

## ВЗАИМОСВЯЗЬ ТРАНСПОЗОНОВ С ТРАНСКРИПЦИОННЫМИ ФАКТОРАМИ В ЭВОЛЮЦИИ ЭУКАРИОТ

© 2019 г. Р. Н. Мустафин

Башкирский государственный университет, Уфа, Россия

E-mail: ruji79@mail.ru

Поступила в редакцию 25.11.2017 г.

После доработки 31.07.2018 г.

Принята к публикации 15.08.2018 г.

Транспозоны являются важными источниками сайтов связывания с транскрипционными факторами, специфическая активация которых характерна для эмбрионального развития животных. В эволюции эукариот одомашнивание генов мобильных элементов привело к возникновению множества новых белков, в том числе транскрипционных факторов, участвующих в управлении дифференцировкой клеток. Транспозоны обильно представлены в межгенной ДНК, интронах и 3' нетранслируемых областях, в том числе вблизи генов транскрипционных факторов, на которые они оказывают регуляторное воздействие. Это способствует взаиморегуляции генов путем активации транспозонов продуктами их экспрессии в последовательных делениях стволовых клеток и представляет собой динамическую биологическую кодировку видоспецифической регуляции онтогенеза. Реализация данной информации возможна благодаря тканеспецифической и стадийспецифической регуляции, опосредованной наличием транспозонов в определенных сайтах генома. В пользу данного предположения говорит важнейшая роль одомашненных генов транспозонов в управлении работой генома и ключевыми этапами развития.

*Ключевые слова:* дифференцировка, стволовые клетки, транскрипционные факторы, транспозоны, эволюция

**DOI:** 10.1134/S004445291901008X

### ВВЕДЕНИЕ

Транспозоны (TE – transposable elements) составляют значительную долю геномов большинства эукариот. В связи с накоплением большого количества данных о составе и расположении TE в геномах, появилось множество теорий, объясняющих их роль в эволюции и онтогенезе. В данном отношении сложились две противоположные концепции. Согласно одной из них, паразитическое поведение TE было выдвинуто как достаточное объяснение их содержания в геномах. Данные исследователи характеризовали TE как эгоистичские (selfish) элементы [1] и геномные паразиты, мало или вообще не влияющие на фенотип [2]. Другая группа исследователей выдвигала концепцию о влиянии TE на функционирование генома хозяина [3]. Они обосновывали свои представления доказательствами того, что TE способствуют геномным нововведениям и появлению регуляторных элементов для хозяина [4, 5]. Большую роль в исследовании роли TE в качестве источников регуляторных последовательностей сыграли отечественные исследователи В.А. Ратнер [6, 7] и Л.З. Кайданов [8], определившие TE как подвижные регуляторные каскады, которые направленно встраиваются в опреде-

ленные сайты генома, что приводит к адаптациям. К настоящему времени появляется все больше данных о том, что TE служат источниками сайтов связывания для транскрипционных факторов (ТФ) и изменяют регуляторные сети развития [9]. Объяснения роли TE в геномах представителями двух крайних концепций имеют право на существование, так каждая из них затрагивает различные характеристики TE. Например, при горизонтальном переносе Р-элементов в геном *Drosophila melanogaster* не было обнаружено эволюционно значимых изменений. Кроме того, идентифицировано множество инсерций TE, вызывающих наследственные болезни. С другой стороны, накоплено достаточно данных, доказывающих, что эукариоты в ходе эволюции используют как регуляторные последовательности TE, так и их гены для собственных нужд.

Физиологическое развитие многоклеточных эукариот характеризуется синергизмом во взаимоотношениях TE и их хозяев. В организмах хозяев существует множество белковых систем, подавляющих активность TE в зрелых дифференцированных клетках, за счет чего поддерживается стабильность функционирования генома. Например, каталитические белки АРОВЕС3А, АРОВЕС3В, АРОВЕС1,

ERCC, TREX1, RB1, HELLS, MEGP2 предотвращают экспрессию ТЕ [10]. Мощным репрессором активности LINE-1 является стабилизирующий белок SIRT6, который взаимодействует с 5'UTR L1 элемента и упаковывает его в транскрипционно репрессированный гетерохроматин [11]. Однако данные системы несовершенны, что может приводить к незапланированной патологической активации ТЕ в онтогенезе, приводя к ряду ассоциированных с возрастом патологий. Например, дисбаланс в управлении ТЕ, ведущий к геномной нестабильности, может привести к развитию злокачественных новообразований [12]. Предполагается, что истощение системы репрессии ТЕ с возрастом может служить причиной старения. С возрастом из областей L1 истощается содержание SIRT6 [11]. Кроме того, поддерживающее метилирование ДНК транспозонов имеет точность приблизительно 99% – с возрастом из-за возможных ошибок маркеры репрессированного хроматина могут теряться, что вызывает дерепрессию ТЕ и лавинообразному росту вызванных этим транспозиций, вызывающих геномную нестабильность и прогрессирующее старение [13]. В связи с этим исследование взаимосвязей ТЕ с геномом хозяина имеет большое значение в современной генетике.

