

Драчкова Ирина Альбертовна

**ВЛИЯНИЕ АССОЦИИРОВАННЫХ С НАСЛЕДСТВЕННЫМИ
ЗАБОЛЕВАНИЯМИ ОДНОНУКЛЕОТИДНЫХ ЗАМЕН В ТАТА-БОКСАХ
НА ВЗАИМОДЕЙСТВИЕ С ТАТА-СВЯЗЫВАЮЩИМ БЕЛКОМ**

03.01.03 – молекулярная биология

Автореферат
диссертации на соискание учёной степени
кандидата биологических наук

Кольцово – 2021

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН).

Научный руководитель:

Савинкова Людмила Кузьминична, кандидат биологических наук, ведущий научный сотрудник, заведующий сектора молекулярно-генетических механизмов белок-нуклеиновых взаимодействий ФГБНУ «Федеральный исследовательский центр Институт Цитологии и Генетики СО РАН»

Официальные оппоненты:

Гуляева Людмила Федоровна, доктор биологических наук, профессор, руководитель лаборатории молекулярных механизмов канцерогенеза ФГБНУ «Федеральный исследовательский центра Фундаментальной и трансляционной медицины»

Жарков Дмитрий Олегович, доктор биологических наук, член-корреспондент РАН, заведующий лабораторией белковой инженерии факультета естественных наук Федерального государственного автономного образовательного учреждения высшего образования «Новосибирский национальный исследовательский государственный университет»

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт биологии гена Российской академии наук

Защита состоится «25» июня 2021 года в 11:00 часов на заседании диссертационного совета Д 208.020.02 при Федеральном бюджетном учреждении науки «Государственный научный центр вирусологии и биотехнологии «Вектор» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по адресу: 630559, р.п. Кольцово, Новосибирской области, тел. (383) 336-60-10.

С диссертацией можно ознакомиться в библиотеке и на сайте ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора <http://www.vector.nsc.ru>.

Автореферат разослан «___» _____ 2021 г.

Ученый секретарь
диссертационного совета
кандидат биологических наук

Т.С. Непомнящих

Актуальность исследования

Детекция функционально значимых однонуклеотидных замен (SNPs) – путь к созданию персонализированной медицины будущего, превентивной и предиктивной [Подколотный *et al.*, 2013, Trovato, 2014]. Особенно сложна задача определения функционально значимых замен и отделения их от нейтральных в регуляторных областях генов, определяющих взаимодействие часто не одного, а нескольких транскрипционных факторов. Большую часть из ≈ 500 аннотированных регуляторных SNP ($\sim 10\%$), по нашим данным [Савинкова *и др.*, 2009], составляют замены в ТАТА-боксе – сайте связывания ТАТА-связывающего белка (ТАТА-binding protein, TBP). Общепризнанным критерием ТАТА-содержащих промоторов генов эукариот (далее – критерий Бухера) является контекстно-зависимая сумма элементов позиционно-частотной $f_i; \xi$ матрицы [Bucher, 1990], элементы которой выражены в натуральных логарифмических единицах. Среди ТАТА-содержащих промоторов отдельно можно выделить промоторы с каноническим ТАТА-боксом с консенсусом ТАТАWAWR, и ТАТА-подобные элементы, имеющие один или два mismatches относительно консенсуса ТАТАWAWR [Rhee and Pugh, 2012]. Очевидно, что, как и у дрожжей [Huisinga and Pugh, 2004], у человека регуляция промоторов, содержащих канонические ТАТА-боксы и ТАТА-подобные элементы, должна существенно отличаться, а замены, переводящие последовательность сайта связывания TBP из класса «канонический ТАТА-бокс» в класс «ТАТА-подобный элемент» или «ТАТА-несодержащая последовательность» и наоборот, должны быть функционально значимы.

Как показывают исследования, сайты связывания TBP не ограничиваются каноническим определением ТАТА-содержащих промоторов [Bucher, 1990], и эксперименты по иммунопреципитации хроматина методом ChIP-seq обнаруживают их перед всеми сайтами старта транскрипции в геномах эукариот [Yang *et al.*, 2011; Rhee and Pugh, 2012; Choukrallah *et al.*, 2012]. TBP необходим для инициации транскрипции всех TFIID-зависимых промоторов, поскольку его

важнейшей функцией является изгибание промоторной ДНК [Burley and Roeder, 1996], создающее возможность для TFIIВ зафиксировать изогнутую конформацию субкомплекса TBP-TATA [Lee and Hahn, 1995; Bernués *et al.*, 1996; Nikolov *et al.*, 1995], без чего невозможна посадка РНК-полимеразы II [Deng and Roberts, 2007].

Одним из современных и многообещающих подходов к поиску функционально значимых замен в регуляторных районах является компьютерный анализ контекстно-зависимых количественных характеристик специфической биологической активности сайтов [Пономаренко, 2017]. Разработанная в рамках этого подхода модель и уравнение пошагового связывания TBP с TATA-боксом позволяет предсказать влияние SNP в районе [−70; −20] перед стартом транскрипции на экспрессию гена [Пономаренко *и др.*, 2008]. Аннотированные SNPs в TATA-боксах, выбранные нами для экспериментальной верификации данной модели, являются причиной таких моногенных заболеваний, как альфа- и бэта-талассемии различной степени тяжести, гемофилия В Лейдена, а также делают вклад в развитие мультифакторных заболеваний, таких как ожирение, гипертония, онкологические и сердечно-сосудистые заболевания. Оценка влияния аннотированных SNPs на аффинность взаимодействия TBP/TATA, способов реализации изменений аффинности, приводящих к определенным биохимическим изменениям, представляет научную значимость и практический интерес. Наличие надежно экспериментально подтвержденной биоинформатической модели взаимодействия TBP со своими сайтами связывания позволит анализировать функциональную значимость неаннотированных SNPs

Цели и задачи исследования.

Целью настоящей работы было установление влияния функционально значимых SNP в сайте связывания TBP на взаимодействие с TBP, а также экспериментальная верификация модели и уравнения пошагового связывания TBP/TATA [Пономаренко *и др.*, 2008], разработанной в ФИЦ ИЦиГ СО РАН.

Для этого были поставлены следующие **задачи**:

1. Создание коллекции аннотированных SNPs TATA-боксов, для которых показана возможность участия в формировании наследственной предрасположенности человека к заболеваниям, и ее анализ с использованием уравнения пошагового связывания TBP.
2. Экспериментальное определение аффинности взаимодействия, (K_D), TBP с ODN, идентичными TATA-боксам канонических аллелей референсного генома и минорных аллелей.
3. Статистическая обработка экспериментальных данных и оценка соответствия прогнозов изменения аффинности TBP *in silico*, сделанных с использованием уравнения пошагового связывания TBP, экспериментальным данным *in vitro*.
4. Экспериментальное определение константы скорости ассоциации k_a и константы скорости диссоциации k_d комплексов TBP/TATA с ODN, идентичными TATA-боксам референсных и минорных аллелей.

Научная новизна

Впервые показано, что функционально значимые SNPs в сайтах связывания TBP вызывают изменения аффинности TBP/TATA, связанные с изменением скоростей ассоциации и диссоциации комплексов TBP/TATA, и эти изменения соответствуют биохимическим проявлениям заболеваний.

Впервые показана функциональная разница во взаимодействии с TBP консенсусного TATA-бокса, TATA-подобного элемента и TATA-несодержащей последовательности.

Теоретическая и практическая значимость работы

Дополнены функциональные характеристики сайтов связывания TBP. Показано, что для консенсусных TATA-боксов с последовательностью TATAWAWR характерна k_a порядка $10^5 \text{M}^{-1}\text{s}^{-1}$, для TATA-подобных элементов k_a порядка $10^4 \text{M}^{-1}\text{s}^{-1}$, а для TATA-несодержащих последовательностей k_a порядка $10^3 \text{M}^{-1}\text{s}^{-1}$.

