

УДК 577.218

РОЖДЕНИЕ НОВОЙ ИЗОФОРМЫ POU2F1 У ПРИМАТОВ ЗА СЧЕТ ИСПОЛЬЗОВАНИЯ МОБИЛЬНЫХ ГЕНЕТИЧЕСКИХ ЭЛЕМЕНТОВ

© 2022 г. Б. М. Льянова¹, А. П. Котнова^{1,*}, А. А. Макарова¹, академик РАН Ю. В. Ильин¹, член-корреспондент РАН С. Г. Георгиева¹, А. Г. Степченко¹, Е. В. Панкратова¹

Поступило 14.12.2021 г.

После доработки 17.01.2022 г.

Принято к публикации 17.01.2022 г.

Появление новых генов и функций имеет первостепенное значение в возникновении новых видов животных. Так, инсерция мобильного элемента Tigger 2 в последовательность функционального гена *POU2F1* у приматов привела к образованию новой химерной примат-специфичной изоформы *POU2F1Z*, трансляция которой активируется при клеточном стрессе. Ее мРНК обнаружена для всех видов обезьян, начиная с макака. Анализ фрагментов копии Tigger2, соответствующих человеческому экзону Z, показал, что сайты сплайсинга экзона Z гомологичны у человека и у большинства обезьян, за исключением лемуров и галаго. Стоп-кодон, привнесенный в мРНК последовательностью Tigger2, присутствует у всех приматов, начиная с макака. Внутренний ATG-кодон также есть у всех приматов, за исключением лемуров и галаго. В процессе эволюции в копию Tigger2 встраивались другие МГЭ, преимущественно типа SINE. В ходе эволюции изменялось как место расположения, так и численность мобильных элементов SINE в пределах гена *POU2F1*. Начиная с макака, паттерн расположения SINE-элементов внутри копии Tigger2 в исследуемом участке гена *POU2F1* закрепился и далее сохранялся в неизменном виде у остальных приматов и человека, что может свидетельствовать о его функциональной значимости.

Ключевые слова: ген *POU2F1*, мобильные генетические элементы, карцинома головы

DOI: 10.31857/S2686738922020147

POU2F1 сверхэкспрессируется во многих злокачественных опухолях и опухолевых клеточных линиях. При этом высокий уровень белка *POU2F1* коррелирует с крайне неблагоприятным прогнозом течения заболевания [1–7]. Так, в работе Sharpe и соавт. [3] показан повышенный уровень экспрессии *POU2F1* в клеточных линиях карциномы головы и шеи по сравнению с образцами неопухолевых клеток тех же тканей у тех же пациентов. Нокдаун *POU2F1* приводил к существенному снижению пролиферации и роста в клеточных линиях этих типов опухолей. Показано, что *POU2F1* регулирует регенерацию и опухолевую трансформацию в клеточной линии HCT116 (клетки опухоли прямой кишки) [5] и других опухолей [4, 6–9].

Интересно, что уровень обнаруженного белка *POU2F1*, зачастую не коррелирует с уровнем соответствующей мРНК. Так, например, в костном

мозге обнаружено высокое содержание мРНК *POU2F1*, но соответствующего белка в этой ткани незначительное количество. В прямой кишке, наоборот, количество соответствующей мРНК сравнительно небольшое, а количество белка – одно из самых высоких. Это возможно, если экспрессия *POU2F1* регулируется в значительной степени на уровне трансляции.