Недифференцированные клетки являются плюрипотентными и могут активировать различные белок-кодирующие гены. По мере прогрессирования клеточной дифференцировки экспрессируются гены, специфичные для данного типа клеток, а также репрессируются неспецифичные гены – в результате теряется плюрипотентность и достигается ограниченное дифференцированное состояние. На ингибирование клеточной пластичности оказывают воздействие эпигенетические факторы и ТФ [14]. В регуляции экспрессии данных факторов ключевую роль играют ТЕ, формирующие динамичные регуляторные сети [15]. Для каждого вида, характеризующегося специфичным составом генома, кариотипа, состава и расположения ТЕ относительно белок-кодирующих генов, характерна конкретная особая регуляторная сеть. В эволюции ТЕ являются важными источниками возникновения сайтов связывания с транскрипционными факторами (TFBS – transcription factors binding sites) [16–20]. Новые белки в эволюции могут формироваться путем одомашнивания генов ТЕ геномом хозяина [5, 21–26]. ТЕ участвуют в создании ретрокопий генов, некоторые из которых активно экспрессируются [27, 28]. При экзонизации транспозонов, расположенных в интронах [29–31] и 3'UTR [32, 33], могут образовываться новые белки. Кроме того, ТЕ являются важными источниками siРНК [34], микроРНК [35–40] и длинных некодирующих РНК (нкРНК) [41, 42]. Таким образом, ТЕ могут играть роль в регуляции экспрессии генов при клеточной дифференцировке в связи со способностью ТЕ к транспозиции под действием раз-

личных сигналов, а также образованием нкРНК, оказывающих влияние на транскрипцию. Экспансии ТЕ в геномах ведут к существенным перестройкам регуляторных сетей и комбинаторной организации транскрипционного управления генами [43]. Регуляторные последовательности, происходящие из ТЕ, особенно интересны с точки зрения эволюции из-за их потенциала способствовать дивергенции экспрессии генов между видами. В эволюции ТЕ представляют собой часть геномов, которая наиболее быстро меняется. В связи с этим каждый вид характеризуется специфическим составом, расположением и количеством копий различных ТЕ. Например, при сравнении геномов человека и мыши стало очевидно, что 99% белок-кодирующих генов являются гомологами, а 80% генов полностью идентичны. В то же время только 13% ТЕ мыши обнаружены в геноме человека, а 48% ТЕ человека выявляются в геноме мыши [44].

#### ЭВОЛЮЦИЯ ИСТОЧНИКОВ САЙТОВ СВЯЗЫВАНИЯ С ФАКТОРАМИ ТРАНСКРИПЦИИ

В формировании геномов эукариот важную роль сыграло появление новых функций для ТЕ за счет наличия или появления под влиянием мутаций в их составе регуляторных последовательностей, что позволяет ТЕ ремоделировать генные сети [45]. Предполагается, что экспансия ТЕ является основным фактором генной регуляции. Выявлено множество инсерций ТЕ, которые были использованы для нужд генома хозяина в связи с предоставлением TFBS. Данные TFBS, произошедшие из ТЕ, служат в качестве энхансеров и промоторов [17]. Например, в геноме человека обнаружено 280000 потенциальных регуляторных элементов, возникших в результате инсерций SINE, LINE, LTR-содержащих и ДНК-транспозонов. У человека не менее 20% регуляторных последовательностей генов, сходных с другими видами, были кооптированы из ТЕ [16]. В недавних исследованиях в геноме человека было выявлено 794 972 TFBS, произошедших из эндогенных ретровирусов HERV. Анализ кластеризации показал, что HERV/LTR могут быть сгруппированы в соответствии с шаблонами связывания с ТФ. HERV/LTR группы ограничены плюрипотентными ТФ (например, SOX2, POU5F1, NANOG), ТФ эмбриональной эндодермы/мезодермы (например, GATA4/6, SOX17, FOXA1/2) и гематопоэтические ТФ (например, SPI1(PU1), GATA1/2, TAL1) [20].

Ретровирусные LTR содержат множество функциональных регуляторных элементов, необходимых для инициации и контроля транскрипции и могут существенно изменить экспрессию проксимальных генов. Кроме того, так как многие отдельные классы ретровирусов различаются по регуляторным последовательностям, содержащимся внут-

ри их LTR, они оказались весьма универсальными при переработке разнообразных транскрипционных программ. На протяжении всей эволюции млекопитающих распространение ретроэлементов вызвало перераспределение сайтов связывания для ряда регуляторов транскрипции, включая факторы плюрипотентности OCT4 и NANOG, белки инсуляторы CTCF, нейронные репрессоры NRSF/REST и другие [45]. Аналогично экспансия транспозонов MER20 и RLTR13D5 способствовала экспрессии в эндометрии и трофобласте для плацентарной транскрипции генов, необходимой для развития беременности [46]. Оказалось, что использование ретроэлементов в качестве альтернативных промоторов — распространенный феномен, который используется преимущественно в эмбриональных клетках, но также в некоторой степени в тканях взрослых организмов. При раннем эмбриональном развитии, в частности, до 20% транскриптома иницируется из ретро-транспозонов. Данные альтернативные промоторы проявляют склонность к тканеспецифической активности. Во многих случаях эти ретро-TE были кооптированы хозяином посредством экзонизации [45]. Экзонизацией называют процесс приобретения генами новых экзонов из некодирующих последовательностей, главным образом интронов. Инсерции или точковые мутации часто способствуют образованию альтернативных сайтов сплайсинга, вызывая включение новых последовательностей в качестве экзонов или удлинение существующих экзонов [31].