С учетом полученных результатов на основе уравнения пошагового связывания [Пономаренко и др., 2008] был создан web-сервис

SNP_TATA_COMPARATOR [Рассказов *и др.*, 2013] <http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>, позволяющий оценивать сродство промоторной области генов человека (-20-70) к ТВР. Сервис работает с последовательностями из базы данных Gencode и позволяет сравнивать каноническую последовательность референсного генома и индивидуальную последовательность, несущую замену, что может быть использовано для доклинической проверки SNPs ТАТА-боксов на функциональную значимость.

Основные положения, выносимые на защиту

1. Функционально значимые SNPs в сайтах связывания ТВР вызывают изменения аффинности ТВР/ТАТА, выраженной равновесной константой диссоциации K_D .
2. Экспериментально измеренная аффинность ТВР/ТАТА, выраженная – $\text{Ln}(K_D)$, и изменения аффинности (δ) = $-\text{Ln}[K_{D, \text{ТАТА SNP}}] - \{-\text{Ln}[K_{D, \text{ТАТА}}]\}$ достоверно коррелируют с прогнозами, сделанными с использованием уравнения пошагового связывания.
3. Снижение аффинности ТВР/ТАТА происходит преимущественно за счет снижения константы скорости ассоциации. В 70% случаев снижение k_a сопровождается компенсирующим его влияние на аффинность снижением скорости диссоциации комплекса ТВР/ТАТА.
4. Изменения аффинности ТВР/ТАТА соответствуют биохимическим проявлениям SNPs и реализуются через изменения как скоростей ассоциации, так и скоростей диссоциации комплексов ТВР/ТАТА.

Апробация работы

По материалам диссертации опубликовано 5 статей в рецензируемых журналах, индексируемых в международных базах Web of science и/или Scopus. Результаты работы были представлены на 7 научных конференциях: «Физико-химическая биология» BGRS - Новосибирск, 2008; Новосибирск, 2010 и Новосибирск, 2012. Crimean meeting, IIIrd International conference, dedicated to N.W. Timofeeff-Ressovsky "Modern problems of genetics, radiobiology, radioecology

and evolution". Alushta, 9-14 october, 2010. Первая Международная научно-практическая конференция "Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине" Санкт-Петербург, 23-26 ноября 2010 года. IInd International conference Phisico-Chemical Biology dedicated to the 85th anniversary of academician Dmitri G. Knorre, July 25-29, 2011г. Первые Международные Беккеровские чтения» (27-29 мая 2010), Волгоград.

Вклад автора

Трансформация клеток *E. coli* плазмидой, несущей ген TBP, наращивание биомассы, адаптация протокола очистки, выделение и анализ белка, мечение, очистка и анализ олигонуклеотидов, разработка протоколов реакции связывания TBP/ТАТА и количественного EMSA для равновесных и кинетических экспериментов, расчет равновесных констант диссоциации K_D и констант скорости ассоциации и диссоциации k_a и k_d проведены автором самостоятельно.

Разработка протокола кинетического анализа взаимодействия TBP/ТАТА методом поверхностного плазмонного резонанса на биосенсоре ProteOn XPR36 проведена совместно с к.б.н. Шеховцовым С.В. Биоинформационный анализ нуклеотидных последовательностей и статистический анализ данных проводился совместно с д.б.н. М.П. Пономаренко.

Структура и объем диссертации

Работа содержит список сокращений, введение, обзор литературы, материалы и методы, результаты и обсуждение, заключение, выводы и список литературы. Работа изложена на 186 страницах, содержит 20 рисунков и 7 таблиц. Список цитируемой литературы включает 352 ссылки.

Материалы и методы

В работе использовали полноразмерный TBP человека с природной аминокислотной последовательностью и двуцепочечные меченые ^{32}P ODN длиной 26 п.н., идентичные референсным и минорным аллелям ТАТА-боксов генов человека. Равновесные константы диссоциации, K_D , константы скорости

ассоциации (k_a) и константы скорости диссоциации (k_d) комплексов TBP/ODN анализировали методом количественного EMSA. Количественный анализ радиоавтографов, полученных на фосфоимиджере Molecular Imager PharoFX Plus (Bio-Rad) проводили с помощью программы Quantity One – 4.5.0 (Bio-Rad). Расчет констант проводили в условиях нелинейной регрессии в программе GraphPad Prism 5. Предсказание величины $-\ln [K_{D, TATA}](S^0)$ сродства TBP к ТАТА-боксу сделано с использованием уравнения пошагового связывания [Пономаренко и др., 2008].

Для референсной последовательности $S^0 = \{s_1; s_2, \dots, s_{i-1}; s_i, s_{i+1}, \dots, s_{25}, s_{26}\}$ и последовательности $S^\# = \{s_1; s_2, \dots, s_{i-1}; s_i, s_{i+1}, \dots, s_{25}, s_{26}\}$ аллельного варианта ТАТА-бокса с заменой $s_i > t_i$ длиной 26 п. н. определяли величину $-\ln[K_{D, TATA}](S^0)$ сродства TBP к ТАТА по формуле (1):

$$-\ln[K_{D, TATA}](S^0) = 10.9 + 0.15PWM_{TATA}(S^0) - 0.2\ln[K_{D, \text{онДНК}}](S^0) - 0.23\ln[K_{D, \text{днДНК}}](S^0). \quad (1) \quad [\text{Пономаренко и др., 2008}]$$

Изменение сродства TBP к ТАТА вычисляли по формуле (2):

$$\delta\{-\ln[K_{D, TATA}](S^0; s_i > t_i)\} = -\ln[K_{D, TATA}](S^\#) - \{-\ln[K_{D, TATA}](S^0)\}, \quad (2)$$

Статистический анализ проводили с помощью пакета STATISTICA.

Результаты и обсуждение

Создание коллекции аннотированных SNPs в сайтах связывания TBP, для которых показана возможность участия в формировании наследственной предрасположенности человека к заболеваниям. Анализ последовательностей ТАТА-боксов in silico с использованием уравнения пошагового связывания и прогноз влияния SNP на аффинность TBP/ТАТА.

Коллекция собрана путем поиска оригинальных статей в Pubmed по ключевым словам и в открытых базах данных. Она содержит 34 промотора генов человека, в которых описано 42 замены в (-20-70) районе. Из них 5 промоторов принадлежат генам, повреждение которых вызывает моногенные заболевания: α -, β -, δ -талассемии (гены α - и β -глобинового кластера), гемофилию Лейдена (ген

F9), синдром Жильбера (*UGT 1A1*). За исключением последнего заболевания, к которому приводит полиморфизм rs3064744:(TA)₇>(TA)₈(-38TA>(TA)₂) в гене *UGT 1A1* с высокой (32,53%) MAF (Minor Allele Frequency), остальные 10 замен в генах *HBB*, *HBD*, *F9* и псевдогене *HBZP1* α-глобинового кластера являются редкими мутациями. Они были выявлены у пациентов и их родственников при секвенировании заинтересованных генов, известных патогенных SNPs в которых не было найдено при скрининге. Патогенность данных мутаций была подтверждена в экспериментах с репортерными конструкциями, и (или) анализом мРНК в клетках пациентов.

Остальные замены – это SNPs в известных генах-кандидатах мультифакторных заболеваний. Ассоциация 2-х из них – rs 2814778:T>C (-33T>C) в промоторе гена *DARC* с устойчивостью к малярии, и rs1143627:C>T (-31C>T) в промоторе гена *IL1B* с предрасположенностью к раку легких – показана с помощью GWAS. Это чрезвычайно распространенные аллельные варианты – по данным проекта «1000 геномов» их MAF равна 26,637% и 47,244%, соответственно.

