

УДК 577.21

мРНК *PRPF19* КОДИРУЕТ МАЛУЮ ОТКРЫТУЮ РАМКУ СЧИТЫВАНИЯ, КОТОРАЯ ВАЖНА ДЛЯ ЖИЗНЕСПОСОБНОСТИ КЛЕТОК ЧЕЛОВЕКА

© 2024 г. Н. М. Шепелев^{1, 2}, А. О. Курочкина²,
академик РАН О. А. Донцова^{1, 2, 3, 4}, М. П. Рубцова^{1, 2, *}

Поступило 18.12.2023 г.

После доработки 25.12.2023 г.

Принято к публикации 28.12.2023 г.

Высокопроизводительное рибосомное профилирование продемонстрировало трансляцию тысяч малых открытых рамок считывания, расположенных в 5'-нетранслируемых областях матричных РНК (вышележащих ОРС). Вышележащая ОРС может как выполнять регуляторную функцию за счет влияния на трансляцию нижележащей основной ОРС, так и кодировать малый функциональный белок или микробелок. В данной работе мы показали, что 5'-нетранслируемая область мРНК *PRPF19* содержит вышележащую ОРС, которая транслируется в клетках человека. Инактивация данной вышележащей ОРС снижает жизнеспособность клеток человека.

Ключевые слова: рибосомное профилирование, вышележащая ОРС, вОРС, микробелок, микропептид.

DOI: 10.31857/S2686738924020079, **EDN:** WFODRJ

ВВЕДЕНИЕ

Изобретение метода высокопроизводительного рибосомного профилирования позволило обнаружить трансляцию множества молекул РНК, а также участков матричных РНК (мРНК), которые прежде считались некодирующими [1]. Последующие исследования раскрыли важную роль многих продуктов трансляции “некодирующих” РНК [2–4]. Интересным примером “некодирующих” участков мРНК являются малые открытые рамки считывания (малые ОРС), расположенные в их 5'-нетранслируемой области (5'-НТО). Далее мы будем называть их вышележащие ОРС по аналогии с upstream ORF или uORF в англоязычной литературе [1]. Особенность функционирования вышележащих ОРС заключается в том, что они могут как регулировать экспрессию гена, влияя на трансляцию нижележащей основной ОРС гена, так и кодировать

малые функциональные белки, роль которых может быть напрямую не связана с функцией кодируемого геном белка [5, 6].

Одним из известных генов, содержащих регуляторные вышележащие ОРС, является *ATF4*, кодирующий главный фактор транскрипции, необходимый для восстановления клеток в условиях стресса [7]. мРНК гена *ATF4* обладает двумя вышележащими ОРС, консервативными среди млекопитающих, которые подавляют трансляцию основной ОРС в нормальных условиях, но при этом способствуют ее трансляции при различных стрессах [8].

Интересный пример вышележащей ОРС расположен в гене *ASNSD1*, вовлеченном в биосинтез аспарагина. Данная вышележащая ОРС кодирует малый белок *ASDURF* [9]. *ASDURF* является одной из 12 субъединиц кошаперонного комплекса PAQ-сомы [9], который вовлечен в биогенез ряда белковых комплексов [10].

В работах по полногеномному скринингу генов человека при помощи системы *CRISPR-Cas9*, выявлен ген *PRPF19*, который критически необходим для жизнеспособности клеток различных линий [11, 12]. Белок *PRPF19* представляет собой высококонсервативный фактор сплайсинга, который включен в состав комплекса *Prp19/NTC* [13]. Основную функцию этого комплекса связывают с процессингом РНК и ответом на повреждение

¹Институт биоорганической химии имени академиков М. М. Шемякина и Ю. А. Овчинникова РАН, Москва, Россия

²Московский государственный университет имени М. В. Ломоносова, химический факультет, Москва, Россия

³Научно-исследовательский институт физико-химической биологии имени А. Н. Белозерского, Москва, Россия

⁴Сколковский институт науки и технологий, центр молекулярной и клеточной биологии, Москва, Россия

*E-mail: mprubtsova@gmail.com

ДНК [13]. В данной работе, чтобы выяснить роль вышележащей ОРС в мРНК гена *PRPF19* мы проверили, может ли транслироваться в клетках данная вышележащая ОРС при помощи репортерной конструкции, и изучили, как ее инактивация влияет на жизнеспособность клеток.

