

## Кандидатные SNP-маркеры, изменяющие сродство TBP к промоторам Y-связанных генов *CDY2A*, *SHOX*, *ZFY*, снижают ряд показателей репродуктивного потенциала мужчин

М.П. Пономаренко<sup>1,2</sup>✉, Е.Б. Шарыпова<sup>1</sup>, И.А. Драчкова<sup>1</sup>, Л.К. Савинкова<sup>1</sup>, И.В. Чадаева<sup>1</sup>, Д.А. Рассказов<sup>1</sup>, П.М. Пономаренко<sup>1</sup>, Л.В. Осадчук<sup>1</sup>, А.В. Осадчук<sup>1</sup>

<sup>1</sup> Федеральное исследовательское учреждение Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

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

✉ e-mail: pon@bionet.nsc.ru

**Аннотация.** Репродуктивный потенциал – уровень физического и психического состояния организма, позволяющий при достижении социальной зрелости воспроизводить здоровое потомство. В узком биомедицинском смысле определение включает комплекс функциональных показателей репродуктивной системы, но в более широком смысле его можно рассматривать как совокупность физиологических, поведенческих, адаптивных, ментальных, антропометрических и генетических характеристик особи, способствующих размножению. Целью настоящей работы было расширить область применимости созданного ранее Web-сервиса SNP\_TATA\_Z-tester для поиска кандидатных маркеров однонуклеотидного полиморфизма (SNP) на хромосоме Y человека, связанных с мужским репродуктивным потенциалом (МРП). В поиске кандидатных SNP-маркеров для МРП мы сосредоточились на генах хромосомы Y человека. Изучены 35 SNP в промоторах генов *CDY2A*, *SHOX* и *ZFY*, представляющих все три типа генов хромосомы Y человека: уникальный, псевдоаутосомный и паралог гена хромосомы X человека соответственно. Предсказаны 11 кандидатных SNP-маркеров ослабления МРП из-за изменения сродства TATA-связывающего белка (TBP) к этим промоторам. Выборочно верифицированы *in vitro* величины сродства «TBP-промотор», предсказанные в этой работе. Установлена достоверная корреляция ( $r = 0.94$ ,  $p < 0.005$ ) между ними и результатами измерения *in vitro* сродства TBP человека к олигонуклеотидам, идентичным сайтам TBP-связывания исследуемых промоторов. Проведя поиск в базе данных PubMed по ключевым словам, мы нашли клиническое описание патологических состояний человека, соответствующих изменению экспрессии генов, несущих предсказанные нами кандидатные SNP-маркеры. Среди них оказались такие патологии, как нарушение сперматогенеза (*ZFY*: rs1388535808 и rs996955491), задержка полового созревания (*CDY2A*: rs200670724), нарушения эмбриогенеза (*SHOX*: rs28378830) и непропорционально низкий рост с деформациями Маделунга (*SHOX*: rs1452787381). Они свидетельствуют, что в случае SNP-промоторов генов хромосомы Y человека следует ожидать изменений широкого круга показателей МРП, выходящих далеко за рамки генетического контроля собственно мужской репродуктивной функции.

Ключевые слова: человек; репродуктивный потенциал; хромосома Y; ген; промотор; TATA-связывающий белок (TBP); TBP-связывающий сайт (TATA-бокс); однонуклеотидный полиморфизм (SNP); SNP-маркер; верификация *in vitro*.

**Для цитирования:** Пономаренко М.П., Шарыпова Е.Б., Драчкова И.А., Савинкова Л.К., Чадаева И.В., Рассказов Д.А., Пономаренко П.М., Осадчук Л.В., Осадчук А.В. Кандидатные SNP-маркеры, изменяющие сродство TBP к промоторам Y-связанных генов *CDY2A*, *SHOX*, *ZFY*, снижают ряд показателей репродуктивного потенциала мужчин. *Вавиловский журнал генетики и селекции*. 2020;24(7):785-793. DOI 10.18699/VJ20.674

## Candidate SNP-markers altering TBP binding affinity for promoters of the Y-linked genes *CDY2A*, *SHOX*, and *ZFY* are lowering many indexes of reproductive potential in men

M.P. Ponomarenko<sup>1,2</sup>✉, E.B. Sharypova<sup>1</sup>, I.A. Drachkova<sup>1</sup>, L.K. Savinkova<sup>1</sup>, I.V. Chadaeva<sup>1</sup>, D.A. Rasskazov<sup>1</sup>, P.M. Ponomarenko<sup>1</sup>, L.V. Osadchuk<sup>1</sup>, A.V. Osadchuk<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

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

✉ e-mail: pon@bionet.nsc.ru

**Abstract.** Reproductive potential is the most important conditional indicator reflecting the ability of individuals in a population to reproduce, survive and develop under optimal environmental conditions. As for humans, the concept of reproductive potential can include the level of the individual's mental and physical state, which allows them to reproduce healthy offspring when they reach social and physical maturity. Female reproductive potential has been investigated in great detail, whereas the male reproductive potential (MRP) has not received the equal amount of attention as yet. Therefore, here we focused on the human Y chromosome and found candidate single-nucleotide polymorphism (SNP) markers of MRP. With our development named Web-service SNP\_TATA\_Z-tester, we examined *in silico* all 35 unannotated SNPs within 70-bp

proximal promoters of the three Y-linked genes, *CDY2A*, *SHOX* and *ZFY*, which represent all types of human Y-chromosome genes, namely: unique, pseudo-autosomal, and human X-chromosome gene paralogs, respectively. As a result, we found 11 candidate SNP markers for MRP, which can significantly alter the TATA-binding protein (TBP) binding affinity for promoters of these genes. First of all, we selectively verified *in vitro* the values of the TBP-promoter affinity under this study, Pearson's linear correlation between predicted and measured values of which were  $r = 0.94$  (significance  $p < 0.005$ ). Next, as a discussion, using keyword search tools of the PubMed database, we found clinically proven physiological markers of human pathologies, which correspond to a change in the expression of the genes carrying the candidate SNP markers predicted here. These were markers for spermatogenesis disorders (*ZFY*: rs1388535808 and rs996955491), for male maturation arrest (*CDY2A*: rs200670724) as well as for disproportionate short stature at Madelung deformity (e.g., *SHOX*: rs1452787381) and even for embryogenesis disorders (e.g., *SHOX*: rs28378830). This indicates a wide range of MRI indicators, alterations in which should be expected in the case of SNPs in the promoters of the human Y-chromosome genes and which can go far beyond changes in male fertility.

Key words: human; reproductive potential; chromosome Y; gene; promoter; TATA-binding protein (TBP); TBP-binding site (TATA-box); single nucleotide polymorphism (SNP); candidate SNP marker; verification *in vitro*.

**For citation:** Ponomarenko M.P., Sharypova E.B., Drachkova I.A., Savinkova L.K., Chadaeva I.V., Rasskazov D.A., Ponomarenko P.M., Osadchuk L.V., Osadchuk A.V. Candidate SNP-markers altering TBP binding affinity for promoters of the Y-linked genes *CDY2A*, *SHOX*, and *ZFY* are lowering many indexes of reproductive potential in men. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2020;24(7):785-793. DOI 10.18699/VJ20.674 (in Russian)

## Введение

Следуя идеям R.N. Chapman (1931) и E.R. Pianka (1976), популяционные экологи используют концепцию репродуктивного потенциала в качестве самого важного показателя шансов воспроизвести потомков и довести их до репродуктивного возраста при наилучших условиях как на уровне индивида, так и на уровне популяции (Axelsson et al., 2010). Теория ожидаемой продолжительности жизни (Bowles, 1998) связывает репродуктивный потенциал, качество и продолжительность жизни человека с предрасположенностью к болезням и стрессам, закодированной в его геноме. Наконец, прогресс медицины, достижения науки, развитие технологий и улучшение образования позволяют увеличить репродуктивный потенциал суперпопуляции людей, хотя рост народонаселения, урбанизация, загрязнение окружающей среды и эпидемии могут его уменьшить.

