# ABSTRACT

of the dissertation for the degree
Doctor of philosophy (Ph.D.)
«6D060700-Biology»
**Belkozhayev Ayaz Maratovich**

## «The interaction of microRNA with mRNA genes having nucleotide repeats»

**General characteristics of the work**

The dissertation work is devoted to the study of the interaction of miRNAs and some piRNAs with mRNA genes having nucleotide repeats and the search for new effective associations, as well as miRNA expression profiling in a human cell line model for the development of diagnostic biomarkers for diseases associated with trinucleotide disorders.

**Relevance of the topic of research**

Expansion (increase in the number) of nucleotide repeats in functionally significant sequences can lead to the development of diseases, primarily to disorders in the functioning of the nervous system. An increase in the number of nucleotide repeats leads to changes in protein function and its toxic accumulation, which leads to neurodegenerative diseases. For example, Huntington's disease is caused by an increase in the number of cytosine-adenine-guanine (CAG) trinucleotide repeats in the *HTT* gene. This leads to the formation of a polyglutamine chain at the N-terminus of the mutant huntingtin protein. In addition, it has been established that a number of pathological processes of a neurodegenerative nature are mediated by the dysregulation of some miRNAs. Neurodegenerative diseases caused by the expansion of nucleotide (especially trinucleotide) repeats are characterized by memory impairment and dementia, which leads to disability of patients and premature death.

According to a report by the World Health Organization, about a billion people worldwide suffer from neurodegenerative diseases. Today, mental retardation, myotonic dystrophy, spinocerebellar ataxia, motor neuron disease, dementia, Huntington's chorea, etc. are an urgent scientific and medical problem since there are no methods for early diagnosis and effective methods of treating these diseases to this day. In recent years, many scientific studies have been carried out that show the possibility of using miRNA and piRNA molecules as potential biomarkers of neurodegenerative diseases caused by the expansion of nucleotide repeats.

miRNA and piRNA are short non-coding RNA molecules (17-25 and 24-32 nucleotide length, respectively) that play an important role in the post-transcriptional regulation of genes. miRNAs are highly expressed in the brain, so these molecules are considered key regulators of neuronal development and plasticity. piRNAs play an important role in germline development, epigenetic modifications, and maintenance of genome integrity, but new research suggests they may play an important role in neurodegenerative disorders. miRNA and piRNA molecules, whose binding sites are located on repeat sequences of target genes, may be involved in the regulation of repeat expansion and related pathogenic mechanisms, but the exact mechanisms remain poorly understood to date.

To date, there is no database of miRNA and piRNA and associated target genes with nucleotide repeats. In addition, miRNA expression in neurodegenerative diseases caused by the expansion of nucleotide repeats is not well understood. In this regard, the determination of interactions between miRNA, piRNA, and mRNA genes with the expansion of nucleotide repeats, as well as the study of the expression of these molecules in normal and pathological conditions of the nervous system, is an urgent and promising direction in the study of the pathogenesis of neurodegenerative diseases.

**The aim of the work:** to study the binding of miRNAs, as well as some piRNAs with nucleotide repeats in mRNA of target genes and the level of expression of miRNAs and their target genes in a model human cell line.

**Tasks of the work:**

1. Determine the features of miRNA interaction with trinucleotide repeats of mRNA genes in the CDS region;

2. Determine the features of miRNA interaction with tri- and dinucleotide repeats of mRNA genes in the UTR region;

3. Determine the features of the interaction of piRNAs with trinucleotide repeats of mRNA genes in the CDS and UTR regions;

4. Obtain a model cell line for Huntington's disease to study miRNA expression and their target genes;

5. To study the aggregation of the Htt protein in normal and pathological conditions in the obtained model cell lines;

6. Determine miRNA expression in normal and pathological conditions in the obtained model cell lines;

7. Determine the expression of target genes of differentially expressed miRNAs in normal and pathological conditions in the obtained model cell lines.

**Objects of the research:** nucleotide sequences of miRNA, piRNA and mRNA of human genes; SH-SY5Y cells (neuroblastoma cell line); HTT-Q23 and HTT-Q74 cell lines.

**Subject of the study:** Characterization of the interaction of miRNAs and piRNAs with nucleotide repeats in mRNA target genes and expression of miRNA and their target genes in a model human cell line.

**Research methods:** *In silico* methods, cell culture method, transformation, transfection, Western and dot blotting, immunofluorescence analysis, HTG EdgeSeq miRNA whole transcriptome analysis, HTG EdgeSeq Reveal statistical analysis package and quantitative PCR.

**Scientific novelty of the research.**

In the course of the study, associations of miRNAs and some piRNAs with mRNAs of genes responsible for the development of nucleotide repeat diseases with binding sites in the 5'UTR, CDS, and 3'UTR regions were established for the first time.