МАТЕРИАЛЫ И МЕТОДЫ

Культуры клеток. Клетки *HEK293T* (производная линия из почки эмбриона человека) и *HAP1* (практически гаплоидная производная линия из хронического миелолейкоза человека) растили в среде *DMEM/F12* и *IMDM* соответственно (Gibco) с добавлением аланил-глутамин (GlutaMAX, Gibco), 10%-й эмбриональной телячьей сыворотки (FBS, Gibco) в присутствии 100 единиц/мл пенициллина и 100 мкг/мл стрептомицина (Gibco) при 37 °C в атмосфере 5% CO₂. Конфлюентность культур исследовали с помощью инвертированного микроскопа. Клетки были проверены на отсутствие микоплазмы (MycroReport, «Евроген»).

Клетки трансфицировали согласно протоколу производителя, используя набор Lipofectamine 3000 Transfection Reagent (Invitrogen). С помощью этого же набора получали лентивирусные частицы согласно примечанию по применению от производителя.

Для анализа конкуренции клетки трансдуцировали вирусными частицами, кодирующими геновую РНК/*Cas9* и ген зеленого флуоресцентного белка (*EGFP*), примерно с 50%-й эффективностью и оценивали процент выживших клеток с флуоресценцией *EGFP* на проточном цитометре MACSQuant Analyser (Miltenyi Biotec) в течение нескольких суток.

Быстрая амплификация 5'-концов кДНК (5'RACE) 5'RACE проводили согласно протоколу производителя (5'RACE System for Rapid Amplification of cDNA Ends, version 2.0, Invitrogen). РНК выделяли из клеток *HEK293T* при помощи набора PureLink RNA Mini Kit (Invitrogen). Для получения кДНК и последующей амплификации использовали олигонуклеотиды 5'-GTCTTCCCTCTCTTCTTGC-3' и 5'-GGTTAGCACAGTGGCTTTGTC-3' соответственно.

Плазмиды и конструкции. Плазмида для проверки трансляции вышележащей ОРС сделана на основе вектора pHaloTag-*EGFP* (Addgene #86629), в котором последовательность HaloTag заменили на последовательность, соответствующую 5'-концу основной изоформы мРНК гена *PRPF19* со вставкой последовательности HiBiT перед стоп-кодonom вышележащей ОРС.

Плазмида для внесения случайных мутаций с помощью *CRISPR-Cas9* сделана на основе вектора *LeGO-iG2* (Addgene #27341), в которую перед

SFFV промотором поместили последовательность из вектора *lentiGuide-Puro* (Addgene #52963) от U6-промотора до остова геновой РНК с терминатором транскрипции. Направляющую последовательность геновой РНК клонировали в полученный вектор путем отжига олигонуклеотидов и их лигирования в вектор после рестрикции *BsmBI*. Использовали следующие олигонуклеотиды: 5'-CACCGCTACGCTAGCATCGCTCGGC-3' и 5'-AAACGCCGAGCGATGCTAGCGTAGC-3', 5'-CACCGAGCGCTACGCTAGCATCGCT-3' и 5'-AAACAGCGATGCTAGCGTAGCGCTC-3' для вышележащей ОРС *PRPF19*; 5'-CACCGCGCGACGACTCAACCTAGTC-3' и 5'-AAACGACTAGGTTGAGTCGTCGCGC-3' для отрицательного контроля; 5'-CACCGATACGTGCAATTTCACCAGA-3' и 5'-AAACTCTGGTGAATTTGCACGTATC-3', 5'-CACCGTGGCCAGGTTGCGGTTCGCAG-3' и 5'-AAACCTGCGACCGCAACCTGGCCAC-3' для положительных контролей (на ген *PCNA*).

