

БИОИНФОРМАЦИОННЫЙ ПОИСК ВЗАИМОСВЯЗАННЫХ СЦЕНАРИЕВ В ОРГАНИЗАЦИИ КОДИРУЮЩЕЙ И НЕКОДИРУЮЩЕЙ БЕЛОК ДНК В ОТВЕТ НА ОКИСЛИТЕЛЬНЫЙ СТРЕСС

Шкурат Т.П., Адигеев М.Г., Пономарева Н.С., Бутенко А.И., Ломтева С.В., Панич А.Е.

Южный федеральный университет, Ростов-на-Дону, Россия (344006, г. Ростов-на-Дону, ул. Большая Садовая, 105/42), e-mail: tshkurat@yandex.ru

Данное исследование посвящено биоинформационному поиску новых регуляторных структур в некодирующей ДНК, расположенной вокруг паттернов генов, существенно изменивших уровень экспрессии в ответ на окислительный стресс. Выдвинута гипотеза, что все гены, повысившие экспрессию в ответ на окислительный стресс, могут иметь одинаковые мотивы в некодирующей ДНК. Для поиска мотивов была создана интегрированная коллекция баз данных транскрипционных сайтов связывания – JASPAR, TRANSFAC, Hocomoco TF Homo sapiens, Uniprobe TF Mus musculus. Исследованы два типа регуляторных областей: промоторные области и последовательности с захватом потенциальных цис-регуляторных модулей. В регуляторных областях всех генов, повысивших экспрессию в ответ на окислительный стресс, в отличие от генов, не изменивших уровень экспрессии, выявлены транскрипционные факторы семейств SOX(1-30) и HX (A, B, C, D).

Ключевые слова: некодирующая ДНК, геномика, окислительный стресс, ДНК-микрочипы, межгенная ДНК, сайт связывания транскрипционного фактора, мотив ДНК.

SEARCH FOR RELATED SCENARIOS IN THE ORGANIZATION OF THE CODING AND NON-CODING DNA IN RESPONSE TO OXIDATIVE STRESS. BIOINFORMATICS APPROACH

Shkurat T.P., Adigeev M.G., Ponomareva N.S., Butenko A.I., Lomteva S.V., Panich A.E.

Southern Federal University, Rostov-on-Don, Russia (344006, Rostov-on-Don, Bolshaya Sadovays st. 105/42) , e-mail: tshkurat@yandex.ru

This study focuses on bioinformatics search for new regulatory structures in the non-coding DNA, located around the patterns of gene expression levels changed significantly in response to oxidative stress. Hypothesized that all of the genes increase the expression in response to oxidative stress may have the same motifs in non-coding DNA. To search for motifs created an integrated collection database of transcription binding sites - JASPAR, TRANSFAC, Hocomoco TF Homo sapiens, Uniprobe TF Mus musculus. Two types of regulatory regions: the promoter region and the sequence with the capture of potential cis-regulatory modules. In the regulatory regions of genes increase the expression in response to oxidative stress, in contrast to the gene expression level did not change, families of transcription factors identified SOX (1-30) and HX (A, B, C, D).

Key words: gene expression, DNA microarrays, noncoding DNA, oxidative stress, transcription factor, sites of transcription factor binding, DNA motif.

Введение

Известно, что более 95% нуклеотидных последовательностей в геномах млекопитающих не кодируют белки. В последние годы все больше появляется работ об участии некодирующей белок ДНК в различных регуляторных процессах [4]. Целостное понимание всех функций генома еще далеко от завершения, особенно важно выяснение роли некодирующих белок и РНК последовательностей ДНК, механизмов регуляции альтернативного сплайсинга. Систематический анализ транскрипционного профиля, данные о механизмах регуляции и поиски взаимосвязанных сценариев взаимодействия кодирующей

и некодирующей белок ДНК являются очень важным фундаментом для понимания системной биологии и молекулярного патогенеза различных заболеваний человека.

