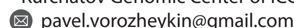
Вавиловский журнал генетики и селекции. 2022;26(8):819-825
DOI 10.18699/VJGB-22-99

Original Russian text https://sites.icgbio.ru/vogis/

The context signals of mitochondrial miRNAs (mitomiRs) of mammals

O.V. Vishnevsky[1], P.S. Vorozheykin[2] ✉, I.I. Titov[1, 2, 3]

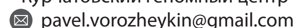
[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

✉ pavel.vorozheykin@gmail.com

Abstract. MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression at the post-transcriptional level in the cytoplasm and play an important role in a wide range of biological processes. Recent studies have found that the miRNA sequences are presented not only in the cytoplasm, but also in the mitochondria. These miRNAs (the so-called mitomiRs) may be the sequences of nuclear or mitochondrial origin; some of them are involved in regulation of the mitochondrial gene functions, while the role of others is still unknown. The identification of nucleotide signals, which are unique to mitomiRs, may help to determine this role. We formed a dataset that combined the experimentally discovered mitomiRs in human, rat and mouse. To isolate signals that may be responsible for the mitomiRs' functions or for their translocation from or into mitochondria a context analysis was carried out for the sequences. For three species in the group mitomiRs/non-mitomiRs and the group of all miRNAs from the miRBase database statistically overrepresented 8-letter motifs were identified (p-value < 0.01 with Bonferroni correction for multiple comparisons), for these motifs the patterns of the localization in functionally important regions for different types of miRNAs were found. Also, for the group mitomiRs/non-mitomiRs we found the statistically significant features of the miRNA nucleotide context near the Dicer and Drosha cleavage sites (Pearson's χ^2 test of independence for the first three positions of the miRNA, p-value < 0.05). The observed nucleotide frequencies may indicate a more homogeneous pri-miRNA cleavage by the Drosha complex during the formation of the 5′ end of mitomiRs. The obtained results can help to determine the role of the nucleotide signals in the origin, processing, and functions of the mitomiRs.
Key words: miRNA; pre-miRNA; mitomiR; mitochondrion.

For citation: Vishnevsky O.V., Vorozheykin P.S., Titov I.I. The context signals of mitochondrial miRNAs (mitomiRs) of mammals. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):819-825. DOI 10.18699/VJGB-22-99

Контекстные сигналы в митохондриальных микроРНК млекопитающих

О.В. Вишневский[1], П.С. Ворожейкин[2] ✉, И.И. Титов[1, 2, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

✉ pavel.vorozheykin@gmail.com

Аннотация. МикроРНК – это малые некодирующие РНК, которые регулируют экспрессию генов на посттранскрипционном уровне в цитоплазме, и, таким образом, играют важную роль в большом числе биологических процессов. Последние исследования обнаружили присутствие последовательностей микроРНК не только в цитоплазме, но и внутри митохондрий. Такие микроРНК (так называемые митомиры, mitomiRs) могут иметь ядерное или митохондриальное происхождение, при этом для некоторых из них установлена роль в регулировании функций митохондриальных генов, а для большинства она пока неизвестна. Выявление нуклеотидных сигналов, уникальных для митомиров, может помочь определить эту роль. В нашей работе составлена выборка экспериментально обнаруженных митомиров человека, мыши и крысы. С целью выделения сигналов, которые могут быть ответственны за функционирование митомиров и за их транспортировку в митохондрии или из них, осуществлен контекстный анализ для полученных последовательностей митомиров. Для трех видов в группе данных митомиры/не-митомиры и в группе всех микроРНК из базы miRBase выявлены статистически перепредставленные 8-буквенные мотивы (уровень значимости $p < 0.01$ с учетом поправки Бонферрони на множественность сравнения). Для этих мотивов обнаружены закономерности их локализации в функционально значимых участках для разных типов микроРНК. Для рас-

© Vishnevsky O.V., Vorozheykin P.S., Titov I.I., 2022
This work is licensed under a Creative Commons Attribution 4.0 License

сматриваемой группы митомиры/не-митомиры также обнаружены статистически значимые особенности нуклеотидного состава последовательностей микроРНК возле границ разрезания комплексами Drosha/Dicer (критерий независимости χ^2 Пирсона для первых трех позиций микроРНК с уровнем значимости $p < 0.05$). Наблюдаемые частоты нуклеотидов, предположительно, могут указывать на наличие у митомиров (в сравнении с не-митомирами) более однородного разрезания прай-миРНК комплексом Drosha при формировании 5′-конца последовательностей. Результаты работы могут быть полезными для выявления сигналов, принимающих участие в возникновении, процессинге и функциях митомиров.
Ключевые слова: микроРНК; пре-миРНК; митомир; митохондрия.