Детекция люциферазной активности, опосредованной HiBiT. Для детекции активности люциферазы, образующейся при связывании большой субъединицы люциферазы *LgBiT* и пептида *HiBiT*, использовали набор Nano-Glo HiBiT Lytic Detection System (Promega) в 96-луночном формате согласно протоколу производителя. Интенсивность люминесценции (RLU) каждого образца определяли с помощью Victor X5 plate reader (Perkin Elmer).

Статистический анализ. Статистический анализ проводили с использованием программного обеспечения GraphPad Prism 6.0 (GraphPad, США). Статистическую значимость определяли с помощью *t*-теста не менее чем для трех повторов.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Согласно базе данных Ensemble основной изоформой мРНК гена *PRPF19* является ENST00000227524. Простым способом выявления ОРС в РНК является анализ интересующей последовательности при помощи транскриптомного браузера Trips-Viz [14], который позволяет визуализировать данные рибосомного профилирования. Используя Trips-Viz, можно распределить совокупные данные рибосомного профилирования из разных исследований по трем возможным рамкам считывания.

Чтобы оценить, с какими участками основной изоформы мРНК гена *PRPF19* связаны транслирующие рибосомы (рис. 1), мы выбрали в Trips-Viz режим однозначного отнесения прочтений из рибосомного профилирования, таким образом, все отображенные данные должны относиться только к транскриптам гена *PRPF19*.

Мы обнаружили, что помимо аннотированной основной ОРС в рамке считывания № 3, отмеченной синим (см. рис. 1), в 5'-НТО основной изоформы мРНК также присутствует отмеченный красным профиль, который относится к рамке считывания № 1 (см. рис. 1) и содержит только один стоп-кодон, который расположен в начале транскрипта. Все это свидетельствует о том, что в 5'-НТО мРНК гена *PRPF19* может транслироваться вышележащая ОРС. Интересно, что при этом в 5'-НТО отсутствует старт-кодон AUG, поэтому трансляция такой вышележащей ОРС должна начинаться с неканонического старт-кодона.

Чтобы подтвердить, что трансляция вышележащей ОРС гена *PRPF19* начинается с неканонического старт-кодона, мы выбрали только данные профилирования в Trips-Viz, полученные при остановке рибосом на стадии инициации трансляции в условиях обработки клеток харрингтонином или лактимидомицином (рис. 2).

Полученный суммарный профиль иницирующих рибосом из разных исследований показывает, что трансляция вышележащей ОРС может начинаться с нескольких старт-кодонов. Красные пики в области 5'-НТО мРНК на графике соответствуют кодонам CUG, GUG, AUC, которые близки к AUG и могут выступать в качестве старт-кодонов при инициации трансляции [15].

Вероятность трансляции вышележащей ОРС в мРНК гена *PRPF19* высока, о чем свидетельствует

сопоставимое и даже большее содержание рибосом, особенно иницирующих на ней, в сравнении с основной ОРС. Для этого мы подтвердили, что в клетках человека *HAP1* присутствует только одна изоформа мРНК гена *PRPF19*. Для этого мы получили кДНК, используя два олигонуклеотида, комплементарные 7 и 8 экзонам *PRPF19*, и провели анализ изоформ мРНК гена *PRPF19* методом 5'RACE. Экзоны 7 и 8 содержатся во всех аннотированных в Ensemble транскриптах гена *PRPF19*, которые содержат хотя бы часть первого экзона основной изоформы, где расположена предполагаемая вышележащая ОРС. Мы смогли обнаружить только последовательность, которая полностью соответствует основной изоформе мРНК гена *PRPF19* из Ensemble. Таким образом, данные рибосомного профилирования могут относиться только к одной мРНК гена *PRPF19*.

