МРНТИ 34.15.25

Belkozhayev A.M.[1], Niyazova R.E.[2], Ivashchenko A.T.[3]

[1]Al-Farabi Kazakh National University, Kazakhstan, Almaty, e-mail: ayaz_jarkent@mail.ru
[2, 3]Scientific Research Institute of Biology and Biotechnology Problems, Kazakhstan, Almaty

THE PREDICTION OF miRNAS BINDING SITES IN CDS mRNA GENES HAVING TRINUCLEOTIDE REPEATS

In human diseases and physiology, the function of miRNAs is expanding; however, especially nucleotide repeats disorder the majority of miRNA – driven regulatory structure is remaining uncertain. The aim of this work is to reveal which candidate genes of nucleotide repeat diseases and in which degrees can interact with miRNA. We present results on the interaction of 2567 miRNAs with mRNA 102 candidate genes of having nucleotide repeats using the MirTarget program. miRNAs binding sites in the CDS mRNAs of 36 genes from 102 candidate genes with nucleotide repeats have been shown. Among miRNAs that bind with high energy to mRNA genes with nucleotide repeats, we choose five miRNAs that have binding sites for two or more genes: miR-3656 (ARX, EP400, HTT, NCOR2); miR-3960 (ARX, CACNA1I, HTT); miR-1322 (ATN, EP400, GIGYF2, HTT, NCOR2); miR-1281 (CACNA1I, HRC, HTT); miR-4279 (CACNA1I, NCOR2). It was determined that considering miRNAs binding sites are located mainly in regions with CAG, GCG, GAG repeats. Neurological disorders are known to be caused by an increased number of CAG, GCG, GAG repeats, typically in coding regions of otherwise unrelated proteins. Better understanding of interaction specificity of miRNAs and genes promises to offer further in sights into the pathogenic pathways of trinucleotide repeats expansion disorders.

Key words: miRNA, mRNA, coding sequence, binding site, trinucleotide repeat.

[1]Белкожаев А.М., [2] Ниязова Р.Е., [3]Иващенко А.Т.

[1]әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, г. Алматы, e-mail: ayaz_jarkent@mail.ru
[2,3]Биология және биотехнология мәселелері ғылыми-зерттеу институты, Қазақстан, г. Алматы

Тринуклеотидтік қайталымдары бар гендердің mRNA-МЕН miRNA-ның CDS-те байланысуының болжамы

Адамның ауруларында және физиологиясында miRNA – ның фунуциясы кеңейе түсуде, дегенмен әсіресе нуклеотидтік қайталымдардың бұзылысы miRNA негізіндегі реттеуші құрылымның көпшілігі анықталмаған болып қалады. Нуклеотидті қайталану бұзылыстарының кеңеюі емделмейтін және ақырында өлімге әкеліп соғатын, басым көпшілігі тұқым қуалайтын неврологиялық аурулардың тобын құрайды. Сондықтан нуклеотидті қайталанатын аурулардың кандидат гендерін және қандай деңгейде miRNA-мен өзара әрекеттесе алатындығын анықтау қажет. Осыған байланысты MirTarget бағдарламасының көмегімен нуклеотидті қайталанатын 102 кандидатты гендердің mRNA- мен 2567 miRNA- дың өзара әрекеттесу нәтижелері көрсетілді. Нуклеотидтік қайталымдары бар 102 кандидат гендердің 36 гендерінің mRNA – мен miRNA – ның CDS – те байланысуы анықталды. Нуклеотидтік қайталымдары бар гендердің mRNA-мен жоғарғы энергияда байланысатын 5 miRNA – лар таңдап алынды: miR-3656 (ARX, EP400, HTT, NCOR2); miR-3960 (ARX, CACNA1I, HTT); miR-1322 (ATN, EP400, GIGYF2, HTT, NCOR2); miR-1281 (CACNA1I, HRC, HTT); miR-4279 (CACNA1I, NCOR2). miRNA – лардың байланысатын аймақтары негізінен CAG, GCG, GAG қайталанатын аумақтарда орналасқаны анықталды. Неврологиялық бұзылыстар CAG, GCG, GAG қайталану санының артуымен, әдетте кодтау аймақтарында басқа да байланыссыз белоктардың пайда болуымен байланысты. miRNA – дың және гендердің өзара әрекеттесу ерекшеліктерін жақсы түсіну тринуклеотидтік қайталанатын бұзылулардың патогендік жолдарымен танысуға мүмкіндік береді.

Түйін сөздер: miRNA, mRNA, кодталатын тізбек, байланысатын сайт, тринуклеотидті қайталым.

[1]Белкожаев А.М., [2] Ниязова Р.Е., [3]Иващенко А.Т.
[1]Казахский национальный университет имени аль-Фараби, Казахстан, Алматы қ. , e-mail: ayaz_jarkent@mail.ru
[2, 3]Научно-исследовательский институт проблем биологии и биотехнологии, Казахстан, Алматы қ.

Предсказывание сайтов связывания miRNA в CDS mRNA генов имеющих тринуклеотидные повторы