TE оказались источниками почти половины функциональных регуляторных последовательностей в геноме человека. TE расположены приблизительно в 44% открытых областях хроматина (участках хроматина, свободных от нуклеосом и гиперчувствительных к DNКазе I). При этом в открытых областях хроматина у приматов обнаруживается до 63% TE. В исследовании Jacques et al. для определения, в какой степени TE способствуют созданию транскрипционной сети человека, были использованы наборы данных консорциума Энциклопедии элементов ДНК (ENCODE) с определением расположения активных регуляторных элементов в клетках более 40 различных тканей человека. Использование данного ресурса позволило измерить вклад всех классов повторяющихся последовательностей и системно охарактеризовать воздействие, которое TE оказывают на ландшафт хроматина человека. Полученные результаты показали, что последовательности, произошедшие из TE, внесли сотни тысяч новых регуляторных элементов в линию приматов и изменили ландшафт транскрипции генома человека. Для системного исследования вклада TE в регуляторные сети человека по целому ряду типов клеток человека были использованы данные гиперчувствительных к DNКазе-I сайтов (DHS), созданные в Университете Вашингтона и в Университете Дюка как часть ENCODE. Преиму-

щество использования этих карт доступности хроматина заключается в выделении активных последовательностей ДНК. Хотя доступность не приравнивается к регуляторной функции, Jaques et al. основывались на этих наборах данных для измерения глобального профиля активности всех классов последовательностей, полученных от транспозонов, и системно охарактеризовали влияние TE на ландшафт хроматина человека. Начиная с 106 наборов данных DHS, был проведен обширный контроль свойств 75 наборов данных, определяющих в общей сложности 11848530 открытых областей хроматина в 41 типе клеток человека, выделенных из нормальных, эмбриональных и злокачественных тканей. Полученные данные DHS в дальнейшем сгруппированы по типам клеток в 1643643 различных области открытого хроматина. При измерении перекрытия с повторяющимися элементами обнаружено, что 725610 (44.1%) DHS областей перекрываются с экземплярами 4 основных классов TE (ERV, LINE, SINE и ДНК транспозоны). При разделении областей DHS на основании наличия или отсутствия гомологичных последовательностей в ортологичных локусах у других видов обнаружено, что эта доля достигает 63.1% для TE, встроенных в последовательности, специфичные для приматов [9].

Тысячи последовательностей, произошедших из TE, активируются в специфических типах клеток. Данная активность связана с экспрессией соседних генов, специфичных для типов клеток. Таким образом, TE, в частности ERV (эндогенные ретровирусы), внесли сотни тысяч новых регуляторных элементов в линию приматов и изменили ландшафт транскрипции человека [9]. Изменения регуляции экспрессии генов, в отличие от изменений в самих генах, играют значительную роль в морфологической эволюции. Регуляция генов во многом зависит от TFBS. При исследовании TFBS у 29 видов млекопитающих показано, что геномы кооптируют фрагменты TE для использования в качестве генной регуляторной последовательности в широких масштабах [16].

TE являются основными игроками в эволюции геномов. Эффекты их передвижений варьируют от нокаута гена до более тонких — изменения экспрессии гена. Так как TE могут содержать TFBS, они могут переписывать новые гены в существующие транскрипционные сети. Показано, например, что TE значительно амплифицировали количество последовательностей, соответствующих E2F TFBS в процессе видообразования капусты. В результате у некоторых видов капусты не менее 85% последовательностей, соответствующих консенсусам E2F TFBS, находятся в пределах TE. Транспозоны вызывают прямой эффект на опосредованную E2F генную регуляцию, так как последовательности внутри TE связывают E2F *in vivo*. TE, расположенные вблизи генов, могут непосред-

ственно служить в качестве промоторов генов. Локализация ТЕ вдали генов оказывает косвенный эффект за счет связывания с белком E2F и уменьшения его количества [18].

Энхансеры – это последовательности TFBS, усиливающие транскрипцию генов. Данные области обычно имеют длину 150–500 п.н. и могут располагаться далеко от стартового сайта транскрипции. Сравнение геномов 29 видов млекопитающих позволило обнаружить более 280000 консервативных некодирующих элементов, произошедших из ТЕ с потенциально регуляторной функцией. Многие из них могут быть энхансерами. Например, у млекопитающих энхансерами служат: LF-SINE для гена *Isl1*, CORE-SINE – для *Pomc*, MIR – для *Tal1*, AmnSINE – для *SATB2*, AmnSINE – для *Fgf8*, MaLR – для *Pomc*. У человека LINE1 служит энхансером для гена *apoa*, Alu – для *K19*; у мыши B2 SINE – для гормона роста; у кукурузы ДНК-ТЕ *Hopscotch* – для *tb1* [17]. В геноме человека энхансеры, произошедшие из MIR (диспергированные повторы млекопитающих) ТЕ, являются богатыми источниками сайтов связывания ТФ, что объясняет их сохранение в эволюции. Произошедшие из MIR энхансеры являются важной *cis*-регуляторной платформой, по которой MIR могут осуществлять регуляторную функцию в геноме человека [19]. Наиболее активными энхансерами в возникновении неокортекса эмбрионов мышей в возрасте 14.5 дней оказались семейства повторов MER130. Было показано, что MER130 имеют значение в регуляции формирования коры головного мозга, располагаясь вблизи с генами, важными для развития и функционирования неокортекса. MER130, неавтономный ТЕ, возникший у четвероногих или, возможно, у лопастеперых рыб, что предшествовало появлению коры больших полушарий. Предполагается, что встраивание регуляторных последовательностей транспозонов MER130 в качестве энхансеров способствовало образованию неокортекса в эволюции [47]. ТЕ могут участвовать в регуляции экспрессии генов на транскрипционном и посттранскрипционном уровнях различными способами. За счет перемещений ТЕ в эволюции формировались регуляторные сети, специфичные для определенных видов. При этом видоспецифические ТЕ характеризуются группировкой вокруг генов, вовлеченных в развитие и регуляцию транскрипции [5].