Our study presents a novel approach to modeling Huntington's diseases *in vitro* by transfection pEGFP-Q23 and -Q74 plasmids into the human SH-SY5Y cell line.

Differential expression of 354 miRNAs between normal and Huntington's disease was revealed in model cell lines HTT-Q23 and HTT-Q74.

Using a computer program, 18 target genes responsible for the development of neurodegenerative diseases were identified for 354 differentially expressed miRNAs.

Quantitative PCR for 7 of these genes revealed differential expression between the normal and Huntington's disease in model cell lines HTT-Q23 and HTT-Q74.

**The theoretical significance of the work**

The thesis characterizes the associations of miRNA and piRNA with nucleotide repeats in mRNA genes responsible for the development of neurodegenerative diseases. The expression profiles of miRNAs and their target genes were studied in model cell lines HTT-Q23 (normal) and HTT-Q74 (Huntington's disease). The obtained data complement the theoretical knowledge in this area. The results of the work shed light on the molecular mechanisms of neurodegenerative diseases caused by the expansion of nucleotide repeats.

**The practical value of the study**

The knowledge gained can serve as a basis for the development of new methods of prevention, early diagnosis, and alternative treatments for neurodegenerative diseases.

**The main provisions for defense:**

- It has been shown that miRNA and piRNA can regulate the expression of genes responsible for diseases associated with the expansion of nucleotide repeats by association with trinucleotide repeats in the 5'UTR, CDS, 3'UTR regions, and with dinucleotide repeats in the 3'UTR region;

- Plasmids pEGFP-Q23 and -Q74 have been shown to be stably transfected into the human SH-SY5Y cell line;

- Protein Q74-Htt is more prone to aggregation than Q23-Htt. Q74-Htt aggregates are located closer to the nucleus, while Q23-Htt protein is evenly distributed in the cytoplasm;

- It was found that 354 miRNAs were differentially expressed between the HTT-Q23 and HTT-Q74 cell lines. For most of them, most of them showed a decrease in expression in the cells of the model of Huntington's disease;

- Using computer analysis, it was found that among 354 differentially expressed miRNAs, 9 miRNAs (miR-3687, miR-612, miR-4417, miR-4261, miR-504-3p, miR-126-5p, miR-411-5p, miR-889-3p and miR-22-5p) have binding sites in the 5'UTR, CDS, and 3'UTR regions of mRNA of 18 genes that play an important role in neurodegenerative diseases;

- Quantitative PCR showed that among these 18 genes, 7 genes (*ATN1, GEMIN4, EFNA5, CSMD2, CREBBP, ATXN1* and *B3GNT2)* are differentially expressed in the HTT-Q74 model cell line of Huntington's disease.

**The main results of research and conclusions**

1. miR-1273f, miR-1322, miR-1181, miR-4258, miR-1281, miR-6833-5p, miR-877-3p, miR-3960, miR-4458, miR-1910-5p, miR-4302, miR-1260a, miR-8083, and miR-1908-3p have binding sites in the CDS region with mRNA trinucleotide repeats (ACC, CAG, CGG, CCG, AGG, UCC, CCC) of 36 genes responsible for the development of neurodegenerative diseases, with high binding free energies (-84 kJ/mole and higher).

2. miR-4258, miR-3960, miR-211-3p and miR-3155b have binding sites in the 5'-UTR and 3'-UTR regions with CGG, GCC and CUG mRNA trinucleotide repeats of 34 genes responsible for the development of neurodegenerative diseases, with high binding free energies (-87 kJ/mole and above). Among the miRNAs discovered by *Londin* et al., ID00372.5p-miR, ID01508.5p-miR, ID00296.3p-miR, ID01702.3p-miR, and ID00522.5p-miR have binding sites in regions with CAG, CGG and CUG trinucleotide repeats of mRNA genes *ATXN2, FMN2, MN1, ADRBK1, BRSK2, C11orf87, FMR1, BLMH* and *DMPK*, with high binding free energies (-121 kJ/mole and above). In addition, miR-466, ID00436.3p-miR, miR-574-5p, and ID00470.5p-miR have binding sites in regions with GU and AC mRNA dinucleotide repeats of genes responsible for the development of neurodegenerative (15 genes) and oncological (18 genes) and diabetic diseases (9 genes).

3. piR-32860, piR-28515 and piR-65782 have binding sites in CDS regions with CAG, CGG and GCC trinucleotide repeats of mRNA genes *AR, ATN1, CASKIN1, DLX6, DMRTA2, E2F4, FOXE1, FOXF2, HTT, IRF2BPL, MAML3, MNX1, NKX2, RAI1, SMARCA2, TBP, ZIC5, ZNF384, ZNF703* and *ZSWIM6*, and in the 5'UTR - *BCL11A, BCL2L11, GLS, GNB2, KIF3B, MAB21L1, NDRG3, SBF1, SMAD9* and *WBP4* with high binding free energies (-118 kJ/ mole and above).