Introduction

The main pathways of miRNA biogenesis, starting at a cell's nucleus and ending in the cytoplasm, have been studied quite well to date (Bartel, 2018). Studying the nucleotide context of microRNAs and their precursors (pri-/pre-miRNAs) established the presence of signals that can affect the functions of miRNAs as well as their maturation at different stages of biogenesis. The nucleotide sequence of a miRNA can both directly determine its functions and affect the 5′-end cleavage accuracy by Drosha/Dicer complexes, thus forming site-specifically modified miRNAs having a shift in the so-called "seed region", a region from 2 to 7 miRNA nucleotides responsible for its addressing (Starega-Roslan et al., 2015a, b; Rolle et al., 2016).

The presence of motifs in the single-stranded ends (UG; CNNC) and in the basal stem of the pri-miRNAs (CUC/GHG) or in the terminal loop (GU) of the pre-miRNA hairpin can lead to blocking or, conversely, to facilitating miRNA processing (Auyeung et al., 2013; Fang, Bartel, 2015; Nguyen et al., 2015; Starega-Roslan et al., 2015a, b; Rolle et al., 2016; Vorozheykin, Titov, 2020). Apart from the nucleus and cytoplasm, these small RNA sequences, as well as the proteins of their processing complexes, are found in organelles, for example, in mitochondria (Kren et al., 2009; Bandiera et al., 2011; Wang et al., 2015). These observations show there are possibly new pathways for miRNA biogenesis inside a mitochondrion as well as ways for transportation of mature miRNAs between the cytoplasm and mitochondria by yet unknown transport complexes. The existence of such mitochondrial miRNAs (so-called mitomiRs) raises questions about their evolutionary origin and their functions inside and outside organelles and whether they have the structural features enabling their functions and transportation inside or outside mitochondria.

This paper is a review of published materials devoted to experimentally observed miRNAs in a mitochondria. For selected mitomiRs, their sequences' contextual features have been evaluated to investigate the possible influence of nucleotide signals on the origin, processing, and functions of the mitomiRs.

Materials and methods

For our review, miRNA sequences of *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* from the miRBase database (http://miRBase.org, edition 22.1) (Kozomara et al., 2019) were selected. The total number of the included sequences comprised 5398.

The information about mitomiRs was obtained from the articles, whose authors experimentally investigated, applying the RT-qPCR, microarray, qRT-PCR methods, microRNA localization inside and outside the mitochondria of different organisms and tissues (Kren et al., 2009; Bian et al., 2010; Bandiera et al., 2011; Barrey et al., 2011; Mercer et al., 2011; Das et al., 2012; Sripada et al., 2012; Wang et al., 2015). Based on these publications, two sets of sequences were formed for human, mouse, and rat: mitomiRs (652 miRNA sequences observed in mitochondria) and all other miRNAs from the miRBase database (4766 sequences) hereinafter called non-mitomiRs).

To study the features of the sequences of these two groups, a search for statistically overrepresented (p-value < 0.01 with Bonferroni correction for multiple comparisons) oligonucleotide motifs was carried out using the ARGO software (Vishnevsky, Kolchanov, 2005) to perform a *de novo* search for motifs in the 15-letter code for mitomiRs/non-mitomiRs sample pairs and for all miRNAs from the miRBase database. When searching for motifs in the microRNAs from the miRBase database, the software estimated the expected proportion of random sequences with a mononucleotide frequency composition similar to that of an analyzed sample containing a motif for random reasons.

For the obtained motifs, an assessment to estimate their similarity within each of the considered groups and between the two groups was performed. For every pair of motifs, the Jaccard similarity coefficient was calculated as $\frac{N_{\text{simil}}}{N_{\text{total}}}$, where N_{simil} is the number of all 4-letter nucleotide sequences corresponding to both motifs. The coefficient takes a value from 0 to 1, where 0 indicates a complete difference between the two motifs, and 1 – a complete match.

To estimate the probability of obtaining the Jaccard coefficient for random reasons, the method proposed in (Real, Vargas, 1996) was applied where the random value of the Jaccard coefficient is assumed to be distributed according to the binomial law (up to normalization). For the identified motifs (found in 616 mitomiRs and 4043 non-mitomiRs), an analysis of their localization in miRNA sequence and an analysis of the nucleotide context were performed to identify the heterogeneity of miRNA cleavage from a precursor by the Drosha and Dicer complexes. The localization analysis of all microRNAs from the miRBase database involved the random positions selected within the sequences of an analyzed sample and used as a "contrast" sample.

