

на правах рукописи

АРКОВА ОЛЬГА ВЛАДИМИРОВНА

**Изучение потенциально функционально
значимых полиморфизмов ТАТА-боксов
промоторов генов липидного метаболизма
человека**

03.02.07 - Генетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Новосибирск – 2017

Работа выполнена в секторе молекулярно-генетических механизмов белок - нуклеиновых взаимодействий Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН), г.Новосибирск.

Научный руководитель: кандидат биологических наук, зав. сектором молекулярно-генетических механизмов белок-нуклеиновых взаимодействий

Савинкова Людмила Кузьминична

ИЦиГ СО РАН, г. Новосибирск

Официальные оппоненты:

доктор биологических наук, проф.

Гуляева Людмила Федоровна

ФГБНУ НИИ молекулярной биологии и биофизики, г. Новосибирск.

доктор биологических наук,

Рыкова Елена Юрьевна

ФГБНУ Институт химической биологии и фундаментальной медицины г.

Новосибирск

Ведущее учреждение:

Федеральное бюджетное учреждение науки Государственный научный центр вирусологии и биотехнологии «Вектор», Новосибирская область, г. Кольцово.

Защита диссертации состоится «___» _____ 201___ года на заседании диссертационного совета по защите диссертаций Д 003.011.01 на базе Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в конференц-зале Института по адресу: проспект Лаврентьева, 10, г.Новосибирск, 630090, тел. (383) 3634906, факс (383) 3331278, email: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института: www.bionet.nsc.ru

Автореферат разослан «___» _____ 201___ г.

Ученый секретарь

диссертационного совета,

доктор биологических наук

Т.М. Хлебодарова

Общая характеристика работы

Актуальность

Изучение молекулярных механизмов генетической предрасположенности к различным заболеваниям человека является одной из фундаментальных проблем генетики, молекулярной генетики и медицины. Одним из наиболее активно развивающихся направлений в этой области является выявление однонуклеотидных полиморфизмов (SNP—однонуклеотидный полиморфизм, Single nucleotide polymorphism) кандидатных генов, ассоциированных с каким-либо наследственным заболеванием, и установление связи между определенным вариантом последовательности ДНК и патологией. Оказалось, что кодирующие белки области генома, ОНП в которых приводят к аминокислотной замене и нарушению функции белка, гораздо более изучены, чем регуляторные области, SNP в которых могут изменять сайты связывания транскрипционных факторов. Это может приводить к аллель-специфической экспрессии генов, различиям в чувствительности к факторам внутренней и внешней среды, разной предрасположенностью к заболеваниям и чувствительности к терапии и т. п. В связи с этим выявление SNP регуляторных районов генов, в частности, промоторов, и изучение молекулярных механизмов, посредством которых SNP сайтов связывания транскрипционных факторов могут оказывать влияние на фенотипические признаки, является актуальным и перспективным направлением.

Ранее в секторе молекулярно-генетических механизмов белок-нуклеиновых взаимодействий ИЦиГ были получены экспериментальные данные о зависимости аффинности ТСБ к ТАТА-боксам некоторых генов эукариот от последовательностей ТАТА-боксов (Савинкова и др., 2007), которые были проанализированы в лаборатории теоретической генетики ИЦиГ СО РАН Пономаренко М.П. В результате анализа в 2008 г. была разработана модель и уравнение пошагового связывания ТСБ с ТАТА-боксом (Пономаренко и др., 2008), позволяющее прогнозировать сродство ТСБ/ТАТА, основанное на анализе последовательности ДНК ТАТА-боксов. Затем была собрана электронная коллекция SNP ТАТА-боксов с фланкирующими последовательностями, ассоциированных с наследственными заболеваниями, которые были подтверждены клинически и молекулярно-биологически, и с помощью разработанного уравнения сделаны прогнозы изменения сродства ТСБ/ТАТА. А в нашем секторе проведена экспериментальная верификация сделанных прогнозов. Эксперименты показали высокую корреляцию экспериментальных результатов с прогнозами, сделанными *in silico*. Коэффициент линейной корреляции, $r = 0.822$ при $\alpha < 10^{-7}$ (Savinkova et al., 2013).

Полученные результаты позволили нам приступить к решению следующей задачи - компьютерному формированию библиотеки неисследованных SNP ТАТА-боксов промоторов генов человека, отобранных из

доступных баз данных, анализу *in silico* последовательностей ДНК ТАТА-боксов с помощью разработанного на основании уравнения пошагового связывания (Пономаренко и др., 2008) компьютерного инструмента SNP_TATA Comparator (Рассказов и др., 2013), и отбору наиболее подходящих примеров для экспериментальной верификации *in vitro* и *in vivo* сделанных прогнозов *in silico* по влиянию SNP на сродство ТСБ/ТАТА.

Цель работы заключалась в формировании библиотеки неисследованных однонуклеотидных полиморфизмов ТАТА-боксов промоторов генов липидного метаболизма человека и их анализе *in silico*, *in vitro* и *in vivo* для выявления потенциально функционально значимых.

Для достижения цели были поставлены следующие задачи:

1. Формирование электронной библиотеки SNP и компьютерный анализ неаннотированных однонуклеотидных полиморфизмов ТАТА-боксов промоторов генов липидного метаболизма человека.
2. Определение скоростей образования и распада комплексов ТСБ/ТАТА анцестральных и минорных аллелей с помощью метода EMSA.
3. Анализ сродства ТСБ к ТАТА-боксам анцестральных и минорных аллелей генов липидного метаболизма
4. Определение в режиме реального времени кинетических характеристик комплексов ТСБ/ТАТА методами остановленной струи и резонансного переноса энергии.
5. Получение плазмидных конструкций с анцестральными аллелями и минорными, SNP-содержащими ТАТА-боксами промоторов генов *LEP* и *CYP2A6*, и изучение их влияния на экспрессию репортерного гена *LUC*.