Для части SNPs ассоциация с предрасположенностью к заболеваниям показана с помощью статистического анализа групп индивидуумов с заболеванием и без. Так, -17g>c в гене *ABCA1* повышает риск инфаркта миокарда у пациентов с ИБС (ишемической болезнью сердца) в 3.5 раза [Santamarina-Fojo *et al.*, 2000; Zwarts *et al.*, 2002]. Rs7586110:t>g (-57t>g) в гене *UGT1A7* повышает риск рака полости рта у курильщиков в 8-10 раз [Lankisch *et al.*, 2005], а rs10895068:G>A (+331G>A) в гене рецептора прогестерона *PGR* повышает риск рака эндометрия в два раза, особенно у женщин с повышенной массой тела.

Для остальных SNPs предположение о связи с заболеваниями сделано на основе анализа распространенных SNPs в этих генах, имеющих сходное биохимическое проявление, либо известной связи биохимического проявления с заболеванием.

Следует заметить, что для 40 из 42 SNPs коллекции показано биохимическое проявление замены – или в экспериментах с репортерными конструкциями, или анализом мРНК или белка в клетках пациентов.

Для всех замен был выполнен анализ *in silico* с использованием уравнения пошагового связывания и сделан прогноз влияния SNP в TATA-боксе на аффинность TBP/TATA. Результаты см. в *Таблице 1*, ниже.

Наиболее значительное снижение сродства предсказано в случае замен в TATA-боксах генов *HBB* (CATAAAAG на CAAAAAAG с $\delta=-1,46$, и на CATGAAAG с $\delta=-1,24$), *CYP2A6* (TATAAAGG на TAGAAAGG с $\delta=-1,49$), *SOD1* (TATAAAGT на TGTAAGT с $\delta=-1,17$), *MBL2* (TATATAGC на TACATAGC с $\delta=-1,11$) и *TPI* (TATATAAG на TATAGAAG с $\delta=-1,03$). Как можно видеть, все эти замены находятся в несомненных с точки зрения последовательности и позиции относительно старта транскрипции (-36-20, согласно EPD) TATA-боксах (или, по крайней мере, TATA-подобных элементах), при этом все они, кроме замены TACATAGC в гене *MBL* относятся к т.н. «запрещенным», их частота согласно данным Бухера не превышает 1% [Bucher, 1990]. Цветовые обозначения позиций PWM_{TATA} см. в подписи к *Таблице 2*.

Максимальное увеличение сродства TBP к TATA предсказано нами для замен в промоторах генов *IL1B* TATAAAAA ($\delta= +1,18^*$), *F3* TTTATAGT ($\delta= +0,52^*$, и псевдогена *ψγE* TATATAGC ($\delta= +0,84^*$). Во всех трех случаях замена расположена в самом TATA-боксе. Например, замена -31C>T, приводящая к восстановлению канонического TATA-бокса TATAAAAA взамен TATA-подобного элемента CATAAAAA в промоторе гена *IL1β* ($\delta = +1.18$), ассоциирована с повышенным риском развития немелкоклеточного рака легкого [Zienolddiny *et al.*, 2004] и рака печени [Wang *et al.*, 2003], а также с повышенной чувствительностью к *Helicobacter pylori* [Hamajima *et al.*, 2001].

Авторы, описавшие замену TTTATAGC> TTTATAGT (-21C>T) в TATA-подобном элементе гена тканевого фактора (*F3*) [Arnaud *et al.*, 2000], показали, что SNP, которые усиливают экспрессию гена *F3*, повышают риск тромбоза.

и инфаркта миокарда. Поскольку эта замена находится в промоторном элементе, авторы высказали предположение о ее функциональной значимости. Вопрос же о направлении изменений (повышении или снижении экспрессии гена) остался открыт. Это одна из 3-х экспериментально изученных замен, для которых прогноз, сделанный с помощью PWM_{TATA} [Bucher, 1990] и с помощью модели пошагового связывания качественно отличается – см. *Таблицу 2*.

Для замен во фланках ТАТА-бокса модель пошагового связывания предсказывает значительно меньшее изменение аффинности, чем для замен, расположенных в ТАТА-боксе: δ от -0,28* до +0,21*. Выраженное влияние многих таких замен на экспрессию, непропорциональное прогнозу повреждения ТАТА-бокса, вероятно, объясняется наличием во фланках ТАТА-бокса неидентифицированных сайтов связывания других транскрипционных факторов.

Анализ показал, что прогноз влияния SNPs на аффинность TBP/ТАТА качественно соответствует литературным данным о биохимическом проявлении SNPs. В 32 из 42 случаев (74%) эти прогнозы оказались статистически достоверными ($\alpha < 0.05$). Таким образом, для экспериментальной верификации модели пошагового связывания и характеристики влияния замен на взаимодействие TBP/ТАТА мы получили 32 функционально значимых SNPs в сайтах связывания TBP.

Таблица 1 Предсказание изменения сродства ТВР к ТАТА (δ) при наличии SNPs в ТАТА-боксах, для которых показана возможность участия в формировании наследственной предрасположенности человека к заболеваниям.

Ген	Аллель	MAF, % (1000 genome)	Последовательность, 5'-3'	δ	$\pm\delta_{5\%}$	Биохимическое проявление	Заболевание или предрасположенность	Ссылка
Уменьшение сродства ТВР к ТАТА								
<i>HBV</i>	референсная rs33980857:T>A (-30T>A) rs34598529:A>G (-29A>G) rs33980857:T>C (-30T>C) rs33931746:A>G (-28A>G) rs33981098:A>G (-31A>G) rs33931746:A>C (-28A>C) NG 000007.3:g.70518A>T (-27A>T)	редкая 0,0399 Редкая Редкая Редкая Редкая нд	cagggctgggCATAAAAGtcagggca cagggctgggCAAAAAAGtcagggca cagggctgggCATGAAAGtcagggca cagggctgggCAAAAAAGtcagggca cagggctgggCATAGAAGtcagggca cagggctgggCGTAAAAAGtcagggca cagggctgggCATACAAGtcagggca cagggctgggCATAATAGtcagggca	-1,46* -1,24* -0,98* -0,71* -0,60* -0,53* -0,36*	0,08	Недостаток β -глобина	β -талассемия	[Fei <i>et al.</i> , 1988] [Antonarakis <i>et al.</i> , 1984] [Cai <i>et al.</i> , 1989] [Orkin <i>et al.</i> , 1983] [Takahara <i>et al.</i> , 1986] [Poncz <i>et al.</i> , 1982] [Badens <i>et al.</i> , 1999]
<i>HBD</i>	референсная rs35518301:A>G (-31A>G)	редкая	acaggaccagCATAAAAGgcagggca acaggaccagCGTAAAGgcagggca	-0,61*	0,08	Недостаток δ -глобина	δ -талассемия	[Frischknecht and Dutly, 2005] [Frischknecht <i>et al.</i> , 2006]
<i>CYP2A6</i>	референсная rs28399433:T>G (-48T>G)	10,026	tttcaggcagTATAAAGGcaaacac tttcaggcagTAGAAAGGcaaacac	-1,49*	0,10	Недостаток никотинооксидазы	Рак легкого	[Pitarque <i>et al.</i> , 2001]
<i>SOD1</i>	референсная rs7277748:A>G (-27A>G)	3,934	aggtctggccTATAAAGTagtcgcgg aggtctggccTGTAAAGTagtcgcgg	-1,17*	0,09	Недостаток супероксид-дисмутазы	Амиотрофический латеральный склероз	[Niemann <i>et al.</i> , 2007]
<i>MBL2</i>	референсная rs72661131:T>C (-35T>C)	0,140	catctatttcTATATAGCctgcaccc catctatttcTACATAGCctgcaccc	-1,11*	0,11	Снижение опсонизации патогенов	Повторные инфекции	[Boldt <i>et al.</i> , 2006]
<i>TPI</i>	референсная rs1800202:T>G (-24T>G)	0,569	cgcggcgcctcTATATAAGtgggcagtc cgcggcgcctcTATAGAAAGtgggcagtc	-1,03*	0,09	Недостаток TPI	Невроз, анемия	[Watanabe <i>et al.</i> , 1996]
<i>DARC</i>	референсная rs2814778:T>C (-33T>C)	26,637	ccttggcTCTTATCTTggaagcacag ccttggcTCTTACCTTggaagcacag	-0,78*	0,07	Отсутствие рецептора хемокинов, через которые <i>Plasmodium vivax</i> проникает в эритроцит	Устойчивость к малярии	[Tournamille <i>et al.</i> , 1995]
<i>EDH17B2</i>	референсная rs201739205:A>C (-27A>C)	0,852	cgaagcaggtGATATCAAgccagag cgaagcaggtGATATCCAgccagag	-0,60*	0,08	Недостаток (55%) регулятора баланса «эстрон \leftrightarrow эстрадиол»	Рак молочной железы	[Peltoketo <i>et al.</i> , 1994]