Базовая концепция предиктивно-превентивной персонализированной медицины (Trovato, 2014), предполагающая достоверные различия между больными и здоровыми добровольцами по встречаемости у них клинических маркеров однонуклеотидного полиморфизма (SNP) (Varzari et al., 2016), указывает на возможность по расшифрованным индивидуальным геномам людей улучшить качество и продолжительность жизни как их, так и потомков. По SNP-маркерам генов, контролирующим показатели репродуктивного потенциала в геноме пациента, врач может прогнозировать способность к зачатию, предрасположенность к репродуктивным нарушениям, стрессорную реактивность и хронические болезни, способные ухудшить его репродуктивное здоровье, качество жизни и долголетие, а также рекомендовать адекватный образ жизни, профилактические мероприятия, диету и форму взаимодействия пациента с врачом во избежание развития нежелательного жизненного сценария.

Крупнейший научный проект XXI в., «1000 геномов» (Telenti et al., 2016), как основа предиктивно-превентивной персонализированной медицины, уже выявил сотни миллионов SNP, собранных в базе данных dbSNP (Sherry et al., 2001), между десятками тысяч расшифрованных индивидуальных геномов людей и эталонным (референсным)

геномом человека в базе данных Ensembl (Cunningham et al., 2019). Наконец, база данных dbWGF (Wu et al., 2016) собирает, систематизирует и приоритизирует сведения о каждом из всех 10 млрд потенциально возможных SNP человека в качестве первоосновы анализа индивидуального генома пациента в предиктивно-превентивной персонализированной медицине (Trovato, 2014).

Поскольку врачебное решение на основе наличия или отсутствия тех или иных SNP-маркеров в геноме пациента нацелено на его здоровье, продолжительность и качество жизни, то приемлемы только биомедицинские SNP-маркеры, клинически доказанные сравнением больных и условно здоровых людей (Varzari et al., 2016). Учет необходимых затрат времени, ручного труда и финансов для каждого такого теста исключает возможность оценить на практике проявление каждого из 10 млрд SNP человека (Wu et al., 2016) в патогенезе каждой из 55 тыс. болезней человека, согласно Международной классификации болезней и сопутствующих проблем со здоровьем людей, МКБ-11 (Pocai, 2019). Однако остается неясным, необходимо ли клинически проверять все SNP человека при условии, что дилемма J.B.S. Haldane (1957) и теория нейтральной эволюции (Kimura, 1968) предсказывают нейтральность абсолютного большинства SNP человека. Сейчас для будущего клинического теста чаще всего вручную эвристически выбирают кандидатный SNP-маркер заданной патологии среди всех SNP генов, уже связанных с этой болезнью (Varzari et al., 2016). Выбор таких кандидатных SNP-маркеров можно сделать объективнее, быстрее и прицельнее за счет малозатратного биоинформатического выявления абсолютного большинства нейтральных SNP во всем референсном геноме человека (Ponomarenko M. et al., 2017). Хотя точность биоинформатических прогнозов для SNP все еще остается ниже порога их применимости в клинической практике (Yoo et al., 2015), она непрерывно увеличивается с каждым годом (Putlyayeva et al., 2018).

Сейчас наилучшую точность биоинформатических прогнозов достигли для SNP белок-кодирующих частей генов (Amberger et al., 2015), чье негативное влияние неустранимо ни сменой образа жизни, ни терапией (Mitsuasu et al., 1998). В то же время значительно меньше

исследованы SNP регуляторных районов генов (Zerbino et al., 2015), лишь модулирующих уровень экспрессии, что корректируемо как сменой образа жизни, так и лекарствами (Ponomarenko M. et al., 2013)). Поэтому регуляторные SNP сайта связывания TATA-связывающего белка (ТВР, каноническая форма сайта – TATA-бокс,  $\approx 15\%$  TATA-содержащих генов у человека), который необходим перед стартом любого транскрипта (Martianov et al., 2002; Rhee, Pugh, 2012) и который модулирует транскрипцию пропорционально сродству ТВР-промотор (Mogno et al., 2010), выглядят многообещающими в силу их предсказуемости и пользы для медицины (Ponomarenko M. et al., 2013, 2017).

Ранее мы создали Web-сервисы SNP\_TATA\_Comparator (Ponomarenko M. et al., 2015) и SNP\_TATA\_Z-tester (Sharypova et al., 2018) для оценки статистической значимости влияния SNP на сродство ТВР к промоторам генов человека и других эукариот с использованием библиотеки BioPerl (Stajich et al., 2002), доступа к базе данных Ensembl (Cunningham et al., 2019) и с помощью других источников данных соответственно. На этой методической основе мы предсказали кандидатные SNP-маркеры ожирения (Arkova et al., 2015), агрессивности (Chadaeva et al., 2016), нарушений циркадного ритма (Ponomarenko P. et al., 2016), атеросклероза (Ponomarenko M. et al., 2019), аутоиммунных заболеваний (Ponomarenko M. et al., 2016), болезни Альцгеймера (Ponomarenko P. et al., 2017), резистентности к противоопухолевой химиотерапии (Turnaev et al., 2016), предрасположенности к социальному доминированию и подчинению у людей (Chadaeva et al., 2019).

Целью настоящей работы было расширить область применимости наших Web-сервисов (Ponomarenko M. et al., 2015; Sharypova et al., 2018) для оценки SNP сайтов ТВР-связывания посредством поиска кандидатных SNP-маркеров на хромосоме Y человека, связанных с показателями мужского репродуктивного потенциала (МРП). Эта методология была использована при изучении женского репродуктивного потенциала (Chadaeva et al., 2018), тогда как МРП еще не получил должного внимания.

## Материалы и методы

**Биоинформатический анализ данных.** Мы исследовали последовательности ДНК  $S = \{s_{-70} \dots s_{-1}\}$  длиной 70 п.о. перед стартами транскрипции ( $TSS \equiv s_0$ ; где  $s_i \in \{a, c, g, t\}$ ) промоторов белок-кодирующих генов хромосомы Y человека, как показано на рис. 1 и описано ранее, например в недавней работе (Ponomarenko M. et al., 2019). Стрелками на рис. 1 указано, как с помощью (а) Web-сервиса UCSC Genome Browser (Haeussler et al., 2015) и (б) базы данных dbSNP (Sherry et al., 2001) мы вводили исходные данные в наш (в) Web-сервис SNP\_TATA\_Z-tester (Sharypova et al., 2018), где окно “Result” содержит результат обработки этих исходных данных в рамках нашей биоинформатической модели трехшагового связывания ТВР/промотор (Ponomarenko P. et al., 2008).

