

# Экспериментальное изучение влияния редких полиморфизмов ТАТА-боксов промоторов генов *HBB*, *HBD* и *F9* человека на кинетику взаимодействия с ТАТА-связывающим белком

Е.Б. Шарыпова<sup>1</sup>, И.А. Драчкова<sup>1</sup>, Е.В. Кашина<sup>1</sup>, Д.А. Рассказов<sup>1</sup>, П.М. Пономаренко<sup>2</sup>, М.П. Пономаренко<sup>1</sup>,  
Н.А. Колчанов<sup>1,3</sup>, Л.К. Савинкова<sup>1</sup> 

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

<sup>2</sup> Университет Ла-Верна, Ла-Верна, Калифорния, США

<sup>3</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Гены человека *HBB*, *HBD* и *F9* относятся к системе кроветворения. Недостаток или избыток продуктов этих генов является причиной наследственных талассемий разной тяжести и гемофилии В Лейдена. Ранее нами было показано, что ряд аннотированных однонуклеотидных полиморфизмов (SNPs) ТАТА-боксов этих генов, ассоциированных с возникновением  $\beta$ - и  $\delta$ -талассемий и гемофилии В Лейдена влияет на взаимодействие с ТАТА-связывающим белком (ТВР), изменение взаимодействия пропорционально изменению количества продуктов этих генов. В настоящей работе исследуется влияние редких неаннотированных SNPs ТАТА-боксов этих генов на взаимодействие с ТАТА-связывающим белком. Для изучения кинетических характеристик образования комплексов ТВР/ТАТА *in vitro* используются двуцепочечные олигодезоксинуклеотиды, идентичные ТАТА-содержащим участкам промоторов генов *HBB*, *HBD* и *F9* (нормальным и минорным аллелям), и рекомбинантный ТВР человека. Показано, что замена  $-25A > C$  (rs281864525) и делеция  $-25AA$  (rs63750953) ТАТА-боксов гена  $\beta$ -глобина одинаково влияют на аффинность ТВР/ТАТА, которая падает в обоих случаях в три раза, но значительно различается влияние этих SNPs на скорость образования комплексов ТВР/ТАТА:  $-25A > C$  приводит к снижению скорости в пять раз, а делеция ( $-25AA$ ) – к снижению скорости более чем в семь раз. Различное влияние замен оказывается и на прочность комплексов ТВР/ТАТА. Если в случае  $-25A > C$  прочность комплексов увеличивается в 1.8 раза, то при делеции  $-25AA$  прочность комплексов возрастает в 2.4 раза, несмотря на уменьшение сродства ТАТА-связывающего белка к ТАТА-боксу. Сравнение изменения экспериментальных значений сродства ( $K_D$ ) ТВР/ТАТА нормальных и минорных аллелей с прогнозируемыми показало, что данные хорошо коррелируют друг с другом: коэффициент линейной корреляции  $r = 0.94$  ( $\alpha < 0.0001$ ). Комплексный подход к изучению редких полиморфизмов может привести к определению наиболее чувствительных маркеров орфанных заболеваний, что внесет вклад в разработку надежных и быстрых методов их диагностирования и лечения.

Ключевые слова: однонуклеотидный полиморфизм; гены кроветворения; редкие полиморфизмы; орфанные заболевания; промотор; ТАТА-бокс.

## 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

Е.Б. Шарыпова<sup>1</sup>, И.А. Драчкова<sup>1</sup>, Е.В. Кашина<sup>1</sup>,  
Д.А. Рассказов<sup>1</sup>, П.М. Пономаренко<sup>2</sup>, М.П. Пономаренко<sup>1</sup>,  
Н.А. Колчанов<sup>1,3</sup>, Л.К. Савинкова<sup>1</sup> 

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

<sup>2</sup> University of La Verne, La Verne, California, USA

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

Human genes *HBB*, *HBD* and *F9* belong to the hematopoiesis system. The deficiency or excess of these genes' products is the cause of hereditary thalassemias of various severity and haemophilia B Leyden. Previously, it was shown that a number of annotated single-nucleotide polymorphisms of TATA boxes of these genes associated with the occurrence of  $\beta$ - and  $\delta$ -thalassemia affect the interaction with the TATA-binding protein, the interaction changing proportionally with the change in the number of gene products. In the present work, we investigate the effect of rare not annotated single-nucleotide polymorphisms (SNPs) of TATA boxes of these genes with an unknown manifestation on the TATA-binding protein interaction. To study the kinetic characteristics of TBP/TATA complex formation *in vitro*, double-stranded oligodeoxynucleotides identical to the TATA-containing portions of the promoters of the *HBB*, *HBD* and *F9* genes ("normal" and minor alleles) and recombinant human TBP were used. It was shown that the TATA-box SNP of  $-25A > C$  (rs281864525) and the deletion of the  $-25AA$  (rs63750953) TATA-box of the  $\beta$ -globin gene have the same effect on the TBP/TATA affinity, which decreases 3-folds in both cases. However, the effect of these substitutions on the rate of the TBP/TATA complex formation is significantly different: SNP  $-25A > C$  decreases the rate 5-fold, and the deletion decreases the rate more than 7-fold. The influence of substitutions on the strength of the TBP/TATA complexes has a different effect. If in the case of SNP  $-25A > C$  the strength of the complexes increases 1.8-fold, then in the case of the  $-25AA$  deletion, the strength of the complexes increases 2.4-fold, even though the affinity of the TATA-binding protein to the TATA box decreases. A comparison of experimental values of affinity ( $K_D$ ) of the TBP/TATA

Received 31.07.2017

Accepted for publication 03.11.2017

© AUTHORS, 2018

E.B. Sharypova and I.A. Drachkova equal contribution to manuscript preparation.

 e-mail: savinkl@mail.ru

complexes of "normal" and minor alleles with the predicted has shown that data correlate well with each other. The coefficient of linear correlation  $r = 0.94$  ( $\alpha < 0.0001$ ). A comprehensive approach to the study of rare polymorphisms may lead to the identification of the most sensitive markers of orphan diseases, which will contribute to the development of reliable and rapid methods for their diagnosis and treatment.