Перед изучением вышележащей ОРС необходимо подтвердить, что она действительно транслируется. С этой целью мы поместили последовательность, соответствующую 5'-концу мРНК гена *PRPF19*, включая начало ОРС белка PRPF19, перед ОРС зеленого флуоресцентного белка (*EGFP*). В данной репортерной конструкции мы добавили на С-конец предполагаемого пептида последовательность белковой метки *HiBiT*, состоящую из 11 аминокислот (рис. 3А). *HiBiT* позволяет детектировать наличие пептида в образце за счет люминесценции, которая возникает при комплементации неактивной субъединицы люциферазы *LgBiT*

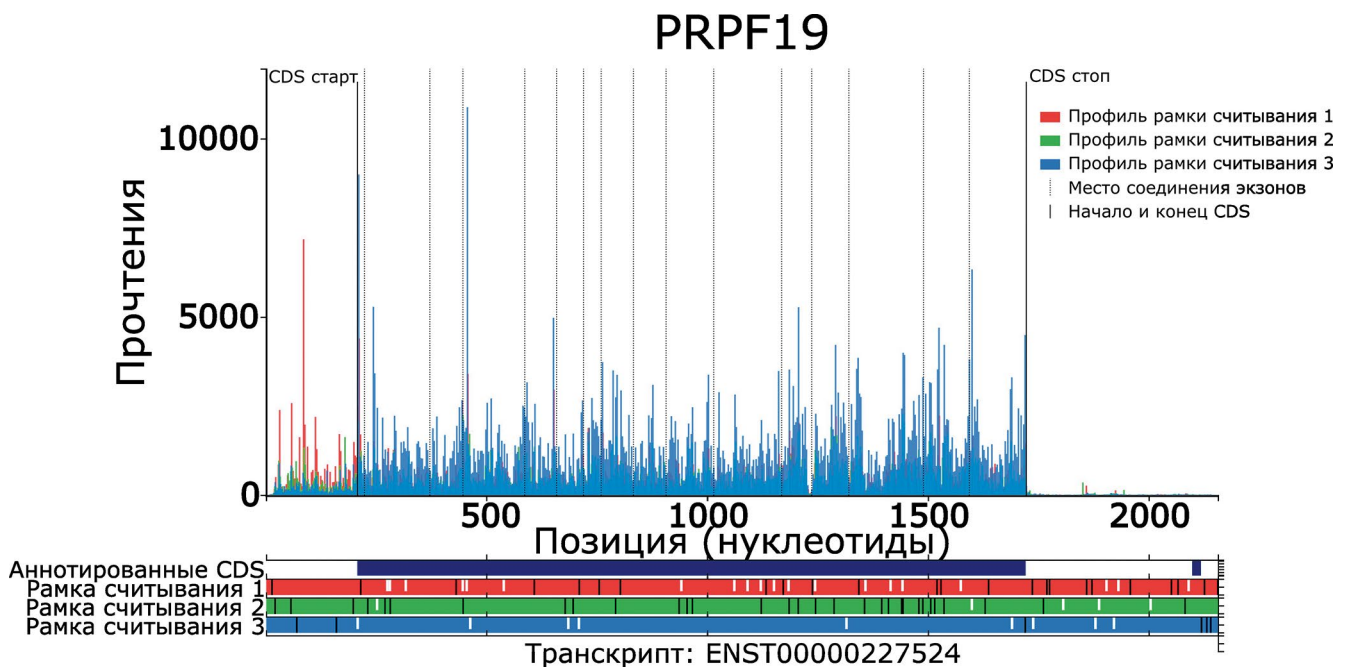


Рис. 1. Рибосомное профилирование основной изоформы мРНК гена *PRPF19* человека (ENST00000227524) согласно базе Ensemble, отображенное в транскриптомном браузере Trips-Viz. На схеме внизу отмечены разными цветами три возможные рамки считывания и соответствующие им профили на графике. На схеме рамок считывания черным отмечены стоп-кодона (UGA, UAA, UAG), белым отмечены старт-кодона (AUG).

