

<https://doi.org/10.17116/molgen20193702151>

## Роль транспозонов в дифференцировке стволовых клеток

Р.Н. МУСТАФИН<sup>1,2</sup>

<sup>1</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Башкирский государственный медицинский университет», 450008, Уфа, Россия;

<sup>2</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Башкирский государственный университет», 450076, Уфа, Россия

В тотипотентных и плюрипотентных клетках эмбриона, а также в стволовых клетках в постнатальном периоде активация мобильных генетических элементов обязательна и является основой управления экспрессии генов в последовательных клеточных делениях с их изменением специфично в зависимости от ткани и стадии развития. Предполагается, что в эволюции многоклеточных эукариот отбираются оптимальные соотношения в расположении транспозонов относительно экзонов и интронов генов хозяев. Данные соотношения, специфичные для каждого вида, могут быть основой для управления последовательной дифференцировкой стволовых клеток, необходимой для развития целостного организма. Несмотря на то что дифференцировка клеток в онтогенезе контролируется весьма консервативным набором генов, мобильные генетические элементы участвуют в тонкой настройке регуляторных сетей, управляющей экспрессией данных генов, что отражается в фенотипических особенностях каждого вида. Транспозоны служат важнейшими источниками специфически активируемых в эмбриональном развитии регуляторных элементов многоклеточных, таких как сайты связывания с транскрипционными факторами, энхансеры и сайленсеры, промоторы, инсуляторы, сайты альтернативного сплайсинга, некодирующие РНК. Кроме того, мобильные генетические элементы участвуют в возникновении и эволюции новых белок-кодирующих генов путем экзонизации, одомашнивания, формирования ретрогенов. При достижении предопределенных размеров органов в терминально дифференцированных клетках организмов активируются молекулярные системы, блокирующие дальнейший каскад активации транспозонов — несовершенство данных систем может быть причиной старения и ассоциированных с возрастом болезней в связи с патологической активацией мобильных генетических элементов. Предполагается, что выявление тканеспецифических механизмов наследуемой активации транспозонов в стволовых клетках, а также их патологической активации в терминально дифференцированных клетках позволит найти способы борьбы со старением.

**Ключевые слова:** мобильные генетические элементы (МГЭ), некодирующие РНК (нкРНК), сайты связывания с транскрипционными факторами (TFBS), стволовые клетки (СК), тканеспецифичность (ТС), транскрипционные факторы (ТФ).

### ИНФОРМАЦИЯ ОБ АВТОРЕ:

**Мустафин Рустам Наилевич** — кандидат биологических наук, младший научный сотрудник, лаборатория ПЦР-анализа, ФГБОУ ВПО «Башкирский государственный университет», 450076, РФ, г. Уфа, ул. Заки Валиди, 32; e-mail: ruji79@mail.ru; <https://orcid.org/0000-0002-4091-382X>.

### КАК ЦИТИРОВАТЬ:

Мустафин Р.Н. Роль транспозонов в дифференцировке стволовых клеток. *Молекулярная генетика, микробиология и вирусология* 2019;37(2):51-57.

<https://doi.org/10.17116/molgen20193702151>

## The role of transposable elements in the differentiation of stem cells

R.N. MUSTAFIN<sup>1,2</sup>

<sup>1</sup>Bashkir State Medical University, 450008, Ufa, Russia;

<sup>2</sup>Bashkir State University, 450076, Ufa, Russia

Activation of mobile genetic elements is a prerequisite for controlling the expression of genes in successive cell divisions with their change specifically depending on the tissue and developmental stage. These patterns of activations are characteristic of totipotent and pluripotent cells of the embryo, as well as for stem cells in the postnatal period. We have suggested that in the evolution of multicellular eukaryotes, optimal ratios are selected in the arrangement of transposons relative to exons and introns of host genes. These ratios, specific to each species, can be the basis for controlling the sequential differentiation of stem cells necessary for the development of the whole organism. Despite the fact that cell differentiation in ontogenesis is controlled by a very conservative set of genes, mobile genetic elements are involved in the fine-tuning of regulatory networks that control the expression of these genes, which is reflected in the phenotypic characteristics of each species. Transposons are important sources of genome structures that are actively used to regulate the embryonic development of multicellular. These structures include binding sites with transcription

**Для корреспонденции:** Мустафин Рустам Наилевич — канд. биол. наук, младший научный сотрудник, лаборатория ПЦР-анализа, ФГБОУ ВПО «Башкирский государственный университет», 450076, РФ, Уфа, ул. Заки Валиди, 32; e-mail: ruji79@mail.ru; <https://orcid.org/0000-0002-4091-382X>

**For correspondence:** Mustafin Rustam Nailevich — 450076, Ufa, Zaki Validi street, 32; e-mail: ruji79@mail.ru; <https://orcid.org/0000-0002-4091-382X>

© Р.Н. Мустафин, 2019

factors, enhancers and silencers, promoters, insulators, alternative splicing sites, non-coding RNA. Moreover, transposons are involved in the emergence and evolution of new protein-coding genes through exonization, domestication, and the formation of retrogenes. Activation of transposons is needed to regulate the differentiation and reproduction of cells in the body; however, in terminally differentiated cells, upon reaching predetermined sizes of organs, molecular systems are activated that block a further cascade of transposon activation. We have suggested that imperfection of systems aimed at specific suppression of transposon activity in mature cells may be the cause of aging and age-related diseases due to the pathological activation of mobile genetic elements. Identifying the tissue-specific mechanisms of inherited transposon activation in stem cells, as well as their pathological activation in terminally differentiated cells, can be the basis for finding ways to fight aging.

