

Локалізація сайтів 5'-CG-3', 5'-CNG-3', 5'-GC-3' та 5'-GNC-3' у послідовностях мікроРНК

В. А. Галицький, С. В. Комісаренко

Інститут біохімії імені О. В. Палладіна НАН України
Вул. Леонтовича, 9, Київ, Україна, 01601

volha@biochem.kiev.ua; svk@biochem.kiev.ua

Мета. Встановлення закономірностей у локалізації динуклеотидів 5'-CG-3' та тринуклеотидів 5'-CNG-3' у послідовностях мікроРНК (miRNA). **Методи.** Біоінформатичний та статистичний аналіз послідовностей зрілих miRNA. **Результати.** Один із найвищих піків концентрації сайту 5'-CG-3' локалізується у місці розташування нуклеотиду 10 з 5'-кінця miRNA. Рівень концентрації сайту 5'-CG-3' у даному піку більш ніж учетверо перевищує середню концентрацію даного динуклеотиду у геномі. Інші еволюційно консервативні піки у послідовності miRNA локалізуються в районі нуклеотидів 16 та 21. Піки концентрації сайту 5'-CNG-3' відповідають нуклеотидам 4 і 8. Приблизно половина послідовностей miRNA містить два чи більше сайтів 5'-CG-3' та 5'-CNG-3', відстань між якими становить у більшості випадків 1 або 3–9 нуклеотидів. **Висновки.** Піки концентрації сайтів 5'-CG-3' і 5'-CNG-3' не зосереджені в межах seed-ділянки, а розподілені уздовж усієї послідовності miRNA. Однак в одній і тій же молекулі miRNA дані сайти розташовані в межах одного витка подвійної спіралі. Така картина розподілу зазначених сайтів може вказувати на участь у miRNA-залежному метилюванні ДНК усієї послідовності miRNA.

Ключові слова: мікроРНК, динуклеотид 5'-CG-3', тринуклеотид 5'-CNG-3', розподіл сайтів.

Вступ. МікроРНК (miRNA) являють собою некодуєчі РНК, які використовуються клітинами для регуляції експресії генів, боротьби з вірусами, пригнічення поширення транспозонів та інших мобільних елементів у геномі. Більшість miRNA транскрибується зі своїх власних генів, котрі містять генний промотор і регуляторні сайти, однак близько 40 % miRNA походять з інтронів білок-кодуєчих генів, а окремі – навіть з екзонів [1].

Послідовності, що містять miRNA, транскрибуються з утворенням попередників miRNA (pri-miRNA). Останні, які є паліндромами, формують недосконалі шпильки, що спершу перетворюються ядерною рибонуклеазою III Drosha на дволанцюгові pre-miRNA довжиною близько 60 нуклеотидів

[2]. У цитоплазму pre-miRNA надходять завдяки активному транспорту за участі експорину 5 [3] і надалі розщеплюються рибонуклеазою Dicer з утворенням фрагментів довжиною 22–23 нуклеотиди. Дані фрагменти потрапляють у комплекс RISC (RNA interfering silencing complex), де розплітаються за допомогою білка сімейства Argonaute. Один із ланцюгів, а саме той, 5'-кінець якого термодинамічно є менш стабільним [4, 5], залишається надалі асоційованим з РНК-зв'язувальним доменом (PAZ-доменом) цього білка [6] і являє собою зрілу miRNA.

Для посттранскрипційного miRNA-залежного сайленсингу у тваринних клітинах достатньо, щоб з транскриптом-мішенню взаємодіяв не увесь ланцюг молекули miRNA, як у рослинних клітин [7], а так звана seed-ділянка – послідовність, розташована з 1-го або 2-го по 7-й чи 8-й нуклеотид на 5'-кінці