В изучении болезней и физиологии человека функция miRNA расширяется, однако роль miRNA при нарушениях, связанных с нуклеотидными повторами, остается неопределенной. Нарушения нуклеотидных повторов представляют собой группу доминантно наследуемых неврологических заболеваний, которые неизлечимы и в конечном итоге приводят к летальному исходу. Следовательно, необходимо выявить, какие гены-кандидаты заболеваний, связанных с нуклеотидными повторами и в какой степени могут взаимодействовать с miRNA. В связи с этим, мы представляем результаты по предполагаемым взаимодействиям 2567 miRNA с mRNA 102 генов-кандидатов, имеющих нуклеотидные повторы, полученные с использованием программы MirTarget. Показаны сайты связывания miRNA в mRNA CDS 36 генов из 102 генов-кандидатов с нуклеотидными повторами. Среди miRNA, связывающихся с высокой энергией взаимодействия с генами, имеющих нуклеотидные повторы, мы выбрали пять miRNA, которые имеют сайты связывания для двух или более генов: miR-3656 (ARX, EP400, HTT, NCOR2); miR-3960 (ARX, CACNA1I, HTT); miR-1322 (ATN, EP400, GIGYF2, HTT, NCOR2); miR-1281 (CACNA1I, HRC, HTT); miR-4279 (CACNA1I, NCOR2). Установлено, что рассматриваемые сайты связывания miRNA расположены в основном в областях, имеющих повторы CAG, GCG, GAG. Известно, что неврологические расстройства вызваны повышенным количеством повторов CAG, GCG, GAG, обычно в кодирующих областях других неродственных белков. Лучшее понимание специфичности взаимодействия miRNA и генов обещает более детально рассмотреть патогенные пути нарушений экспансии тринуклеотидных повторов.

Ключевые слова: miRNA, mRNA, кодирующая последовательность, сайт связывания, тринуклеотидный повтор.

Abbreviations

mRNA – messenger ribonucleic acids; CDS – coding sequence; UTR – untranslated region; TRED – triplet repeat expansion diseases; FXTAS – fragile X-associated tremor/ataxia syndrome; SCA – several spinocerebellar ataxias; FRDA – Friedreich's ataxia; DM1 – dystrophy Myotonic 1; HD – Huntington disease

1. Introduction

About half of the human genome is composed of repeated sequences of various types; of these, short tandem repeats, such as trinucleotide repeats, represent a substantial portion [1]. Of these, specific trinucleotide repeats located in non-coding and coding regions of individual genes implicated in these disorders are strongly over-represented [2]. Short tandem repeats act as a trigger in over 20 neurodegenerative and neuromuscular human disorders collectively known as triplet repeat expansion diseases (TREDs). These disorders include fragile X-associated tremor/ataxia syndrome (FXTAS), several spinocerebellar ataxias (SCA), Huntington disease (HD), dentatorubral-pallidoluysian atrophy *(DRPLA)*, Friedreich's ataxia (FRDA) and dystrophy Myotonic 1 (DM1) [3, 4].

FXTAS is a neurodegenerative disorder characterized by progressive intention tremor (parkinsonism), gait ataxia and cognitive decline. An expansion of CGG trinucleotide repeats in the 5′-UTR of *FMR1* causes different neuropathological conditions based on the number of CGG repeats [5,6].

Specific links between miRNA (mRNA-inhibitory RNA) regulation and SCAs have been described in several studies. Specifically, the research on SCA1 has revealed that some miRNAs can regulate the level of *ATXN1* protein. It has been reported that miR-19a, miR-101, and miR-130a may bind to the 3′UTR of *ATXN1* and suppress the translation of *ATXN1* [7,8]. SCA1 is a dominantly inherited and fatal neurodegenerative disease resulting from an over expansion of CAG repeats within the *Ataxin-1 (ATXN1)* gene [9].

HD is a dominantly inherited progressive neurodegenerative disorder that results from a mutation that expands the polymorphic trinucleotide (CAG) tract in *HTT*. The average control CAG tract size in the general population is 17–20 repeats. However, in HD patients, one of the two copies of the gene has a CAG tract that has expanded to 36 repeats or more [10]. All these diseases have a predominantly hereditary component and are characterized by the expansion of existing

trinucleotide repeats upon transmission from parent to off-spring [11].

The profile of miRNAs in CAG trinucleotide repeat disorders is scarcely described. However, miRNA dysregulation has been identified in these diseases, and miRNA-related interference with gene expression is considered to be involved in their pathogenesis. A better understanding of microRNAs functions and means of manipulation promises to offer further insights into the pathogenic pathways of CAG repeat expansion disorders, to point out new potential targets for drug intervention and to provide some of the much needed etiopathogenic therapeutic agents [12,13].

Recent studies have demonstrated that interest in miRNA biogenesis and function is growing rapidly and that deep-sequencing technologies in combination with various micro array analyses have become routinely used in the investigation of short RNAs, including studies on miRNA deregulation in trinucleotide repeat expansion disorders [14].

Regulatory miRNAs play a fundamental role in the majority of biological processes, such as cell development, proliferation, differentiation, apoptosis and signal transduction. However, the biological function of most miRNAs remains to be uncovered [15]. Both the altered miRNA expression and the deregulation of genes controlled by miRNAs have been linked to many disorders, such as cancer and cardiovascular, metabolic, and neurological disorders, including trinucleotide repeat expansion disorders [16].

Neurons exhibit considerable sensitivity to mutations in huntingtin containing an elongated polyQ chain. Global changes in gene expression induced by the presence of the mutations in huntingtin as well as the widespread changes in neuronal miRNAs levels have been described in HD patients [17]. Therefore, miRNA deregulation has been recognized as a hallmark of HD and other polyQ diseases. These candidate genes may be targets for miRNA (mRNA-inhibiting RNA) that regulate their expression. It is impossible to experimentally identify how a well-known miRNA can interact with more than 20,000 genes and their isoforms. Therefore, it is required with the help of computational technologies to predict the target genes of certain miRNAs and then to test them experimentally. The purpose of this work is to establish the characteristics of the interaction of miRNAs included in the miRBase with mRNA of genes having nucleotide repeats in the coding sequence (CDS).