**Keywords:** mobile genetic elements (MGE), non-coding RNA (ncRNA), review, transcription factor binding sites (TFBS), stem cells (SC), tissue specificity (TS), transcription factors (TF).

#### INFORMATION ABOUT AUTHOR:

Mustafin Rustam Nailevich – e-mail: ruji79@mail.ru; <https://orcid.org/0000-0002-4091-382X>

#### TO CITE THIS ARTICLE:

Mustafin RN. The role of transposable elements in the differentiation of stem cells. *Molekulyarnaya Genetika, Mikrobiologiya i Virusologiya (Molecular Genetics, Microbiology and Virology)*. 2019;37(2): 51-57 (Russian). <https://doi.org/10.17116/molgen20193702151>

## Введение

При дифференцировке стволовых клеток (СК) в эмбриональном развитии многоклеточных животных происходят репрессия неспецифических и активация специфических генов, за счет чего ограничивается пластичность развития. Обнаружено, что изменения метилирования ДНК, модификаций гистонов [1], особенностей экспрессии длинных некодирующих РНК (днРНК) [2] и микроРНК [3], активации сайтов связывания с транскрипционными факторами (ТФ) [4], происходящие при дифференцировке СК, связаны с активацией мобильных генетических элементов (МГЭ) [5]. МГЭ составляют значительную долю геномов всех животных и растений — их сохранение в эволюции не случайно, так как многие МГЭ используются геномами хозяев для регуляции экспрессии клеточных генов в последовательных делениях [6]. В зиготе происходит глобальное деметилирование генома, при котором стираются эпигенетические метки. При последовательных делениях клеток происходит ремоделирование всего эпигенома, с тканеспецифическими и стадийспецифическими особенностями данного процесса [7]. В связи с этим представление о существовании некоего эпигенетического наследования, не связанного с изменением структуры генома, нелогично и необоснованно. В качестве базиса для глобального перепрограммирования характера метилирования ДНК и модификаций гистонов могут быть использованы только последовательности нуклеотидов в геноме, представляющие собой особую по структуре и функции кодировку, активация которой в каждом клеточном делении связана с предыдущими этапами регуляторных изменений экспрессии конкретных генов. Данный код наиболее вероятно связан с динамическими структурами, способными активироваться под действием внутриклеточных и внеклеточных сигналов, что обосновывает особенности дифференцировки клеток во времени и пространстве. Наиболее подходящими структурами в данном отношении являются МГЭ, играющие ключевую роль в возникновении значительной доли регуляторных последовательностей [4], днРНК [2] и микроРНК животных [8] и растений [9]. Значительное распространение и сохранение в эволюции МГЭ в геномах эукариот обосновывают правомерность представленных суждений. Имеется ряд исследований, доказывающих закономерную активацию МГЭ в эмбриогенезе [10, 11], а также в постнатальном развитии

сохранение потенциала активации МГЭ в СК [12], сопряженного с экспрессией днРНК [13] (домены которых происходят от МГЭ [14]). Более того, истощение специфических МГЭ и днРНК на ранних этапах эмбриогенеза приводит к полной остановке дальнейшего развития, что говорит о ключевой регуляторной роли активации МГЭ в управлении дифференцировкой клеток путем контроля экспрессии специфических генов [2, 11].

МикроРНК и днРНК играют важную роль в управлении плюрипотентности, что обосновывает роль МГЭ в качестве регуляторов экспрессии генов при дифференцировке клеток, так как МГЭ являются важным источником возникновения микроРНК и днРНК [5]. Кроме того, в онтогенезе характерна активация промоторов, происходящих от МГЭ [15], а экспрессия, например, LTR-содержащих ретроэлементов является необходимым условием плюрипотентности [7], в то время как изменение пластичности СК при их дифференцировке сопровождается активацией ретротранспозонов, не содержащих LTR [16].

#### Активация транспозонов в эмбриогенезе

В геномах разных видов эукариот содержатся различные по составу, количеству и расположению друг к другу и белок-кодирующим генам МГЭ, что позволяет предположить их влияние на проявление фенотипа и особенности изменения экспрессии генов при развитии. Транскрипция L1 *in vivo* в бластоцисте мыши была описана еще в 1993 г. Была обнаружена обильная экспрессия L1 размером около 8 кб со смысловой цепью, полиаденилированием, 5'-концом и двумя открытыми рамками считывания (ORF — Open Reading Frame) [17]. В 1994 г. описана дифференциальная экспрессия L1 в половых и соматических клетках семенников в онтогенезе мыши. Особый интерес представила коэкспрессия полноразмерной смысловой цепи L1 РНК и L1-кодируемого белка в лептотене и зиготене сперматоцитов на 14-й день постнатального развития. Экспрессия в мейотической профазе предшествовала разрыву нити во время хромосомной рекомбинации, что дало возможность для инсерции L1 в новый локус ДНК хромосом в клеточном типе и способствовало распространению L1 в последующих поколениях [18]. В 1995 г. при помощи иммуногистохимического анализа была выявлена экспрессия L1 у постимплантационных эмбрионов мыши. Полученные результаты показали, что экспрессия МГЭ контролируется

жесткой регуляторной программой во времени и пространстве [19].