Каждый SNP анализировали независимо от остальных таким образом, что если было предсказано неустойчивое изменение сродства ТВР к минорному варианту промотора с этим SNP в сравнении с нормой, то он исключался из дальнейшего анализа (данные не показаны). Для оставшихся SNP с помощью стандартного поиска в базе

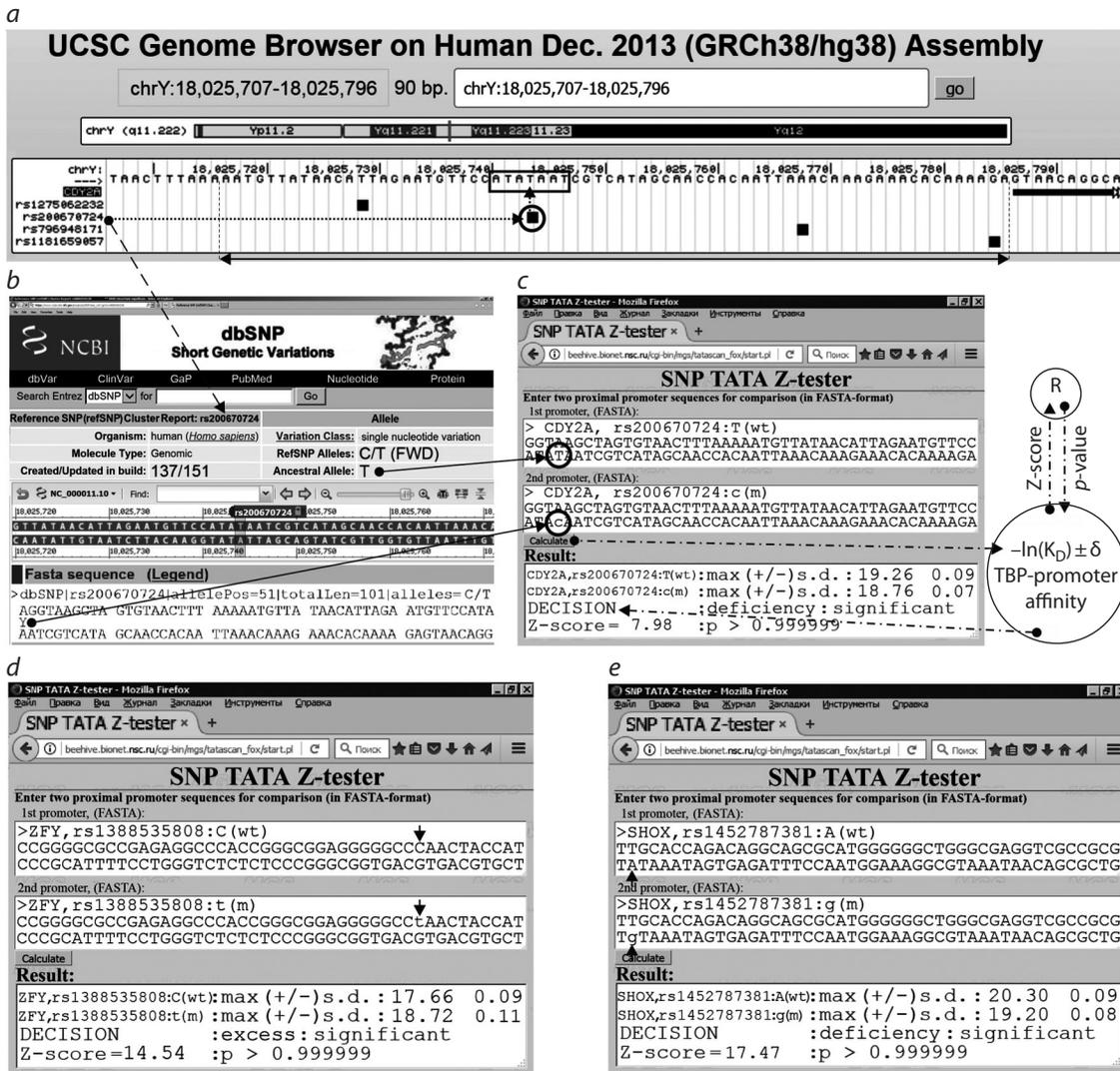
данных PubMed (Lu, 2011) по ключевым словам находили сообщения о клинически значимых показателях МРП, которые соответствовали значимым изменениям экспрессии генов, несущих предсказанные кандидатные SNP, вследствие достоверного изменения сродства ТВР к промоторам этих генов.

**Эксперимент *in vitro*.** Рекомбинантный ТВР человека (hТВР) экспрессировали в клетках *Escherichia coli* BL21(DE3) с плазмиды pAR3038-hТВР (любезно предоставлена проф. В. Puhg, The Pennsylvania State University, Pennsylvania, США) согласно (Peterson et al., 1990; Pugh, 1995), за исключением концентрации ИПТГ (1 мМ вместо 0.1 мМ) и времени индукции (3 ч вместо 1.5 ч).

В работе использовали олигодезоксирибонуклеотиды (ODN) длиной 26 п. о., синтезированные и дополнительно очищенные электрофорезом в ПААГ («Биоссет», Новосибирск). Для получения меченых двуцепочечных ODN обе цепи метили  $^{32}\text{P}$ -АТФ («Биосан», Новосибирск) с помощью T4-полинуклеотидкиназы («СибЭнзим», Новосибирск), отжигали при 95 °C (в эквимольном соотношении) и медленно (не менее 3 ч) охлаждали до комнатной температуры. Отожженные дуплексы очищали и анализировали электрофорезом в 15 % ПААГ в неденатурирующих условиях с последующей радиоавтографией на фосфоримиджере Molecular Imager PharosFX Plus (Bio-Rad) (Drachkova et al., 2005). Немеченые дуплексы получали так же и использовали без дополнительной очистки в ПААГ.

Для определения равновесной константы диссоциации ( $K_D$ ), а также времени полураспада ( $t_{1/2}$ ) констант скорости образования ( $k_a$ ) и распада ( $k_d$ ) комплексов ТВР-ODN проводили не менее трех экспериментов по связыванию ТВ-ODN при 25 °C в буфере, содержащем 20 мМ HEPES-KOH (pH 7.6), 5 мМ  $\text{MgCl}_2$ , 70 мМ KCl, 1 мМ DTT, 100 мкг/мл BSA, 0.01 % NP-40, и 5 % глицерина с фиксированной концентрацией 2 нМ активного ТВР. Каждый эксперимент включал 32 реакции связывания (8 временных точек, каждая с четырьмя концентрациями ODN). Четверка реакций связывания (1 временная точка) запускалась одновременно добавлением ТВР к реакционной смеси, содержащей ODN, и немедленным переносом в термостат на 25 °C. По окончании реакции связывания все реакционные смеси наносили на ПААГ одновременно при напряженности поля 10 В/см. Комплексы ТВР-ODN отделяли от несвязанного ODN с помощью количественного метода электрофоретической задержки (EMSA). Электрофорез проводили в нативном 5 % ПААГ на Tris-глициновом буфере (pH 8.3) в течение 40 мин при температуре 10 °C и напряженности поля 25 В/см.

Затем гели высушивали и экспонировали с экраном Imaging Screen-K (Kodak) для фосфоимиджера Molecular Imager PharosFX Plus (Bio-Rad). Экран сканировали на фосфоимиджере и с помощью программы Quantity One – 4.5.0 (Bio-Rad) проводили количественный анализ радиоавтографов. Полученные зависимости концентрации комплексов ТВР-ODN от концентраций ODN вводили как исходные данные в общедоступный Web-сервис GraphPad Prism 5 (<http://graphpad-prism.software.informer.com/5.01>), результатом которого были величины  $K_D$ ,  $t_{1/2}$ ,  $k_a$  и  $k_d$  анализируемых комплексов ТВР-ODN, а также оценки стандартных ошибок средних ( $\pm$  SEM) для этих величин.



**Fig. 1.** Analysis of the 70 base long promoter (↔) upstream of the transcription start of the human *CDY2A* gene with four unannotated SNPs (■). One of them (rs200670724; dashed arrow) has been proposed as a candidate marker for human male maturation delay and poorer reproduction potential caused by poorer expression of the gene (Stahl et al., 2012) associated with the presence of the SNP in the TBP binding site (boxed).