Нами была обнаружена новая изоформа *POU2F1Z*, трансляция которой многократно возрастает при клеточном стрессе. Одной из функций этой изоформы является защита клеток от повреждающего действия различных видов стресса [10]. мРНК *POU2F1Z* экспрессируется во всех тканях и клетках человека, но на максимальном уровне она обнаруживается в коре головного мозга и в лимфоидных клетках. В частности, высокий уровень экспрессии белка *POU2F1Z* был обнаружен нами в опухолевых клетках лимфомы Беркитта Namalwa. Интересно, что у человека обнаружены сразу две мРНК (MT294127 и AK091438.1), с которых транслируется один и тот же белок *POU2F1Z* (UniProt KBP14859-3). Эти мРНК транскрибируются с двух альтернативных промоторов гена *POU2F1* и отличаются 5-концевыми последовательностями. По сравнению с хорошо

¹Федеральное государственное бюджетное учреждение науки Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук (ИМБ РАН), Москва, Россия

*e-mail: alina_kotnova@mail.ru

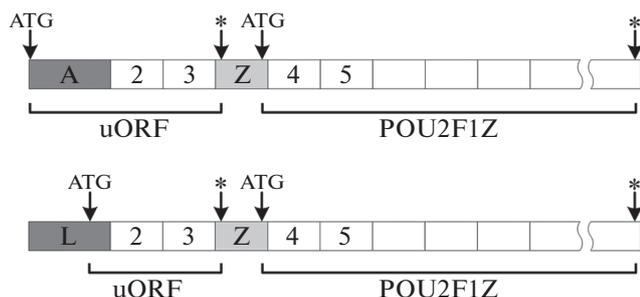


Рис. 1. Схема строения двух изоформ мРНК POU2F1, с которых транслируется одинаковый белок POU2F1Z. Экзон Z формируется из последовательности транспозона Tigger2. uORF – короткая открытая рамка считывания на 5'-конце мРНК, трансляция которой начинается с ATG-кодов альтернативных экзонов A(U) или L. Звездочкой обозначены стоп-кодона. Начало трансляции изоформы POU2F1Z обозначено ATG в области экзона Z.

изученными изоформами *POU2F1A* и *POU2F1L*, *POU2F1Z* содержит дополнительный экзон Z, расположенный между экзонами 3 и 4. При сплайсинге этот экзон привносит в последовательность мРНК стоп-кодон, а следом – ATG-кодон, совпадающий с “основной” рамкой считывания белка POU2F1. Для получившейся мРНК с этого привнесенного триплета ATG синтезируется изоформа POU2F1Z с новым N-концевым доменом, а с ATG-кодона, расположенного в первом экзоне, синтезируется короткий пептид (рис. 1).

В нормальных условиях синтез этого короткого пептида (uORF) ингибирует трансляцию последующей длинной рамки считывания (белка POU2F1Z), но в условиях клеточного стресса синтез короткого пептида подавляется и активируется трансляция белка POU2F1Z [10].

Ранее мы показали, что активация трансляции POU2F1Z начинается под действием повреждающих агентов и защищает клетки от гибели [10]. Мы предполагаем, что такой механизм защиты может существенно повышать устойчивость раковых клеток к действию химиотерапевтических препаратов и немедикаментозным способам лечения злокачественных опухолей.

В данной работе мы исследовали возникновение изоформы POU2F1Z в процессе эволюции, а также провели поиск и сравнительный анализ этой изоформы у животных подтипа Craniata.

В рамках исследования происхождения экзона Z мы провели выравнивание его последовательности с геномами различных млекопитающих и обнаружили в хромосомах у разных видов довольно много мест встраивания последовательностей, имеющих значимую гомологию с экзонами Z. Мы предположили, что экзон Z может представлять собой мобильный элемент, и на следующем этапе прогнали его последовательность через базу дан-

ных мобильных элементов в GeneBank NCBI. Мы обнаружили, что экзон Z, с которого транскрибируется изоформа POU2F1Z, представляет собой фрагменты копии ДНК-транспозона Tigger2, встроившегося в ген POU2F1 между 3 и 4 экзонами, и соответствует участку 2150 – 2461 п.н. данного транспозона.

Tigger2 – это древний ДНК транспозон, который обнаруживают в геномах млекопитающих, в том числе человека. Длина полного реконструированного варианта составляет 2718 п.н. Он включает ген транспозазы (с 603 по 2483 п.н.) и фланкирован инвертированными повторами длиной 24 п.н. Полный вариант Tigger 2 в геномах млекопитающих не был обнаружен, однако повсеместно обнаруживаются различные его фрагменты [11].

