

ОСОБЕННОСТИ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЗРЕЛЫХ микроРНК МОГУТ ВЛИЯТЬ НА СРОДСТВО К БЕЛКАМ AGO2 И AGO3 ЧЕЛОВЕКА

© 2011 г. Н. А. Омелянчук¹, П. М. Пономаренко^{1,2}, М. П. Пономаренко^{1*}

¹Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, 630090

²Новосибирский государственный университет, Новосибирск, 630090

Поступила в редакцию 25.03.2010 г.

Принята к печати 12.07.2010 г.

Зрелые микроРНК (miРНК) длиной от 20 до 24 н. относят к эндогенным малым некодирующим РНК. miРНК определяют взаимодействие комплекса белков (RISC), связывание которого с miРНК программирует способность RISC подавлять трансляцию мРНК, которые содержат сайты, комплементарные miРНК. Это взаимодействие приводит к ингибированию трансляции мРНК, если в этот комплекс входит белок Ago3, в случае Ago2 возможно также разрезание мРНК в центре гетеродуплекса мРНК/miРНК. С помощью созданной нами ранее компьютерной системы ACTIVITY мы проанализировали литературные данные о сродстве последовательностей зрелых miРНК человека к белкам Ago2 и Ago3. Обнаружено, что сродство miРНК к обоим белкам тем выше, чем больше содержание тетра-нуклеотидов YRHB вблизи 3'-конца miРНК ($r = 0.613$, $\alpha < 0.025$), и связывание miРНК с Ago2 по сравнению с Ago3 растет с увеличением содержания тетра-нуклеотидов RHHK вблизи центра miРНК ($r = 0.501$, $\alpha < 0.05$). На основе этих двух закономерностей мы вывели формулы для прогноза величин сродства зрелых miРНК к белкам Ago2 и Ago3 и достоверно на независимых данных предсказали сродство к каждому из них для канонических ($\alpha < 0.00025$) и неканонических miРНК ($\alpha < 0.05$).

Ключевые слова: микроРНК, нуклеотидная последовательность, RISC, белки Ago, сродство, прогноз.

THE NUCLEOTIDE SEQUENCE FEATURES OF THE MATURE microRNA SEEM TO BE RESPONSIBLE FOR THE AFFINITY TO HUMAN AGO2 AND AGO3 PROTEINS, by N. A. Omelyanchuk¹, P. M. Ponomarenko², M. P. Ponomarenko^{1,*} (¹Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia; ²Novosibirsk State University, Novosibirsk, 630090 Russia; *e-mail: pon@bionet.nsc.ru). Mature miRNA of 20–24 nt in length are the endogenous sncRNA. They programs RISC to regulate functioning of mRNA with complimentary sites for these miRNAs. In case of Ago3 protein present in human RISC miRNAs direct inhibition of translation, whereas in case of Ago2 is in RISC, than mRNA cleavage in the middle of miRNA/mRNA heteroduplex is also possible. Using ACTIVITY system, that we developed earlier, we analyzed published data on miRNA affinity to human Ago2 and Ago3 proteins. We found increase in miRNA affinity to both Ago2 and Ago3 with the increase of the YRHB tetranucleotide abundance near 3'-end of these miRNAs ($r = 0.613$, $\alpha < 0.025$). We also found that miRNA tendency to bind Ago2 in favor of Ago3 increases with the RHHK tetranucleotide abundance near miRNA center ($r = 0.501$, $\alpha < 0.05$). Using these two findings we proposed two formulas to predict miRNA affinity to Ago2 and Ago3 proteins based on the YRHB and RHHK abundances within this arbitrary miRNA. Thereby we made reliable predictions of miRNA affinity to these proteins in RISC for both canonical ($\alpha < 0.00025$) and non-canonical ($\alpha < 0.05$) miRNAs in comparison with independent experimental data.

Keywords: microRNA, nucleotide sequence, RISC, argonaute proteins (Ago), affinity, prediction.

Принятые сокращения: miРНК – микроРНК; мнкРНК – малая некодирующая РНК; мРНК – матричная РНК; н. – нуклеотид (в качестве единицы длины нуклеотидной последовательности РНК); RISC – комплекс белков, связывание которого с miРНК программирует его биологическую активность на ингибирование трансляции мРНК-мишеней (сокращение от “RNA-Induced Silencing Complex”); MPSS – метод измерения содержания РНК в клетке (сокращение от “Massively Parallel Signature Sequencing”).

* Эл. почта: pon@bionet.nsc.ru

ВВЕДЕНИЕ

МикроРНК (miРНК) относят к эндогенным малым некодирующим РНК. Сначала РНК-полимераза II транскрибирует ген miРНК с образованием первичного транскрипта, из которого затем вырезается предшественник и, наконец, получается зрелая miРНК длиной 20–24 н. [1, 2]. Зрелая miРНК человека взаимодействует с белками Ago1, Ago2, Ago3 или Ago4 семейства Argonaute [3] в составе комплекса белков (RISC), связывание которого с miРНК определяет его способность ингибировать трансляцию мРНК-мишеней и способность препятствовать трансляции тех мРНК, в которых есть сайт, комплементарный этой miРНК [1, 2]. Если в состав RISC входит Ago3, то ингибируется сам процесс трансляции, если вместо Ago3 включен Ago2, то возможно также разрезание мРНК в центре miРНК/мРНК-гетеродуплекса [4, 5]. В известной пространственной структуре комплекса мРНК/мнкРНК/Ago *Pyrococcus furiosus* [6] белок Ago связывает 3'-конец мнкРНК (малая некодирующая РНК) [6].

Зрелая miРНК в клетке может быть представлена ее основной последовательностью (“каноническая форма”) и модификациями этой канонической формы за счет замен, делеций и/или вставок нуклеотидов (“неканонические формы”) [5]. В клетках присутствует многократно больше копий канонических miРНК, чем неканонических. Неканонические формы miРНК общепринято ассоциировать с ошибками биогенеза и с биохимическими модификациями зрелой канонической miРНК. Итогом секвенирования библиотек зрелых miРНК, специфичных для Ago2 или для Ago3, стало обнаружение как неразличимо равного сродства к обоим этим белкам в случае большинства miРНК, так и достоверного предпочтения одного из этих белков перед другим в случае остальных miРНК [5].

В этой работе мы применили созданную ранее компьютерную систему ACTIVITY [7] для анализа парных данных “последовательность нуклеотидов, количественная величина”. С ее помощью мы ранее выявляли достоверные особенности контекста, например, ТАТА-боксы эукариот [8], филамента RecA/ДНК [9], горячих точек мутаций ДНК и сайтов 3'-разрезания пре-мРНК [10], мишеней мобильных элементов [11], сайтов регуляции транскрипции [12] и полиморфизмов, ассоциированных с болезнями человека [13]. Показана возможность адаптации прогнозов ACTIVITY к условиям различных опытов [14] и совместимость этих прогнозов с другими источниками [15].

В предыдущей работе [16] с помощью ACTIVITY мы нашли корреляцию между микрочип-измерениями [17] содержания зрелых канонических miРНК *Arabidopsis thaliana* и тетра nukлеотидов WRHW, DRYD (номенклатура IUPAC-IUB, [18]) в этих miРНК. Прогноз содержания miРНК в клетках на основе этой корреляции [16] был в достоверном со-

гласии с независимыми результатами метода измерения содержания РНК в клетке (MPSS) [19]. Этим мы впервые показали, что, несмотря на очень малую длину ≈ 20 н., miРНК способны нести биологически значимые контекстные мотивы (DRYD вблизи 3'-конца, WRHW вблизи центра), положение и количество которых в последовательности miРНК достоверно коррелируют с содержанием этой miРНК в клетке. Контекстные характеристики могут определять и другие аспекты функционирования miРНК, например, в работе [20] на клетках HeLa по различию зрелых miРНК miR-29a и miR-29b человека также нашли два контекстных мотива (U в центральной позиции 10 и 3'-конец AGUGUU), определяющих преимущественную локализацию miR-29b в ядре клетки.