Получено множество свидетельств того, что ТЕ служат важными источниками TFBS в ключевых механизмах переключения экспрессии генов при дифференцировке клеток в онтогенезе [16–20]. Можно предположить, что особенности расположения данных TFBS, а также локализация самих ТЕ в геномах могут участвовать в эпигенетической регуляции работы генома. Это опосредовано образованием из транскриптов ТЕ при их процессинге siРНК [34] и микроРНК [35–40]. Данные нкРНК

регулируют экспрессию ТЕ, из которых они образованы, а также белок-кодирующих генов, содержащих комплементарные им последовательности. В эволюции ТЕ являются важными источниками возникновения новых белков за счет одомашнивания генов ТЕ [5, 21–26], образования ретрокопий и экзонизацией ТЕ [27–33]. За счет этого многие гены содержат последовательности ТЕ и способны комплементарно связываться с нкРНК, происходящими из ТЕ, что опосредует формирование сложных регуляторных сетей эпигенетического управления онтогенезом.

### ТРАНСПОЗОНЫ КАК ИСТОЧНИКИ БЕЛОК-КОДИРУЮЩИХ ГЕНОВ В ЭВОЛЮЦИИ

В связи с глобальной распространенностью ТЕ в геномах эукариот, их последовательности используются как для использования регуляторных последовательностей в управлении экспрессии генов хозяев, так и в образовании новых генов. В эволюции эукариот гены ТЕ могут быть одомашнены и использованы для нужд хозяина [5, 21–26]. ТЕ способствуют образованию ретрокопий существующих белок-кодирующих генов [27, 28]. Перемещения ТЕ могут служить причиной эволюции белок-кодирующих генов путем экзонизации при inserции в интрон [29–31] или вблизи 3'UTR гена [32, 33]. Новые варианты белков, образованные в результате экзонизации, способствуют появлению новой функции. Если приобретенное свойство позволяет организмам лучше адаптироваться, альтернативный сплайсинговый вариант может полностью заместить предшествующий [31]. Характерно тканеспецифическое использование экзонов, происходящих из ТЕ, что говорит о возможной роли транспозонов в управлении дифференцировкой клеток [29, 30]. В данном отношении важное значение имеет экзонизация межгенных ТЕ за счет образования изоформ с альтернативным 3'UTR. Так как копии ТЕ содержатся в 3'UTR многих генов, то произошедшие из них микроРНК могут регулировать клеточный уровень большого числа неродственных мРНК с образованием таксонспецифических микроРНК [32, 33].

Сохранение последовательностей ТЕ в составе генов имеет важное значение в обеспечении их специфической регуляции. Этим можно объяснить сохранение в эволюции ретрокопий, дублированных при помощи ретротранспозонов. Около 10% всех белок-кодирующих генов содержат более одного ретрогена. Наличие большого количества ретрокопий является одним из способов, обеспечивающих эукариот разнообразием белков [27]. При этом ретрогены отличаются от родительских генов наличием фланкированных специфичных для ТЕ повторов, что позволяет стадиспецифически и тканеспецифически регулировать их экспрессию [28]. Регуляция обеспечивается взаимо-

действием с нкРНК, произошедшими из ТЕ [33], что обеспечивается способностью siРНК специфически изменять метилирование ДНК и модификации гистонов [48]. МикроРНК, помимо посттранскрипционного воздействия, также способны оказывать влияние на транскрипцию путем взаимодействия с гистоновыми деацетилазами и ДНК-метилтрансферазами [49, 50]. Дополнительные регуляторные взаимосвязи ретрокопии приобретают при инсерции в них новых ТЕ. Многие ретрогены активно экспрессируются и могут сохранять функцию родительского гена (субфункционализация), развить новую функцию (неофункционализация) или полностью заместить родительский ген [28].

Одомашнивание генов ТЕ оказалось распространенным явлением у всех эукариот. Приобретенные хозяином гены транспозонов часто становятся консервативными в ходе эволюции, выполняя важные функции в работе генома. Наиболее характерный пример – возникновение гена теломеразы, произошедшей, как предполагается, из обратной транскриптазы (RT) LINE-1 в связи с их значительным сходством [23]. Благодаря образованию теломер геномы всех эукариот отличаются от прокариотических универсальным свойством дискретности генома.

