

## РЕВИЗИЯ ФУНКЦИОНАЛЬНО ЗНАЧИМЫХ И АКТИВНО ЭКСПРЕССИРУЮЩИХСЯ ДЛИННЫХ НЕКОДИРУЮЩИХ РНК

© 2024 г. Д. О. Кони́на<sup>а, \*</sup>, М. Ю. Скоблов<sup>а</sup>

<sup>а</sup>Лаборатория функциональной геномики, Медико-генетический научный центр  
им. акад. Н.П. Бочкова, Москва, 115522 Россия

\*e-mail: darya.konina@phystech.edu

Поступила в редакцию 19.09.2023 г.

После доработки 10.12.2023 г.

Принята к публикации 17.12.2023 г.

Длинные некодирующие РНК (long non-coding RNA, lncРНК) участвуют в регуляции многих клеточных процессов, а их экспрессия характеризуется высокой тканеспецифичностью. В отличие от них белоккодирующие гены, в частности гены “домашнего хозяйства”, активно экспрессируются во всех тканях. Нами исследована функциональная значимость широкоэкспрессирующихся lncРНК. Мы проанализировали экспериментальные данные по скринингу lncРНК в клетках человека и выявили 18 некодирующих РНК, которые имели не только высокий уровень экспрессии, но и были функционально активны в большинстве тканей. Детальный анализ этих lncРНК показал, что они имеют широкий спектр молекулярных функций: от контроля повреждения кардиомиоцитов до переключения классов макрофагов. На основании полученных экспериментальных данных можно сделать вывод, что большинство исследованных нами lncРНК кодирует небольшие, но функционально значимые пептиды. Среди 18 lncРНК высоким кодирующим потенциалом обладали NEAT1, SNHG1, SNHG7, SNHG12, SNHG15, SNHG16, MIR17HG, LINC00680, LINC00263 и LINC00339. Кроме того, обнаружено, что EPB41L4A-AS1, CRNDE, SNHG6, LINC00493 и LINC01420 могут функционировать как за счет первичной структуры РНК, так и за счет кодируемого ими функционального пептида.

**Ключевые слова:** CRISPR-интерференция, длинные некодирующие РНК, короткая открытая рамка считывания

DOI: 10.31857/S0026898424030137, EDN: JBZHXL

### ВВЕДЕНИЕ

С открытием большого класса длинных некодирующих РНК (lncРНК) началась эпоха активного исследования их функций. В результате были идентифицированы и подробно изучены такие известные lncРНК, как XIST, MALAT1, HOTAIR, GAS5, HOTTIP, TERRA и FIRRE [1–3]. Из проведенных на сегодняшний день исследований известно, что lncРНК обладают способностью изменять архитектуру хроматина, регулировать сборку и функционирование бимолекулярных конденсатов, влиять на стабильность и процессинг мРНК, а также регулировать различные сигнальные пути

[4–6]. Описанные примеры демонстрируют широкий спектр молекулярных функций, связанных с участием lncРНК во многих ключевых клеточных процессах [7–10].

Согласно данным проекта GENCODE (версия 44) [11], в геноме человека аннотировано сравнимое число генов lncРНК (19928) и белоккодирующих генов (19393). Следовательно, кроме исследования функций конкретных lncРНК, необходимо проводить широкомасштабный функциональный анализ аннотированных транскриптов генов lncРНК.

Один из наиболее распространенных подходов – использование нокдаун библиотек генов с последующим анализом клеточного фенотипа. В крупномасштабных скринингах нокдаун проводят с использованием различных технологий. Так, для анализа роли 2231 lncРНК в процессе деления клеток HeLa использовали малые интерферирующие РНК (siРНК) [12], для выявления функционально значимых генов среди 285 lncРНК в дермальных фибробластах человека – антисмысловые олигонуклеотиды [13], а для изучения влияния 214 lncРНК на процессы деления и поддержания плюрипотентности эмбриональных стволовых клеток

Сокращения: CRISPRi – CRISPR-интерференция; НК (housekeeping) – (ген) “домашнего хозяйства”; lncРНК – длинные некодирующие РНК; FPKM (fragments per kilobase per million mapped reads) – число фрагментов на 1 тыс. оснований транскрипта на 1 млн картированных прочтений; TPM (transcripts per million) – число транскриптов на миллион прочтений; TSS (transcription start site) – сайт начала транскрипции; kОРС – короткая открытая рамка считывания; shРНК – короткие шпилечные РНК; sgРНК – направляющие РНК; snoРНК – малые ядрышковые РНК; miРНК – микроРНК; БКГ – белоккодирующий ген.

мышей – короткие шпилечные РНК (shРНК) [14]. В результате этих исследований было получено множество функционально значимых транскриптов, управляющих процессами жизнедеятельности клеток. Тем не менее более подробное изучение молекулярных функций выявленных генов остается задачей для будущего рассмотрения.

В представленной работе проведен ретроспективный анализ данных крупномасштабного скрининга на основе CRISPR-интерференции (CRISPRi) [15]. Мы выявили и охарактеризовали активно экспрессирующиеся функциональные lncРНК, а затем проанализировали их молекулярные функции, основываясь на результатах работ, опубликованных за последние несколько лет. Мы также оценили кодирующий потенциал отобранных lncРНК-хитов. В результате выявлен широкий спектр молекулярных функций lncРНК, в том числе связанных с кодированием пептидов.

## МЕТОДЫ

**Методы анализа данных крупномасштабного скрининга.** Для анализа результатов крупномасштабного скрининга аннотированных транскриптов использовали данные исследования функциональных локусов lncРНК в клетках человека с помощью метода CRISPRi [15]. Скрининг с использованием технологии CRISPRi проведен для 16401 lncРНК человека. Аннотации транскриптов lncРНК были взяты из Ensembl build 75 [16], MiTranscriptome [17], каталога lncРНК человека [18] и набора специфичных для мозга lncРНК [19].

Для отбора функционально значимых генов lncРНК использовали значения уровня экспрессии и количественного показателя нокдауна, которые были измерены S. Liu с соавт. [15] для 16401 lncРНК в клеточных линиях HEK293T, K562, MCF7, MDA-MB-231, HeLa и U87. Уровень экспрессии для каждой lncРНК определяли как среднее значение числа фрагментов на 1 тыс. оснований транскрипта на 1 млн картированных прочтений (fragments per kilobase per million mapped reads, FPKM) повторных образцов по данным РНК-секвенирования открытых источников данных: HEK293T (GSE56010), HeLa (GSE30567, GSE33480, GSE23316), K562 (GSE30567, GSE33480, GSE23316), MCF7 (GSE30567, GSE33480), MDAMB231 (GSE73526, GSE45732). Данные по РНК-секвенированию клеточной линии U87 были взяты из работы [15].

Количественный показатель нокдауна определял нормализацию обогащения направляющих РНК (single guide RNA, sgРНК) для таргетной lncРНК по общему числу удвоенных клеток. Такой параметр был выбран для нормализации обогащения sgРНК в клеточных линиях с разной скоростью роста. Таким образом, количественный

показатель нокдауна отражает положительное или отрицательное влияние нокдауна гена на пролиферацию клеток.

Наиболее функционально значимыми считали lncРНК с высоким уровнем экспрессии, а также со значимым эффектом нокдауна во всех исследованных клеточных линиях. Для отбора наиболее функционально значимых lncРНК были построены распределения двух выбранных параметров lncРНК в клеточных линиях. Далее на основе наблюдаемого правостороннего распределения значений уровня экспрессии предполагали, что lncРНК высоко экспрессируется в выбранной клеточной линии, если ее измеренный уровень экспрессии превышает средний уровень экспрессии генов в соответствующей клеточной линии: выше – (1), ниже – (0). В то же время на основе симметричного распределения значений количественного показателя нокдауна мы предполагали, что нокдаун lncРНК оказывает значимый эффект на пролиферацию клеток в выбранной клеточной линии, если ее измеренный показатель нокдауна превышает стандартное отклонение этого параметра в соответствующей клеточной линии: выше – (1), ниже – (0). Таким образом, для каждой lncРНК оценили в общей сложности 12 параметров. Затем был определен “показатель значимости” lncРНК, который включал сумму значений 12 оцениваемых параметров. Полученное значение показателя значимости lncРНК отражает интегральную оценку как для количественного показателя нокдауна, так и для уровня экспрессии. Для lncРНК, оценка показателя значимости которых составила более 6, в соответствии с экспериментальными данными гарантирован наблюдаемый функционально значимый эффект нокдауна на пролиферацию клеток и высокий уровень экспрессии не менее чем в одной клеточной линии. Однако для исследования и последующего углубленного биоинформатического анализа мы выбрали более высокое пороговое значение, которое установили на уровне  $\geq 8$ , – для отбора lncРНК-хитов. Дополнительно учитывали, чтобы расстояние от сайта начала транскрипции (transcription start site, TSS) lncРНК до ближайшего TSS белоккодирующего гена было не менее 1000 п.н. (ввиду особенностей метода CRISPRi) и чтобы не было пересечений нуклеотидной последовательности lncРНК с экзонами белоккодирующих генов.

**Биоинформатический анализ lncРНК-хитов.** Для отобранных lncРНК-хитов провели биоинформатический анализ, включающий оценку уровня экспрессии в тканях человека, степень консервативности, степень изученности и кодирующий потенциал.

Уровень экспрессии lncРНК в тканях человека оценивали с использованием базы данных Genotype-Tissue Expression (GTEx v8) [20]. GTEx v8

содержит данные высокопроизводительного секвенирования для 54 образцов ткани от 948 доноров, сРНК-, ДНК- и иммунопреципитацией хроматина по крайней мере для 70 образцов на ткань. Для каждой lncРНК оценивали общую медиану по значениям экспрессии в единицах числа транскриптов на миллион прочтений (transcripts per million, TPM) в разных тканях человека.