Таблица 1 продолжение

Ген	Аллель	MAF, % (1000 genome)	Последовательность, 5'-3'	δ	$\pm\delta_{5\%}$	Биохимическое проявление	Заболевание или предрасположенность	Ссылка
<i>CYP2B6</i>	референсная rs34223104:T>C (-82T>C&	1, 637	gatgaaatTTTATAACAGggtgcaga gatgaaatTT C ATAACAGggtgcaga	-0,59*	0,08	Избыток фермента	Рак печени	[Zukunft <i>et al.</i> , 2005]
<i>F9</i>	референсная NG 007994.1:g.4975 G>C (-48G>C)	нд	acagctcagcTTGTACTTtggtacaa acagctcagcTT C TACTTtggtacaa	-0,49*	0,11	Низкий уровень фактора IX свертывания крови	Гемофилия В Лейдена	[Reijnen <i>et al.</i> , 1992]
<i>F7</i>	референсная -32A>C	нд	ggagggcAGAGAACTTTgcccgtcagt ggagggc C GAGAACTTTgcccgtcagt	-0,15*	0,07	Умеренно сниженный уровень фактора VII свертывания	Умеренное снижение свертываемости	[Kavlie A., 2003]
<i>APOA1</i>	референсная -27A>C	нд	tggtctgcagaCATA A ATAggccctgc tggtctgcagaCATA C ATAggccctgc	-0,41 ^a	0,08	Накопление холестерина	Атеросклероз	[Matsunaga <i>et al.</i> 1999]
<i>ERβ</i>	референсная rs35036378:T>G (-59T>G)	0, 919	cctctcgggtcTTTAAAGgaagaagg cctctcgggtc G TTAAAGgaagaagg	-0,31*	0,07	Недостаток рецептора (50% от нормы)	Гипертония	[Philips <i>et al.</i> , 2005; Philips <i>et al.</i> , 2007]
<i>ABCA1</i>	референсная -17g>c	68, 812 _b	aatcTATAAAAGgaactagtctcggc aatcTATAAAAGgaacta c tctcggc	-0,28*	0,06	Недостаток белка A1	Инфаркт миокарда у лиц с ИБС	[Santamarina-Fojo <i>et al.</i> , 2000]; [Zwarts <i>et al.</i> , 2002]
<i>UGT1A7</i>	референсная rs7586110:t>g (-57t>g)	29, 772	tccacttacTATATTATaggagctta tccac g tacTATATTATaggagctta	-0,17*	0,09	Снижение детоксикации экзоканцерогенов	Рак полости рта	[Lankisch <i>et al.</i> , 2005]
<i>GH1</i>	референсная rs11568827:g>Δ (-31g>Δ)	4, 553	gscaggTATAAAAAgggsscsaaga gscag_ TATAAAAAgggsscsaaga	-0,13*	0,09	Недостаток гормона роста	Низкорослость	[Horan <i>et al.</i> , 2003]
<i>Cx-40</i>	референсная rs35594137:g>a (-44g>a)	18, 830	aagggcagacatacgATTAAAAAgag aa a gagcagacatacgATTAAAAAgg	-0,12*	0,09	Недостаток (35%) коннексина-40	Аритмия предсердия	[Groenewegen <i>et al.</i> , 2003]
<i>GSTM3</i>	rs1332018:c>a (-63c>a) референсная (-63c)	67, 977 32, 023	ttatgt a gggTATAAAGCccctccc ttatgt c gggTATAAAGCccctccc	-0,11*	0,10	Недостаток фермента, 12% от нормы	Повреждение ДНК	[Liu <i>et al.</i> , 2005]
<i>UCP3</i>	референсная rs1800849:t>c (-55t>c)	18, 610	agcc c gtgtgTATAAGACcagtgcca agcc t gtgtgTATAAGACcagtgcca	-0,05	0,11	Снижена скорость метаболизма	Избыточный вес	[Otabe <i>et al.</i> , 2000]
<i>KLK1</i>	референсная rs36217621:c>g (-19c>g)	6, 689	gataagggcTTTAAAAg c ctcccca gataagggcTTTAAAAg g ctcccca	-0,05	0,06	Недостаток калликреина	Заболевания почек	[Slim <i>et al.</i> , 2002]

Таблица 1 продолжение

Таблица 1 продолжение								
Ген	Аллель	MAF, % (1000 genome)	Последовательность, 5'-3'	δ	$\pm\delta_{5\%}$	Биохимическое проявление	Заболевание или предрасположенность	Ссылка
STAR	референсная rs16887226:c>t (-33c>t)	0, 839	cttcagcggggggaCATTTAAGacgca cttcagtggggggaCATTTAAGacgca	-0,05	0,06	Недостаток альдостерона	Гипертония	[Casal <i>et al.</i> , 2006]
TM	референсная rs13306848:g>a (-33g>a)	2, 299	ggccgggcaCTTATAAAActcgagccc ggccaggcaCTTATAAAActcgagccc	-0,04	0,10	Недостаток тромбомодулина	Тромбофлебит	[Le Flem <i>et al.</i> , 1999]
UGT 1A	референсная rs3064744:(TA) ₇ >(TA) ₈ (-38TA>(TA) ₂)	32, 53	gcca (TA) ₇ Agtaggag			Снижение катаболизма эндогенных веществ крови Редкий аллель	Синдром Жильбера	[Kaniwa <i>et al.</i> , 2005]
	rs3064744:(TA) ₇ >(TA) ₉ (-38TA>(TA) ₃)	нд	gcca (TA) ₈ Agtaggag	-0,01	0,06		Синдром Жильбера	[Kaniwa <i>et al.</i> , 2005]
	rs3064744:(TA) ₇ >(TA) ₆ (-38TA> Δ)	нд	gcca (TA) ₉ Agtaggag	-0,01				
		нд	gcca (TA) ₆ Agtaggag	+0,01			Норма	[Kaniwa <i>et al.</i> , 2005]
Увеличение родства ТВР к ТАТА								
RMRP	референсная rs1886600894:c>g (-24c>g)	0, 286	tcttttagggcTATAAAATa _c tactct tcttttagggcTATAAAATa _g tactct	+0,07	0,08	Недостаток эндорибонуклеазы	Гипоплазия хряща	[Bonafè <i>et al.</i> , 2002]
TF	референсная rs8177186:g>t (-34g>t)	14, 137	ccggggaatggAATAAAGGgacgcggg ccgggtaatggAATAAAGGgacgcggg	+0,07	0,07	Нарушение транспорта железа	Анемия у женщин	[Lee <i>et al.</i> , 2001]
AGT	референсная rs5050:a>c (-20a>c)	17, 592	accscctcagcTATAAATAgggca _t tcg accscctcagcTATAAATAgggc _c tcg	+0,07	0,09	Пол-специфическая экспрессия Избыток регулятора тонуса сосудов	Гипертония у женщин Дефекты развития малых сосудов мозга	[Schmidt <i>et al.</i> , 2004] [Zhao <i>et al.</i> , 1999]
PRTN3	референсная rs2074639:c>t (-7c>t)	18, 21	gggcTATAAGAGgagccttga _{cc} gtgg gggcTATAAGAGgagccttga _t cgtgg	+0,09*	0,08	Избыток миелобластина	Лейкемия	[Zimmer <i>et al.</i> , 1992]
NOS2A	референсная NG 011470.1:g.4980 T>C (-21t>c)	нд	atgggggtgagTATAAATAc _{tt} ccttgg atgggggtgagTATAAATAc _c tccttgg	+0,21*	0,10	Избыток NO-синтазы	Восприимчивость к инфекциям	[Burgner <i>et al.</i> , 2003]
PGR	референсная rs10895068:G>A (+331G>A)	1, 378	aaagtccgggaGATAAAGGagccgcgt aaagtccgggaGATAAAG _A agccgcgt	+0,30*	0,07	Избыток В-формы рецептора	Рак эндометрия	[Liu <i>et al.</i> , 2001]
BCR	референсная -48A> Δ	нд	cggcgcgggcccccTTTAAG _A ggcccgcg cggcgcgggcccccTTTAAG __ ggcccgcg	+0,50*	0,08	Повышенный выход гиперактивной тирозинкиназы	Более тяжелая форма лейкоза	[Lifshitz <i>et al.</i> , 1988]