Одомашнивание гена *env* оболочки LTR-содержащих ретровирусов у млекопитающих привело к возникновению в геноме хозяина нового гена *ERVW-1*, кодирующего белок синцитин, который необходим для регуляции образования объединенного слоя клеток при имплантации зародыша на поверхности матки [24]. У животных, растений и грибов обнаружено большое количество белок-кодирующих генов, образованных путем одомашнивания генов ТЕ. В геноме человека идентифицировано 47 генов, произошедших непосредственно от ТЕ, из них 43 образованы из транспозазы [5]. Из древней транспозазы возникли гены факторов транскрипции FHY3 и FAR1, которые обеспечивают реакцию на свет у растений [22]. У позвоночных выявлено более 1000 генов, которые содержат домены транспозонного происхождения – RETRA (Retroviral or Transposon-associated) [21]. То есть ТЕ способны образовывать структуры белков, оказывающие регуляторное воздействие при взаимодействии с ДНК. Например, транспозазы образуют узлы и петли в местах взаимодействия с ДНК, способствуя организации и компактизации генома. Многие белки ТЕ-происхождения оказались важнейшими транскрипционными факторами [5]. Учитывая ключевую роль ТЕ в обеспечении геномов эукариот TFBS [16–20], можно предположить, что ТЕ служат универсальными источниками в построении регуляторных сетей геномов. Это обеспечивается благодаря наличию в последовательностях ТЕ повторов, образующих доменные структуры в пространстве и способных взаимодействовать с другими молекулами. Данное свойство универ-

сально как для транскриптов ТЕ, так и для их белковых продуктов. В частности, выявлено, что фрагменты ТЕ, встроенные в последовательности длинных нкРНК, способствуют их функционированию. Последовательности ТЕ используются в качестве сайтов распознавания и для белков, и для нуклеиновых кислот. Это объясняет значительную распространенность ТЕ в структуре длинных нкРНК. У человека, например, 83% экзонов всех длинных нкРНК содержат не менее одного фрагмента ТЕ [51].

Универсальные свойства повторов ТЕ приводят к тому, что их одомашнивание может способствовать образованию белков, обеспечивающих глобальное регулирование генома, что подчеркивает высокий потенциал ТЕ в регуляции онтогенетического развития. Примером является происхождение от обратной транскриптазы LINE ретротранспозонов теломеразы. Одомашненный ген теломеразы в эволюции стал консервативным для всех эукариот, обеспечивая формирование теломер на концах хромосом [52]. ТЕ оказались источниками белков HDP 1 и HDP 2, произошедших из ДНК-связывающего белка и транспозазы ДНК-транспозона Harbinger [26]. Из рога-подобной транспозазы произошел центромер-связывающий белок CENP-B [21]. Одомашненные из транспозонов MIR (mammalian-wide interspersed repeats) инсульторы используются для модуляции генных регуляторных сетей путем рекрутирования транскрипционных комплексов и ферментов модификации хроматина у человека [25].

Таким образом, в эволюции ТЕ способствуют возникновению новых генов путем экзонизации, дубликаций и одомашнивания. ТЕ оказались важными источниками ТФ [5], что, наряду с их важным значением в образовании TFBS, говорит о ключевой функции ТЕ в регуляции экспрессии генов. Данная функция связана с ролью ТЕ в качестве источников нкРНК.

#### ТРАНСПОЗОНЫ – ИСТОЧНИКИ ЭПИГЕНЕТИЧЕСКОГО УПРАВЛЕНИЯ

Расположение ТЕ в геноме оказывает влияние на экспрессию соседних генов в связи с эпигенетическим изменением структуры хроматина под влиянием происходящих из транспозонов нкРНК, модифицирующих гистоны [49, 50]. Процессинг транскриптов ТЕ приводит к образованию siРНК [55] и микроРНК [35–40]. Более того, ТЕ могут служить непосредственно в качестве генов длинных нкРНК, а образованные из них нкРНК участвуют в регуляции экспрессии генов при дифференцировке клеток [42]. Продукты ТЕ обладают также потенциалом влияния на структуру хроматина. Например, белки HDP1/2, произошедшие от ДНК-транспозонов, используются в ацетилтрансферазном комплексе гистонов для деметилирования ДНК [26]. Геном человека на 10% состоит из после-

довательностей SINE, представленных миллионами копий. При этом SINE транспозоны формируют регуляторную сеть и оказывают существенное влияние на хроматиновый ландшафт своих “хозяев” [53].

Значение TE в качестве источников TFBS, новых белков и некодирующих РНК отображает сложные эволюционные процессы, в которых TE используются в формировании регуляторных сетей управления онтогенезом. Выявлено, например, что первый промотор гена *Aebp2* произошел от ретро-транспозонов независимо в линиях приматов и грызунов. Данный промотор не метилирован в сперме, метилирован в зрелых яйцеклетках и частично метилирован у эмбрионов. Для первого промотора гена *Aebp2* также характерны различные уровни метилирования ДНК в разных органах взрослых особей [54]. TE служат основными источниками siРНК у растений и животных. В последние годы интенсивно изучается сайленсинг, индуцированный siРНК, которые транскрибируются из последовательностей TE. Данный сайленсинг действует путем РНК-направленного метилирования ДНК (RdDM) и участвует в поддержании целостности генома. Он используется для супрессии генов, не относящихся к TE. Выявлено, например, что siR815, происходящая из TE, оказывает противоположное действие на два аллеля *WRKY45-1* и *WRKY45-2* гена, кодирующего транскрипционный фактор *WRKY45* у риса, что имеет важное значение в устойчивости к бактериальному заболеванию растения. Экспрессия аллеля *WRKY45-1*, активируемая при помощи siR815, способствует резистентности к *Xanthomonas oryzae pv. oryzae*. В то же время на аллель *WRKY45-2*, вызывающий чувствительность к бактерии, siR815 оказывает ингибирующее воздействие, что говорит о защитной роли произошедшей от TE siРНК для хозяина [55].