Key words: single nucleotide polymorphism; hematopoiesis genes; rare polymorphisms; orphan diseases; promoter; TATA-box.

#### КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Шарыпова Е.Б., Драчкова И.А., Кашина Е.В., Рассказов Д.А., Пономаренко П.М., Пономаренко М.П., Колчанов Н.А., Савинкова Л.К. Экспериментальное изучение влияния редких полиморфизмов TATA-боксов промоторов генов *HBB*, *HBD* и *F9* человека на кинетику взаимодействия с TATA-связывающим белком. Вавиловский журнал генетики и селекции. 2018;22(1):145-152. DOI 10.18699/VJ18.342

#### HOW TO CITE THIS ARTICLE:

Sharypova E.B., Drachkova I.A., Kashina E.V., Rasskazov D.A., Ponomarenko P.M., Ponomarenko M.P., Kolchanov N.A., Savinkova L.K. 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)

Новая база данных однонуклеотидных вариантов всего генома человека (dbWGFP) содержит более 8 млрд замен, среди которых только 15 % расположены в кодирующих районах, большинство же представляют SNPs регуляторных районов (Wu et al., 2016). Около 29 % всех выявленных SNPs встречаются с частотой менее 1 % и относятся к редким (MacArthur et al., 2014). В группу редких (орфанных) заболеваний включают талассемии, мукополисахаридозы (нарушение обмена гликозаминогликанов), гемофилию В Лейдена, амиотрофический латеральный склероз, врожденные мышечные дистрофии, врожденный гиперинсулинизм, дефицит факторов свертывания крови и др. Достоверная информация о встречающихся с низкой частотой SNPs, делающих вклад в патогенез орфанных заболеваний, необходима для того, чтобы своевременно начать лечение, предотвратить развитие тяжелых осложнений и реакцию на используемые лекарства и отделить SNPs, делающие вклад в патогенез, от ложноположительных. Например, в работе (Bell et al., 2011) приводится результат анализа 406 опубликованных SNPs, ассоциированных с тяжелыми заболеваниями, которые обнаружили в 104 секвенированных геномах человека. Оказалось, что либо 27 % обнаруженных SNPs (у 122 человек) относятся к общим полиморфизмам, либо вообще не хватает доказательств их патогенности.

Парадокс орфанных заболеваний заключается в том, что несмотря на то, что каждое из них встречается редко, в совокупности все «редкие» больные составляют огромную группу. Например, в Европе болезни считаются редкими, если они затрагивают менее пяти человек из 10000; в Бразилии это определение аналогично определению Всемирной организации здравоохранения, поскольку оно затрагивает менее 65 из 100000 человек, а на Тайване редкое заболевание определяется как заболевание, встречающееся менее чем у 1 из 10000 лиц (Dharssi et al., 2017). Особенно важно то, что 75 % редких заболеваний наблюдается у детей, и 30 % из них умирают в течение первых пяти лет жизни (MacArthur et al., 2014). Редкие варианты часто характеризуют этнические группы, входят в состав разных гаплотипов.

Несмотря на впечатляющие результаты GWAS, пока описаны функции небольшого количества SNPs (Cavalli et al., 2016), причастных к регуляции генов. Общая характеристика заключается в том, что функционально значимые SNPs локализованы в сайтах связывания транскрипционных факторов и аллели отличаются по их способности связывать транскрипционные факторы и регулировать гены (Keinan, Clark, 2012). Редкие варианты не обнаруживаются в основном при полногеномных исследованиях (GWAS), так как они отфильтровываются на промежуточных этапах контроля для снижения количества ложноположительных SNPs.

Развитие современных методов и подходов секвенирования геномов приближает время, когда результаты GWAS будут основаны на полном секвенировании индивидуальных геномов и позволят оценивать вклад редких регуляторных вариантов в общее заболевание. Но для доказательства их функционального влияния могут потребоваться сложные функциональные тесты и компьютерное моделирование (Kheradpour et al., 2013). Точность прогнозов, сделанных на основании компьютерного моделирования, очень мала (Dooms, 2017). Актуальность теста во многом зависит от базовой биологической структуры при оценке связи между группой редких генетических вариантов и болезнью. Каждый статистический тест для редких вариантов основан на относительно сложных предположениях, включающих математическое преобразование механизмов болезни. А в настоящее время еще мало известно об основных биологических механизмах, которые могут сильно различаться между заболеваниями (Persyn et al., 2017).

Таким образом, исследование редко встречающихся вариаций последовательностей регуляторной ДНК конкретных генов является актуальной задачей, поскольку понимание влияния генетической вариации на регуляцию экспрессии генов остается главной проблемой в расшифровке транскрипционного регуляторного кода.

Ранее мы изучали влияние аннотированных SNPs (связь которых с определенными заболеваниями показана молекулярно-биологически и клинически) TATA-боксов генов

*HBB*, *HBD*, *F9*, *MBL2*, *F3* и других, на кинетические и термодинамические характеристики образования комплексов ТВР/TATA *in vitro*. Из литературных данных известно, что эти SNPs вызывают моногенные заболевания человека, такие как талассемии различной тяжести, гемофилию В Лейдена, иммунодефицит, амиотрофический латеральный склероз и др. (Savinkova et al., 2013; Drachkova et al., 2014). С помощью уравнения трехшагового связывания (Пономаренко и др., 2008) для предсказания аффинности взаимодействия ТВР с TATA-боксом, построенного на основании наших экспериментальных данных, нами проанализированы *in silico* последовательности TATA-боксов здоровых индивидуумов и SNP-содержащие TATA-боксы, ассоциированные с повышенным риском возникновения наследственных заболеваний человека. Сделанные прогнозы изменения аффинности ТВР/TATA мы экспериментально верифицировали с использованием метода задержки ДНК в геле в максимально стандартизованных условиях и получили довольно высокий коэффициент линейной корреляции для соотношения прогноз/эксперимент:  $r = 0.822$  при  $\alpha < 10^{-7}$  (Savinkova et al., 2013). Мы провели экспериментальное исследование кинетических параметров взаимодействия ТВР с TATA-боксами и впервые показали, что при взаимодействии ТВР с SNP-содержащими TATA-боксами скорости образования комплексов ТВР/TATA уменьшаются в 8–36 раз по сравнению со здоровыми индивидуумами (Drachkova et al., 2014), скорости диссоциации комплексов ТВР/TATA уменьшаются в меньшей степени: в 2–6 раз. Полученные нами результаты обнаружили также хорошее совпадение между влиянием SNPs на взаимодействие ТВР/TATA в условиях *in vitro* с литературными данными по изменению экспрессии соответствующих генов у пациентов *in vivo* и по тяжести заболеваний, ассоциированных с данными SNPs (Драчкова и др., 2010).