Общепринято, что все стадии процессинга miРНК (транскрипция \rightarrow биогенез \rightarrow диссоциация miРНК/RISC \rightarrow деградация miРНК) влияют на содержание miРНК в клетке. Однако в последнее время накапливается все больше и больше данных, что во многих случаях лимитирующим звеном в этой цепи может быть диссоциация miРНК/RISC [20–24]. Так, при овуляции в клетках яичника мыши за кратковременным импульсным приростом уровня лютеинизирующего гормона наблюдали такие же импульсные появления первичных транскриптов с генов miРНК *Mir132* и *Mir212* [21]. В свою очередь, зрелые miРНК с этих транскриптов регистрировали на высоком уровне долгое время спустя после исчезновения их предшественников. Время полужизни некоторых зрелых miРНК может достигать 14 дней [22], тогда как биогенез от первичного транскрипта до зрелой miРНК может протекать настолько быстро, что промежуточные продукты не всегда можно зарегистрировать [2]. Таким образом, в клетке зрелые miРНК большую часть их жизни связаны с RISC. Комплекс мнкРНК с RISC обладает высокой стабильностью [23]. Освобождение miРНК от RISC приводит к немедленной деградации этой miРНК и регулируется наличием ее мРНК-мишени [24]. Если же лимитирующая стадия накопления miРНК в клетке – диссоциация miРНК/RISC [20–24], то из найденной нами ранее достоверной корреляции между контекстом и содержанием miРНК в клетке [16] с необходимостью следует корреляция между контекстом miРНК и сродством miРНК/RISC. В предлагаемой статье мы проверяем именно это следствие нашей предыдущей работы.

С помощью ACTIVITY мы проанализировали величины сродства miРНК к белкам Ago2 и Ago3 из RISC человека [5]. В результате мы обнаружили увеличение сродства miРНК к обоим Ago2 и Ago3 при увеличении содержания YRHW [18] вблизи 3'-конца зрелой miРНК ($r = 0.613$, $\alpha < 0.025$). Нами также показано, что предпочтение miРНК к Ago2 перед Ago3 растёт с увеличением количества RHHK вблизи центра miРНК ($r = 0.501$, $\alpha < 0.05$). Используя эти закономерности, мы вывели формулы прогноза срод-

ства последовательностей зрелых miРНК к белкам Ago2 и Ago3 по содержанию URNB и RHNK в определенных районах этих последовательностей. С помощью этих формул мы впервые достоверно предсказали сродство канонических ($\alpha < 0.00025$) и неканонических ($\alpha < 0.05$) miРНК к белкам Ago2 и Ago3 на независимых данных из работы [5].

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Исследуемые литературные данные о сродстве miРНК к белкам Ago2 и Ago3 человека

Мы анализировали для 28 зрелых канонических (табл. 1) и 48 неканонических (табл. 2) miРНК, опубликованных в работе [5, дополнение], последовательности $S = \{s_1 \dots s_{22}\}$ длины 22 н., где s_{22} – 3'-конец miРНК. Поскольку последовательности зрелых канонических miРНК имели длину не менее 22 н., то мы игнорировали нуклеотиды 5'-конца канонических miРНК длиннее 22 н. и неканонические miРНК короче 22 н. Величины A_2 и A_3 сродства miРНК к белкам Ago2 и Ago3 человека соответственно, изменявшиеся в пределах от 2.30 до 9.43 натуральных логарифмических (ln) единиц для всех $28 + 48 = 76$ miРНК, брали из дополнения к работе [5]. Всего мы проанализировали 76 наборов $\{S_n, A_{2,n}, A_{3,n}\}_{1 \leq n \leq 76}$, для каждой n -й зрелой miРНК: ее последовательность длиной 22 н. от 3'-конца и две величины сродства, к Ago2 и к Ago3.

Поиск корреляций между контекстными особенностями последовательностей зрелых miРНК и величинами их сродства к Ago2 и Ago3 белкам

Мы искали корреляции порядка $S = \{s_1 \dots s_i \dots s_{22}\}$ нуклеотидов $s_i \in \{A, U, G, C\}$ в miРНК с величинами A_2 и A_3 ее сродства к Ago2 и Ago3, соответственно, с помощью ACTIVITy [7]. Ранее мы описали важные различия в применении системы ACTIVITy в случае анализа экспериментальных данных о связывании ДНК/белок [8], филаментации [9], мутагенезе [10], мишени мобильного элемента [11], сайтах регуляции транскрипции [12], точечном полиморфизме [13], *in vitro* селекции аптамеров ДНК [14], многошаговых процессах [15] и содержании miРНК в клетке [16]. Новшеством этой работы стали три важные особенности применения ACTIVITy [7–16] в случае дивергенции гомологичных белков от их общего предка (Ago2 ← предок → Ago3 [3]) с использованием величин их сродства к зрелым miРНК [5].

Во-первых, для анализа обоих Ago2 и Ago3 одновременно мы заменили независимые измерения A_2 и A_3 [5] двумя вспомогательными оценками ($\Sigma = (A_2 + A_3)/2$ и $\Delta = (A_2 - A_3)/2$) на их основе. Мы эвристически ассоциировали оценку Σ с ко-адаптацией контекстов древних miРНК и общего предка белков Ago2 и Ago3 на ранних этапах их ко-эволюции,

оценку Δ – с ко-адаптацией контекстов потомков этих древних miРНК при эволюционной дивергенции Ago2 и Ago3. Благодаря филогенетической инерции [25], сохранение современными miРНК этих двух разных особенностей их контекста соответствует одному из выводов статьи [5], данные которой мы использовали для нашего анализа: большинство miРНК имеют неразличимо равное сродство к обоим Ago2 и Ago3, тогда как остальные имеют значимое предпочтение к одному из этих белков перед другим.

Во-вторых, поскольку исходные экспериментальные данные содержали как близкородственные miРНК (например, табл. 1: № 24–26: hs-miR-19a, hs-miR-19b-1 и hs-miR-19b-2), так и отдаленные, то для преодоления этой естественной их несбалансированности мы сформировали репрезентативный набор их представителей для анализа с помощью ACTIVITy. Из указанных авторами работы [5] самых надежных случаев № 1–20 (табл. 1) мы выбрали лишь случаи минимальных и максимальных величин A_2, A_3, Σ, Δ и частот нуклеотидов A, U, G, C, W, R, M в miРНК (табл. 1, № 1–12). Поэтому набор данных для анализа представлял диапазоны исследуемых величин и контекстное разнообразие канонических miРНК [5], а не естественную гетерогенность результатов измерений. Система ACTIVITy не анализировала канонические miРНК № 13–28 и неканонические miРНК.