В 1996 г. обнаружена высокая частота активных ретротранспозиций в культурах клеток млекопитающих (мыши и человека) [20]. В 1998 г. выявлены ретротранспозиции семи полноразмерных активных элементов T(F) подсемейства L1 в культурах клеток мыши. При этом было показано, что у мыши и человека количество активных L1-элементов значительно отличается [21]. В 2000 г. описано накопление множества ретротранспозиционных событий L1 в культуре клеток человека [22].

В 2003 г. было показано, что L1-EGFP могут перемещаться *in vivo* на ранних стадиях развития трансгенных мышей [23]. В 2007 г. в экспериментах *in vitro* было обнаружено накопление инсерций LINE-1 в ЭСК, сопровождающееся подавлением активности определенных генов. На основании этого был предложен механизм управления работы генома при помощи МГЭ, так как специфическое изменение экспрессии генов под влиянием МГЭ в клетках сопровождалось их дифференцировкой и изменением функций [24]. Можно предположить, что механизм, при котором активация определенных МГЭ в очередном клеточном делении в зависимости от ткани и стадии развития вызывает специфическое регулирование экспрессии генов, является универсальной системой для управления дифференцировкой клеток. Причиной может быть высокая чувствительность специфических МГЭ к определенным внутриклеточным (стадия развития) и внеклеточным (пространственное расположение) сигналам, за счет чего возможен видоспецифический наследуемый каскад активаций МГЭ, начиная с первого деления зиготы, когда образующиеся в каждом новом типе клеток сигналы могут активировать строго определенные транспозоны, что влечет за собой дальнейшие реакции с активацией других МГЭ и генов хозяев. Данная система, несмотря на динамичность за счет чувствительности МГЭ к стрессовым [25] и гормональным воздействиям [26], способна стабилизироваться на уровне вида.

Белки, кодируемые L1-элементами, могут действовать *in trans* для мобилизации SINE, таких как Alu. В 2011 г. была обнаружена экспрессия нескольких подсемейств Alu-элементов в недифференцированных эмбриональных СК человека. Наиболее активная экспрессия наблюдалась для молодого подсемейства Y Alu-элементов. Было выявлено также, что экспрессия L1s происходила главным образом у элементов, расположенных внутри генов, свидетельствуя о наличии эпигенетического контроля ретротранспозонов в эмбриональных СК [10]. В 2016 г. выявлена реактивация транскрипции определенных классов ретротранспозонов в раннем эмбриональном развитии животных по линиям СК. Профили специфических МГЭ могут свидетельствовать о клеточной идентичности, что говорит о важном значении МГЭ в управлении экспрессией генов при дифференцировке клеток. Активная транскрипция МГЭ способствует также поддержанию плюрипотентности [7].

В 2015 г. было сделано предположение, что МГЭ способствуют тканеспецифической и линиеспецифической регуляции экспрессии генов в онтогенезе на основании свидетельств того, что LTR-содержащие МГЭ тканеспецифически регулируют соседние гены. При использовании корреляции паттернов экспрессии для восемнадцати типов тканей было выявлено тканеспецифическое расщепление экспрессии генов по шестидесяти двум различным классам LTR. Данные паттерны были специфичны для ре-

тровирусных инсерций, так как те же гены у видов без LTR не проявляли сходных эффектов [27]. В индуцированных плюрипотентных стволовых клетках человека (hiPSC) также выявлена эндогенная мобилизация Alu, SVA и L1-элементов при их репрограммировании [28]. Было показано, что различные днРНК, образующиеся из HERVH, не только экспрессируются, но и необходимы для плюрипотентности ЭСК человека [5]. В 2012 г. получены подтверждения того, что многие транскрипты эмбрионов мыши двухклеточной стадии инициируются из LTR, происходящих от ERV. Это стало свидетельством регуляции судьбы клеток у плацентарных млекопитающих при помощи LTR МГЭ [11]. В 2014 г. показано, что транскрипты, полученные из LTR у мышей, вносят весомый вклад в усложнение ядерного транскриптома ЭСК. Некоторые транскрипты, полученные из LTR, связаны с энхансерами и вовлечены в поддержание плюрипотентности [29]. Наконец, в 2016 г. было доказано ключевое значение ассоциированной с эндогенными ретровирусами GLN, MERVL, ERVK длинной некодирующей РНК *LincGET* в регуляции эмбриогенеза. Оказалось, что истощение днРНК *LincGET* у эмбриона мыши вызывает остановку развития на двухклеточной стадии. *LincGET* ассоциирована с эндогенными ретровирусами (ERV) GLN, MERVL, ERVK и формирует РНК-белковый комплекс с ILF2, FUBP1, hnRNP, вызывая ингибирование альтернативного сплайсинга РНК и активацию цис-регуляторной активности GLN, MERVL, ERVK [2].

В 2012 г. К. Lee и соавт. работали с гипотезой о том, что структурные изменения в геноме параллельны стадиям и органоспецифическим фенотипам в сочетании с дифференциальной транспозиционной активностью ретроэлементов на линии мышей C57BL/6J. Было показано, что геномы в клетках печени больше, чем в других органах, что совпадало с распространением МГЭ. При этом наблюдалось дифференцированное увеличение размера генома печени с возрастом, достигая пика в 5 нед [30]. Исследователи идентифицировали геномные библиотеки массивов МГЭ, предположив, что неслучайные конфигурации МГЭ говорят об их функции. Были протестированы особенности перегруппировки массивов МГЭ в зависимости от возраста и ткани и выявлены изменения в конфигурации массивов в геномах кожи и головного мозга всех мышей в возрасте 6 нед и старше, а также в сердце и печени в 29 нед. На основании полученных данных высказано предположение, что преобразования некоторых массивов МГЭ в пространстве и времени могут способствовать управлению характеристикой информационной системы генома [31].