В рамках анализа уровня экспрессии оценивали принадлежность гена lncРНК к генам “домашнего хозяйства” (housekeeping, НК), идентифицированным в рамках проекта FANTOM5 [21]: (0) – ген не относится к НК, (1) – ген относится к НК. Группа НК-генов идентифицирует группу генов, имеющих равномерный характер экспрессии среди 945 образцов клеток и тканей человека.

Степень консервативности lncРНК оценивали с использованием геномного браузера UCSC [22]. На треке phastCons100way показано множественное выравнивание 100 видов позвоночных и оценка степени консервативности с использованием метода phastCons из пакета PHAST для всех 100 видов. Множественные выравнивания были получены с использованием Multiz и других инструментов выравнивания геномной последовательности, доступных в UCSC браузере.

Анализ зависимости эффекта нокдауна lncРНК от типа клеток. Для определения функционально значимых lncРНК, нокдаун которых приводил к значимо различному влиянию на рост разного типа клеток, выбирали lncРНК с явно выраженным разнонаправленным эффектом нокдауна на рост клеток не менее чем в двух клеточных линиях.

Степень изученности lncРНК оценивали по числу статей, относящихся к lncРНК (этот параметр

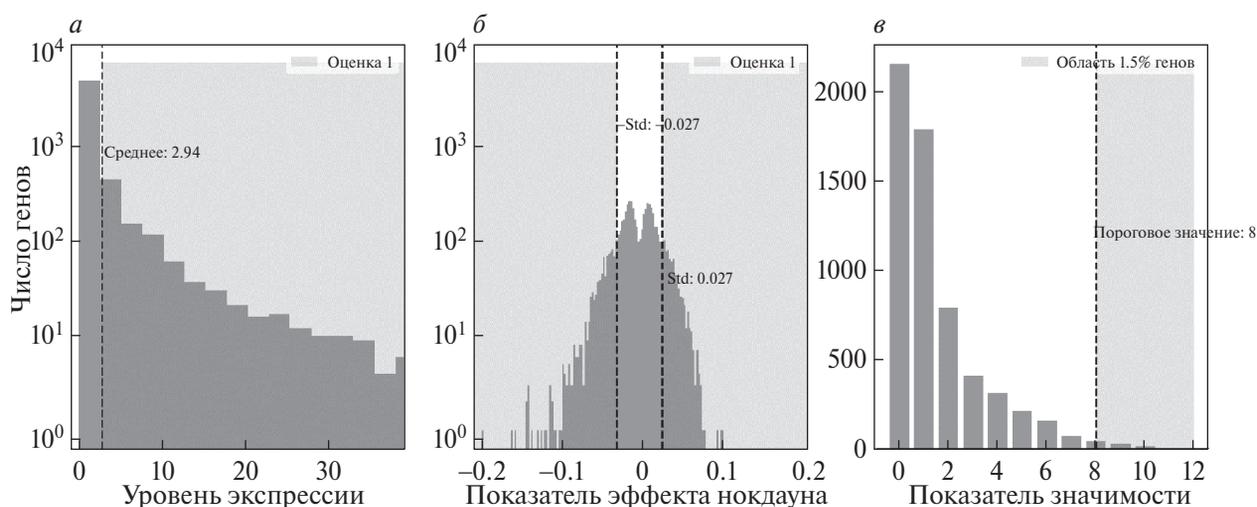
определяли по ключевым словам в базе PubMed по состоянию на июль 2023 года) и опубликованных, начиная с 2017 года.

Оценка кодирующего потенциала lncРНК включала исследование общедоступных экспериментальных данных рибосомного профилирования с использованием веб-браузеров GWIPS-viz [23] и Trips-Viz [24] (обозначали как (0) – отсутствие взаимодействия с рибосомами, (1) – наличие взаимодействия с рибосомами). Также мы провели анализ данных базы Lncprep (<http://www.shenglilabs.com/LncPrep/>) [25] и CPC2.0 (<http://cpc2.gao-lab.org/>) [26] для оценки кодирующего потенциала методами CPC2 и CPAT, основанными на анализе нуклеотидной последовательности генов, а также наличия экспериментального подтверждения трансляции по данным рибосомного профилирования (число наборов данных). Дополнительно в рамках анализа кодирующего потенциала lncРНК учитывали наличие опубликованных результатов, которые подтверждают трансляцию пептида с выбранной lncРНК: (0) – соответствующих публикаций нет, (1) – такие публикации есть.

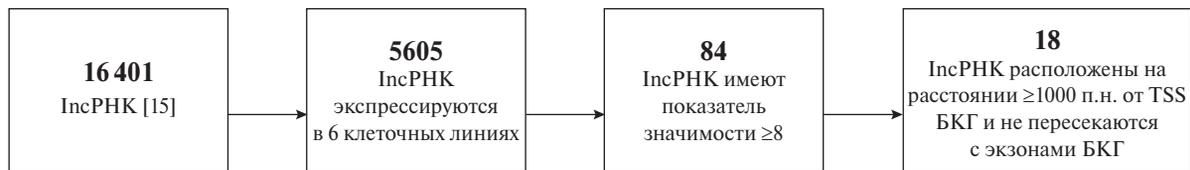
## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### *Необходимые для роста клеток высокоэкспрессирующиеся lncРНК: анализ данных CRISPRi*

Для поиска новых потенциально значимых lncРНК мы провели анализ опубликованных данных крупномасштабного скрининга их генов. В анализ были включены данные CRISPRi-скрининга для 16401 lncРНК, полученные в 6 клеточных



**Рис. 1.** Процесс отбора lncРНК. *а* – Анализ экспрессии lncРНК в клеточной линии НЕК293Т. Выделенная область представляет lncРНК со значимым уровнем экспрессии. *б* – Усредненные данные по показателю эффекта нокдауна lncРНК в клеточной линии НЕК293Т. Выделенная область включает lncРНК, нокдаун которых значимо влияет на пролиферацию клеток. *в* – Показатель значимости lncРНК. Выделенная область включает lncРНК, которые были выбраны для дальнейшего анализа.



**Рис. 2.** Этапы отбора lncРНК с указанием критериев отбора и количества выбранных генов на каждом этапе. Обозначения: БКГ – белоккодирующие гены, TSS (transcription start site) – сайт начала транскрипции.

линиях: HEK293T, K562, MCF7, MDA-MB-231, HeLa, U87 [15]. В отличие от S. Liu с соавт. [15], наша цель заключалась в выявлении универсальных lncРНК, обладающих значимым клеточным фенотипом независимо от типа клеток.

На начальном этапе обобщили экспериментальные данные и отметили, что во всех 6 клеточных линиях экспрессируется только 5605 lncРНК. Далее оценили уровень экспрессии и значимость эффекта нокдауна для отдельной lncРНК в каждой исследуемой клеточной линии. Таким образом, для каждой lncРНК в 6 клеточных линиях оценивали 12 параметров. Если значение признака для отдельной lncРНК превышало предопределенный порог, то при оценке ему присваивалось значение 1, в противном случае – 0 (см. рис. 1а,б).

Далее суммировали оценки всех 12 признаков для каждой lncРНК в отдельности. Такую суммарную оценку обозначили как “показатель значимости”. Из исходного набора lncРНК найдено примерно 1.5% (84), для которых показатель значимости  $\geq 8$  (см. рис. 1в). При отборе генов также учитывали расстояние от TSS lncРНК до TSS ближайшего белоккодирующего гена ( $>1000$  п.н.) и отсутствие перекрытия lncРНК с экзонами белоккодирующего гена. В результате выбрано 18 lncРНК-хитов (рис. 2, табл. 1): NEAT1, SNHG12, EPB41L4A-AS1, SNHG16, SNHG1, SNHG7, MIR17HG, SCARNA10, CRNDE, LINC00680, SNHG25, SNHG6, LINC00339, LINC00263, SNHG30, LINC00493, SNHG15 и LINC01420.

#### *Характеристика lncРНК-хитов и их молекулярные фенотипы*

С целью охарактеризовать хиты lncРНК мы провели их комплексный анализ по различным параметрам, включая уровень экспрессии в тканях человека, эволюционную консервативность последовательности, степень изученности и кодирующий потенциал. Обобщенные данные по характеристикам хитов lncРНК приведены в табл. 1.

Для выявленных генов проведен анализ уровня экспрессии в тканях человека с использованием данных проекта GTEx v8. Для каждой lncРНК оценивали общую медиану по значениям экспрессии в единицах TPM в разных тканях человека. Дополнительно учитывали принадлежность lncРНК-хита к группе генов НК, идентифицированным в рамках проекта FANTOM5 [21]. Также провели

поиск корреляции между показателями экспрессии и данными о зависимом от типа клеток влиянии нокдауна lncРНК на пролиферацию [15].

В целом, для выбранных lncРНК-хитов регистрировали высокий уровень экспрессии (общая медиана TPM) в тканях человека. Для lncРНК, таких как NEAT1, SNHG12, EPB41L4A-AS1, SNHG1, SNHG16, SNHG7, SNHG6, LINC00339, LINC00493, SNHG15 и LINC01420, максимальный уровень экспрессии (медиана TPM для одной ткани) соответствовал значениям, которые характерны для белоккодирующих генов. Для некоторых lncРНК, включая NEAT1, SNHG1 и SNHG6, выявлен высокий уровень экспрессии (более 50 TPM) во всех тканях GTEx v8.