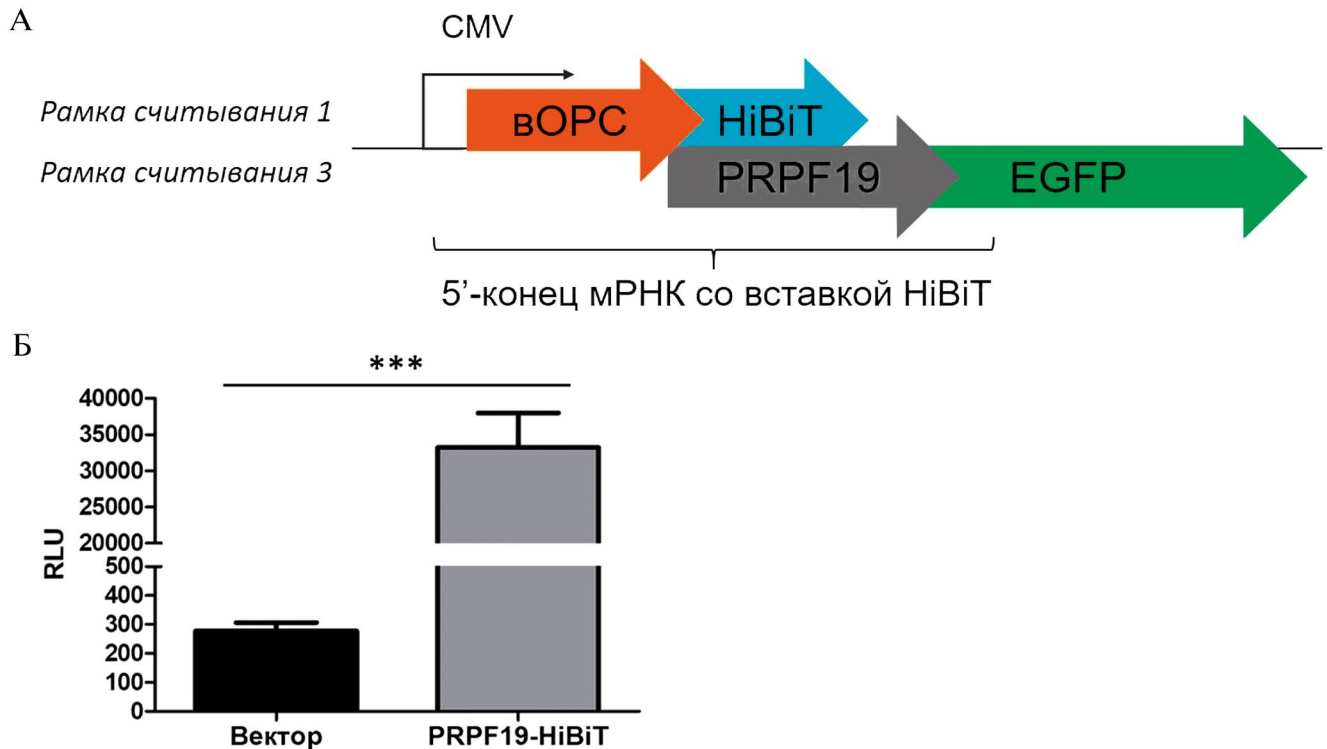


Рис. 3. Подтверждение трансляции вышележащей ОРС гена *PRPF19* с помощью репортерной конструкции: схема репортерной конструкции. Часть ОРС белка *PRPF19* соединена с ОРС *EGFP*. вОРС — вышележащая ОРС (А); измерение люциферазной активности, опосредованной HiBiT, в лизатах клеток после трансфекции репортерными конструкциями. Показаны средние значения и стандартные отклонения для 4 независимых трансфекций. RLU — относительные световые единицы; *** — p -значение < 0.001 (Б).

