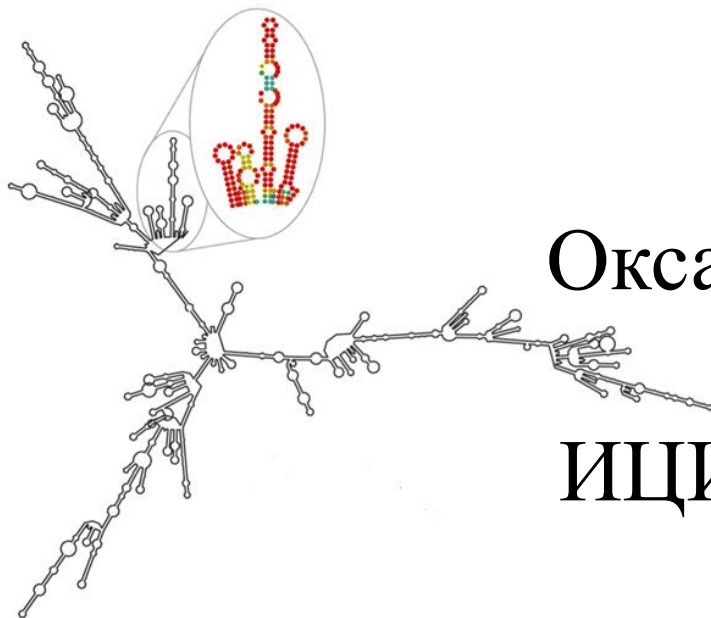
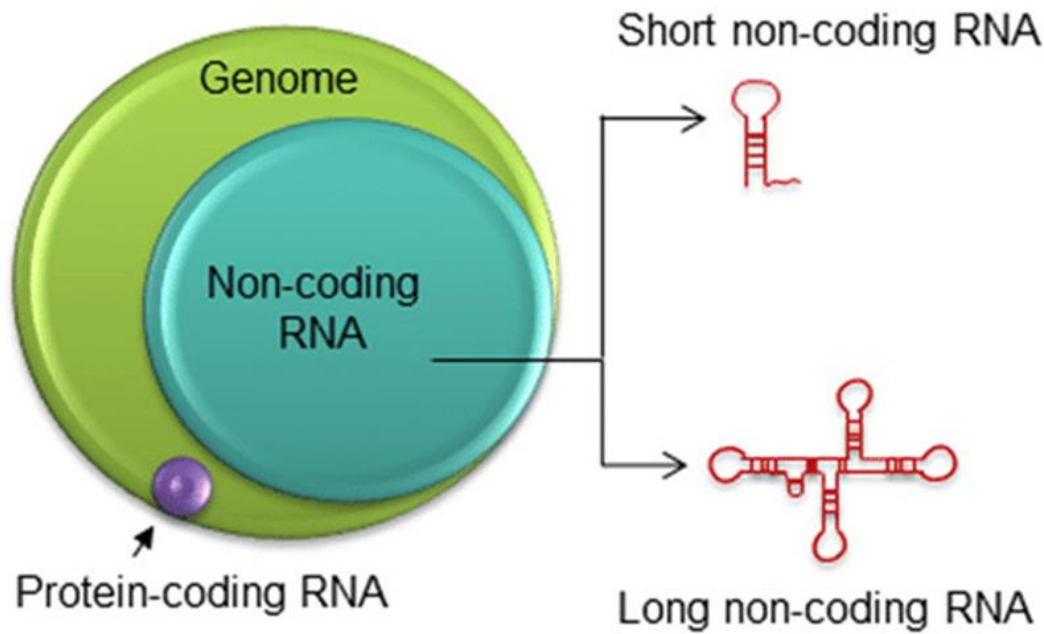


Кодирующие длинные некодирующие РНК



Оксана Волкова

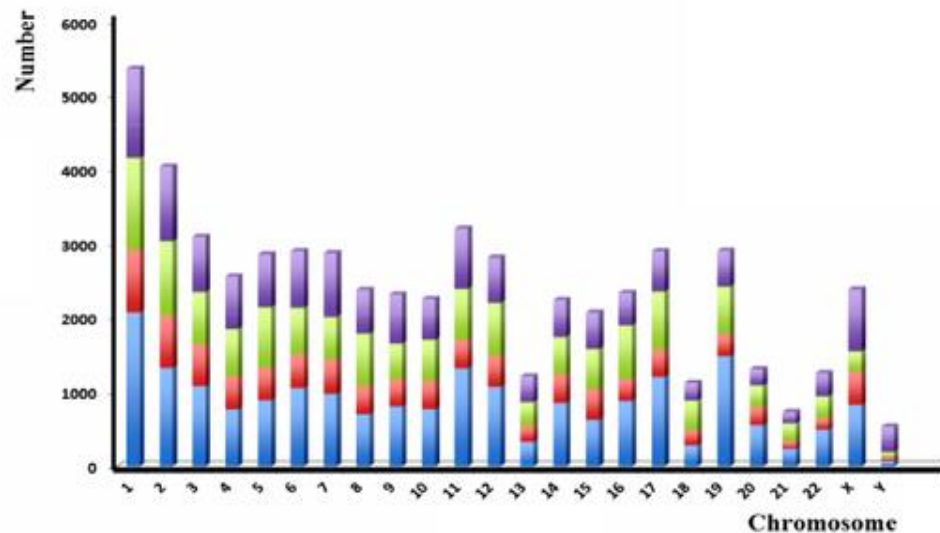
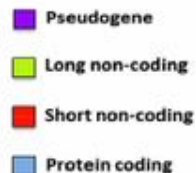
ИЦИГ СО РАН
2019



50–80% ДНК
транскрибируется
2% транскриптов
транслируется

(По данным FANTOM и ENCODE в геноме человека
число **генов некодирующих РНК** превышает число **белок-кодирующих генов** (~23000 и ~20000)

Type of RNA



Total RNAs = 57,952

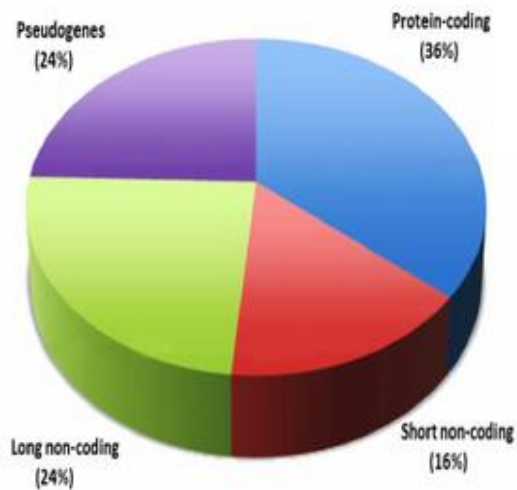
■ Protein-coding (20,805)

■ Long non-coding (13,870)

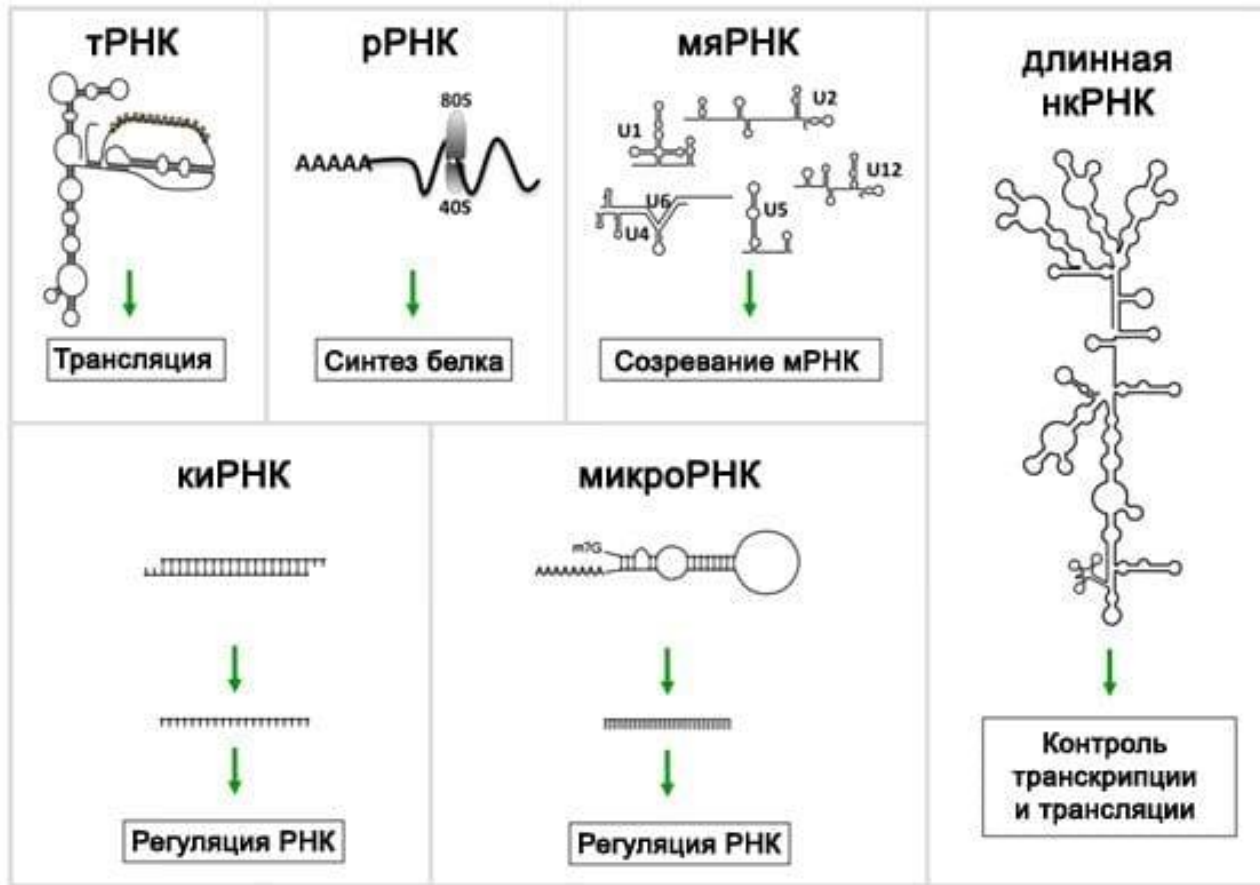
■ Short non-coding (9,096)