Однако встречались lncРНК-хиты и с тканеспецифичным типом экспрессии – для них получены средние (10–50 TPM) или низкие ( $<10$  TPM) значения уровня экспрессии. К ним относятся CRNDE, EPB41L4A-AS1, SNHG7, LINC00339, SNHG30 и MIR17HG. Заметим, что тканеспецифичные lncРНК экспрессируются на значимом уровне только в одной ткани. Особенно четко эта особенность проявилась в случае lncРНК CRNDE, для которой рассчитанный в GTEx v8 уровень экспрессии во всех проанализированных образцах тканей составлял 1.37 TPM, в то время как в семенниках был максимальным – 52.2 TPM.

Профиль равномерной экспрессии, характерный для НК-генов, как правило, не связан напрямую с высоким уровнем экспрессии. Например, NEAT1 характеризуется высоким уровнем экспрессии, но при этом не относится к НК-генам. Для lncРНК EPB41L4A-AS1, наоборот, выявлен средний уровень экспрессии, хотя ее ген принадлежит к НК. Сходный с генами НК профиль экспрессии во всех тканях наблюдали для LINC00680, SNHG12, SNHG16, SNHG25, LINC01420, LINC00493 и SCARNA10. Результаты проведенного нами анализа профилей экспрессии lncРНК-хитов подтвердили данные CRISPRi о наличии специфичных для типа клеток эффектов нокдауна в случае CRNDE, SNHG7, LINC00339 и SNHG30. Для LINC00493 раньше также экспериментально продемонстрировано наличие зависимости эффекта нокдауна от выбранного типа клеток [27]. Кроме того, при анализе данных CRISPRi для LINC00263, SCARNA10, SNHG6, SNHG25 и SNHG15 мы выявили специфичный для конкретного типа клеток эффект

Таблица 1. Характеристика lncРНК-хитов<sup>а</sup>

Ген	Ensemble_ID	Уровень экспрессии (TPM)	НК	Зависимость от типа клеток	Консервативность (PhastCons)	Число публикаций (>2017)	Степень изученности	Молекулярные взаимодействия <sup>б</sup>
<i>NEAT1</i>	ENSG00000245532	210.55	0	0	0.05	1231	Высокая	РНК-белок, РНК-miРНК, РНК-мРНК, РНК-ДНК
<i>SNHG12</i>	ENSG00000197989	18.81	0	0	0.12	147	Средняя	РНК-miРНК
<i>EPB41L4A-AS1</i>	ENSG00000224032	35.99	1	0	0.07	28	Средняя	РНК-белок, РНК-miРНК, РНК-ДНК
<i>SNHG16</i>	ENSG00000163597	15.41	0	0	0.05	271	Высокая	РНК-белок, РНК-miРНК, РНК-ДНК
<i>SNHG1</i>	ENSG00000255717	61.92	0	0	0.17	300	Высокая	РНК-белок, РНК-miРНК, РНК-ДНК
<i>SNHG7</i>	ENSG00000233016	25.10	0	1	0.07	182	Средняя	РНК-белок, РНК-miРНК, РНК-мРНК, РНК-ДНК
<i>MIR17HG</i>	ENSG00000215417	0.58	0	0	0.27	119	Средняя	РНК-белок, РНК-miРНК, РНК-мРНК
<i>SCARNA10</i>	ENSG00000239002	—	—	1	0.59	8	Низкая	РНК-белок
<i>CRNDE</i>	ENSG00000245694	1.36	0	1	0.31	229	Высокая	РНК-белок, РНК-miРНК, РНК-ДНК
<i>LINC00680</i>	ENSG00000215190	4.86	0	0	0.19	10	Низкая	РНК-белок, РНК-miРНК, РНК-мРНК
<i>SHNG25</i>	ENSG00000266402	3.88	0	1	0.76	8	Низкая	РНК-белок, РНК-miРНК, РНК-snoРНК
<i>SNHG6</i>	ENSG00000245910	154.65	1	1	0.07	151	Средняя	РНК-белок, РНК-miРНК, РНК-ДНК
<i>LINC00339</i>	ENSG00000218510	13.23	1	1	0.06	37	Средняя	РНК-miРНК
<i>LINC00263</i>	ENSG00000235823	5.93	0	1	0.11	11	Низкая	РНК-белок, РНК-miРНК, РНК-ДНК
<i>SNHG30</i>	ENSG00000257270	11.40	0	1	0.15	2	Низкая	Не установлено
<i>LINC00493</i>	ENSG00000232388	37.94	1	1	0.10	3	Низкая	Не установлено
<i>SNHG15</i>	ENSG00000232956	12.01	0	1	0.04	109	Средняя	РНК-белок, РНК-miРНК, РНК-ДНК
<i>LINC01420</i>	ENSG00000204272	30.83	1	0	0.10	15	Низкая	РНК-белок, РНК-miРНК

<sup>а</sup>Здесь и далее: оттенки серого обозначают тепловую карту полученных значений характеристик lncРНК-хитов. Интервалы значений для выбора цвета зависят от характеристики; цвета выбраны от светло-серого до темно-серого – от наименьших значений к наибольшим.

<sup>б</sup>Обозначения: miРНК – микроРНК; snoРНК – малая ядершковая РНК.

нокдауна, который предстоит дополнительно подтвердить экспериментально.

По результатам оценки степени консервативности lncРНК-хитов выявлена невысокая степень консервативности нуклеотидных последовательностей отобранных генов lncРНК, что соответствует данным по эволюции lncРНК [28]. Примечательно, что высокая степень консервативности последовательностей наблюдалась для SCARNA10 и SHNG25. Полученные результаты можно объяснить тем, что SCARNA10 происходит из интронной области белоккодирующего гена NCAPD2, а значительная часть нуклеотидной последовательности SHNG25 перекрывается с генами малых ядрышковых РНК (small nucleolar RNA, snoРНК): SNORD104 и SNORA50C. В общем случае для отобранных генов lncРНК замечено, что их экзонные последовательности более консервативны, чем интронные области. Такой паттерн характерен для генов семейства SNHG (SNHG12, SNHG1, SNHG16, SNHG6, SNHG30) и lncРНК MIR17HG, где высококонсервативные экзонные области способствуют последующей транскрипции snoРНК или микроРНК (miРНК). В случае SNHG15 область транскрипции SNORA9 включена в его третий интрон, что объясняет сравнительно меньшую степень консервативности гена. С другой стороны, lncРНК, такие как CRNDE, LINC00680, LINC00263 и LINC01420, нами были идентифицированы как консервативные гены (со средним показателем PhastCons > 0.10). Важно отметить, что эти гены не пересекаются с интронными последовательностями белоккодирующих генов или транскрипционными областями snoРНК или miРНК.

Мы также установили степень изученности lncРНК-хитов, используя для оценки характеристики число статей (поиск проведен по ключевым словам в PubMed), вышедших с момента публикации данных исследования CRISPRi [15]. По результатам проведенного анализа публикационной активности установлено, что, несмотря на появление новых технологий, позволяющих ученым исследовать потенциально значимые новые гены, в научном сообществе сохраняется тенденция к более глубокому изучению уже известных. В связи с этим большинство публикаций посвящено таким хорошо известным lncРНК, как NEAT1, SNHG1 и CRNDE, что привело к высокой степени их изученности (>200 публикаций). Первые публикации, посвященные среднеизученным генам (20–200 публикаций), появились в период 2013–2019 гг. К группе среднеизученных генов относятся EPB41L4A-AS1, LINC00339 и SNHG12. Гены SCARNA10, LINC00680, SHNG25, LINC00263, SNHG30, LINC01420 и LINC00493 образуют группу генов с относительно низкой степенью изученности (<20 публикаций). Тем не менее выявленный уровень публикационной активности подчеркивает

значимую роль этих lncРНК-хитов в регуляторном ландшафте генома.

Молекулярные партнеры и функции для lncРНК-хитов проанализированы нами по данным, опубликованным в научной литературе (табл. S1 Дополнительных материалов см. на сайте [http://www.molecbio.ru/downloads/2024/3/supp\\_Konina\\_rus.pdf](http://www.molecbio.ru/downloads/2024/3/supp_Konina_rus.pdf)). В результате проведенного анализа установлено, что большинство отобранных lncРНК выполняют свои функции, действуя как “губки” для miРНК. В качестве примера можно привести NEAT1 [29], SNHG12 [30], SNHG16 [31], SNHG1 [32], SNHG7 [33], MIR17HG [34], CRNDE [35], SNHG6 [36], LINC00339 [37], LINC00263 [38] и LINC01420 [39]. Некоторые из выявленных lncРНК-хитов выступают в роли эпигенетических регуляторов, взаимодействуя с белками PRC2 или DNMT1/2 и образуя с генами-мишенями триплексы lncРНК-ДНК-РНК. Примером такого взаимодействия являются NEAT1 [40], SNHG1 [41], SNHG7 [42], CRNDE [43], SNHG6 [44] и SNHG15 [45].

Кроме того, можно выделить подгруппу lncРНК-хитов, которая взаимодействует с различными ферментами и факторами транскрипции, тем самым принимая активное участие в регуляции процессов транскрипции, процессинга и трансляции РНК и в то же время играя ключевую роль в активации сигнальных путей. К этой функциональной категории относятся NEAT1 [46], SNHG1 [47], SNHG7 [48], SCARNA10 [49], CRNDE [50], LINC00680 [51], SNHG6 [52] и LINC01420 [53]. Следует отметить NEAT1 и LINC01420, которые участвуют в образовании биомолекулярных конденсатов. NEAT1 связывается с белками параспеклов [54], а LINC01420 взаимодействует с белками Р-гранул [55]. К числу редких межмолекулярных взаимодействий, обнаруженных для lncРНК-хитов, относится образование дуплексов lncРНК-мРНК. Такие взаимодействия были обнаружены для LINC00680 [56], SNHG7 [57], NEAT1 [58] и MIR17HG [59].