цевой части. Таким образом, данная малая ОРС и кодируемый ей белок могут быть важны для жизнеспособности клетки.

Далее мы оценили, влияют ли мутации в последовательности данной вышележащей ОРС на жизнеспособность клетки. Для этого мы выбрали две наиболее специфичные гидовые РНК, нацеленные на середину вышележащей ОРС, так что расстояние от старта транскрипции или от начала основной ОРС гена *PRPF19* до ближайшего места разрезания составляло 71 и 128 нуклеотидов соответственно, и использовали их для внесения случайных мутаций с помощью системы CRISPR-Cas9. Таким образом, возникающие мутации не должны влиять на транскрипцию гена *PRPF19* или трансляцию основной ОРС, но будут инактивировать вышележащую ОРС при сдвиге рамки считывания. Мы провели эксперимент по анализу конкуренции между клетками дикого типа и клетками с набором образующихся случайных мутаций в вышележащей ОРС при экспрессии гидовой РНК (рис. 5). В качестве положительного контроля использовали гидовые РНК, нацеленные на ген *PCNA*, который необходим для репликации ДНК, а в качестве отрицательного — гидовую РНК, не имеющую комплементарного участка в геноме.

Мы обнаружили, что экспрессия гидовых РНК, инактивирующих вышележащую ОРС в *PRPF19*, но не контрольной гидовой РНК, снижает жизнеспособность клеток (см. рис. 5).

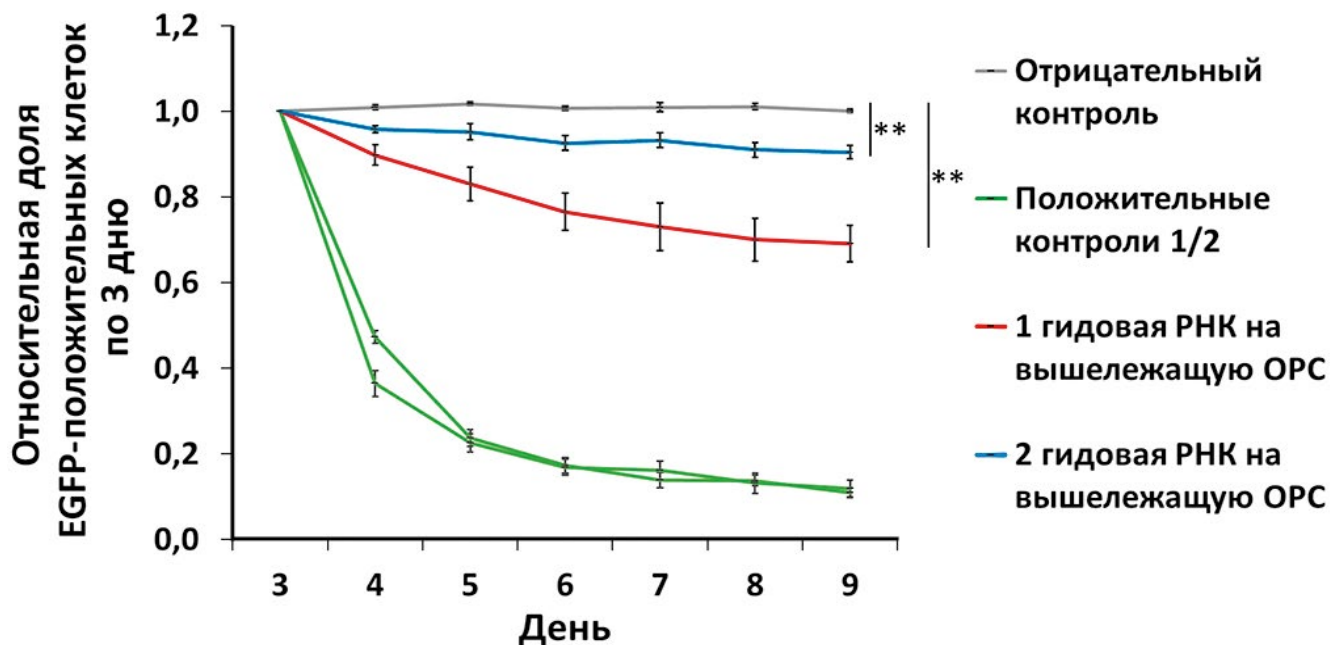
Полученные в работе данные указывают на то, что вышележащая ОРС в мРНК гена *PRPF19* может транслироваться в клетках человека *in vitro*. При этом кодируемый ею малый белок консервативен среди млекопитающих. Мутации в вышележащей ОРС приводят к снижению жизнеспособности клеток человека. Таким образом, либо потеря малого белка, либо нарушение регуляторной активности вышележащей ОРС приводит к снижению жизнеспособности клеток человека.