■ Pseudogenes (14,181)

Distribution of RNA types within the Human Genome



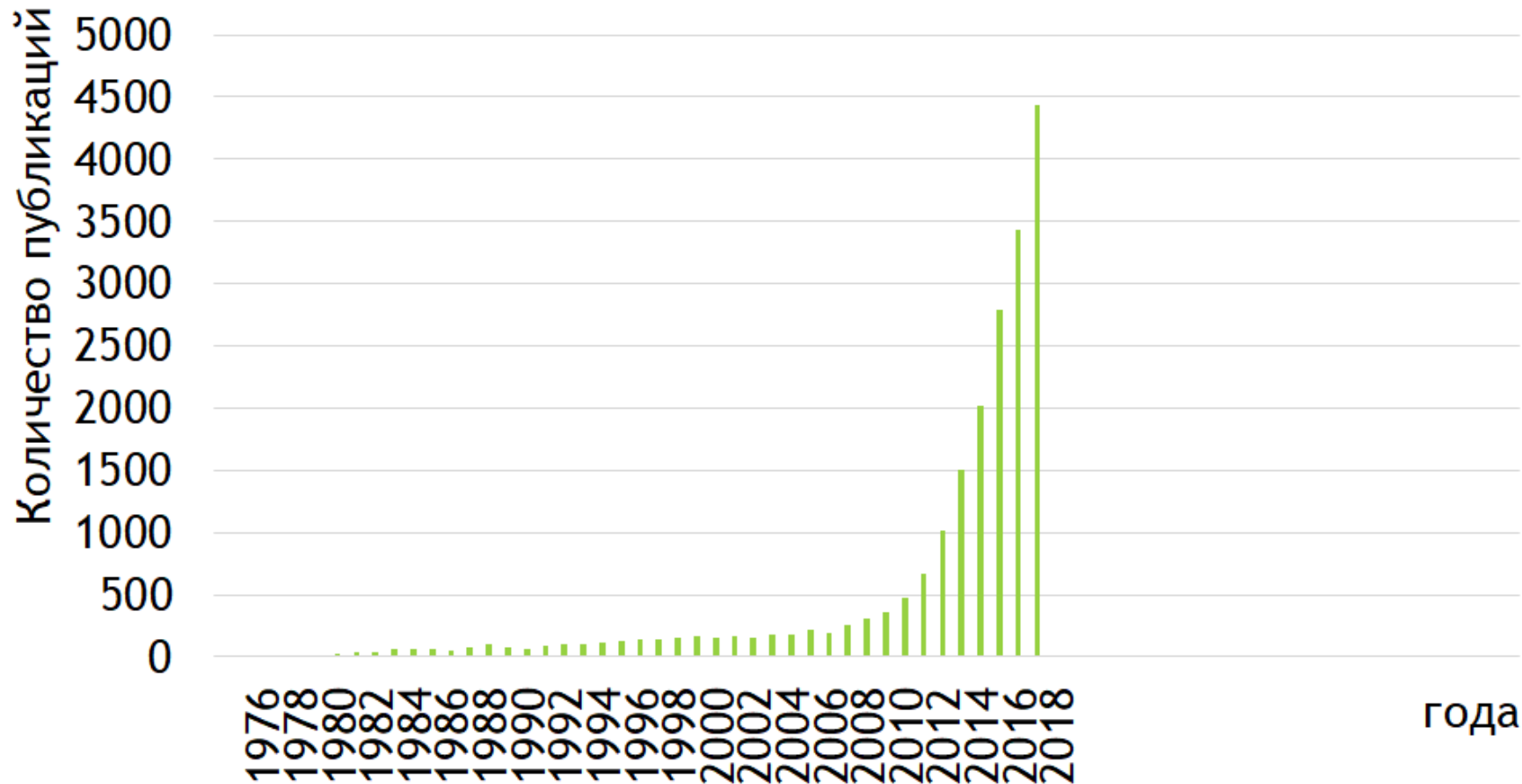
Схематичное строение и основные функции некоторых нкРНК



Основные виды некодирующих РНК

Вид нкРНК	Англоязычное название	Число генов в геноме человека	Примеры	Длина (нукл.)	Год открытия
Все гены: 58 288					
Транспортные РНК (тРНК)	Transfer RNA, tRNA	387 = 0,7%	tRNAPhe, tRNAala...	76–90	1965
Рибосомные РНК (рРНК)	Ribosomal RNA, rRNA	544 = 0,9%	28S, 5.8S, 18S, 5S	200–5000	1980
Малые ядерные РНК (мяРНК)	Small nuclear RNA, snRNA	1 900 = 3,3%	U1, U6, U12...	100–150	1966
Малые ядрышковые РНК (мякРНК)	Small nucleolar RNA, snoRNA	943 = 1,6%	SNORD13, SNORD22, SNORD118...	70–120	1996
Длинные некодирующие РНК (днРНК)	Long non-coding RNA, lncRNA	15 778 = 27,1%	XIST, ANRIL, HOTAIR, H19, TUG1	200–15 000	≈1990
Малые некодирующие РНК: 9 534 = 15,8%					
МикроРНК	MicroRNA, miRNA, miR	1881 = 3,2%	miR-21, miR-143/145	19–25	1993
Короткие интерферирующие РНК (киРНК)	Small interfering RNA, siRNA	Не имеют собственных генов!	киРНК вирусного происхождения	21–25	1998

Количество публикаций Medline trend for **long+non-coding+rna**



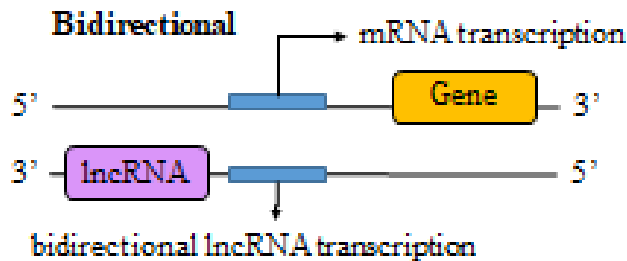
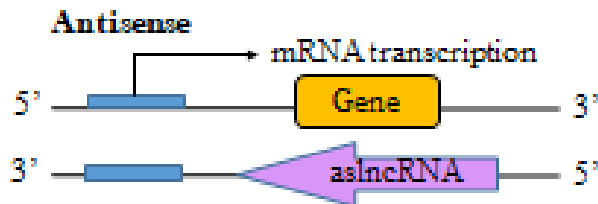
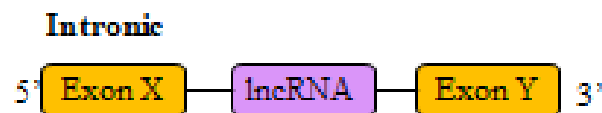
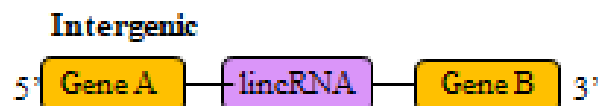
Long non-coding RNA databases

deepBase	Identification, expression, evolution and function of long non-coding RNAs (lncRNAs), small RNAs and circular RNAs from deep-sequencing data
LNCipedia	a comprehensive compendium of human long non-coding RNAs
lncRNAdb	The Reference Database For Functional Long Noncoding RNAs
lncRNAWiki	a wiki-based, publicly editable and open-content platform for community non-coding RNAs (lncRNAs)
lncBook	a comprehensive collection of 270,044 human lncRNAs and systematic annotation by multi-omics data integration, function annotation and disease
MONOCLdb	The MOnse NOnCode Lung database provides the annotations and expression of long non-coding RNAs (lncRNAs) involved in influenza and SARS-CoV
NONCODE	An integrated knowledge database dedicated to ncRNAs, especially lncRNAs
lncRNome	a comprehensive searchable biologically oriented knowledgebase for long non-coding RNAs in Humans.
NRED	a database of long noncoding RNA expression.
C-It-Loci	a tool to explore and to compare the expression profiles of conserved lncRNAs across three organisms
MiTranscriptome	a catalog of human long poly-adenylated RNA transcripts derived from high-throughput RNA-Seq data from over 6,500 samples, spanning diverse
lncRNA Evolution Browser	this site contains alignments and evolutionary metrics of conserved lncRNAs