Благодаря разнообразным взаимодействиям lncРНК-хиты участвуют в регуляции фундаментальных клеточных процессов и вносят вклад в развитие патогенеза заболеваний. Например, они вовлечены в онкологические, воспалительные и нейродегенеративные процессы, противовирусный ответ, развитие сердечно-сосудистых заболеваний и диабета.

*Высокий белоккодирующий потенциал  
lncРНК-хитов: результаты рибосомного  
профайлинга*

Поскольку lncРНК могут транслироваться в небольшие функциональные пептиды, мы решили провести оценку кодирующего потенциала исследуемых lncРНК-хитов. Для этого использовали

Таблица 2. Оценка белоккодирующего потенциала lncРНК-хитов

Ген	TRIPS/ GWIPS	ЭП <sup>а</sup>	Размер пептида (а.о.)	CPAT	CPC2	Ribo-Seq (число наборов)
<i>NEAT1</i>	1	0	106	0.76	0.59	4
<i>SNHG12</i>	1	0	76	0.08	0.07	5
<i>EPB41L4A-AS1</i>	1	1	120	0.26	0.17	8
<i>SNHG16</i>	1	1	55	0.12	0.03	46
<i>SNHG1</i>	1	0	83	0.03	0.06	56
<i>SNHG7</i>	1	0	51	0.08	0.20	54
<i>MIR17HG</i>	1	0	31	0.35	0.13	5
<i>SCARNA10</i>	1	0	35	–	0.05	11
<i>CRNDE</i>	1	1	84	0.18	0.35	11
<i>LINC00680</i>	1	0	40	0.13	0.04	44
<i>SNHG25</i>	0	0	57	0.01	0.17	0
<i>SNHG6</i>	1	1	40	0.02	0.02	156
<i>LINC00339</i>	1	0	38	0.37	0.02	58
<i>LINC00263</i>	1	0	123	0.27	0.70	51
<i>SNHG30</i>	1	0	50	0.01	0.02	46
<i>LINC00493</i>	1	1	95	0.30	0.40	110
<i>SNHG15</i>	1	0	49	0.13	0.17	58
<i>LINC01420</i>	1	1	68	–	0.04	–

<sup>а</sup>Экспериментальное подтверждение.

данные рибосомного профайлинга (ribosome profiling, Ribo-Seq), доступные в веб-браузерах GWIPS-viz [23] и Trips-Viz [24], и определяли возможность взаимодействия целевых lncРНК с рибосомами: (0) – нет взаимодействия, (1) – есть. Дополнительно для комплексной оценки возможности трансляции lncРНК-хитов использовали методы оценки белоккодирующего потенциала CPC2 и CPAT, а также сведения о числе наборов данных рибосомного профайлинга, подтверждающих трансляцию lncРНК, из баз данных LncPep (<http://www.shengliilabs.com/LncPep/>) [25] и CPC2 (<http://cpc2.gao-lab.org/>) [26]. Кроме того, провели анализ публикаций по lncРНК-хитам с целью выяснить, для каких из них получено экспериментальное подтверждение трансляции пептидов: (0) – релевантных публикаций нет, (1) – релевантные публикации есть. Результаты анализа представлены в табл. 2.

Анализ данных из GWIPS-viz и Trips-Viz показал, что большинство lncРНК-хитов, за исключением SNHG25, имеют выраженные пики сигнала взаимодействия с рибосомой. Эти пики свидетельствуют о накоплении рибосом на перспективных сайтах начала трансляции lncРНК. Кроме того, используя данные из вышеуказанных веб-браузеров, мы определили примерную длину возможных коротких открытых рамок считывания (кОРС). Так, для NEAT1, EPB41L4A-AS1, LINC00263 и LINC00493 были выявлены кодирующие области длиной около

300 н., а для MIR17HG, SCARNA10 и LINC00339 – примерно в три раза меньше (около 90 н.).

Информация из LncPep, отражающая число наборов данных рибосомного профилирования, в которых была обнаружена трансляция выбранной lncРНК, подтвердила результаты, полученные с помощью GWIPS-viz и Trips-Viz. Достоверное подтверждение трансляции, согласно данным LncPep, получено для SNHG6 и LINC00493 – 156 и 110 наборов данных соответственно. Для остальных lncРНК-хитов число наборов данных варьировалось от 4 до 60.

Заметим, что значения кодирующего потенциала, полученные методами CPC2 и CPAT по шкале от 0 до 1, для lncРНК-хитов были низкими. Наибольшие значения по CPAT составили 0.76 для NEAT1 и 0.37 для LINC00339, а по CPC2 – 0.69 для LINC00263 и 0.59 для NEAT1. Диапазон значений для остальных lncРНК варьировался от 0.005 до 0.3.

В результате проведенного нами анализа литературы выявлено, что для некоторых выбранных нами lncРНК-хитов есть экспериментальные доказательства трансляции. Это относится к EPB41L4A-AS1 [60], CRNDE [61], SNHG6 [62], LINC00493 [63] и LINC01420 [64], для которых описаны кОРС и трансляция с них пептидов.

Однако для подтверждения высокого кодирующего потенциала NEAT1, SNHG1, SNHG7, SNHG12, SNHG15, SNHG16, MIR17HG, LINC00680, LINC00263 и LINC00339, а также для

оценки вклада транслируемых и транскрибируемых последовательностей вышеуказанных lncРНК в наблюдаемые клеточные и молекулярные фенотипы необходимы дальнейшие экспериментальные исследования. Стоит заметить, что при оценке кодирующего потенциала lncРНК с использованием методов на основе моделей машинного обучения важно соблюдать осторожность, так как при обучении эти модели используют аннотации lncРНК в качестве отрицательных примеров, что может привести к неточностям в оценках, полученных методами CPAT и CPC2. Использование данных рибосомного профайлинга для предварительной оценки возможности трансляции новых lncРНК дает более точный результат, чем существующие программные инструменты на основе методов машинного обучения.

*Функциональный анализ пептидов,  
транслируемых с lncРНК*

В настоящее время исследование функций пептидов, транслируемых с lncРНК, — актуальная задача молекулярной биологии. Роль пептидов остается малоизученной. Мы проанализировали функции пептидов, транслируемых с lncРНК-хитов. При поиске соответствующих публикаций мы нашли экспериментальные данные, полученные

для EPB41L4A-AS1 [60], CRNDE [61], SNHG6 [62], LINC00493 [63] и LINC01420 [64]. Транслируемые с них пептиды имеют размер от 40 а.о. до 120 а.о. Их можно разделить на три группы: структурные (NBDY, CRNDEP) — участвующие в образовании биомолекулярных конденсатов; митохондриальные (TIGA1, SMIM26) — связанные с процессами в митохондриях; сигнальные (SNHG6 ORF#2) — вовлеченные в сигнальные пути.

Продемонстрировано, что NBDY участвует в формировании Р-телец, а CRNDEP — в образовании стрессовых гранул. Доказано, что TIGA1 и SMIM26 вовлечены соответственно в регуляцию стабильности микротрубочек и процесс митохондриального импорта глутатиона. Недавно был открыт пептид SNHG6 ORF#2 и подтверждена его функция в активации сигнального пути TGF- $\beta$ /SMAD (табл. 3).

Установлено влияние CRNDEP, SMIM26, NBDY и SNHG6 ORF#2 на такие фундаментальные клеточные процессы, как миграция и пролиферация. Из этого следует, что трансляция пептидов может лежать в основе функций, ранее приписываемых соответствующим lncРНК. Функциональная оценка роли транслируемых и транскрибируемых последовательностей lncРНК требует экспериментального подтверждения.

**Таблица 3.** Функции пептидов, которые транслируются с коротких открытых рамок считывания lncРНК-хитов

lncРНК	Пептид	Функция	Источник
EPB41L4A-AS1	TIGA1	TIGA1 — митохондриальный пептид, выполняющий роль важнейшего посредника между микротрубочками и митохондриями. Его влияние на стабильность микротрубочек достигается за счет взаимодействия с $\alpha$ -тубулином. Истощение TIGA1 приводит к дестабилизации микротрубочек, что вызывает функциональное нарушение потенциал-зависимого анионного канала и провоцирует клеточный стресс. Описанный процесс, в свою очередь, активирует сигнальный путь p38 MAPK и усиливает накопление HIF-1 $\alpha$ .	[60]
CRNDE	CRNDEP	CRNDEP участвует в клеточной пролиферации, поскольку его эндогенная экспрессия повышена в высокопролиферирующих тканях. Кроме того, при искусственной оверэкспрессии CRNDEP может стимулировать образование стрессовых гранул.	[61]
SNHG6	SNHG6 ORF#1 и ORF#2	С предполагаемых открытых рамок считывания (ORF#1 и ORF#2) SNHG6 транслируются два пептида. Транслируемый пептид SNHG6 ORF#2 активирует путь TGF- $\beta$ /SMAD и способствует миграции клеток и ЭМП в ЭСКЧ. Предполагается участие пептидов, кодируемых SNHG6, в развитии стромальных и эпителиальных клеток эндометрия и связанных с этим гинекологических заболеваний.	[62]
LINC00493	SMIM26	Митохондриальный пептид SMIM26, кодируемый LINC00493, выступает в качестве белка-супрессора опухолей, особенно раковых. SMIM26 образует комплекс с AGK-SLC25A11 для поддержания митохондриального импорта глутатиона и ингибирования метастазирования светлоклеточной почечно-клеточной карциномы, опосредованного активацией сигнального пути AGK/АКТ.	[63]
LINC01420	NBDY	NBDY регулирует цитоплазматические рибонуклеопротеиновые гранулы, известные как Р-тела, и стабильность репортерных генов.	[64]

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Развитие исследований в области генома человека показало, что значительная часть транскриптов lncРНК играет гораздо большую роль в биологических процессах, чем считали ранее. В последние годы понимание регуляторной роли lncРНК в развитии организмов расширилось, в то время как функциональная значимость установлена менее чем для 10% транскриптов lncРНК.

