промотора гена *HSP70* дрозофилы. Было замечено (Fire *et al.*, 1984), что после связывания RNAPII с анкерным комплексом вокруг ТАТА-бокса на этой основе формируется иной стабильный ДНК/белковый-комплекс, названный преинициаторным из-за начала транскрипции только после его самосборки. Позднее (Schmidt *et al.*, 1989) у дрожжей нашли ген ТАТА-связывающего полипептида ДНК-связывающей субъединицы TFIID (ТАТА-связывающий белок, ТВР). Параллельно для ТАТА бокса рентгеноструктурным анализом были установлены трехмерные структуры свободной В-спирали ДНК (Drew *et al.*, 1981) и комплекса ТАТА-бокса с

ТВР (Kim J. *et al.*, 1993; Kim Y. *et al.*, 1993). Наконец, были измерены (Hahn *et al.*, 1989) уровни  $10^{-9}$  М специфического сродства ТВР к ТАТА-боксу и  $10^{-5}$  М неспецифического – к произвольной ДНК.

В настоящее время ТАТА-боксы – один из самых изученных регуляторных сигналов геномов эукариот (Ponomarenko *et al.*, 2013). Хотя опыт с антителами против ТВР (Wieczorek *et al.*, 1998) продемонстрировал возможность инициации транскрипции *in vitro* без ТВР, все еще не было найдено гена эукариот с такой инициацией транскрипции *in vivo*. Поэтому ТВР/ТАТА-комплекс общепринято считать обязательным для преинициаторного комплекса RNAPII якорем на ДНК (Auble, 2009). С помощью микрочипов была построена полногеномная карта экспериментально доказанных ТАТА-боксов для 17181 генов человека (Yang *et al.*, 2011). Описано более 50 связанных с заболеваниями SNPs (Single nucleotide polymorphisms) ТАТА-боксов промоторов генов человека (Савинкова и др., 2009). В этой связи исследование проявления полиморфизма регуляторных районов генов на уровне их экспрессии имеет важное фундаментальное и прикладное значение. С этой целью был создан Web-сервис SNP\_TATA\_Comparator для сравнения ТАТА-боксов индивидуальных геномов пациентов с референсным геномом человека – URL=<http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>.

## МАТЕРИАЛЫ И МЕТОДЫ

Web-сервис SNP\_TATA\_Comparator анализирует последовательность ДНК  $\{s_{-140} \dots s_{-1}\}$  нуклеотидов  $s \in \{a, t, g, c\}$  длиной 140 п.о. непосредственно перед стартом транскрипции генов человека. Он автоматически берет ее из базы данных Ensembl (Flicek *et al.*, 2011) референсного генома человека, т. е. варианта генома, общепринятого в качестве стандарта для сравнительного анализа с использованием разметки генов из базы данных GENCODE (Harrow *et al.*, 2012), как показано на рисунке на примере связанного с  $\beta$ -талассемией человека SNP A-31g гена *HBB*  $\beta$ -гемоглобина.

Web-сервис SNP\_TATA\_Comparator был написан на языке программирования Perl. Он доступен по URL=<http://beehive.bionet.nsc.ru/>

[cgi-bin/mgs/tatascan/start.pl](http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl). Сначала пользователь получает пустую форму запроса «выбор гена» (рис., слева). После ввода пользователем (непрерывная стрелка) кода или ID интересующего его гена SNP\_TATA\_Comparator по команде «Search» находит (штриховые стрелки) в базе данных Ensembl (Flicek *et al.*, 2011) все документы об этом гене и предлагает пользователю их список для выбора одного из них с целью дальнейшего анализа. Затем (точечные стрелки) по команде «Search» Web-сервис находит в базе данных GENCODE (Harrow *et al.*, 2012) все старты транскрипции этого гена и дает пользователю их список для выбора одного из них для анализа его кор-промотора. Наконец, по команде «GetSeq» он находит (пунктирные стрелки) в базе данных Ensembl (Flicek *et al.*, 2011) фрагмент длиной 140 п.о., локализованный в референсном геноме человека (Вып. 68) непосредственно перед выбранным стартом транскрипции.

Итак, пользователь автоматически получает две копии интересующего его фрагмента референсного генома человека: в качестве стандарта для предстоящего сравнения (окно «Base sequence») и (окно «Editable sequence») для внесения в копию этого стандарта интересующих пользователя индивидуальных отличий.