Задача исследования, частью которого является настоящая работа, заключается в изучении влияния редких неаннотированных SNPs TATA-боксов генов человека, участвующих в кроветворении, на скорости образования и распада комплексов ТВР/TATA *in vitro* и определении влияния на экспрессию репортерного гена *ex vivo* с целью выявления среди них потенциально функционально значимых SNPs. Нами формируется библиотека для определения неаннотированных редких SNPs, которые будут проанализированы с использованием обновления SNP\_TATA\_Z-tester на основе разработанного нами Web-сервиса SNP\_TATA\_Comparator (Ponomarenko et al., 2015), отобраны потенциально функционально значимые SNPs для экспериментальной верификации *in vitro* и *ex vivo*. В настоящей работе представлены результаты экспериментального определения *in vitro* влияния неаннотированных редких SNPs TATA-боксов, отобранных *in silico*, на кинетические характеристики образования комплексов ТВР/TATA. Следующим этапом будет определение действия этих верифицированных редких SNPs TATA-боксов на экспрессию в культуре клеток репортерного гена *Luc*. На основании всех полученных результатов будут сделаны выводы об их возможном потенциале функциональной значимости и предположение о фенотипическом проявлении.

## Материалы и методы

**Получение рекомбинантного ТВР.** В работе использовали рекомбинантный ТВР человека (ТВР), экспрессированный в клетках *E. coli* BL21(DE3) с плазмиды pAR3038-hТВР (любезно предоставленной проф. В. Puhg, The Pennsylvania State University, University Park, Pennsylvania, США), выделение которого описано ранее (Драчкова и др., 2010).

**Получение меченых  $^{32}\text{P}$  олигонуклеотидов.** Использовали олигонуклеотибы (ODN, длиной 26 п. о., синтезированные и дополнительно очищенные электрофорезом в ПААГ (Biosset, г. Новосибирск). Для получения меченых двуцепочечных ODN обе цепи метили  $^{32}\text{P}$ -АТФ с помощью Т4-полинуклеотидкиназы (SibEnzyme, Novosibirsk), отжигали при 95 °С (в эквимольном соотношении) и медленно (не менее 3 ч) охлаждали до комнатной температуры. Получение меченых и немеченых двуцепочечных ODN подробно описано в работе (Драчкова и др., 2010). В данной работе использовали ODN, идентичные WT и SNP-содержащим TATA-боксам промоторов генов человека, последовательности которых приведены ниже.

### Ген *HBB*:

WT – 5'-caggctgggCATAAAgtagggca-3';  
SNP – –25A/C – 5'-caggctgggCATAAAcgtcaggca-3';  
Del –25 AA – 5'-caggctgggCATAAgtcaggcaga-3'.

### Ген *HBD*:

WT – 5'-acaggaccagCATAAAaggcaggca-3';  
SNP – –30T/C – 5'-acaggaccagCACAAAaggcaggca-3'.

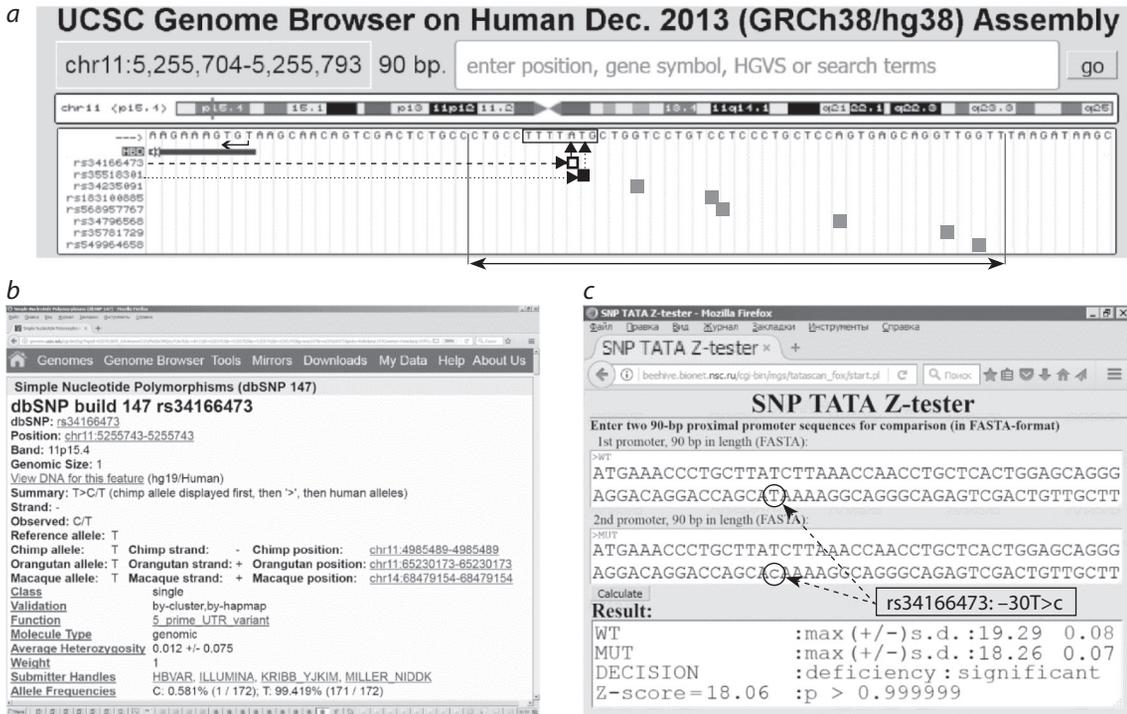
### Ген *F9*:

WT – 5'-ttggTACAACtaatgacctacca-3';  
SNP – –41C/A – 5'-ttggTACAAAtaatgacctacca-3'.