Научная новизна

Впервые в мире проведено компьютерно-экспериментальное исследование, включающее компьютерное предсказание и экспериментальную верификацию неаннотированных однонуклеотидных замен ТАТА-боксов промоторов генов человека. Компьютерный анализ для прогноза значимости последовательностей был проведен с использованием разработанного в ИЦиГ СО РАН компьютерной системы SNP_TATA_Comparator. Экспериментальная верификация *in vitro* с использованием метода задержки ДНК в геле, метода остановленной струи и флуоресцентного резонансного переноса энергии, и *in vivo* на культуре клеток показала высокую корреляцию опыт/прогноз ($r=0.89$). В работе впервые получены кинетические характеристики комплексов ТСБ с олигодезоксинуклеотидами, идентичными ТАТА-боксам с фланкирующими нуклеотидами анцестральных и минорных аллелей ряда генов человека, которые по прогнозам могут быть

ассоциированы с повышенным риском возникновения наследственных нарушений энергетического метаболизма.

Основные положения, выносимые на защиту:

1. Экспериментальная верификация показала эффективность использования Web сервиса SNP_TATA_Comparator для предсказания новых потенциально функционально значимых мутаций ТАТА-боксов
2. Мутации ТАТА-боксов, изменяющие средство ТАТА-связывающего белка, влияют на экспрессию генов, находящихся под регуляцией этих ТАТА-боксов содержащих промоторов.
3. Формирование комплексов ТАТА-связывающего белка с ТАТА-боксами в режиме реального времени происходит как по прямому пути, так и через промежуточные стадии, количество которых возрастает с увеличением концентрации ТАТА-связывающего белка и длины олигонуклеотидов.

Теоретическая и практическая значимость исследования

Теоретическая значимость работы заключается в том, что экспериментально доказана эффективность использования разработанного компьютерного метода SNP_TATA_Comparator для предсказания потенциально значимых мутаций ТАТА-боксов промоторов генов человека. Теоретическое значение работы также заключается в том, что определен механизм реализации мутационных изменений ТАТА-боксов реальных промоторов генов липидного метаболизма человека.

Практическая значимость исследований заключается в том, что полученные данные могут использоваться для доклинической проверки врачами кандидатных SNP-маркеров ТАТА-боксов в предиктивно-превентивной персонализированной медицине.

Вклад автора

Основные результаты получены автором самостоятельно. Автор принимал личное участие в планировании, проведении и обсуждении всех экспериментов, по результатам которых написана диссертация.

Биоинформационный анализ нуклеотидных последовательностей проводился совместно с к.б.н. М.П. Пономаренко. Выращивание химически компетентных клеток и их трансформация проводились совместно с И.А. Драчковой. Оценка взаимодействия ТСБ с ТАТА-боксами с помощью FRET проводилась совместно с к.х.н. Н.А. Кузнецовым. Плазмидные конструкции получены совместно с Е.Б. Шарыповой. Результаты по влиянию плазмидных конструкций на активность гена люциферазы в культуре клеток получены Е.В. Кашиной.

Публикации

Материал диссертации представлен в 14 работах, из них 10 статей и 4 тезиса. 5 статей опубликованы в журналах, индексируемых в Web-of-Science, 9 статей - в журналах SCOPUS, в РИНЦ - 4.

Апробация работы

Результаты данной работы были представлены и обсуждены: на конференции BGRS «Bioinformatics of Genome Regulation and Structure\Systems Biology - BGRS\SB-2014» (Новосибирск, 2014); на 7 съезде Российского общества медицинских генетиков (Санкт-Петербург, 2015); на конференции АМСА - Актуальные проблемы вычислительной и прикладной математики (Новосибирск, 2015); на конференции BGRS «Bioinformatics of Genome Regulation and Structure - BGRS-2016» (Новосибирск, 2016).

Структура и объем диссертации

Диссертация состоит из введения, обзора литературы, описания материалов и методов исследования, результатов и обсуждений, списка цитированной литературы, содержащего 243 источников (в том числе на иностранном языке 228). Объем диссертации составляет 187 страниц. Диссертация включает 21 рисунок, 10 таблиц и 4 схемы, 5 приложений.

Материалы и методы

Отбор неаннотированных SNP генов липидного метаболизма человека проводили из базы данных dbSNP NCBI (<http://www.ncbi.nlm.nih.gov/SNP/>). Достоверность изменения средства ТСБ/ТАТА при заменах в ТАТА-боксе исследуемых генов измеряли при помощи SNP_TATA Comparator (<http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl>). Последовательности ТАТА-боксов с фланками, использованные в настоящей работе, брали из БД Eukaryotic promotor database (<http://epd.vital-it.ch>).

Выделение рекомбинантного ТАТА-связывающего белка проводилось в три стадии по методике В.Ф. Pugh (Pugh, 1995).

Синтезированные ОДН (олигодезоксинуклеотиды) («BIOSYN», Новосибирск) для измерения кинетики методом «задержки ДНК в геле» (далее EMSA) метили фосфором ³²P и использовали для связывания с ТСБ. Количественный анализ радиоавтографов проводили с помощью компьютерной программы Quantity One – 4.5.0 (Bio-Rad).

Для анализа связывания ТСБ/ТАТА методом FRET на спектрометре, работающем в режиме stopped-flow (связывание изучали на базе Института ХБиФМ СО РАН ЛИМБ), использовали ОДН, синтезированные с флуоресцентной парой красителей Cy3/Cy5 и TAMRA/FAM на 5'-концах. Для анализа полученных данных и построения кинетической модели взаимодействия ТСБ с дуплексами ДНК, а также для расчета констант скоростей всех

элементарных реакций использовали программу Dynafit (Biokin, USA) (Kuzmic, 1996).

Для экспериментального изучения влияния выявленных SNP TATA-боксов на экспрессию репортерного гена *LUC* использовали линии клеток эукариот HCT116 и MCF7. Для получения репортерных конструкций TATA-содержащие участки промоторов (WT и минорные аллели) генов *Lep* и *CYP2A6* человека были клонированы в беспромоторный вектор pGL4.10 [luc2] (Promega, США). Для клонирования необходимые фрагменты ДНК были амплифицированы при помощи ПЦР с использованием праймеров, содержащих сайты рестрикции Xho I и Hind III, в качестве матрицы использовали ДНК клеток человека HCT-116 и MCF-7. Активность люциферазы измеряли с помощью двойного люциферазного набора для анализа (Dual-Luciferase, Promega, USA) согласно инструкции производителя. Беспромоторный вектор pGL4.10 использовали в качестве отрицательного контроля.