В-третьих, поскольку мы проверяли следствие найденной нами ранее [16] корреляции между содержанием miРНК в *A. thaliana* и содержанием WRHW и DRYD в miРНК, то здесь ACTIVITy вновь анализировала все возможные тетра-нуклеотиды $z_1 z_2 z_3 z_4$, ядра нуклеации гетеродуплекса miРНК/мРНК [26], вычисляя для каждого из $15^4 = 50625$ тетра-нуклеотидов взвешенные содержания [16]:

$$[z_1 z_2 z_3 z_4]_F(s_1 \dots s_i \dots s_{22}) = \sum_{i=1}^{19} F(i) \prod_{k=1}^4 \delta(s_{i+k-1} \in z_k), \quad (1)$$

здесь: $z_k \in \{A, U, G, C, W, R, M, K, Y, S, B, V, H, D, N\}$ [18]; $F(i)$ – вес $z_1 z_2 z_3 z_4$ в i -ой позиции miРНК, линейно-аддитивная аппроксимация его вклада в сродство miРНК к белку на основе эвристического правила: “чем выше $F(i)$, тем больше вклад”; $\delta(\text{истина}) = 1$, $\delta(\text{ложь}) = 0$.

В качестве примера наибольшего вклада центра miРНК в сродство к белкам Ago на рис. 1 разрывной линией показан Δ -образная весовая функция $F(i)$. Непрерывной V -образной линией показан пример функции взвешивания для случая наибольшего влияния 3'-конца miРНК на ее сродство к белкам Ago. Всего ACTIVITy [7–16] анализировала 90 V - и 90 Δ -образных $F(i)$ с одним пиком. При этом положение, ширину и степень асимметрии пика изменяли, что привело к $2 \times 90 \times 15^4 \approx 10^7$ вариантам $[z_1 z_2 z_3 z_4]_F$.

Таблица 1. Сродство канонических miРНК к белкам Ago2 и Ago3 человека [5], A₂ и A₃, и результаты его анализа с помощью АСТВИТУ [7]

№	код miРНК	последовательность miРНК*	Исходные экспериментальные данные из работы [5, дополнение]			Анализ (№ 1–12) и независимый контроль (№ 13–28)				In silico прогноз	
			A ₂	A ₃	$\Delta = (A_2 - A_3)/2$	[RHNK] _A	$\Sigma = (A_2 + A_3)/2$	[YRHB] _V	$A_2 = \Sigma + \Delta$	$A_3 = \Sigma - \Delta$	
1	hs-miR-342	uCUCACACAGAAAUCGCACCCGU	7.81	4.85	1.48	1.95	6.33	0.77	7.02	6.13	
2	hs-miR-21	UAGCUUAUCAGACUGAUGUUGA	8.34	7.24	0.55	1.59	7.79	1.45	7.76	7.24	
3	hs-miR-378	ACUGGACUUGGAGUCAGAAGGC	4.97	5.89	-0.46	1.26	5.43	0.00	5.62	5.45	
4	hs-miR-629	GUUCUCCSAACGUAAGCCAGC	5.28	7.98	-1.35	0.91	6.63	0.37	5.94	6.13	
5	hs-miR-92b	UAUUGCACUCGUCCCGCCUCC	6.48	8.98	-1.25	0.22	7.73	1.06	6.52	7.42	
6	hs-miR-221	AGCUACAUUGUCUGUGGGUUU	6.08	5.26	0.41	2.37	5.67	0.50	6.88	5.56	
7	hs-miR-29c	UAGCACSAUUUGAAAUCGGUUA	6.08	6.05	0.02	1.36	6.06	0.42	6.25	5.97	
8	hs-miR-210	CUGUGCGUGUGACAGCGGCUGA	6.81	7.96	-0.57	0.59	7.38	1.32	7.06	7.58	
9	hs-let-7d	CUAUACGACCUUGUCUCCUUUCU	6.49	6.46	0.01	1.53	6.47	1.00	7.11	6.66	
10	hs-miR-99b	CACCCGUAGAACCGACCUUGCG	8.55	6.70	0.92	1.43	7.62	2.09	8.54	8.18	
11	hs-miR-191	CAACGGAAUCCCAAAAGCAGCU	6.40	7.47	-0.53	1.05	6.94	0.82	6.62	6.67	
12	hs-miR-425	aAUGACACGAUCACUCCCGUUGA	7.33	8.50	-0.59	0.10	7.92	1.99	7.71	8.74	
13	hs-miR-93	aCUGCUGAGCUAGCACUCCCCGA	5.69	6.62	-0.46	0.38	6.16	1.13	6.70	7.43	
14	hs-miR-532	CAUGCCUUGAGUGUAGGACCGU	7.21	8.45	-0.62	0.71	7.83	1.40	7.23	7.63	
15	hs-miR-28	CACUAGAUUGUGAGCUCCUGGA	6.15	5.26	0.45	1.40	5.71	1.04	7.11	6.78	
16	hs-miR-25	CAUUGCACUUGUCUGGUCUGA	6.56	6.15	0.20	1.69	6.36	1.06	7.28	6.66	
17	hs-miR-29a	UAGCACCAUCUGAAAUCGGUUA	6.05	5.89	0.08	1.36	5.97	0.42	6.25	5.97	
18	hs-miR-23a	AUCACAUUGCCAGGGAUUUCCA	5.49	6.05	-0.28	1.15	5.77	0.30	5.97	5.91	
19	hs-miR-26a-1	UUCAAGUAUCCAGGAUAGGCU	5.35	6.05	-0.35	0.11	5.70	0.58	5.82	6.84	
20	hs-miR-26a-2	UUCAAGUAUCCAGGAUAGGCU	5.35	6.05	-0.35	0.11	5.70	0.58	5.82	6.84	
21	hs-miR-92a-1	UAUUGCACUUGUCCCGGCCUGU	7.95	8.03	-0.04	1.01	7.99	1.06	6.93	7.01	
22	hs-miR-92a-2	UAUUGCACUUGUCCCGGCCUGU	7.94	8.03	-0.04	1.01	7.98	1.06	6.93	7.01	
23	hs-miR-484	UCAGGCUCAGUCCCUCCCGGAU	7.69	7.58	0.06	0.00	7.63	1.00	6.31	7.45	
24	hs-miR-19a	uGUGCAAAUCUAUGCAAAACUGA	5.72	5.30	0.21	1.42	5.51	0.10	5.85	5.50	
25	hs-miR-19b-1	uGUGCAAAUCCAUAGCAAAACUGA	5.71	5.29	0.21	0.63	5.50	0.10	5.44	5.91	
26	hs-miR-19b-2	uGUGCAAAUCCAUAGCAAAACUGA	5.72	5.29	0.22	0.63	5.50	0.10	5.44	5.91	
27	hs-miR-501	AAUGCACCCGGGCAAGGAUUCU	5.68	5.57	0.05	0.20	5.62	0.47	5.71	6.63	
28	hs-miR-150	UCUCCCAACCCUUUGUACCCAGUG	8.79	9.43	-0.32	0.00	9.11	0.73	5.96	7.10	
Коэффициент линейной корреляции, r (№ miРНК или все)			0.750 (1–12), 0.501 (13–28)			0.861 (1–12), 0.613 (13–28)			0.661	0.664	
Статистическая значимость, α (№ miРНК или все)			0.005 (1–12), 0.05 (13–28)			0.0005 (1–12), 0.025 (13–28)			0.00025	0.00025	

* Анализ — заглавные буквы.