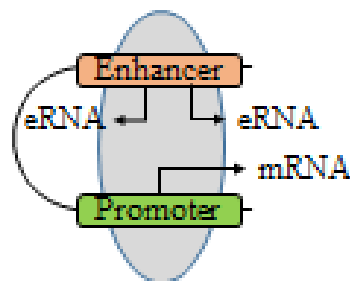
Noncode DB statistics

Species	Number of lncRNA transcripts	Number of lncRNA genes
Human	172,216	96,308
Mouse	131,697	87,774
Cow	23,515	22,227
Rat	24,879	22,127
Chimp	18,004	12,790
Gorilla	18,539	15,095
Orangutan	15,178	13,106
Rhesus	9,128	6,010
Opossum	27,167	17,795
Platypus	11,210	9,163
Chicken	12,850	9,527
Zebrafish	4,852	3,503
Fruitfly	42,848	15,543
Celegans	3,154	2,552
Yeast	55	52
Arabidopsis	3,763	3,472
Pig	29,585	17,811
Total	548,640	354,855

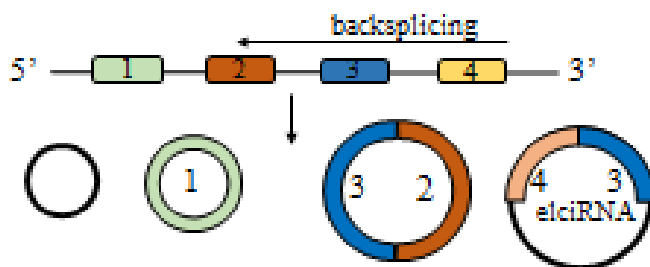
Классификация длинных некодирующих РНК

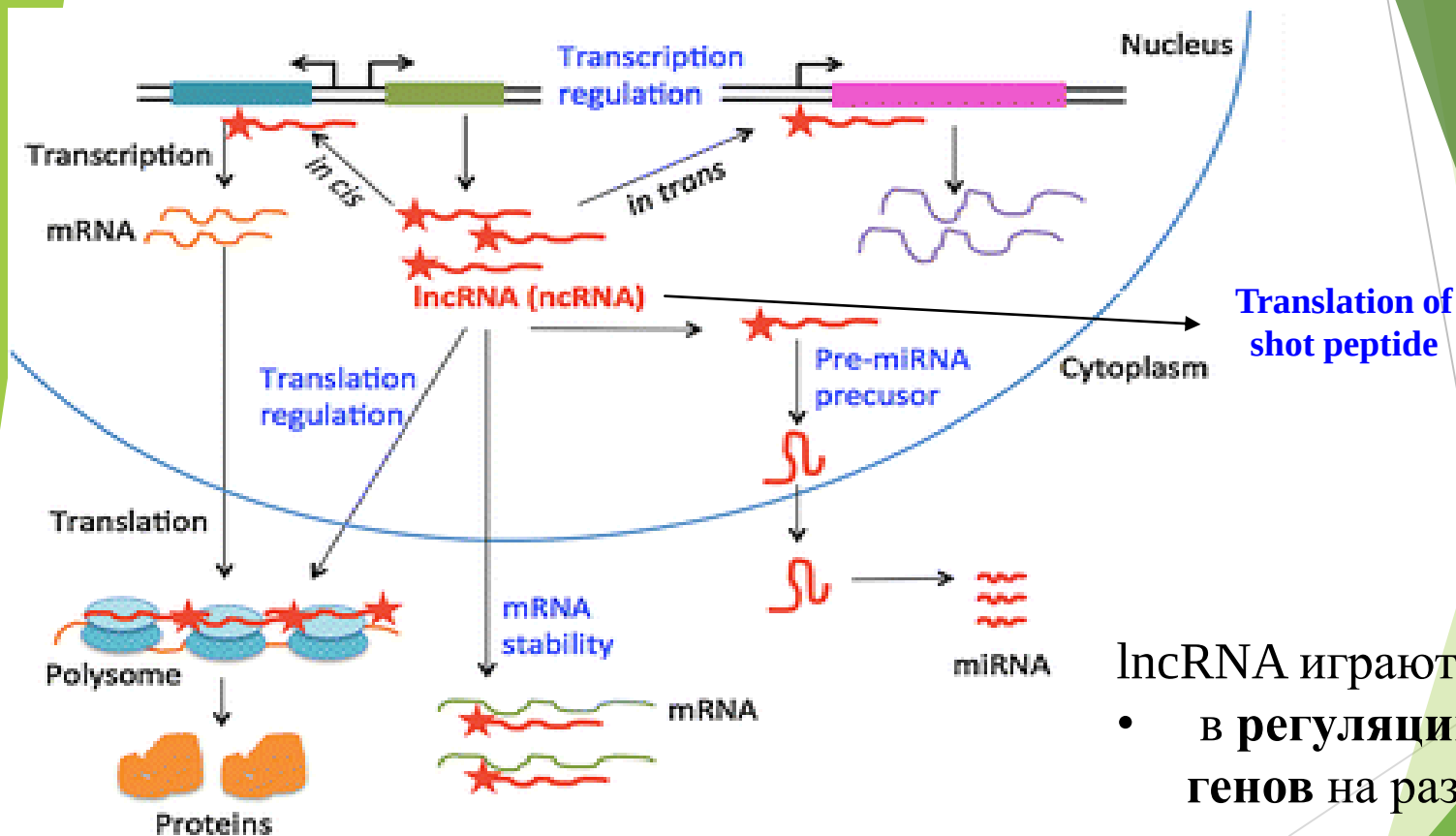


Enhancer RNA



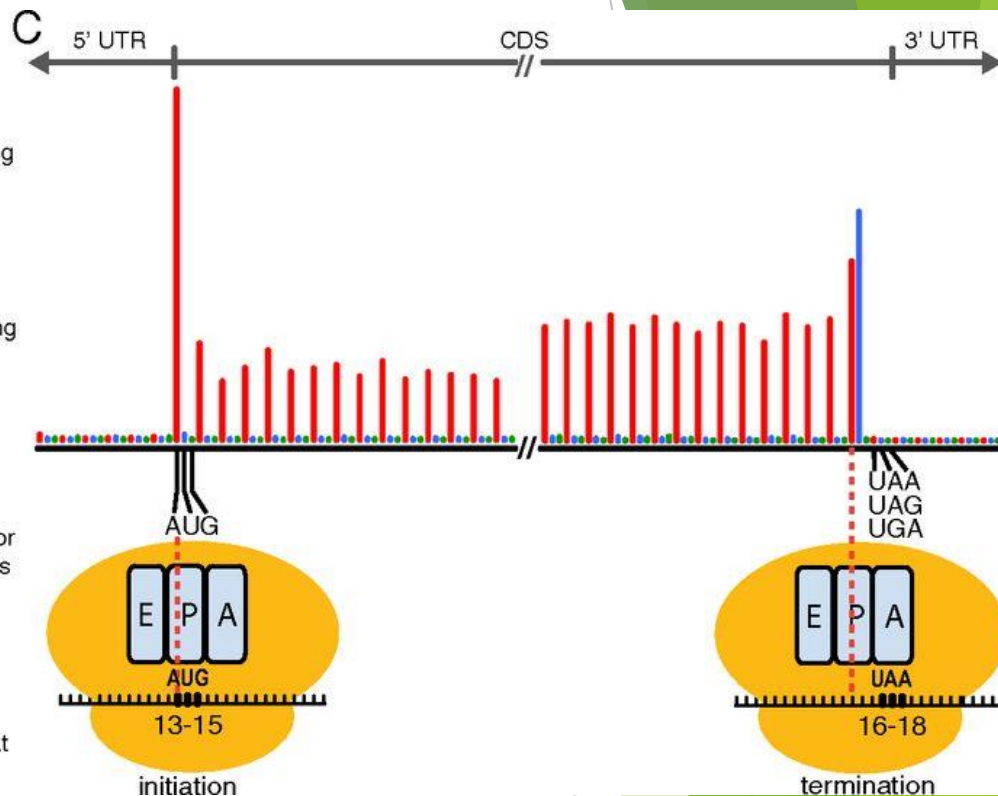
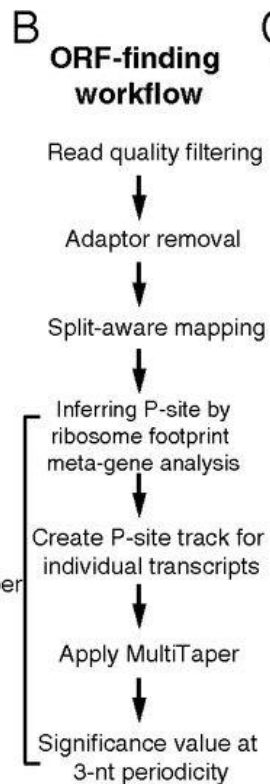
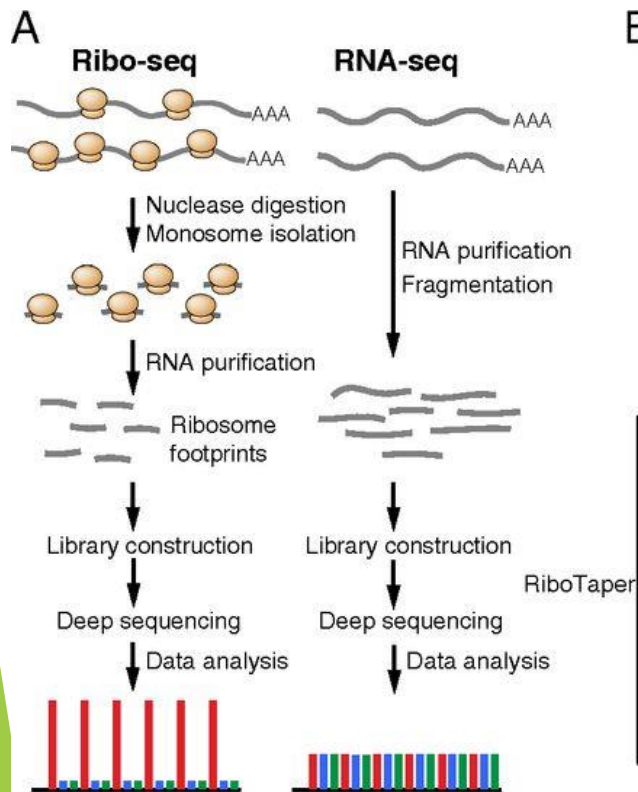
General formation of circRNAs





lncRNA играют важную роль

- в регуляции экспрессии генов на разных уровнях,
- а так же могут быть транслированы



Данные профилирования рибосом (Ribo-seq) указывают на то, что некоторые из цитоплазматических **днРНК могут транслироваться**

Могут работать как **бифункциональные РНК**, наделенные и кодирующими и регуляторными функциями

Транслируемые днРНК предпочтительно локализуются в **цитоплазме**, тогда как **нетранслируемые днРНК** преимущественно локализуются в **ядре**

Эффективность трансляции цитоплазматических днРНК почти сравнима с эффективностью мРНК

Функциональную значимость этих пептидов показана биоинформатическими и экспериментальными методами (Ruiz-Orera et al., 2014; Pan et al., 2018; Choi et al., 2018; Matsumoto, Nakayama, 2018)

Длинные некодирующие РНК могут

- функционировать как **РНК**,
- **транслироваться**,
- быть **бифункциональными**

днРНК	Кодирующие	Бифункциональные	Некодирующие
Функция	Пептид	Пептид + РНК	РНК
Примеры	MLN DWORF LINC00961	HOXB-AS3	XIST HOTAIR HOTTIP p21-lincRNA Fendrr

На данный момент известны **16 малых пептидов** размером до 90 аминокислот
Seo-Won Choi, Hyun-Woo Kim, Jin-Wu Nam (2018) The small peptide world in long noncoding RNAs. Briefings in Bioinformatics, bby055, <https://doi.org/10.1093/bib/bby055>

- Показано, что **днРНК** содержат либо внутренние сайты посадки рибосом (**IRES**),
- либо короткие открытые рамки считывания (**sORF/ORC**),
- и другие элементы трансляции, которые могут быть транслированы в пептиды с физиологическими функциями, которые могут быть также эволюционно консервативны.



Функциональное назначение любой lncРНК определяется ее

- нуклеотидной последовательностью
- и вторичной структурой (сайты связывания для белков)
- третичные пространственные структуры (база для сборки и осуществления ферментативной активности многокомпонентных белковых комплексов)

- lncRNA **обогащены консервативными элементами последовательности**,
- Часто когда последовательность человеческой lncRNA консервативна у другого вида позвоночных, нет транскрипции lncRNA в ортологичной геномной области.

Межгенные и антисмысловые lncRNA более консервативны, чем интронные, как и сплайсированные по сравнению с транскриптами одного экзона