#### **Активация транспозонов в стволовых клетках в постнатальном развитии**

Репрессия большинства МГЭ является не полной, в связи с чем их последовательности транскрибируются [32]. В 2005 г. были описаны ретротранспозиции LINE-1 нейрональных СК, выделенных из гиппокампа взрослых крыс. Полученные ретротранспозиционные события могли изменять экспрессию генов нейронов. Исследователи предположили, что данные события могут лежать в основе диверсификации соматических нейронов при формировании головного мозга [33]. Обнаруженная активация МГЭ в клетках, обладающих определенной степенью плюрипотентности, может говорить об их значении, подобно таковому в эмбриональных СК, в управлении дифференцировкой за счет регуляции транскрипции специфических генов хозяев. В 2009 г. было выявлено, что нейрональные СК,

изолированные из головного мозга плода и взятые из ЭСК человека, проявляли ретротранспозиции L1s *in vitro*, вносящие вклад в индивидуальный соматический мозаицизм [12].

Нейрогенез в гиппокампе находится под контролем генной регуляторной сети, ТФ, микроРНК, днРНК и различных сигнальных путей. Возрастает количество доказательств того, что днРНК, участвующие в регуляции генных сетей, контролирующих развитие и функционирование различных тканей, играют ключевую роль в нейрогенезе гиппокампа. Анализ экспрессии транскриптов в нервной системе показал обилие днРНК, проявляющих ограниченную в пространстве и динамичную во времени экспрессию [34]. Было выявлено 849 различных нкРНК, экспрессирующихся в головном мозге мышей, большинство из которых были ассоциированы со специфическими нейроанатомическими областями, типами клеток или субклеточными структурами. Сравнение профилей экспрессии нкРНК и связанных с ними белоккодирующих генов выявило сложные взаимосвязи, которые в сочетании со специфическими профилями экспрессии не согласуются с представлениями о том, что они представляют собой «транскрипционный шум» или артефакты ремоделирования хроматина [35]. Большинство днРНК перекрывают гены нейрогенеза и разделяют с ними почти идентичную модель экспрессии, предполагающую, что днРНК контролируют кортикогенез путем настройки экспрессии детерминантов соседних участков в клетках [13].

Достижения в геномике позвоночных позволили выявить тысячи локусов, кодирующих днРНК. Хотя был достигнут прогресс в выяснении регуляторных функций днРНК, мало известно об их происхождении и эволюции. При исследовании вклада МГЭ в состав и регулирование днРНК у людей, мыши и рыбы данио, было выявлено, что МГЭ встречаются более чем в  $2/3$  зрелых транскриптов днРНК и составляют значительную часть их общих последовательностей (около 30% у человека) [36]. Существует существенное межвидовое различие охвата и типов МГЭ, встроенных в днРНК. МГЭ вносят сигналы, необходимые для биогенеза многих днРНК, включая 30 000 уникальных сайтов инициации транскрипции, сплайсинга или полиаденилирования у человека. Подмножество МГЭ, встроенных в днРНК, подвергаются редактированию и формируют вторичные структуры, важные для функционирования, а многие зрелые днРНК полностью состоят из последовательностей МГЭ [36]. Около 83% доменов днРНК содержат хотя бы один фрагмент МГЭ [14]. Более того, сами МГЭ используются как гены днРНК, регулирующие дифференцировку клеток. Например, у человека HERVН экспрессируется в днРНК, которая поддерживает идентичность ЭСК [37].

Активация днРНК в нейрональных СК в гиппокампе [13, 34, 35] может иметь прямое отношение к обнаруженным в них явлениям активаций и транспозиций МГЭ [12, 33]. Было сделано предположение, что перемещения МГЭ в головном мозге могут создавать генетически различные субпопуляции клеток, за счет чего формируются особенности структуры и функции нейронов в разных структурах ЦНС [32]. Соматический мозаицизм по перемещениям L1-элементов выявлен в головном мозге не только человека и грызунов, но и в ЦНС дрозофилы [38]. Данный процесс считается потенциальным источником генотипических вариаций среди нейронов. При исследовании ретротранспозиций в единичных клетках при помощи RC-seq в индивидуальных нейронах гиппокампа и глиальных клетках, а так-

же нейронах коры головного мозга, было обнаружено 13,7 соматических инсерций на один нейрон гиппокампа. Инсерции L1 в нейронах гиппокампа специфически обогащали транскрибируемые в нейрональных СК энхансеры и гены гиппокампа, повышая вероятность их функциональной значимости [39]. Закономерные наследуемые перемещения МГЭ в нейрональных СК используются для управления дифференцировкой клеток, после чего активность МГЭ блокируется специфическими механизмами. К примеру, выявлен стабилизирующий белок SIRT6, являющийся мощным репрессором активности L1. Данный белок взаимодействует с 5'UTR L1-элемента и вызывает его упаковку в транскрипционно репрессированный гетерохроматин. Однако механизмы, сдерживающие дальнейшие перемещения МГЭ в зрелых дифференцированных клетках, не являются совершенными, что способствует вероятным новым, но незапрограммированным перемещениям МГЭ и может служить причиной старения [40].