Для выбора редких SNPs используется обновление SNP\_TATA\_Z-tester (рис. 1) на основе ранее разработанного нами Web-сервиса SNP\_TATA\_Comparator (Ponomarenko et al., 2015). Обновление касается только интерфейса ввода данных и не затрагивает предсказания величин сродства ТВР/TATA.

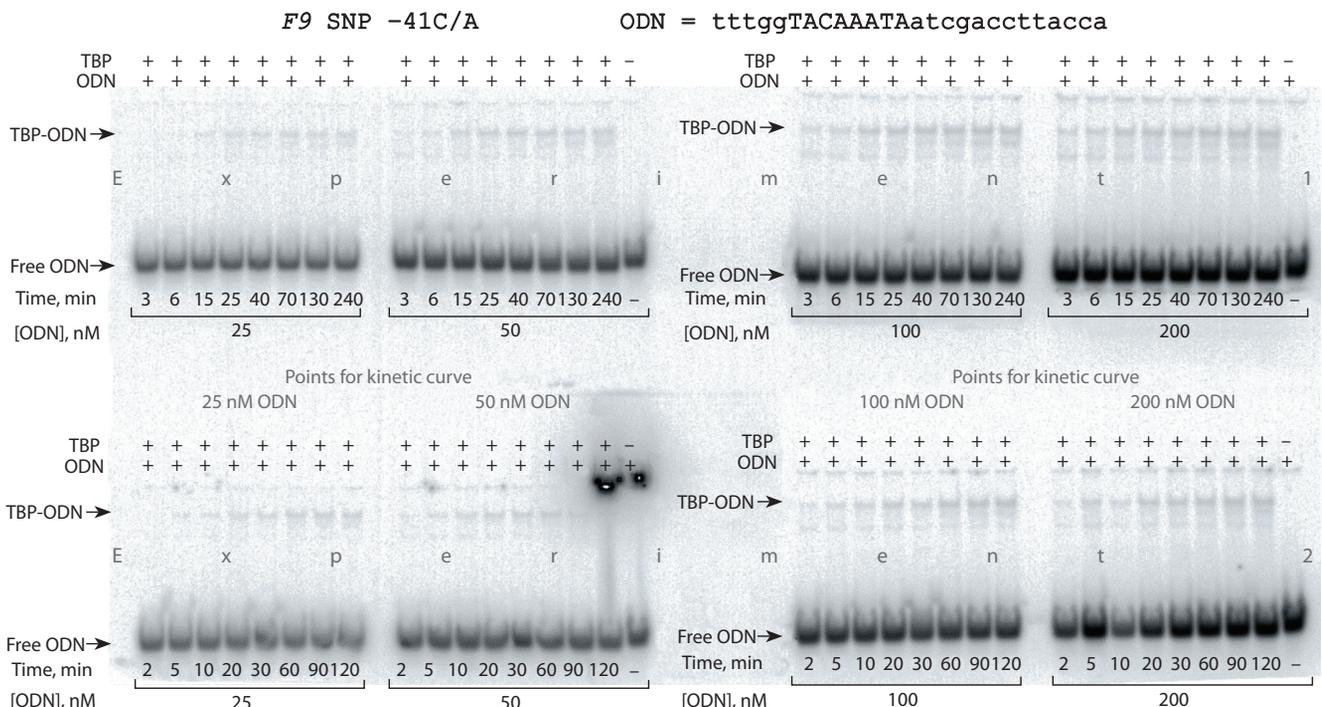
**Экспериментальное определение кинетических характеристик образования комплексов ТВР с олигонуклеотидами.** Эксперименты по изучению кинетических характеристик связывания ТВР с ODN выполняли при постоянной концентрации ТВР и изменении концентрации ODN во времени до установления равновесия при 25 °С (рис. 2 и 3) в буфере связывания следующего состава: 20 мМ Hepes-KOH (pH = 7.6), 5 мМ MgCl<sub>2</sub>, 50 мМ KCl, 1 мМ DTT, 100 мкг/мкл BSA, 0.01 % NP-40, 5 % глицерин. Для отделения комплексов ТВР/ODN от несвязавшегося ODN использовали метод задержки ДНК в геле (EMSA). Электрофорез проводили в 5 % ПААГ на Трис-глициновом буфере (pH 8.3) при температуре 10 °С и напряженности поля 25 В/см в течение 40 мин (Drachkova et al., 2014). Все эксперименты выполняли в максимально стандартизованных условиях.

**Статистический анализ данных.** Сравнение прогнозируемых и экспериментальных значений аффинности комплексов ТВР/TATA нормальных и минорных аллелей проводили с использованием пакета программ Statistica (Statsoft™, США).