Results and discussion

In the reviewed publications, 652 unique miRNA identifiers were mentioned. 272 sequences from the found mitomiRs can be characterized as highly reliable, since they were either ad-

ditionally verified by RT-qPCR/qRT-PCR methods, or in the data of microarray experiments, they were observed in greater numbers inside mitochondria than outside them.

It is worth mentioning seven mitomiRs, whose sequences fully present in the human mitochondrial genome: hsa-miR-1974, hsa-miR-1977, hsa-miR-1978, hsa-miR-4461, hsa-miR-4463, hsa-miR-4484, hsa-miR-4485-3p, and that can serve as an additional confirmation of their validity. At the same time, due to miRNA sequence and mitochondrial tRNA imposition, references to the following miRNAs such as hsa-miR-1974, hsa-miR-1977, hsa-miR-1978 were removed from the miRBase database. The hsa-miR-4461 microRNA was also removed from the database since the experimental data obtained for it did not meet the miRNA-annotation requirements. Hence, the sequences that did not correspond to the currently known miRNA biogenesis pathways but could be formed through unknown non-canonical pathways had been excluded from the miRBase database.

For further investigation and comparison of mitomiR characteristics, a sample of non-mitomiRs of a total of 4766 sequences was also used in the study. It includes all human, mouse, and rat miRNAs from the considered miRBase database, excluding the selected mitomiRs.

Using the ARGO software (Vishnevsky, Kolchanov, 2005), all the two miRNAs groups (mitomiRs/non-mitomiRs group, and all the miRNAs included in the miRBase database) were analyzed. For each of the groups, 40 (Table 1) and 44 (Table 2) 8-nucleotide IUPAC motifs were selected, each having a statistically significant difference in occurrence in miRNA samples in each of the groups ($p < 0.01$ with Bonferroni correction for multiple comparisons). For the motifs within each of the groups, as well as for the motifs from different groups, the Jaccard similarity coefficient average value (averaged over motif pairs excluding zero values) for all three calculations did not exceed 0.02.

For two motifs, KTGCANDK from the mitomiRs/non-mitomiRs group and KTGYABDD from the group of all microRNAs, a maximum Jaccard coefficient of (0.3) and a minimum probability to observe it for random reasons of (0.81) were obtained. These motifs were found in 193 sequences from the mitomiRs/non-mitomiRs group and in 315 sequences in the group of all miRNAs from the miRBase

Table 1. Motifs that have a statistically significant difference in occurrence between the mitomiRs/non-mitomiRs groups

No.	Motif	% of mitomiRs	% of non-mitomiRs	Significance level, *p*	No.	Motif	% of mitomiRs	% of non-mitomiRs	Significance level, *p*
1	CAKTSHAN	8.44	0.87	3.8×10^{-26}	21	MWCMBAVH	9.82	2.86	9.7×10^{-8}
2	KTGCANDK	8.90	1.26	5.6×10^{-21}	22	RKTGYWBH	11.81	3.98	1.2×10^{-7}
3	HASHWSBD	28.53	11.89	5.8×10^{-21}	23	GYHSHBDG	18.10	7.97	1.7×10^{-7}
4	WMAGKGCD	6.29	0.54	1.8×10^{-20}	24	YWMCMTBT	5.21	0.85	3.8×10^{-7}
5	MNTVCANK	13.96	3.40	3.6×10^{-20}	25	KKVAACMH	5.98	1.17	8.9×10^{-7}
6	HRVRNTSH	34.97	17.14	1.4×10^{-18}	26	CTNVRBTS	9.66	3.13	2.1×10^{-5}
7	KBAGGTWG	5.21	0.41	8.5×10^{-17}	27	CTRKNBVW	14.88	6.34	2.4×10^{-5}
8	AGSAVCWY	5.21	0.41	8.7×10^{-17}	28	RCABCMHH	6.13	1.40	5.7×10^{-5}
9	RHASHWSB	20.86	7.93	9.1×10^{-16}	29	YCMYWMMM	6.29	1.48	7.5×10^{-5}
10	RCADTSDH	9.97	2.13	7.9×10^{-15}	30	SAGVAMHN	8.13	2.45	2.0×10^{-4}
11	RSTRRDTT	8.13	1.44	4.5×10^{-14}	31	WKMYCMKA	5.21	1.06	2.1×10^{-4}
12	WMDWSCWB	15.49	5.08	1.2×10^{-13}	32	NMYASDGS	10.43	3.79	3.5×10^{-4}
13	HSVYDGDN	44.02	26.58	1.9×10^{-12}	33	KGARNMCY	5.52	1.22	4.0×10^{-4}
14	WRMACWTB	6.13	0.85	3.5×10^{-12}	34	TSRGWSDG	5.98	1.46	9.8×10^{-4}
15	CCHKBWGD	9.36	2.20	2.3×10^{-11}	35	WCCHBTHS	6.60	1.77	1.3×10^{-3}
16	GBYWVYWG	12.12	3.63	6.5×10^{-11}	36	SAVWSSCW	6.13	1.59	2.9×10^{-3}
17	KGYWNASW	10.74	3.03	4.8×10^{-10}	37	STRHDGTT	5.06	1.11	3.5×10^{-3}
18	CADKGNTD	8.13	1.79	5.8×10^{-10}	38	NGGCWMDS	7.06	2.07	4.1×10^{-3}
19	GWGSTNVY	9.66	2.60	4.2×10^{-9}	39	HCYBRRCT	5.37	1.26	5.3×10^{-3}
20	WSCAKSWR	6.44	1.24	3.5×10^{-8}	40	YSTSRSTS	5.98	1.55	5.8×10^{-3}