### Транспозоны как источники регуляторных структур геномов

Регуляция МГЭ при помощи ТФ осуществляется за счет наличия в них специфических повторов, позволяющих взаимодействовать с ДНК-связывающими доменами. Данное свойство объясняет феномен одомашнивания МГЭ и сохранения в эволюции в связи с использованием их регуляторных последовательностей для регуляции экспрессии генов хозяев. Например, ТФ PLZF действует как эпигенетический регулятор в зародышевых и гемопоэтических СК. При этом основными мишенями для PLZF являются L1-ретротранспозоны. Опосредованное при помощи PLZF метилирование ДНК вызывает сайленсинг L1 полной длины и ингибирует ретротранспозиции [41]. Сравнение геномов 29 млекопитающих позволило обнаружить более 280 000 консервативных некодирующих элементов, произошедших из МГЭ с потенциально регуляторной функцией, многие из которых могут служить энхансерами. До 25% белок-кодирующих генов содержат МГЭ в их промоторах и нетранслируемых областях [42]. МГЭ играют важную роль в формировании регуляторной сети и хроматинового ландшафта их хозяев. МГЭ оказались источниками почти половины активных элементов в геноме человека, в открытых областях хроматина находятся около 44% МГЭ, и эта доля достигает 63% для областей, относящихся к приматам. Тысячи последовательностей, произошедших из МГЭ, активируют экспрессию соседних генов, специфичных для типов клеток [43]. Важной *cis*-регуляторной платформой для управления функционированием генома человека служат энхансеры, произошедшие из MIR МГЭ и являющиеся богатыми источниками сайтов связывания с ТФ (TFBS — transcription factor binding sites) [44].

Фундаментальным признаком плюрипотентности оказалась транскрипция промоторов ERV. У мыши и человека около 30% всех стартовых сайтов транскрипции оказались внутри МГЭ, характеризуясь тканеспецифическими паттернами экспрессии. Наибольшее количество и разнообразие стартовых сайтов транскрипции, ассоциированных с МГЭ, экспрессируют эмбриональные ткани. Данный факт позволяет сделать предположение о важной роли МГЭ в специфической регуляции экспрессии генов в зависимости от типа клеток и стадии развития [7]. Около 66% образованных из МГЭ сайтов связывания с ТФ являются специфичными для эпигенетического ландшафта типов клеток [4]. Геном человека содержит более полумиллиона LTR, которые мо-

гут рассматриваться как мобильные регуляторные модули. Многие из них были сохранены в эволюции в качестве элементов контроля транскрипции экспрессии клеточных генов. Было показано, что ERV LTR функционируют в определенных типах клеток и обладают потенциалом в качестве промоторов для конкретных типов клеток [15]. МГЭ обладают потенциалом перераспределения генной регуляции в зависимости от специфического контекста. Путем кодирования определенных сайтов связывания их эффекты могут быть ограничены тканевым контекстом экспрессии в соответствии с ДНК-связывающими молекулами [27].

На протяжении всей эволюции млекопитающих распространение МГЭ вызывало перераспределение сайтов связывания для ряда регуляторов транскрипции, включая факторы плюрипотентности OCT4 и NANOG, белки инсультаторы CTCF, нейронные репрессоры NRSF/REST и другие. Аналогично экспансия МГЭ MER20 и RLTR13D5 способствовала экспрессии в эндометрии и трофобласте плацентарной транскрипции генов, необходимых для развития беременности [45]. Оказалось, что использование МГЭ в качестве альтернативных промоторов происходит преимущественно в эмбриональных клетках. При раннем эмбриональном развитии, в частности, до 20% транскриптома инициируется из ретротранспозонов. Данные альтернативные промоторы проявляют склонность к тканеспецифической активности. Во многих случаях эти ретротранспозоны были кооптированы хозяином посредством экзонизации, при этом они транскрибируются и сплайсируются вниз по течению экзонов гена. В результате химерные транскрипты потенциально кодируют изоформы с профилями пространственных или временных ограничений экспрессии [7, 45]. У различных многоклеточных обнаружено множество LTR-содержащих TE, которые содержат регуляторные элементы, влияющие на транскрипцию генов хозяина. В геноме человека было выявлено 794 972 сайта связывания с ТФ, происходящих от HERV. Анализ кластеризации показал, что HERV/LTR могут быть сгруппированы в соответствии с шаблонами связывания с ТФ (плюрипотентности, гемопоэза) [46].

#### **Транспозоны как источники новых белков в эволюции и онтогенезе**

МГЭ, встроенные в интроны, используются для образования новых изоформ белков путем экзонизации. Наиболее активными в данном отношении являются SINE, содержащие множество латентных сплайсинговых сигналов. Например, в геноме человека не менее 5% всех альтернативных экзонов образуются путем экзонизации Alu [47]. Альтернативный сплайсинг экзонизированных МГЭ может проявлять тканеспецифические особенности экспрессии [48]. У человека некоторые экзоны, происходящие из Alu-элементов, могут проявлять строгое тканеспецифическое использование их транскриптов [49]. Еще в 2007 г. у мыши были выявлены тканеспецифические изоформы 28 генов, генерируемых путем экзонизации МГЭ [50]. Важным способом эволюции белков с возможностью использования новых вариантов при дифференцировке тканей оказалось альтернативное полиаденилирование 3'UTR. Оказалось, что более чем для 50% генов человека характерно множество изоформ с альтернативным 3'UTR, что способствует тканеспецифической и стадийспецифической регуляции экспрессии [51]. Экзонизация межгенных МГЭ путем введения новых терминальных экзонов и сайтов полиаденилирования имела значение в эволюции для пост-

транскрипционной регуляции в разных тканях и органах [47]. Сохранение данных изоформ в эволюции и их использование логичны: МГЭ, входящие в состав 3'UTR — канонических мишеней для связывания микроРНК у животных, становятся удобными системами саморегуляции, так как МГЭ служат источниками микроРНК [8, 9].