Для моделирования неспецифического патогенеза нами был выбран окислительный стресс. Ранее нами было показано, что однократное воздействие повышенного давления кислорода на новорожденных крыс изменяет внутриклеточный метаболизм и формирует качественно новое соотношение про- и антиоксидантных систем в организме, которое сохраняется длительное время и наблюдается у потомков первого поколения [5; 6]. Клетки новорожденных более приспособляемы к изменениям физиологических условий, чем клетки зрелого организма. Предполагается, что обработка новорожденных животных низкой дозой гипербарической оксигенации облегчит формирование широкой кислородчувствительной нейронной сети, которая, в свою очередь, ведет к увеличению сопротивления животных окислительному стрессу в дальнейшем онтогенезе [1]. Полногеномный анализ транскрибирующихся последовательностей ткани мозга этих животных показал, что в ранние генетические механизмы преадаптации к окислительному стрессу вовлечены гены, которые образуют, по меньшей мере, один экспрессионный генетический паттерн и регулируют 6 ключевых процессов: окислительный обмен веществ; синаптическую передачу; внутриклеточный транспорт; регулирование апоптоза и пролиферации; регулирование мембранной проницаемости для ионов, формирование мембранных потенциалов и модификацию межклеточных контактов [2].

Экспрессия гена зависит от двух основных типов сигналов: один с участием транскрипционных факторов, который инициирует транскрипцию генов, и другой, который регулирует перевод зарождающейся мРНК. Эти посттранскрипционные события играют важную и еще не полностью понятную роль в регуляции экспрессии генов и клеточного поведения.

Цель исследования

Биоинформационный поиск новых регуляторных структур в промоторных и цис-регуляторных областях некодирующей белок ДНК, участвующих в механизмах экспрессии генов в ответ на окислительный стресс.

Материал и методы исследования

Объектом исследования служили белые аутбредные крысы *Rattus norvegicus*, которые подвергались воздействию повышенного давления кислорода в режиме 0,2 МПа – 1 час. Животных декапитировали через три часа после окончания воздействия агента, исследовали экспрессию генов в лобной доле коры головного мозга. Результаты полногеномных исследований транскриптом описаны нами ранее [2].

Для биоинформационного анализа были сформированы две группы генов. Первая группа – гены с повышенной экспрессией на окислительный стресс. Вторая группа – гены, не изменившие уровень экспрессии в ответ на окислительный стресс. В первую группу вошли следующие гены: NM_017138:Lamr1, NM_053440:Stmn2, NM_057207:Sv2b, NM_053339:Acox3, Scn7a, Trpc3, Nid2:NM_213627, Herc1, Ssc1:BC085795, Golph2:NM_023977, Actr1a, Crebzf,Pdk4:NM_053551, Mrpl3, Api5, Zfx1b, Snrpb:BC083694, Snrpb.

Во вторую – все гены локализованные на 20 хромосоме *Rattus norvegicus*. Исследовали регуляторные области 264 генов, доступные в KnownGene (за вычетом первого и последнего генов).

В обеих группах были исследованы два типа регуляторных областей. Промоторные области – длина 1000, 800 п.н. до старта транскрипции (СТ), и 200 п.н. после окончания транскрипции и последовательности с захватом потенциальных цис-регуляторных модулей (ЦРМ) – длина 5000, 4000 п.н. до СТ, 1000 п.н. после окончания транскрипции. Биоинформационный поиск был направлен на поиск «мотивов» ДНК, локализованных рядом с генами первой группы. Под «мотивом» распознавания здесь и далее понимается способ описания набора сходных олигонуклеотидов, специфически распознаваемых и связываемых конкретным регуляторным белком.

Для поиска мотивов нами была создана интегрированная коллекция транскрипционных сайтов (ТС) связывания, на основе баз данных JASPAR, TRANSFAC, UCSC ENCODE, Hocomoco TF Homo sapiens, Uniprobe TF Mus musculus. Интегрированная коллекция Uniprobe включала 272 мотива для транскрипционных факторов Mus musculus. Коллекция мотивов hocomoco TF Homo sapiens, полученная методом интеграции данных различных источников, содержала 332 мотива для 321 транскрипционных факторов.