2. Materials and Methods

The nucleotide sequences of mRNAs of 102 human genes were taken from the GenBank database (http://www.ncbi.nlm.nih.gov), and 2567 miRNAs were taken from miRBase (http://mirbase.org). The search for target genes for miRNA was performed using the MirTarget program [18]. This program defines the features of binding: a) the localization of miRNA binding sites in the 5'UTRs, the CDSs and the 3'UTRs of the mRNAs; b) the free energy of hybridization (ΔG, kJ/mole). The ratio ΔG/ΔGm (%) was determined for each site (Δ Gm equals the free energy of miRNA binding with its perfect complementary nucleotide sequence). For analyzing and formatting sequences of genes, we used the sequence manipulation suite program (https://bioinformatics.org/sms). The miRNA binding sites located on the mRNAs had ΔG/ΔGm ratios of 85% and more. The MirTarget program takes account of hydrogen bonds between adenine (A) and uracil (U), guanine (G) and cytosine (C), G and U, and A and C. The distances between A and C were equal to those between G and C, A and U, and G and U [19]. To prediction the secondary structure of RNA, the software RNA fold was used (http://rna.tbi.univie.ac.at).

3. Results and Discussion

The search of 2567 human miRNAs binding with 102 mRNAs of human genes having nucleotide repeats in the coding region (CDS) using the MirTarget program has been completed. miRNAs binding sites in the CDS mRNAs of 36 genes from 102 candidate genes with nucleotide repeats have been shown. Out of human miRNAs, only 1023 miRNAs interacted with mRNAs of 36 genes. Based on the results, 3189 miRNAs binding sites are predicted in the CDS with ΔG/ΔGm values equal to 85 % and more.

Of the 102 genes with nucleotide repeats, we select eight genes that are targets of three or more miRNAs: *ARX* (miR-3656, miR-3960, miR-4258, miR-4274); *ATN1* (miR-1322, miR-4271, miR-4281, miR-7150); *CACNA1I* (miR-1268a, miR-1281, miR-3960, miR-4279, miR-6784-5p); *EP400* (miR-1322, miR-1587, miR-3656, miR-4463, miR-4478, miR-4505, miR-4651, miR-6763-5p, miR-8071); *GIGYF2* (miR-1322, miR-483-3p, miR-6894-3p); *HRC* (miR-1281, miR-6891-3p, miR-877-3p); *HTT* (miR-1281, miR-1322, miR-3656, miR-3665, miR-3960, miR-7704); *NCOR2* (miR-1322, miR-3656,

miR-4279, miR-4297, miR-4481, miR-4486, miR-4516). Data given in table 1 show the characteristics of miRNA binding with mRNA of target genes in the CDS.

According to the table miRNA binding sites in mRNA of *ARX, ATN1, CACNA1I, EP400, GIGYF2, HRC*, *HTT, NCOR2* genes are located in the CDS. One or two of the miRNAs bind with each gene in repeat nucleotide sequences: miR-3960 and miR-4258 with *ARX* (GCC)/(CGG), miR-1322 with *ATN1* (CAG), miR-1281 with *CACNA1I* (AGG), miR-1322 with *EP400* (CAG) and *GIGYF2* (CAG), miR-1281 with *HRC* (AGG), miR-1322 and miR-3665 with *HTT* (CAG)/(CGC) and *NCOR2* (CAG).

Table 1 – Characteristics of miRNAs binding sites in CDS mRNA genes with nucleotide repeats

Gene	Characteristics of binding sites
ARX	miR-3656,125,85,-89,17; miR-3960, 1511, 91, -114, 20; miR-4258, 513, 85, -87, 17; miR-4274, 925,85, -89, 18
ATN1	miR-1322, 1693, 87, -89, 19; miR-4271, 1461, 85, -91, 19; miR-4281, 1146, 86, -95, 18; miR-7150, 1139, 85, -87, 18
CACNA1I	miR-1268a, 91, 86, -93, 18; miR-1281, 1820, 86, -84, 17; miR-3960, 6179, 96, -121, 20; miR-4279, 5321, 90, -82, 16; miR-6784-5p, 1436, 87, -101, 20
EP400	miR-1322, 8313, 85, -87, 19; miR-1587, 4736, 85, -97, 20; miR-3656, 252, 89, -93, 17; miR-4463, 4727, 87, -87, 17; miR-4478, 438, 86, -82, 17; miR-4505, 4748, 87, -91, 18; miR-4651, 240, 85, -101, 20; miR-6763-5p, 426, 86, -95, 19; miR-8071, 237, 85, -97, 20
GIGYF2	miR-1322, 3383, 87, -89, 19; miR-483-3p, 813, 87, -99, 21; miR-6894-3p, 3352, 89, -104, 21
HRC	miR-1281, 768, 87, -85, 17; miR-6891-3p, 762, 85, -97, 21; miR-877-3p, 777, 85, -109, 21
HTT	miR-1281, 8052, 86, -84, 17; miR-1322, 197, 87, -89, 19; miR-3656, 344, 87, -91, 17; miR-3665, 267, 86, -93, 18; miR-3960, 268, 92, -115, 20; miR-7704, 264, 88, -101, 19
NCOR2	miR-1322, 1813, 87, -89, 19; miR-3656, 4744, 85, -89, 17; miR-4279, 1240, 88, -80, 16; miR-4297, 347, 90, -78; miR-4481, 442, 86, -82, 17; miR-4486, 5854, 87, -87, 17; miR-4516, 3496, 87, -87, 17

Note: miRNA; the beginning of binding site; the ΔG/ΔGm (%); the free energy change (ΔG, kJ/mole); length of miRNA (nt)

From this results in miR-3960 and miR-4258 bind with $(GCC)_8$/$(GCT)_1$ and $(CGG)_{11}$/$(CAG)_2$ repeats of *ARX* gene. The binding sites of miR-3960 in mRNA of the *ARX* gene are located between 1505 and 1532 nucleotide sequences with a start in 1511 nt. The miR-4258 binds in the region with $(CGG)_{11}$/$(CAG)_2$ repeats.

mir-1322 and mir-3665 bind with CDS mRNA of the *HTT* gene. mir-1322 binding sites are located in a region with $(CAG)_{20}$ and $(CAA)_1$ repeats between 197 – 259 nt. mir-3665 binds in the region with $(CGC)_7$ repeat located from 267 to 288 nt. Obtained data indicate that *ATN1, EP400, GIGYF2, HTT, NCOR2* genes are targets for miR-1322, and *ARX, HTT,CACNA1I* and *HRC* genes are targets for miR-3656, miR-3960 and miR-1281.