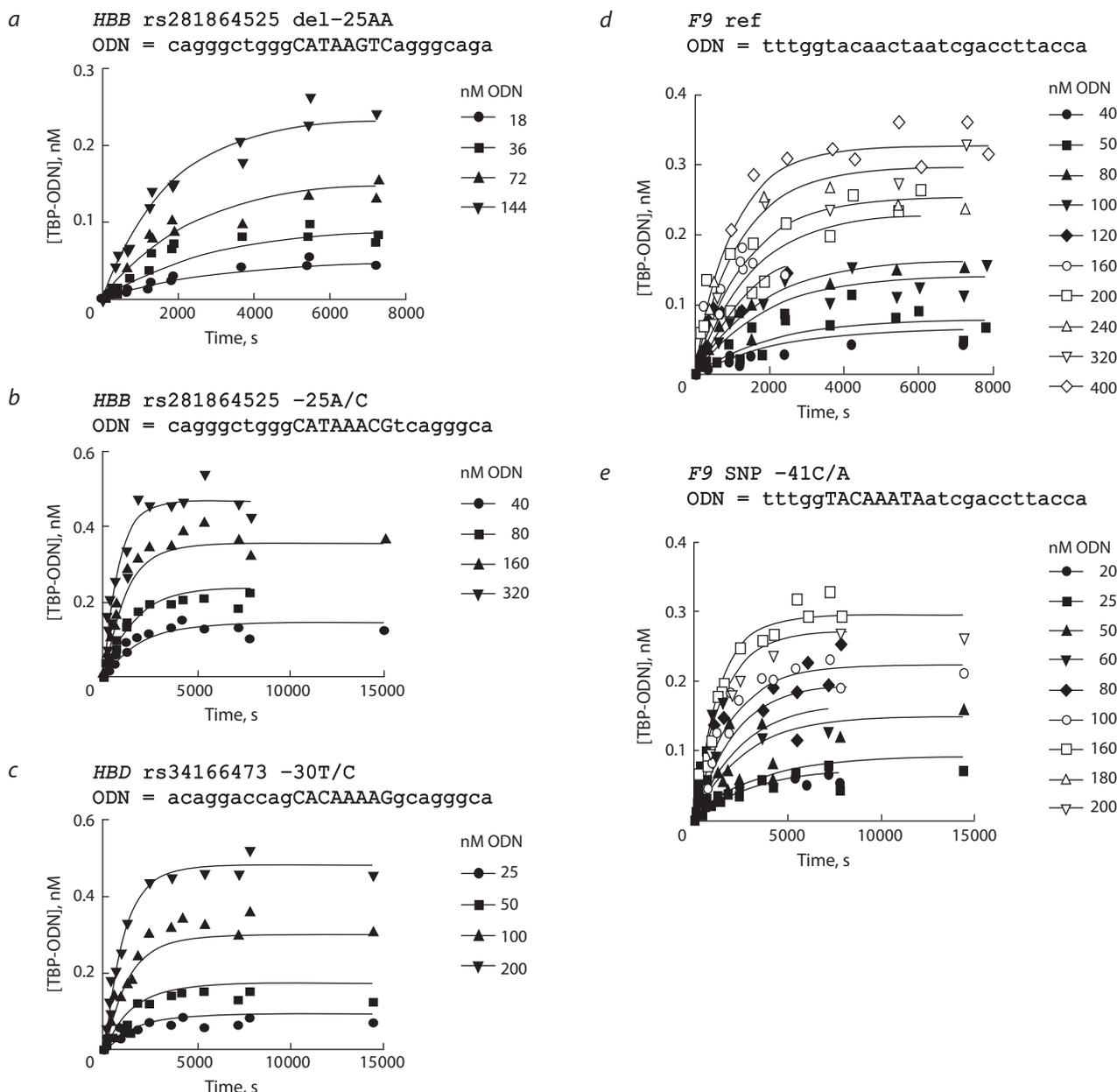


**Fig. 1.** Prediction of the candidate SNP marker rs34166473 (■, dashed arrow) in the [-70; -20] (↔) region relative to the transcription start (designated as →) of the human HBD gene for the hemoglobin δ chain with the SNP\_TATA\_Z-tester service pack of our earlier Web service SNP\_TATA\_Comparator (Ponomarenko et al., 2015).

(a) Visualization of the promoter analyzed with the UCSC Genome Browser Web service (Haeussler et al., 2015). ■, the known clinical SNP marker rs35518301 for δ thalassemia and malaria resistance associated with deficiency of the hemoglobin δ chain (Frischknecht, Dutly, 2005). The marker is located next to rs34166473. (b) Description of rs34166473 in dbSNP (Sherry et al., 2001). (c) SNP\_TATA\_Z-tester. SNP: ■, unannotated SNPs estimated to be insignificant for the TBP affinity to the human HBD promoter (data not shown). Line DECISION of the Result window: the rs34166473 marker at issue can also cause δ chain deficiency; therefore, it may be a candidate SNP marker similarly to the known rs35518301.



**Fig. 2.** Electrophoretic images obtained in the construction of kinetic curves of TBP binding to the ODN identical to the TATA box of the F9 promoter (minor allele). TBP-ODN complexes were separated from unbound ODN by the gel retardation method (EMSA).



**Fig. 3.** Kinetic isotherms of TBP binding to ODNs identical to the TATA boxes of *HBB*, *HBD*, and *F9* promoters. The binding isotherms and  $k_a$  and  $k_d$  values were derived from electrophoretic images (Fig. 2) with GraphPad Prism 5 software.

## Результаты и обсуждение

В таблице представлены результаты экспериментов по определению констант ассоциации и диссоциации, характеризующих скорости образования и распада комплексов ТВР с ТАТА-боксами промоторов генов *HBB*, *HBD* и *F9*, содержащими редкие неаннотированные SNPs (rs281864525, rs63750953, rs341664473 и rs750827465). Для сравнения в таблицу включены также ранее определенные (Drachkova et al., 2014) значения кинетических характеристик влияния SNPs (rs33931746, rs33931746) на взаимодействие ТВР/ТАТА, для которых показана связь с заболеваниями β- и δ-талассемиями.