Длинные некодирующие РНК участвуют в межмолекулярных взаимодействиях

A lncRNA-DNA



RNA pol II

Negative expression

Pre-initiation complex



Positive expression



Epigenetic modulation

B lncRNA-RNA



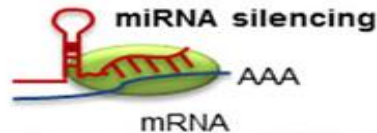
Spliced variant

Exon 1 Exon 4

Alternative splicing



Degradation



Translation inhibition

C lncRNA-protein



Modulation of protein activity

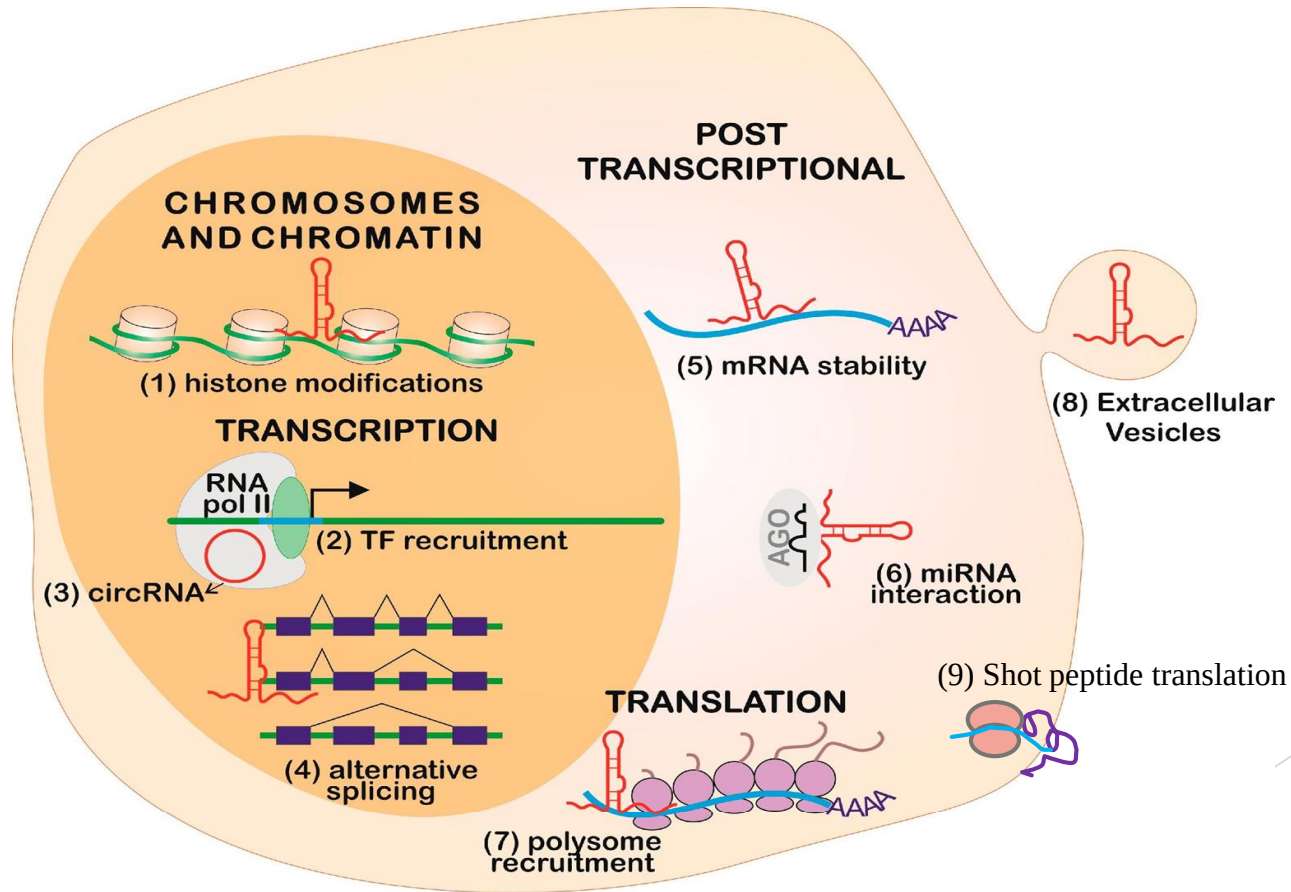


Protein localization

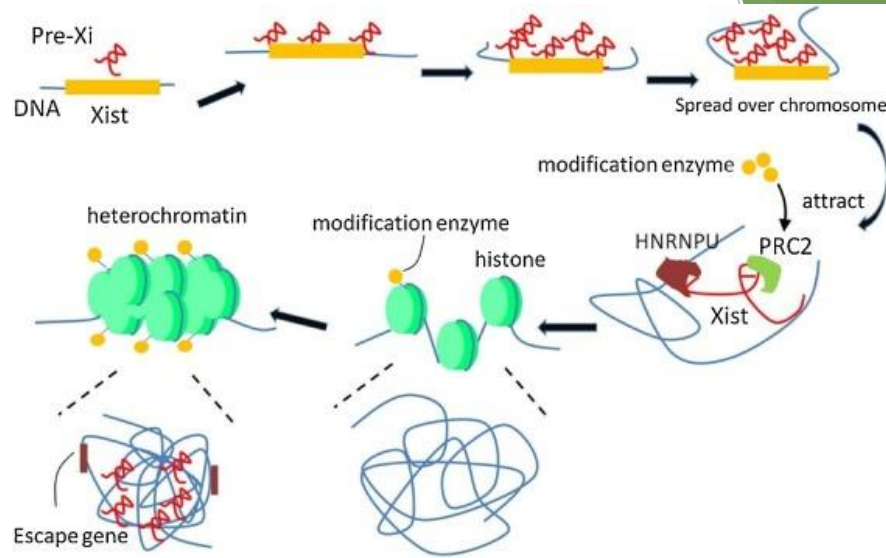


Structural and/or organization role

Длинные некодирующие РНК регулируют экспрессию генов



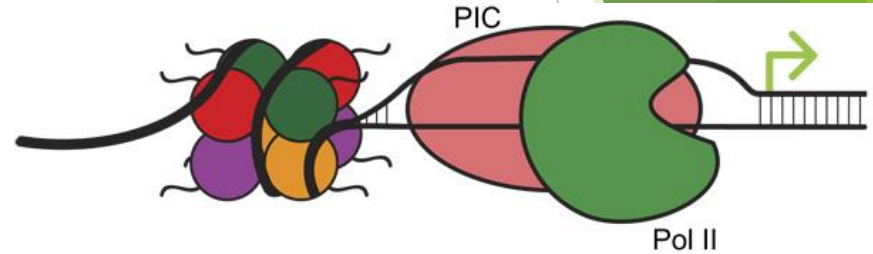
Процесс инактивации X-хромосомы начинается с распространением **Xist-lncRNA** от Xic по всей хромосоме, где Xist-РНК индуцирует образование и расширение области гетерохроматина.



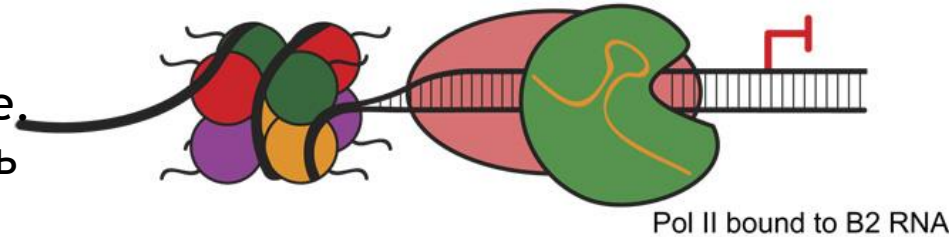
Гетерохроматин инактивированной X-хромосомы характеризуется не только связью с Xist-РНК, но также наличием особого варианта гистона H2A (macroH2A), гипoaцетилированием гистонов H3 и H4, убиквитинированием гистона H2A и метилированием специфических позиций на гистоне H3, а также метилирования ДНК. Этих модификаций к самораспространению обеспечивает поддержание инактивированного состояния X-хромосомы в течение многих клеточных делений