TE являются важнейшими источниками возникновения микроРНК у животных [35, 36, 38, 39] и растений [37, 40]. Например, в 2011 г. у мыши выявлено 141 микроРНК, произошедших из TE, у человека – 226, у резуса – 115 [35]. При анализе базы данных микроРНК и TE животных в 2011 г. было выявлено 2392 микроРНК, произошедших от различных повторов, главным образом TE [36]. В 2013 г. были выявлены 1213 новых локусов микроРНК, созданных инсерциями TE [38], в 2014 г. – 1494 новых микроРНК, непосредственно произошедших из TE [39]. У растений большинство микроРНК гомологичны или даже идентичны известным TE [37, 40]. Происхождение многих микроРНК от TE говорит о ключевой роли транспозонов в регуляции экспрессии генов в индивидуальном развитии. Регуляция генов опосредуется способностью микроРНК оказывать воздействие на экспрессию белок-кодирующих генов не только путем взаимодействий с целевыми мРНК, но и за счет влияния

на гистоновые деацетилазы и ДНК-метилтрансферазы [48, 50].

Согласно данным крупных исследовательских проектов, основная часть генома любого живого организма транскрибируется. Например, у человека транскрибируется более 75% генома [56]. Большая часть транскриптов представлена длинными нкРНК, которые регулируют транскрипцию, сплайсинг, деградацию РНК и трансляцию [57]. Отличительными свойствами длинных нкРНК является высокая тканевая специфичность. Они влияют на изменение химической и пространственной структуры хроматина при клеточной дифференцировке. Многие длинные нкРНК полифункциональны и оказывают воздействие на пространственное размещение в ядре целого ряда TF, участвующих в пролиферации и миграции клеток [58]. Оказалось, что LTR-содержащие TE могут служить непосредственно в качестве генов длинных нкРНК [42], а для большинства генов длинных нкРНК транспозоны составляют более 60% последовательностей их транскриптов [41], что говорит о ключевой роли TE в эволюционных механизмах формирования регуляторных сетей геномов.

## ЗАКЛЮЧЕНИЕ

Таким образом, получены многочисленные данные о важной роли TE в формировании и регуляции генных регуляторных сетей эукариотических организмов. У многоклеточных животных данная регуляция характеризуется активацией специфических TFBS транспозонного происхождения в эмбриональном развитии при дифференцировке стволовых клеток. Данные особенности активации TFBS сопровождаются специфическими изменениями экспрессии генов, необходимыми для формирования органов и тканей. Динамичность последовательностей TE, связанная с их способностью к перемещениям, обосновывает логичность их использования геномом хозяев для систем регуляции генных сетей в развитии. Данные системы обладают видоспецифическими особенностями, что отражается в проявлении фенотипических признаков. Основой для формирования систем регуляции генных сетей в эволюции служат перемещения TE с изменением особенностей активации определенных генов при помощи TFBS и некодирующих РНК транспозонного происхождения. В эволюции эукариот TE оказались важными источниками возникновения новых белков, в том числе TF, участвующих в регуляции функционирования генома в развитии организмов. Это говорит о функции TE в формировании динамичных генных сетей. Поддержание данной функции и ее преобразования обеспечиваются образованием некодирующих РНК при процессинге транскриптов TE.

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Все применимые международные, национальные и/или институциональные принципы ухода и использования животных были соблюдены.

Настоящая статья не содержит каких-либо исследований с участием людей в качестве объектов исследований.