Note. % of mitomiRs is a proportion of mitomiR sequences containing the motif; % of non-mitomiRs is a proportion of non-mitomiR sequences containing the motif.
Here and in the Table 2: The table includes only those motifs whose significance level was $p < 0.01$ with Bonferroni correction for multiple comparisons.

Table 2. Motifs having a statistically significant difference in occurrence between the group of all miRBase base miRNAs and random sequences of similar mononucleotide composition as analyzed miRNAs

No.	Motif	% of miRBase	% of random sequences	Significance level, *p*	No.	Motif	% of miRBase	% of random sequences	Significance level, *p*
1	YNCKBYCB	12.09	5.36	1.1×10^{-71}	23	HRHABYRC	5.63	2.99	2.8×10^{-15}
2	BYNCYKYC	11.26	4.94	2.0×10^{-67}	24	GCKSVKBK	6.28	3.46	3.1×10^{-15}
3	AWRYRHWY	6.33	2.10	1.3×10^{-59}	25	KTGYABDD	5.63	3.01	1.0×10^{-14}
4	RHARHRHW	11.79	5.90	1.0×10^{-50}	26	GTWDWHYV	5.15	2.76	9.9×10^{-13}
5	NCKKYCBB	11.09	5.45	1.2×10^{-49}	27	RHBTKTGH	5.94	3.36	1.8×10^{-12}
6	WDYAYDKW	9.21	4.16	2.1×10^{-49}	28	HWYVYAYR	6.11	3.49	3.0×10^{-12}
7	RHAWWYRY	5.06	1.63	4.2×10^{-48}	29	NRRMRSSA	8.40	5.29	4.4×10^{-12}
8	VGGMDVNG	11.92	6.10	4.8×10^{-48}	30	RVKGGMRV	7.88	4.88	5.0×10^{-12}
9	YRTANANV	5.04	1.86	3.9×10^{-37}	31	CBKCYCNV	5.76	3.33	2.2×10^{-10}
10	NHYYVCAG	9.47	4.81	8.4×10^{-37}	32	KCCNBBKC	5.89	3.48	2.0×10^{-9}
11	BTBYCYKY	9.21	4.66	5.7×10^{-36}	33	HATHNYWY	5.70	3.35	3.1×10^{-9}
12	YKHCTYYH	7.97	3.89	2.5×10^{-33}	34	NKGWTDTH	5.06	2.89	9.3×10^{-9}
13	ANBGHWDH	16.11	10.11	4.4×10^{-33}	35	ASDHAVWW	5.37	3.15	2.6×10^{-8}
14	CDGKVNNN	38.28	30.07	7.4×10^{-29}	36	BCDGTKHY	5.30	3.11	5.5×10^{-8}
15	RRMDGNAR	9.63	5.33	5.0×10^{-28}	37	WGDRMHKG	8.99	6.10	1.0×10^{-7}
16	GRGRHDGD	9.10	4.94	6.7×10^{-28}	38	WWWTYRBD	5.41	3.27	7.6×10^{-7}
17	DYAYDGTN	6.02	2.82	6.2×10^{-26}	39	TBTMMYHY	5.30	3.30	5.1×10^{-5}
18	WHAYAHNS	6.24	3.07	2.0×10^{-23}	40	KSRGNBAG	6.31	4.12	6.0×10^{-5}
19	NSDTNTHT	9.10	5.32	1.5×10^{-20}	41	HMCMKYCH	5.44	3.50	6.9×10^{-4}
20	TVYNYVCA	6.39	3.38	1.3×10^{-18}	42	GWSGVDMN	7.88	5.54	1.9×10^{-3}
21	DRYBTKTG	5.35	2.72	1.0×10^{-16}	43	GWGHKBAB	5.08	3.26	4.1×10^{-3}
22	TGBRRWKW	5.70	2.98	1.3×10^{-16}	44	TWVTDWRH	5.19	3.37	9.7×10^{-3}