Далее мы провели исследование с целью определить, в какой именно момент эволюционного развития животных подтипа Craniata Tigger 2 встроился в ген *POU2F1*.

Для этого был произведен анализ последовательностей ортологов *POU2F1* животных подтипа Craniata, содержащихся в базе данных NCBI. Всего нами было проанализировано 279 видов. Из базы данных Dfam [12] была получена скрытая марковская модель для транспозона Tigger2. При помощи HMMER (ver.3.3.2) [13] был произведен поиск фрагментов Tigger2 в полученных из NCBI последовательностях *POU2F1*. Для данного анализа мы посчитали необходимым использовать именно метод поиска, основанный на скрытых марковских моделях, реализуемый в HMMER, поскольку показано, что он более чувствителен (по сравнению с blastn и другими методами, основанными на поиске гомологий с единственной последовательностью) в задачах поиска фрагментов древних транспозонов [14].

Из базы NCBI были получены файлы с аннотациями референсных геномов попавших в нашу выборку видов. Из полученных аннотаций были извлечены фрагменты, соответствующие ортологам POU2F1. Полученные аннотации, а также результаты HMMER были наложены на соответствующие референсные сиквенсы генов.

Мы обнаружили, что изоформа POU2F1Z является примат-специфической и отсутствует у других видов животных. Кроме того, в отличие от остальных приматов, Tigger2 не был обнаружен в гене *POU2F1* долгопятов.

Мы показали, что во всех исследуемых генах *POU2F1* приматов встроена копия Tigger2 разбита на несколько частей и имеет делецию в области 1341 (–44) – 2119 п.н. В генах *POU2F1* лемуру и галаго делетированная копия Tigger2 разбита на три части, тогда как в генах *POU2F1* остальных приматов – на четыре. Известно, что кодирующая область транспозона Tigger2 расположена на