Результаты и обсуждение

В работе было проведено исследование *in silico* кор-промоторных районов 68 генов пищевого поведения человека. Для 62 генов было обнаружено 388 SNP, для шести генов SNP не было обнаружено. Для 296 из 388 SNP было оценено недостоверное изменение сродства ТСБ к кор-промоторам относительно референсного генома человека (hg19). Для остальных 91 SNP прогнозируется достоверное увеличение или уменьшение ТСБ/ДНК-сродства. Наиболее сильное изменение сродства ТСБ/TATA прогнозировалось для нескольких генов, участвующих в липидном метаболизме, среди которых для дальнейшей работы были отобраны исследуемые нами TATA-боксы промоторов.

Определение кинетических и термодинамических характеристик комплексов ТСБ-TATA анцестральных и минорных аллелей с помощью метода EMSA

Для определения кинетических констант (k_a и k_d) взаимодействия ТСБ с TATA-содержащими ОДН длиной 26 пн, характеризующих скорости образования и распада комплексов, использовали меченые ОДН. Результаты связывания ТСБ с ОДН, соответствующими «нормальным» и SNP-содержащим вариантам TATA-боксов, методом EMSA представлены в Таблице 1.

Leptin. В Таблице 1 представлены значения k_a и k_d , характеризующие скорости образования и распада комплексов ТСБ с TATA-боксом в норме и при содержании SNP, ассоциированных с ожирением, а также времена полураспада комплексов, характеризующие их прочность, и изменения свободной энергии Гиббса, характеризующие возможность спонтанного протекания реакции взаимодействия ТСБ/ОДН.

Таблица 1. Кинетические характеристики взаимодействия ТСБ/ОДН (определение констант связывания (k_a) и диссоциации (k_d) комплексов (ТСБ/ОДН) для референсных генов (Hg 19) и минорных аллелей.

Ген	ТАТА-бокс 5'-3' нить ДНК, 26 п.о.	Норма (Hg 19)/SNP	$k_a(M^{-1}c^{-1})$ скорости образования комплексов	$k_d(c^{-1})$ скорости распада комплексов	$-\Delta G$ (kcal/mol)	$t_{1/2}$ мин
LEP	gcTATAAGag	Hg 19	$(2.4 \pm 0.6) \cdot 10^3$	$(1.8 \pm 0.6) \cdot 10^{-4}$	9.71 ± 0.9	64.2 ± 12
	acTATAAGag	-38G>A, rs200487063	$(3.5 \pm 0.5) \cdot 10^3$	$(1.5 \pm 0.2) \cdot 10^{-4}$	10.04 ± 1	77 ± 9
	gcTATAAGtg	-30A>T, rs34104384	$(1.1 \pm 0.2) \cdot 10^4$	$(8 \pm 1) \cdot 10^{-4}$	9.73 ± 1	14.4 ± 4
	gcTGTAAAGag	-35A>G, rs201381696	$(5.6 \pm 0.8) \cdot 10^3$	$(1.3 \pm 0.2) \cdot 10^{-3}$	9.04 ± 0.9	8.9 ± 1
ABCA9	aTATACAAat	Hg 19	$(7 \pm 1) \cdot 10^3$	$(4.6 \pm 0.8) \cdot 10^{-4}$	9.79 ± 0.8	25.1 ± 7
	aTATGCAAat	-37A>G, rs367781716	$(2.9 \pm 0.4) \cdot 10^3$	$(4.9 \pm 0.7) \cdot 10^{-4}$	9.23 ± 1.2	23.6 ± 7
GCG	ttTATATActc	Hg 19	$(7 \pm 1) \cdot 10^4$	$(5 \pm 1) \cdot 10^{-4}$	11.1 ± 1	23.1 ± 5
	ttTATACACTc	-41T>C, rs183433761	$(3 \pm 1) \cdot 10^4$	$(6 \pm 1) \cdot 10^{-4}$	10.5 ± 0.9	19.3 ± 4
TPI	tcTATATAAg	Hg 19	$(2.0 \pm 0.1) \cdot 10^5$	$(6.8 \pm 0.8) \cdot 10^{-4}$	11.5 ± 0.9	16.8 ± 3
	tcTATAGAAg	-24T>G, rs1800202	$(4.6 \pm 0.5) \cdot 10^3$	$(4.9 \pm 0.6) \cdot 10^{-4}$	9.4 ± 1.3	23.57 ± 3
CYP2A6	gTATAAAGg	Hg 19	$(3.0 \pm 0.3) \cdot 10^4$	$(7 \pm 1) \cdot 10^{-4}$	10.4 ± 1	16.5 ± 3
	gTAGAAAGg	-34T>G, rs28399433	$(2.2 \pm 0.3) \cdot 10^3$	$(2.1 \pm 0.5) \cdot 10^{-4}$	9.57 ± 1.1	55 ± 13

Из представленных результатов можно увидеть, что SNP (мутация) -38G>A в кор-промоторе гена *LEP* приводит к увеличению скорости образования комплексов ТСБ/ТАТА на 50%, ОНП -30A>T приводит к увеличению скорости образования комплексов почти в 5 раз, а SNP -35A>G – к увеличению в 24 раза. Скорость распада комплексов для мутации -38G>A повысилась на 20%, для -30A>T уменьшилась в 4.4 раза, а для -35A>G повысилась в 14 раз. Аффинность (K_D) ТСБ к ОДН с SNP -38G>A увеличилась на 50%, с SNP -30A>T (Рис.1) - увеличилась примерно на 12%, а для SNP -35A>G аффинность снизилась в 2.8 раза (для WT $K_D = 80$ nM, а для замены -35A>G = 230 nM). При этом минимальное время полужизни комплексов, 8.9 мин, наблюдалось для SNP -35A>G, что вызвано увеличением скорости их распада в 14 раз относительно нормы и очень низкой аффинностью.