## СПИСОК ЛИТЕРАТУРЫ

1. Doolittle W.F., Sapienza C. Selfish genes, the phenotype paradigm and genome evolution // *Nature*. 1980. V. 284. № 5757. P. 601–603.
2. Orgel L.E., Crick F.H. Selfish DNA: the ultimate parasite // *Nature*. 1980. V. 284. № 5757. P. 604–607.
3. McClintock B. The significance of responses of the genome to challenge // *Science*. 1984. V. 226. № 4676. P. 792–801.
4. Davidson E.H., Britten R.J. Regulation of gene expression: possible role of repetitive sequences // *Science*. 1979. V. 204. № 4397. P. 1052–1059.
5. Feschotte C. Transposable elements and the evolution of regulatory networks // *Nat. Rev. Genet.* 2008. V. 9. P. 397–405.
6. Ратнер В.А., Васильева Л.А. Мобильные генетические элементы и количественные признаки у дрозофилы: факты и гипотезы // *Генетика*. 1992. Т. 28. № 11. С. 15–29.
7. Ратнер В.А., Васильева Л.А. Роль мобильных генетических элементов в микроэволюции // *Генетика*. 1992. Т. 28. № 12. С. 5–16.
8. Кайданов Л.З., Галкин А.П., Иовлева О.В., Сиделева О.Г. Направленный характер перемещения по геному мобильного элемента хобо в длительно селективируемой линии *Drosophila melanogaster* // *Цитология и генетика*. 1996. Т. 30. № 1. С. 23–30.
9. Jacques P.E., Jeyakani J., Bourgue G. The majority of primate-specific regulatory sequences are derived from transposable elements // *PLoS Genet.* 2013. V. 9. № 5. e1003504.
10. Rodic N., Burns K.H. Long Interspersed Element-1 (LINE-1): Passenger or Driver in Human Neoplasms? // *PLOS Genetics*. 2013. Vol. 9. Is. 3. e1003402.
11. Van Meter M., Kashyap M., Rezazadeh S., Geneva A.J., Morello T.D., Seluanov A., Gorbunova V. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age // *Nat. Commun.* 2014. V. 5. P. 5011.
12. Мустафин Р.Н., Хуснутдинова Э.К. Взаимосвязь эпигенетических факторов в механизмах старения и малигнизации // *Успехи физиологических наук*. 2017. Т. 48. № 2. С. 72–99.
13. Галицкий В.А. Эпигенетическая природа старения // *Цитология*. 2009. Т. 51. № 5. С. 388–397.
14. Patel T., Hobert O. Coordinated control of terminal differentiation and restriction of cellular plasticity // *eLife*. 2017. V. 6. e24100.
15. Мустафин Р.Н. Некодирующие части генома как основа эпигенетической наследственности // *Вавиловский журнал генетики и селекции*. 2017. Т. 21. С. 742–749.
16. Lowe C.B., Haussler D. 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome // *PLoS One*. 2012. V. 7. № 8. e43128.
17. de Souza F.S., Franchini L.F., Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong // *Mol. Biol. Evol.* 2013. V. 30. № 6. P. 1239–1251.
18. Henaff E., Vives C., Desvoyes B., Chaurasia A., Payet J., Gutierrez C., Casacuberta J.M. Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species // *Plant J.* 2014. V. 77. P. 852–862.
19. Jjingo D., Conley A.B., Wang J., Marino-Ramirez L., Lunyak V.V., Jordan I.K. Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression // *Mob. DNA*. 2014. V. 5. P. 5–14.
20. Ito J., Suqimoto R., Nakaoka H., Yamada S., Kimura T., Hayano T., Inoue I. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses // *PLoS Genet.* 2017. V. 13. № 7. e1006883.
21. Zdobnov E.M., Campillos M., Harrington E.D., Torrents D., Bork P. Protein coding potential of retroviruses and other transposable elements in vertebrate genomes // *Nucleic Acids Res.* 2005. V. 33. P. 946–954.
22. Lin R., Ding L., Casola C., Ripoll D.R., Feschotte C., Wang H. Transposase-derived transcription factors regulate light signaling in Arabidopsis // *Science*. 2007. V. 318. P. 1302–1305.
23. Kopera H.C., Moldovan J.B., Morrish T.A., Garcia-Perez J.L., Moran J.V. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase // *Proc. Natl. Acad. Sci. USA*. 2011. V. 108. P. 20345–20350.
24. Dupressoir A., Lavialle C., Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation // *Placenta*. 2012. V. 33. P. 663–671.
25. Wang J., Vicente-Garcia C., Seruggia D., Molto E., Fernandez-Minan A., Neto A., Lee E., Gomez-Skarmeta J.L., Montoliu L., Lunyak V.V., Jordan I.K. MIR retrotransposons sequences provide insulators to the human genome // *Proc. Natl. Acad. Sci. USA*. 2015. V. 112. P. 4428–4437.
26. Duan C.G., Wang X., Pan L., Miki D., Tang K., Hsu C.C., Lei M., Zhong Y., Hou Y.J., Wang Z., Zhang Z., Mangrauthia S.K., Xu H., Zhang H., Dilkes B., Tao W.A., Zhu J.K. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation // *Cell. Res.* 2017. V. 27. № 2. P. 226–240.
27. Grandi F.C., Rosser J.M., Newkirk S.J., Yin J., Jiang X., Xing Z., Whitmore L., Bashir S., Ivics Z., Izsvak Z., Ye P., Yu Y.E., An W. // Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res.* 2015. V. 25. P. 1135–1146.
28. Kubiak M.R., Makalowska I. Protein-Coding Genes' Retrocopies and Their Functions // *Viruses*. 2017. V. 9. № 4. pii: E80.