**Ген *HBB*.** Как известно, гемоглобин взрослого человека состоит из двух α- и двух β-цепей (α<sub>2</sub>β<sub>2</sub>). Порядок расположения генов совпадает с порядком их экспрессии в

процессе онтогенеза: сначала в кластере β-глобинов расположен ген эмбриональной ε-цепи, за ним – два гена фетальных γ-цепей, и затем гены δ- и β-цепей взрослого организма (Schechter, 2008). Регуляция транскрипции этого семейства генов, ведущая к их экспрессии, зависит от стадии в онтогенезе, осуществляется через проксимальные и дистальные регуляторные элементы ДНК (Levings et al., 2001). Один из наиболее важных дистальных регуляторных элементов – локус контрольной области (LCR) (Bank et al., 2005), расположенный на расстоянии от 6 до 22 т.п.н. левее ε-глобинового гена. У взрослого организма уровень экспрессии *HBB* гораздо выше, чем *HBD*. На долю α<sub>2</sub>β<sub>2</sub> приходится 97 % общего гемоглобина человека, а на α<sub>2</sub>δ<sub>2</sub>, которые содержат δ-цепи, кодируемые геном *HBD*, – менее 3 % (Schechter, 2008). По аминокис-

Kinetic ( $k_{ar}$  and  $k_d$ ) and thermodynamic ( $K_D$ ) parameters of the TBP–TATA interaction determined by the gel retardation assay

Gene	SNP	Prediction		Experiment						
		$-\ln(K_D)$	$\Delta$ , ln-ед.	$-\ln(K_D)$	$K_D$ , nM	$\Delta$ , ln-un	$k_{ar}$ , ( $M^{-1}c^{-1}$ ) $\times 10^3$	$k_d$ , ( $c^{-1}$ ) $\times 10^{-4}$	$t_{1/2}$ , min	$-\Delta G$ , kcal/mol
HBB	WT	19.20	0.0	16.81	50 $\pm$ 0.7	0.0	14 $\pm$ 1	7.1 $\pm$ 0.7	16 $\pm$ 2	10 $\pm$ 1
	–28A > G (rs33931746)	18.34	–0.9	14.40	560 $\pm$ 80	–2.4	1.2 $\pm$ 0.1	6.6 $\pm$ 0.9	18 $\pm$ 2	8.5 $\pm$ 0.9
	–28A > C (rs33931746)	18.63	–0.6	14.51	500 $\pm$ 100	–2.3	0.4 $\pm$ 0.1	2.0 $\pm$ 0.3	58 $\pm$ 9	8.6 $\pm$ 1.4
	–25A > C (rs281864525)	18.73	–0.5	15.71	150 $\pm$ 20	–1.1	2.7 $\pm$ 0.3	4.1 $\pm$ 0.5	28 $\pm$ 2	8.5 $\pm$ 0.8
	Del –25AA (rs63750953)	18.61	–0.6	15.71	150 $\pm$ 40	–1.1	1.9 $\pm$ 0.4	3.0 $\pm$ 0.5	38 $\pm$ 3	8.5 $\pm$ 0.8
HBD	WT	19.29	0.0	17.14	36 $\pm$ 4	0.0	23 $\pm$ 3	8.0 $\pm$ 2	14 $\pm$ 4	10 $\pm$ 1
	–31A > G (rs35518301)	18.65	–0.6	16.12	100 $\pm$ 20	–1.0	2.0 $\pm$ 0.1	2.0 $\pm$ 0.4	60 $\pm$ 10	9.5 $\pm$ 0.9
	–30T > C (rs34166473)	18.27	–1.0	15.02	300 $\pm$ 60	–2.12	1.8 $\pm$ 0.2	5.5 $\pm$ 0.5	21 $\pm$ 2	8.1 $\pm$ 0.8
F9	WT	18.86	0.0	15.05	580 $\pm$ 20	0.0	2.4 $\pm$ 0.2	1.4 $\pm$ 0.1	8.3 $\pm$ 1	8.5 $\pm$ 0.6
	–34C > A (rs750827465)	19.32	0.5	16.22	90 $\pm$ 20	1.2	2.7 $\pm$ 0.3	2.4 $\pm$ 0.4	48 $\pm$ 4	8.8 $\pm$ 0.9

The frequency of the minor rs281864525 allele is 0.007%. The frequencies of other alleles  $\leq 0.1\%$  by definition.  $k \pm$ , standard deviation;  $K_D = k_d/k_a$ ;  $\Delta$ , difference between TBP affinities to the ODNs containing and not containing the SNP expressed in log units;  $\Delta = -\ln[K_{D,TATMut}] - (-\ln[K_{D,TATA}])$ .

лотной последовательности  $\delta$ -цепи высоко гомологичны  $\beta$ -цепям (93 %): отличаются только 10 из 147 аминокислот (Steinberg, Adams, 1991).

Мутации в кодирующей и регуляторной областях гена  $\beta$ -глобина приводят к наследственным нарушениям синтеза гемоглобина –  $\beta$ -талассемиям различной тяжести и серповидно-клеточной анемии (Galanello, Origa, 2010). Ранее роль гемоглобина  $\delta$  считали несущественной, но авторы работы (Moleirinho et al., 2013) в результате комплексного анализа, основанного на эмпирических и гаплотипных исследованиях, пришли к выводу о его важной роли в регуляции кластера этих генов.

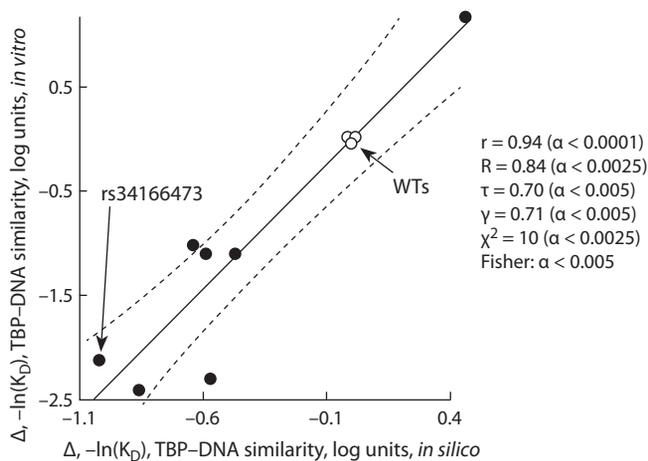
На примере SNP –25A > C и делеции –25AA TATA-бокса промотора гена  $\beta$ -гемоглобина (HBB) хорошо видны различия в механизмах реализации изменений в последовательности ДНК. Аффинность TBP/TATA изменяется в одинаковой мере в обоих случаях: для аллеля «дикого типа»  $K_D = 50$  нМ, для обоих минорных аллелей  $K_D = 150$  нМ, но прочность комплексов отличается для SNP –25A > C в 1.7 раза, а для делеции –25AA – в 2.3 раза (время полураспада:  $t_{1/2} = 28$  мин и  $t_{1/2} = 38$  мин соответственно). Скорость образования комплексов для SNP –25A > C уменьшается более чем в пять раз по сравнению с анцестральным аллелем:  $k_a = (1.4 \pm 0.1)10^4 M^{-1}c^{-1}$  для WT, а для минорного (C) –  $k_a = (2.7 \pm 0.3)10^3 M^{-1}c^{-1}$ . Для SNP –28A > C и –28A > G ранее было показано уменьшение скорости образования комплексов TBP/TATA в 34 и 12 раз соответственно и снижение сродства практически до неспецифического уровня (500 и 560 нМ) (Drachkova et al., 2014), что вызывало у носителей тяжелые формы  $\beta$ -талассемии.