Таблица 1 продолжение

Ген	Аллель	MAF, % (1000 genome)	Последовательность, 5'-3'	δ	$\pm\delta_{5\%}$	Биохимическое проявление	Заболевание или предрасположенность	Ссылка
<i>F3</i>	референсная rs563763767:C>T (-21C>T)	0,080	gссггсссТТТАТАG C гсгсггггсгс gссггсссТТТАТАG T гсгсггггсгс	+0,52	0,09	Избыток тканевого фактор	Инфаркт миокарда	[Arnaud <i>et al.</i> , 2000]
<i>HBZP1</i>	референсная NG 000006. 1:g.20573t>c (-70t>c)	нд	ctgссacaccсacattat t agaaaat ctgссacaccсacattat c agaaaat	+0,56*	0,22	Транскрипция псевдогена ингибирует гены кластера α - глобинов: μ , $\alpha 1$, $\alpha 2$	α -талассемия	[De Gobbi <i>et al.</i> , 2006]
<i>CRYGEP</i>	референсная rs55999272:c>t (-28 c >t)	29,792	gссctсctgctata c agсссгсгсгс gссctсctgctata t agсссгсгсгс	+0,84*	0,12	Повышенное количество белка псевдогена	Врожденная катаракта	[Brakenhoff <i>et al.</i> , 1994]
<i>IL1B</i>	референсная rs1143627:C>T (-31 C >T)	47,244	ttttgaaagc C ATAAAAAcagсгagg ttttgaaagc T ATAAAAAcagсгagg	+1,18*	0,08	Избыток интерлейкина -1 β	Рак легких	[Zienolddiny <i>et al.</i> , 2004]

Обозначения: ТАТА-бокс[#] - заглавные БУКВЫ; Δ - делеция; (ТА)₂ и ₃ - соответственно ТАТА и ТАТАТА, -82T>C&)-номера позиций промотора гена *CYP2B6*, кодирующего цитохром P450 2B6, общеприняты от старта трансляции [Zukunft *et al.*, 2005],], ^a – расчет SNP_TATA_Comparator, ^b – в выборке голландцев с ИБС [Zwarts *et al.*, 2002]

Значимые изменения сродства ТВР к ТАТА отмечены в таблице символом «*». Зеленым цветом выделены гены, SNPs в которых вызывают моногенные заболевания, светло-зеленым – гены, ассоциация SNPs которых с предрасположенностью к заболеваниям показана с помощью GWAS, желто-зеленым – гены, ассоциация SNPs которых с предрасположенностью к заболеваниям показана с помощью статистического анализа групп индивидуумов с заболеванием и без.

Экспериментальное определение равновесной константы диссоциации, (K_D), константы скорости ассоциации k_a и константы скорости диссоциации k_d TBP с олигонуклеотидами, идентичными референсным и минорным аллелям TATA-боксов генов человека

Среди референсных последовательностей выбранных для экспериментального исследования генов было 2 консенсусных TATA-бокса, 7 TATA-подобных элементов и 2 TATA-несодержащих последовательности.

Показано, что функционально значимые SNPs в TATA-подобных элементах (см. Таблицу 2) вызывают изменение аффинности TBP/TATA в 2 – 9 раз.

Замена в консенсусном TATA-боксе гена *TPI*, выявленная у гетерозигот с гаплотипом -5A/G -8G/A -24T/G и сниженной до 54-75% активностью TPI в эритроцитах: [Mohrenweiser, 1987, Humphries *et al.*, 1999], вызывает снижение аффинности TBP/TATA в 31 раз.

Замена во фланкирующей последовательности консенсусного TATA-бокса гена *NOS2A*, как и ожидалось, вызывает меньшее изменение аффинности TBP/TATA, чем замены в самом TATA-боксе – всего на 12%.

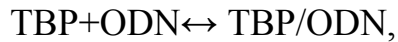
Замена -26G>C в TATA-несодержащей последовательности -30 района гена *FIX*, ассоциированная с гемофилией Лейдена, разрушает сайт связывания HNF4, находящийся на месте TATA-бокса [Reijnen *et al.*, 1992], и вопреки прогнозу не влияет на исходно низкое (510nM) сродство TBP к данной последовательности.

Для 9 генов были выполнены кинетические эксперименты. Было проанализировано взаимодействие с TBP 25-и последовательностей, из которых было 5 консенсусных TATA-боксов, 6 TATA-несодержащих последовательностей, и 14 последовательностей, первоначально оцененных как TATA-подобные элементы. Полученные кинетические характеристики (k_a и k_d) перекрываются с областью значений, характерных для взаимодействия TBP/TATA с учетом разницы в последовательности TATA-бокса [Jung, Mikata, and Lippard, 2001; Hoopes, LeBlanc, and Hawley, 1988; Strahs *et al.*, 2003].

Эксперименты показали, что для разработанной нами методики анализа (кинетика

связывания, количественный EMSA, детекция ^{32}P на фосфоимиджере)
 взаимодействие описывается простым одностадийным механизмом

$$k_a$$



$$k_d$$

так что $K_D = k_d/k_a$

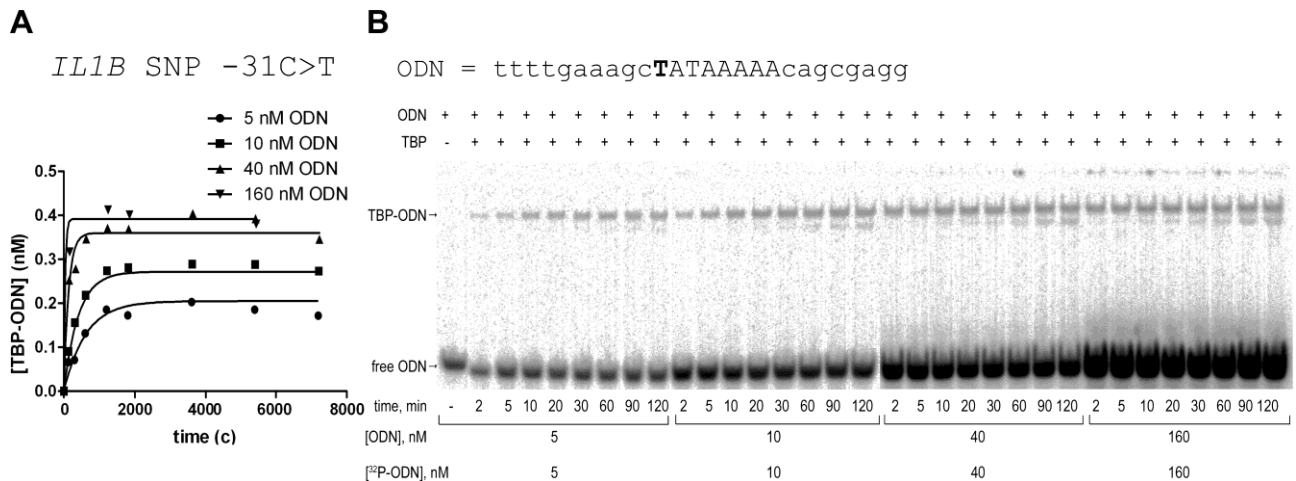


Рис. 1. Типичная схема постановки эксперимента по кинетике связывания ТВР с ТАТА-содержащим ОДН методом количественного EMSA