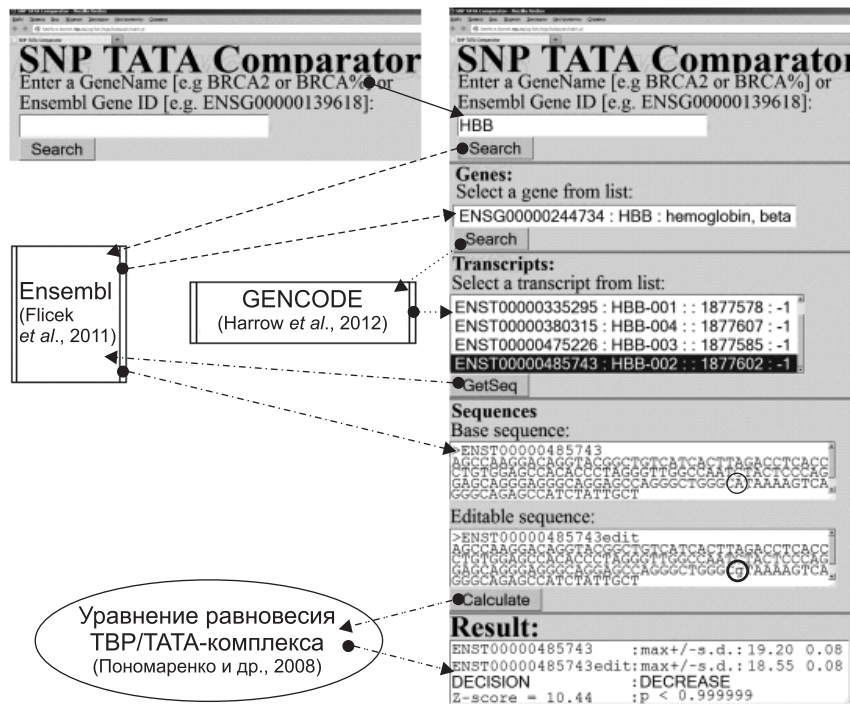
По команде «Calculate» с помощью формул (1–4) SNP\_TATA\_Comparator сравнивает (стрелки «штрих-точка-точка») введенный пользователем частный вариант генома с референсным геномом человека (Вып. 68, Flicek *et al.*, 2011). Окно «Result» содержит результат работы Web-сервиса SNP\_TATA\_Comparator для интересующего пользователя SNP ТАТА-боксов человека. Этот результат интерпретируется согласно приведенному ниже описанию указанных выше формул (1–4).

Прежде всего, Web-сервис SNP\_TATA\_Comparator анализирует единообразно независимо один от другого оба варианта генома человека, «референсный» и «индивидуальный». Каждое положение скользящего окна длиной 26 п.о.  $\{s_{m-13} \dots s_m \dots s_{m+12}\}$  в  $m$ -й позиции характеризуется оценкой равновесной константы диссоциации  $K_D$  комплекса между ДНК и ТАТА-связывающим белком (ТВР), выраженной в натуральных логарифмических единицах ( $\ln$ ) так называемой «аффинности»:

$$\left\{ \begin{aligned}
 &-\ln[K_D](\{\zeta_i\}) = 10,9 - 0,2(\ln[K_{D,dsDNA}](\{\zeta_i\}) + PWM(\{\zeta_i\}) - \ln[K_{D,ssDNA}](\{\zeta_i\})) \\
 &-\ln[K_{D,dsDNA}](\{\zeta_i\}) = \text{mean} [35,1 + 3,4 \sum_{1 \leq k \leq 11} \omega(\zeta_j \zeta_{j+1}) - 0,8 \sum_{\substack{k \leq j \leq k+13 \\ \zeta_j \zeta_{j+1} = TA}} F_{TA}(j)]; \\
 &PWM(\{\zeta_i\}) = \max [ \sum_{1 \leq k \leq 11} \sum_{k+5 \leq j \leq k+14} \text{weight}_{\zeta_j} ]; \\
 &-\ln[K_{D,TBP/ssDNA}](\{\zeta_i\}) = \text{mean} [14,5 + 0,9 \sum_{\substack{k=k_{TATA} \\ \gamma \in \{\zeta-\zeta\}}} \sum_{\substack{k \leq j \leq k+13 \\ \gamma_j \gamma_{j+1} \in WR}} F_{WR}(j) + 2,5 \sum_{\substack{k \leq j \leq k+13 \\ \gamma_j \gamma_{j+1} \in TV}} F_{TV}(j)];
 \end{aligned} \right. \quad (1)$$