В настоящее время для выявления функционально значимых lncРНК разработаны разные высокопроизводительные технологии. Нами проанализированы результаты одного из таких крупномасштабных скрининговых исследований, направленного на изучение функциональных локусов lncРНК в клетках человека.

На данный момент уже выявлено и детально изучено множество тканеспецифичных lncРНК. Эта особенность lncРНК косвенно свидетельствует об их важной роли в регуляции процессов и функций, характерных для клеток определенных тканей. В качестве примера можно привести MUNC, которая экспрессируется в мышечной ткани и участвует в дифференцировке клеток мышц [65], и PCGEM1, специфичную для ткани предстательной железы и ассоциированную с раком простаты [66]. Однако здесь мы сосредоточили внимание на широкоэкспрессирующихся lncРНК и их функциональных характеристиках, включая уровень экспрессии, степень консервативности и изученность. Кроме того, мы проанализировали взаимосвязь между показателями экспрессии и клеточными функциями, а также оценили кодирующий потенциал lncРНК.

Анализ экспериментальных данных, полученных в результате исследований функциональных локусов lncРНК на основе CRISPRi [15], позволил отобрать 18 lncРНК, включающих как малоизученные, так и хорошо известные гены. Мы подтвердили высокий уровень экспрессии отобранных lncРНК-хитов с помощью данных из GTEx v8. Более того, мы обнаружили отсутствие прямой корреляции между принадлежностью к группе НК-генов и высоким уровнем экспрессии. Еще одним важным наблюдением стало то, что выявленные эффекты нокдауна, специфичные для конкретного типа клеток, частично согласуются с данными по образцам тканей из GTEx v8.

Для отобранных lncРНК общая консервативность нуклеотидных последовательностей была невысокой, однако определены отдельные высококонсервативные участки, такие как сегменты экзонных последовательностей, области транскрипции snoРНК и miРНК. Ретроспективный анализ публикаций подтвердил функциональную значимость выбранных lncРНК.

Выявленные функции lncРНК-хитов согласуются с результатами оригинального комплексного

исследования, посвященного изучению связи между lncРНК, модифицирующими клеточный рост, и их роли в патогенезе онкологических заболеваний. Мы определили основные механизмы, с помощью которых lncРНК реализуют свои функции, включая конкурентное связывание с miРНК и участие в сигнальных путях. Заметим, что большинство выбранных генов входят в семейство генов-хозяев snoРНК (SNHG). У человека более 50% snoРНК происходят из генов-хозяев, причем около 80% из них относятся к семейству SNHG. В настоящее время это семейство насчитывает более 20 генов. Среди них наиболее изучены SNHG1, SNHG6 и SNHG7. По результатам многочисленных исследований установлено, что это онкогены, вовлеченные в канцерогенез многих типов злокачественных опухолей и патогенез различных заболеваний.

Результаты многих исследований подтверждают, что NEAT1 – функционально консервативная lncРНК, которая высоко экспрессирована в клетках и тканях млекопитающих. NEAT1 взаимодействует с различными молекулярными партнерами, включая белки, miРНК, mРНК и ДНК. Доказана ее функциональная значимость, в частности при различных онкологических заболеваниях, где NEAT1 участвует в ответе клеток на повреждение ДНК [29]. Имеются и данные о том, что NEAT1 влияет на развитие нейродегенеративных заболеваний, включая болезнь Хантингтона, боковой амиотрофический склероз и болезнь Паркинсона [67]. Кроме того, NEAT1 является глобальным регулятором, который действует путем активации врожденного иммунитета при вирусной инфекции [68]. NEAT1 считается перспективной терапевтической мишенью и диагностическим биомаркером. Клинические испытания lncРНК NEAT1 начались в 2021 году. Ее клиническая значимость для онкологии уже доказана – экспрессия NEAT1 коррелирует с ответом пациентов на анти-PD-1/PD-L1-терапию при меланоме и глиобластоме [69].

Другой пример – новая перспективная lncРНК LINC00339. Многочисленные экспериментальные данные свидетельствуют о ее повышенной экспрессии при эндометриозе, апоптозе кардиомиоцитов, остеопорозе и онкологии [37, 70, 71]. Кроме того, выявлена корреляция между LINC00339 и клиническими показателями у онкологических больных: стадией заболевания, наличием метастаз в лимфатические узлы, степенью патологии и др. Также доказано, что LINC00339 влияет на базовые клеточные процессы опухолевых клеток, такие как пролиферация, подвижность и инвазивность, способствуя росту опухоли. Спектр механизмов, обеспечивающих функционирование LINC00339, включает привлечение miРНК, взаимодействие с транскрипционными факторами, участие в сигнальных путях Wnt/ $\beta$ -катенин, MAPK и RhoA. Значимые эффекты LINC00339 при различных заболеваниях свидетельствуют о возможности ее

использования в качестве диагностического маркера и терапевтической мишени [37].

В ходе анализа данных, полученных с помощью CRISPRi, мы также отобрали для дальнейших экспериментальных исследований ряд малоизученных lncРНК, таких как SCARNA10, LINC00680, SHNG25, LINC00263, SNHG30, LINC01420, LINC00493.

Еще одна актуальная задача молекулярной биологии – оценка белоккодирующего потенциала lncРНК. Исследования широкого спектра организмов выявили существование kОПС, кодируемых практически всеми классами РНК. Во многих работах продемонстрирована трансляция пептидов с kОПС, расположенных в мРНК, интронах пре-мРНК, lncРНК и даже в первичных транскриптах miРНК и рибосомной РНК. Наиболее перспективным источником таких рамок считывания считаются lncРНК. Проведенный нами анализ данных рибосомного профайлинга показал, что большинство lncРНК с высокой вероятностью транслирует небольшие пептиды. Рассчитанная нами длина kОПС lncРНК-хитов согласуется со средним значением в 24 кодона для lncРНК, исследованных J. Courso с соавт. [72]. Оценки, полученные на основе моделей CPAT и CPC2, как правило, носят случайный характер и могут быть ошибочными. Например, для LINC01420 с экспериментально подтвержденной трансляцией функционального пептида значение, полученное с помощью CPC2, составляло 0.04. Подобные результаты указывают на основную проблему моделей машинного обучения при выборе обучающего набора данных.

Для пяти lncРНК-хитов: EPB41L4A-AS1 [60], CRNDE [61], SNHG6 [62], LINC00493 [63] и LINC01420 [64] – мы нашли подробные описания функций транслируемых пептидов. Так, NBDY и CRNDEP участвуют в образовании биомолекулярных конденсатов, TIGA1 и SMIM26 связаны с митохондриальными процессами, а SNHG6 ORF#2 активирует сигнальный путь TGF- $\beta$ /SMAD. Кроме того, большинство идентифицированных пептидов участвует в регуляции основных клеточных функций. Таким образом, в результате анализа уже опубликованных данных по трансляции lncРНК стало понятно, что назрела необходимость пересмотра их аннотации и функций в масштабе всего генома.

Нами выявлен ряд lncРНК с высокой вероятностью трансляции малых пептидов: NEAT1, SNHG1, SNHG7, SNHG12, SNHG15, SNHG16, MIR17HG, LINC00680, LINC00263 и LINC00339. Экспериментальное исследование их возможной трансляции позволит углубить наши знания о регуляторной сети генома.

Анализ истории версий GENCODE показывает, что аннотации генов человека постоянно обновлялись как для белоккодирующих генов, так и для lncРНК [73]. В настоящее время прежнее

преобладание числа белоккодирующих генов над общим числом генов lncРНК изменилось на обратное – преобладание lncРНК. Версия GENCODE26 (V26), выпущенная в год публикации данных крупномасштабного исследования [15], содержала в общей сложности 58219 аннотированных генов, из которых белоккодирующих было 19817 и lncРНК-кодирующих – 15787. Текущая версия GENCODE44 (V44), выпущенная в 2023 году, содержит в общей сложности 62700 аннотаций, из которых 19396 белоккодирующих и 19922 lncРНК. Выявленная корреляция между паттернами экспрессии lncРНК и их способностью кодировать пептиды расширяет наше понимание транскрипционного потенциала генома и его функциональных участков. Возможно, что в следующей версии аннотации генома человека число белоккодирующих генов снова превзойдет таковое для lncRNA. Не исключены изменения в номенклатуре типов транскриптов и генов.

Исследование выполнено в рамках Государственного задания Министерства науки и высшего образования Российской Федерации для Медико-генетического научного центра.

В данной статье не содержится исследований с участием животных, выполненных кем-либо из авторов. В данной статье не содержится исследований с участием людей, выполненных кем-либо из авторов.

Авторы заявляют об отсутствии конфликта интересов.

## ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Дополнительные материалы: Функции lncРНК-хитов (табл. S1, см на сайте [http://www.molecbio.ru/downloads/2024/3/supp\\_Konina\\_rus.pdf](http://www.molecbio.ru/downloads/2024/3/supp_Konina_rus.pdf)).

## СПИСОК ЛИТЕРАТУРЫ

1. Mattick J.S., Amaral P.P., Carninci P., Carpenter S., Chang H.Y., Chen L.L., Chen R., Dean C., Dinger M.E., Fitzgerald K.A., Gingeras T.R., Guttman M., Hirose T., Huarte M., Johnson R., Kanduri C., Kapranov P., Lawrence J.B., Lee J.T., Mendell J.T., Mercer T.R., Moore K.J., Nakagawa S., Rinn J.L., Spector D.L., Ulitsky I., Wan Y., Wilusz J.E., Wu M. (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **24**, 430–447.
2. Wang K.C., Chang H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell.* **43**, 904–914.
3. Ponting C.P., Haerty W. (2022) Genome-wide analysis of human long noncoding RNAs: a provocative review. *Annu. Rev. Genomics Hum. Genet.* **23**, 153–172.
4. Garcia-Padilla C., Duenas A., Garcia-Lopez V., Aranea A., Franco D., Garcia-Martinez V., Lopez-Sanchez C. (2022) Molecular mechanisms of lncRNAs in the dependent regulation of cancer and their potential therapeutic use. *Int. J. Mol. Sci.* **23**, 764.