Для биоинформационного анализа был выбран релиз генома *Rattus norvegicus* rn4 (Baylor 3.4/rn4) [<http://genome.ucsc.edu/cgi-bin/hgGateway>].

В качестве модели области связывания использовалась концепция гомотипических кластеров сайтов связывания транскрипционных факторов (ССТФ), представляющих собой группы сайтов связывания транскрипционных факторов. Для оценки статистической значимости кластера сайтов связывания использовали модель г-скан с фиксированным порогом на вхождение мотива. Статистическая значимость г-скана оценивалась как вероятность обнаружить не менее заданного числа ССТФ на отрезке заданной длины, при условии, что хотя бы 1 ССТФ уже обнаружен (т.к. в принятом определении кластер содержит как минимум один ССТФ).

Для достоверной оценки присутствия гомотипического кластера конкретного мотива в первой группе (гены с повышенной экспрессией к окислительному стрессу) использовали процедуру сравнения со второй группой (контролем). Для контрольной выборки генов были построены аналогичные наборы промоторов и цис-регуляторных модулей. Далее подсчитывали гомотипические кластеры мотивов в первой и второй группе и оценивали шанс p случайно выбрать в геноме такой ген, чтобы его регуляторная область содержала хотя бы один кластер (для промоторов) и кластер со значимостью не хуже 0.0005 для цис-регуляторных модулей.

Используя p с помощью биномиального распределения можно оценить P-value как значимость того факта, что в исследуемой выборке генов кластеры ССТФ содержат не менее k регуляторных областей, где k – реально обнаруженное число последовательностей с кластерами ССТФ (для промоторов) / не менее чем 0.0005-значимыми кластерами ССТФ для цис-регуляторных модулей.

Результаты исследования и их обсуждение

В таблице 1 представлен фрагмент оценки гомотипического кластера для группы генов, чувствительных к воздействию повышенного давления кислорода. В полном варианте таблицы в верхней строке были расположены 605 мотивов.

В результате биоинформационного анализа гомотипических кластеров мотивов выявлены следующие значимые транскрипционные факторы:

- транскрипционные факторы, значимые с P-value не хуже 0.05;
- среди них дополнительно выделены те транскрипционные факторы, для которых число генов со значимыми г-сканами составляет не менее половины исследуемой выборки (т.е. не менее 9);
- транскрипционные факторы, значимые с P-value не хуже 0.01.

Как видно из представленных результатов (таблица 2), достоверно чаще перед первой группой генов встречались транскрипционные факторы сайтов связывания белков семейства SOX. Семейство SOX включает более 30 транскрипционных факторов [7], для которых характерна гептаметрическая последовательность – (A/T) (A/T) CAA(A/T)G (рис.1).

Семейство SOX-белков – одно из самых важных в регуляции развития как позвоночных, так и беспозвоночных животных. Биологические функции этих белков в процессе эмбрионального и постэмбрионального развития были определены в различных тканях и клетках млекопитающих. Как правило, разделение на группы связано с особенностями функционирования в определенных тканях. Они инициируют программу дифференцировки и активируют тканеспецифическую экспрессию. Семейство белков SOX не имеет единой биологической функции, и среди их многочисленных ролей SOX-транскрипционные

факторы обладают свойствами индукции или подавления различных клеточных процессов. Кроме этого, показано, что SOX-белки принимают участие в процессах репарации.

HXA₃, heroxilin A3, что в отсутствие цитозольной глутатионпероксидазы в клетках из внутренне активной липоксигеназы из 12 (12 (S)-LOX) преобразуется 12S-гидропероксиэйкозатетраеновую кислоту (12 (S)-HPETE) на биологически активные эпоксидных heroxilin A3 (HXA3).

HLX экспрессируется во время эмбриогенеза в тканях мезодермального происхождения, особенно висцеральной мезенхимы, скелетных миобластов, склеротома и конечностей мезенхимы [3].