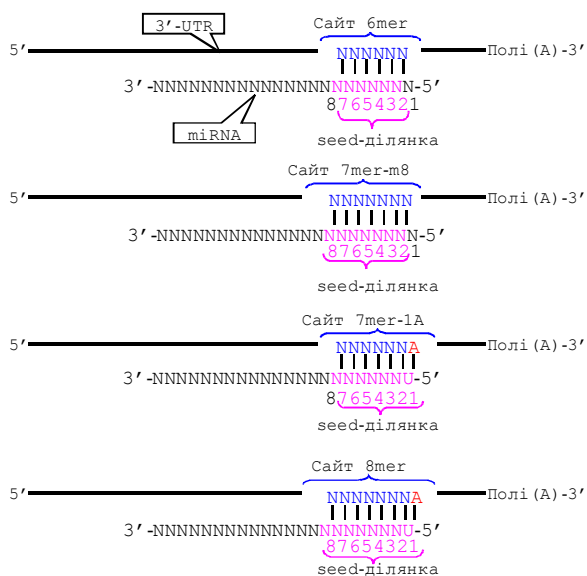


Рис. 1. Типи seed-ділянки

miRNA (рис. 1) [8, 9]. У цьому разі miRNA діє у складі не комплексу RISC, а miRNA-рибонуклеопротеїну (miRNP – miRNA-ribonucleoprotein) і спричиняє не розщеплення мРНК, а пригнічення її трансляції після стадії ініціації [10]. Є чотири основних типи seed-ділянки: секстимерна (6mer, утворена нуклеотидами з 2-го по 7-й), септимерна m8 (7mer-m8, з 2-го по 8-й), септимерна 1A (7mer-1A, з 1-го по 7-й) та октамерна (8mer, з 1-го по 8-й), причому в обох останніх випадках першим нуклеотидом з 3'-кінця у сайті зв'язування miRNA є аденін (рис. 1) [9, 11]. Коли взаємодія seed-ділянки з мішенню є неповною (seed-ділянка виявляється надто короткою, утворює петлю чи містить нуклеотида, не комплементарні мішені), miRNA все ж може проявити свою функцію – за умови, що кілька з її нуклеотидів, котрі займають положення 12–17, також зв'язуються з мішенню [11].

Одним із авторів цієї роботи раніше встановлено, що динуклеотиди 5'-CG-3' та тринуклеотиди 5'-CNG-3' (літерою N позначено будь-який з нуклеотидів) у послідовностях miRNA зустрічаються достовірно значно частіше, ніж у середньому в геномі [12, 13]. Це дало підставу для припущення, що miRNA гібридизуються з ДНК-мішенями і викликають у них метилювання *de novo* залишків цитозину у сайтах, комплементарно зв'язаних з динуклеотидами 5'-CG-3' та тринуклеотидами 5'-CNG-3', перебуваючими у складі miRNA [13]. Також показа-

но, що деяка кількість послідовностей miRNA містить два і більше сайтів 5'-CG-3' або 5'-CNG-3' [13]. У зв'язку з цим, а також зважаючи на неповну взаємодію miRNA з мішенню при посттранскрипційному сайленсингу, постало завдання з'ясувати, чи проявляються певні закономірності у локалізації зазначених сайтів та у відстані між ними. Одночасно, враховуючи наявність у базах даних значно більшої кількості послідовностей miRNA, аніж на час попереднього нашого дослідження, є нагода уточнити значення середньої концентрації сайтів 5'-CG-3' та 5'-CNG-3' у miRNA.

Матеріали і методи. Нуклеотидні послідовності зрілих miRNA людини, миші, пацюка – загалом 1639 – отримано з бази даних miRBase, <http://www.mirbase.org/> [14].

Послідовності аналізували *in silico* з використанням програмних пакетів Discovery Studio Gene 1.5 (виробництво Accelrys Inc.); Vector NTI Advance 10 (Invitrogen Corp.); Lasergene 6 (DNA STAR Inc.), а також Microsoft Excel 2003; статистичне опрацювання результатів – за допомогою програми Microsoft Excel 2003.

Вміст гуаніну і цитозину у досліджуваних послідовностях (v) визначали за формулою

$$v = \frac{a}{s},$$

де a – кількість гуанінових або цитозинових нуклеотидів; s – сумарна кількість нуклеотидів.

Оскільки в послідовності із z нуклеотидів міститься $z - 1$ можливих динуклеотидів, середня концентрація динуклеотидів у m послідовностях, що загалом містять s нуклеотидів, повинна обчислюватися за формулою

$$\eta = \frac{x}{s - m},$$

де η – середня концентрація динуклеотидів; x – кількість виявлених динуклеотидів.

Виходячи з міркування, що в послідовності із z нуклеотидів міститься $z - 2$ можливих тринуклеотидів, середню концентрацію тринуклеотидів у m послідовностях, що загалом містять s нуклеотидів, можна вирахувати за формулою

$$\varphi = \frac{y}{s - 2m},$$

Таблиця 1
Вміст нуклеотидів у геномі та досліджених послідовностях miRNA

Вид	Кількість послідовностей	Вміст нуклеотиду в послідовностях miRNA, %				Вміст G=C у геномі, % [15]	Розрахункова концентрація сайтів 5'-C(N)G-3' та 5'-G(N)C-3' у випадковій послідовності геному
		A	U	G	C		
<i>Homo sapiens</i>	718	22,58 ± 0,66	28,27 ± 0,71	26,53 ± 0,69	22,62 ± 0,66	42,00	4,41
<i>Mus musculus</i>	591	23,67 ± 0,73	27,63 ± 0,77	26,82 ± 0,76	21,88 ± 0,71	42,20	4,45
<i>Rattus norvegicus</i>	330	23,47 ± 0,98	28,93 ± 1,04	25,65 ± 1,01	21,95 ± 0,95	43,90	4,82
Всього	1639	23,15 ± 0,44	28,18 ± 0,47	26,46 ± 0,46	22,22 ± 0,43	–	–