Как видно из Таблицы 1, самое большое увеличение скорости образования и распада комплексов ТСБ/ТАТА наблюдается для мутации -35A>G (в 24 раза и 14 раз, соответственно), вносящего изменение в последовательность ТАТА-бокса и заменяющего самый консервативный «А» на «G» (Caron et al., 1991). Это приводит к падению аффинности практически до неспецифического уровня (230 nM) и уменьшению времени жизни комплексов более чем 7 раз, что по литературным данным ассоциируется с ожирением и сопутствующей гипертензией (Caron et al., 1991). Следует отметить, что такие низкие значения аффинности ТСБ/ТАТА мы наблюдали и ранее (Savinkova et al., 2013): SNP -29A>G ТАТА-бокса промотора гена β -глобина приводил к еще большему снижению значения K_D - до 390 nM, что ассоциировалось с клинически доказанной промежуточной β -талассемией.

(на 30%) скорости диссоциации. Период полураспада комплексов возрастает в 1,4 раза. Когда ТАТА-бокс не изменен, период полураспада составляет 17 минут, а при наличии SNP в ТАТА-боксе период полураспада составляет 23 мин. Эта мутация приводит к уменьшению $-\Delta G$ на 2,1 ккал/моль. Существуют дефицитные по TPI люди, имеющие недостаток мРНК, у них активность фермента в эритроцитах и других клетках 2-10% от нормы; пациенты имели неврологические и мышечные расстройства и гемолитическую анемию. Наблюдались и другие носители гетерозиготных аллелей, которые имели умеренное снижение активности TPI в естественных условиях, на 26-50% (Watanabe et al., 1996). Было показано, что мутация $-24T>G$ (очень редко встречается, его распространенность составляет менее 1%) (Patikoglou et al., 1999), и это приводит к самым значительным, связанным с SNP, расстройствам - 43-кратное снижение скорости ассоциации ТСБ/ТАТА и наибольшее снижение сродства среди всей рассмотренной группы генов.

CYP2A6: В гене *CYP2A6* при замене $-34T>G$ rs28399433 падают скорости и ассоциации комплекса ТСБ/ТАТА (в 13,6 раза), и диссоциации его (в 3,5 раз). Уменьшение скорости ассоциации ТСБ/ТАТА с 30 до $2,2 \text{ M}^{-1}\text{s}^{-1}$ показывает, что комплексы образуются медленнее, но дольше не распадаются.

Таблица 2. Экспериментальная верификация прогнозируемых *in silico* значений сродства (K_D) ТСБ/ОДН и влияние на него неаннотированных SNP.

hg19	5'-3' нить ДНК, 26 п.о.	Аллель (WT, rs)	Прогноз		Эксперимент		
			$-\ln[K_D]$	δ	$K_D, \text{нМ}$	$-\ln[K_D]$	δ
LEP	atcgggccgcTATAAGagggcgggc	WT	19.43		80±20	16.34	
	atcgggcc a cTATAAGagggcgggc	-38G>A, rs200487063	19.85	0.42	40±10	17.03	0.69
	atcgggccgcTATAAG t ggggcgggc	-30A>T, rs34104384	19.70	0.27	70±20	16.47	0.13
	atcgggccgcT G TAAAGagggcgggc	-35A>G, rs201381696	18.22	-1.21	230±10	15.29	-1.05
ABCA9	gctcagaaaTATACAAataattgcaa	WT	19.66		70±20	16.47	
	gctcagaaaTAT C CAAataattgcaa	-37A>G, rs367781716	18.73	-0.93	170±30	15.59	-0.88
GCG	cactgctttTATATACTctccagctc	WT	20.89		8±2	18.64	
	cactgctttTATAC A CTctccagctc	-41T>C, rs183433761	20.28	-0.61	18±4	17.83	-0.81
CYP2A6	tttcaggcagTATAAAGgcaaacccac	WT	20.03		25±4	17.5	
	tttcaggcagTA G AAAGgcaaacccac	-34T>G, rs28399433	18.53	-1.5	100±30	16.12	-1.38
TPI	cgcggcgctcTATATAAgtgggcagt	WT	20.11		3±0.3	19.62	
	cgcggcgctcTATA G Agtgggcagt	-24T>G, rs1800202	19.08	-1.03	21±4	17.68	-1.94

В Таблице 2 приведены количественные прогнозы ($\ln K_D$) сродства ТСБ/ТАТА анцестральных аллелей и величина их изменений (δ) относительно референсного аллеля. В таблице также приведены экспериментально полученные значения сродства ТСБ/ТАТА и его изменений (δ). Как можно видеть их таблицы, наблюдается достаточно высокое соответствие прогнозов и экспериментальных

данных (сравнивать δ прогноз и эксперимент). Коэффициент линейной корреляции, $r = 0.89$.

Мы определили *in silico* все неаннотированные SNP промотора гена *LEP* человека, полученные в результате выполнения проекта «1000 Genomes project» (Colonna et al., 2014). Поиск проводился в районе от [-20 до -70] от старта транскрипции, где располагаются все экспериментально определенные ТАТА-боксы. Мы обнаружили SNP-маркер дефицита лептина, rs201381696, который был клинически связан с индуцированным диетой ожирением и стеатозом (Zhang et al., 2015), и два предполагаемых SNP, rs200487063 и rs34104384, для избытка лептина, которые коррелируют с индуцированным ожирением гипертонией, что было определено ранее (Tümer et al., 2007; Beltowski, 2008). Экспериментально мы показали, что сродство (K_D) ТСБ к промотору референсного гена равнялось 80 нМ. В случае аллеля -35G (rs201381696), сродство было ниже почти в 3 раза, $K_D = 230$ нМ. Оно надежно ($Z = 17$, $\alpha < 10^{-7}$) указывает на дефицит лептина у носителей этого SNP. Величина сродства ТСБ в случае аллелей -38A (rs200487063) и -30T (rs34104384), была выше на 50% и 11 %, соответственно. В последнем случае разница находится на границе достоверности.