Gata 3 контролирует многие гены регуляции роста клеток, иммунной функции и ингибиции ади- погенеза.

SP100 фактор полипотентного белка S-100, который в основном встречается в нейронах головного мозга.

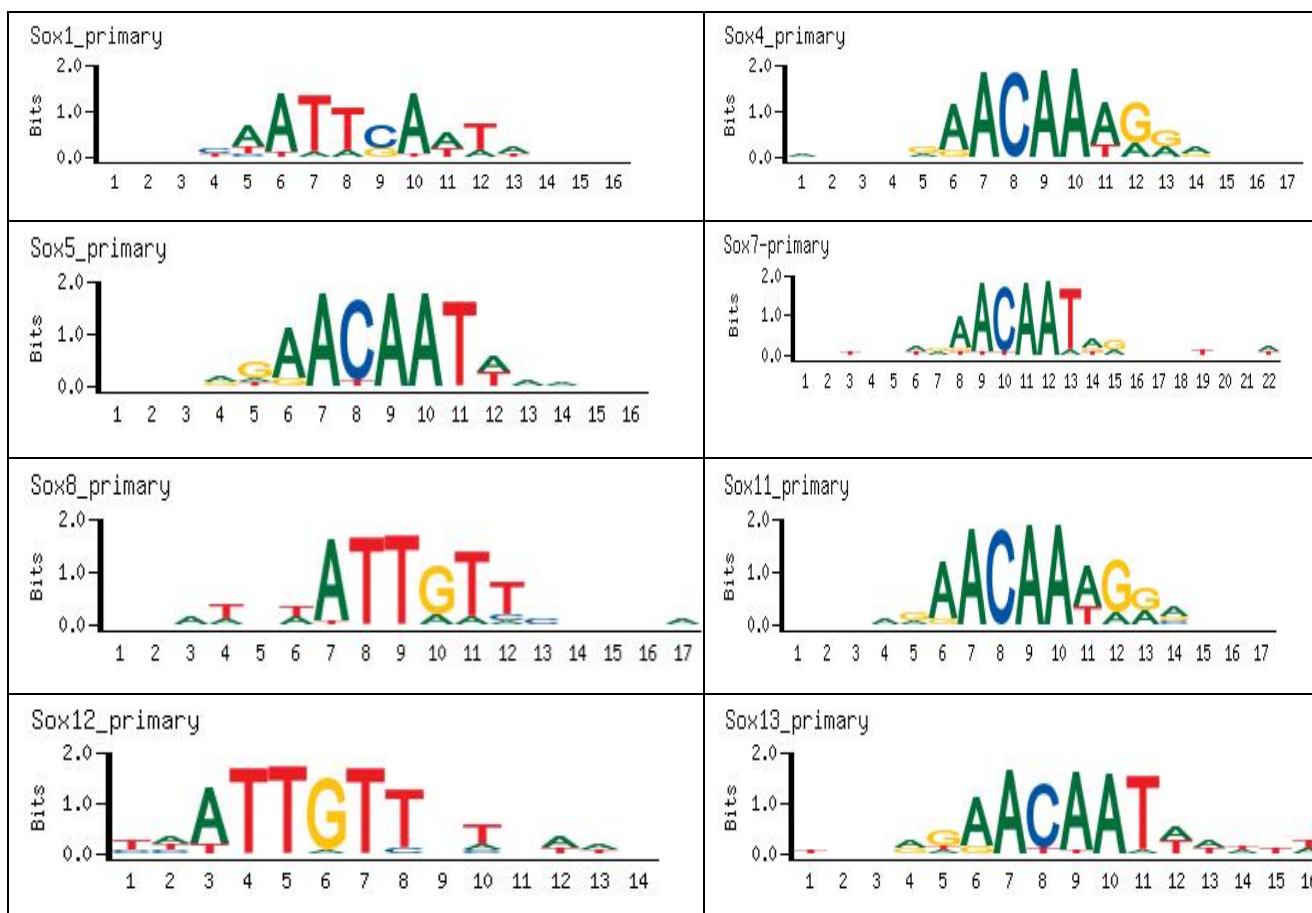


Рис. 1. Консенсусные последовательности транскрипционных факторов семейства SOX [(A/T) (A/T) CAA(A/T)G] по Guimont P,ea 2007.