МГЭ играют ключевую роль в формировании ретрогенов путем дупликации. При этом ретрокопии содержат специфичные последовательности МГЭ, за счет которых они были получены. В результате создается возможность более тонкого регулирования новых генов, которые могут быть функциональными. Например, ретрокопии в геномах млекопитающих содержат специфичные для LINE-1 сайт расщепления эндонуклеазы TTTT/AA и поли(A) хвост во фланкирующих областях [52]. Таким образом, экзонизация и дупликация генов при помощи МГЭ являются источником возможностей управления экспрессии в последовательных клеточных делениях, отражаясь в эволюционном усложнении. Учитывая важную роль МГЭ в качестве источников регуляторных последовательностей, можно прийти к выводу, что у эукариот использование МГЭ геномами хозяев носит глобальный характер. При этом большинство универсальных характеристик транспозонов становятся присущими геномам их хозяев. В десятках исследований с применением строгих критериев функциональности было выявлено, что гены МГЭ были одомашнены геномами различных грибов, растений и животных. При этом транспозазы оказались источниками важнейших ДНК-связывающих белков и ТФ, в том числе участвующих в регуляции дифференцировки клеток при развитии целостного организма [53]. В крупномасштабном исследовании геномов различных позвоночных животных было выявлено, что более 1000 белок-кодирующих генов были одомашнены из МГЭ [54]. Интересно, что одомашненные гены МГЭ используются в важнейших молекулярных процессах, жизненно необходимых для функционирования генома хозяина. Это подчеркивает ключевую глобальную роль МГЭ в регуляции онтогенеза эукариот. Например, белки HDP1 и HDP2 произошли из белков ДНК-транспозона Harbinger и одомашнены хозяевами для использования в ацетилтрансферазном комплексе гистонов при деметилировании ДНК [55]. Предполагается, что фермент теломераза, имеющий значительную гомологию с обратной транскриптазой L1, появился путем одомашнивания генов МГЭ [23]. У дрозофилы белок инсультатор BEAF-32 полностью произошел из транспозазы hAT. У делящихся дрожжей центромерсвязывающие белки Abp1 и его два паралога, участвующие в сегрегации хромосом, произошли из *rogo*-подобных транспозаз [53].

#### **Транспозоны как источники эпигенетических факторов**

Исходя из того, что активность МГЭ необходима для поддержания плюрипотентности и меняется при дифференцировке СК, необходимо определить — являются ли МГЭ «водителями» данного процесса или находятся под влиянием более сложных эпигенетических процессов, источники которых до сих пор не известны. Ключом к разгадке данного вопроса является факт обнаружения важной роли МГЭ в качестве источников нкРНК, являющихся основными регуляторами эпигенетических настроек генома при дифференцировке клеток. Таким образом, можно предположить, что в природе существуют системы управления экспрессией клеток на эпигенетическом уровне, «водителями» которых являются МГЭ. Дан-

ные системы наиболее логично объясняют «эпигенетический код, не связанный с изменением последовательностей ДНК», как принято считать ряду современных авторов. Имеется параллелизм между преобразованиями генома под действием эпигенетических факторов и изменениями свойств клеток. Зигота и ее дочерние клетки тотипотентны. Геном зиготы характеризуется глобальным деметилированием, когда стираются эпигенетические метки. В последующих делениях ограничивается потенциал дифференцировки и устанавливается характерный для каждого из линий клеток паттерн метилирования, сопровождающийся реактивацией транскрипции МГЭ и происходящих из них нкРНК в раннем эмбриональном периоде [7]. Причиной реактивации МГЭ может быть их чувствительность к стрессу [25] и воздействию различных внутриклеточных и внеклеточных сигналов и гормональным воздействиям [26]. Можно предположить, что образующиеся при слиянии ядер сперматозоида и яйцеклетки сигнальные молекулы могут быть причиной последовательного каскада активаций МГЭ, ведущих к регулированию экспрессии специфических генов, что необходимо для формирования организма в целом. Видоспецифический набор и расположение МГЭ в геноме могут служить основой своеобразной «кодировки», позволяющей эпигенетически перепрограммировать характер экспрессии генов в зависимости от стадии развития и типа ткани. Составными элементами реализации данного «кода» могут быть сложные взаимосвязи происходящих от МГЭ регуляторов, ТФ, микроРНК и днРНК. Получены доказательства происхождения микроРНК от МГЭ [3, 8, 9]. Доказано, что МГЭ встречаются более чем в  $2/3$  всех транскриптов днРНК и составляют значительную часть их полных последовательностей [36], а некоторые МГЭ могут использоваться непосредственно как гены днРНК [37]. Длинные нкРНК играют ключевую роль в дифференцировке нормальных тканей путем их ТС экспрессии в транскриптоме человека. ТС днРНК путем спаривания оснований с различными мРНК связаны с тканевой дифференцировкой. Это свидетельствует о том, что тканевая специфичность — важный фактор, контролирующий взаимодействия днРНК-мРНК [56]. Выявлено непосредственное влияние МГЭ на метилирование генома [7]. В развитии многоклеточного организма микроРНК играют решающую роль в регуляции дифференцировки клеток. Транскрипционная активность микроРНК меняется тканеспецифично, вызывая перепрограммирование СК