Among miRNAs that bind with high energy to mRNA genes with nucleotide repeats, we choose five miRNAs that have binding sites for two or more genes: miR-3656 (*ARX, EP400, HTT, NCOR2*); miR-3960 (*ARX, CACNA1I, HTT*); miR-1322 (*ATN, EP400, GIGYF2, HTT, NCOR2*); miR-1281 (*CACNA1I, HRC, HTT*); miR-4279 (*CACNA1I, NCOR2*). As can be seen from the above data, miR-3656, miR-1322 can interact with a lot of genes compared to miR-4279.

Depicted below table 2 shows the compliance of miRNAs binding in the nucleotide repeats located regions in CDS mRNA genes. It was determined that considering miRNAs binding sites are located mainly in regions with CAG, GCG, GAG repeats. Oligonucleotides of binding sites located in CDSs can encode polyglutamine and polyalanine depending on the open reading frame.

Expansions of CAG trinucleotide repeats (CAG repeats) in coding regions of human genes cause neurodegenerative disorders by generating proteins with elongated polyglutamine (polyQ) stretches. This group of disorders includes Huntington's disease, dentatorubral – pallidoluysian atrophy, spinal bulbar muscular atrophy, and the spinocerebellar ataxia types 1, 2, 3, 6 and 7 [20, 21]. For instance, the HD (*IT15*) gene, which encodes huntingtin, a 350 kDa protein of unknown function, is located on the human chromosome 4 and consists of 67 exons. The disease – causing mutation is a

CAG repeat expansion located within exon 1 of the *HD* gene (*HD* exon 1). The CAG repeat is translated into a polyQ stretch. The disease manifests itself when the polyQ stretch exceeds the critical length of 37 glutamines (pathological threshold), whereas 8–35 glutamine residues in huntingtin are tolerated by neuronal cells [22, 23].

mRNAs genes of *ATN1, ATXN1, ATXN2, BRD4 CELF3, EP400, FOXP2, GIGYF2, HTT, MAML3, MN1, MEF2A, NCOR2, SMARCA2, TNRC6B, TOX3, TNRC6A, RUNX2, ZNF384* and *MLL2* interact with miR-1322. It was previously shown that miR-1322 had arranged binding sites in the CDSs of orthologous *MAMLD1*, *MAML2* and *MAML3* genes. Binding sites encode a polyglutamine oligopeptide ranging from six to 47 amino acids in length. Data indicated the importance of conserved nucleotide sequences of miR-1322 binding sites and not only the amino acid sequence corresponding to oligopeptides of the encoded protein [24]. miR-1281 binds in mRNA of *CACNA1I, PVRL1* and *EGR1,* genes. Moreover, mRNAs genes of *ARX* and *PHOX2B* interact with miR-4258 (Table 2).

Table 2 – Accordance of binding sites to nucleotide repeats of genes

miRNA	Gene	Position	Nucleotidesequence	Codon number
miR-1322	*ATN1*	1693	QQQQQQQQQQQQQQQQ	$CAG_{14}CAA_{2}$
	ATXN1	1578	QQQQQQQQQQQQHQHQQQQQQQQQQQQQQQ	$CAG_{27}CAT_{2}$
	ATXN2	658	QQQQQQQQQQQQQQQQQQQQQQQQ	$CAG_{22}CAA_{1}$
	BRD4	2533	PQQPPPPPQQPPPPPPPQQQQQPPPPPPPP	$CAG_{9}CCG_{21}$
	CELF3	1893	QQQQQQQQQQQQQQQ	$CAG_{12}CAA_{3}$
	EP400	8313	QQQQQQQQQQQQQQQQQ	$CAG_{15}CAA_{2}$
	FOXP2	912	QQQQQQQQQQQQQQQQQQ	$CAG_{13}CAA_{5}$
	GIGYF2	3068	QQQQQQQQQ	$CAG_{7}CAA_{2}$
	HTT	197	QQQQQQQQQQQQQQQQQQQQ	$CAG_{20}CAA_{1}$
	MAML3	2220	QQQQQQQQQQQQQQQQQQQQQQQQQQQQQK	CAG_{24} $CAA_{3}AAA_{1}$
	MN1	2534	QQQQQQQQQQQQQQQQQQQQQQQQQQQQ	CAG_{23} CAA_{5}
	MEF2A	1852	QQQQQQQQQQQQQPPP	$CAG_{13}CCG_{3}$
	NCOR2	1813	QQQQQQQQQQQQQQQQQ	$CAA_{2}CAG_{15}$
	SMARCA2	766	QQQQQQQQQQQQQQQQQQQQQQ	$CAG_{18}CAA_{4}$
	TNRC6B	4172	QQQQQQQQQMM	$CAG_{9}ATG_{2}$
	TOX3	1509	QQQQQQQQQQQQQQQ	$CAG_{10}CAA_{5}$
	TNRC6A	400	QQQQQQQQQQQQQQQQ	$CAG_{14}CCA_{2}$
	RUNX2	512	QQQQQQQQQQQQQQQQQQ	$CAG_{14}CAA_{4}$
	ZNF384	1780	QQQQQQQQQQQQQQQQQ	$CAG_{15}CCA_{2}$
	MLL2	9820	QQQQQQQQ	$CAG_{7}CAA_{1}$