Таблица 2. Сродство неканонических miРНК к Ago2 и к Ago3 [5] и его прогноз (формулы (4), (5))

Каноническая miРНК	Неканоническая miРНК*	Ago2, A ₂		Ago3, A ₃		(A ₂ + A ₃)/2		A ₂ – A ₃	
		[5]	(4)	[5]	(5)	[5]	(4),(5)	[5]	(4), (5)
hs-miR-150	CUCCCAACCCUUGUACCAGUGU	2.64	5.80	2.56	6.94	2.60	6.37	0.07	-1.14
	UCUCCCAACCCUUGUACCAGUGU	6.27	5.80	6.63	6.94	6.45	6.37	-0.35	-1.14
	UCUCCCAACCCUUGUACCAGUGA	5.89	5.80	6.50	6.94	6.20	6.37	-0.61	-1.14
	UCUCCCAACCCUUGUACCAGUGC	4.95	5.80	5.57	6.94	5.26	6.37	-0.62	-1.14
	UCUCCCAACCCUUUUACCAGUG.	3.85	5.96	3.40	7.10	3.63	6.53	0.45	-1.14
	UCUCCCAACCCUUGUACCAGUU	3.43	5.96	2.56	7.10	3.00	6.53	0.87	-1.14
	UCUCCCAACCCUUGUACCAGUA	3.14	5.96	4.26	7.10	3.70	6.53	-1.13	-1.14
	UCUCCCAACCCUUGCACCAGUG	2.89	5.96	2.77	7.10	2.83	6.53	0.12	-1.14
	UCUCCCAACCCUCGUACCAGUG	2.48	5.96	3.14	7.10	2.81	6.53	-0.65	-1.14
	uGUCUCCCAACCCUUGUACCAGU.	2.40	6.21	2.30	7.24	2.35	6.73	0.10	-1.03
hs-miR-92a	UAUUGCACUUGUCCCGGCCUUU	4.28	6.93	3.56	7.01	3.92	6.97	0.72	-0.08
	UAUUGCACUUGUCCCGGCCUGA	3.74	6.93	4.20	7.01	3.97	6.97	-0.47	-0.08
	UAUUGCACUUGUCCCGGCCUGUA	3.61	5.80	4.17	6.08	3.89	5.94	-0.56	-0.28
	UAUUGCACUUGUCCCGGCCUGU	3.58	7.15	3.18	6.78	3.38	6.97	0.41	0.37
	UAUUGCACUUUCCCGGCCUGU	3.26	6.79	3.40	6.87	3.33	6.83	-0.14	-0.08
	UAUUGCACUUGUCCCGGCCUCU	2.83	6.93	2.77	7.01	2.80	6.97	0.06	-0.08
	UAUUGCACUUGUCCCGGCCUGUU	2.48	7.15	3.85	7.42	3.17	7.29	-1.37	-0.27
hs-miR-484	UCAGGCUCAGUCCCCUCCCGUU	3.53	6.31	4.08	7.45	3.81	6.88	-0.55	-1.14
	uCAGGCUCAGUCCCCUCCCGAAU	4.92	6.33	6.41	7.36	5.67	6.85	-1.49	-1.03
	UCAGGCUCAGUCCCCUCCCGAA	3.33	4.97	4.39	6.10	3.86	5.54	-1.06	-1.13
	uCAGGCUCAGUCCCCUCCCGAAA	2.89	4.97	4.28	6.10	3.59	5.54	-1.39	-1.13
hs-miR-425	uCAGGCUCAGUCCCCUCCCGAUA	2.64	6.28	4.96	7.42	3.80	6.85	-2.32	-1.14
	AaAUGACACGAUCACUCCCGUUGA	2.30	7.71	3.50	8.74	2.90	8.23	-1.19	-1.03
	AUGACACGAUCACUCCCGUUGA	4.42	7.71	5.19	8.74	4.81	8.23	-0.77	-1.03
	AAUGACACGAUCACUCCCGUUG.	5.96	7.59	4.57	8.51	5.27	8.05	1.38	-0.92
	AAUGACACUAUCACUCCCGUUGA	3.64	7.70	4.65	8.73	4.15	8.22	-1.02	-1.03
hs-miR-532	AAUGACACGAUCACUCCCUUUGA	3.00	6.44	2.40	7.58	2.70	7.01	0.60	-1.14
	CAUGCCUUGAGUGUAGGACCGA	2.83	7.23	3.47	7.63	3.15	7.43	-0.63	-0.40
	cAUGCCUUGAGUGUAGGACCGUA	3.09	6.41	4.52	6.74	3.81	6.58	-1.43	-0.33
	CAUGCCUUGAGUGUAGUACCGU	2.71	8.45	2.83	8.84	2.77	8.65	-0.13	-0.39
hs-miR-210	cAUGCCUUGAGUGUAGGACCGUU	2.64	7.76	2.64	8.09	2.64	7.93	0.00	-0.33
	CAUGCCUUGAGUGUAUGACCGU	2.40	8.84	2.83	9.23	2.62	9.04	-0.44	-0.39
	CUGUGCGUGUGACAGCGGCUGAU	3.09	9.03	3.61	9.35	3.35	9.19	-0.52	-0.32
	CUGUGCGUGUGACAGCGGCUGAA	2.71	7.68	3.43	8.00	3.07	7.84	-0.73	-0.32
hs-miR-92b	CUGUGCGUGUGACAUCGGCUGA	2.64	7.06	2.83	7.58	2.74	7.32	-0.19	-0.52
	CUGUGCGUGUGACAUCGGCUGA	2.56	7.53	2.94	7.83	2.75	7.68	-0.38	-0.30
	UAUUGCACUCGUCCCGGCCUCA	3.37	6.52	6.08	7.42	4.73	6.97	-2.72	-0.90
hs-miR-21	UAUUGCACUCGUCCCGGCCUCU	2.77	6.52	5.35	7.42	4.06	6.97	-2.58	-0.90
	uAUUGCACUCGUCCCGGCCUCCU	2.30	5.48	5.81	6.40	4.06	5.94	-3.51	-0.92
	UAGCUUAUCAGACUGAUGUUGAC	7.34	9.00	4.13	8.11	5.74	8.56	3.21	0.89
hs-miR-93*	UAGCUUAUCAGACUGAUGUUGAA	2.94	7.65	3.43	6.77	3.19	7.21	-0.49	0.88
	ACUGCUGAGCUAGCACUCCCG	4.66	6.64	5.58	7.23	5.12	6.94	-0.92	-0.59
hs-miR-29c	aCUGCUGAGCUAGCACUCCCGU	3.50	6.70	3.56	7.43	3.53	7.07	-0.06	-0.73
hs-miR-191	UAGCACCACUGAAAUCGGUUA	4.30	6.25	4.36	5.97	4.33	6.11	-0.05	0.28
hs-miR-221	CAACGGAAUCCCAAAGCAGCUG	5.15	5.53	4.98	5.54	5.07	5.54	0.17	-0.01
hs-miR-99b	AGCUACAUGUCUGCGGGUUUC	4.51	6.78	3.22	5.58	3.87	6.18	1.29	1.20
hs-let-7d	CACCCGUAAGACCGACCUUGC GA	5.04	7.29	3.00	6.89	4.02	7.09	2.05	0.40
	CUAUACGACCUGCUGCCUUUCA	2.64	7.11	2.64	6.66	2.64	6.89	0.00	0.45
Линейная корреляция, r (α)		-0.01 (>0.9)		-0.1 (>0.3)		-0.18 (>0.2)		0.468 (<0.001)	
Ранговая корреляция Кендалла, τ (α)		-0.11 (>0.2)		-0.1 (>0.3)		-0.17 (>0.09)		0.222 (<0.05)	

* Отличие от канонической – заглавные буквы.

Итак, с помощью автоматического режима ACTIVITY мы проверили каждый из $\approx 10^7$ вариантов $[z_1 z_2 z_3 z_4]_F$, рассчитывая каждый вариант независимо от других при каждом из двух наборов “входных данных” для этой системы, $\{S_n, \Sigma_n = (A_{2;n} + A_{3;n})/2\}_{1 \leq n \leq 12}$ и $\{S_n, \Delta_n = (A_{2;n} - A_{3;n})/2\}_{1 \leq n \leq 12}$.