здесь: 10,9 ln – величина неспецифического сродства TBP/ДНК ( $\approx 10^{-5}$  М), взятая из эксперимента (Hahn *et al.*, 1989); 0,2 – коэффициент, взятый из работы Пономаренко с соавт. (2008) и численно равный нормированному на число

этапов отношению длин TATA-бокса и «скользящего окна» (здесь:  $(15 \text{ п.о.}/26 \text{ п.о.})/3 = 0,2$ , стехиометрический коэффициент каждого из трех последовательных шагов образования TBP/ДНК-комплекса);  $\text{weight}_{\zeta}$  и  $k_{TATA}$  – матрица



**Рис.** Применение Web-сервиса SNP\_TATA\_Comparator на иллюстративном примере сравнительной оценки для связанной с  $\beta$ -талассемией SNP A-31g (o) промотора гена *HBB*, кодирующего  $\beta$ -цепь гемоглобина человека.

Вверху слева – результат ввода в Интернет адреса <http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl> исходно пустой формы «входных данных» Web-сервиса SNP\_TATA\_Comparator. После ввода в нее (стрелка) информации о целевом гене SNP\_TATA\_Comparator по команде «Search» находит (штриховые стрелки) в базе данных Ensembl (Flicek *et al.*, 2011) все ее документы и предлагает пользователю их список для выбора одного из них для анализа. Затем (точные стрелки) по команде «Search» SNP\_TATA\_Comparator находит в базе данных GENCODE (Harrow *et al.*, 2012) все старты транскрипции этого гена и предлагает пользователю их список для выбора одного из них для анализа его кор-промотора. Далее по команде «GetSeq» он берет (пунктирные стрелки) из базы данных Ensembl (Flicek *et al.*, 2011) фрагмент длиной 140 п.о. референсного генома человека (вып. 68) и предлагает пользователю две его копии: в окне «Base sequence» в качестве стандарта предстоящего сравнения и в окне «Editable sequence» для внесения пользователем в эту копию анализируемых им отличий (o) от указанного стандарта. Наконец, SNP\_TATA\_Comparator по команде «Calculate» (стрелки «штрих-точка-точка») сравнивает введенный пользователем вариант геномной ДНК с референсным геномом человека с помощью формул (1–4) и представляет пользователю результат этого сравнительного анализа, интерпретируемого согласно разделу «Материалы и методы».

Бухера (Bucher, 1990), вес нуклеотида  $\zeta$  в  $k$ -й позиции TATA-box ( $-1 \leq k \leq 13$  относительно канонического варианта  $T_0A_1T_2A_3A_4A_5G_6$  и позиция с максимумом PWM-скора этих весов в «окне сканирования»;  $F_{TA}(i)$ ,  $F_{WR}(i)$ , и  $F_{TV}(i)$  – веса динуклеотидов TA, WR, и TV в  $i$ -х позициях TATA-бокса, взятые из работ (Пономаренко и др., 1997; Ponomarenko *et al.*, 1999);  $\omega(\zeta_i \zeta_{i+1})$ , ширина малой бороздки в ангстремах (Karas *et al.*, 1996), взятая из базы данных (Колчанов и др., 1998); « $-\zeta$ » означает «комплементарный к  $\zeta$ ».

Формула (1) описывает следующие шаги формирования TBP/TATA-комплекса: (i) TBP скользит вдоль ДНК (Coleman, Pugh, 1995) в силу их неспецифического сродства (Hahn *et al.*, 1989) → (ii) остановка скольжения на TATA-боксе (Berg, von Hippel, 1987; Bucher, 1990) → (iii) эндотермическая (Powell *et al.*, 2002) стабилизация TBP/ДНК-комплекса увеличением изгиба оси двойной спирали ДНК от  $19^\circ$  (Drew *et al.*, 1981) до  $90^\circ$  (Kim J. *et al.*, 1993; Kim Y. *et al.*, 1993).