Note. % of miRBase is a proportion of miRBase sequences containing the motif; % of random sequences is an expected by ARGO a proportion of random sequences with a mononucleotide frequency composition similar to analyzed sample containing a motif for random reasons.

database. Apart these two motifs, the Jaccard coefficient for all other considered pairs did not exceed 0.13 with the probability of observing corresponding coefficients for random reasons being less than 0.001, so the observed data showed a low degree of motif coincidence both within each of the groups and between the two groups.

The observed differences in the nucleotide composition between the samples of mitomiRs and non-mitomiRs can act as specific signals for mitomiR processing, e. g., for recognition and transportation of sequences to/from mitochondria by transport complexes or for the implementation of mitomiR specific functions through direct binding to targets in mitochondrial or cellular DNAs. At the same time, the motifs found in the group of all miRNAs may correspond to the signals common for the processing and functioning of miRNAs, regardless of their localization.

For both considered groups the first motif position tended to be located at the beginning of a miRNA (Fig. 1), so the maximum proportion of sequences was observed for the motifs with their start being at positions 1–3. For the obtained observations, a statistically significant dependence of a miRNA-sequence type on the positions of a motif start (Pearson's χ^2 test of independence, *p*-values 4.46×10^{-2} and 6.58×10^{-5} for the mitomiRs/non-mitomiRs group and the miRbase miRNAs/random positions in miRNAs group, respectively). At the same time, for the mitomiRs, in contrast to the other samples, a significant reduction in the number of miRNAs whose motif start was located at positions 8–10 of a microRNA was observed. A possible reason for that was that the so-called seed region of all miRNAs (both mitomiRs and non-mitomiRs) is the most conservative and significant region in terms of its functionality, and therefore the considered 8-letter conservative motifs often take this region.

In contrast to this, the motifs starting from 8 to 10 position in a miRNA are often localized in the region of the so-called additional seed (~13–16), which is supposedly less conservative

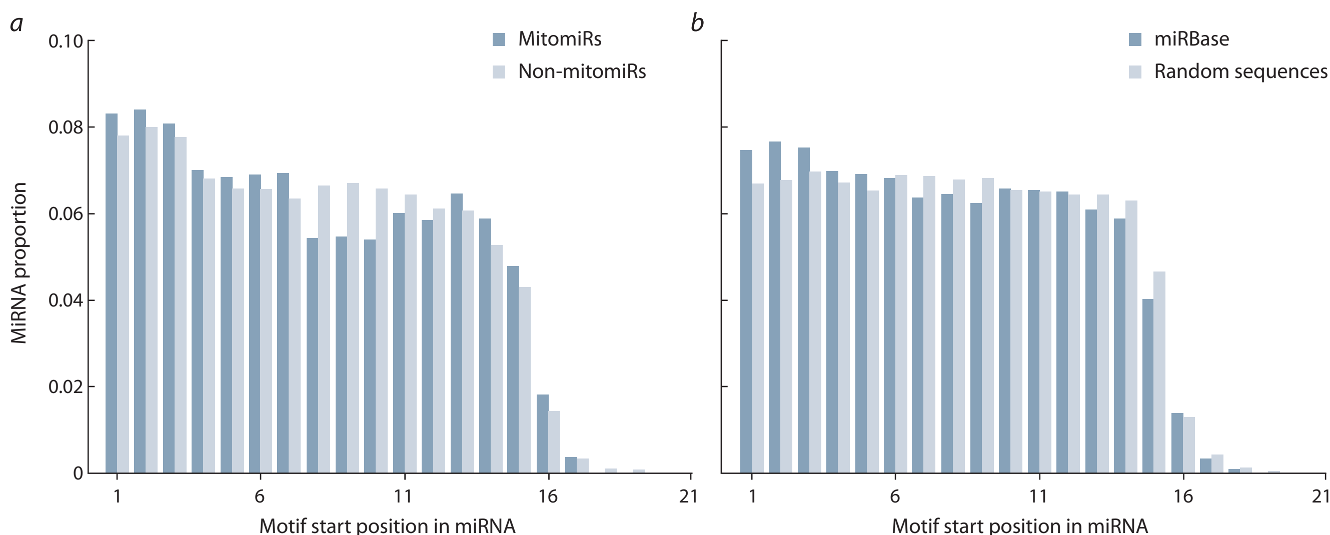


Fig. 1. MiRNAs proportion depending on the motif-start positions found in the sequences of mitomiRs and non-mitomiRs (*a*) and in the miRNAs from the miRBase database (*b*).