5. Yoon J.H., Abdelmohsen K., Gorospe M. (2013) Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* **425**, 3723–3730.
6. Karakas D., Ozpolat B. (2021) The role of LncRNAs in translation. *Noncoding RNA*. **7**, 16.
7. Statello L., Guo C.J., Chen L.L., Huarte M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118.
8. Guh C.Y., Hsieh Y.H., Chu H.P. (2020) Functions and properties of nuclear lncRNAs—from systematically mapping the interactomes of lncRNAs. *J. Biomed. Sci.* **27**, 44.
9. Oo J.A., Brandes R.P., Leisegang M.S. (2022) Long non-coding RNAs: novel regulators of cellular physiology and function. *Pflugers Arch.* **474**, 191–204.
10. DiStefano J.K. (2018) The emerging role of long noncoding RNAs in human disease. *Methods Mol. Biol.* **1706**, 91–110.
11. Frankish A., Carbonell-Sala S., Diekhans M., Jungreis I., Loveland J.E., Mudge J.M., Sisu C., Wright J.C., Arnan C., Barnes I., Banerjee A., Bennett R., Berry A., Bignell A., Boix C., Calvet F., Cerdán-Vélez D., Cunningham F., Davidson C., Donaldson S., Dursun C., Fatima R., Giorgetti S., Giron C.G., Gonzalez J.M., Hardy M., Harrison P.W., Hourlier T., Hollis Z., Hunt T., James B., Jiang Y., Johnson R., Kay M., Lagarde J., Martin F.J., Gómez L.M., Nair S., Ni P., Pozo F., Ramalingam V., Ruffier M., Schmitt B.M., Schreiber J.M., Steed E., Suner M.M., Sumathipala D., Sycheva I., Uszczyńska-Ratajczak B., Wass E., Yang Y.T., Yates A., Zafrulla Z., Choudhary J.S., Gerstein M., Guigo R., Hubbard T.J.P., Kellis M., Kundaje A., Paten B., Tress M.L., Flicek P. (2023) GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.* **51**, D942–D949.
12. Stojic L., Lun A.T.L., Mascalchi P., Ernst C., Redmond A.M., Mangei J., Barr A.R., Bousgouni V., Bakal C., Marioni J.C., Odom D.T., Gergely F. (2020) A high-content RNAi screen reveals multiple roles for long noncoding RNAs in cell division. *Nat. Commun.* **11**, 1851.
13. Ramilowski J.A., Yip C.W., Agrawal S., Chang J.C., Ciani Y., Kulakovskiy I.V., Mendez M., Ooi J.L.C., Ouyang J.F., Parkinson N., Petri A., Roos L., Severin J., Yasuzawa K., Abugessaisa I., Akalin A., Antonov I.V., Arner E., Bonetti A., Bono H., Borsari B., Brombacher F., Cameron C.J., Cannistraci C.V., Cardenas R., Cardon M., Chang H., Dostie J., Ducoli L., Favorov A., Fort A., Garrido D., Gil N., Gimenez J., Guler R., Handoko L., Harshbarger J., Hasegawa A., Hasegawa Y., Hashimoto K., Hayatsu N., Heutink P., Hirose T., Imada E.L., Itoh M., Kaczkowski B., Kanhere A., Kawabata E., Kawaji H., Kawashima T., Kelly S.T., Kojima M., Kondo N., Koseki H., Kouno T., Kratz A., Kurowska-Stolarska M., Kwon A.T.J., Leek J., Lennartsson A., Lizio M., López-Redondo F., Luginbühl J., Maeda S., Makeev V.J., Marchionni L., Medvedeva Y.A., Minoda A., Müller F., Muñoz-Aguirre M., Murata M., Nishiyori H., Nitta K.R., Noguchi S., Noro Y., Nurtdinov R., Okazaki Y., Orlando V., Paquette D., Parr C.J.C., Rackham O.J.L., Rizzo P., Sánchez-Martinez D.F., Sandelin A., Sanjana P., Sempke C.A.M., Shibayama Y., Sivaraman D.M., Suzuki T., Szumowski S.C., Tagami M., Taylor M.S., Terao C., Thodberg M., Thongjuea S., Tripathi V., Ulitsky I., Verardo R., Vorontsov I.E., Yamamoto C., Young R.S., Baillie J.K., Forrest A.R.R., Guigó R., Hoffman M.M., Hon C.C., Kasukawa T., Kauppinen S., Kere J., Lenhard B., Schneider C., Suzuki H., Yagi K., de Hoon M.J.L., Shin J.W., Carninci P. (2020) Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res.* **30**, 1060–1072.
14. Guttman M., Donaghey J., Carey B.W., Garber M., Grenier J.K., Munson G., Young G., Lucas A.B., Ach R., Bruhn L., Yang X., Amit I., Meissner A., Regev A., Rinn J.L., Root D.E., Lander E.S. (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. **477**, 295–300.
15. Liu S.J., Horlbeck M.A., Cho S.W., Birk H.S., Malatesta M., He D., Attenello F.J., Villalta J.E., Cho M.Y., Chen Y., Mandegar M.A., Olvera M.P., Gilbert L.A., Conklin B.R., Chang H.Y., Weissman J.S., Lim D.A. (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*. **355**, aah7111.
16. Yates A., Akanni W., Amode M.R., Barrell D., Billis K., Carvalho-Silva D., Cummins C., Clapham P., Fitzgerald S., Gil L., Girón C.G., Gordon L., Hourlier T., Hunt S.E., Janacek S.H., Johnson N., Juettemann T., Keenan S., Lavidas I., Martin F.J., Maurel T., McLaren W., Murphy D.N., Nag R., Nuhn M., Parker A., Patricio M., Pignatelli M., Rahtz M., Riat H.S., Sheppard D., Taylor K., Thormann A., Vullo A., Wilder S.P., Zadissa A., Birney E., Harrow J., Muffato M., Perry E., Ruffier M., Spudich G., Trevanion S.J., Cunningham F., Aken B.L., Zerbino D.R., Flicek P. (2016) Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716.
17. Iyer M.K., Niknafs Y.S., Malik R., Singhal U., Sahu A., Hosono Y., Barrette T.R., Prensner J.R., Evans J.R., Zhao S., Poliakov A., Cao X., Dhanasekaran S.M., Wu Y.M., Robinson D.R., Beer D.G., Feng F.Y., Iyer H.K., Chinnaiyan A.M. (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208.
18. Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega B., Regev A., Rinn J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927.
19. Liu S.J., Nowakowski T.J., Pollen A.A., Lui J.H., Horlbeck M.A., Attenello F.J., He D., Weissman J.S., Kriegstein A.R., Diaz A.A., Lim D.A. (2016) Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67.
20. Carithers L.J., Moore H.M. (2015) The genotype-tissue expression (GTEx) project. *Biopreserv. Biobank.* **13**, 307–308.
21. FANTOM Consortium and the RIKEN PMI and CLST (DGT); Forrest A.R., Kawaji H., Rehli M., Baillie J.K., de Hoon M.J., Haberle V., Lassmann T., Kulakovskiy I.V., Lizio M., Itoh M., Andersson R., Mungall C.J., Meehan T.F., Schmeier S., Bertin N., Jørgensen M., Dimont E., Arner E., Schmidl C., Schaefer U., Medvedeva Y.A., Plessy C., Vitezic M., Severin J., Sempke C., Ishizu Y., Young R.S., Francescato M., Alam I., Albanese D., Altschuler G.M., Arakawa T., Archer J.A., Arner P., Babina M., Rennie S., Balwiercz P.J., Beckhouse A.G., Pradhan-Bhatt S., Blake J.A., Blumenthal A., Bodega B., Bonetti A., Briggs J., Brombacher F., Burroughs A.M., Califano A., Cannistraci C.V., Carbajo D., Chen Y., Chierici M., Ciani Y., Clevers H.C., Dalla E., Davis C.A., Detmar M., Diehl A.D., Dohi T., Drabløs F., Edge A.S., Edinger M., Ekwall K., Endoh M., Enomoto H., Fagiolini M., Fairbairn L., Fang H., Farach-Carson M.C., Faulkner G.J., Favorov A.V., Fisher M.E., Frith M.C., Fujita R., Fukuda S., Furlanello C., Furino M., Furusawa J., Geijtenbeek T.B., Gibson A.P., Gingeras T., Goldowitz D., Gough J., Guhl S.,