Поскольку алгоритм ACTIVITY описан многократно [7–16], отметим коротко самое важное. Всего для каждого из $\approx 10^7$ вариантов $[z_1 z_2 z_3 z_4]_F$ оценивали пять типов корреляций. Прежде всего, проверяли наличие линейной корреляции между $[z_1 z_2 z_3 z_4]_F$ и Ψ (здесь и далее: $\Psi \in \{\Sigma, \Delta\}$), поскольку она дает самый ценный количественный прогноз. Поскольку этот прогноз не устойчив к неоднородности данных, то с помощью критериев χ^2 и Фишера дополнительно оценивали корреляции между знаками отклонений $[z_1 z_2 z_3 z_4]_F$ и Ψ от их среднеарифметических значений, обеспечивающие самый устойчивый малоценный прогноз “выше/ниже среднего”. Наконец, оценивали ранговые корреляции Спирмена и Кендалла, промежуточные по балансу между ценностью и устойчивостью прогнозов. Совместный учет разных типов корреляций компенсирует недостатки одних из них достоинствами других.

ACTIVITY также дополнительно проверяла 6 критериев однородности данных: равномерность распределений $[z_1 z_2 z_3 z_4]_F$ и Ψ ; нормальность отклонений Ψ от ее простой регрессии $\Psi([z_1 z_2 z_3 z_4]_F) = \zeta_0 + \zeta_1 [z_1 z_2 z_3 z_4]_F$ и отклонений $[z_1 z_2 z_3 z_4]_F$ от сопряженной простой регрессии $[z_1 z_2 z_3 z_4]_F(\Psi) = \zeta_0 + \zeta_1 \Psi$; независимость знаков отклонений $\{\Psi - (\zeta_0 + \zeta_1 [z_1 z_2 z_3 z_4]_F)\}$ и $\{[z_1 z_2 z_3 z_4]_F - (\zeta_0 + \zeta_1 \Psi)\}$. Итак, всего анализировали $5 + 6 = 11$ статистических критериев.

С той же целью ACTIVITY применяла метод bootstrap [27] для оценки каждого из 11 статистических критериев на 7 наборах данных: на всем введенном наборе экспериментальные данные (табл. 1: miРНК № 1–12) и на каждом из шести его поднаборов 50%-ного объема (по 6 из 12 miРНК № 1–12) с наибольшими, наименьшими и ближайшими к среднеарифметическим значениям независимо для случаев величин Ψ и $[z_1 z_2 z_3 z_4]_F$.

В результате проверки 11 критериев (m) на 7 наборах данных (n) получали $11 \times 7 = 77$ уровней значимости $\alpha_{m,n}$. С помощью нечеткой логики Задэ [28] ACTIVITY отображала каждый $\alpha_{m,n}$ на единую шкалу $[-1; 1]$, позитивные оценки $q_{m,n}$ которой означают достоверное выполнение m -ого критерия на n -ом наборе данных ($\alpha_{m,n} < 0.05$), негативные – недостоверное ($\alpha_{m,n} > 0.05$), “0” шкалы соответствует общепринятому порогу достоверности ($\alpha_{m,n} = 0.05$):

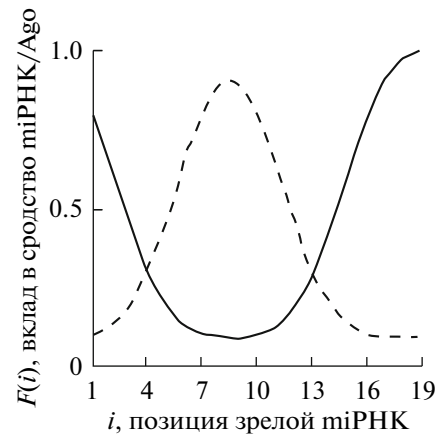


Рис. 1. Линейно-аддитивная аппроксимация (формула 1) вклада $F(i)$ тетра nukлеотида $z_1 z_2 z_3 z_4$ с началом в i -ой позиции miРНК, $\{s_i \in z_1, s_{i+1} \in z_2, s_{i+2} \in z_3, s_{i+3} \in z_4\}$, в средство этой miРНК к Ago-белкам с помощью эвристического правила: “чем выше $F(i)$, тем больше вклад”. Иллюстративные примеры: Λ-образный пик – наибольший вклад центрального района miРНК в ее средство к белкам (всего 90 вариантов с разными положениями, шириной и степенью асимметрии пика), V-образная линия – наибольший вклад 3'-конца miРНК (всего 90 вариантов).

$$q_{m,n}([z_1 z_2 z_3 z_4]_F; \Psi) = \begin{cases} 1 & \text{если } \alpha_{m,n} \leq 0.01; \\ 1.3 - 28.3\alpha_{m,n} + 55.6\alpha_{m,n}^2 & \text{если } 0.01 \leq \alpha_{m,n} \leq 0.1; \\ -1 & \text{если } \alpha_{m,n} \geq 0.1. \end{cases} \quad (2)$$

Шкала (2) задана параболой по трем $(-1; 0.1)$, $(0; 0.05)$ и $(1; 0.01)$ ее точкам $(q; \alpha)$. Так все $[z_1 z_2 z_3 z_4]_F$ получили по 77 частных оценок $q_{m,n}([z_1 z_2 z_3 z_4]_F; \Psi)$ выполнения 11 критериев на 7 наборах данных [28]. В терминах теории полезности для принятия решений [29] оценку полезности учета $[z_1 z_2 z_3 z_4]_F$ в прогнозе Ψ дало среднеарифметическое этих частных оценок:

$$Q([z_1 z_2 z_3 z_4]_F; \Psi) = \frac{\sum_{m=1}^{11} \sum_{n=1}^7 q_{m,n}([z_1 z_2 z_3 z_4]_F; \Psi)}{11 \times 7}. \quad (3)$$

Наибольшая позитивная $Q([z_1 z_2 z_3 z_4]_F; \Psi)$ указала на тот тетра nukлеотид $z_1 z_2 z_3 z_4$ и с тем его линейно-аддитивным вкладом F в средство miРНК к Ago2 и Ago3, Ψ , которым нашлись самые однородные данные с самым большим количеством самых достоверных корреляций.

Проверка найденных особенностей контекста miРНК с помощью независимых данных

Сначала для найденных ACTIVITY по $\{S_n, \Sigma_n\}_{1 \leq n \leq 12}$ и $\{S_n, \Delta_n\}_{1 \leq n \leq 12}$ лучших $Q([\xi_1 \xi_2 \xi_3 \xi_4]_{\Phi}; \Sigma)$ и