днРНК В2 напрямую ингибирует транскрипцию

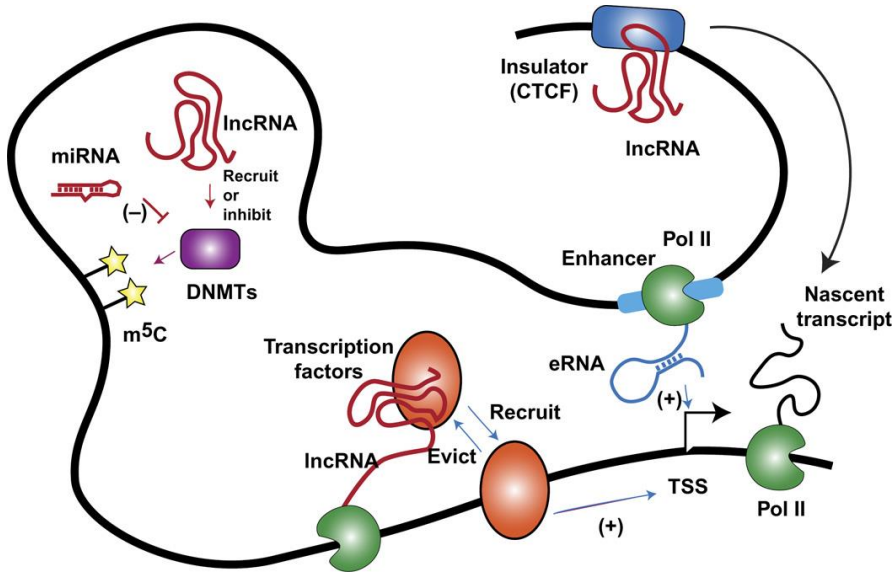
В отсутствие В2 может быть собран функциональный закрытый преинициаторный комплекс (PIC). Этот комплекс может расплавить дуплекс дцДНК, образуя открытый преинициаторный комплекс, который способствует инициации транскрипции.



Напротив, днРНК В2 непосредственно связывается с РНК Pol II в нефункциональном закрытом комплексе. РНК В2 не позволяет Pol II устанавливать функциональные контакты с ДНК и тем самым предотвращает расплавление дцДНК и образование открытых комплексов.



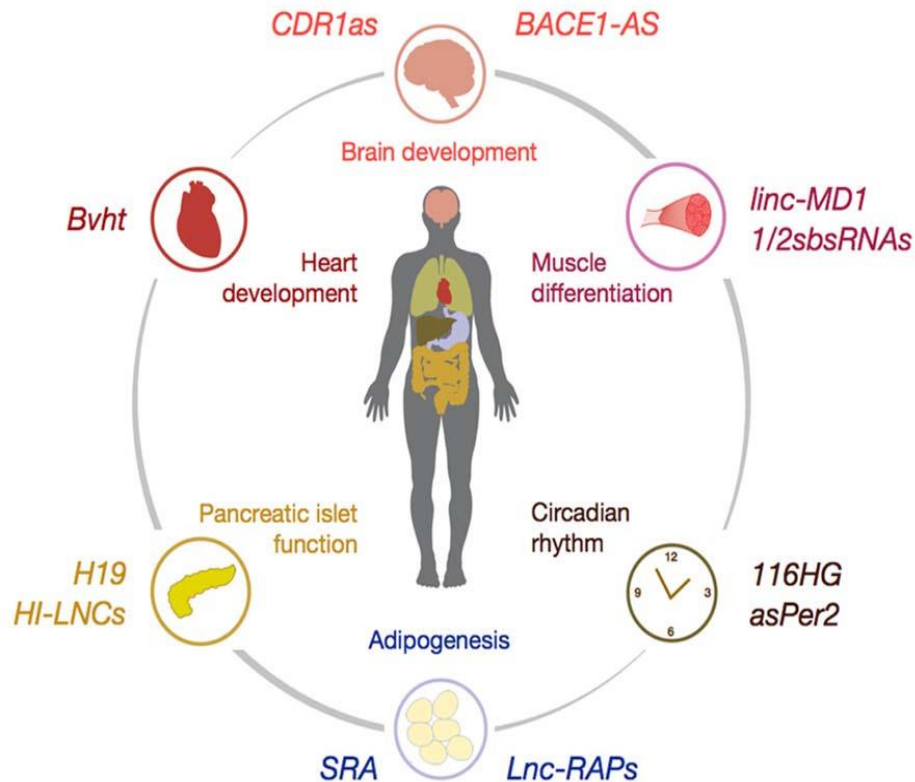
днкРНК регулируют транскрипцию через ДНК-связывающие белки



- lncRNAs взаимодействуют с ДНК-метилтрансферазами (DNMTs), что приводит к ингибированию их в локусах хроматина. Изменение уровня метилирования ДНК лиет на локальную транскрипцию.
- Взаимодействия lncRNA с транскрипционным фактором могут способствовать или препятствовать посадке транскрипционных факторов

lncRNA, транскрибируемая из энхансерной области, может способствовать петлеобразованию хроматина и активации генов.