**Ген HBD.** Для минорного аллеля C (–30T > C) гена гемоглобина (HBD) аффинность уменьшается более чем в восемь раз ( $K_D = 36$  нМ для анцестрального аллеля и  $K_D = 300$  нМ для минорного аллеля). Скорость распада комплексов,  $k_d$ , снижается в 1.5 раза для минорного аллеля по сравнению с «диким типом» и, соответственно,

увеличивается в 1.5 раза время полураспада комплексов TBP/TATA (14 мин и 21 мин соответственно), т. е. увеличивается прочность комплексов, несмотря на то, что их аффинность уменьшилась почти до неспецифического уровня. Возможно, это является одним из факторов, компенсирующих *in vivo* низкую аффинность TBP/TATA.

**Ген F9.** Известно, что у пациентов с тяжелым течением гемофилии В Лейдена в детстве наблюдается низкий уровень фактора свертывания крови (F9) (Picketts et al., 1994), который после полового созревания увеличивается и приходит в норму. Установлено, что тяжелое течение заболевания вызывается группой однонуклеотидных замен, расположенных вокруг старта транскрипции гена (Kurachi, Davie, 1982). Как видно из таблицы, для минорного аллеля A (–34C > A) гена фактора коагуляции (F9) скорость образования комплексов увеличилась более чем в два раза по сравнению с аллелем «дикого типа» ( $k_a = (2.7 \pm 0.3)10^3 M^{-1}c^{-1}$  и  $k_a = (1.3 \pm 0.2)10^3 M^{-1}c^{-1}$  соответственно), уменьшилась скорость распада комплексов в 1.6 раза и время их жизни. Все эти изменения согласуются с увеличением аффинности TBP/TATA для минорного аллеля ( $K_D = 90$  нМ) в 3.2 раза в результате SNP –34C > A, улучшающего последовательность TATA-бокса по сравнению с анцестральным ( $K_D = 290$  нМ). Об этом же свидетельствует небольшое увеличение свободной энергии Гиббса (см. таблицу), освободившейся в результате улучшения последовательности TATA-бокса. В результате этой замены можно ожидать увеличения количества фактора коагуляции, о чем более определенно можно будет говорить после экспериментов *ex vivo*.

Сравнение экспериментальных значений сродства ( $K_D$ ) комплексов TBP/TATA «нормальных» и минорных аллелей с прогнозируемыми с помощью обновления SNP\_TATA\_Z-tester на основе ранее разработанного Web-сервиса SNP\_TATA\_Comparator (Ponomarenko et al., 2015), показало, что данные хорошо коррелируют друг с другом. Коэффициент линейной корреляции  $r = 0.94$  ( $\alpha < 0.0001$ ).



**Fig. 4.** Significant correlation between deviations (●, Δ) of  $-\ln(KD)$  TBP-DNA affinities from the standard value (○, WT) predicted with the SNP\_TATA\_Z-tester Web service (X-axis) and the same deviations measured by the gel retardation assay (Y-axis).

Dashed curves indicate borders of the 95% confidence interval for the regression line. Fisher,  $r$ ,  $\gamma$ ,  $\chi^2$ ,  $R$ ,  $\tau$ , and  $\alpha$ : exact Fisher's test for  $2 \times 2$  tables; coefficients of the linear, generalized Goodman and Kruskal correlation, chi-square dichotomous test, Spearman-Kendal rank correlations, and their statistical significance as estimated with the Statistica software (Statsoft™, United States); rs34166473, an exemplary SNP marker for  $\delta$  thalassemia predicted in this work.

На рис. 4 представлены результаты сравнения прогнозируемых и экспериментально измеренных *in vitro* значений аффинности TBP/TATA.

Таким образом, мы определили кинетические характеристики взаимодействия TBP с TATA-боксами, содержащими редкие неаннотированные SNPs, которые могут быть потенциально функционально значимыми и ассоциированы с повышенным риском возникновения заболеваний крови. Это предположение основывается на том, что ранее нами было экспериментально показано снижение сродства TBP к ODN, несущим полиморфизмы в TATA-боксе, ассоциированные с аутоиммунными, сердечно-сосудистыми заболеваниями, амиотрофическим латеральным склерозом и другими заболеваниями. Снижение количества мРНК и белка *in vivo* у носителей этих заболеваний экспериментально показано другими авторами (Драчкова и др., 2010). В случае редких неаннотированных полиморфизмов будут осуществлены дополнительные исследования, чтобы подтвердить или опровергнуть выдвинутое предположение.

## Заключение

Результаты работы показывают влияние редких однонуклеотидных полиморфизмов TATA-боксов промоторов генов гемоглобинов  $\beta$  и  $\delta$  и фактора свертывания крови F9 (rs281864525, rs63750953, rs34166473 и rs750827465) на сродство и скорость образования и распада комплексов TBP/TATA. Полученные численные характеристики влияния каждого SNP на образование и диссоциацию комплексов TATA-боксов с TATA-связывающим белком свидетельствуют о потенциальной возможности редких SNPs влиять на различные фенотипические проявления.