Экспериментально охарактеризован не ассоциированный с каким-либо нарушением здоровья человека SNP rs367781716, -37T>C, промотора к альтернативному старту транскрипции в позиции chr17:66985252 (минус-цепь) гена *ABCA9*, который был детектирован экспериментально (Piehler et al., 2002), но не был до сих пор документирован в базе данных GENECODE v.19 (Harrow et al., 2012). Полученный с помощью Web-сервиса прогноз относительного изменения в 2,4 раза кажущейся равновесной константы диссоциации, K_D , комплекса ТСБ с минорным аллелем -37C промотора гена *ABCA9* приведен в Табл. 2. Данные независимых опытов *in vitro* и *in vivo* (Pugh, 2000; Stewart, Stargell, 2001; Mogno et al., 2010) указывают на то, что такому снижению сродства ТСБ/ТАТА соответствует недостаточность у носителя аллеля -37C, продукта *ABCA9* - транспортера АТФ-связывающей кассеты А9. В свою очередь, этот *in silico* прогноз риска недостаточности экспрессии транспортера, предположительно участвующего в липидном гомеостазе, в случае аллеля -37C позволяет предположить rs367781716 в качестве кандидатного SNP-маркера для нарушений липидного обмена.

Мы экспериментально охарактеризовали SNP rs183433761, -41T>C, промотора к альтернативному старту транскрипции в позиции chr2:162152278 гена *GCG*. Полученный в результате выполнения проекта «1000 Genomes project» полиморфизм был охарактеризован *in silico* с помощью того же Web-сервиса, и показал на уменьшение сродства ($\delta = -0.61$) между ТСБ и ТАТА-боксом при его наличии. Относительное изменение кажущейся равновесной константы диссоциации K_D (K_D возросла с 8 до 18 нМ) комплекса ТСБ с минорным аллелем

-41С промотора гена *GCG* в 2,25 раза – также свидетельствует о снижении сродства ($\delta = -0.81$). Такое изменение предположительно может внести вклад в снижение экспрессии глюкогона и, как следствие, - привести к нерегулируемому и не постоянному уровню концентрации глюкозы в крови.

Мутация -34Т>G rs28399433 в ТАТА-боксе гена *CYP2A6* из системы цитохрома P450, привела к существенному снижению сродства ТСБ/ТАТА в 4 раза, что было весьма ожидаемо исходя из прогнозов ($\delta = -1.5$). $\delta = -1.38$ указывает на существенное ухудшение сродства. Так как рядом с исследуемой заменой -34Т>G расположена уже описанная мутация (-48 Т>G), приводящая к снижению сродства ТСБ/ТАТА в 4.7 раза, то можно предполагать, что исследуемая замена -34Т>G будет также вносить вклад в разрушение ТАТА-бокса гена главной никотиноксидазы и ассоциироваться с повышенным риском возникновения воспалительных заболеваний и рака легких.

Замена -24Т>G rs1800202 в гене *TPI* приводит к снижению сродства в 7 раз (K_D возросла с 3 до 21 нМ), при прогнозируемом $\delta = -1.03$ против экспериментального $\delta = -1.94$. Так как фермент триозофосфатизомеразы относится к «генам домашнего хозяйства» и участвует в гликолитическом пути, то логично ожидать, что нарушение метаболизма триозофосфатов может привести к избыточной продукции метаболитов гликолиза. Нарушения в ТАТА-боксе гена триозофосфатизомеразы приводят к снижению уровня мРНК и появлению некоторых болезней (Watanabe et al., 1996).

Анализируя прогнозируемые *in silico* значения сродства, K_D , ТСБ/ОДН (Табл. 2) и K_D , полученные экспериментально, можно видеть, что экспериментальная оценка величины K_D совпала с прогнозом *in silico* в пределах точности используемых расчетов и измерений: δ , характеризующая изменение значения $-\ln K_D$ комплексов ТСБ с референсным и минорным аллелями, равна значимым величинам по прогнозу и в результате экспериментальной верификации. Это означает, что с использованием двух тест-систем, *in silico* и *in vitro*, было установлено, что неаннотированные ранее рассматриваемые SNP достоверно нарушают связывание ТСБ с промоторами генов человека на одном из первых этапов инициации транскрипции этих генов (Ponomarenko et al., 2013a, b). Поэтому исследуемые SNP предлагаются в качестве потенциальных кандидатных маркеров для отслеживания нарушений липидного обмена. Это является прогнозом данной работы, который может быть верифицирован по биомедицинским стандартам и протоколам.

Определение в режиме реального времени кинетических характеристик комплексов ТСБ с ТАТА-содержащими олигонуклеотидами, методами остановленной струи и FRET

Используя рекомбинантный ТСБ человека и флуоресцентно меченные по 5-концам FRET-парой флуорофоров Су3/Су5-дуплексы, идентичные ТАТА-боксу

промотора гена триозофосфатизомеразы (*TPI*) здорового человека и ТАТА-боксу, содержащему SNP -24T>G, методом «остановленной струи» в режиме реального времени исследован механизм процесса образования комплекса ТСБ/ТАТА. Впервые показано, что образование комплексов ТСБ с нормальным ТАТА-боксом происходит в 5.5 раза быстрее, а диссоциация в 31 раз медленнее, чем в случае ТАТА-бокса, содержащего SNP. Из полученных результатов следует, что ТСБ при образовании комплекса одновременно связывает и изгибает ДНК (Arkova et al., 2014).

Кинетика связывания ДНК-дуплексов с ТСБ в режиме реального времени, представленная на рис. 2 и 3, свидетельствует о том, что образование комплекса ТСБ/ТАТА приводит к росту интенсивности флуоресценции Су5. Рост интенсивности FRET-сигнала обусловлен изгибанием ДНК-дуплекса в комплексе с ТСБ, за счет чего происходит сближение остатков флуорофоров Су3 и Су5. Анализ кинетических кривых ДНК-дуплексов показал, что изгибание дуплекса, содержащего ТАТА-боксы дикого типа, происходит при меньших концентрациях ТСБ, чем в случае аллеля «G» ТАТА-боксов (рис. 2 и 3). На основании полученных данных нами предложен кинетический механизм связывания ТСБ с нормальным ТАТА-боксом гена *TPI* и с ТАТА-боксом, содержащим мутацию, описываемый одностадийной схемой: **ТСБ + ТАТА \rightleftharpoons ТСБ/ТАТА**.