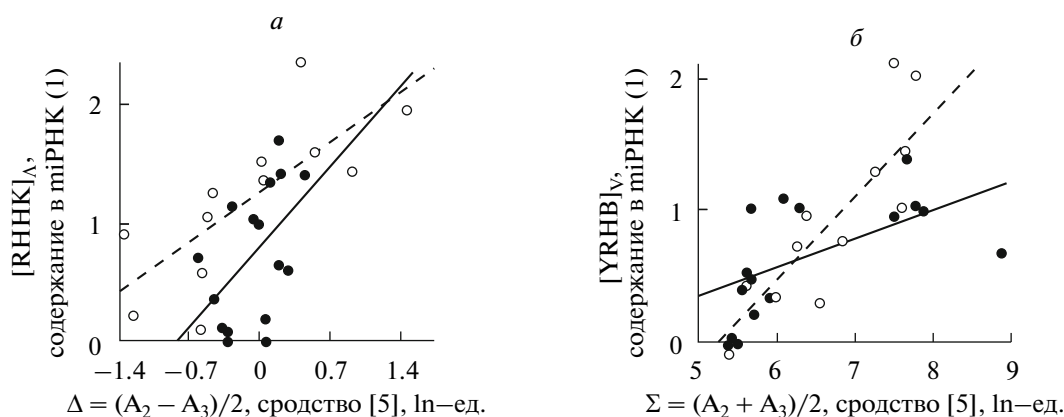


Рис. 2. Результаты автоматического анализа экспериментальных данных [5] с помощью ACTIVITY: анализируемая выборка miРНК № 1–12 из табл. 1 (○, пунктир) и контрольная выборка miРНК № 13–28 (●, линия). *a* – Достоверная линейная корреляция между полуразностью $\Delta = (A_2 - A_3)/2$ величин A_2 и A_3 сродства miРНК к Ago2 и Ago3, соответственно, и содержанием $[RNNK]_{\Delta}$ тетра nukлеотида RNNR вблизи центра miРНК (формула (1), рис. 1: пунктир); *б* – достоверная линейная корреляция между полусуммой $\Sigma = (A_2 + A_3)/2$ и содержанием $[YRNB]_{\Sigma}$ тетра nukлеотида YRNB вблизи 3'-конца miРНК (рис. 1, линия).

$Q([\zeta_1 \zeta_2 \zeta_3 \zeta_4; \Delta]_{\Omega}; \Delta)$ мы проверили линейные корреляции между Σ и $[\xi_1 \xi_2 \xi_3 \xi_4]_{\Phi}$, Δ и $[\zeta_1 \zeta_2 \zeta_3 \zeta_4; \Delta]_{\Omega}$ на независимых контрольных данных $\{S_n; \Sigma_n; \Delta_n\}_{13 \leq n \leq 28}$.

Затем мы провели стандартный тест пермутаций [30]. В каждом из 100 независимых однотипных испытаний перемешивали случайным образом элементы обучающих наборов $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 12}$ и каждую полученную смесь анализировали с помощью ACTIVITY [7].

Наконец, на всех $\{S_n; \Sigma_n; \Delta_n\}_{1 \leq n \leq 28}$ канонических miРНК без оптимизации нашли простые регрессии $\Sigma = \gamma_0 + \gamma_1[\xi_1 \xi_2 \xi_3 \xi_4]_{\Phi}$ и $\Delta = \psi_0 + \psi_1[\zeta_1 \zeta_2 \zeta_3 \zeta_4; \Delta]_{\Omega}$ (линейные сдвиг и масштаб) и, вновь без оптимизации, преобразовали их обратно в $A_2(S) = \Sigma(S) + \Delta(S)$ и $A_3(S) = \Sigma(S) - \Delta(S)$. Проверили выведенные таким путем $A_{2;n}$ и $A_{3;n}$ линейные корреляции между ними и соответственно $A_{2;n}$ и $A_{3;n}$ из $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 28}$. В выведенных формулах $A_2(S) = \Sigma(S) + \Delta(S)$ и $A_3(S) = \Sigma(S) - \Delta(S)$ также оценили значимость их линейных корреляций с множественными регрессиями $A_{2;n}$ и $A_{3;n}$ по обоим $[\xi_1 \xi_2 \xi_3 \xi_4]_{\Phi}$ и $[\zeta_1 \zeta_2 \zeta_3 \zeta_4; \Delta]_{\Omega}$, оптимизированным на $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 28}$.

В заключение предсказали на основе выведенных формул сродство 48 неканонических miРНК к Ago2 и к Ago3, оценив значимость (α) линейной (r) и ранговой Кендалла (τ) корреляций между этими прогнозами и соответствующими измерениями из работы [5].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Результаты автоматического анализа данных с помощью ACTIVITY и их контроль

Обучающие выборки $\{S_n; \Sigma_n\}_{1 \leq n \leq 12}$ и $\{S_n; \Delta_n\}_{1 \leq n \leq 12}$ независимо проанализированы ACTIVITY в автома-

тическом режиме, как описано выше. В итоге, найдена самая высокая оценка $Q([RNNK]_{\Delta}; \Delta) = 0.363$ (формула 3) в случае анализа выборки $\{S_n; \Delta_n\}_{1 \leq n \leq 12}$. Так для предпочтения miРНК к Ago2 перед Ago3 самым важным оказался тетра nukлеотид RNNK [18] с его наибольшим линейно-аддитивным вкладом $F(i)$ в сродство к белкам Ago в центре miРНК (рис. 1: Δ -образный пунктир). Величины $[RNNK]_{\Delta}$ для всех 28 miРНК даны в табл. 1. На рис. 2*a* показана значимая ($r = 0.501$, $\alpha < 0.05$) линейная корреляция между Δ и $[RNNK]_{\Delta}$ также для независимых контрольных данных (miРНК № 13–28). Следующая по величине $Q([RNNK]_{\Sigma}; \Sigma) = 0.343$ указала на тот же тетра nukлеотид RNNK с более узким пиком $F(i)$ в центре зрелых miРНК. Важность последовательности центра зрелой miРНК для различия между Ago2 и Ago3 по их сродству к ней согласуется с различием биологических функций этих белков: в случае Ago2 возможно разрезание мРНК в центре ее гетеродуплекса с miРНК, в случае Ago3 – нет [5]. И так, коадаптацией контекста miРНК и белков Argonaute при дивергенции “Ago2 ← предок → Ago3” [3] была, по-видимому, их селекция на многовариантность способов регуляции трансляции: ингибировать трансляцию (Ago2 и Ago3) и/или необратимо предотвратить ее посредством разрезания мРНК (Ago2).

По выборке $\{S_n; \Sigma_n\}_{1 \leq n \leq 12}$ ACTIVITY нашла наибольшую $Q([YRNB]_{\Sigma}; \Sigma) = 0.358$ для тетра nukлеотида YRNB [18] с наибольшим линейно-аддитивным вкладом $F(i)$ в сродство к обоим Ago2 и Ago3 на 3'-конце miРНК (рис. 1, линия). Это согласуется с пространственной структурой комплекса мРНК/мнкРНК/Ago из *Pyrococcus furiosus* [6], где Ago действительно связывает 3'-конец мнкРНК. По-анalogии с этим известным контактом между Ago и 3'-концом мнкРНК из *Pyrococcus furiosus* [6],

коадаптацией контекстов предков miРНК и предков Ago2 и Ago3 человека на сродство между ними могла быть селекция 3'-концевых районов зрелых miРНК. Величины $[YRNB]_V$ для всех 28 зрелых канонических miРНК [5] представлены в табл. 1. Оценки $[YRNB]_V$ и Δ коррелируют достоверно для независимой контрольной выборки miРНК № 13–28 (рис. 2б: $r = 0.613$, $\alpha < 0.025$).

В стандартном испытании пермутаций [30] для каждой взятой из обучающих данных $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 12}$ случайной miРНК $S_n^\#$ независимо один от другого брали случайные $A_2^\#$ и $A_3^\#$, строили $\{S_n^\#; \Sigma_n^\# = (A_{2;n}^\# + A_{3;n}^\#)/2\}_{1 \leq n \leq 12}$ и $\{S_n^\#; \Delta_n^\# = (A_{2;n}^\# - A_{3;n}^\#)/2\}_{1 \leq n \leq 12}$, которые анализировали с помощью системы ACTIVITY. Всего проведено 100 таких испытаний, ни в одном из которых никакие сочетания RHNK- и YRNB-подобных тетрануклеотидов $z_1 z_2 z_3 z_4$ ни с какими V-или L-образными весовыми функциями $F(i)$ не получили позитивных оценок формулы (3).