For each miRNA with a motif, its position starting from the 5′ end of the sequence was determined. If one motif occurred several times in a microRNA or several motifs occurred once in a microRNA, each occurrence was considered independently and generated a data structure (microRNA, motif position). The graph was normalized for the total number of structures obtained for all motifs. The decrease in observations in the positions whose numbers were greater than 15 was due to the variability in the lengths of miRNA sequences changing from 15 to 28 nucleotides.

in mitomiRs and less often participates in microRNA binding to the target if compared to non-mitomiRs (see Fig. 1, *a*). However, since the sample of all miRNAs mostly consisted of non-mitomiRs, its observation results approximately coincided with those obtained for non-mitomiRs.

For the detected motifs, different localization patterns within a miRNA sequence were observed. One motif could be observed as in several different miRNAs with different localizations within the sequences (e.g., the KTGCANDK motif with a significance level of $p = 5.6 \times 10^{-21}$ started at position 14 from the 5′ end of the hsa-miR-92a-1-5p mitomiR, and at position 2 from the 5′ end of the mmu-miR-19b-3p mitomiR) as within one miRNA, including cases that did not intersect each other (e.g., in the hsa-miR-33a-5p mitomiR, the KTGCANDK motif occurs twice, starting from position 1 and from position 12 from the 5′ end).

Hence, the variability of motif localization in a miRNAs may indicate both the functional importance of these nucleotide signals for these miRNAs and the possible involvement of the signals in microRNA processing, in particular, in the selection and transportation of mitomiR sequences between a mitochondrion and cytoplasm.

For the considered miRNA groups, an increase in the proportion of the sequences where motifs began in positions 1–3 was observed (see Fig. 1), so sequence analysis of this region, but only for those mitomiR and non-mitomiR sequences in which motifs has previously been found, was performed. For the first three positions of the 5′ end of 5p- and 3p-miRNAs, the positional frequencies of nucleotide occurrence were calculated (Fig. 2).

In the mitomiRs from a pre-miRNA's 5p-branch, U was predominantly observed in the first position and G was very rarely found, while A or G were mainly detected in the second position (see Fig. 2, *a*). In non-mitomiRs, an increase in the first position of the number of G and A nucleotides and a decrease in the number of U were observed (see Fig. 2, *b*). For the Drosha cleavage site, an inversion in 2–3 positions between G in non-mitomiRs and A in mitomiRs was detected. For each of the three positions, the nucleotide frequencies showed dependence on a miRNA type (Pearson's χ^2 test of independence, *p*-values 2.89×10^{-31}, 1.03×10^{-28}, and 1.79×10^{-42} for the 1, 2 and 3 positions, respectively), while the third position demonstrated the most significant difference in frequencies between miRNA types, in contrast to the first and second ones.

Comparing the observed nucleotide context of the 5′ ends for mitomiRs/non-mitomiRs with the results of a study that investigated pre-miRNA cleavage accuracy by Drosha and Dicer complexes (Starega-Roslan et al., 2015b), it can be assumed that the Drosha cleavage was more accurate for the mitomiRs from the 5p-branch of pre-miRNAs than for non-mitomiRs, in other words, a more homogeneous 5′ end and a corresponding seed region were formed for mitomiRs, which may be the evidence of greater conservatism of mitomiR functions in comparison to those of non-mitomiRs. The detected signals of mitomiR cleavage homogeneity could be an indication either of the possible existence of a more accurate Drosha-like complex for miRNA processing in mitochondria or of the possible compensation of inaccurate cleavage by the nucleotide composition of the pri-miRNA sequences selected for processing by the Drosha complex.