Continuation of table 2

miRNA	Gene	Position	Nucleotidesequence	Codon number
miR-1281	*CACNA1I*	1820	RRRRRR	AGG_6
	PVRL1	1482	RRRRRRRR	AGG_8
	EGR1	474	RRRRRGGWRQQQQQQQQQQQQQQQQ	$CGG_5GGG_2TGG_1AGG_1CAG_{14}CAA_2$
miR-4258	*ARX*	513	RRRRRRRRRRRQQ	$CGG_{12}CAG_2$
	PHOX2B	1100	QQQQQQRRRRRRRRRRP	$CAG_6CGG_{11}CCG1$
miR-1273f	*ANK3*	12230	TTTTTTATTTTTTTTTTT	$ACT_4GCC_1ACC_{14}$
miR-1181	*ATXN7*	593	RRAAAAAA	CGC_2GCG_6
miR-6833-5p	*DIP2B*	837	FFLLLII	$TTC_2CTC_3ATC_2$
miR-3960	*HOXA13*	405	AAAAAAAAAAAAASSS	$GCA_1GCC_8GCG_2GCT_2TCG_2TCC_1$
miR-7110	*IRS1*	3654	LLPPPHH	$CTC_2CCC_3CAC_2$
miR-1910-5p	*MAPK1*	254	GGAAAAAAAMMPP	$GGC_2GCG_7ATG_2CCG_2$
miR-877-3p	*HRC*	777	RRRRRRR	AGG_8
miR-1260a	*PABPN1*	1284	WRRRRRRQQQ	$TGG_1CGG_6CAG_3$
miR-1908-3p	*ZIC2*	1787	RRRRRRRRGG	CGG_8GGC_1GGG1
miR-574-5p	*NGFR*	186	LLLLLL	CTG_4CTT_2
miR-8083	*SOX3*	1047	AAAAAAQQ	$GCG_3GCA_2GCU_1CAG_2$

From indicated in Table 2 genes, especially *ATN1, ATXN1, ATXN2* plays an important role in diseases associated with trinucleotide repeats. In addition, *EGR1* has thus been revealed as a major mediator and regulator of synaptic plasticity and neuronal activity in both physiological and pathological conditions [25]. The obtained results indicate that the mRNA genes of *ATN1, ATXN1, ATXN2* interacted with miR-1322 with CAG and *EGR1* binds with miR-1281 escorted by GGC AGC repeats. As above we said that in the overwhelming majority of cases, neurological diseases are caused by nucleotide repeats.

As can be seen in Table 2, the binding sites of miR-1322 in mRNA of *ATN1, ATXN1, ATXN2, EP400, HTT, MAML3, MN1, NCOR2, SMARCA2, TNRC6A, RUNX2* and *ZNF384* genes are located in regions with CAG repeats (14-27 times in coding sequence). The $(CAG)_{14}$ and $(CAA)_2$ repeats located between 1687 and 1744 nucleotidesin *ATN1* gene and miR-1322 bind in this region with a start at 1693 nt. There are a lot of repetitions of $(CAG)_{27}$, $(CAT)_2$ codons in *ATXN1* gene and miR-1322 binds in regions with these repeats with a start in 1578 nt. mRNA of *MAML3* gene has $(CAG)_{24}$, $(CAA)_3$, $(AAA)_1$ repeats between 2199 to 2282 nt and miR-1322 binding sites located from 2220 nt. In mRNA of *MN1* and *SMARCA2* genes codons $(CAG)_{23,18}$, $(CAA)_5$ repeated from 2519 to 2605 and 745-813, respectively, and miR-1322 binds in regions with these repeats from 2534 and 766 nt.

mRNAs of *ANK3, ATN1, ATXN1, ATXN2, ATXN7, BRD4, CELF3, FOXP2, MAML3, GIGYF2, DIP2B, EGR1, HOXA13, IRS1, HTT, EP400, MAPK1, HRC, RUNX2, PABPN1, ZIC2, MN1, TNRC6B, PVRL1, MLL2, MEF2A, ZNF384, TOX3, NGFR, TNRC6A, SOX3, SMARCA2, PHOX2B, ARX, CACNA1I* and *NCOR2* genes have miRNA binding sites with ΔG/ΔGm value more than 85% (Table 3**).** From this results, obviously visible that miR-1322 binding sites mainly located in regions with CAG repeats of mRNA of *ATN1, ATXN1, ATXN2, BRD4, CELF3, EP400,FOXP2, GIGYF2, HTT, MAML3, MN1,MEF2A, NCOR2, SMARCA2, TNRC6B, TOX3, TNRC6A, RUNX2, ZNF384, MLL2* and *PHOX2B* genes.