Константы скоростей прямой и обратной реакции представлены в Таблице 3. Видно, что образование комплексов ТСБ с ТАТА-боксом дикого типа происходит в 5.5 раза быстрее ($1.1 \times 10^6 \text{ M}^{-1} \text{ c}^{-1}$), чем с аллелем G ($0.2 \times 10^6 \text{ M}^{-1} \text{ c}^{-1}$), а диссоциация комплексов ТСБ/ТАТА – в 31 раз медленнее ($2.8 \times 10^{-3} \text{ c}^{-1}$ у дикого типа и $8.9 \times 10^{-2} \text{ c}^{-1}$ в случае аллеля «G»). Необходимо отметить, что подобное различие в константах скорости образования и распада комплекса ТСБ/ТАТА приводит к отличию значений равновесных констант диссоциации в 150 раз ($2.7 \times 10^{-9} \text{ M}$ в норме и $0.4 \times 10^{-6} \text{ M}$ при наличии мутации). Различия в значениях

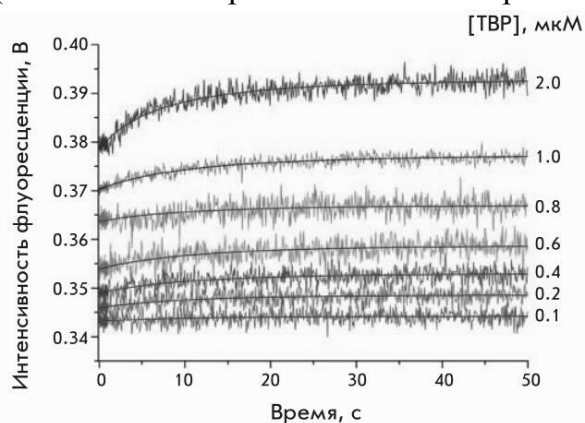


Рис. 2. Кинетика связывания и изгибания ДНК-дуплекса, идентичного последовательности ТАТА-боксов гена *TPI* дикого типа

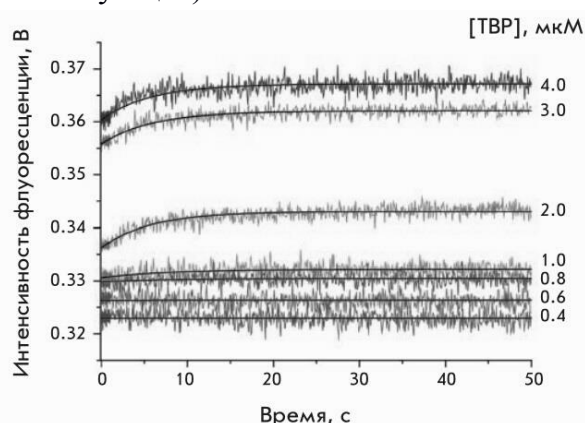


Рис. 3. Кинетика связывания и изгибания ДНК-дуплекса, идентичного последовательности содержащего SNP ТАТА-боксов промотора гена *TPI*

равновесных констант диссоциации (K_D) между нормальным и SNP-содержащим ТАТА-боксом говорит о резком снижении аффинности ТСБ к олигонуклеотидам с измененным ТАТА-боксом.

Полученные данные свидетельствуют о том, что возникшая в результате замены в ТАТА-боксе G/C-пара делает структуру ДНК более жесткой, что затрудняет связывание ТАТА-бокса с ТСБ и образование функционального комплекса, имеющего оптимальную конформацию. Из этого следует, что *in vivo* ген триозофосфатизомеразы, содержащий замену -24T>G в ТАТА-боксе, транскрибируется и экспрессируется менее эффективно. Эти результаты подтверждены клинически (Watanabe et al., 1996).

Сравнение полученных нами и опубликованных данных (Watanabe et al., 1996; Humphries et al., 1999) показывает, что уменьшение в 150 раз сродства ТСБ к SNP-содержащему ТАТА-боксу промотора *TPI* приводит к повышению риска развития ряда заболеваний, связанных с недостатком триозофосфатизомеразы. Несмотря на то, что сродство ТСБ к SNP-содержащему ТАТА-боксу промотора гена *TPI* снижено в 150 раз, активность *TPI* в эритроцитах одних пациентов падает до 3–10% от нормы (Chang et al., 1993), а у некоторых гетерозиготных носителей этого полиморфного аллеля наблюдается умеренное (26–50% от нормы) снижение активности *TPI* (Watanabe et al., 1996).

Следует отметить, что впервые путем регистрации в режиме реального времени взаимодействия ТСБ человека с флуоресцентно меченными $Cy3$ и $Cy5$ ТАТА-содержащими дуплексами показано, что ТСБ быстро связывает и одновременно изгибает ДНК ТАТА-бокса гена *TPI*.

Таблица 3. Определение в режиме реального времени кинетических характеристик комплексов ТСБ/ТАТА промотора гена *TPI* методами FRET и остановленной струи.

Константа	ТАТА-бокс, аллель Т ОНП -24T>G	ТАТА-бокс, аллель G ОНП-24T>G
$k_a, M^{-1}s^{-1}$	$(1.1 \pm 0.1) * 10^0$	$(0.2 \pm 0.1) * 10^0$
k_d, s^{-1}	$(2.8 \pm 0.1) * 10^{-3}$	$(8.9 \pm 1.2) * 10^{-2}$