For the 5p-non-mitomiRs, the context shifted towards heterogeneous cleavage, i.e., more active site-specific miRNA modification, and in this case, the non-mitomiRs could act as a functional-variability factor. It can be assumed that mitochondria do not "tolerate" the variability of "their" miRNAs and eliminated regulatory-sequence isoforms in the course of evolution, so the observed mitomiRs may be the remaining

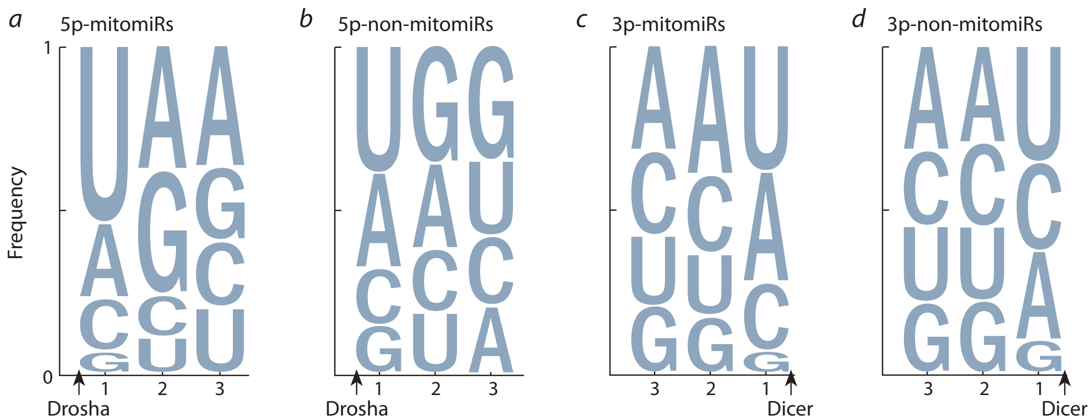


Fig. 2. Nucleotide occurrence frequency for the first three positions starting from the 5′ end of the microRNA from 5p- and 3p-branches of pre-miRNA for mitomiRs (*a*, *c*) and non-mitomiRs (*b*, *d*) samples in the sequences where motifs were found.

The sizes of the letters are proportional to the frequencies. The *X*-axis displays the position numbers in microRNA, starting from the 5′ end. The arrows show the Dicer or Drosha cleavage sites. For positions 1–3 in a 5p-miRNA and for positions 1–2 in a 3p-miRNA presented a statistically significant dependence of miRNA sequence type on nucleotide occurrence frequency for a considered position (Pearson's χ^2 test of independence, *p*-value 2.89×10^{-31}, 1.03×10^{-28}, 1.79×10^{-42}, 1.17×10^{-9}, 3.23×10^{-10} for respective positions).

conservative sequences that originated in the times mitochondrial ancestors were domesticated.

For mitomiRs and non-mitomiRs from the 3p branch of pre-miRNA, no noticeable differences in the nucleotide context were observed, except for the inversion in the first position of the second and third most popular nucleotides (see Fig. 2, *c* and *d*). Statistically significant dependence of positional nucleotide frequencies on a miRNA type was demonstrated only at the first and second positions (Pearson's χ^2 test of independence, *p*-values: $1.17 \times 1.17 \times 10^{-9}$ and 3.23×10^{-10}, respectively). Comparison of the observed nucleotide frequencies against the results obtained by (Starega-Roslan et al., 2015b) did not allow us to make unambiguous conclusions about the cleavage quality of the 5′ end of the 3p-miRNA by the Dicer complex.

Conclusion

In the present study, a sample of experimentally confirmed mitomiRs was formed and a nucleotide analysis of their sequences was performed. For the mitomiRs/non-mitomiRs group and the group of all microRNAs from the miRbase database, statistically overrepresented 8-letter IUPAC motifs within miRNA sequences were found. These motifs demonstrated that mitomiR sequences may represent a new, non-canonical class of miRNAs. While the motifs for the mitomiRs/non-mitomiRs group could act as signals for mitomiR processing, e. g., participating in mitomiR transportation to/from mitochondria or for mitomiR function implementation through binding to targets in mitochondrial or cellular DNAs, the motifs of the group of all microRNAs could correspond to the signals common for the processing and functions of miRNAs, regardless of their localization in the cell.

The nucleotide context of the mitomiRs (if compared to that of the non-mitomiRs) near the 5′ end formed by Drosha/Dicer cleavage could presumably indicate a more uniform formation of the 5′ end of mitomiR sequences and, thus, a more conserved functionality of these sequences.

References

Auyeung V.C., Ulitsky I., McGeary S.E., Bartel D.P. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. 2013;152(4):844-858. DOI 10.1016/j.cell.2013.01.031.

Bandiera S., Rüberg S., Girard M., Cagnard N., Hanein S., Chrétien D., Munnich A., Lyonnet S., Henrion-Caude A. Nuclear outsourcing of RNA interference components to human mitochondria. *PLoS One*. 2011;6(6):e20746. DOI 10.1371/journal.pone.0020746.