- Guler R., Gustincich S., Ha T.J., Hamaguchi M., Hara M., Harbers M., Harshbarger J., Hasegawa A., Hasegawa Y., Hashimoto T., Herlyn M., Hitchens K.J., Ho Sui S.J., Hofmann O.M., Hoof I., Hori F., Huminiński L., Iida K., Ikawa T., Jankovic B.R., Jia H., Joshi A., Jurman G., Kaczkowski B., Kai C., Kaida K., Kaiho A., Kajiyama K., Kanamori-Katayama M., Kasianov A.S., Kasukawa T., Katayama S., Kato S., Kawaguchi S., Kawamoto H., Kawamura Y.I., Kawashima T., Kempfle J.S., Kenna T.J., Kere J., Khachigian L.M., Kitamura T., Klinken S.P., Knox A.J., Kojima M., Kojima S., Kondo N., Koseki H., Koyasu S., Krampitz S., Kubosaki A., Kwon A.T., Laros J.F.J., Lee W., Lennartsson A., Li K., Lilje B., Lipovich L., Mackay-Sim A., Manabe R., Mar J.C., Marchand B., Mathelier A., Mejhert N., Meynert A., Mizuno Y., de Lima Morais D.A., Morikawa H., Morimoto M., Moro K., Motakis E., Motohashi H., Mummery C.L., Murata M., Nagao-Sato S., Nakachi Y., Nakahara F., Nakamura T., Nakamura Y., Nakazato K., van Nimwegen E., Ninomiya N., Nishiyori H., Noma S., Noma S., Noazaki T., Ogishima S., Ohkura N., Ohimiya H., Ohno H., Ohshima M., Okada-Hatakeyama M., Okazaki Y., Orlando V., Ovchinnikov D.A., Pain A., Passier R., Patrikakis M., Persson H., Piazza S., Prendergast J.G.D., Rackham O.J.L., Ramiłowski J.A., Rashid M., Ravasi T., Rizzu P., Roncador M., Roy S., Rye M.B., Saijyo E., Sajantila A., Saka A., Sakaguchi S., Sakai M., Sato H., Savvi S., Saxena A., Schneider C., Schultes E.A., Schulze-Tanzil G.G., Schwegmann A., Sengstag T., Sheng G., Shimoji H., Shimoni Y., Shin J.W., Simon C., Sugiyama D., Sugiyama T., Suzuki M., Suzuki N., Swoboda R.K., 't Hoen P.A.C., Tagami M., Takahashi N., Takai J., Tanaka H., Tatsukawa H., Tatum Z., Thompson M., Toyodo H., Toyoda T., Valen E., van de Wetering M., van den Berg L.M., Verado R., Vijayan D., Vorontsov I.E., Wasserman W.W., Watanabe S., Wells C.A., Winteringham L.N., Wolvetang E., Wood E.J., Yamaguchi Y., Yamamoto M., Yoneda M., Yonekura Y., Yoshida S., Zabierowski S.E., Zhang P.G., Zhao X., Zucchelli S., Summers K.M., Suzuki H., Daub C.O., Kawai J., Heutink P., Hide W., Freeman T.C., Lenhard B., Bajic V.B., Taylor M.S., Makeev V.J., Sandelin A., Hume D.A., Carninci P., Hayashizaki Y. (2014) A promoter-level mammalian expression atlas. *Nature*. **507**, 462–470.
22. Nassar L.R., Barber G.P., Benet-Pages A., Casper J., Clawson H., Diekhans M., Fischer C., Gonzalez J.N., Hinrichs A.S., Lee B.T., Lee C.M., Muthuraman P., Nguy B., Pereira T., Nejad P., Perez G., Raney B.J., Schmelter D., Speir M.L., Wick B.D., Zweig A.S., Haussler D., Kuhn R.M., Haussler M., Kent W.J. (2023) The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.* **51**, D1188–D1195.
23. Michel A.M., Fox G., Kiran A.M., De Bo C., O'Connor P.B., Heaphy S.M., Mullan J.P., Donohue C.A., Higgins D.G., Baranov P.V. (2014) GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.* **42**, D859–D864.
24. Kiniry S.J., Judge C.E., Michel A.M., Baranov P.V. (2021) Trips-Viz: an environment for the analysis of public and user-generated ribosome profiling data. *Nucleic Acids Res.* **49**, W662–W670.
25. Liu T., Wu J., Wu Y., Hu W., Fang Z., Wang Z., Jiang C., Li S. (2022) LncPep: a resource of translational evidences for lncRNAs. *Front. Cell Dev. Biol.* **10**, 795084.
26. Kang Y.J., Yang D.C., Kong L., Hou M., Meng Y.Q., Wei L., Gao G. (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16.
27. Konina D., Sparber P., Viakhireva I., Filatova A., Skoblov M. (2021) Investigation of LINC00493/SMIM26 gene suggests its dual functioning at mRNA and protein level. *Int. J. Mol. Sci.* **22**, 8477.
28. Uszczynska-Ratajczak B., Lagarde J., Frankish A., Guigo R., Johnson R. (2018) Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **19**, 535–548.
29. Farzaneh M., Masoodi T., Ghaedrahmati F., Radoszkiewicz K., Anbiyaiee A., Sheykhi-Sabzehpoush M., Rad N.K., Uddin S., Jooybari S.P.M., Khoshnam S.E., Azizidoost S. (2023) An updated review of contribution of long noncoding RNA-NEAT1 to the progression of human cancers. *Pathol. Res. Pract.* **245**, 154380.
30. Wang X., Jiang Q., Zhang C., Yang Q., Wang L., Zhang J., Wang L., Chen X., Hou X., Han D., Wu J., Zhao S. (2021) Long noncoding RNA SNHG12 is a potential diagnostic and prognostic biomarker in various tumors. *Chin. Neurosurg J.* **7**, 37.
31. Gong C.Y., Tang R., Nan W., Zhou K.S., Zhang H.H. (2020) Role of SNHG16 in human cancer. *Clin. Chim. Acta.* **503**, 175–180.
32. Thin K.Z., Tu J.C., Raveendran S. (2019) Long non-coding SNHG1 in cancer. *Clin. Chim. Acta.* **494**, 38–47.
33. Bian Z., Ji W., Xu B., Huang W., Jiao J., Shao J., Zhang X. (2020) The role of long noncoding RNA SNHG7 in human cancers (Review). *Mol. Clin. Oncol.* **13**, 45.
34. Morelli E., Fulciniti M., Samur M.K., Ribeiro C.F., Wert-Lamas L., Henninger J.E., Gulla A., Aktas-Samur A., To-doerti K., Talluri S., Park W.D., Federico C., Scionti F., Amodio N., Bianchi G., Johnstone M., Liu N., Gramegna D., Maisano D., Russo N.A., Lin C., Tai Y.T., Neri A., Chauhan D., Hideshima T., Shammas M.A., Tassone P., Gryaznov S., Young R.A., Anderson K.C., Novina C.D., Loda M., Munshi N.C. (2023) A MIR17HG-derived long noncoding RNA provides an essential chromatin scaffold for protein interaction and myeloma growth. *Blood*. **141**, 391–405.
35. Ghafouri-Fard S., Safarzadeh A., Hussen B.M., Taheri M., Mokhtari M. (2023) Contribution of CRNDE lncRNA in the development of cancer and the underlying mechanisms. *Pathol. Res. Pract.* **244**, 154387.
36. Ghafouri-Fard S., Khoshbakht T., Taheri M., Shojaei S. (2021) A review on the role of small nucleolar RNA host gene 6 long non-coding RNAs in the carcinogenic processes. *Front. Cell Dev. Biol.* **9**, 741684.
37. Wu Z., Zhang S., Guo W., He Y. (2022) LINC00339: an emerging major player in cancer and metabolic diseases. *Biomed. Pharmacother.* **149**, 112788.
38. Lee W.J., Shin C.H., Ji H., Jeong S.D., Park M.S., Won H.H., Pandey P.R., Tsitsipatis D., Gorospe M., Kim H.H. (2021) hnRNPK-regulated LINC00263 promotes malignant phenotypes through miR-147a/CAPN2. *Cell Death. Dis.* **12**, 290.
39. Cui H., Ruan M., Xu H., Qi J., Ruan L., Gao X., Sun X., Zhang S., Zuo R., Yin Y. (2021) LINC01420 serves as a novel prognostic biomarker and promotes cell proliferation, migration, and invasion by suppressing miR-149-5p in gastric cancer. *Crit. Rev. Eukaryot. Gene Exp.* **31**, 49–58.