Table 3 – Schemes of miRNA binding with mRNA genes having nucleotide repeats

ANK3; miR-1273f;12230;CDS;-89;85;19 5′ – CACUGCC**ACCACCACCACC** – 3′ |||||| ||| |||| || 3′ – GUGACGUUGGAGGUAGAGG – 5′	***MAML3;*** miR-1322;2220;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
ATN1; miR-1322;1693;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***MAPK1;*** miR-1910-5p;254;CDS;-104;85;21 5′ – CG**GCGGCGGGCGCGGGC**CCGG – 3′ |||||||||||||| |||| 3′ – UCCGCCGUCCGUGUCCUGACC – 5′
ATXN1; miR-1322;1569;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***MN1;*** miR-1322;2534;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
ATXN2; miR-1322;658;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	**MLL2;** miR-1322;9820;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAACAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
ATXN7; miR-1181;593;CDS;-112;88;21 5′ – CGCCGCG**CGGCGGCGGCGGCGG** – 3′ || | || ||||||||||||| 3′ – GCCGAGC-CCACCGCCGCUGCC – 5′	***MEF2A;*** miR-1322;1852;CDS;-91;89;19 5′ – **CAGCAGCAGCAGCAGCAG**CC – 3′ ||||| ||||||| || || 3′ – GUCGUAGUCGUCGUAGU-AG- 5′
ARX; miR-4258;513;CDS; -87;85;17 5′ – CAG**CGGCGGCGGCGGCGG** – 3′ | | ||||||||||| || 3′ – GGUUCCGCCACCGCC-CC – 5′	***NGFR;*** miR-4302;667;CDS;-87;85;18 5′ – CACGCUGGG**CCG**ACG**CCG**A – 3′ | ||||||||| ||||||| 3′ – GAGCGACUCGG-UGUGACC – 5′
BRD4; miR-1322;2533; CDS;-87;85;19 5′ – **CAGCAG**CAA**CAGCAG**CCGCC – 3′ ||||| |||||||| ||||| 3′ – GUCGUAGUCGUCGU-AGUAG – 5′	***NCOR2;*** miR-1322; 1813; CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
CACNA1I; miR-1281; 1820;CDS;-84;86;17 5′ – **AGGAGGAGGAGGAGGAGG** – 3′ ||||| |||||||| || 3′ – CCCUC-UCCUCCUCCGCU – 5′	***PABPN1;*** miR-1260a;1284;CDS;-84;85;18 5′ – UG**GCGGCGGCGGCGGCG**GC – 3′ ||||||||| ||||| ||| 3′ – ACCACCGUCUCCACC-CUA – 5′
CELF3; miR-1322; 1872; CDS; -89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***PVRL1;*** miR-1281;1482;CDS;-84;86;17 5′ – AG**GAGGAGGAGGAGGAG**G – 3′ ||||| |||||||| || 3′ – CCCUC-UCCUCCUCCGCU – 5′
DIP2B; miR-6833-5p;837; CDS;-99;85;22 5′ – U**UCCUCCUCAUCA**UCU**UCCUCA**A – 3′ ||||||||| ||||||||| || 3′ – AAAGAGGAG-GGUAGAAGGUGUG – 5′	**PHOX2B;** miR-4258; 1100;CDS; -89;87;17 5′ – **CCGCGGCAGCGGCGGCGG** – 3′ ||| |||||||||| || 3′ – GGUUCCGCCACCGCC-CC – 5′
EGR1; miR-1281;474;CDS;-84;86;17 5′ – G**GGC**GGUGGA**GGCGGC**GG – 3′ ||| || ||||| ||||| 3′ – CCCUCU-CCUCCUCCGCU – 5′	**RUNX2;** miR-1322;512;CDS;-89;87;19 5′ – **CAG**CAA**CAGCAGCAGCAG**C – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
EP400;miR-1322;8313;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAG**CAAC – 3′ ||||| ||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***SOX3;*** miR-8083;1080;CDS;-97;85;21 5′ – G**GCC**GCA**GCCGCC**AUGA**GCC**UG – 3′ ||||||||||||| | |||| 3′ – UCAACGUCGGCAGUUCA-GGAC – 5′

Continuation of table 3

FOXP2; miR-1322;912;CDS;-89;87;19 5′ – CAGCAACAGCAGCAGCAGC – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***SMARCA2***; miR-1322;766;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
GIGYF2; miR-1322;3068;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAACAA**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***TOX3***; miR-1322;1509;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
HTT; miR-1322; 197; CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′	***TNRC6A***; miR-1322;400;CDS;-89;87;19 5′ – **CAGCAGCAGCAG**CAACAGC – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG- 5′
HRC; miR-877-3p; 777; CDS; -99; 85;21 5′ – AG**GAGGAGGAGGAG**GAA**GAG**GA – 3′ | ||||||||||||||||| 3′ – GAC-CCUCCUCCCUCUUCUCCU – 5′	***TNRC6B***; miR-1322; 4172;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG – 5′
HOXA13; miR-3960;405;CDS;-108;86;20 5′ – GCA**GCCGCCGCCGCCGCCGCC**- 3′ | ||||| |||||||||| 3′ – GGG-GGCGGAGGCGGCGGCGG – 5′	***ZIC2***; miR-1908-3p; 1787; CDS;-116;88;23 5′ – **CGGCGGCGG**GG**CGG**G**CGGCGG**G – 3′ ||| |||||||| |||||| || 3′ – GCC-CCGCCUCGGCCGCCGGCC – 5′
IRS1; miR-4458;3654; CDS;-84;85;19 5′ – CUC**CCC**CA**CCC**CCA**CCC**CCU – 3′ ||||||||| ||||||| || 1. 3′ – AAGAAGGUGUGGAUGGA-GA – 5′	***ZNF384***; miR-1322;1780;CDS;-89;87;19 5′ – **CAGCAGCAGCAGCAGCAG**C – 3′ ||||| |||||||| || | 3′ – GUCGUAGUCGUCGUAGUAG- 5′
Note: Gene; miRNA; the beginning of binding site; the free energy change (ΔG, kJ/mole); the ΔG/ΔGm (%); length of miRNA (nt)	

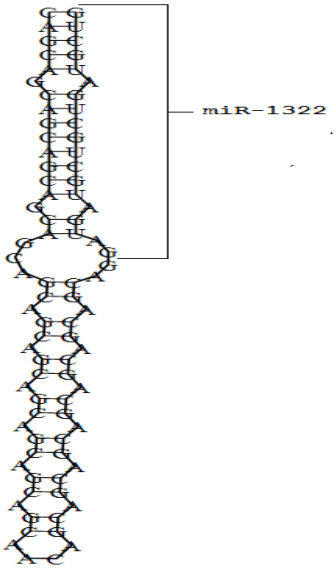
The average free energy of miRNA binding sites in mRNA of all targets genes is equal to -113 kJ/mole. Only miR-1181 and miR-1908-3p interact with free energy more than -112 kJ/mole with mRNA of *ATXN7* and *ZIC2* genes, respectively. The $(CGG)_3$ repeats located between 599 and 670 nucleotides in the *ATXN7* gene and miR-1181 bind in this region with a start at 593 nt. In addition, mRNA of the *ZIC2* gene has $(CGG)_7$ repeat between 1786 to 1816 nt and miR-1908-3p binding sites located from 1787 nt.

Two miRNAs (miR-1181, miR-1908-3p) have free binding energy more than -112 kJ/mole, which gives the reason anticipate to use their interaction with the corresponding target genes as biomarkers for the development of the neurological diseases.