Web-сервис SNP\_TATA\_Comparator находит максимальные оценки  $K_D^{REF}$  и  $K_D^{USER}$  для участка  $[-70; -20]$  локализации всех доказанных к настоящему времени TATA-боксов промоторов эукариот относительно старта транскрипции по отдельности для соответственно

референсного генома человека (Вып. 68) и «введенного» пользователем, как это показано на рис. На этом рисунке можно видеть, что  $K_D^{REF}$  и  $K_D^{USER}$  имеют оценки « $\pm$  s.d.» стандартного отклонения ( $\pm \sigma$ ):

$$\pm \sigma_{\#} = \sqrt{\sum_{j=-13}^{12} \sum_{\xi=1}^3 \ln \left[ \frac{K_{D; s_j \rightarrow \xi}^{\#}}{K_D^{\#}} \right]} / (78 \times 77); \quad (2)$$

здесь:  $\# \in \{REF, USER\}$ ;  $s_j \rightarrow \xi$ , замена нуклеотида  $s_j$  заданной ДНК на другой  $\xi$ .

Формула (2) оценивает величину стандартного отклонения  $K_D^{REF}$  и  $K_D^{USER}$  с помощью всех  $78 = 26 \times 3$  возможных одиночных замен каждого нуклеотида на три других варианта в каждой из 26 позиций «окна сканирования» при том его положении, когда были получены эти максимальные оценки формулы (1). Эти две независимые оценки,  $\sigma_{REF}$  и  $\sigma_{USER}$ , стандартным способом объединяются:

$$\pm \sigma = \sqrt{\sigma_{REF}^2 + \sigma_{USER}^2} \quad (3)$$

На этой основе оценивается 95%-й доверительный интервал,  $\pm \Delta_{95\%} = \tau_{95\%; \nu=78} \sigma$ , для t-критерия Стьюдента недостоверных различий двух сравниваемых вариантов генома, на основе которого SNP\_TATA\_Comparator принимает решение:

$$\left\{ \begin{array}{l} \ln \left[ \frac{K_D^{REF}}{K_D^{USER}} \right] \geq \Delta_{95\%} \xrightarrow{\text{yields}} \text{достоверный ИЗБЫТОК продукта гена;} \\ \ln \left[ \frac{K_D^{REF}}{K_D^{USER}} \right] \leq -\Delta_{95\%} \xrightarrow{\text{yields}} \text{достоверный ДЕФИЦИТ продуктов гена;} \\ \text{ИНАЧЕ} \xrightarrow{\text{yields}} \text{ПРОГНОЗ недостоверный.} \end{array} \right. \quad (4)$$

Результат формулы (4) показан в строке «DECISION» окна «Results» на рис.

Наконец, в качестве дополнительной верификации результата формулы (4) Web-сервис SNP\_TATA\_Comparator оценивает часто используемую Z-статистику (отношение абсолютной разницы оценок к их стандартному отклонению  $|K_D^{REF} - K_D^{USER}| / \sigma$ ) и уровень ее значимости  $p$ . Величины Z и p показаны в нижней строке окна «Results» на рис. и в табл.

## РЕЗУЛЬТАТЫ

Результаты работы Web-сервиса SNP\_TATA\_Comparator по сравнительной оценке связанных

с болезнями человека SNPs промоторов генов представлены на рис. и в табл.

На рис. детально шаг за шагом показан алгоритм применения пользователем Web-сервиса SNP\_TATA\_Comparator на иллюстративном примере A-31g, связанного с  $\beta$ -талассемией SNP промотора гена *HBB*, кодирующего  $\beta$ -цепь гемоглобина (Takahara *et al.*, 1986) – одного из самых изученных наследственных заболеваний.

Как можно видеть, согласно решению SNP\_TATA\_Comparator с помощью формул (1–4), биохимической причиной  $\beta$ -талассемии был дефицит  $\beta$ -цепей гемоглобина, что соответствует данным эксперимента (Takahara *et al.*, 1986). С помощью Web-сервиса SNP\_TATA\_Comparator

**Таблица**  
 Результаты SNP\_TATA\_Comparator по сравнительной оценке SNPs промоторов генов, связанных с болезнями человека