(a) Presentation of the human *CDY2A* promoter with the UCSC Genome Browser (Haeussler et al., 2015). (b) Annotation of rs200670724 SNP in dbSNP (Sherry et al., 2001). (c) The SNP\_TATA\_Z-tester application formerly developed by us. Line "DECISION" in its "Result" window present the prediction of *CDY2A* deficiency with the minor allele of the promoter as compared to the ancestral SNP variant. (R) means the R software (Waardenberg et al., 2015). (d, e) Corresponding predictions of lower male reproductive potentials done by SNP\_TATA\_Z-tester for candidate SNP markers rs1388535808 and rs1452787381 in the *ZFY* and *SHOX* genes, respectively.

**Статистический анализ.** Биоинформатические оценки сродства ТБП к вариантам SNP-промоторов генов хромосомы Y человека сравнивали с экспериментальными величинами сродства ТБП к ODN, идентичным этим промоторам, с помощью STATISTICA (Statsoft™).

### Результаты и обсуждение

Проанализированы все 35 неаннотированных SNP, доступных в выпуске № 151 базы данных dbSNP (Sherry et al., 2001), для проксимальных промоторов длиной 70 п. о. генов *CDY2A*, *SHOX* и *ZFY*, представляющих все три типа генов на хромосоме Y человека: уникальный, псевдоаутосомный и паралог гена хромосомы X человека

соответственно (табл. 1). В результате было выявлено 11 кандидатных SNP-маркеров снижения МРП из-за достоверных изменений сродства ТБП к этим промоторам (см. табл. 1). Рассмотрим сначала наиболее подробно единственный кандидатный SNP-маркер rs200670724 в гене *CDY2A*, связанный с ослаблением МРП, чтобы затем, благодаря аналогиям с этим иллюстративным примером, кратко описать 10 остальных кандидатных SNP-маркеров в генах *ZFY* и *SHOX*.

Уникальный ген *CDY2A* на хромосоме Y человека кодирует хромосома 2A и несет четыре неаннотированных SNP в своем проксимальном промоторе длиной 70 п. о. (см. рис. 1, a), лишь один среди которых, rs2276109 (см.

**Table 1.** Candidate SNP markers of male reproductive potential (MRP) capable of altering the affinity of TBP for the promoters of the human Y chromosome genes (according to our predictions made in this work)

Gene	dbSNP (Sherry et al., 2001)	5'-flank	wt mut	3'-flank	K <sub>D</sub> , nM				Candidate SNP marker	♂	References
					wt mut	Δ	Z	α			
<i>CDY2A</i> (unique male-specific protein-coding gene)	rs200670724	atgttcata	t c	aatcgatca	4 7	<	7	10 <sup>-6</sup>	Increased risks of male maturation arrest	↓	Stahl et al., 2012
	rs1388535808	ggagggggcc	c t	aactaccatc	21 7	>	15	10 <sup>-6</sup>	Increased risks of azoospermia and infertility	↓	Jan et al., 2018
<i>ZFY</i> (paralogous to X-linked genes)	rs996955491	gggggccaa	c a	taccatccc	21 13	>	8	10 <sup>-6</sup>		↓	
	<i>SHOX</i> (pseudo-autosomal gene)	rs1452787381	gtcgcgcgt	a g	taaagtga	2 5	<	17	10 <sup>-6</sup>	Increased risks of disproportionate short stature and Madelung deformity	↓
rs1405831103		gtcgcgcgt	ataa t	atagtgat	2 8	<	22	10 <sup>-6</sup>	↓		
rs1273755135		tcgccgcgt	t c	aaatagtag	2 4	<	17	10 <sup>-6</sup>	↓		
rs375938368		aggctgcgc	g t	tataaatag	1.5 1.9	<	3	10 <sup>-3</sup>	↓		
rs771395540		caggagcca	t c	aggggtctc	62 70	<	2	0.05	↓		
rs28378830		aggggtctc	g a*	agtcaccctg	62 30	>	13	10 <sup>-6</sup>	Increased risks of pathoem- bryogenesis	↓	Brosens et al., 2014
rs894540003		caataggggt	c a	ttcaggtcac	62 23	>	15	10 <sup>-6</sup>		↓	
rs970127768	cgaggtcgc	g a*	cgtataata	1.5 1.3	>	2	0.05	↓			

Note. wt – norm (reference human genome); mut – minor allele; K<sub>D</sub> – equilibrium dissociation constant of TBP-promoter complex; Δ – alteration in the affinity of TBP for promoter, namely: increase (<) and decrease (>); Z – Z-value of Fisher's Z-test; α – statistical significance estimate (α = 1 – p, where: p, probability estimate for random reasons shown in Fig. 1, d); ♂ – alteration in MRP, namely: increase (↑) and decrease (↓). Genes: *CDY2A* – chromodomain Y-linked 2A; *ZFY* – zinc finger protein Y-linked; *SHOX* – short stature homeobox.

\* In addition to the indicated minor allele, other alleles are present, which insignificantly alter the affinity of TBP to the corresponding human promoter.

рис. 1, б), достоверно снижает экспрессию этого гена, согласно нашему прогнозу (см. рис. 1, в). Как представлено в первой строке табл. 1, этот дефицит экспрессии гена *CDY2A* вызван снижением сродства «ТБП-промотор» с 4 нМ в норме до 7 нМ для минорного аллеля rs2276109.

Прежде всего мы экспериментально проверили этот прогноз методом задержки электрофоретической подвижности (EMSA). Электрофореграммам для нормального и минорного аллелей rs2276109 соответствуют рис. 2, а и рис. 2, б, тогда как зависимость концентрации комплексов ТБП-ODN от концентраций ODN, которые были построены на основе этих электрофореграмм, демонстрируют рис. 2, в и 2, г.