Примітка. Жирним шрифтом виділено достовірну ($p \leq 0,05$) відмінність концентрації нуклеотиду від рівня, середнього для геному.

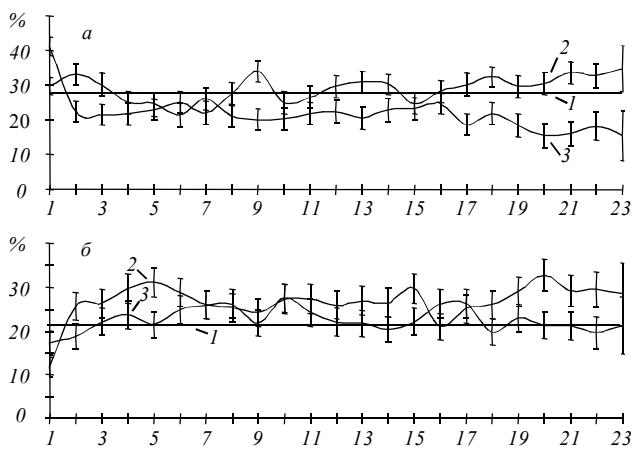


Рис. 2. Розподіл урацилу і аденіну (а – криві 2, 3) та гуаніну і цитозину (б – криві 2, 3) у послідовностях miRNA людини (а, б: 1 – середній рівень у геномі) ($p \leq 0,05$). По осі абсцис – порядковий номер нуклеотиду у послідовності miRNA

де φ – середня концентрація тринуклеотидів; y – кількість виявлених тринуклеотидів.

Результати даних обчислень у статті подано у вигляді

$$P = (\varphi \pm \Delta) \cdot 100 \%,$$

де P – частка у генеральній сукупності; φ – частка у вибірковій сукупності (тобто середня концентрація мононуклеотиду, динуклеотиду чи тринуклеотиду); Δ – гранична похибка частки.

Результати і обговорення. Для послідовностей miRNA характерним є значно вищий вміст гуаніну і значно нижчий – аденіну, ніж для геному загалом (табл. 1). Відповідно, на більшості позицій нуклеотидів у послідовності miRNA з вищою імовірністю буде зустрічатися гуанін і з меншою – аденін порівняно з середнім рівнем даних нуклеотидів у геномі (рис. 2). Щодо вмісту двох інших нуклеоти-

дів, то рівень цитозину у послідовностях miRNA людини та миші достовірно, хоч і ненабагато, вищий, а урацилу – нижчий за середній для геному рівень.

Як показало дослідження локалізації залишків аденіну, урацилу, гуаніну та цитозину, для miRNA в цілому властивий відносно рівномірний розподіл даних нуклеотидів від 5'- до 3'-кінця послідовності, за винятком різко підвищеної концентрації урацилу та різко зниженої концентрації гуаніну, а також цитозину у положенні першого нуклеотиду з 5'-кінця miRNA (рис. 2).

Така особливість пояснюється тим, що урацил, розташований у даній позиції, входить до складу seed-ділянок типу 7mer-1A та 8mer і комплементарно зв'язується з аденіном на 3'-кінці мішені (див. вище). Однак, починаючи вже з положення другого нуклеотиду з 5'-кінця miRNA, концентрація урацилу стає достовірно нижчою навіть за середній для геному рівень, і дана ситуація залишається незмінною до позиції 7-го нуклеотиду, після якої вміст урацилу підвищується і для більшості наступних у напрямі 3'-кінця положень нуклеотидів достовірно не відрізняється від середнього для геному рівня.

При цьому привертає увагу пік концентрації урацилу в ділянці 9-го нуклеотиду, що є характерним як для послідовностей miRNA людини, так і послідовностей miRNA миші та пацюка, тобто еволюційно консервативним.

Лише 20 % досліджених miRNA (328 послідовностей з 1639) виявилися цілковито позбавленими сайтів 5'-CG-3' та 5'-CNG-3'. Також відзначимо, що у складі сайтів 5'-CG-3' та 5'-CNG-3' знаходиться третина від усіх залишків цитозину, які містяться у послідовностях miRNA.