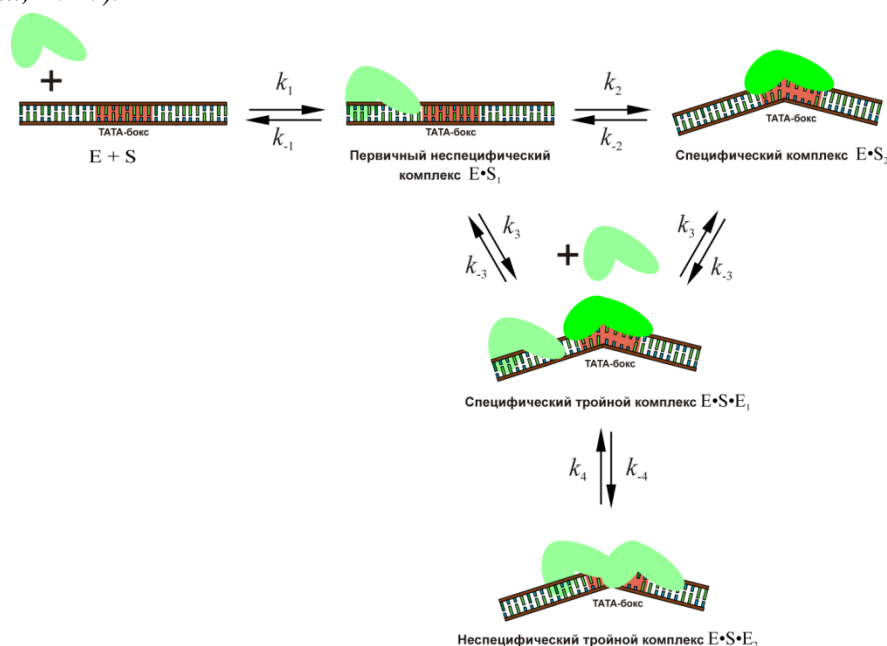
Полученные нами результаты на реальных промоторах генов человека с полноразмерным ТСБ улучшили представление о взаимодействии ТСБ/ТАТА, ключевом в инициации и регуляции транскрипции и синтеза белков в клетках эукариот.

Для более тонкого наблюдения различий в структуре ДНК при взаимодействии с ТСБ в режиме реального времени мы использовали пару флуорофоров TAMRA/FAM с более высоким квантовым выходом и, соответственно, чувствительностью детекции, чем у $Cy3/Cy5$.

Нами впервые в режиме реального времени (в первые миллисекунды взаимодействия) определены конформационные переходы ТАТА-содержащих

ДНК-дуплексов, идентичных анцестральным и минорным аллелям реальных генов *LEP* и *TPI* человека, при образовании комплексов с ТСБ. Полученные данные свидетельствуют о том, что формирование специфического комплекса ТСБ/ТАТА происходит через образование промежуточных комплексов. При повышении концентрации ТСБ и длины дуплексов наблюдается увеличение гетерогенности конформационных переходов ДНК («изгиб — исходное состояние»). На рис. 4 представлена схема взаимодействия ТСБ с ДНК-дуплексами, предложенная на основании полученных результатов.

Рисунок 4. Схема взаимодействия ТСБ с ДНК-дуплексами, содержащими ТАТА-элемент (E – ТСБ; S – ОДН, идентичный ТАТА-боксу в норме и при полиморфизме) (Arkova et al., 2016).



Влияние SNP ТАТА-боксов *LEP* и *CYP2A6* на активность *LUC*

Изучение влияния анцестральных и минорных аллелей генов *LEP* и *CYP2A6*, на транскрипцию репортерного гена *LUC*, показало (Рис. 5), что люминесценция как для *LEP* -35A>G, так и для *CYP2A6* -34T>G значительно ниже, чем люминесценция для WT (дикого) промотора, откуда можно сделать вывод, что наличие таких полиморфизмов ухудшает связывание ТСБ/ТАТА и снижает последующую экспрессию гена *LUC*. Полиморфизм 38G>A rs200487063 «улучшает» ТАТА-бокс, увеличивая сродство ТСБ к нему, что ведет к усилению экспрессии репортерного гена *LUC*. Данное заключение совпадает с предсказанным *in silico* эффектом, а также с экспериментально полученными данными методами EMSA и FRET.

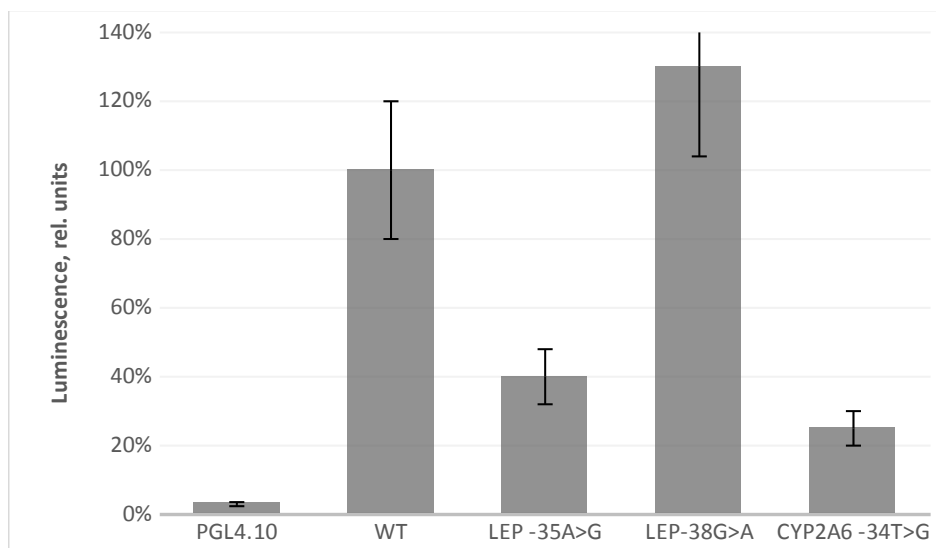


Рисунок 5. Культурально-клеточная верификация некоторых маркерных SNP генов *LEP* (-35A>G rs201381696 и -38G>A rs200487063) и *CYP2A6* (-34T>G rs28399433) на клеточных линиях человека MCF-7 и HCT-116, трансфецированных pGL4.10 вектором, содержащим репортерный ген *LUC* (pGL4.10 – контроль для WT в виде вставки беспромоторного вектора pGL4.10; WT- TATA-бокс дикого типа, контрольный для остальных, мутантных TATA-боксов).

Заключение

Проведенное исследование позволяет говорить о возможности эффективного использования разработанного Web-сервиса SNP_TATA_Comparator для выявления и доклинической проверки кандидатных SNP-маркеров TATA-боксов в предиктивно-превентивной персонализированной медицине.