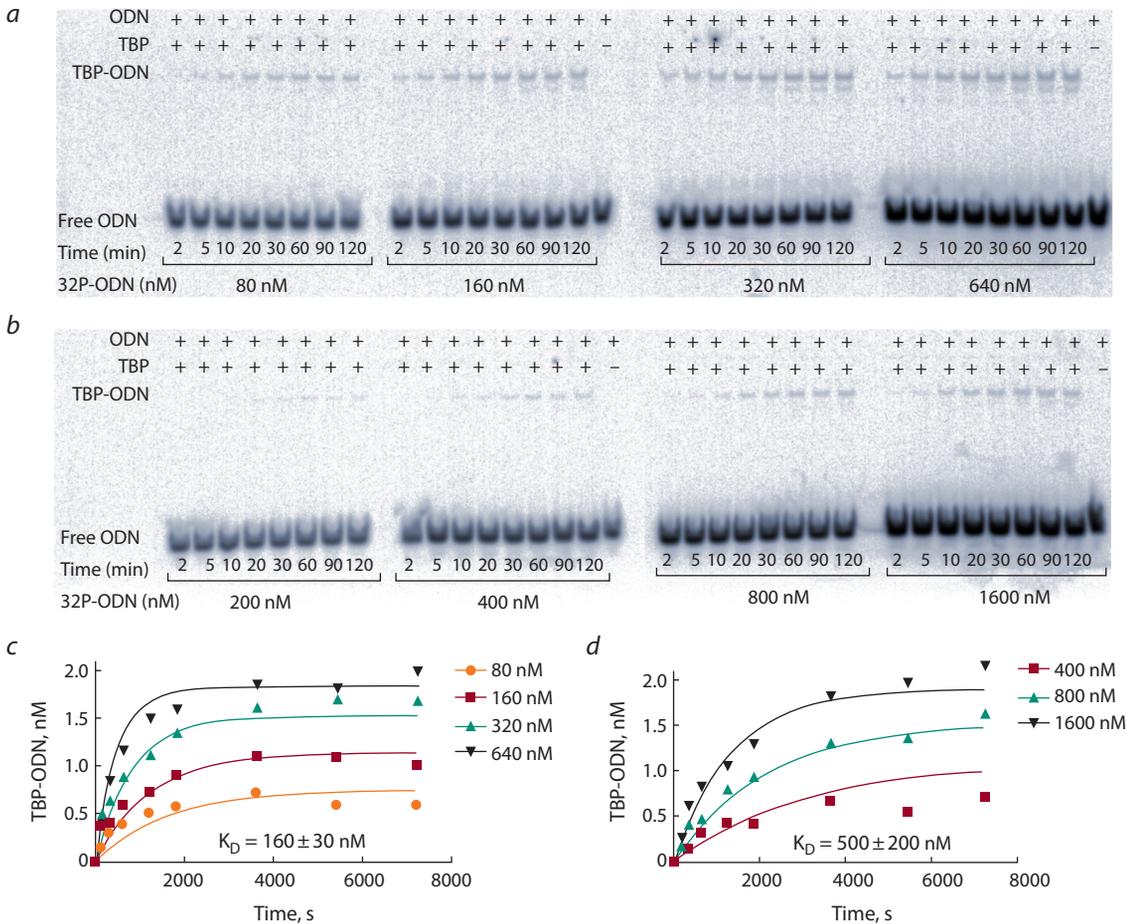
На этих рисунках можно видеть снижение сродства «ТБП-ODN» с 160 нМ в норме до 500 нМ для минорного аллеля rs2276109. Это свидетельствует об адекватности оценок Web-сервиса SNP\_TATA\_Z-tester (Sharypova et al., 2018) для оценки сродства ТБП к промоторам генов на хромосоме Y человека.

Наконец, с помощью поиска по ключевым словам в базе данных PubMed (Lu, 2011) мы нашли клинические данные (Stahl et al., 2012) о дефиците экспрессии гена *CDY2A* при задержке полового созревания. Это позволило нам предложить кандидатный SNP-маркер rs200670724 как генетическую причину снижения МРП вследствие задержки полового созревания мужского организма.

**Ген *ZFY*** (Y-связанный белок «цинковый палец») является паралогом гена *ZFX* на хромосоме X человека. В его промоторе мы нашли два неаннотированных SNP, rs1388535808 и rs996955491, способных вызвать избыток *ZFY*, как представлено на рис. 1, г.

S.Z. Jan с коллегами (2018) сообщили о суперэкспрессии *ZFY* в сперматоцитах как клиническом маркере остановки их мейоза, азооспермии и бесплодия у мужчин. В границах применимости этих клинических наблюдений без гипотез о причинно-следственных связях мы предсказали rs1388535808 и rs996955491 как кандидатные SNP-маркеры, повышающие риск остановки сперматогенеза на стадии мейоза с последующей азооспермией и бесплодием, что безусловно, значительно снижает мужской репродуктивный потенциал (см. табл. 1).

**Ген *SHOX*** (гомеобокс низкорослости) локализован в псевдо-аутосомном районе 1 (PAR1) хромосомы Y человека. В проксимальных промоторах длиной 70 п. о. перед его стартами транскрипции мы нашли пять кандидатных SNP-маркеров для дефицита этого транскрипционного фактора (см. рис. 1, е: rs1452787381) и три кандидатных SNP-маркера для его избытка (см. табл. 1: rs28378830). Согласно табл. 1, они снижают МРП посредством непропорционально низкого роста с деформациями Маделунга (Ramachandrapa et al., 2018) либо нарушений эмбриогенеза (Brosens et al., 2014) соответственно.



**Fig. 2.** Measuring the kinetics of TBP binding to two TATA-containing ODNs identical to the human *CDY2A* gene promoter.

Legend: (a and b) Electropherograms of the wild-typed ancestral and minor alleles of the unannotated SNP rs200670724 under this study, respectively; the concentration of TBP was 2 nM in all the experiments; the concentrations of an ODN containing a tested SNP allele that we used are indicated; (c и d) dependences of reaction rates on ODN concentrations in the cases of the ancestral and minor alleles of the SNP rs200670724, respectively; the  $K_D$  value of the equilibrium dissociation constant was inferred from the dependences of reaction rates on ODN concentrations according to the Web-service GraphPad Prism 5 (<http://graphpad-prism.software.informer.com/5.01>).

Экспериментальные данные *in vitro* для выборочной проверки трех кандидатных SNP-маркеров снижения МРП, rs20067072 (*CDY2A*), rs1452787381 (*ZFY*) и rs1452787381 (*SHOX*), представлены в табл. 2. С использованием метода EMSA были измерены, прежде всего, равновесная константа диссоциации ( $K_D$ ), а также время полураспада ( $t_{1/2}$ ), константы скорости сборки ( $k_a$ ) и распада ( $k_d$ ) комплексов ТБП-ОДН, как это было показано на иллюстративном примере rs200670724 (см. рис. 2).

Достоверные робастные корреляции – линейная Пирсона ( $r$ ), обобщенная Гудмана–Крускала ( $\gamma$ ), ранговые Спирмена ( $R$ ) и Кендалла ( $\tau$ ), – между нашими предсказанными (см. табл. 1) и экспериментальными (см. табл. 2) значениями равновесной константы диссоциации ( $K_D$ ), выраженными в натуральных логарифмических единицах, показаны на рис. 3. Они доказывают адекватность наших прогнозов для величин сродства ТБП к промоторам генов Y-хромосомы человека. В порядке обсуждения можно коротко отметить, что несовпадение между шкалами вертикальной (эксперимент) и горизонтальной (прогноз) осей на рис. 3 соответствует различию в концентрациях ТБП

(т. е. в неконтролируемой доле димеров ТБП, которые имеют  $K_D = 4 \pm 1.5$  нМ (Coleman et al., 1995) и не связывают ДНК) в этой работе (2 нМ) и ранее (Ponomarenko P. et al., 2008) при оптимизации трехшаговой модели связывания «ТБП-промотор» (0.3 нМ), использованной Web-сервисом SNP\_TATA\_Z-tester (Sharypova et al., 2018) для прогнозов в этой работе.

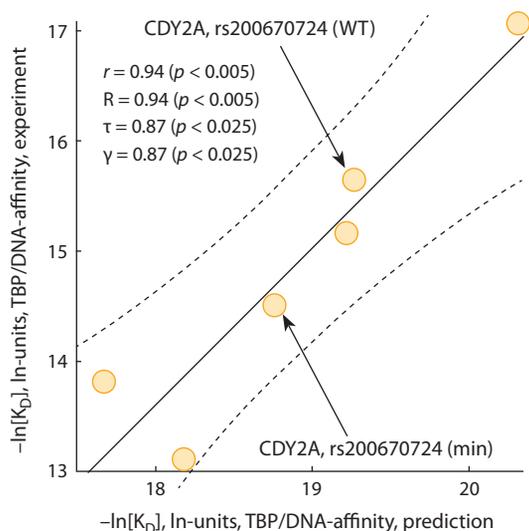
Самая близкая к канонической последовательность ТАТА-бокса наблюдается в промоторе гена *SHOX* с  $K_D = 39 \pm 7$  нМ (см. табл. 2). Минорный вариант SNP rs1452787381 (замена «А» на «G» в ТАТА-боксе) приводит к снижению сродства почти в 7 раз, что обусловлено снижением скорости образования комплексов ( $k_a$ ) в 22 раза. Скорость распада комплексов ( $k_d$ ) снижается в 3 раза, а также увеличивается время жизни комплексов ( $t_{1/2}$ ). Все эти характеристики могут привести к повышенному риску формирования непропорционально низкого роста с деформациями Маделунга.