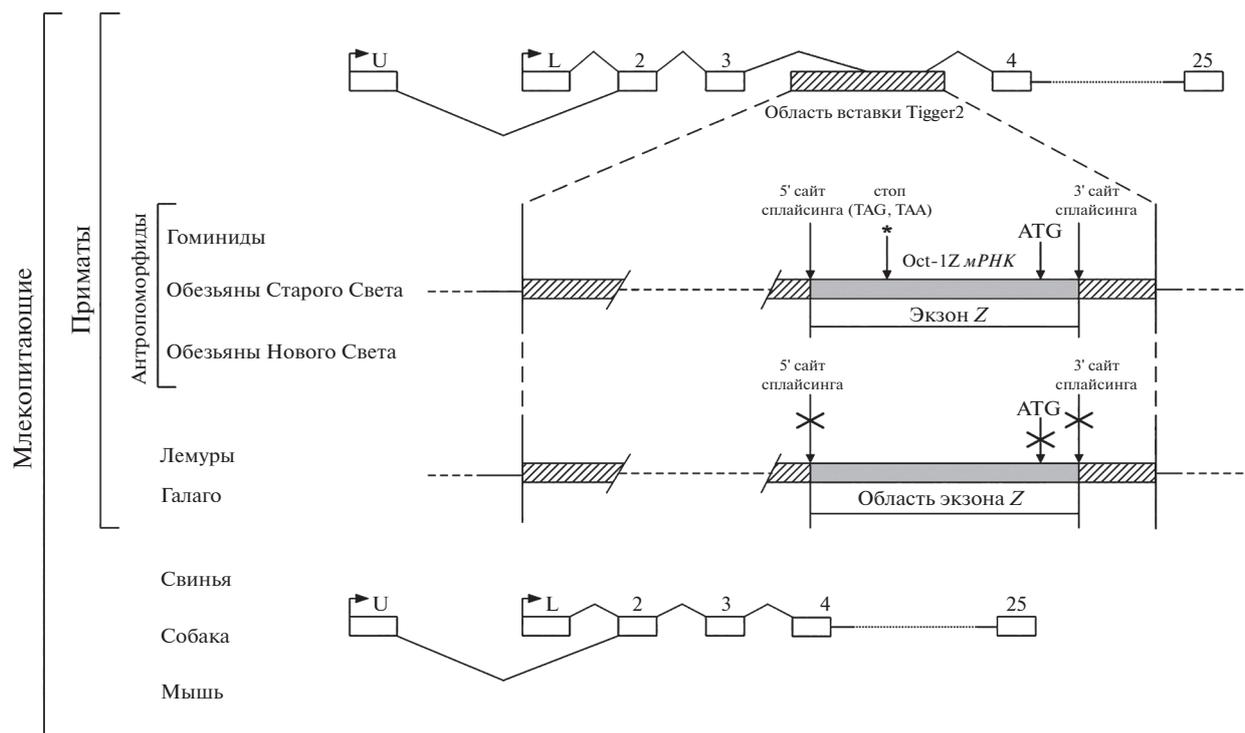


Рис. 2. Фрагмент транспозона Tigger2, встроенный в ген *POU2F1*. Экзоны обозначены прямоугольниками. Стрелками обозначены старты транскрипции. Линиями обозначены участки места сплайсинга. У всех приматов фрагмент Tigger2 содержит делецию (обозначено пунктиром). У всех приматов, кроме лемуров и галаго, фрагмент Tigger2 содержит сайты сплайсинга, стоп-кодон и ATG-кодон. У лемуров и галаго, несмотря на наличие фрагмента Tigger2 в гене *POU2F1*, образование изоформы мРНК *POU2F1Z* невозможно, также как и у других млекопитающих, у которых в гене *POU2F1* отсутствует транспозон Tigger2.

участке 603–2483 пн, из чего следует, что попавшие в состав гена участки Tigger2 не могут кодировать полноценную транспозазу, однако могут сохранять свойства ее отдельных доменов.

На сегодняшний день в базах данных для всех приматов имеются только смоделированные версии последовательностей мРНК *POU2F1* и транскрируемых с них белков. Однако, по данным моделирования, мРНК и белок изоформы *POU2F1Z* экспрессируются уже у макаков – низших представителей обезьян Старого Света.

В дальнейшем для того, чтобы определить, происходит ли экспрессия мРНК и белка изоформы *POU2F1Z* в организмах приматов, мы анализировали участок копии Tigger2, встроенный в ген *POU2F1* приматов и соответствующий человеческому экзону Z.

При анализе сайтов сплайсинга экзона Z человека мы выяснили, что среди обезьян соответствующие последовательности полностью гомологичны последовательностям сайтов сплайсинга экзона Z человека, за исключением ночных обезьян, у которых в акцепторном сайте предположительного экзона Z присутствует одна нуклеотидная замена. У лемуров и галаго имеются значительные отличия от сайтов сплайсинга экзона

Z человека, и можно предположить, что несмотря на наличие фрагмента Tigger2 в этой области гена, в мРНК этот участок не попадает, так как альтернативный сплайсинг этой области невозможен (рис. 2).

Также в исследуемом участке встроенной копии Tigger2 анализировали последовательности предположительных стоп-кодона и ATG-кодона на соответствие таковым в экзоне Z человека. Выяснилось, что стоп-кодоны TAG на данном участке полностью гомологичны для всех приматов, за исключением макаков, у которых предположительным соответствующим стоп-кодоном является TAA (рис. 2).

Соответствующие точки начала трансляции (ATG-кодон) присутствуют у всех приматов, за исключением лемуров и галаго (рис. 2), у которых вместо ATG-кодона присутствуют триплеты CTC и GTG соответственно.