Ген	мРНК	SNP	Геном человека		Заболевание	Результаты SNP_TATA_Comparator				Экспериментально подтверждено (ссылка)
			TATA, REF → USER	K <sub>D</sub> REF ± σ <sub>REF</sub>		K <sub>D</sub> USER ± σ <sub>USER</sub>	Решение	Z	p	
<i>HBB</i>	№ 2	A-31g	gC(A→g)TAAAAg	19,20 ± 0,08	18,55 ± 0,08	β-талассемия	дефицит	10,44	10 <sup>-6</sup>	(Takahara <i>et al.</i> , 1986)
<i>CYP2A6</i>	№ 1	T-48g	agTA(T→g)AAAagg	20,02 ± 0,10	18,52 ± 0,09	рак легких	дефицит	20,70	10 <sup>-6</sup>	(Pitarque <i>et al.</i> , 2001)
<i>SOD1</i>	№ 4	A-27g	ccT(A→g)TAAAgt	19,83 ± 0,09	18,78 ± 0,08	боковой амиотрофический склероз	дефицит	16,67	10 <sup>-6</sup>	(Niemann <i>et al.</i> , 2007)
<i>EDH17B2</i>	№ 3(1)	A-27c	TGATATC(A→c)A	18,14 ± 0,08	17,85 ± 0,09	рак молочной железы	дефицит	4,52	10 <sup>-3</sup>	(Pelotoko <i>et al.</i> , 1994)
<i>DARC</i>	№ 3	T-33c	tcTTA(T→c)CTTgg	19,20 ± 0,08	18,55 ± 0,08	анемия, но устойчивая к малярии	дефицит	10,44	10 <sup>-6</sup>	(Penner, Davie, 1994) (Toumamillt <i>et al.</i> , 1995)
<i>MBL2</i>	№ 1	T-35c	tcTA(T→c)ATAgcc	20,17 ± 0,11	19,28 ± 0,09	риск инсульта ниже, инфекций – выше	дефицит	11,96	10 <sup>-7</sup>	(Boldt <i>et al.</i> , 2006) (Sziller <i>et al.</i> , 2007) (Cervera <i>et al.</i> , 2010)
<i>NOS2</i>	№ 1	t-21c	TATAAATAc(t→c)t	20,17 ± 0,10	20,38 ± 0,10	рассеянный склероз, но устойчивость к малярии	избыток	2,90	10 <sup>-2</sup>	(Охотин и др., 2002) (Clark <i>et al.</i> , 2003)
<i>IL1b</i>	№ 4	C-31t	gc(C→t)ATAAAA	19,21 ± 0,08	20,15 ± 0,09	рак легких, рак печени, язва желудка, гастрит, депрессии	избыток	14,55	10 <sup>-6</sup>	(Wang <i>et al.</i> , 2003) (Wu <i>et al.</i> , 2010) (Martinez-Carrillo <i>et al.</i> , 2010) (Borkowska <i>et al.</i> , 2011)
<i>TAF5L</i>	№ 2	C-25t	cc(C→t)AGCTGAg	15,73 ± 0,07	16,83 ± 0,10	диабет I типа	избыток	17,82	10 <sup>-8</sup>	(Chistiakov <i>et al.</i> , 2005)
<i>F3</i>	№ 3	C-21t	cTTTATAg(c→t)gc	19,59 ± 0,10	19,92 ± 0,10	инфаркт миокарда	избыток	4,57	10 <sup>-3</sup>	(Arnaud <i>et al.</i> , 2000)

были получены аналогичные результаты для генов *SOD1*, *CYP2A6* и *EDH17B2*, дефицит белковых продуктов которых является биохимической причиной развития бокового амиотрофического склероза, рака легких и рака молочной железы соответственно.

В нижней части табл. показаны результаты работы SNP\_TATA\_Comparator в случае диаметрально противоположного биохимического проявления связанных с болезнями человека SNPs промоторов генов *IL1b*, *TAF5L* и *F3*, избыток продуктов которых вызвал патологии. Однако решения, полученные SNP\_TATA\_Comparator с помощью формул (1–4), оказались вновь в согласии с данными экспериментов.

Наконец, в центральной части табл. даны примеры нетривиальных «слабоповреждающих/слабоулучшающих» SNP промоторов генов *DARC*, *MBL2* и *NOS2* человека, вследствие многофункциональности продуктов которых оба, и избыток, и дефицит, оказываются «двуликими», препятствуя одним заболеваниям и способствуя другим. Тем не менее решения (формулы 1–4) предлагаемого Web-сервиса SNP\_TATA\_Comparator снова оказались в согласии с экспериментами.

## ОБСУЖДЕНИЕ

Предсказанный теоретически (формула 1) факт связывания ТВР с ТАТА-боксом за три последовательных шага (Пономаренко и др., 2008) был через год независимо установлен в эксперименте Delgado с соавт. (2009).