Таблиця 2

Концентрація сайтів 5'-CG-3', 5'-CNG-3', 5'-GC-3', 5'-GNC-3' у досліджених молекулах miRNA, %

Вид	Концентрація сайтів 5'-C(N)G-3' та 5'-G(N)C-3' у випадковій послідовності miRNA, обчислена за вмістом G і C у miRNA (за табл.1)	Сайт 5'-CG-3'		Сайт 5'-CNG-3'	Сайт 5'-GC-3'		Сайт 5'-GNC-3'
		У середньому в геномі [15]	miRNA		У середньому в геномі [15]	miRNA	
<i>Homo sapiens</i>	6,00 ± 0,33	1,00	2,66 ± 0,26*	5,81 ± 0,38	4,30	6,06 ± 0,38	3,98 ± 0,32*
<i>Mus musculus</i>	5,87 ± 0,36	1,00	2,55 ± 0,28*	5,70 ± 0,42	3,90	5,89 ± 0,42	4,04 ± 0,36*
<i>Rattus norvegicus</i>	5,63 ± 0,47	1,20	2,46 ± 0,36*	5,81 ± 0,57	4,10	5,22 ± 0,52	3,98 ± 0,47*
У середньому	5,88 ± 0,22	–	2,58 ± 0,17*	5,77 ± 0,25	–	5,83 ± 0,25	3,81 ± 0,21*

Примітка. Жирним шрифтом позначено достовірну ($p \leq 0,05$) відмінність концентрації сайту від рівня, середнього для геному (для сайтів 5'-CNG-3' та 5'-GNC-3' див. табл. 1), зірочкою – від рівня у випадковій послідовності miRNA.

Концентрація динуклеотидів 5'-CG-3' у послідовностях miRNA становить $2,58 \pm 0,17$ %. Розглядаючи даний результат, варто зазначити, що протягом еволюції більшість залишків цитозину в сайтах 5'-CG-3' зазнали транзиції через заміну цитозину на тимін внаслідок дезамінування м⁵C. У наш час, наприклад, у геномі людини динуклеотид 5'-CG-3' зустрічається з частотою 1 % (табл. 2) [15, 16], що в більш ніж у 4 рази менше його очікуваної концентрації у випадковій послідовності геномної ДНК (табл. 1). Видові показники концентрації динуклеотидів 5'-CG-3', а також сайтів 5'-CNG-3', 5'-GC-3' та 5'-GNC-3' (див. далі) у miRNA наведено в табл. 2. При цьому, оскільки досліджуваним послідовностям miRNA властивий підвищений вміст гуаніну та цитозину, сайти 5'-C(N)G-3' повинні зустрічатися у їхньому складі приблизно в 1,2–1,36 рази частіше, аніж у геномі (табл. 2). Тим не менше, лише цим не можна пояснити присутність у послідовностях miRNA більшої порівняно з геномом відносної кількості динуклеотидів 5'-CG-3', тому що дана величина перевищує середній геномний рівень у 2–2,6 рази.

Тринуклеотид 5'-CNG-3' у складі досліджених послідовностей miRNA зустрічається із частотою $5,77 \pm 0,25$ %. Видові показники також представлені в табл. 2.

Відзначимо, що у випадковій геномній послідовності ДНК очікувана концентрація тринуклеотиду 5'-CNG-3' повинна бути такою ж, що й динуклеотиду 5'-CG-3' (див. вище), – зокрема, у людини $0,21 \cdot (0,21 + 0,21 + 0,29 + 0,29) \cdot 0,21 = 0,21 \cdot 1 \cdot 0,21 = 0,0441 = 4,41$ %, де 0,21 – частка цитозину, а також гуаніну;

0,29 – частка аденіну, а також тиміну у геномі людини. Однак концентрація сайтів 5'-CNG-3' достовірно не відрізняється від її рівня у випадковій послідовності miRNA.

Таким чином, підтверджується зроблений раніше [12, 13] висновок, що сайти 5'-CG-3' та 5'-CNG-3' у складі miRNA зустрічаються достовірно частіше, ніж повинні б зустрічатися у випадковій геномній послідовності.

Розподіл динуклеотидів 5'-CG-3' у досліджених послідовностях miRNA характеризується наявністю кількох піків концентрації (рис. 3). Зокрема, еволюційно консервативними є піки в ділянках 10-, 16- та 21-го нуклеотидів від 5'-кінця. Окрім них, спостерігаються піки концентрації даних сайтів у ділянках нуклеотидів 2–3, 5–6 та 12–14.