Таблица 1 – Фрагмент результатов оценки гомотипического кластера мотивов связывания для группы promoter_hocomoco (s1000)

PWMID	SOX5_si	FOXO1_si	PAX6_f2	FOXA1_f1	ALX1_si
TF	SOX5 HUMAN	FOXO1 HUMAN	PAX6 HUMAN	FOXA1 HUMAN	ALX1 HUMAN
Lamr1	0,009	0,100	0,002	0,233	1,000
Stmn2	0,112	0,059	1,000	1,000	0,037
Sv2b	0,001	1,000	0,005	1,000	0,078
Acox3	1,000	0,118	1,000	0,354	1,000
Scn7a	0,126	0,000	0,002	0,000	0,111
Trpc3	1,000	0,002	1,000	0,112	1,000
Nid2	0,001	0,000	1,000	0,011	1,000
Herc1	0,234	0,005	0,127	0,037	1,000
Ssc1	1,000	1,000	1,000	1,000	1,000
Golph2	1,000	1,000	1,000	1,000	0,003
Actr1a	0,034	0,000	1,000	0,090	1,000
Crebzf	0,007	0,000	1,000	0,085	1,000
Pdk4	1,000	0,151	1,000	0,010	1,000
Mrpl3	0,001	0,000	1,000	0,000	1,000
Api5	1,000	0,157	1,000	1,000	0,058
Zfhx1b	0,099	0,005	1,000	0,187	0,067
Snrpb: BC083694	0,000	0,000	1,000	0,154	1,000
Snrpb	1,000	0,079	0,207	0,195	1,000
total_signN	11,000	15,000	5,000	13,000	6,000
average_signV	0,057	0,045	0,069	0,113	0,059
chance_p	0,186	0,443	0,080	0,436	0,121
signN_P-value	0,000	0,001	0,011	0,013	0,016

Значимые транскрипционные факторы для групп CRM_uniprobe (s5000), CRM_hocomoco (s5000), promoter_uniprobe (s1000) и promoter_hocomoco (s1000) представлены в таблице 2.

Таблица 2 – Достоверные отличия локализации транскрипционных факторов перед генами первой группы

Коллекция	P <0.05	P <0.05,г-скан >9	P <0.01
Promoter_Hocomoco (s1000)	PAX6, ALX1, PITX2, OTX2, CRX, CDC5L, PO3F2	FOXA1, CDX2	SOX5, FOXO1,EVI1, FOXO3
Promoter_Uniprobe (s1000)	GATA6, BARX2, HXB6, MEOX1, SH SOX8, HXD3, HXA6, DLX1, ARI5A, PITX2, HXA10, HXA7, DLX3,	TBP, TF7L1, LMX1A	DLX2, ARH2, DBX1

	HXA9, BARH1, SOX15, HXB7, HXC8, SOX13, HXA2, HDX, PDX1, SP100, SOX30, GATA3, PO6F1, HXA1		
CRM_Носомосо (s5000)	SOX9, ARI3A, NFYA, ATF6A, CDX1, ZEP1, SOX5, NFYB, SOX2	EVI1, MBD2, MEF2C, MEF2A	-
CRM_Uniprobe (s5000)	SOX15, SOX30, ZN187, HXD13, PROP1, SOX18, RX, SRY, SOX7, SOX14, HXD1, DLX4, PHX2B, PAX6, HXD10, HBP1, OTP	PO3F2, PO2F1, SOX12	SOX5, SOX13, SOX17, DBX1, HLX-

Таким образом, мы обнаружили, что при первичном воздействии гипербарической оксигенации 0,2 МПа-1ч на клетки мозга выявляется определенный набор транскрипционных факторов, связывающихся с исследуемыми генами и участками их промоторов.