Ранее (Пономаренко и др., 2010) были установлены достоверные корреляции между оценками *in silico* формулы (1) и всеми опубликованными данными 68 экспериментов по измерению влияния локального окружения ТАТА-боксов на средство к ним ТВР в условиях 16 типов клеток из 19 видов эукариот, включая простейшие, дрожжи, растения, животных, а также их вирусы. Уравнение (1) было верифицировано экспериментально как для представительной выборки связанных с заболеваниями человека SNPs (Savinkova *et al.*, 2013), так и отдельно для самых сильных повреждений ТАТА-боксов промоторов генов человека (Drachkova *et al.*, 2011).

Все это вместе взятое обосновывает возможность рассмотрения оценок формулы (1)

для предсказания генетически обусловленных изменений экспрессии генов человека. Эти предсказания могут быть приняты во внимание при анализе индивидуальных отклонений генома пациента от референсного генома человека при диагностике, мониторинге, профилактике и лечении для учета генетической предрасположенности пациента к заболеваниям и чувствительности/устойчивости к определенным лекарственным препаратам и терапевтическим воздействиям в рамках развития персонифицированной медицины. Именно эта прикладная задача была решена в настоящей работе путем создания Web-сервиса SNP\_TATA\_Comparator.

## БЛАГОДАРНОСТИ

Работа была поддержана Госконтрактом № 14.512.11.0094 Минобрнауки РФ и Соглашением № 8740 Минобрнауки РФ; Проектом поддержки ведущих научных школ НШ-5278.2012.4; Интеграционным проектом СО РАН № 136 и Программой Президиума РАН «Молекулярная и клеточная биология» (проект 6.6).

## ЛИТЕРАТУРА

- Колчанов Н.А., Пономаренко М.П., Пономаренко Ю.В. и др. Функциональные сайты геномов про- и эукариот: компьютерное моделирование и предсказание активности // Молекуляр. биология. 1998. Т. 32. Вып. 2. С. 255–267.
- Охотин В.Е., Калинин С.Г., Дудина Ю.В. NO-ергическая трансмиссия и NO как объемный нейротрансмиттер. Влияние NO на механизмы синаптической пластичности и эпилептогенез // Усп. физиол. наук. 2002. Т. 33. № 2. С. 41–55.
- Пономаренко М.П., Савинкова Л.К., Пономаренко Ю.В. и др. Моделирование последовательностей ТАТА-боксов генов эукариот // Молекуляр. биология. 1997. Т. 31. Вып. 4. С. 726–732.
- Пономаренко П.М., Савинкова Л.К., Драчкова И.А. и др. Пошаговая модель связывания ТВР/ТАТА-бокс позволяет предсказать наследственное заболевание человека по точечному полиморфизму // Докл. АН. 2008. Т. 419. Вып. 6. С. 828–832.
- Пономаренко П.М., Суслов В.В., Савинкова Л.К. и др. Точное уравнение равновесия четырех шагов связывания ТВР с ТАТА-боксом для прогноза фенотипического проявления мутаций // Биофизика. 2010. Т. 55. Вып. 3. С. 400–414.
- Савинкова Л.К., Пономаренко М.П., Пономаренко П.М. и др. Полиморфизмы ТАТА-боксов промоторов генов человека и ассоциированные с ними наследственные патологии // Биохимия. 2009. Т. 74. Вып. 2. С. 149–163.