ЗАКЛЮЧЕНИЕ

Итак, 5'-нетранслируемая область мРНК гена *PRPF19* содержит вышележащую ОРС, которая транслируется в клетках человека *in vitro*, как было показано с использованием репортерной конструкции. Инактивация данной вышележащей ОРС снижает жизнеспособность клеток человека. Консервативность белка, кодируемого вышележащей ОРС в мРНК гена *PRPF19*, среди млекопитающих указывает на то, что именно он может быть необходим для



Рис. 4. Анализ консервативности вышележающей OPC гена *PRPF19* среди млекопитающих: выравнивание вышележающей OPC гена *PRPF19* в геномном браузере *CodAlignView* среди некоторых млекопитающих. Указаны положения первого стар-кодона вышележающей OPC у человека по данным на рис. 2 и стоп-кодона (А); Выравнивание последовательностей белков, кодируемых вышележающей OPC в мРНК гена *PRPF19*, среди указанных видов в *MSA4U* [17] often encoding so-called leader peptides. Начало белка выбрано по первому возможному стар-кодону в соответствующей рамке считывания (Б).



жизнеспособности клетки. В то же время нельзя исключать, что данная вышележащая ОРС регулирует трансляцию основной ОРС, и содержание белка *PRPF19* в клетке соответственно. Регуляторная роль вышележащей ОРС в мРНК гена *PRPF19* также может иметь значение для жизнеспособности клетки.

КОНФЛИКТ ИНТЕРЕСОВ

Конфликт интересов у авторов отсутствует.

БЛАГОДАРНОСТИ

Выражаем благодарность Программе развития МГУ ПНР-5 за предоставление доступа к проточному цитометру MACSQuant Analyser, планшетному мультиридеру Victor X5 plate reader и капиллярному секвенатору SeqStudio.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 23-14-00058).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ НОРМ И СТАНДАРТОВ

В данной работе отсутствуют исследования человека или животных.