Чтобы понять, какие именно события вызвали разделение копии Tigger2 внутри гена *POU2F1* на несколько частей, с помощью базы данных мобильных элементов Dfam, был проанализирован участок гена *POU2F1* приматов со встроенной копией транспозона Tigger2. В результате мы обнаружили, что в процессе эволюции в копию

Tigger2 встраивались другие мобильные генетические элементы, преимущественно типа SINE. В ходе эволюции изменялось как место расположения, так и численность мобильных элементов SINE в пределах исследуемого участка. Указанные изменения происходили в исследуемых генах *POU2F1* низших приматов вплоть до макаков. Расположение мобильных генетических элементов SINE, установившееся в исследуемом участке гена *POU2F1* макаков, сохраняется в неизменном виде у остальных приматов – обезьян Старого Света и человека, что может свидетельствовать о его функциональной значимости.

Можно предположить, что именно на данном этапе эволюции произошло попадание участка 2150–2461 п.н. Tigger2 в кодирующую последовательность гена *POU2F1*, что повлекло за собой положительные эволюционные изменения и, в итоге, привело к образованию экзона Z.

Наши результаты показывают, что инсерция мобильного элемента стала частью функционального гена *POU2F1* приматов благодаря пошаговому процессу, который включал в себя транспозицию, утрату способности к дальнейшим перемещениям за счет встраивания внутрь Tigger2 других мобильных элементов, и последующее транскрипционное и трансляционное слияние с геном-мишенью. В нашем исследовании было показано, что новая химерная примат-специфичная изоформа *POU2F1Z* получилась в результате слияния гена фактора транскрипции *POU2F1* и фрагмента гена транспозазы мобильного элемента Tigger 2, давшего начало новому экзону гена *POU2F1* приматов.

Появление новых генов и функций имеет первостепенное значение в формировании новых видов животных. В настоящее время достаточно хорошо изучены различные типы дупликаций в образовании новых генов [15]. Гораздо менее понятным является создание новых генов за счет использования материала из “эгоистичных” мобильных генетических элементов. Анализ подобных эволюционных событий – сложная задача, потому что зачастую это результат небольших замен или структурных перестроек, функциональное значение которых трудно определить.

В случае гена *POU2F1* человека и приматов такое преобразование привело к возникновению новой изоформы *POU2F1Z*, которая защищает клетки от повреждающего действия клеточного стресса, но при этом создает трудности для лечения злокачественных новообразований.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена при поддержке гранта российского научного фонда (грант № 19-14-00365).