Наихудшее сродство «ТБП-ОНД»,  $K_D = 1000 \pm 400$  нМ, было определено для комплексов ТБП с промотором гена *ZFY*, имеющего последовательность (СССААСТАС),

**Table 2.** EMSA-based *in vitro* analysis of a complex of TBP and synthetic 26 bp ODNs identical to natural promoters near the SNPs being tested

Gene dbSNP ID (Sherry et al., 2001)	Allele	ODN, sequence, 26 bp	Prediction	Experiment			
			$-\ln(K_D)$	$K_D$ , nM	$t_{1/2}$ , min	$k_a$ , $M^{-1}s^{-1}$	$k_d$ , $s^{-1} \times 10^{-4}$
<i>CDY2A</i> rs200670724	WT	agaatgttccataTaatcgtcatagc	19.27	160 ± 30	28.2 ± 5.4	2500 ± 300	4.1 ± 0.7
	T-43C	agaatgttccataCaatcgtcatagc	18.76	500 ± 200	64.2 ± 18.2	350 ± 50	1.8 ± 0.6
<i>ZFY</i> rs1388535808	WT	ggcggagggggccCaactaccatccc	17.67	1000 ± 400	50.2 ± 15.8	230 ± 40	2.3 ± 0.6
	C-56T	ggcggagggggccTaaactaccatccc	18.18	2000 ± 1000	57.8 ± 17.8	90 ± 50	2.0 ± 0.7
<i>SHOX</i> rs1452787381	WT	gaggtcgccgcgtAtaaatagtgaga	20.31	39 ± 7	19.3 ± 3.1	17000 ± 2000	6.0 ± 1.0
	A-45G	gaggtcgccgcgtGtaaatagtgaga	19.21	260 ± 70	57.8 ± 14.5	770 ± 80	2.0 ± 0.5

Note:  $t_{1/2}$ , half-life;  $K_D$ , dissociation constant;  $k_a$  and  $k_d$ , assembly and dissociation rate constants, respectively.



**Fig. 3.** The significant robust correlations between the predicted *in silico* and experimentally measured *in vitro* values of the TBP–DNA affinity (in natural logarithmic units).

Legend: solid and dashed lines denote the linear regression and boundaries of its 95 % confidence interval, calculated by means of standard software package STATISTICA (Statsoft™, Tulsa, USA); arrows pinpoint the ancestral (WT) and minor (min) alleles of the SNP being studied (rs200670724 of *CDY2A*), an analysis of which is depicted in Figures 1 and 2 as an example of the application of our Web service SNP\_TATA\_Z-tester (Sharyпова et al., 2018) in this work and its *in vitro* selective verification here;  $r$ ,  $\gamma$ ,  $R$ ,  $\tau$ , and  $p$  are coefficients of linear correlation, Goodman–Kruskal generalized correlation, Spearman's and Kendall's rank correlations, and their  $p$  values, respectively.

лишь весьма отдаленно напоминающую консенсус ТАТА бокса (ТАТААААГ). Замена «С» на «Т» не спасает положение – скорость образования комплексов ( $k_a$ ) снижается в 2.5 раза, и незначительно изменяется скорость их распада ( $k_d$ ) и время жизни ( $t_{1/2}$ ). Можно предположить, что предсказанная нами *in silico* суперэкспрессия минорного аллеля SNP rs1388535808 в сперматоцитах, приводящая к остановке сперматогенеза на стадии мейоза, азооспермии и бесплодию, в условиях *in vivo* обеспечивается участием ТВР-ассоциированных факторов (ТАFs) и других регуляторных белков в образовании транскрипционного комплекса в случае ТАТА-несодержащего промотора гена *ZFY* человека.

Полученные данные о снижении сродства ТВР к промотору гена *CDY2A* более чем в 3 раза для аллеля С со-

ответствуют клиническим данным о дефиците продукта этого гена как генетической причине задержки полового созревания.

### Заключение

В настоящей работе исследованы все 35 SNP в промоторах генов *CDY2A*, *SHOX* и *ZFY*, представлявших все три типа генов хромосомы Y человека: уникальный, псевдоаутосомный и паралог гена хромосомы X человека соответственно. Из них отобраны 11 кандидатных SNP-маркеров, способных изменить сродство ТВР к этим промоторам и таким образом способствовать ослаблению МРП. Прежде всего, мы выборочно верифицировали *in vitro* величины сродства «ТВР-промотор», предсказанные в этой работе, и установили статистически достоверную линейную корреляцию между ними и экспериментально измеренными величинами ( $r = 0.94$ ,  $p < 0.005$ ). Используя поиск по ключевым словам в базе данных PubMed, мы нашли клинические данные о физиологических признаках патологий человека, которые соответствовали изменению экспрессии генов, несущих предсказанные кандидатные SNP-маркеры. Среди них были показатели процесса сперматогенеза (*ZFY*: rs1388535808 и rs996955491), задержки полового созревания (*CDY2A*: rs200670724), непропорционально низкого роста с деформациями Маделунга (например, *SHOX*: rs1452787381) и нарушения процесса эмбриогенеза (например, *SHOX*: rs28378830). В совокупности полученные результаты свидетельствуют, что в случае SNP-промоторов генов хромосомы Y человека следует ожидать изменений широкого круга показателей МРП, выходящих далеко за рамки генетического контроля собственно мужской репродуктивной функции.

Это заключение согласуется с выводом авторов реконструкции коэволюции хромосом X и Y у разных видов (опоссум, бык, крыса, мышь, мармозетка, макака-резус, шимпанзе и человек) на основе данных высокопроизводительного секвенирования (Bellott et al., 2014), что, кроме мужской репродуктивной функции, хромосома Y определяет также гендерный диморфизм в предрасположенности к заболеваниям и жизнеспособность мужчин. Предсказанные и экспериментально подтвержденные здесь изменения сродства ТВР к минорным вариантам SNP генов *CDY2A*, *SHOX* и *ZFY* человека могут снизить МРП лишь при аномальных изменениях уровней белков, кодируемых этими генами, в соответствующих тканях на

определенных стадиях онтогенеза. Это с необходимостью требует клинической проверки, на снижение затрат которой благодаря улучшению ее адресности нацелено наше экспериментально-биоинформатическое исследование.