Результаты исследования показали, что снижение аффинности ТВР/ТАТА происходит преимущественно за счет снижения константы скорости ассоциации – k_a падает в 2,5 – 36 раз. В 70% случаев снижение k_a сопровождается компенсирующим его влияние на аффинность снижением скорости диссоциации комплекса ТВР/ТАТА. Полученные данные не позволяют говорить о том, что при низком сродстве транскрипционного фактора (ТФ) к мишеневому сайту ДНК должна увеличиваться скорость диссоциации комплексов, а скорости ассоциации (k_a) должны быть практически одинаковы для специфических и неспецифических сайтов, см. обзор [Dinant C., 2009]. Как можно видеть из полученных значений k_a и k_d (Табл.2), при изменении сродства могут изменяться обе скорости и, в том числе, при менее специфическом взаимодействии скорость диссоциации может падать, т.е. комплексы становятся более долгоживущими. Возможно, данный факт отражает показанную ранее способность ТВР к одномерному скольжению по ДНК [Coleman and Pugh, 1995, Marklund *et al.*, 2013].

Таблица 2. Влияние SNPs, ассоциированных с заболеваниями, на взаимодействие ТВР/ТАТА

Ген	Аллель	Заболевание	Последовательность, 5'-3'	Прогноз		Эксперимент								Score PWM TATA
						Равновесное связывание			Кинетика связывания					
				-Ln[K _D]	δ	K _D , nM	-Ln[K _D]	δ	k _a (M ⁻¹ s ⁻¹)	k _d (s ⁻¹)	t _{1/2} (min)	K _D (nM)		
HBB	WT	Норма	cagggctgggCATAAAAGtcagggca	18.47		44±3	16.94		(1.4±0.1)×10 ⁴	(7.1±0.7)×10 ⁻⁴	16±2	50±7	-3,51	
	rs33981098:A>G (-31A>G)	β-талассемия	cagggctgggCGTAAAAAGtcagggca	17.87	-0.60	99±9	16.13	-0.81	нд	нд	нд	нд	-7,89	
	rs33980857:T>A (-30T>A)	β-талассемия	cagggctgggCAAAAAAGtcagggca	17.01	-1.46	116±5	15.97	-0.97	(1.7±0.2)×10 ³	(1.9±0.3)×10 ⁻⁴	60±10	110±20	-8,12	
	rs33980857:T>C (-30T>C)	β-талассемия	cagggctgggCACAAAAAGtcagggca	17.49	-0.98	111±8	16.01	-0.93	(8±2)×10 ²	(1.1±0.3)×10 ⁻⁴	100±30	140±60	-7,00	
	rs34598529:A>G (-29A>G)	β-талассемия	cagggctgggCATGAAAAGtcagggca	17.23	-1.24	390±20	14.76	-2.18	(1.1±0.3)×10 ³	(5.3±0.4)×10 ⁻⁴	22±2	500±200	-7,28	
	rs33931746:A>G (-28A>G)	β-талассемия	cagggctgggCATAGAAGtcagggca	17.76	-0.71	87±8	16.26	-0.68	(1.2±0.1)×10 ³	(6.6±0.9)×10 ⁻⁴	18±2	560±80	-8,24	
	rs33931746:A>C (-28A>C)	β-талассемия	cagggctgggCATACAAGtcagggca	17.94	-0.53	300±20	15.02	-1.92	(4.1±0.5)×10 ²	(2.0±0.3)×10 ⁻⁴	58±9	500±100	-8,14	
	NG 000007.3:g.70518A>T (-27A>T)	β-талассемия	cagggctgggCATAATAGtcagggca	18.11	-0.36	64±3	16.56	-0.38	(3.6±0.4)×10 ⁴	(2.2±0.4)×10 ⁻³	5±1	60±10	-7,16	
HBD	WT	Норма	acaggaccagCATAAAAGgcagggca	18.94		46 ± 5	16.89		(2.3±0.3)×10 ⁴	(8±2)×10 ⁻⁴	14±4	36±4	-2,54	
	rs35518301:A>G (-31A>G)	δ-талассемия	acaggaccagCGTAAAAAGgcagggca	18.33	-0.61	120±10	15.94	-0.95	(2.0±0.1)×10 ³	(2.0±0.4)×10 ⁻⁴	60±10	100±20	-6,92	
HBZP1	WT	Норма	ctgccacaccaCattattagaaaat	17.72		152 ±6	15.70		(2.3±0.3)×10 ³	(4.2±0.6)×10 ⁻⁴	28±4	160±30	-11,83	
	NG 000006.1:g.20573T>C (-70T>C)	α-талассемия	ctgccacaccacattatcagaaaat	18.28	+0.56	112 ±6	16.00	+0.30	(2.4±0.2)×10 ³	(2.5±0.3)×10 ⁻⁴	46±6	100±10	-13,78	
CYP2A6	WT	Норма	tttcaggcagTATAAAGGcaaaccac	19.87		17±4	17.89		нд	нд	нд	нд	3,45	
	rs29399433:T>G (-48T>G)	Рак легких	tttcaggcagTAGAAGGcaaaccac	18.38	-1.49	80±20	16.34	-1.55	нд	нд	нд	нд	8,06	
SOD1	WT	Норма	aggtctggccTATAAAGT agtcgcgg	19.21		40±8	17.03		нд	нд	нд	нд	5,54	
	rs7277748:A>G (-27A>G)	АЛС	aggtctggccTGTAAAGT agtcgcgg	18.04	-1.17	170±30	15.59	-1.44	нд	нд	нд	нд	9,92	
MBL2	WT	Норма	catctatttcTATATAGCctgcaccc	19.68		28±4	17.39		(3.0±0.3)×10 ⁴	(1.0±0.1)×10 ⁻³	12±1	34±5	-6,13	
	rs72661131:T>C (-35T>C)	Иммуносупрес-сия	catctatttcTACATAGCctgcaccc	18.57	-1.11	58±9	16.66	-0.73	(1.2±0.1)×10 ⁴	(6.4±0.7)×10 ⁻⁴	18±2	54±7	-9,62	