Заключение

Множество сигналов поступает извне в клетку, и если бы для каждого из них клетка использовала полностью независимые системы, то никакого генома не хватило бы, чтобы зашифровать все необходимое множество продуктов. Поэтому клетка использует несколько путей, чтобы экономить информацию и рационально использовать свой геном. Один из путей заключается в том, чтобы на один продукт нагружать множество разных функций и использовать его в различных клеточных процессах. Другой подход заключается в том, чтобы множество разных процессов сводить в какой-то точке к одному, общему для всех, итоговому процессу, который выполняет роль завершающей стадии для всех них. Этот последний прием является интеграцией множества сигналов в один конечный результат, путь от разного к общему. Согласно современным представлениям, у эукариот наиболее значимым уровнем регуляции экспрессии генов является транскрипция, которая осуществляется с участием огромного разнообразия белков, называемых транскрипционными факторами. В основе функционирования ТФ лежит способность этих белков к специфическим взаимодействиям с ДНК промоторов, энхансеров и других регуляторных элементов генов, а также к белок-белковым взаимодействиям с другими транскрипционными факторами и белками эукариотической клетки. Комплексы транскрипционных факторов с различными со-регуляторами необходимы для активирования или подавления транскрипции генов путем модуляции активности промотора.

Полученные нами данные свидетельствуют в пользу специфически сильного, вероятно кооперативного взаимодействия факторов транскрипции как в промоторной

области, так и в цис-регуляторных модулях. Механизмы одновременной экспрессии генов, вызываемой в данном исследовании окислительным стрессом, в настоящее время полностью не изучены, но могут регулироваться последовательностями, лежащими в близлежащих областях некодирующей ДНК. Данное исследование позволяет сделать вывод, что участки, относящиеся к некодирующей ДНК, несут определенные функции в регуляции ген-генных взаимодействий, полное же определение всех функций ДНК является важной биологической задачей.

Список литературы

1. Гуськов Е.П. [и др.] Мутационные процессы у животных, преадаптированных к окислительному стрессу // Экологическая генетика. – 2009. – Т. VII. – № 1. – С. 40-47.
2. Шкурат Т.П. [и др.] Поиск новых механизмов преадаптации мозга животных к окислительному стрессу, с использованием технологии dna array // Современные проблемы науки и образования. – 2011. – № 5. – URL: www.science-education.ru/99-4888.
3. Allen J.D., Lints T., Jenkins N.A., Copeland N.G., Strasser A., Harvey R.P., and Adams J.A. 1991. Novel murine homeo box gene on chromosome 1 expressed in specific hematopoietic lineages and during embryogenesis. *Genes & Dev.* 5: 509-520.
4. Dunham I., Kundaje A. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. – *Nature* (2012), Vol. 489: p. 57-74.
5. Guskov E.P., Mashkina E.V., Belichenko N.I., Vardun T.V., Volosovtsova G.I., Pokudina I.O., Guskov G.E. and Shkurat T.P. Mutation processes in oxidative stress preadapted animals // *Russian Journal of Genetics: Applied Research* Volume 1, Number 2 (2011), 112-118.
6. Guskov E.P., Shkurat T.P., Shimanskaja E.I., Guskova S.I. Genetic effects of hyperbaric oxygen therapy // *Mutation Research/Genetic Toxicology* Volume 241, Issue 4, August 1990, Pages 341–347.
7. Guimont P, Grondin F, Dubois C: Sox9-dependent transcriptional regulation of the proprotein convertase furin. *American journal of physiology. Cell physiology* 2007, 293:C172-183

Работа выполнена при поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы», госконтракт №14.740.11.0006.

Рецензенты

Усатов А.В., д.б.н., зав. отделом Изменчивости генома Научно-исследовательского института биологии Южного федерального университета, г. Ростов-на-Дону.

Вардуни Т.В., д.п.н., профессор кафедры генетики Южного федерального университета, г.
Ростов-на-Дону.