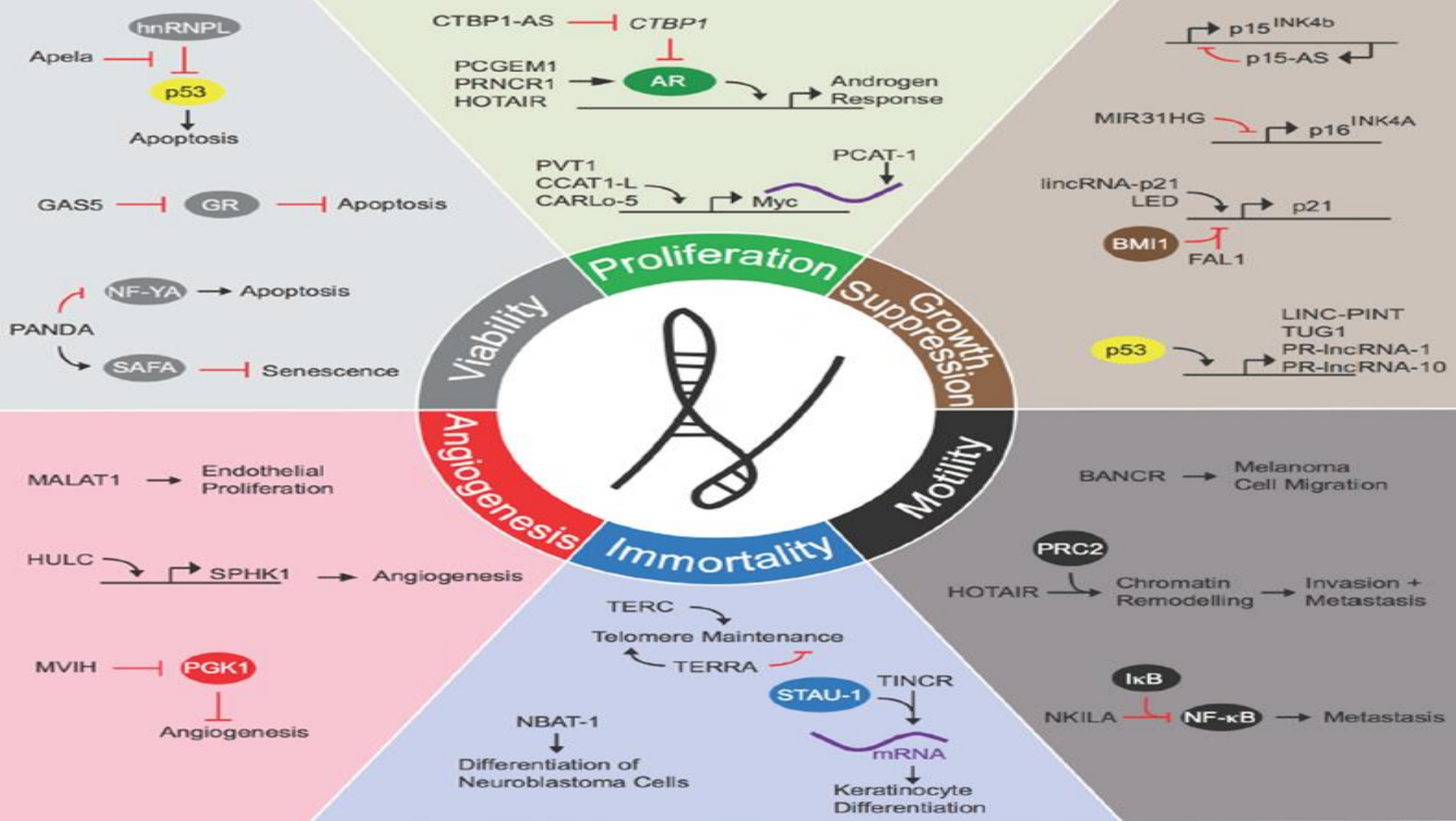
lncRNAs взаимодействуют с CTCF-связывающим фактором (CTCF) инсулятора хроматина и регулируют транскрипцию. Механизм может включать действие CTCF в петле хроматина и ядерной архитектуре.



Базы данных взаимосвязи отдельных днРНК с развитием различных заболеваний

- lncRNAdb,
- LncRANDisease
- Lnc2Cancer
- PLNlncRBase
- EVLncRNAs

ДнРНК являются и биомаркерами и мишенями для терапии

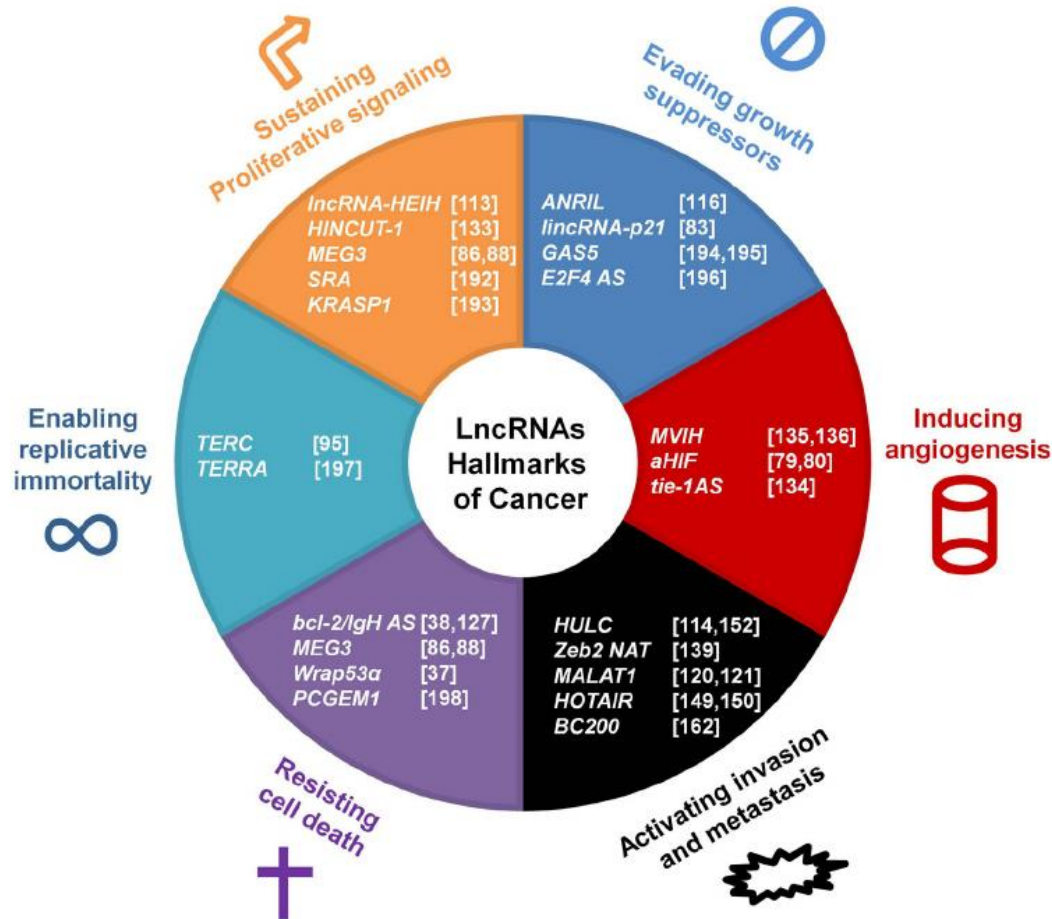


lncРНК	Определяемый процесс/ свойство	Опухоль
hoTair	Злокачественный рост	Плоскостеточная карцинома гортани, гепатоцеллюлярная карцинома, поджелудочная железа, колоректальная карцинома
pcaT-1 (prostate cancer associated Transcript 1)	Злокачественный рост	Колоректальная карцинома
uca1 (urothelial cancer associated 1)	Злокачественный рост	Карцинома мочевого пузыря, плоскоклеточная карцинома пищевода
malaT1	Злокачественный рост	Карцинома поджелудочной железы, колоректальная карцинома
anril	Злокачественный рост	Карцинома желудка
linc00152	Злокачественный рост	Карцинома желудка Прогностические маркеры
hoTair	Метастазирование Рецидивирование	Карцинома толстой кишки, карцинома молочной железы Гепатоцеллюлярная карцинома
anril (antisense non-coding rna in the ink4 locus)	Метастазирование	Немелкоклеточная карцинома легкого

Некоторые днкРНК обладают двойственной ролью — то есть могут выступать и как **онкосупрессоры** и как **онкогены**, в зависимости от конкретного типа опухоли, то есть от клеточного и тканевого окружения