Таблица 2, продолжение													
<i>TP11</i>	WT	Норма	cgcgggcgctcTATATAAGtgggcagct	20.11		4.8±0.7	19.15		$(1.8 \pm 0.1) \times 10^5$	$(7.2 \pm 0.8) \times 10^{-4}$	16±2	4.1±0.5	-2,76
	rs1800202:T>G (-24T>G)	Недостаток TPI	cgcgggcgctcTATAGAAGtgggcagct	19.08	-1.03	150±20	15.71	-3.41	$(5.0 \pm 0.8) \times 10^5$	$(5.1 \pm 0.7) \times 10^{-4}$	23±3	100±20	-6,97
<i>F9</i>	WT	Норма	acagctcagctTGTACTTTtggtacaa	18.24		510±60	14.49		$(2.4 \pm 0.2) \times 10^5$	$(1.4 \pm 0.1) \times 10^{-3}$	8.3±0.6	580±20	-18,26
	NG 007994.1:g.4975 G>C (-48G>C)	Гемофилия В Лейдена	acagctcagctTCTACTTTtggtacaa	17.75	-0.49	500±100	14.51	0.02	$(1.0 \pm 0.1) \times 10^5$	$(5.6 \pm 0.9) \times 10^{-4}$	21±3	560±30	-17,14
<i>IL1B</i>	WT	Норма	ttttgaaagcCATAAAAAcagcgagg	18.67		29±4	17.36		$(1.0 \pm 0.1) \times 10^4$	$(2.1 \pm 0.5) \times 10^{-4}$	60±10	20±4	-3,67
	rs1143627:C>T (-31C>T)	Рак легких	ttttgaaagcTATAAAAAcagcgagg	19.85	+1.18	7±1	18.78	+1.42	$(1.6 \pm 0.2) \times 10^5$	$(8 \pm 1) \times 10^{-4}$	14±2	4.8±0.6	-1,61
<i>F3</i>	WT	Норма	gccgggccTTTATAGCgcgcggggca	18.91		72±9	16.45		$(2.3 \pm 0.2) \times 10^4$	$(1.7 \pm 0.2) \times 10^{-3}$	6.8±0.8	77±9	-5,82
	NG 029366.1:g.5078 C>T (-21C>T)	Тромбофлебит и инфаркт миокарда	gccgggccTTTATAGTgcgcggggca	19.43	+0.52	26±3	17.47	+1.02	$(1.0 \pm 0.1) \times 10^5$	$(1.9 \pm 0.2) \times 10^{-3}$	6.1±0.6	19±1	-6,09
<i>NOS2</i>	WT	Норма	atgggggtgagTATAAATActtcttgg	19.85		1.8±0.2	20.14		$(4.2 \pm 0.4) \times 10^5$	$(8 \pm 1) \times 10^{-4}$	14±2	1.9±0.3	-3,91
	NG 011470.1:g.4980 T>C (-60T>C)	Восприимчи- вость к малярии	atgggggtgagTATAAATAcCtcttgg	20.06	+0.21	1.6±0.3	20.25	+0.11	$(4.0 \pm 0.3) \times 10^5$	$(7.3 \pm 0.9) \times 10^{-4}$	16±2	1.8±0.2	-2,51
<i>AdML</i>	контроль		tgaaggggggcTATAAAAGggggtgg	19.12		4.3±0.7			$(4.4 \pm 0.3) \times 10^5$	$(9 \pm 1) \times 10^{-4}$		2.1±0.2	-0,91

Обозначения: НД – нет данных; TATATAAG – TATA-бок с консенсусной последовательностью, CATAAAAG – сайт связывания ТВР, содержащий неканонические, но допустимые в данных позициях основания, TATAGAAG – сайт связывания ТВР, содержащий запрещенные основания; $(1.8 \pm 0.1) \times 10^5$ – значения, характерные для TATA-боксов с консенсусной последовательностью, $(1.0 \pm 0.1) \times 10^4$ – значения, характерные для TATA-подобных элементов, $(1.0 \pm 0.1) \times 10^3$ – значения, характерные для сайтов связывания ТВР, содержащих редкие и запрещенные основания, либо не являющихся TATA-боксами с точки зрения критерия Бухера (имеющих Score PWM_{TATA} ниже - 8,16), АЛС – амиотрофический латеральный склероз.

Проведенные эксперименты показали, что для кинетики образования и диссоциации комплексов ТВР/ТАТА есть общие закономерности, связанные со степенью гомологии последовательности с ТАТА-боксом. ТАТА-боксы с консенсусной последовательностью ТАТАWAWR имеют самые высокие скорости связывания с ТВР: для них константа скорости связывания k_a порядка $10^5 \text{ M}^{-1}\text{s}^{-1}$, при наномолярной аффинности. Для ТАТА-несодержащих последовательностей k_a на 2 порядка ниже ($10^3 \text{ M}^{-1}\text{s}^{-1}$), при характерной аффинности от 100 до 580 нМ. ТАТА-подобные элементы генов *HBB*, *HBD*, *MBL*, *IL1 β* , *F3* имеют промежуточный порядок k_a , равный $10^4 \text{ M}^{-1}\text{s}^{-1}$ при аффинности 20-77 нМ.

Наши исследования показали, что последовательности, первоначально отнесенные к ТАТА-подобным элементам, но содержащие стерически запрещенные [Patikoglou *et al*, 1999] основания, встречающиеся в $\text{PWM}_{\text{ТАТА}}$ с частотой $< 1\%$, имеют пограничный score $\text{PWM}_{\text{ТАТА}}$ ниже -7,00, и k_a порядка $10^3 \text{ M}^{-1}\text{s}^{-1}$, характерную для ТАТА-несодержащих последовательностей.

Не менее интересна зависимость от последовательности константы скорости диссоциации комплексов k_d . Последовательности, отвечающие консенсусу ТАТАWAWR, имеют k_d в довольно узком диапазоне $7,3 \times 10^{-4} \text{ s}^{-1}$ - $9 \times 10^{-4} \text{ s}^{-1}$ и, соответственно, время жизни комплексов 16-13 мин. Для ТАТА-подобных элементов и ТАТА-несодержащих последовательностей характерен большой разброс параметров – k_d в диапазоне от $(2.2 \pm 0.4) \times 10^{-3} \text{ s}^{-1}$ до $(1.1 \pm 0.3) \times 10^{-4} \text{ s}^{-1}$. Авторы работы [Tora and Timmers, 2010] видят смысл в быстром освобождении консенсусных ТАТА-боксов высокорегулируемых генов (именно для скорости регуляции) и «липкости» неспецифической последовательности, на которой сидит ТВР генов домашнего хозяйства.

Статистическая обработка экспериментальных данных и оценка соответствия прогнозов изменения аффинности ТВР in silico экспериментальным данным in vitro – определение коэффициентов линейной

корреляции.

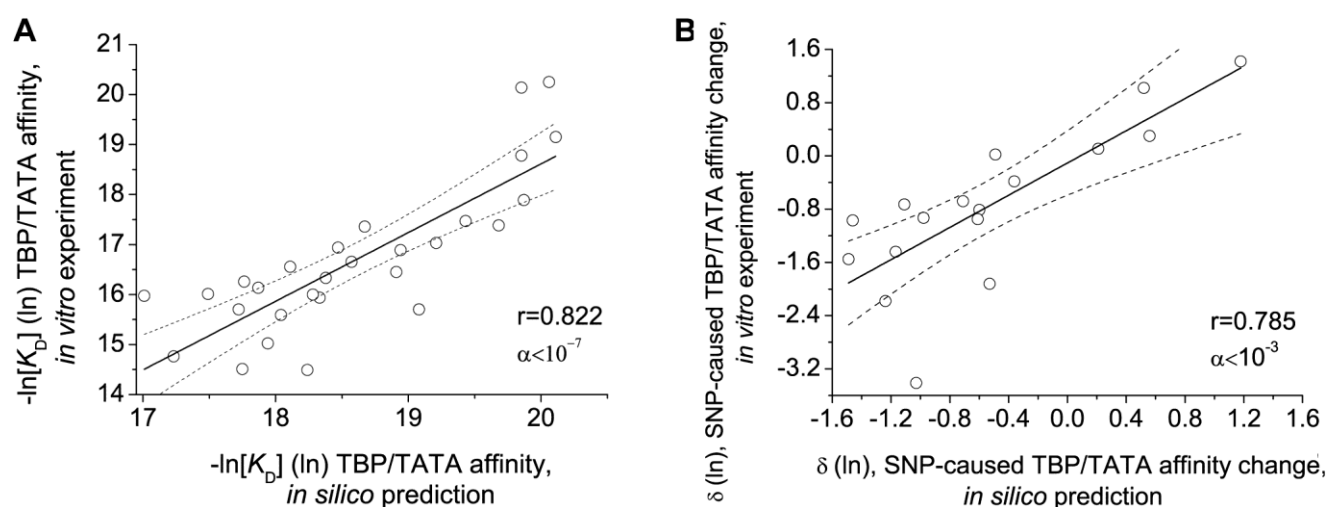


Рис 2. Экспериментально измеренная аффинность ($-\ln K_D$) (А) и изменения аффинности (δ) (В) в результате SNPs в ТАТА-боксах имеют высокую корреляцию с прогнозами *in silico*