Как показала экспериментальная верификация с использованием различных методов, предсказанные *in silico* с использованием разработанного Web сервиса SNP_TATA_Comparator, SNP TATA-боксов действительно обладают потенциалом функциональной значимости. Эксперименты по определению влияния SNP на сродство ТСБ/ТАТА, проведенные в равновесных и неравновесных условиях и в режиме реального времени с использованием методов остановленной струи и FRET, показывают хорошую корреляцию с прогнозами *in silico*. Результаты, полученные в режиме реального времени, показали, что взаимодействие ТСБ с ТАТА-содержащими ОДН зависит от длины и концентрации ОДН, и может происходить как по прямому, так и разветвленному пути. Как показали эксперименты *in vivo* на культуре клеток аденокарциномы кишечника и молочной железы, влияние выявленных однонуклеотидных замен на экспрессию репортерного гена *LUC* статистически значимо коррелирует с влиянием на сродство ТСБ/ТАТА.

Выводы

1. Проведенное экспериментальное исследование показало эффективность использования разработанного в ИЦиГ Web сервиса SNP-TATA-Comparator для предсказания новых потенциально функционально значимых мутаций ТАТА-боксов промоторов генов человека.

2. Экспериментальное изучение связывания ТАТА-связывающего белка с олигонуклеотидами, идентичными ТАТА-боксам референсных и минорных аллелей ряда генов липидного метаболизма показало, что минимальное увеличение сродства, на 14%, оказывает мутация -30A>T (rs34104384) ТАТА-боксов промотора гена *LEP*, максимальное, в 2 раза, мутация -38G>A (rs200487063) ТАТА-боксов этого гена. Максимальное уменьшение сродства в 7 раз оказывает мутация -24T>G (rs1800202) ТАТА-боксов гена *TPI*. Экспериментальные данные коррелируют с прогнозами, сделанными с помощью разработанного Web сервиса ($r=0.89$).

3. Изменение наследственной информации в результате мутаций в ТАТА-боксов реализуется через изменение скоростей образования и распада комплексов ТАТА-связывающего белка с ТАТА-боксов. Скорости образования изменяются от 0.2 до 14 раз. Мутация -35A>G (rs201381696) промотора гена *LEP*, нарушающая последовательность ТАТА-боксов, приводит к максимальному увеличению скорости распада комплексов - в 7 раз.

4. Изучение в режиме реального времени образования комплексов ТАТА-связывающего белка с мечеными FRET-парой флуорофоров TAMRA и FAM олигонуклеотидами, идентичными ТАТА-боксов промоторов референсных генов *LEP* и *TPI* и их минорным аллелям, показало, что образование комплексов может происходить как по прямому пути, так и по разветвленному, через промежуточные стадии, вероятность возникновения которых возрастает с увеличением длины олигонуклеотидов и концентрации ТАТА-связывающего белка.

5. На культуре клеток человека HCT116 показана корреляция влияния ($r=0.89$) мутаций ТАТА-боксов на сродство ТАТА-связывающего белка к ТАТА-боксов и на экспрессию репортерного гена *LUC* относительно «дикого типа» (WT): мутация -38G>A (rs200487063) ТАТА-боксов промотора гена *LEP*, повышающая сродство в 2 раза, в линии клеток HCT116 повышала экспрессию репортерного гена *LUC* в 1.5 - 1.8 раза, мутация -35A>G (rs201381596), снижающая сродство в 3 раза, снижала экспрессию в среднем в 2.5 - 3 раза; мутация -34T>G (rs28399433) ТАТА-боксов промотора гена *CYP2A6*, снижающая сродство в 4 раза, также в 4 раза снижала экспрессию репортерного гена *LUC*.

Список основных публикаций по теме диссертации

Статьи в журналах из списка ВАК:

1. 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. V. 9. N. 2. P.276-306.. DOI: 10.2741/S488. IF = 3.970.
2. **Arkova O.**, Kuznetsov N., Fedorova O., Savinkova L. A real-time study of the interaction of TBP with a TATA box-containing duplex identical to an ancestral or minor allele of human gene LEP or TPI. // Journal of Biomolecular Structure and Dynamics, 2016; 25:1-20. IF = 2.300.
 3. Ponomarenko M.P., **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. IF = 5.695.
 4. Chadaeva I.V., Ponomarenko M.P., Rasskazov D.A., Sharypova E.B., Kashina E.V., Matveeva M.Y., 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): S3, IF = 3.867.
 5. **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. IF = 3.867.
 6. 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. Suslov V, Savinkova L, Kolchanov N. // Biomed Res Int. 2015; 5:359835. IF = 1.579.
 7. **Arkova O.V.**, Kuznetsov N.A., Fedorova O.S., Kolchanov N.A., Savinkova L.K. Real Time Interaction between TBP and the TATA box of the Human Triosephosphate Isomerase Gene promoter in the Norm and Pathology. // Acta Naturae. 2014;6(2):36-40, IF = 1.777.
 8. Турнаев И.И., Рассказов Д.А., **Аркова О.В.**, Пономаренко М.П., Пономаренко П.М., Савинкова Л.К., Колчанов Н.А. Гипотетические SNP-маркеры, значимо изменяющие оценки сродства ТАТА-связывающего белка к промоторам онкогенов VEGFA, ERBB2, IGF1R, FLT1 KDR, MET – мишеней для химиотерапии. // Мол. биол. 2016. 50(1), 161-73, РИНЦ =0.786.
 9. **Аркова О.В.**, Драчкова И.А., Аршинова Т.В., Рассказов Д.А., Суслов В.В., Пономаренко П.М., Пономаренко М.П., Колчанов Н.А., Савинква Л.К. Прогноз и верификация влияния SNP rs367781716 на взаимодействие ТАТА-связывающего белка с промотором гена ABCA9 человека. // Вавиловский журнал генетики и селекции. 2015, 19(6), 675-681.
 10. **Аркова О.В.**, Пономаренко М.П., Аршинова Т.В., Савинкова Л.К. Механизм влияния полиморфизмов кор-промоторов генов пищевого поведения человека на взаимодействие с ТАТА-связывающим белком. Мед. генетика, 2015, т. 14, №2, с. 10-11.