- Arnaud E., Barbalat V., Nicaud V. *et al.* Polymorphisms in the 5' regulatory region of the tissue factor gene and the risk of myocardial infarction and venous thromboembolism: the ECTIM and PATHROS studies. *Etude Cas-Temoin de l'Infarctus du Myocarde. Paris Thrombosis case-control Study // Arterioscler. Thromb. Vasc. Biol.* 2000. V. 20. No. 3. P. 892–898.
- Auble D.T. The dynamic personality of TATA-binding protein // *Trends Biochem. Sci.* 2009. V. 34. No. 2. P. 49–52.
- Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters // *J. Mol. Biol.* 1987. V. 193. No. 4. P. 723–750.
- Boldt A.B., Culpi L., Tsuneto L.T. *et al.* Diversity of the MBL2 gene in various Brazilian populations and the case of selection at the mannose-binding lectin locus // *Hum. Immunol.* 2006. V. 67. No. 9. P. 722–734.
- Borkowska P., Kucia K., Rzezniczek S. *et al.* Interleukin-1 $\beta$  promoter (-31T/C and -511C/T) polymorphisms in major recurrent depression // *J. Mol. Neurosci.* 2011. V. 44. No. 1. P. 12–16.
- Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences // *J. Mol. Biol.* 1990. V. 212. No. 4. P. 563–578.
- Cervera A., Planas A.M., Justicia C. *et al.* Genetically-defined deficiency of mannose-binding lectin is associated with protection after experimental stroke in mice and outcome in human stroke // *PLoS ONE.* 2010. V. 5. No. 2. P. e8433.
- Chistiakov D.A., Chernisheva A., Savost' anov K.V. *et al.* The TAF5L gene on chromosome 1q42 is associated with type 1 diabetes in Russian affected patients // *Autoimmunity.* 2005. V. 38. No. 4. P. 283–293.
- Clark I.A., Rockett K.A., Burgner D. Genes, nitric oxide and malaria in African children // *Trends Parasitol.* 2003. V. 19. No. 8. P. 335–337.
- Coleman R.A., Pugh B.F. Evidence for functional binding and stable sliding of the TATA binding protein on non-specific DNA // *J. Biol. Chem.* 1995. V. 270. No. 23. P. 13850–13859.
- Davison B.L., Egly J.M., Mulvihill E.R., Chambon P. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences // *Nature.* 1983. V. 301. No. 5902. P. 680–686.
- Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism // *Biochemistry.* 2009. V. 48. No. 8. P. 1801–1809.
- Drachkova I.A., Ponomarenko P.M., Arshinova T.V. *et al.* *In vitro* examining the existing prognoses how TBP binds to TATA with SNP associated with human diseases // *Health.* 2011. V. 3. No. 9. P. 577–583.
- Drew H.R., Wing R.M., Takano T. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics // *Proc. Natl Acad. Sci. USA.* 1981. V. 78. No. 4. P. 2179–2183.
- Fire A., Samuels M., Sharp P.A. Interactions between RNA polymerase II, factors, and template leading to accurate transcription // *J. Biol. Chem.* 1984. V. 259. No. 4. P. 2509–2516.
- Flicek P., Amode M.R., Barrell D. *et al.* Ensembl 2011 // *Nucl. Acids Res.* 2011. V. 39. Database issue. P. D800–D806.
- Hahn S., Buratowski S., Sharp P.A., Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non consensus DNA sequences // *Proc. Natl Acad. Sci. USA.* 1989. V. 86. No. 15. P. 5718–5722.
- Harrow J., Frankish A., Gonzalez J.M. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project // *Genome Res.* 2012. V. 22. No. 9. P. 1760–1774.
- Karas H., Knuppel R., Schulz W. *et al.* Combining structural analysis of DNA with search routines for the detection of transcription regulatory elements // *Comput. Applic. Biosci.* 1996. V. 12. No. 5. P. 441–446.
- Kim J.L., Nikolov D.B., Burley S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element // *Nature.* 1993. V. 365. No. 6446. P. 520–527.
- Kim Y., Gieger J.H., Hahn S., Sigler P.B. Crystal structure of a yeast TBP/TATA-box complex // *Nature.* 1993. V. 365. No. 6446. P. 512–520.
- Lifton R., Goldberg M., Karp R., Hogness D. The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications // *Cold Spring Harb. Symp. Quant. Biol.* 1978. V. 42. Pt. 2. P. 1047–1051.
- Martinez-Carrillo D.N., Garza-Gonzalez E., Betancourt-Linares R. *et al.* Association of IL1B -511C/-31T haplotype and *Helicobacter pylori vacA* genotypes with gastric ulcer and chronic gastritis // *BMC Gastroenterol.* 2010. V. 10. P. 126.
- Niemann S., Broom W.J., Brown R.H. Jr. Analysis of a genetic defect in the TATA box of the SOD1 gene in a patient with familial amyotrophic lateral sclerosis // *Muscle Nerve.* 2007. V. 36. No. 5. P. 704–707.
- Parker C.S., Topol J. A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene // *Cell.* 1984. V. 37. No. 1. P. 273–283.
- Peltoketo H., Piao Y., Mannermaa A. *et al.* A point mutation in the putative TATA box, detected in nondiseased individuals and patients with hereditary breast cancer, decreases promoter activity of the 17 beta-hydroxysteroid dehydrogenase type 1 gene 2 (EDH1B2) in vitro // *Genomics.* 1994. V. 23. No. 1. P. 250–252.
- Penner C.G., Davie J.R. Transcription factor GATA-1-multi-protein complexes and chicken erythroid development // *FEBS Lett.* 1994. V. 342. No. 3. P. 273–277.
- Pitarque M., von Richter O., Oke B. *et al.* Identification of a single nucleotide polymorphism in the TATA box of the CYP2A6 gene: impairment of its promoter activity // *Biochem. Biophys. Res. Commun.* 2001. V. 284. No. 2. P. 455–460.
- Ponomarenko M., Mironova V., Gunbin K., Savinkova L. Hogness Box // *Brenner's Encyclopedia of Genetics.* 2nd edn. / Eds S. Maloy, K. Hughes. San Diego: Academic Press, Elsevier Inc., 2013. V. 3. P. 491–494.
- Ponomarenko M.P., Ponomarenko J.V., Frolov A.S. *et al.* Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins // *Bioinformatics.* 1999. V. 15. No. 7/8. P. 687–703.
- Powell R.M., Parkhurst K.M., Parkhurst L.J. Comparison of TATA-binding protein recognition of a variant and consensus DNA promoters // *J. Biol. Chem.* 2002. V. 277. No. 10. P. 7776–7784.