Звернемо увагу на те, що не всі піки розподілу сайтів 5'-CG-3' локалізуються у межах seed-ділянки. Це, на нашу думку, може вказувати на гібридизацію з послідовністю-мішенню при miRNA-залежному метилюванні ДНК усієї молекули miRNA, а не лише якоїсь її частини.

Розподілу тринуклеотидів 5'-CNG-3' у досліджених послідовностях miRNA притаманні еволюційно консервативні піки концентрації, що припадають на ділянки 4-го та 8-го нуклеотидів від 5'-кінця (рис. 4). Також спостерігаються піки концентрації даних сайтів у ділянках нуклеотидів 6, 11–13 та 17–19, які, однак, не у всіх видів організмів, miRNA яких досліджували, достовірно відрізняються від рівня концентрації тринуклеотидів 5'-CNG-3' у випадковій послідовності геному.

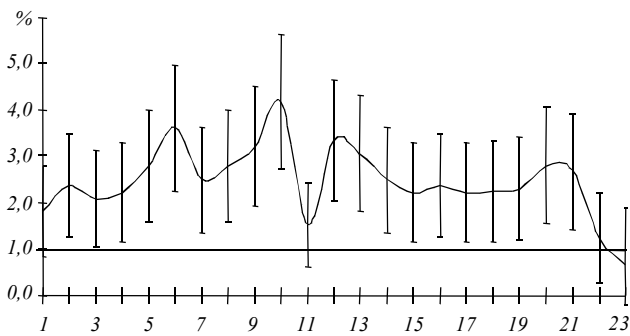


Рис. 3. Розподіл сайтів 5'-CG-3' (хвиляста лінія) у послідовностях miRNA людини ($p \leq 0,05$). Пряма лінія – середня концентрація даного сайту у геномі. По осі абсцис – порядковий номер нуклеотиду у послідовності miRNA

Динуклеотид 5'-GC-3' у складі послідовностей miRNA зустрічається з частотою $5,83 \pm 0,25$ %. Видові показники концентрації зазначеного динуклеотиду наведено в табл. 2. При цьому важливо відмітити, що ці показники достовірно перевищують частоти, з якими даний динуклеотид зустрічався б у випадковій послідовності геному, однак вони достовірно не відрізняються від концентрації сайту 5'-GC-3' у випадковій послідовності miRNA відповідного виду. Таким чином, високий рівень зустрічальності даного динуклеотиду у складі miRNA міг би пояснюватися присутністю більших кількостей гуаніну і цитозину.

Проте картина розподілу динуклеотидів 5'-GC-3' у досліджених послідовностях miRNA є досить своєрідною і характеризується наявністю еволюційно консервативних піків концентрації даних сайтів у позиціях нуклеотидів 3, 5–7, 14–15 та 20 від 5'-кінця (рис. 5). Причому пік, що припадає на нуклеотиди 5–7, є особливо вираженим – концентрація сайтів 5'-GC-3' у його межах у кілька разів перевищує середній рівень у геномі.

Звернемо увагу на те, що даний пік, а також пік у ділянці 3-го нуклеотиду локалізуються в межах seed-ділянки. Це може вказувати на роль динуклеотидів 5'-GC-3', які тут містяться, у забезпеченні більшої термодинамічної стабільності взаємодії seed-ділянки miRNA з послідовністю-мішенню (пари гуанін–цитозин є термодинамічно стійкішими, ніж пари аденін–урацил, оскільки утворені трьома водневими зв'язками, а не двома). Інші піки концентрації даного сайту не є значними, хоча від рівня, властивого випадковій геномній послідовності, до-

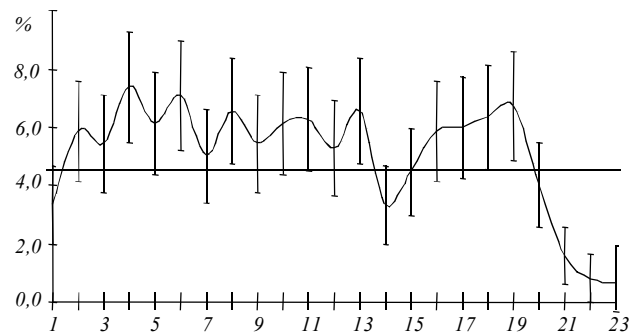


Рис. 4. Розподіл сайтів 5'-CNG-3' (хвиляста лінія) у послідовностях miRNA людини ($p \leq 0,05$). Пряма лінія – середня концентрація даного сайту у геномі. По осі абсцис – порядковий номер нуклеотиду у послідовності miRNA

стовірно відрізняються. Можливо, сайти 5'-GC-3', розташовані тут, беруть участь у забезпеченні більшої термодинамічної стабільності 3'-кінця молекул miRNA, важливої для правильного розплітання дволанцюгового попередника miRNA у комплексі RISC (див. вище).