## Список литературы / References

- Amberger J., Bocchini C., Schiettecatte F., Scott A., Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;43(D1):D789-D798. DOI 10.1093/nar/gku1205.
- Arkova O.V., Ponomarenko M.P., Rasskazov D.A., Drachkova I.A., Arshinova T.V., Ponomarenko P.M., Savinkova L.K., Kolchanov N.A. Obesity-related known and candidate SNP markers can significantly change affinity of TATA-binding protein for human gene promoters. *BMC Genomics.* 2015;16(Suppl. 13):S5. DOI 10.1186/1471-2164-16-S13-S5.
- Axelsson J., Bonde J.P., Giwercman Y.L., Rylander L., Giwercman A. Gene-environment interaction and male reproductive function. *Asian J. Androl.* 2010;12(3):298-307. DOI 10.1038/aja.2010.16.
- Bellott D.W., Hughes J.F., Skaletsky H., Brown L.G., Pyntikova T., Cho T.-J., Koutseva N., Zaghlul S., Graves T., Rock S., Kremitzki C., Fulton R.S., Dugan S., Ding Y., Morton D., Khan Z., Lewis L., Buhay C., Wang Q., Watt J., Holder M., Lee S., Nazareth L., Alfoldi J., Rozen S., Muzny D.M., Warren W.C., Gibbs R.A., Wilson R.K., Page D.C. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature.* 2014;508(7497):494-499. DOI 10.1038/nature13206.
- Bowles J.T. The evolution of aging: a new approach to an old problem of biology. *Med. Hypotheses.* 1998;51(3):179-221. DOI 10.1016/s0306-9877(98)90079-2.
- Brosens E., de Jong E., Barakat T., Eussen B., D'Haene B., De Baere E., Verdin H., Poddighe P., Galjaard R., Gribnau J., Brooks A., Tibboel D., de Klein A. Structural and numerical changes of chromosome X in patients with esophageal atresia. *Eur. J. Hum. Genet.* 2014;22:1077-1084. DOI 10.1038/ejhg.2013.295.
- Chadaeva I.V., Ponomarenko M.P., Rasskazov D.A., Sharypova E.B., Kashina E.V., Matveeva M.Yu., Arshinova T.V., Ponomarenko P.M., Arkova O.V., Bondar N.P., Savinkova L.K., Kolchanov N.A. Candidate SNP markers of aggressiveness-related complications and comorbidities of genetic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics.* 2016;17(Suppl. 14):995. DOI 10.1186/s12864-016-3353-3.
- Chadaeva I., Ponomarenko P., Rasskazov D., Sharypova E., Kashina E., Zhechev D., Drachkova I., Arkova O., Savinkova L., Ponomarenko M., Kolchanov N., Osadchuk L., Osadchuk A. Candidate SNP markers of reproductive potential are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics.* 2018;19(Suppl. 3):19. DOI 10.1186/s12864-018-4478-3.
- Chadaeva I., Ponomarenko P., Rasskazov D., Sharypova E., Kashina E., Kleshchev M., Ponomarenko M., Naumenko V., Savinkova L., Kolchanov N., Osadchuk L., Osadchuk A. Natural selection equally supports the human tendencies in subordination and domination: a genome-wide study with *in silico* confirmation and *in vivo* validation in mice. *Front. Genet.* 2019;10:73. DOI 10.3389/fgene.2019.00073.
- Chapman R.N. Animal Ecology with Special Reference to Insects. N. Y.; London: McGraw-Hill Book Co Inc, 1931.
- Coleman R., Taggart A., Benjamin L., Pugh B. Dimerization of the TATA binding protein. *J. Biol. Chem.* 1995;270:13842-13849.
- Cunningham F., Achuthan P., Akanni W., Allen J., Amode M., Armean I., Bennett R., Bhai J., Billis K., Boddu S., Cummins C., Davidson C., Dodiya K., Gall A., Giron C., Gil L., Grego T., Haggerty L., Haskell E., Hourlier T., Izuogu O., Janacek S., Juettmann T., Kay M., Laird M., Lavidas I., Liu Z., Loveland J., Marugan J., Maurel T., McMahon A., Moore B., Morales J., Mudge J., Nuhn M., Ogeh D., Parker A., Parton A., Patricio M., Abdul Salam A., Schmitt B., Schuilenburg H., Sheppard D., Sparrow H., Stapleton E., Szuba M., Taylor K., Threadgold G., Thormann A., Vullo A., Walts B., Winterbottom A., Zadissa A., Chakiachvili M., Frankish A., Hunt S., Kostadima M., Langridge N., Martin F., Muffato M., Perry E., Ruffier M., Staines D., Trevanion S., Aken B., Yates A., Zerbino D., Flicek P. Ensembl 2019. *Nucleic Acids Res.* 2019;47(D1):D745-D751. DOI 10.1093/nar/gky1113.
- Drachkova I.A., Lysova M.V., Repkova M.N., Prokuda O.V., Sokolenko A.A., Arshinova T.V., Kobzev V.F., Iamkovoï V.I., Savinkova L.K. Interaction of proteins from general transcription complex RNA polymerase II with oligoribonucleotides. *Mol. Biol. (Mosk).* 2005;39(1):139-146.
- Haeussler M., Raney B.J., Hinrichs A.S., Clawson H., Zweig A.S., Karolchik D., Casper J., Speir M.L., Haussler D., Kent W.J. Navigating protected genomics data with UCSC Genome Browser in a box. *Bioinformatics.* 2015;31(5):764-766. DOI 10.1093/bioinformatics/btu712.
- Haldane J.B.S. The cost of natural selection. *J. Genet.* 1957;55:511-524. DOI 10.1007/BF02984069.
- Jan S.Z., Jongejan A., Korver C.M., van Daalen S.K.M., van Pelt A.M.M., Repping S., Hamer G. Distinct prophase arrest mechanisms in human male meiosis. *Development.* 2018;145(16):dev160614. DOI 10.1242/dev.160614.
- Kimura M. Evolutionary rate at the molecular level. *Nature.* 1968;217(5129):624-626. DOI 10.1038/217624a0.
- Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford). 2011;2011:baq036. DOI 10.1093/database/baq036.
- Martianov I., Viville S., Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science.* 2002;298(5595):1036-1039. DOI 10.1126/science.1076327.
- Mitsuyasu H., Izuhara K., Mao X., Gao P., Arinobu Y., Enomoto T., Kawai M., Sasaki S., Dake Y., Hamasaki N., Shirakawa T., Hopkin J. Ile50Val variant of IL4R-alpha upregulates IgE synthesis and associates with atopic asthma. *Nat. Genet.* 1998;19:119-120. DOI 10.1038/472.
- Mogno I., Vallania F., Mitra R.D., Cohen B.A. TATA is a modular component of synthetic promoters. *Genome Res.* 2010;20(10):1391-1397. DOI 10.1101/gr.106732.110.
- Peterson M.G., Tanese N., Pugh B.F., Tjian R. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science.* 1990;248(4963):1625-1630.
- Pianka E.R. Natural selection of optimal reproductive tactics. *Am. Zool.* 1976;16(4):775-784. www.jstor.org/stable/3882142.
- Pocai B. The ICD-11 has been adopted by the World Health Assembly. *World Psychiatry.* 2019;18(3):371-372. DOI 10.1002/wps.20689.
- Ponomarenko M., Arkova O., Rasskazov D., Ponomarenko P., Savinkova L., Kolchanov N. Candidate SNP markers of gender-biased autoimmune complications of monogenic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front. Immunol.* 2016;7:130. DOI 10.3389/fimmu.2016.00130.
- Ponomarenko M., Mironova V., Gunbin K., Savinkova L. Hogness Box. In: Maloy S., Hughes K. (Eds.). Brenner's Encyclopedia of Genetics. 2nd edn. San Diego: Academic Press, Elsevier Inc. 2013;3:491-494. DOI 10.1016/B978-0-12-374984-0.00720-8.
- Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP\_TATA\_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed. Res. Int.* 2015;2015:35983004625. DOI 10.1155/2015/359835.
- Ponomarenko M., Rasskazov D., Chadaeva I., Sharypova E., Drachkova I., Ponomarenko P., Oshchepkova E., Savinkova L., Kolchanov N. Candidate SNP markers of atherosclerosis that may significantly change the affinity of the TATA-binding protein for the human gene promoters. *Russ. J. Genet.* 2019;55(9):1137-1151. DOI 10.1134/s1022795419090114.