- Savinkova L.K., Drachkova I.A., Arshinova T.V. *et al.* An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein // *PLoS ONE*. 2013. V. 8. No. 2. P. e54626.
- Schmidt M.C., Kao C.C., Pei R., Berk A.J. Yeast TATA-box transcription factor gene // *Proc. Natl Acad. Sci. USA*. 1989. V. 86. No. 20. P. 7785–7789.
- Sziller I., Babula O., Hupuczi P. *et al.* Mannose-binding lectin (MBL) codon 54 gene polymorphism protects against development of pre-eclampsia, HELLP syndrome and pre-eclampsia-associated intrauterine growth restriction // *Mol. Hum. Reprod*. 2007. V. 13. No. 4. P. 281–285.
- Takahara Y., Nakamura T., Yamada H. *et al.* A novel mutation in the TATA box in a Japanese patient with beta + -thalassemia // *Blood*. 1986. V. 67. No. 2. P. 547–550.
- Tournamille C., Colin Y., Cartron J.P., Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals // *Nat. Genet*. 1995. V. 10. No. 2. P. 224–228.
- Wang Y., Kato N., Hoshida Y. *et al.* Interleukin-1beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection // *Hepatology*. 2003. V. 37. No. 1. P. 65–71.
- Wieczorek E., Brand M., Jacq X., Tora L. Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II // *Nature*. 1998. V. 393. No. 6681. P. 187–191.
- Wu K.S., Zhou X., Zheng F. *et al.* Influence of interleukin-1 beta genetic polymorphism, smoking and alcohol drinking on the risk of non-small cell lung cancer // *Clin. Chim. Acta*. 2010. V. 411. No. 19/20. P. 1441–1446.
- Yang M.Q., Laflamme K., Gotea V. *et al.* Genome-wide detection of a TFIID localization element from an initial human disease mutation // *Nucl. Acids Res*. 2011. V. 39. No. 6. P. 2175–2187.

**SNP\_TATA\_COMPARATOR: WEB SERVICE  
FOR COMPARISON OF SNPs WITHIN GENE PROMOTERS ASSOCIATED  
WITH HUMAN DISEASES USING THE EQUILIBRIUM EQUATION  
OF THE TBP/TATA COMPLEX**

**D.A. Rasskazov<sup>1</sup>, K.V. Gunbin<sup>1</sup>, P.M. Ponomarenko<sup>1</sup>,  
O.V. Vishnevsky<sup>1,2</sup>, M.P. Ponomarenko<sup>1</sup>, D.A. Afonnikov<sup>1,2</sup>**

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Novosibirsk National Research State University, Novosibirsk, Russia

**Summary**

Web-service SNP\_TATA\_Comparator has been developed for using experimentally proven *in silico* evaluation of *in vivo* biochemical manifestations of SNPs in the core promoter regions of human genes associated with diseases on the base of the TBP/TATA complex equilibrium equation (TBP is TATA binding protein). Data of individual deviations from the reference human genome (Ensembl, rel. 68, i.e. the variant of human genome commonly accepted as datum in comparison analysis) are accessible for free in the real-time mode. Also, data from the GENCODE database (rel. 17) are automatically invoked. The reported Web service SNP\_TATA\_Comparator is designed for personalized medicine using individual genomes in diagnosis, monitoring, prevention, and treatment.

**Key words:** SNP, TATA box, TBP, gene expression, human diseases.