Комплексный подход к изучению редких полиморфизмов может привести к определению наиболее чувствительных маркеров орфанных заболеваний, что внесет вклад в разработку надежных и быстрых методов их диагностирования и лечения. Это будет также способствовать лучшему пониманию молекулярных механизмов самой начальной стадии процесса реализации генетической информации – взаимодействию TBP с TATA-боксом и влиянию на него изменений структуры промоторной ДНК.

## Acknowledgments

This work was supported by State Budgeted Project 0324-2018-0019. The development of the service pack of the Web service was supported by State Budgeted Project 0324-2018-0017. Computerized data analysis was supported by the Russian Federation Ministry of Education and Science, Program for improving the competitiveness of leading Russian universities among leading research and educational centers in the world (Project 5-100) and Integration Project 0324-2018-0021.

## Conflict of interest

The authors declare no conflict of interest.

## References

- Bank A., O'Neill D., Lopez R., Pulte D., Ward M., Mantha S., Richardson C. Role of intergenic human  $\gamma$ - $\delta$ -globin sequences in human hemoglobin switching and reactivation of fetal hemoglobin in adult erythroid cells. *Ann. N. Y. Acad. Sci.* 2005;1054:48-54.
- Bell C.J., Dinwiddie D.L., Miller N.A., Hateley S.L. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* 2011;3:65ra4. DOI 10.1126/scitranslmed.3001756.
- Cavalli M., Pan G., Nord H., Wallerman O., Arzt E.W., Berggren O., Elvers I., Eloranta M.-L., Rönblom L., Toh K.L., Wadelius C. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. *Hum. Genet.* 2016;135:485-497.
- Dharssi S., Wong-Rieger D., Harold M., Terry S. Review of 11 national policies for rare diseases in the context of key patient needs. *Orphanet J. Rare Dis.* 2017;12:63.
- Dooms M. From promising molecules to orphan drugs: Early clinical drug development. *Intractable Rare Dis. Res.* 2017;(1):29-34. DOI 10.5582/irdr.2016.01094.
- Drachkova I.A., Arshinova T.V., Ponomarenko P.M., Merkulova T.I., Kolchanov N.A., Savinkova L.K. Effect of TATA box polymorphisms in the human  $\beta$ -globin gene promoter associated with  $\beta$ -thalassemia on the interaction of the TATA-binding protein. *Vestnik VOGiS = Herald Vavilov Society for Geneticists Breeding Scientists.* 2010;14(4):698-705. (in Russian)
- Drachkova I., Savinkova L., Arshinova T., Ponomarenko M., Peltek S., Kolchanov N. The mechanism by which TATA-box polymorphisms associated with human hereditary diseases influence interactions with the TATA-binding protein. *Hum. Mutat.* 2014;35(5):601-608. DOI 10.1002/humu.22535.
- Frischknecht H., Dutly F. Two new delta-globin mutations: Hb A2-Ninive [ $\Delta$ 133(H11)Val-Ala] and a delta(+)-thalassemia mutation [-31 (A→G)] in the TATA box of the delta-globin gene. *Hemoglobin.* 2005;29(2):151-154.
- Galanello R., Origa R. Beta-thalassemia. *Orphanet J. Rare Dis.* 2010; 5:11.
- 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.
- Keinan A., Clark A.G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*. 2012;336:740-743.
- Kheradpour P., Ernst J., Melnikov A., Rogov P., Wang L., Zhang X., Alston J., Mikkelsen T.S., Kellis M. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*. 2013;23(5):800-811. DOI 10.1101/gr.144899.112.
- Kolchanov N.A., Ponomarenko P.M., Savinkova L.K., Drachkova I.A., Lysova M.V., Arshinova T.V., Ponomarenko M.P. 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.
- Kurachi K., Davie E.W. Isolation and characterization of a cDNA coding for human factor IX. *Proc. Natl. Acad. Sci. USA*. 1982;79(21):6461-6464.
- Levings P.P., Bungert J. The human beta-globin locus control region. *Eur. J. Biochem*. 2002;269:1589-1599.
- MacArthur D.G., Manolio T.A., Dimmock D.P., Rehm H.L., Shendure J., Abecasis G.R., Adams D.R., Altman R.B., Antonarakis S.E., Ashley E.A., Barrett J.C., Biesecker L.G., Conrad D.F., Cooper G.M., Cox N.J., Daly M.J., Gerstein M.B., Goldstein D.B., Hirschhorn J.N., Leal S.M., Pennacchio L.A., Stamatoyannopoulos J.A., Sunyaev S.R., Valle D., Voight B.F., Winckler W., Gunter C. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469-476. DOI 10.1038/nature13127.
- Moleirinho A., Seixas S., Lopes A.M., Bento C., Prata M.J., Amorim A. Evolutionary constraints in the  $\beta$ -globin cluster: the signature of purifying selection at the  $\delta$ -globin (*HBD*) locus and its role in developmental gene regulation. *Genome Biol. Evol*. 2013;5(3):559-571. DOI 10.1093/gbe/evt029.
- Persyn E., Karakachoff M., Scouarnec S. Le, Clézio C. Le, Campion D., French Exome Consortium Schott J.-J., Redon R., Bellanger L., Dina C. DoEstRare: A statistical test to identify local enrichments in rare genomic variants associated with disease. *PLoS ONE*. 12(7). DOI 10.1371/journal.pone.0179364.
- Picketts D.J., Mueller C.R., Lillicrap D. Transcriptional control of the factor IX gene: analysis of five cis-acting elements and the deleterious effects of naturally occurring hemophilia B Leyden mutations. *Blood*. 1994;84(9):2992-3000.
- 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; 2015:359835. DOI 10.1155/2015/359835.
- Savinkova L., Drachkova I., Arshinova T., Ponomarenko P., Ponomarenko M., Kolchanov N. 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;8(2):22-23.
- Schechter A.N. Hemoglobin research and the origins of molecular medicine. *Blood*. 2008;112(10):3927-3938. DOI 10.1182/blood-2008-04-078188.
- 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.
- Steinberg M., Adams J.G. 3rd Hemoglobin A2: origin, evolution, and aftermath. *Blood*. 1991;78:2165-2177.
- 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;pii: baw024. DOI 10.1093/database/baw024.