Статистический анализ данных показывает, что значения равновесных K_D с высокой достоверностью коррелируют с прогнозами: коэффициенты линейной корреляции, $r = 0.822$ для прогнозов значений равновесных K_D ($-\ln K_D$) и $r = 0.785$ для прогнозов изменения равновесных K_D (δ) при SNPs в ТАТА-боксах ($\delta = -\ln[K_{D,TATA Mut}] - (-\ln[K_{D,TATA}])$), при значимости $\alpha < 10^{-7}$ и $\alpha < 10^{-3}$, соответственно (см. Рис 2). Результаты позволяют говорить о применимости уравнения пошагового связывания ТВР/ТАТА для анализа сайтов связывания ТВР и выявления потенциально функционально значимых SNPs.

В то же время наблюдаемая недооценка предсказаний аффинности ТВР и мутационных изменений ТАТА-боксов *in silico* может указывать на вероятность кооперативного влияния контекстных повреждений ДНК на участке связывания ТВР\ТАТА.

SNPs ТАТА-боксов, могут, действительно, повреждать совпадающие с этими ТАТА-боксами контакты кор-промоторов с димерами НЗ-Н4 гистоновых октамеров. Обнаруженные нами достоверные систематические занижения прогнозов *in silico*, относительно экспериментальных измерений мутационного повреждения ТАТА-боксов, соответствуют влиянию на них повреждения перекрывающихся с ними контактов ДНК/(НЗ-Н4) [Richmond and Davey, 2003,

Gershenzon, Trifonov, and Ioshikhes, 2006], аналогично детально описанному ранее случаю композиционного элемента NFATp/AP-1 [Kel *et al.*, 1999]. Это указывает на возможность композиционного элемента TATA/(H3-H4)(H2A-H2B)₂(H3-H4) кор-промоторов генов эукариот, который до сих пор не учитывался в *in silico* моделях.

Заключение

Предпринятое экспериментально-биоинформатическое исследование, имеющее целью всесторонний анализ функционально значимых SNPs, впервые показало, каким образом SNPs в TATA-боксах оказывают влияние на фенотипические признаки человека.

Экспериментально подтверждено, что SNPs, для которых прогнозы изменения сродства TBP/TATA, сделанные с помощью уравнения пошагового связывания, статистически достоверны, являются функционально значимыми; а изменения аффинности TBP/TATA соответствуют биохимическим проявлениям SNPs и реализуются через изменения как скоростей ассоциации, так и скоростей диссоциации комплексов TBP/TATA.

Традиционно полагается, что на TATA-несодержащих промоторах специфичность взаимодействия TBP достигается с использованием механизмов прямого и непрямого взаимодействия с активаторными белками и TAFs, а на TATA-содержащих промоторах определяющим является ДНК-белковое взаимодействие. Однако, даже для TATA-боксов, отвечающих консенсусу TATAWAWR, характерная константа скорости связывания k_a порядка $10^5 \text{M}^{-1}\text{s}^{-1}$, что на три порядка ниже скорости простой бимолекулярной реакции, ограниченной скоростью диффузии [von Hippel and Berg, 1989]. По-видимому, наши кинетические данные подтверждают показанный исследовательским коллективом E. Nogales факт того, что собственная аффинность TATA бокса играет ограниченную роль в привлечении TBP на промотор [Louder *et al.*, 2016], а TFIIID, динамично связывая протяженный интерфейс кор-промоторных элементов (обнаруженных в кор-промоторах всех экспериментально исследованных нами

генов за исключением *MBL2*), играет роль «молекулярной линейки» для точного позиционирования TBP [Patel *et al.*, 2018]. Более того, в последних работах [Patel *et al.*, 2018, Bhuiyan and Timmers, 2019] идет речь о смене парадигмы с «связывание кор-промотора с помощью TFIID всегда начинается или направляется связыванием TBP с промоторной ДНК» на «связывание ТАТА с помощью TBP является последним, а не первым шагом в связывании кор-промотора с помощью TFIID».

Выводы

1. Создана коллекция аннотированных SNPs в сайтах связывания TBP, для которых показана возможность участия в формировании наследственной предрасположенности человека к заболеваниям. Коллекция содержит 34 промотора, в которых описано 42 замены в (-20-70) районе. Из них 5 промоторов (12 патогенных SNPs) принадлежат генам, повреждение которых вызывает моногенные заболевания, ассоциация 2-х SNPs подтверждена GWAS, ассоциация 3-х SNPs показана с помощью статистического анализа групп индивидуумов с заболеванием и без. Для остальных SNPs предположение о связи с заболеваниями сделано на основе известной связи биохимического проявления с заболеванием. Анализ показал, что прогноз влияния SNPs на аффинность TBP/ТАТА качественно соответствует литературным данным о биохимическом проявлении SNPs.
2. Впервые экспериментально показано, что функционально значимые SNPs в сайтах связывания TBP вызывают изменение аффинности TBP/ТАТА, выраженной равновесной константой диссоциации K_D , в диапазоне 1,5 – 31 раз.
3. Статистическая обработка экспериментальных данных показала, что экспериментально измеренная аффинность TBP/ТАТА, выраженная $-\ln(K_D)$, и изменения аффинности (δ) = $-\ln[K_{D, \text{ТАТА SNP}}] - \{-\ln[K_{D, \text{ТАТА}}]\}$ достоверно коррелируют с прогнозами, сделанными с использованием уравнения пошагового

связывания: $r = 0.822$ для $-\ln(K_D)$ и $r = 0.785$ для изменения (δ), при значимости $\alpha < 10^{-7}$ и $\alpha < 10^{-3}$, соответственно.

4. Впервые показано, что снижение аффинности ТВР/ТАТА происходит преимущественно за счет снижения константы скорости ассоциации – k_a падает в 2,5 – 36 раз. В 70% случаев снижение k_a сопровождается компенсирующим его влияние на аффинность снижением скорости диссоциации комплекса ТВР/ТАТА.

Основные публикации

1. 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):601-8. IF 5.359
2. 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. // PLoSOne. 2013; 8(2):e54626. IF 2.766
3. Драчкова И.А., Шеховцов С.В., Пельтек С.Е., Пономаренко П.М., Аршинова Т.В., Пономаренко М.П., Меркулова Т.И., Савинкова Л.К., Колчанов Н.А. Изучение взаимодействия ТВР человека с ТАТА-элементом промотора гена *NOS2A* с использованием метода поверхностного плазмонного резонанса. // Вавиловский журнал Генетики и селекции, 2012, 16(2):391-396. IF 0.472
4. Drachkova I.A., Ponomarenko P.M., Arshinova T.V., Ponomarenko M.P., Suslov V.V., Savinkova L.K., Kolchanov N.A. *In vitro* examining the existing prognoses how TBP binds to TATA with SNP associated with human diseases. // Health, 2011, 3(9):577-583. IF 0.73
5. Savinkova L.K., Ponomarenko M.P., Ponomarenko P.M., Drachkova I.A., Lysova M.V., Arshinova T.V., Kolchanov N.A. TATA box polymorphisms in human gene promoters and associated hereditary pathologies. // Biochemistry (Mosc), 2009, 74(2):117-29. IF 1.724