Тринуклеотид 5'-GNC-3' у складі досліджених послідовностей miRNA зустрічається з частотою $3,81 \pm 0,21$ %. Видові значення також наведено в табл. 2, при цьому варто зазначити, що вони достовірно менші, ніж частоти, з якими даний сайт повинен зустрічатися у випадковій послідовності геному та miRNA відповідного виду. У картині розподілу даних сайтів (рис. 6) немає піків, які б достовірно перевищували середній для геному рівень концентрації тринуклеотиду 5'-GNC-3'.

Наявність у складі miRNA більшої, ніж середня в геномі, концентрації сайтів 5'-CNG-3' (у той час як сайти 5'-GNC-3' зустрічаються на рівні, нижчому від середнього для них у геномі) вказує на те, що з певних причин для послідовностей miRNA важлива не сама по собі присутність гуаніну та цитозину у тому чи іншому тринуклеотиді, а 5'-3'-порядок їхнього поєднання і, отже, сайти 5'-CNG-3' виконують якусь біологічну роль у miRNA.

Концентрація сайтів 5'-GC-3' перебуває на середньому для miRNA рівні (хоча і є певні позиції нуклеотидів, де вона значно вища), тоді як, нагадаємо, концентрація сайтів 5'-CG-3' у кілька разів перевищує середнє для геному значення. Крім того, сайти 5'-CG-3' у послідовностях miRNA зустрічаються частіше, ніж цього варто очікувати, виходячи з притаманного послідовностям miRNA більшо-

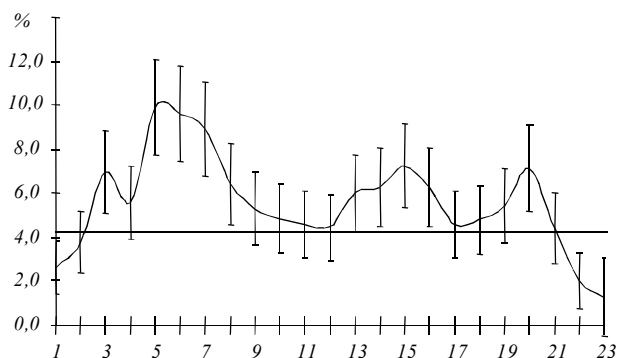


Рис. 5. Розподіл сайтів 5'-GC-3' (хвиляста лінія) у послідовностях miRNA людини ($p \leq 0,05$). Прямая лінія – середня концентрація даного сайту у геномі. По осі абсцис – порядковий номер нуклеотиду у послідовності miRNA

го, аніж загалом для геному, вмісту гуаніну та цитозину. Це означає, що і динуклеотиди 5'-CG-3' у складі miRNA теж відіграють певну важливу біологічну роль. У попередніх роботах висловлено гіпотезу, що сайти 5'-CG-3' та 5'-CNG-3' комплементарні сайтам метилювання у послідовностях ДНК, з якими miRNA гібридизуються у процесі транскрипційного сайленсингу, ініціюючи метилювання ДНК *de novo* [12, 13].

Близько половини miRNA (805 із 1639 послідовностей) містять у своєму складі два або більше сайтів 5'-C(N)G-3'. Відстань між ними переважно становить 1 або 3–9 нуклеотидів. Це означає, що сусідні сайти 5'-C(N)G-3' розташовані в основному у межах одного витка подвійної спіралі (як відомо, дуплекси ДНК–РНК формують А-спіраль, на один виток якої припадає 10,7 п. н.). Цікаво, що таким же виявився розподіл відстаней і між сусідніми сайтами 5'-G(N)C-3'.

Висновки. Послідовності miRNA людини, миші і пацюка характеризуються підвищеним вмістом гуаніну та цитозину і зниженим – аденіну та урацилу у порівнянні з середнім вмістом даних нуклеотидів у геномі. Тим не менше, картина розподілу урацилу характеризується різким піком, що припадає на перший нуклеотид з 5'-кінця miRNA, а також піком у точці розташування 9-го нуклеотиду.