[57]. МикроРНК регулируют экспрессию белоккодирующих генов не только посттранскрипционно, но и за счет регуляции гистоновых деацетилаз и ДНК-метилтрансфераз, специфически изменяя характер метилирования участков ДНК [58]. ДнРНК также функционируют в качестве эпигенетических регуляторов индивидуального развития путем связывания с гистонмодифицирующими комплексами, РНК-полимеразой II и ДНК-связывающими белками (в том числе ТФ) [59]. Учитывая важную роль МГЭ в качестве источников функциональных доменов днРНК, а также ДНК-связывающих белков в эволюции, можно предположить, что данные преобразования геномов под влиянием МГЭ являются универсальными механизмами формирования регуляторных сетей управления онтогенезом. Обнаружено, например, что Alu внедряются в гены днРНК, способствуя их взаимодействию с мРНК за счет образования коротких несовершенных спариваний нуклеотидов [60]. Тканевая специфичность экспрессии днРНК превышает таковую для белоккодирующих генов и характеризуется взаимосвязью с МГЭ в управлении дифференцировкой СК [5, 56].

## Заключение

Получены многочисленные данные, свидетельствующие о том, что МГЭ способствуют тканеспецифическому и стадийспецифическому управлению экспрессии генов, выступая в качестве источников сайтов связывания с ТФ, промоторов, инсультаторов и эпигенетических регуляторов (микроРНК, днРНК). Активация МГЭ на разных стадиях эмбрионального развития может носить обязательный характер, участвуя в регуляции важнейших процессов, необходимых для дифференцировки и последующего деления с участием эпигенетических факторов. При достижении предопределенного размера органов во взрослом организме активируются системы, подавляющие дальнейший каскад активаций МГЭ. Однако в стволовых клетках взрослых организмов активация определенных МГЭ необходима как для поддержания плюрипотентности, так и для специфической дифференцировки. Незапланированные активации МГЭ в зрелых дифференцированных клетках вследствие несовершенства систем, направленных на подавление транспозонов, могут быть причиной глобальных дегенеративных процессов, ведущих к старению.

**Автор заявляет об отсутствии конфликта интересов.**

## ЛИТЕРАТУРА/REFERENCES

1. Patel T, Hobert O. Coordinated control of terminal differentiation and restriction of cellular plasticity. *eLife*. 2017;6:e24100.
2. Wang J, Li X, Wang L, et al. A novel long intergenic noncoding RNA indispensable for the cleavage of mouse two-cell embryos. *EMBO Rep*. 2016;17:1452-1470.
3. Borchert GM, Holton NW, Williams JD, et al. Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. *Mobile Genetic Elements*. 2011;1:8-17.
4. Sundaram V, Cheng Y, Ma Z, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res*. 2014;24:1963-1976.
5. Ramsay L, Marchetto MC, Caron M, et al. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genomics*. 2017;18:214-226.
6. Мустафин Р.Н., Хуснутдинова Э.К. Некодирующие части генома как основа эпигенетической наследственности. *Вавиловский журнал генетики и селекции*. 2017;21:742-749.
7. Gerdes P, Richardson SR, Mager D., Faulkner GJ. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol*. 2016;17:100.
8. Gim J, Ha H, Ahn K, et al. Genome-Wide Identification and Classification of microRNAs derived from repetitive elements. *Genomic Inform*. 2014;12:261-267.
9. Lorenzetti AP, A de Antonio GY, Paschoal AR, Domingues DS. Plant TE-MIR DB: a database for transposable element-related microRNAs in plant genomes. *Funct Integr Genomics*. 2016;16:235-242.
10. Macia A, Munoz-Lopez M, Cortes JL, et al. Epigenetic control of retrotransposons expression in human embryonic stem cells. *Mol Cell Biol*. 2011;31:300-316.

11. Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012;487:57-63.
12. Coufal NG, Garcia-Perez J, Peng GE, et al. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009;460:1127-1131.
13. Aprea J, Penninger S, Dori M, et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*. 2013;32:3145-3160.
14. Johnson R, Guigo R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA*. 2014;20:959-976.
15. Conley AB, Piriyaopongsa J, Jordan IK. Retroviral promoters in the human genome. *Bioinformatics*. 2008;24:1563-1567.
16. Marchetto MC, Narvaiza I, Denli AM, et al. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature*. 2013;503:525-529.
17. Packer A, Manova K, Bachvarova RF. A discrete LINE-1 transcript in mouse blastocysts. *Dev Biol*. 1993;157:281-283.
18. Branciforte D, Martin SL. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol*. 1994;14:2584-2592.
19. Trelogan SA, Martin SL. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci USA*. 1995;92:1520-1524.
20. Moran JV, Holmes SE, Naas TP. High frequency retrotransposition in cultured mammalian cells. *Cell*. 1996;87:917-927.
21. De Berardinis RJ, Goodier JL, Ostertag EM, Kazazian HH. Rapid amplification of a retrotransposons subfamily is evolving the mouse genome. *Nat Genet*. 1998;20:288-290.
22. Wei W, Morrish TA, Alisch RS. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal Biochem*. 2000;284:435-438.
23. Prak ET, Dodson AW, Farkash EA, Kazazian HH. Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci USA*. 2003;100:1832-1837.
24. Garcia-Perez JL, Marchetto MC, Muotri AR, et al. LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet*. 2007;16:1569-1577.
25. Grandbastien MA. LTR retrotransposons, handy hitchhikers of pant regulation and stress response. *Biochim Biophys Acta*. 2015;1849:403-416.
26. Lu Y, Feng F, Yang Y, et al. LINE-1 ORF-1p functions as a novel androgen receptor co-activator and promotes the growth of human prostatic carcinoma cells. *Cell Signal*. 2013; 25:479-489.
27. Pavlicev M, Hiratsuka K, Swaggart KA, et al. Detecting endogenous retrovirus-driven tissue-specific gene transcription. *Genome Biol Evol*. 2015;7:1082-1097.
28. Klawitter S, Fuchs NV, Upton KR, et al. Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat Commun*. 2016;7:10286-10301.
29. Fort A, Hashimoto K, Yamada D, et al. Deep transcriptome profiling of mammalian stem cells supports maintenance. *Nat Genet*. 2014;46:558-566.
30. Lee KH, Chiu S, Lee YK, et al. Age-dependent and tissue-specific structural changes in the C57BL/6J mouse genome. *Exp Mol Pathol*. 2012;93:167-172.
31. Lee KH, Yee L, Lim D, et al. Temporal and spatial rearrangements of a repetitive element array on C57BL/6J mouse genome. *Exp Mol Pathol*. 2015;98:439-445.
32. Faulkner GJ. Retrotransposons: mobile and mutagenic from conception to death. *FEBS Lett*. 2011;585:1589-1594.
33. Muotri AR, Chu VT, Marchetto MC, et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005;435:903-910.
34. Deng B, Cheng X, Li H, et al. Microarray expression profiling in the denervated hippocampus identifies long noncoding RNAs functionally involved in neurogenesis. *BMC Mol Biol*. 2017;18:15.
35. Mercer TR, Dinger ME, Sunkin SM, et al. Specific expression of long non-coding RNAs in the mouse brain. *Proc Natl Acad Sci USA*. 2008;105:716-721.
36. Kapusta A, Kronenberg Z, Lynch VJ, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*. 2013;9:e1003470.
37. Lu X, Sachs F, Ramsay L, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*. 2014;21:423-425.
38. Richardson SR, Morell S, Faulkner GJ. L1 retrotransposons and somatic mosaicism in the brain. *Annu Rev Genet*. 2014;48:1-27.
39. Upton KR, Gerhardt DJ, Jesuadian JS, et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell*. 2015;161:22-39.
40. Van Meter M, Kashyap M, Rezazadeh S, et al. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun*. 2014;5:5011.
41. Puszyk W, Down T, Grimwade D, et al. The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. *EMBO J*. 2013;32:1941-1952.
42. De Souza FS, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong. *Mol Biol Evol*. 2013;30:1239-1251.
43. Jacques PE, Jeyakani J, Bourgue G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet*. 2013;9:e1003504.
44. Jjingo D, Conley AB, Wang J, et al. Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob DNA*. 2014;5:14.
45. Mak KS, Burdach J, Norton LJ, et al. Repression of chimeric transcripts emanating from endogenous retrotransposons by a sequence-specific transcription factor. *Genome Biol*. 2014;15:R58.
46. Wang J, Xie G, Singh M, et al. Primate-specific endogenous retrovirus-driven transcription defines naïve-like stem cells. *Nature*. 2014;516:405-409.
47. Tajnik M, Vigilante A, Braun S, et al. Inergetic Alu exonisation facilitates the evolution of tissue-specific transcript ends. *Nucleic Acids Res*. 2015;43:10492-10505.
48. Sela N, Kim E, Ast G. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biol*. 2010;11:6:R59.
49. Lin L, Shen S, Tye A, et al. Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet*. 2008;4:e1000225.
50. Mersch B, Sela N, Ast G. SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements. *BMC Genet*. 2007;8:78.
51. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*. 2013;14:496-506.
52. Tan S, Cardoso-Moreira M, Shi W, et al. LTR-mediated retroposition as a mechanism of RNA-based duplication in metazoans. *Genome Res*. 2016;26:1663-1675.
53. Feschotte C. The contribution of transposable elements to the evolution of regulatory networks. *Nat Rev Genet*. 2008;9:397-405.
54. Zdobnov EM, Campillos M, Harrington ED, et al. Protein coding potential of retroviruses and other transposable elements in vertebrate genomes. *Nucleic Acids Res*. 2005;33:946-954.
55. Duan CG, Wang X, Pan L, et al. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. *Cell Res*. 2017;27:226-240.
56. Iwakiri J, Terai G, Hamada M. Computational prediction of lncRNA-mRNA interactions by integrating tissue specificity in human transcriptome. *Biol Direct*. 2017;12:15.
57. Du Z, Yang C, Rothschild MF, Ross J. Novel microRNA families expanded in the human genome. *BMC Genomics*. 2013;14:98-105.
58. 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;11:315-328.
59. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription. *Sci Adv*. 2017;3:eao2110. <https://doi.org/10.1126/sciadv.aao2110>
60. Hadjiargyrou M, Delilhas N. The Intertwining of Transposable Elements and Non-Coding RNAs. *Int J Mol Sci*. 2013;14:13307-13328.

Поступила 18.02.18

Received 18.02.18

После доработки 18.02.18

Revised 18.02.18

Принята к публикации 15.03.18

Accepted 15.03.18