Barrey E., Saint-Auret G., Bonnamy B., Damas D., Boyer O., Gidrol X. Pre-microRNA and mature microRNA in human mitochondria. *PLoS One*. 2011;6(5):e20220. DOI 10.1371/journal.pone.0020220.

Bartel D.P. Metazoan microRNAs. *Cell*. 2018;173(1):20-51. DOI 10.1016/j.cell.2018.03.006.

Bian Z., Li L.-M., Tang R., Hou D.-X., Chen X., Zhang C.-Y., Zen K. Identification of mouse liver mitochondria-associated miRNAs and their potential biological functions. *Cell Res*. 2010;20(9):1076-1078. DOI 10.1038/cr.2010.119.

Das S., Ferlito M., Kent O.A., Fox-Talbot K., Wang R., Liu D., Raghavachari N., Yang Y., Wheelan S.J., Murphy E., Steenbergen C. Nuclear miRNA regulates the mitochondrial genome in the heart. *Circ. Res*. 2012;110(12):1596-1603. DOI 10.1161/CIRCRESAHA.112.267732.

Fang W., Bartel D.P. The menu of features that define primary microRNAs and enable *de novo* design of microRNA genes. *Mol. Cell*. 2015;60(1):131-145. DOI 10.1016/j.molcel.2015.08.015.

Kozomara A., Birgaoanu M., Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1):D155-D162. DOI 10.1093/nar/gky1141.

Kren B.T., Wong P.Y.-P., Sarver A., Zhang X., Zeng Y., Steer C.J. MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *RNA Biol*. 2009;6(1):65-72. DOI 10.4161/rna.6.1.7534.

Mercer T.R., Neph S., Dinger M.E., Crawford J., Smith M.A., Shearwood A.-M.J., Haugen E., Bracken C.P., Rackham O., Stamatoyannopoulos J.A., Filipovska A., Mattick J.S. The human mitochondrial transcriptome. *Cell*. 2011;146(4):645-658. DOI 10.1016/j.cell.2011.06.051.

Nguyen T.A., Jo M.H., Choi Y.-G., Park J., Kwon S.C., Hohng S., Kim V.N., Woo J.-S. Functional anatomy of the human microprocessor. *Cell*. 2015;161(6):1374-1387. DOI 10.1016/j.cell.2015.05.010.

Real R., Vargas J. M. The probabilistic basis of Jaccard's index of similarity. *Syst. Biol*. 1996;45(3):380-385. DOI 10.1093/sysbio/45.3.380.

Rolle K., Piwecka M., Belter A., Wawrzyniak D., Jeleniewicz J., Barciszewska M.Z., Barciszewski J. The sequence and structure determine the function of mature human miRNAs. *PLoS One*. 2016; 11(3):e0151246. DOI 10.1371/journal.pone.0151246.

Sripada L., Tomar D., Prajapati P., Singh R., Singh A.K., Singh R. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS One*. 2012;7(9):e44873. DOI 10.1371/journal.pone.0044873.

Starega-Roslan J., Galka-Marciniak P., Krzyzosiak W.J. Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. *Nucleic Acids Res*. 2015a;43(22):10939-10951. DOI 10.1093/nar/gkv968.

Starega-Roslan J., Witkos T., Galka-Marciniak P., Krzyzosiak W. Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. *Int. J. Mol. Sci*. 2015b;16(12):8110-8127. DOI 10.3390/ijms16048110.

Vishnevsky O.V., Kolchanov N.A. ARGO: a web system for the detection of degenerate motifs and large-scale recognition of eukaryotic promoters. *Nucleic Acids Res*. 2005;33(Web Server Iss.):W417-W422. DOI 10.1093/nar/gki459.

Vorozheykin P.S., Titov I.I. Erratum to: How animal miRNAs structure influences their biogenesis. *Russ. J. Genet*. 2020;56(8):1012-1024. DOI 10.1134/S1022795420220019.

Wang W.-X., Visavadiya N.P., Pandya J.D., Nelson P.T., Sullivan P.G., Springer J.E. Mitochondria-associated microRNAs in rat hippocampus following traumatic brain injury. *Exp. Neurol*. 2015;265:84-93. DOI 10.1016/j.expneurol.2014.12.018.

ORCID ID

I.I. Titov orcid.org/0000-0002-2691-3292

Acknowledgements. The OV and IT works were supported by the Russian State Budgetary Project No. FWNR-2022-0020.

The IT work was supported by the Kurchatov Genomic Center of ICG SB RAS (075-15-2019-1662).

Conflict of interest. The authors declare no conflict of interest.

Received September 7, 2022. Revised November 17, 2022. Accepted November 17, 2022.