29. Lin L., Shen S., Tye A., Cai J.J., Jiang P., Davidson B.L., Xing Y. Diverse splicing patterns of exonized Alu elements in human tissues // *PLoS Genet.* 2008. V. 4. № 10. e1000225.
30. Sela N., Kim E., Ast G. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates // *Genome Biol.* 2010. V. 11. R59.
31. Schmitz J., Brosius J. Exonization of transposed elements: A challenge and opportunity for evolution // *Biochimie.* 2011. V. 93. P. 1928–1934.
32. Макарова Ю.А., Крамеров Д.А. Некодирующие РНК // *Биохимия.* 2007. Т. 72. С. 1427–1448.
33. Tajnik M., Vigilante A., Braun S., Hanel H., Luscombe N.M., Ule J., Zarnack K., Konig J. Inergenic Alu exonisation facilitates the evolution of tissue-specific transcript ends // *Nucleic Acids Res.* 2015. V. 43. P. 10492–10505.
34. Zhang H., Tao Z., Hong H., Chen Z., Wu C., Li X., Xiao J., Wang S. Transposon-derived small RNA is responsible for modified function of WRKY45 locus // *Nat. Plants.* 2016. V.2. P. 16016–16023.
35. Yuan Z., Sun X., Liu H., Xie J. MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes // *PLoS ONE.* 2011. V. 6. Iss. 3. e17666.
36. Borchert G.M., Holton N.W., Williams J.D., Hernan W.L., Bishop I.P., Dembosky J.A., Elste J.E., Gregoire N.S., Larson E.D. Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins // *Mobile Genetic Elements.* 2011. V. 1 № 1. P. 8–17.
37. Li Y., Li C., Xia J., Jin Y. Domestication of transposable elements into microRNA genes in plants // *PLoS One.* 2011. V. 6. e19212.
38. Roberts J.T., Cooper E.A., Favreau C.J. Formation from transposable element insertions and noncoding RNA mutations // *Mobile Genetic Elements.* 2013. V. 1. № 6. e27755.
39. Gim J., Ha H., Ahn K., Kim D.S., Kim H.S. Genome-Wide Identification and Classification of microRNAs derived from repetitive elements // *Genomic Inform.* 2014. V.12. № 4. P. 261–267.
40. Lorenzetti A.P., A de Antonio G.Y., Paschoal A.R., Domingues D.S. Plant TE-MIR DB: a database for transposable element-related microRNAs in plant genomes // *Funct. Integr. Genomics.* 2016. V. 16. P. 235–242.
41. Kapusta A., Feschotte C. Volatile evolution of long non-coding RNA repertoires: mechanisms and biological implications // *Trends Genet.* 2014. V. 30. № 10. P. 439–452.
42. Lu X., Sachs F., Ramsay L., Jacques P.E., Goke J., Bourgue G., Ng H.H. The retrovirus HERVH is a long non-coding RNA required for human embryonic stem cell identity // *Nat. Struct. Mol. Biol.* 2014. V. 21. № 4. P. 423–425.
43. Testori A., Caizzi L., Cutrupi S., Friard O., Bortoli M.D., Cora D., Caselle M. The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors // *BMC Genomics.* 2012. V. 13. P. 400.
44. Polavarapu N., Marino-Ramirez L., Landsman D., McDonald J.F., Jordan I.K. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA // *BMC Genomics.* 2008. V. 9. P. 226. doi: .10.1186/1471-2164-9-226
45. Mak K.S., Burdach J., Norton L.J., Pearson R.C., Crossley M., Funnell A.P. Repression of chimeric transcripts emanating from endogenous retrotransposons by a sequence-specific transcription factor // *Genome Biol.* 2014. V. 15. № 4. R 58.
46. Chuong E.B., Rumi M.A., Soares M.J., Baker J.C. Endogenous retroviruses function as species-specific enhancer elements in the placenta // *Nat. Genet.* 2013. V. 45. № 3. P. 325–329.
47. Notwell J.H., Chung T., Heavner W., Bejerano G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex // *Nat. Commun.* 2015. V. 6. P. 6644.
48. Zhang G., Esteve P., Chin H.G., Terragni J., Dai N., Correa I.R., Pradhan S. Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation // *Nucleic Acids Research.* 2015. V. 43. P. 6112–6124.
49. Morita S., Horii T., Kimura M., Ochiya T., Tajima S., Hatada I. MiR-29 Represses the Activities of DNA Methyltransferases and DNA Demethylases // *Int J Mol Sci.* 2013. V. 14. P. 14647–14658.
50. Samantarrai D., Dash S., Chhetri B., Mallick B. Genomic and epigenomic cross-talks in the regulatory landscape of miRNAs in breast cancer // *Mol. Cancer Res.* 2013. V. 11. № 4. P. 315–328.
51. Johnson R., Guigo R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs // *RNA.* 2014. V. 20. № 7. P. 959–976.
52. Garavis M., Gonzalez C., Villasante A. On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution // *Genome Biol. Evol.* 2013. V. 5. P. 1142–1150.
53. Ichiyangi K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs // *Genes Genet. Syst.* 2013. V. 88. № 1. P. 19–29.
54. Kim H., Bakshi A., Kim J. Retrotransposon-derived promoter of Mammalian Aebp2 // *PLoS One.* 2015. V. 10. № 4. e0126966.
55. Zhang H., Tao Z., Hong H., Chen Z., Wu C., Li X., Xiao J., Wang S. Transposon-derived small RNA is responsible for modified function of WRKY45 locus // *Nat. Plants.* 2016. V. 2. P. 16016–16023.
56. Djebali S., Davis C.A., Merkel A. et al. Landscape of transcription in human cells // *Nature.* 2012. V. 489. № 7414. P. 101–108.
57. Fitzgerald K.A., Caffrey D.R. Long noncoding RNAs in innate and adaptive immunity // *Current opinion in immunology.* 2014. V. 26. P. 140–146.
58. Лисицын Н.А., Черный А.А., Карнов В.Л., Берестень С.Ф. Роль длинных некодирующих РНК в процессе канцерогенеза // *Молекулярная биология.* 2015. Т. 49. № 4. С. 561–570.



## **The Relationship between Transposons and Transcription Factors in the Evolution of Eukaryotes**

**R. N. Mustafin**

*Bashkir State University, Ufa, Russia*

*E-mail: ruji79@mail.ru*

Transposons are important sources of binding sites for transcription factors whose specific activation characterized the embryonic development of animals. In the evolution of eukaryotes, molecular domestication of the mobile genetic elements led to the emergence of multiple novel proteins, including the transcription factors involved in the control of cell differentiation. Transposons are abundantly present in intergenic DNA, introns and 3R-untranslated regions, specifically, near the transcription factor genes which they regulate. This promotes gene interregulation through transposon activation by the products of their expression in successive divisions of stem cells, thus representing a dynamic biological encoding of the species-specific ontogenetic regulation. This program can be implemented due to tissue- and stage-specific regulation mediated by the presence of transposons at specific genomic sites. The crucial role of domesticated transposon genes in controlling genome operation and key developmental stages provides a sound argument in favor of this assumption.

*Key words:* differentiation, stem cells, transcription factors, transposons, evolution