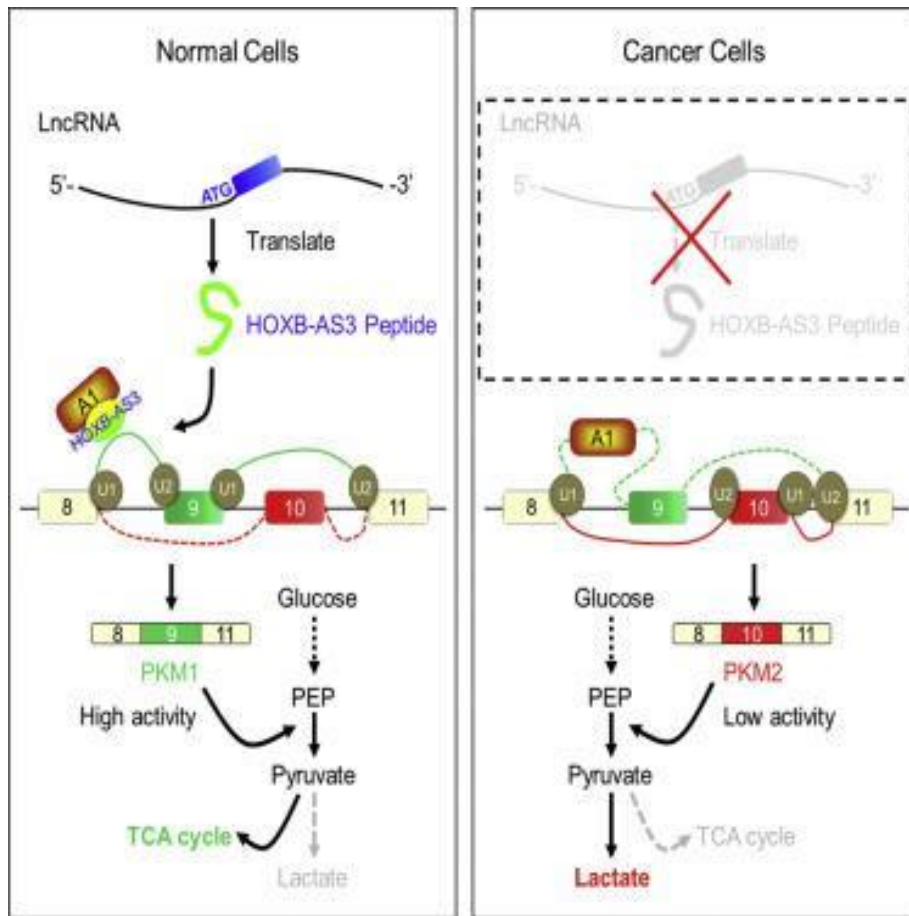
онкоген	Двойственная роль	онкосупрессор
ANRIL, HOTAIR,PCAT1	H19, BANCR, ncRAN, MALAT1, XIST	MEG3,PCAT29, lncRNA-LET

lncRNA могут использоваться как маркеры заболеваний



- lncRNA могут использоваться как **маркеры** и
- для **терапии** заболеваний, например, как «эндогенные губки»

Тип рака	Маркеры	Биологическая жидкость
Плоскоклеточная карцинома полости рта	HOTAIR	Слюна
Мелкоклеточный рак легкого	XIST MALAT1	Сыворотка Цельная кровь
Рак желудка	LINC00152, H19 LINC00152, PVT1	Плазма Желудочный сок
Рак толстой кишки	HOTAIR	Цельная кровь
Рак мочевого пузыря	SncmtRNA, AsncmtRNA	Моча
Рак предстательной железы	MALAT1 PSA3, MALAT1, XLOC007697	Плазма Моча



Пептид , транслируемый с lncRNA HOXB-AS3, регулирует сплайсинг фермента пируваткиназы М (PKM), фермент с большей активностью направляет по пути окислительного фосфорилирования, низкой-гликолиза

Раковые клетки проявляют отчетливый метаболический фенотип, характеризующийся повышенным гликолизом и образованием лактата вместо митохондриального окислительного фосфорилирования независимо от наличия кислорода

A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Molecular Cell* 2017. 68, 171–184. doi.org/10.1016/j.molcel.2017.09.015

► Предсказание методами *in silico* потенциально транслируемых из большого множества длинных некодирующих РНК (днРНК) и выявление соответствующих им малых пептидов методами экспериментальной биологии является актуальной комплексной задачей последних 5-10 лет.

► На данный момент известны **16 малых пептидов** размером до 90 аминокислот, синтезируемых по матрице **13 днРНК** человека, мыши, данио-рерио, дрозофилы и сои

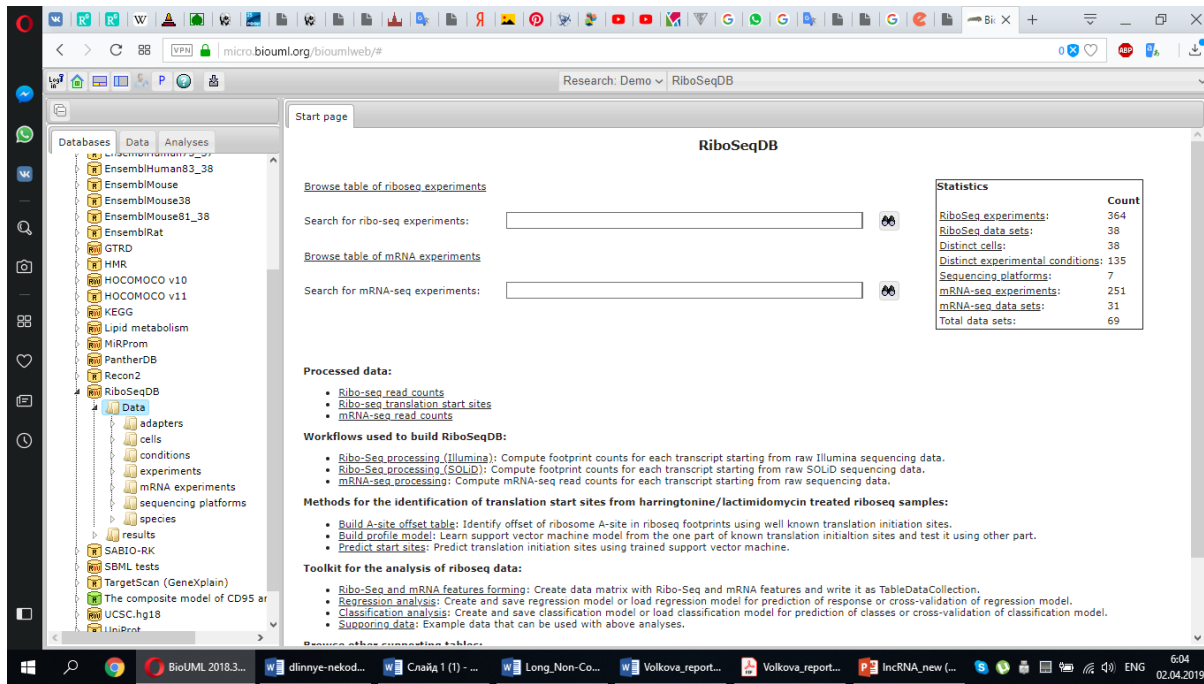
(Seo-Won Choi, Hyun-Woo Kim, Jin-Wu Nam (2018) The small peptide world in long noncoding RNAs. Briefings in Bioinformatics, bby055, <https://doi.org/10.1093/bib/bby055>).

► Из восьми человеческих и мышинных пептидов больше половины задействовано в регуляции работы мышечных клеток, а три ассоциированы с раковыми процессами.

Основные результаты, полученные в ходе реализации проекта **РФФИ**
2014-2016, 14-04-01284

Построение модели и предсказание эффективности трансляции мРНК
млекопитающих по ее первичной последовательности на основе
данных по рибосомному профилированию

1. Создана база данных RiboSeqDB. Она содержит 38 наборов данных (364 образца) Ribo-seq и 31 набор mRNA-seq (251 образец) для 38 линий и типов клеток человека, мыши и крысы



The screenshot displays the RiboSeqDB web application interface. The top navigation bar includes tabs for 'Databases', 'Data', and 'Analyses'. The left sidebar lists various databases and data sources, with 'RiboSeqDB' selected under the 'Data' tab. The main content area, titled 'RiboSeqDB', features a 'Start page' section with links to 'Browse table of riboseq experiments' and 'Browse table of mRNA experiments'. Below these are search fields for ribo-seq and mRNA-seq experiments. A 'Statistics' table is displayed on the right, showing the following data:

Statistics	Count
RiboSeq experiments:	364
RiboSeq data sets:	38
Distinct cells:	38
Distinct experimental conditions:	135
Sequencing platforms:	7
mRNA-seq experiments:	251
mRNA-seq data sets:	31
Total data sets:	69

The interface also includes sections for 'Processed data' (listing Ribo-seq read counts, Ribo-seq translation start sites, and mRNA-seq read counts), 'Workflows used to build RiboSeqDB' (describing Ribo-Seq processing, Ribo-Seq processing (SOLID), and mRNA-seq processing), 'Methods for the identification of translation start sites from harringtonine/lactimidomycin treated riboseq samples' (listing Build A-site offset table, Build profile model, and Predict start sites), and 'Toolkit for the analysis of riboseq data' (listing Ribo-Seq and mRNA features forming, Regression analysis, Classification analysis, and Supporting data).