При априорно ожидаемой частоте $P_0 = 0.05$ иного исхода по случайным причинам и при допущении его в 101-ом испытании закон бинома дает оценку вероятности $p = 0.036$ такого результата теста пермутаций [30]. Это означает, что соответствие между обучающими данными $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 12}$ [5] и найденными по ним содержаниями тетрануклеотидов $[RHNK]_\Lambda$ и $[YRNB]_V$ в зрелых канонических miРНК оказалось достоверным ($\alpha < 0.05$).

Вывод формул прогноза сродства канонических miРНК к белкам Ago2 и Ago3 человека

По всем miРНК (№ 1–28, табл. 1) найдены простые регрессии вспомогательных оценок без их оптимизации (линейные сдвиг и масштаб): $\Sigma = 5.54 + 1.35[YRNB]_V$ и $\Delta = -0.57 + 0.52[RHNK]_\Lambda$. Затем, мы преобразовали их вновь без оптимизации обратно к измеримым величинам [5], $A_2 = \Sigma + \Delta$ и $A_3 = \Sigma - \Delta$, с помощью двух следующих формул:

$$A_2(S) = 4.97 + 1.35[YRNB]_V(S) + 0.52[RHNK]_\Lambda(S), \quad (4)$$

$$A_3(S) = 6.11 + 1.35[YRNB]_V(S) - 0.52[RHNK]_\Lambda(S). \quad (5)$$

Прогнозы формул (4) и (5) для всех 28 зрелых канонических miРНК представлены в табл. 1 и на рис. 3. Они достоверны: $r = 0.661$ ($\alpha < 0.00025$) и $r = 0.664$ ($\alpha < 0.00025$).

Также обнаружено достоверное согласие формул (4) и (5) с множественными линейными регрессиями A_2 и A_3 по $[YRNB]_V$ и $[RHNK]_\Lambda$, соответственно $r = 0.953$ и $\alpha < 10^{-14}$, $r = 0.963$ и $\alpha < 10^{-15}$ (данные не показаны). В этих линейных регрессиях $[RHNK]_\Lambda$ и

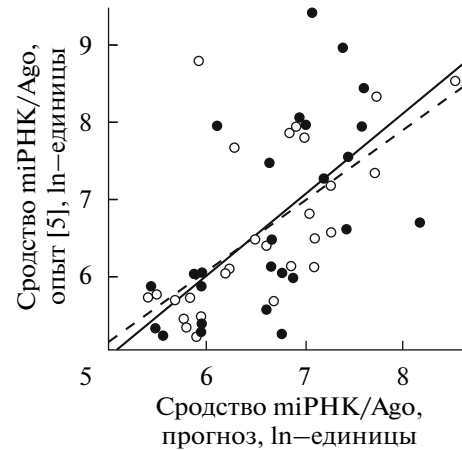


Рис. 3. Результаты измерения A_2 и A_3 [5] сродства зрелых канонических miРНК к белкам человека, соответственно Ago2 (○, пунктир) и Ago3 (●, линия), достоверно линейно коррелируют с прогнозами этого сродства по формулам (4) и (5) на основе нуклеотидных последовательностей зрелых miРНК.

$[YRNB]_V$ объясняли $10 \pm 14\%$ и $69 \pm 14\%$ вариансы сродства miРНК к Ago2 соответственно. В случае Ago3 эти показатели составляли $44 \pm 15\%$ и $50 \pm 15\%$. В случае Ago2 вклад $[RHNK]_\Lambda$, эвристически ассоциированный нами с дивергенцией “Ago2 ← предок → Ago3” [3], оказался недостоверным, поскольку величина 10% этого вклада в вариансу измерений A_2 [5] не превышает доверительного интервала $\pm 14\%$ регрессионной оценки этой величины. В случае Ago3 вклады обоих $[RHNK]_\Lambda$ и $[YRNB]_V$ были, напротив, равноценными: разность $50 - 44 = 6\%$ между величинами их вкладов в вариансу измерения A_3 [5] соответствует точности $\pm 15\%$ регрессионной оценки каждого из этих вкладов. Это означает, что из двух гомологов Ago2 и Ago3, по сродству к miРНК белок Ago2 оказался более похож на их общего предка. Это согласуется с независимым результатом филогенетического анализа всех известных аминокислотных последовательностей белков семейства Argonaute [3].

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Итак, стандартный тест пермутаций [30] указал на значимость соответствия, прежде всего, между данными $\{S_n; A_{2;n}; A_{3;n}\}_{1 \leq n \leq 12}$ [5] и содержанием тетрануклеотидов $[RHNK]_\Lambda$ и $[YRNB]_V$ в зрелых канонических miРНК, полученного в результате анализа этих данных с помощью ACTIVITY. Кроме того, эти контекстные особенности зрелых канонических miРНК человека достоверно коррелируют с измерениями сродства miРНК к белкам Ago RISC человека как на подготовленных репрезентативных данных для их *in silico* анализа, так и на оставшихся неоднородных независимых данных из опыта [5]. Наконец,

они согласуются с известной пространственной структурой комплекса мРНК/мнкРНК/Ago из *Purococcus furiosus* [6], различием по воздействию RISC с содержанием Ago2 и Ago3 на мРНК-мишень (разрезание и/или ингибирование [5]) и филогенетическим древом всех известных аминокислотных последовательностей белков семейства Ago [3]. Существенно, что эти две закономерности выбраны из гигантского массива 10^7 однотипных с ними потенциальных закономерностей (1), каждую из которых подвергали 77 испытаниям при уровне значимости $\alpha < 0.05$ [7] на малой выборке из 12 miРНК [5]. Поэтому нельзя совсем исключить случайность их выбора системой АСТИВІТУ. Однако вероятность случайного выбора можно оценить с помощью следующих расчетов.

Прежде всего, согласно распределению редких событий Пуассона следует ожидать не менее четырех ($77 \times 0.05 \approx 4$) случайных положительных исходов в 77 испытаниях и верхнюю 5%-ную доверительную границу $4 \times (1 + t_{\alpha < 0.05; v = 77}) = 4 \times (1 + 1.6649) \approx 11$ их количества. Этой границе соответствует $77 - 11 = 66$ негативных исходов, благодаря которым ($11 \ll 66$) формула (3) даст в целом негативную оценку полезности $Q < 0$. Чтобы получить позитивную оценку необходимо 50% позитивных из 77 исходов при значимости $\alpha < 0.05$. Закон бинома оценивает вероятность этого величиной:

$$P(Q(\xi) > 0) = \sum_{n=38}^{77} C_{77}^n \times 0.05^n \times (1 - 0.05)^{77-n} < 10^{-28}.$$

Тогда неравенство Бонферрони ограничивает искомую вероятность верхней оценкой:

$$P\{Q([z_1 z_2 z_3 z_4]_F; \Psi) > 0\} < P\{Q([z_1 z_2 z_3 z_4]_F; \Delta) > 0\} + P\{Q([z_1 z_2 z_3 z_4]_F; \Sigma) > 0\} < 2 \times 10^7 \times P\{Q(\xi) > 0\} < 2 \times 10^7 \times 10^{-28} < 10^{-20}.$$

В порядке дискуссии о возможных применениях формул (4) и (5), выведенных нами для канонических miРНК, мы предсказали с их помощью также сродство 48 неканонических miРНК к белкам Ago2 и Ago3 и сравнили эти прогнозы с измерениями [5] (табл. 2). Видно, что для неканонических miРНК мы не нашли ни линейных (r), ни ранговых (τ) корреляций ни для Ago2 ($r = -0.01$, $\alpha > 0.9$ и $\tau = -0.11$, $\alpha > 0.2$), ни для Ago3 ($r = -0.1$, $\alpha > 0.3$ и $\tau = -0.1$, $\alpha > 0.3$), ни для полусуммы их сродства к miРНК, которую мы эвристически ассоциировали с их общим предком ($r = -0.18$, $\alpha > 0.2$ и $\tau = -0.17$, $\alpha > 0.09$). Это указывает на отсутствие какой-либо эволюционной селекции Ago2, Ago3 и их общего предка на сродство к неканоническим miРНК. Действительно, содержание канонических miРНК в клетках много больше по сравнению с неканоническими miРНК, последние к то-

му же принято ассоциировать с ошибками биогенеза miРНК или их особыми модификациями [5].