4. Plasmids pEGFP-Q23 and -Q74 are stably transfected into the human cell line SH-SY5Y. As a result of transfection, model cell lines of Huntington's disease and normal were obtained.

5. Protein Q74-Htt is more prone to aggregation than Q23. The Q74-Htt aggregates were located closer to the nucleus, while the Q23-Htt protein was more evenly distributed in the cytoplasm.

6. 354 miRNAs are differentially expressed between the HTT-Q23 and HTT-Q74 cell lines ($p < 0.05$). Expression of 126 miRNAs significantly increased and 228 miRNAs significantly decreased in the cells of the Huntington's disease model compared to the control.

7. Among 354 differentially expressed miRNAs in the Huntington's disease cell line, 9 miRNAs (miR-3687, miR-612, miR-4417, miR-4261, miR-504-3p, miR-126-5p, miR-411-5p, miR-889-3p and miR-22-5p) have binding sites in the 5'UTR, CDS and 3'UTR regions of mRNA 18 genes that play an important role in neurodegenerative diseases. Among 18 target genes, 7 genes were differential expressed in the HTT-Q74 model cell line of Huntington's disease (*ATN1, EFNA5, CREBBP* genes with $p < 0.05$ and *GEMIN4, CSMD2, ATXN1, B3GNT2* genes with $p < 0.01$).

The associations of microRNAs and piRNAs with target genes identified in this work may play an important role in the development of neurodegenerative diseases. Further studies should confirm the discovered functional relationships by revealing the mechanisms of pathological processes, which will serve as the basis for the development of new diagnostic strategies and alternative therapies.

Differentially expressed miRNAs in model cell lines can be considered as potential biomarkers (early diagnosis, prognosis, treatment monitoring, etc.) of Huntington's disease, as well as other neurodegenerative diseases. Further research in cohorts of patients and healthy controls is needed to develop such biomarkers.

**Connection with the plan of the main scientific works**

The dissertation work was carried out within the framework of the project "Development of test systems for early diagnosis of cardiovascular, oncological and neurodegenerative diseases based on miRNA associations and their target genes" № AP05132460 of the ministry of science and higher education of the Republic of Kazakhstan (2018-2020). The experimental part of the research work was carried out as part of a scientific internship at the Life Sciences Laboratory, Sandwich, UK (2020-2023).

**Approbation of work:** The dissertation materials were presented and discussed at the following scientific conferences:

1. International scientific conference "Prospects for the development of biology, medicine and pharmacy" (Shymkent, Kazakhstan, December 07-08, 2018);
2. International scientific conference of students and young scientists "FARABI ALEMI" (Almaty, Kazakhstan, 2019-2022);
3. International scientific conference "Biotechnology: state of the art and perspectives" (Moscow, Russia, February 25-27, 2019);
4. International scientific conference "Moscow Conference on Computational Molecular Biology (MCCMB)" (Moscow, Russia, July 27-30, 2019);
5. International scientific conference "Biotechnology: Science and Practice" (Sevastopol, Russia, September 16-20, 2019);
6. International scientific conference "Fundamental research and innovations in molecular biology, biotechnology, biochemistry" (Almaty, Kazakhstan, November 28-29, 2019);
7. International scientific conference "Innovations in life sciences: collection of materials of the II International Symposium" (Belgorod, Russia, May 19-20, 2020);
8. International scientific conference "1st Central Asia Genomics Symposium" (Tashkent, Uzbekistan, December 09-10, 2021);
9. International Scientific Conference "UK Society for Extracellular Vesicles" (Edinburgh, Scotland, December 01-02, 2022);
10. International scientific conference “Asfen. Forum, new generation-2023" (Almaty, Kazakhstan, June 05-06, 2023).
11. The International Conference of Young Scientists “Fundamental and applied research in molecular biology, biochemistry and biotechnology” dedicated to the 40th anniversary of the founding of M.Aitkhozhin Institute of Molecular Biology and Biochemistry (Almaty, Kazakhstan, November 17, 2023).

**Publications:** The results of the dissertation have been published in 20 scientific journals, including 3 articles in international journals with a non-zero impact factor cited in Scopus and Web of Knowledge; 3 articles from the list of the quality assurance committee in the field of science and higher education; 14 theses in the materials of international conferences.

**Personal contribution of the author:** Conducting experiments on the research work, analyzing the results, and writing the dissertation were carried out with the personal participation of the author.

**The structure of the dissertation:** The dissertation is presented on 130 pages and consists of sections: definitions, designations and abbreviations, introduction, literature review, materials and methods, results and discussion, conclusion, and a list of 345 references; contains 31 tables, 21 figures.