Динуклеотиди 5'-CG-3' і 5'-GC-3' та тринуклеотид 5'-CNG-3' у послідовностях miRNA зустрічаються значно частіше, ніж у випадковій геномній послідовності. При цьому рівень зустрічальності сайтів 5'-CG-3' у кілька разів перевищує випадко-

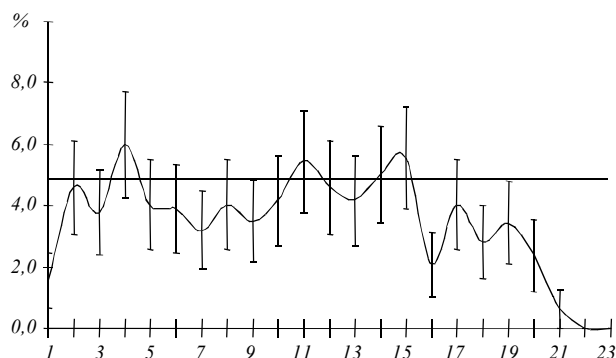


Рис. 6. Розподіл сайтів 5'-GNC-3' (хвиляста лінія) у послідовностях miRNA людини ($p \leq 0,05$). Прямая лінія – середня концентрація даного сайту у геномі. По осі абсцис – порядковий номер нуклеотиду у послідовності miRNA

вий рівень, а також рівень, який варто очікувати, враховуючи присутність у складі miRNA більшого, аніж загалом для геному, вмісту гуаніну та цитозину. Концентрація сайтів 5'-GC-3' та 5'-CNG-3' у miRNA достовірно не відрізняється від рівня, визначеного для випадкової послідовності miRNA.

Локалізація динуклеотидів 5'-CG-3' та тринуклеотидів 5'-CNG-3' у miRNA характеризується кількома піками, у яких рівень даних сайтів у кілька разів перевищує середній для геному. Частина цих піків є еволюційно консервативними. Піки концентрації сайтів 5'-CG-3' та 5'-CNG-3' не зосереджуються в межах seed-ділянки, а розподілені уздовж усієї послідовності miRNA. Це може вказувати на участь у передбачуваному miRNA-залежному метилюванні ДНК усієї послідовності miRNA.

Динуклеотиди 5'-GC-3', на відміну від сайтів 5'-CG-3' та 5'-CNG-3', локалізуються головним чином у межах seed-ділянки miRNA і, ймовірно, відіграють певну роль у забезпеченні стабільності зв'язування seed-ділянки з послідовністю-мішенню.

V. A. Halyskiy, S. V. Komisarenko

5'-CG-3', 5'-CNG-3', 5'-GC-3' and 5'-GNC-3' site localization in microRNA sequences

Palladin Institute of Biochemistry, NAS of Ukraine
9, Leontovycha Str., Kyiv, Ukraine, 01601

Summary

Aim. Discovery of localization pattern of 5'-CG-3' dinucleotide and 5'-CNG-3' trinucleotide in microRNA (miRNA) sequences. **Methods.** Bioinformatic and statistic analysis of mature miRNA sequences. **Results.** One of the highest peaks of 5'-CG-3' dinucleo-

tides concentration is located close to the nucleotide 10 from 5'-end of miRNA. The level of 5'-CG-3' in this peak exceeds more than four times the average genomic frequency of this dinucleotide. Other evolutionary conservative peaks are located around the nucleotides 16 and 21 in miRNA sequence. Peaks of 5'-CNG-3' trinucleotides concentration correspond to the nucleotides 4 and 8. Approximately 50 % of miRNA sequences contain two or more 5'-CG-3' or 5'-CNG-3' sites. The distance between these sites amounts mostly to 1 or 3–9 nucleotides. **Conclusions.** The peaks of 5'-CG-3' and 5'-CNG-3' sites concentration are located not only within seed region, but also along the full length of miRNA sequence. Nevertheless, these sites are disposed in the same miRNA molecule mainly within one turn of double helix. Such site localization pattern may be an evidence of participation of the whole miRNA sequence in miRNA-dependent DNA methylation.

Keywords: microRNA, 5'-CG-3' dinucleotide, 5'-CNG-3' trinucleotide, site pattern.