СПИСОК ЛИТЕРАТУРЫ

1. *Vázquez-Arreguín K., Tantin D.* The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks // *Biochim Biophys Acta*. 2016. V. 1859. № 6. P. 792–804.
2. *Maddox J., Shakya A., South S., et al.* Transcription factor Oct1 is a somatic and cancer stem cell determinant // *PLoS Genet*. 2012. V. 8. P. e1003048.
3. *Sharpe D.J., Orr K.S., Moran M., et al.* POU2F1 activity regulates HOXD10 and HOXD11 promoting a proliferative and invasive phenotype in head and neck cancer // *Oncotarget*. 2014. V. 5. № 18. P. 8803–8815.
4. *Qian J., Kong X., Deng N., et al.* OCT1 is a determinant of synbindin-related ERK signalling with independent prognostic significance in gastric cancer // *Gut*. 2015. V. 64. P. 37–48.
5. *Vázquez-Arreguín K., Bensard C., Schell J.C., et al.* Oct1/Pou2f1 is selectively required for colon regeneration and regulates colon malignancy // *PLoS Genet*. 2019. V. 15. № 5. P. e1007687.
6. *Xiao S., Liao S., Zhou Y., Jiang B., Li Y., Xue M.* High expression of octamer transcription factor 1 in cervical cancer // *Oncol Lett*. 2014. V. 7. P. 1889–1894.
7. *Kuzmanov A., Johansen P., Hofbauer G.* FBXO25 Promotes Cutaneous Squamous Cell Carcinoma Growth and Metastasis through Cyclin D1 // *J. Invest Dermatol*. 2020. V. 140. P. 2496–2504.
8. *Pankratova E.V., Stepchenko A.G., Portseva T., Mogila V.A., Georgieva S.G.* Different N-terminal isoforms of Oct-1 control expression of distinct sets of genes and their high levels in Namalwa Burkitt’s lymphoma cells affect a wide range of cellular processes // *Nucleic Acids Res*. 2016. V. 44. № 19. P. 9218–9230.
9. *Pankratova E.V., Stepchenko A.G., Krylova I.D., Portseva T.N., Georgieva S.G.* The regulatory interplay between Oct-1 isoforms contributes to hematopoiesis and the isoforms imbalance correlates with a malignant transformation of B cells // *Oncotarget*. 2018. V. 9. № 52. P. 29892–29905.
10. *Stepchenko A.G., Portseva T.N., Glukhov I.A., Kotnova A.P., Georgieva S.G., Pankratova E.V.* Primate-specific stress-induced transcription factor POU2F1Z protects human neuronal cells from stress // *Sci. Rep*. 2021. V. 11. P. 18808.
11. *Smit A.F., Riggs A.D.* Tiggers and DNA transposon fossils in the human genome // *Proc. Natl. Acad. Sci. U S A*. 1996. V. 93. № 4. P. 1443–1448.
12. *Storer J., Hubley R., Rosen J., Wheeler T.J., Smit A.F.* The Dfam community resource of transposable element families, sequence models, and genome annotations // *Mob. DNA* 2021. V. 12. № 1. P. 2.
13. *Wheeler T.J., Eddy S.R.* Hmmer: DNA homology search with profile HMMs // *Bioinformatics*. 2013. V. 29. № 19. P. 2487–2489.
14. *Wheeler T.J., Clements J., Eddy S.R., et al.* Dfam: a database of repetitive DNA based on profile hidden Markov models // *Nucleic Acids Res*. 2013. P. 41.
15. *Bizzotto S., Walsh C.A.* Making a Notch in the Evolution of the Human Cortex // *Dev. Cell*. 2018. V. 45. № 5. P. 548–550.

THE BIRTH OF A NEW ISOFORM OF POU2F1 IN PRIMATES THROUGH THE USE OF EGOISTIC MOBILE GENETIC ELEMENTS

B. M. Lyanova^a, A. P. Kotnova^a, A. A. Makarova^a, Academician of the RAS Yu. V. Ilyin^a, Corresponding Member of the RAS S. G. Georgieva^a, A. G. Stepchenko^a, and E. V. Pankratova^{a,#}

^a*Engelhardt Institute of Molecular Biology, RAS, Moscow, Russian Federation*

[#]*e-mail: panliz@mail.ru*

The emergence of new genes and functions is of paramount importance in the emergence of new animal species. Thus, the insertion of the mobile element Tigger 2 into the sequence of the functional gene POU2F1 in primates led to the formation of a new chimeric primate-specific isoform POU2F1Z, the translation of which is activated under cellular stress. Its mRNA has been found in all types of monkeys, starting with monkeys. Analysis of the fragments of the Tigger2 copy corresponding to the human exon Z showed that the splicing sites of exon Z are homologous in humans and in most monkeys, with the exception of lemurs and galago. The stop codon introduced into the mRNA by the Tigger2 sequence is present in all primates, starting with monkeys. The internal ATG codon is also present in all primates, with the exception of lemurs and galagos. In the course of evolution, other MGEs, mainly of the SINE type, were built into the Tigger2 copy. In the course of evolution, both the location and the number of mobile SINE elements within the POU2F1 gene changed. Starting with macaques, the pattern of the arrangement of SINE elements within the Tigger2 copy in the studied region of the POU2F1 gene was fixed and then remained unchanged in other primates and humans, which may indicate its functional significance.

Keywords: POU2F1 gene, mobile genetic elements, head carcinoma