- Ponomarenko M., Rasskazov D., Chadaeva I., Sharypova E., Ponomarenko P., Arkova O., Kashina E., Ivanisenko N., Zhechev D., Savinkova L., Kolchanov N. SNP\_TATA\_Comparator: genomewide landmarks for preventive personalized medicine. *Front. Biosci. (Schol Ed.)*. 2017;9(2):276-306. DOI 10.2741/S488.
- Ponomarenko P., Chadaeva I., Rasskazov D., Sharypova E., Kashina E.V., Drachkova I., Zhechev D., Ponomarenko M., Savinkova L., Kolchanov N. Candidate SNP markers of familial and sporadic Alzheimer's diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front. Aging Neurosci.* 2017;9:231. DOI 10.3389/fnagi.2017.00231.
- Ponomarenko P., Rasskazov D., Suslov V., Sharypova E., Savinkova L., Podkolodnaya O., Podkolodny N., Tverdokhlebov N., Chadaeva I., Ponomarenko M., Kolchanov N. Candidate SNP markers of chromopathologies are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BioMed Res. Int.* 2016;2016:8642703. DOI 10.1155/2016/8642703.
- Ponomarenko P., Savinkova L., Drachkova I., Lysova M., Arshinova T., Ponomarenko M., Kolchanov N. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl. Biochem. Biophys.* 2008;419:88-92. DOI 10.1134/S1607672908020117.
- Pugh B.F. Purification of the human TATA-binding protein, TBP. *Methods Mol. Biol.* 1995;37:359-367. DOI 10.1385/0-89603-288-4:359.
- Putlyaeva L.V., Demin D.E., Korneev K.V., Kasyanov A.S., Tatosyan K.A., Kulakovskiy I.V., Kuprash D.V., Schwartz A.M. Potential markers of autoimmune diseases, alleles rs115662534(T) and rs548231435(C), disrupt the binding of transcription factors STAT1 and EBF1 to the regulatory elements of human CD40 gene. *Biochemistry (Mosc)*. 2018;83(12):1534-1542. DOI 10.1134/S0006297918120118.
- Ramachandrapa S., Kulkarni A., Gandhi H., Ellis C., Hutt R., Roberts L., Hamid R., Papageorghiou A., Mansour S. SHOX haploinsufficiency presenting with isolated short long bones in the second and third trimester. *Eur. J. Hum. Genet.* 2018;26:350-358. DOI 10.1038/s41431-017-0080-4.
- Rhee H., Pugh B. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*. 2012;483(7389):295-301. DOI 10.1038/nature10799.
- Sharypova E., Drachkova I., Kashina E., Rasskazov D., Ponomarenko P., Ponomarenko M., Kolchanov N., Savinkova L. An experimental study of the effect of rare polymorphisms of human HBB, HBD and F9 promoter TATA boxes on the kinetics of interaction with the TATA-binding protein. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2018;22(1):145-152. DOI 10.18699/VJ18.342. (in Russian)
- Sherry S.T., Ward M.H., Kholodov M., Baker J., Phan L., Smigielski E.M., Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-311. DOI 10.1093/nar/29.1.308.
- Stahl P., Mielnik A., Barbieri C., Schlegel P., Paduch D. Deletion or underexpression of the Y-chromosome genes *CDY2* and *HSFY* is associated with maturation arrest in American men with nonobstructive azoospermia. *Asian J. Androl.* 2012;14(5):676-682. DOI 10.1038/aja.2012.55.
- Stajich J., Block D., Boulez K., Brenner S., Chervitz S., Dagdigan C., Fuellen G., Gilbert J., Korf I., Lapp H., Lehvaslaiho H., Mantsala C., Mungall C., Osborne B., Pocock M., Schattner P., Senger M., Stein L., Stupka E., Wilkinson M., Birney E. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 2002;12(10):1611-1618. DOI 10.1101/gr.361602.
- Telenti A., Pierce L.C., Biggs W.H., di Iulio J., Wong E.H., Fabani M.M., Kirkness E.F., Moustafa A., Shah N., Xie C., Brewerton S.C., Bulsara N., Garner C., Metzker G., Sandoval E., Perkins B.A., Och F.J., Turpaz Y., Venter J.C. Deep sequencing of 10,000 human genomes. *Proc. Natl. Acad. Sci. USA*. 2016;113(42):11901-11906. DOI 10.1073/pnas.1613365113.
- Trovato G.M. Sustainable medical research by effective and comprehensive medical skills: overcoming the frontiers by predictive, preventive and personalized medicine. *EPMA J.* 2014;5(1):14. DOI 10.1186/1878-5085-5-14.
- Turnaev I., Rasskazov D., Arkova O., Ponomarenko M., Ponomarenko P., Savinkova L., Kolchanov N. Hypothetical SNP markers that significantly affect the affinity of the TATA-binding protein to VEGFA, ERBB2, IGF1R, FLT1, KDR, and MET oncogene promoters as chemotherapy targets. *Mol. Biol. (Mosc)*. 2016;50(1):161-173. DOI 10.7868/S0026898416010201.
- Vazari A., Deyneko I.V., Tudor E., Turcan S. Polymorphisms of glutathione S-transferase and methylenetetrahydrofolate reductase genes in Moldavian patients with ulcerative colitis: Genotype-phenotype correlation. *Meta Gene*. 2016;7:76-82. DOI 10.1016/j.mgene.2015.12.002.
- Waardenberg A.J., Basset S.D., Bouveret R., Harvey R.P. CompGO: an R package for comparing and visualizing Gene Ontology enrichment differences between DNA binding experiments. *BMC Bioinformatics*. 2015;16:275.
- Wu J., Wu M., Li L., Liu Z., Zeng W., Jiang R. dbWGFP: a database and web server of human whole-genome single nucleotide variants and their functional predictions. Database (Oxford). 2016;2016:baw024. DOI 10.1093/database/baw024.
- Yoo S., Jin C., Jung D., Choi Y., Choi J., Lee W., Lee S., Lee J., Cha S., Kim C., Seok Y., Lee E., Park J. Putative functional variants of XRCC1 identified by RegulomeDB were not associated with lung cancer risk in a Korean population. *Cancer Genet.* 2015;208(1-2):19-24. DOI 10.1016/j.cancergen.2014.11.004.
- Zerbino D.R., Wilder S.P., Johnson N., Juettemann T., Flicek P.R. The ensembl regulatory build. *Genome Biol.* 2015;16(1):56. DOI 10.1186/s13059-015-0621-5.

#### ORCID ID

M.P. Ponomarenko orcid.org/0000-0003-1663-318X  
I.V. Chadaeva orcid.org/0000-0002-2724-5441  
D.A. Rasskazov orcid.org/0000-0003-4795-0954  
P.M. Ponomarenko orcid.org/0000-0003-2715-9612  
L.V. Osadchuk orcid.org/0000-0002-7597-9204  
A.V. Osadchuk orcid.org/0000-0002-4210-7354

**Acknowledgements.** The concept, supervision of the work, computerized predictions, and their biological interpretation (LVO and AVO) were supported by the Russian Science Foundation, project 19-15-00075. The manuscript was written by MPP and IVCh, support by the Russian Federation Ministry of Science and Education, Program for Improvement of the Competitive Ability of Russian Universities among the Global Research and Educational Centers (project 5-100). The work with the Web service (PMP and DAR) and the *in vitro* experiment (EBSH, IAD, and LKS) were supported by State Budgeted Projects 0324-2019-0040-5-01 and No. 0324-2019-0042-5-01, respectively.

**Conflict of interest.** The authors declare no conflict of interest.

Received December 09, 2019. Revised July 13, 2020. Accepted July 15, 2020.