СПИСОК ЛИТЕРАТУРЫ

1. Brar G. A., Weissman J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis // *Nat Rev Mol Cell Biol.* 2015. V. 16. № 11. P. 651–664.
2. Rubtsova M., Naraykina Y., Vasilkova D., et al. Protein encoded in human telomerase RNA is involved in cell protective pathways // *Nucleic Acids Research.* 2018. V. 46, № 17. P. 8966–8977.
3. Chugunova A., Loseva E., Mazin P., et al. *LINC00116* codes for a mitochondrial peptide linking respiration and lipid metabolism // *Proc. Natl. Acad. Sci. U.S.A.* 2019. V. 116, № 11. P. 4940–4945.
4. Sergiev P. V., Rubtsova M. P. Little but Loud. The Diversity of Functions of Small Proteins and Peptides – Translational Products of Short Reading Frames // *Biochemistry Moscow.* 2021. V. 86, № 9. P. 1139–1150.
5. Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression // *WIREs RNA.* 2014. V. 5, № 6. P. 765–768.
6. Renz P. F., Valdivia-Francia F., Sandoel A. Some like it translated: small ORFs in the 5'UTR // *Experimental Cell Research.* 2020. V. 396, № 1. P. 112229.
7. Ameri K., Harris A. L. Activating transcription factor 4 // *The International Journal of Biochemistry & Cell Biology.* 2008. V. 40, № 1. P. 14–21.
8. Vattem K. M., Wek R. C. Reinitiation involving upstream ORFs regulates *ATF4* mRNA translation in mammalian cells // *Proc. Natl. Acad. Sci. U.S.A.* 2004. V. 101, № 31. P. 11269–11274.
9. Cloutier P., Poitras C., Faubert D., et al. Upstream ORF-Encoded ASDURF Is a Novel Prefoldin-like Subunit of the PAQosome // *J. Proteome Res.* 2020. V. 19, № 1. P. 18–27.
10. Pinard M., Cloutier P., Poitras C., et al. Unphosphorylated Form of the PAQosome Core Subunit RPAP3 Binds Ribosomal Preassembly Complexes to Modulate Ribosome Biogenesis // *J. Proteome Res.* 2022. V. 21, № 4. P. 1073–1082.
11. Meyers R. M., Bryan J. G., McFarland J. M., et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells // *Nat Genet.* 2017. V. 49, № 12. P. 1779–1784.
12. Behan F. M., Iorio F., Picco G., et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens // *Nature.* 2019. V. 568, № 7753. P. 511–516.
13. Yin J., Zhu J.-M., Shen X.-Z. New insights into pre-mRNA processing factor 19: A multi-faceted protein in humans // *Biology of the Cell.* 2012. V. 104, № 12. P. 695–705.
14. Kiniry S. J., O'Connor P. B. F., Michel A. M., et al. Trips-Viz: a transcriptome browser for exploring Ribo-Seq data // *Nucleic Acids Research.* 2019. V. 47, № D1. P. D847–D852.
15. Cao X., Slavoff S. A. Non-AUG start codons: Expanding and regulating the small and alternative ORFeome // *Experimental Cell Research.* 2020. V. 391, № 1. P. 111973.
16. Jungreis, I., Lin, M. F., Chan, C. S., et al. CodAlignView: A tool for visualizing protein-coding constraint. 2016.
17. Egorov A. A., Atkinson G. C. uORF4u: a tool for annotation of conserved upstream open reading frames // *Bioinformatics.* 2023. V. 39, № 5. P. btad323.

***PRPF19* mRNA ENCODES A SMALL OPEN READING FRAME THAT IS IMPORTANT FOR VIABILITY OF HUMAN CELLS**

**N. M. Shepelev^{a, b}, A. O. Kurochkina^b, Academician of the RAS O. A. Dontsova^{a, b, c, d},
M. P. Rubtsova^{a, b, #}**

^a*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russian Federation*

^b*Lomonosov Moscow State University, Department of Chemistry, Moscow, Russian Federation*

^c*A. N. Belozersky Institute of Physico-Chemical Biology, Moscow, Russian Federation*

^d*Skolkovo Institute of Science and Technology, Center for Molecular and Cellular Biology, Moscow, Russian Federation*

[#]*E-mail: mprubtsova@gmail.com*

High-throughput ribosome profiling demonstrated the translation of thousands small open reading frames located in the 5' untranslated regions of messenger RNAs (upstream ORFs). Upstream ORF can both perform a regulatory function by influencing the translation of the downstream main ORF, and encode a small functional protein or microprotein. In this work, we showed that the 5' untranslated region of the *PRPF19* mRNA encodes an upstream ORF that is translated in human cells. Inactivation of this upstream ORF reduces the viability of human cells.

Keywords: ribosome profiling, upstream ORF, uORF, microprotein, micropeptide.