B. A. Галицкий, С. В. Комисаренко

Локализация сайтов 5'-CG-3', 5'-CNG-3', 5'-GC-3' и 5'-GNC-3' в последовательностях микроРНК

Резюме

Цель. Установление закономерностей в локализации динуклеотидов 5'-CG-3' и тринуклеотидов 5'-CNG-3' в последовательностях микроРНК (miRNA). **Методы.** Биоинформатический и статистический анализ последовательностей зрелых miRNA. **Результаты.** Один из наиболее высоких пиков концентрации сайта 5'-CG-3' локализуется в месте расположения нуклеотида 10 с 5'-конца miRNA. Уровень концентрации сайта 5'-CG-3' в данном пике более чем в четыре раза превышает среднюю концентрацию этого динуклеотида в геноме. Другие эволюционно консервативные пики в последовательности miRNA локализуются в районе нуклеотидов 16 и 21. Пики концентрации сайта 5'-CNG-3' соответствуют нуклеотидам 4 и 8. Приблизительно половина последовательностей miRNA содержит два или больше сайтов 5'-CG-3' или 5'-CNG-3', расстояние между которыми в большинстве случаев составляет 1 или 3–9 нуклеотидов. **Выводы.** Пики концентрации сайтов 5'-CG-3' и 5'-CNG-3' не сосредоточены в пределах seed-участка, а распределены по всей длине последовательности miRNA. Однако в одной и той же молекуле miRNA данные сайты расположены в пределах одного витка двойной спирали. Такая картина распределения упомянутых сайтов может указывать на участие в miRNA-зависимом метилировании ДНК всей последовательности miRNA.

Ключевые слова: микроРНК, динуклеотид 5'-CG-3', тринуклеотид 5'-CNG-3', распределение сайтов.

ПЕРЕЛІК ЛІТЕРАТУРИ

- Rodriguez A., Griffiths-Jones S., Ashurst J. L., Bradley A. Identification of mammalian microRNA host genes and transcription units // *Genome Res.*–2004.–**14**, N 10A.–P. 1902–1910.
- Lee Y., Ahn C., Han J., Choi H., Kim J., Yim J., Lee J., Provost P., Radmark O., Kim S., Kim V. N. The nuclear RNase III Drosha initiates microRNA processing // *Nature.*–2003.–**425**, N 6956.–P. 415–419.
- Lund E., Guttinger S., Calado A., Dahlberg J. E., Kutay U. Nuclear export of microRNA precursors // *Science.*–2004.–**303**, N 5654.–P. 95–98.
- Khvorova A., Reynolds A., Jayasena S. D. Functional siRNAs and miRNAs exhibit strand bias // *Cell.*–2003.–**115**, N 2.–P. 209–216.
- Schwarz D. S., Hutvagner G., Du T., Xu Z., Aronin N., Zamore P. D. Asymmetry in the assembly of the RNAi enzyme complex // *Cell.*–2003.–**115**, N 2.–P. 199–208.
- Ma J. B., Ye K., Patel D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain // *Nature.*–2004.–**429**, N 6989.–P. 318–322.
- Rhoades M. W., Reinhart B. J., Lim L. P., Burge C. B., Bartel B., Bartel D. P. Prediction of plant microRNA targets // *Cell.*–2002.–**110**, N 4.–P. 513–520.
- Lai E. C. miRNAs: whys and wherefores of miRNA-mediated regulation // *Curr. Biol.*–2005.–**15**, N 12.–P. R458–R460.
- Lewis B. P., Burge C. B., Bartel D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets // *Cell.*–2005.–**120**, N 1.–P. 15–20.
- Kim J., Krichevsky A., Grad Y., Hayes G. D., Kosik K. S., Church G. M., Ruvkun G. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons // *Proc. Natl Acad. Sci. USA.*–2004.–**101**, N 1.–P. 360–365.
- Grimson A., Farh K. K., Johnston W. K., Garrett-Engle P., Lim L. P., Bartel D. P. MicroRNA targeting specificity in mammals: determinants beyond seed pairing // *Mol. Cell.*–2007.–**27**, N 1.–P. 91–105.
- Halytskyi V. A. Mechanism of the initiation of DNA methylation *de novo* by small RNA // *Eur. J. Cancer Suppl.*–2007.–**5**, N 4.–P. 75.
- Halytskyi V. A. Hypothesis of initiation of DNA methylation *de novo* and allelic exclusion by small RNAs // *Cell Tissue Biol.*–2008.–**2**, N 2.–P. 97–106.
- Griffiths-Jones S., Grocock R. J., van Dongen S., Bateman A., Enright A. J. miRBase: microRNA sequences, targets and gene nomenclature // *Nucleic Acids Res.*–2006.–**34**, Database issue.–P. D140–D144.
- Jabbari K., Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies // *Gene.*–2004.–**333**.–P. 143–149.
- Scarano E., Iaccarino M., Grippo P., Parisi E. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos // *Proc. Natl Acad. Sci. USA.*–1967.–**57**, N 5.–P. 1394–1400.

UDC 577.214.5

Received 28.04.11