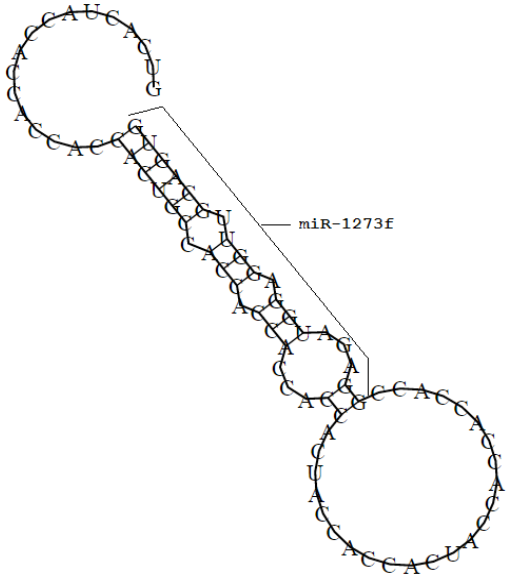
By these genes taking into account that they contained CAG repeats in their coding regions and excessive recurrence of these dysfunctions may lead to the emergence of hereditary neurodegenerative disease, detected these associations of miRNA and genes could be used as markers for the diagnosis of hereditary neurodegenerative disease.

Cytosine – adenine-guanine (CAG) repeat expansions in the coding regions of 19 polyglutamine (polyQ) genes (*ATN1, ATXN1, ATXN2, BRD4, GIGYF2, CELF3, MAML3, HTT, EP400, RUNX2, MLL2, MEF2A, ZNF384, TOX3, TNRC6A, TNRC6B, TOX3, SMARCA2,NCOR2, PHOX2B)* interact with miR-1322. By using the software of RNA fold web server (http://rna.tbi.univie.ac.at), we determined the secondary structures of a fragment of single-stranded RNA sequences with miRNAs binding sites (Figure 1**)**.

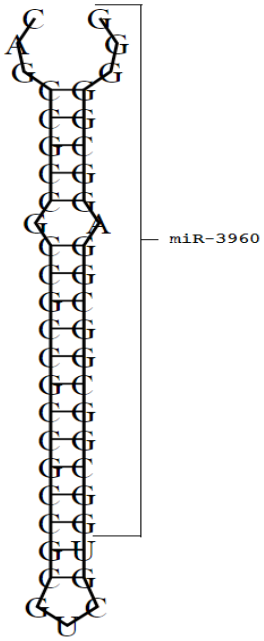
The figure 1 show the formation of bonds between complementary base pairs in sites with the binding sites of miR-1322 in mRNA of *ATXN2* gene*,* miR-1273f in mRNA of *ANK3* gene, miR-3960 in mRNA of *HOXA13* gene and miR-4258 in mRNA genes of *ARX* gene. The binding sites contain nucleotide repeats CAG, ACC, and CGG, which correspond to oligopeptides polyglutamine, polythreonine and polyalanine, respectively. The figures clearly show the preferential formation of bonds with corresponding miRNAs.



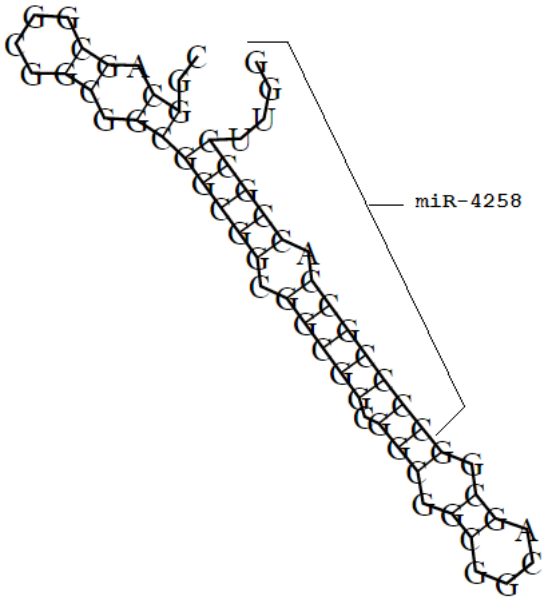
A – miR-1322 binding site in mRNA ATXN2 gene



B – miR-1273f binding site in mRNA ANK3 gene



C – miR-3960 binding sitein mRNA HOXA13 gene



D – miR-4258 binding site in mRNA ARX gene

Figure 1 – Binding sites of miR-1322, miR-1273f, miR-3960, miR-4258 in CDS mRNA genes of ATXN2, ANK3, HOXA13 and ARX

Conclusion

Thus, the article presents the results of the prediction of binding sites in mRNA genes having trinucleotide repeats. Binding sites for considering miRNAs are located in the CDS and encoded oligopeptides with glutamine, threonine, alanine, proline, arginine and leucine repeats. Binding sites have ΔG/ΔGm values from 85% to 92%. The secondary structures of a fragment of single – stranded RNA and miRNAs sequences demonstrate binding of miRNAs in predicted binding site. The binding sites of miRNAs and their targets have been identified for a set of candidate genes for nucleotide repeat disorders.

The results indicate that these binding sites have a large free energy of miRNA interaction with mRNA genes having nucleotide repeats.

Moreover, there is a several number of miRNAs that bind with mRNA genes having nucleotide repeats and includes miRNAs have more target genes. The studies highlighted above demonstrate that spinocerebellar ataxias plays an important role in diseases associated with trinucleotide repeats. Given the above mentioned data, binding sites of miRNAs and genes anticipated for using as markers for the diagnosis of hereditary neurodegenerative disease. Taking into account that these genes contain nucleotide repeats in their coding regions, excessive recurrence of these dysfunctions may lead to the emergence of hereditary neurodegenerative disease.

Conflict of interest

All authors have read and are familiar with the content of the article and do not have a conflict of interest.

Acknowledgments

We are grateful to PyrkovaA.Yu. to performing calculations on the program MirTarget.