Sharipov R.N., Yevshin I.S., Kondrakhin Y.V., Volkova O.A. RiboSeqDB - a repository of selected human and mouse ribosome footprint and RNA-seq data. Virtual Biology. 2014. N2. P.37-46. DOI: 10.12704/vb/e18http://micro.biouml.org/bioumlweb/riboseqdb.html

3. Разработаны методы оценки эффективности трансляции (translation efficiency, TE) и проверена их точность. Разработаны методы определения доверительного интервала и относительной погрешности TE. Подобраны оптимальные параметры обработки “сырых” данных Ribo-seq и mRNA-seq.

4. На основе метода, описанного в (Ingolia et al., 2011), разработан модуль для платформы BioUML, позволяющий на основе данных Ribo-seq находить в РНК сайты инициации трансляции и давать количественную оценку TE.

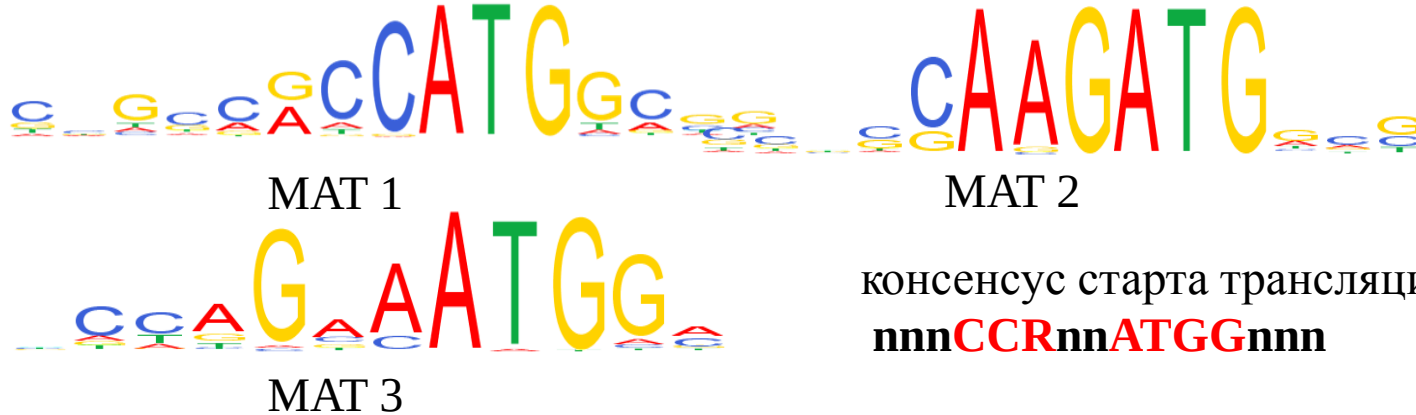
5. Проведен статистический анализ отдельных характеристик последовательностей мРНК человека и мыши; выявлена их взаимосвязь с соответствующими данными Ribo-seq.

Отобраны **трансляционно-значимые характеристики мРНК, влияние на эффективность трансляции** которых наиболее статистически значима и одинакова для всех трех проанализированных выборок. На основании регрессионного анализа построена предварительная модель инициации трансляции.

6. Выявлены параметры мРНК значимые для эффективности инициации трансляции (translation initiation efficiency, TIE) и TE в целом, на основе дискриминантного и регрессионного анализа белок-кодирующих и длинных некодирующих РНК (lncRNA), а также высоко- и низко- транслируемых мРНК.

Эти параметры позволяют надежно, с высокой точностью **разделять белок-кодирующие и некодирующие РНК** независимо от используемого метода анализа.

7. Метод весовых матриц протестирован для **предсказания новых сигналов** регуляции эффективности трансляции, неаннотированных в базах данных.



8. Разработан тулkit (набор инструментов) с инструкцией для анализа данных Ribo-seq и оценки TE и его веб-интерфейс в платформе BioUML.

9. Разработана математическая модель инициации трансляции и трансляции в целом с учетом выявленных трансляционно-значимых характеристик 5'-НТП

Сравнительный анализ белок кодирующих мРНК и длинных некодирующих РНК

позволяют **различать белок-кодирующие и некодирующие РНК с высокой точностью (96%)**

Dataset	True classification rates	
	lncRNA sample = lncRNA(6)	lncRNA sample = lncRNA(12)
mRNA sample	0.954	0.949
lncRNA sample	0.964	0.972
Union of mRNA and lincRNA samples	0.960	0.965

Предсказание потенциально транслируемых областей в транскриптах lncRNA

88,3% идентифицированных OPC lncRNA имели длину менее 150 нт, максимальная длина для 97,8% идентифицированных OPC достигала 300 нт.

Полученные результаты хорошо согласуются с нашим предположением, что lncRNA кодируют короткие пептиды

8. Разработан тулkit (набор инструментов) с инструкцией для анализа данных Ribo-seq и оценки TE и его веб-интерфейс в платформе BioUML.

9. Разработана математическая модель инициации трансляции и трансляции в целом с учетом выявленных трансляционно-значимых характеристик 5'-НТП.

- ▶ 1. **Volkova OA**, Kondrakhin YuV., Kashapov TA, Sharipov RN Comparative analysis of protein-coding and long non-coding transcripts based on RNA sequence features. *JBCB*. 2018. 16:02, 1840013. (IF 0.97)
- ▶ 2. Kondrakhin YuV., Sharipov RN, **Volkova OA** Start codons and mRNA translation efficiency prediction using Ribo-Seq-based weight matrices. *Acta Natutae*. 2017. 1:9, 25. (IF 1.79)
- ▶ 3. Yevshin IS Sharipov, RN, **Volkova OA** Assessment of translation efficiency from Ribo-Seq and mRNA-Seq data. *Acta Natutae*. 2017. 1:9, 28. (IF 1.79)
- ▶ Sharipov RN, Yevshin IS, Kondrakhin YuV, **Volkova OA** Ribo-Seq DB - a repository of selected human and mouse Ribo-Seq and RNA-Seq data. *Acta Natutae*. 2017. 1:9, 27. (IF 1.79)
- ▶ 4. **Volkova OA**, Kondrakhin YuV, Yevshin IS, Valeev TF, Sharipov RN Assessment of translational importance of mammalian mRNA sequence features based on Ribo-seq and mRNA-seq data. *JBCB*. 2016. 14:02. (IF 0.97)
- ▶ 5. **Волкова ОА**, Кондрахин ЮВ, Шарипов РН Оценка трансляционной значимости характеристик нуклеотидной последовательности мРНК млекопитающих на основе данных рибосомного профилирования. *Вавиловский журнал генетики и селекции* 2016. 20(6):779-786. (ИФ РИНЦ 0.514)
- ▶ 6. Sharipov RN, Yevshin IS, Kondrakhin YuV., **Volkova OA** RiboSeqDB - a repository of selected human and mouse ribosome footprint and RNA-seq data. *Virtual Biology*. 2014. 2, 37-46, ISSN: 2306-8140. <http://virtual-biology.org/vb/article/view/18>

Задачи:

1. Предсказать потенциально транслируемые в клетках человека, длинные некодирующие РНК (lncRNA/днРНК) с использованием собственных математических моделей.
2. Проанализировать информацию из баз данных GEO, SRA, ENCODE, и др., характеризующую представленность различных днРНК в клетках отдельных линий.
3. Проанализировать информацию из протеомных баз данных по представленности в отдельных клеточных линиях фрагментов, гомологичных потенциально транслируемым РНК.
4. Отобрав наиболее перспективные для экспериментальной проверки днРНК, создать генно-инженерные конструкции, содержащие потенциально транслируемые последовательности, для внедрения в живые системы с целью получения пептидов в результате трансляции.
5. Верифицировать полученные пептиды с помощью методов масс-спектрометрии перед получением специфичных к ним антител.
6. Выявить пептиды - продукты трансляции днРНК в клетках человека -- с использованием специфичных антител и экстракта суммарного клеточного белка.
7. Сравнить наличие и количества выявленных пептидов в клетках человека с патологиями рака.
8. Исследовать наличие искомых пептидов в отобранных клетках опухолей человека, культивируемых в иммунодефицитных мышах "SciD" в условиях SPF-вивария ФИЦ ИЦиГ СО РАН.