40. Ma F., Lei Y.Y., Ding M.G., Luo L.H., Xie Y.C., Liu X.L. (2020) LncRNA NEAT1 interacted with DNMT1 to regulate malignant phenotype of cancer cell and cytotoxic T Cell infiltration via epigenetic inhibition of p53, cGAS, and STING in lung cancer. *Front. Genet.* **11**, 250.
41. Yu X., Song M.S., Rong P.Z., Chen X.J., Shi L., Wang C.H., Pang Q.J. (2022) LncRNA SNHG1 modulates adipogenic differentiation of BMSCs by promoting DNMT1 mediated Opg hypermethylation via interacting with PTBP1. *J. Cell. Mol. Med.* **26**, 60–74.
42. Zhang G.D., Gai P.Z., Liao G.Y., Li Y. (2019) LncRNA SNHG7 participates in osteosarcoma progression by down-regulating p53 via binding to DNMT1. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 3602–3610.
43. Xie S.C., Zhang J.Q., Jiang X.L., Hua Y.Y., Xie S.W., Qin Y.A., Yang Y.J. (2020) LncRNA CRNDE facilitates epigenetic suppression of CELF2 and LATS2 to promote proliferation, migration and chemoresistance in hepatocellular carcinoma. *Cell Death. Dis.* **11**, 676.
44. Yang H., Jia W., Zhang D.P., Xu Y.X., Wang F., Wang N., Zhu D.F. (2022) Long non-coding RNAs and small nucleolar RNA host gene 6 contribute to the depression-like behavior of hypothyroid mice by promoting methylation of the brain-derived neuropathic factor promoter that is mediated by DNA methyltransferase 1. *J. Physiol. Pharmacol.* **73**(5). doi: 10.26402/jpp.2022.5.09
45. Ma Z., Huang H., Wang J., Zhou Y., Pu F., Zhao Q., Peng P., Hui B., Ji H., Wang K. (2017) Long non-coding RNA SNHG15 inhibits P15 and KLF2 expression to promote pancreatic cancer proliferation through EZH2-mediated H3K27me3. *Oncotarget.* **8**, 84153–84167.
46. Hu X., Li F., He J., Yang J., Jiang Y., Jiang M., Wei D., Chang L., Hejtmanec J.F., Hou L., Ma X. (2021) LncRNA NEAT1 recruits SFPQ to regulate MITF splicing and control RPE cell proliferation. *Invest. Ophthalmol. Vis. Sci.* **62**, 18.
47. Shen Y., Liu S., Fan J., Jin Y., Tian B., Zheng X., Fu H. (2017) Nuclear retention of the lncRNA SNHG1 by doxorubicin attenuates hnRNPC-p53 protein interactions. *EMBO Rep.* **18**, 536–548.
48. Zhang J., Zhang R., Ye Y. (2021) Long non-coding RNA (LncRNA) SNHG7/ Eukaryotic translation initiation factor 4 gamma 2 (EIF4G2) involves in the malignant events of ovarian cancer cells with paclitaxel resistant. *Bioengineered.* **12**, 10541–10552.
49. Wu Y., Sun Y., Xu B., Yang M., Wang X., Zhao X. (2023) SCARNA10 regulates p53 acetylation-dependent transcriptional activity. *Biochem. Biophys. Res. Commun.* **669**, 38–45.
50. Zhang F., Wang H., Yu J., Yao X., Yang S., Li W., Xu L., Zhao L. (2021) LncRNA CRNDE attenuates chemoresistance in gastric cancer via SRSF6-regulated alternative splicing of PICALM. *Mol. Cancer.* **20**, 6.
51. Sun X., Wang R., Tan M., Tian X., Meng J. (2021) LncRNA LINC00680 promotes lung adenocarcinoma growth via binding to GATA6 and canceling GATA6-mediated suppression of SOX12 expression. *Exp. Cell. Res.* **405**, 112653.
52. Lan Z., Yao X., Sun K., Li A., Liu S., Wang X. (2020) The interaction between lncRNA SNHG6 and hnRNPA1 contributes to the growth of colorectal cancer by enhancing aerobic glycolysis through the regulation of alternative splicing of PKM. *Front. Oncol.* **10**, 363.
53. Zhai H., Zhang X., Sun X., Zhang D., Ma S. (2020) Long non-coding RNA LINC01420 contributes to pancreatic cancer progression through targeting KRAS proto-oncogene. *Dig. Dis. Sci.* **65**, 1042–1052.
54. Taiana E., Ronchetti D., Todoerti K., Nobili L., Tassone P., Amodio N., Neri A. (2020) LncRNA NEAT1 in paraspeckles: a structural scaffold for cellular DNA damage response systems? *Noncoding RNA.* **6**, 26.
55. Na Z., Luo Y., Schofield J.A., Smelyansky S., Khitun A., Muthukumar S., Valkov E., Simon M.D., Slavoff S.A. (2020) The NBDY microprotein regulates cellular RNA decapping. *Biochemistry.* **59**, 4131–4142.
56. Ren J., Li Y., Wuermanbieke S., Hu S., Huang G. (2022) N(6)-methyladenosine (m(6)A) methyltransferase METTL3-mediated LINC00680 accelerates osteoarthritis through m(6)A/SIRT1 manner. *Cell Death. Discov.* **8**, 240.
57. Jing L., Li S., Wang J., Zhang G. (2019) Long non-coding RNA small nucleolar RNA host gene 7 facilitates cardiac hypertrophy via stabilization of SDA1 domain containing 1 mRNA. *J. Cell. Biochem.* **120**, 15089–15097.
58. Feng Y., Gao L., Cui G., Cao Y. (2020) LncRNA NEAT1 facilitates pancreatic cancer growth and metastasis through stabilizing ELF3 mRNA. *Am. J. Cancer Res.* **10**, 237–248.
59. Wu X., Qiu L., Feng H., Zhang H., Yu H., Du Y., Wu H., Zhu S., Ruan Y., Jiang H. (2022) KHDRBS3 promotes paclitaxel resistance and induces glycolysis through modulated MIR17HG/CLDN6 signaling in epithelial ovarian cancer. *Life Sci.* **293**, 120328.
60. Liao M., Liao W., Xu N., Li B., Liu F., Zhang S., Wang Y., Wang S., Zhu Y., Chen D., Xie W., Jiang Y., Cao L., Yang B.B., Zhang Y. (2019) LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. *EBioMedicine.* **41**, 200–213.
61. Szafron L.M., Balcerak A., Grzybowska E.A., Pienkowska-Grela B., Felisiak-Golabek A., Podgorska A., Kulesza M., Nowak N., Pomorski P., Wysocki J., Rubel T., Dansonka-Mieszkowska A., Konopka B., Lukasik M., Kupryjanczyk J. (2015) The novel gene CRNDE encodes a nuclear peptide (CRNDEP) which is overexpressed in highly proliferating tissues. *PLoS One.* **10**, e0127475.
62. Zou Q., Du X., Zhou L., Yao D., Dong Y., Jin J. (2023) A short peptide encoded by long non-coding RNA small nucleolar RNA host gene 6 promotes cell migration and epithelial-mesenchymal transition by activating transforming growth factor-beta/SMAD signaling pathway in human endometrial cells. *J. Obstet. Gynaecol. Res.* **49**, 232–242.
63. Yeasmin F., Imamachi N., Tanu T., Taniue K., Kawamura T., Yada T., Akimitsu N. (2021) Identification and analysis of short open reading frames (sORFs) in the initially annotated noncoding RNA LINC00493 from human cells. *J. Biochem.* **169**, 421–434.
64. D’Lima N.G., Ma J., Winkler L., Chu Q., Loh K.H., Corpuz E.O., Budnik B.A., Lykke-Andersen J., Saghatelian A., Slavoff S.A. (2017) A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* **13**, 174–180.
65. Mueller A.C., Cichewicz M.A., Dey B.K., Layer R., Reon B.J., Gagan J.R., Dutta A. (2015) MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol. Cell. Biol.* **35**, 498–513.
66. Srikantan V., Zou Z., Petrovics G., Xu L., Augustus M., Davis L., Livezey J.R., Connell T., Sesterhenn I.A., Yoshino K., Buzard G.S., Mostofi F.K., McLeod D.G.,

- Moul J.W., Srivastava S. (2000) PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc. Natl. Acad. Sci. USA*. **97**, 12216–12221.
67. Li K., Wang Z. (2023) lncRNA NEAT1: key player in neurodegenerative diseases. *Ageing Res. Rev.* **86**, 101878.
68. Morchikh M., Cribier A., Raffel R., Amraoui S., Cau J., Severac D., Dubois E., Schwartz O., Bennasser Y., Benkirane M. (2017) HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. *Mol. Cell.* **67**, 387–399. e385.
69. Toker J., Iorgulescu J.B., Ling A.L., Villa G.R., Gadet J., Parida L., Getz G., Wu C.J., Reardon D.A., Chiocca E.A., Mineo M. (2023) Clinical importance of the lncRNA NEAT1 in cancer patients treated with immune checkpoint inhibitors. *Clin. Cancer Res.* **29**, 2226–2238.
70. Holdsworth-Carson S.J., Churchill M., Donoghue J.F., Mortlock S., Fung J.N., Sloggett C., Chung J., Cann L., Teh W.T., Campbell K.R., Luwor R., Healey M., Montgomery G., Girling J.E., Rogers P.A.W. (2021) Elucidating the role of long intergenic non-coding RNA 339 in human endometrium and endometriosis. *Mol. Hum. Reprod.* **27**, gaab010.
71. Chen X.F., Zhu D.L., Yang M., Hu W.X., Duan Y.Y., Lu B.J., Rong Y., Dong S.S., Hao R.H., Chen J.B., Chen Y.X., Yao S., Thynn H.N., Guo Y., Yang T.L. (2018) An Osteoporosis Risk SNP at 1p36.12 acts as an allele-specific enhancer to modulate LINC00339 expression via long-range loop formation. *Am.J. Hum. Genet.* **102**, 776–793.
72. Couso J.P., Patraquim P. (2017) Classification and function of small open reading frames. *Nat. Rev. Mol. Cell. Biol.* **18**, 575–589.
73. Jalali S., Gandhi S., Scaria V. (2016) Navigating the dynamic landscape of long noncoding RNA and protein-coding gene annotations in GENCODE. *Hum. Genomics.* **10**, 35.

## REVISION OF FUNCTIONALLY RELEVANT AND WIDELY EXPRESSED LONG NON-CODING RNAs

D. O. Konina<sup>1</sup>, \*, M. Y. Skoblov<sup>1</sup>

<sup>1</sup>Laboratory of Functional Genomics, Research Centre for Medical Genetics, Moscow, 115522 Russia

\*e-mail: darya.konina@phystech.edu

Long non-coding RNAs (lncRNAs) are involved in many cellular processes while displaying high tissue specificity. In contrast, protein-coding genes, including the category of housekeeping ones, exhibit broad expression patterns. The aim of this study was to highlight the functional importance of widely expressed lncRNAs. We analyzed experimental data from cell-growth screen of lncRNA loci in human cells, which allowed us to identify 18 lncRNA hits. Notably, these lncRNAs were not only widely expressed in most human tissues, but also played functional roles within them. Detail investigation revealed them encompass a variety of molecular functions, from cardiomyocyte damage controlling to macrophage class switching. Interestingly, experimental data highlighted the fact that a significant part of these lncRNAs encoded small but functional polypeptides. A set of lncRNAs, NEAT1, SNHG1, SNHG7, SNHG12, SNHG15, SNHG16, MIR17HG, LINC00680, LINC00263 and LINC00339, that were highly likely to be translated into small polypeptides was identified. Additionally, for EPB41L4A-AS1, CRNDE, SNHG6, LINC00493, and LINC01420, a dual function associated with both the RNA sequences and small proteins they encoded was established.

**Keywords:** CRISPRi, long non-coding RNA, short open reading frame