Funding. The work was carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the grant №AP05132460.

References

1 Jasinska A., Krzyzosiak W.J. Repetitive sequences that shape the human transcriptome // FEBS Lett. – 2004. – Vol. 567. – P. 136-41.

2 MarzenaW., Wlodzimierz J. CAG repeat RNA as an auxiliary toxic agent in polyglutamine disorders // RNA Biology. – 2011. – Vol. 8.4.- P.565-571.

3 Liquori C.L., Ricker K., Moseley M.L., Jacobsen J.F., Kress W., Naylor S.L. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9 // Science. – 2001. – Vol. 293. 864-867.

4 Matsuura T., Fang P., Lin X., Khajavi M., Tsuji K., Rasmussen A. Somatic and germline instability of the ATTCT repeat in spinocerebellar ataxia type 10 // American Journal of Human Genetics. – 2004. – Vol. 74(6). – P. 1216–1224.

5 Gohel D., Sripada L., Prajapati P., Singh K., Roy M., Kotadia D., Tassone F., Charlet-Berguerand N., Singh R. FMR poly G alters mitochondrial transcripts level and respiratory chain complex assembly in Fragile X associated tremor/ataxia syndrome FXTAS // Biochim Biophys Acta Mol Basis Dis. – 2019. – Vol. 0925-4439 (19). – P. 30062-6.

6 Maureen A., Leehey, M.D. Fragile X-associated Tremor/Ataxia Syndrome (FXTAS) // Clinical Phenotype, Diagnosis and Treatment. - 2009; – Vol.57(8). – P. 830–836.

7 Patrick A., Marc S. Applied RNAi From Fundamental Research to Therapeutic Applications // Antiviral Gene Therapy Research Unit, School of Pathology.– 2014. – ISBN 978-1-908230-67-6.

8 Danny B., Catherine L., Cyntia B., Guillaume T., Julie M., Xavier R. An Out-of-frame Overlapping Reading Frame in the Ataxin-1 Coding Sequence Encodes a Novel Ataxin-1 Interacting Protein // J Biol Chem. – 2013. – Vol. 288 (30). – P. 21824–21835.

9 Aaron M., Carrie S., Austin F., Orion R., Marija C., Brain D. Neurotrophic Factor (BDNF) Delays Onset of Pathogenesis in Transgenic Mouse Model of Spinocerebellar Ataxia Type 1 (SCA1) // Front Cell Neurosci. – 2018. – Vol.12.– PMC. 6348256.

10 Simon C., Alexandre M., Anna R., Jeffrey B., Stefanie L. CAG Expansion in the Huntington Disease Gene Is Associated with a Specific and Targetable Predisposing Haplogroup // Am J Hum Genet. – 2009. – Vol. 84 (3). – P. 351–366.

11 Kushal J. Keith T. RNA biology of disease-associated micro satellite repeat expansions // Rohilla and Gagnon Acta Neuropathological Communications. – 2017. – Vol.5:63.–doi. 10.1186/s40478-017-0468-y.

12 Dumitrescu L., Popescu B.O. MicroRNAs in CAG trinucleotide repeat expansion disorders // an integrated review of the literature. – 2015. – Vol.14(2). – P. 176-93

13 Helen B., Cynthia T., MurrayA. Brief History of Triplet Repeat Diseases // Methods Mol Biol. – 2013. – Vol. 1010. – P. 3-17.

14 Chen, P.S., Su J.L., Hung, M.C. Dysregulation of microRNAs in cancer // Journal of Biomedical Science. -2012. – doi: 10.1186/1423-0127-19-90.

15 Cary N., Keisuke I. A. Macro View of MicroRNAs: The Discovery of MicroRNAs and Their Role in Hematopoiesis and Hematologic Disease // Int Rev Cell Mol Biol. – 2017. – Vol.334. – P. 99 –175.

16 Ivashchenko A., Niyazova R. MicroRNA. Function, properties, application // Ed. KazNU. – 2016. – ISBN 9786010423855. – P. 317.

17 Cann C., Holohan E.E., Das S., Dervan A., Larkin A., Lee J.A. Rodrigues V., Parker R., Ramaswami M. The Ataxin-2 protein is required for microRNA function and synapse-specific long-term olfactory habituation // Proc. Natl. Acad. Sci. U.S.A. – 2011. – Vol. 108. – P.655–662.

18 Ivashchenko A.T., Pyrkova A.Y., Niyazova R.Y. Prediction of miRNA binding sites in mRNA // Bioinformation. – 2016. – Vol. 12. – P. 237-240.

19 Kool E.T. Hydrogen bonding, base stacking, and steric effects in DNA replication // Annual Review of Biophysics and Biomolecular Structure. – 2001. Vol. 30. – P. 1–22.

20 Wanker E.E., Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations // Biol. Chem. – 2000. –Vol. 381. – P. 937–942.

21 Gusella J.F., MacDonald M.E. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease // Nature Rev. Neurosci. – 2000. – Vol. 1.– P. 109–115.

22 Scherzinger E., Sittler A., Schweiger K., Heiser V., Lurz R., Hasenbank R., Bates G.P., Lehrach H., Wanker E.E. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology // Proc. Natl Acad. Sci. USA. -1999. – Vol. 96. – P. 4604–4609.

23 Waelter S., Boeddrich A., Lurz R., Scherzinger E., Lueder G., Lehrach H., Wanker E.E., Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation // Mol. Biol. Cell – 2001. – Vol. 12. – P. 1393–1407.

24 Niyazova R., Berillo O., Atambayeva Sh., Pyrkova A., Alybaeva A., Ivashchenko A. miR-1322 Binding Sites in Paralogous and Orthologous Genes // Biomed Research International. – 2015. – Vol. 2015 – P. 1-7.

25 Duclot F., Kabbaj M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders // Front BehavNeurosci. – 2017. –Vol. 11. – P. 35.