Тем не менее, прогноз разности величин сродства miРНК/Ago2 и miРНК/Ago3 оказался достоверным в случае обоих типов корреляций ($r = 0.488$, $\alpha < 0.001$ и $\tau = 0.222$, $\alpha < 0.05$). Это означает, что все копии определенной неканонической miРНК распределены между ее комплексами с белками Ago2 или Ago3 пропорционально предсказанному с помощью формул (4) и (5) величинам сродства этой miРНК к Ago2 и к Ago3.

Итак, прогнозы формул (4) и (5) для канонических и неканонических miРНК в целом соответствуют общепринятым представлениям о взаимодействии и эволюции макромолекул. Поэтому они могут быть полезны для конструирования miРНК-подобных регуляторов генов.

Работа частично поддержана Интеграционным проектом Сибирского отделения РАН № 119 “Постгеномная биоинформатика: компьютерный анализ и моделирование молекулярно-генетических систем” и Программами РАН № 21 “Фундаментальные науки – медицине” (проект 26 “Информационно-компьютерные подходы к поиску оптимальных фармакологических стратегий коррекции патологических состояний организма”), № 22 “Молекулярная и клеточная биология” (проект 8 “Системная биология: компьютерно-экспериментальные подходы”) и № 23 “Биологическое разнообразие” (проект 29 “Биоинформатика генетической изменчивости: исследование влияния мутаций на молекулярно-генетические системы организмов”).

СПИСОК ЛИТЕРАТУРЫ

1. Lee R.C., Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*. **294**, 862–864.
2. Bartel D. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*. **116**, 281–297.
3. Murphy D., Dancis B., Brown J.R. 2008. The evolution of core proteins involved in microRNA biogenesis. *BMC Evol. Biol.* **8**, 92.
4. Meister G., Landthaler M., Patkaniowska A., et al. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*. **15**, 185–197.
5. Azuma-Mukai A., Oguri H., Mituyama T., et al. 2008. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc. Natl. Acad. Sci. USA*. **105**, 7964–7969.
6. Song J.J., Smith S.K., Hannon G.J., et al. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*. **305**, 1434–1437.
7. Ponomarenko M.P., Kolchanova A.N., Kolchanov N.A. 1997. Generating programs for predicting the activity of functional sites. *J. Comput. Biol.* **4**, 83–90.
8. Пономаренко М.П., Савинкова Л.К., Кель А.Э., Колчанов Н.А. 1997. Компьютерное моделирование последовательностей ТАТА-боксов промоторов эукариот. *ДАН*. **355**, 557–561.

9. Пономаренко М.П., Пономаренко Ю.В., Титов И.И. и др. 1998. Предпочтительность ResA-филамента к последовательностям ДНК коррелирует с генетическим кодом. *ДАН*. **363**, 122–125.
10. Колчанов Н.А., Пономаренко М.П., Пономаренко Ю.В. и др. 1998. Функциональные сайты геномов про- и эукариот: компьютерное моделирование и предсказание активности *Молекуляр. биология*. **32**, 255–267.
11. Пономаренко М.П., Пономаренко Ю.В., Колчанов Н.А. 1999. Вклад сигналов и анти-сигналов в мутационный спектр сайта встраивания td-интрона. *Биофизика*. **44**, 655–663.
12. Ponomarenko M.P., Ponomarenko J.V., Frolov A.S., et al. 1999. Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins. *Bioinformatics*. **15**, 687–703.
13. Васильев Г.В., Меркулов В.М., Кобзев В.Ф. и др. 2000. Точковые мутации в районе 663–666 п.н. интрона 6 гена триптофаноксигеназы, связанные с рядом психических расстройств, разрушают сайт связывания фактора транскрипции YY1. *Молекуляр. биология*. **34**, 214–222.
14. Ponomarenko J.V., Furman D.P., Frolov A.S., et al. 2001. ACTIVITY: a database on DNA/RNA sites activity adapted to apply sequence-activity relationships from one system to another. *Nucleic Acids Res.* **29**, 284–287.
15. Пономаренко П.М., Савинкова Л.К., Драчкова И.А. и др. 2008. Пошаговая модель связывания ТВР/ТАТА бокс позволяет предсказать наследственные заболевания человека по точечным полиморфизмам. *ДАН*. **419**, 828–832.
16. Пономаренко М.П., Омелянчук Н.А., Катохин А.В., Колчанов Н.А. 2006. Содержание микроРНК в *Arabidopsis thaliana* коррелирует с встречаемостью тетра nukлеотидов WRHW и DRYD. *Инф. Вест. ВО-ЛуС*. **10**, 304–311.
17. Axtell M., Bartel D. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell*. **17**, 1658–1673.
18. IUPAC-IUB commission on biochemical nomenclature (CBN) 1971. Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. *J. Mol. Biol.* **55**, 299–310.
19. Lu C., Tej S.S., Luo S., Haudenschild C.D., Meyers B.C., Green P.J. 2005. Elucidation of the small RNA component of the transcriptome. *Science*. **309**, 1567–1569.
20. Hwang H.W., Wentzel E.A., Mendell J.T. 2007. A hexanucleotide element directs microRNA nuclear import. *Science*. **315**, 97–100.
21. Fiedler S.D., Carletti M.Z., Hong X., Christenson L.K. 2008. Hormonal regulation of MicroRNA expression in periovulatory mouse mural granulosa cells. *Biol. Reprod.* **79**, 1030–1037.
22. van Rooij E., Olson E.N. 2007. MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets. *J. Clin. Invest.* **117**, 2369–2376.
23. Martinez J., Tuschl T. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980.
24. Chatterjee S., Grosshans H. 2009. Active turnover modulates mature microRNA activity in *Caenorhabditis elegans*. *Nature*. **461**, 546–549.
25. Sakamoto M., Lloyd G.T., Benton M.J. 2010. Phylogenetically structured variance in felid bite force: the role of phylogeny in the evolution of biting performance. *J. Evol. Biol.* **23**, 463–478.
26. Porschke D., Eigen M. 1971. Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* **62**, 361–381.
27. Hayes K.G., Perl M.L., Efron B. 1989. Application of the bootstrap statistical method to the tau-decay-mode problem. *Phys. Rev. D. Part. Fields*. **39**, 274–279.
28. Zadeh L.A. 1965. Fuzzi sets. *Inform. Control*. **8**, 338–353.
29. Fishburn P.C. 1970. *Utility theory for decision making*. N.Y: John Wiley & Sons, 227 p.
30. Sohn I., Owzar K., George S.L., Kim S., Jung S.H. 2009. A permutation-based multiple testing method for time-course microarray experiments. *BMC Bioinformatics*. **10**, 336.