ABSTRACT
of the dissertation for the degree
Doctor of Philosophy (Ph.D.)
6D070100-Biotechnology

Aisina Dana Evgenievna

miRNA interaction with mRNA of genes participating in the development of breast cancer

General characteristics of the work

The work is devoted to the study of interaction of miRNA with mRNA genes involved in development of breast cancer, and the search for new diagnostic markers based on associations of miRNA and candidate genes for early diagnosis of breast cancer.

Relevance of the topic

Breast cancer (BC) has one of the first places among all cancers worldwide. The statistics of recent years demonstrate an intense, steady increase in the incidence and mortality from breast cancer in Kazakhstan and other countries.

Recently, there have been many studies on the diagnosis of breast cancer using miRNA, which play a key role in the post-transcriptional regulation of genes involved in biological processes, such as proliferation, differentiation, angiogenesis, migration, apoptosis and carcinogenesis. Genes associated with breast cancer may be targets for miRNA, which regulate gene expression at the post-transcriptional level. Currently in the world there is still no complete database of genes and miRNA associated with breast cancer. Therefore, it is necessary to determine miRNA associations with genes indicating breast cancer and its subtypes, which can be biomarkers for the diagnosis of breast cancer subtypes and play a key role in its emergence and development.

The aim of the work: study characteristics of miRNA interaction with mRNA genes involved in the development of breast cancer.

Tasks of the work:

1. Create databases on miRNA and genes involved in the development of breast cancer.
2. To establish the characteristics of miRNA binding sites in the mRNA of the *E2F* family genes and the participation of miR-1322 in the development of breast cancer.
3. Determine characteristics of miRNA interaction with mRNA genes involved in the development of breast cancer HER2 subtype.
4. Determine characteristics of miRNA interaction with mRNA genes involved in the development of breast cancer luminal A and B subtypes.
5. Determine characteristics of miRNA interaction with mRNA genes involved in the development of breast cancer triple-negative subtype.
6. Establish features of interaction of miRNA-5p and miRNA-3p pairs with mRNA of their target genes.

7. Establish associations of miRNA and mRNA genes to develop methods for diagnosing breast cancer.

Objects of the research: nucleotide sequences of miRNA and mRNA of human genes involved in the development of breast cancer.

Subject of the study: characteristics of miRNA binding sites in mRNA of genes involved in the development of breast cancer.

Methods of the study. Computational methods of miRNA binding sites prediction based on the modeling of hydrogen bonds by the MiRTarget program.

Scientific novelty of the study. The scientific novelty and originality of the study consists in firstly established characteristics of interaction of miRNA with mRNA genes associated with the development of breast cancer. The miRNA interaction sites with mRNA have been identified in 5'UTR, CDS and 3'UTR mRNA target genes.

For the first time, out of 602 candidate breast cancer genes, only the mRNA of *E2F3* gene have been established that it contains multiple sites for 22 miRNA in CDS. It has been found that relationship of the binding sites of this mRNA with miRNA has been formed long ago and has been stable for tens of millions of years of species divergence. For the first time, cluster organization of miRNA binding sites in mRNA of candidate breast cancer genes has been established.

For the first time, characteristics of miRNA binding sites in mRNA candidate genes breast cancer of HER2, luminal A and B, triple-negative (basal-like) subtypes has been determined.

For the first time, associations of miRNA and mRNA genes involved in the development of breast cancer has been identified, which are proposed as diagnostic markers.

The theoretical significance of the work is to establish quantitative characteristics of miRNA interaction with mRNA genes of breast cancer subtypes and to identify the cluster organization of binding sites. This overlapping of miRNA binding sites creates a competition among miRNA for the binding site and reduces the proportion of nucleotide binding sites.

The practical value of the study is to develop foundations of a method for the early diagnosis of breast cancer subtypes using miRNA associations and target genes.

The results are used to read the subject "Bionanotechnologies in diagnosis and therapy".

Basic statements for the defence:

Of the 602 candidate breast cancer genes, 325 candidate genes including specific disease subtype genes are miRNA targets and their expression can be regulated by these miRNAs.

The miRNA binding sites are located on the 5'UTR, CDS and 3'UTR mRNA candidate breast cancer genes. The distribution of miRNA binding sites in 5'UTR, CDS or 3'UTR is specific for each gene.

Most candidate genes are targeted by one or more miRNA binding sites which are located in mRNA separately from each other. In mRNAs of other candidate genes, miRNA binding sites are located with overlapping nucleotide sequences forming clusters of binding sites for two or more miRNAs.

For subtypes of HER2, luminal A and B, triple-negative, there are specific associations of miRNA and mRNA of candidate genes of different subtypes.

Several miRNA-5p and miRNA-3p pairs originating from the same pre-miRNA interact fully with the mRNA of their target genes.

The main results of research and conclusions:

The results obtained in this work give reason to make the following points.

The established characteristics of the interaction of miRNA with mRNA of the genes of the *E2F* transcription factor family demonstrate the possibility of regulating the expression of the *E2F* family genes using many miRNAs. The *E2F1-E2F3* genes involved in the cell cycle and the *E2F4-E2F8* genes involved in apoptosis are affected to different degrees by miRNA.

Genes have been identified that are involved in the development of breast cancer and are targets for miR-1322. A feature of the mRNA of these genes is the encoding of the miR-1322 binding sites for homo-oligopeptides.

The characteristics of the interaction of miRNA with mRNA genes are specific for each subtype of breast cancer, and can be used in the selection of associations for the early diagnosis of breast subtypes. In each group of candidate genes, there are miRNA associations and their target genes, which have an increased degree of interaction and can serve as markers for the development of methods for the early diagnosis of HER2, luminal A and B, triple-negative subtypes.

The average free energy of the interaction of miRNA with mRNA of candidate genes of all subtypes is greater in 5'UTR and CDS compared to 3'UTR, which suggests the preferred binding of miRNA to 5'UTR and CDS.

The organization of the binding sites of miRNA with mRNA in clusters containing two or more binding sites with overlapping of their nucleotide sequences, can significantly reduce the proportion of nucleotide sequences of binding sites in the nucleotide sequence of mRNA. The imposition of miRNA binding sites creates competition between miRNAs for binding sites in the cluster, since the RISC complex with miRNA, which has a large amount of free interaction energy, will not allow another RISC complex to bind to miRNA, which has a weaker interaction with mRNA.

The *FOXF2, PLPPR3, KIAA2026, GLYCTK*, and *CCDC42B* genes from 17,508 human genes are miRNA targets that bind completely complementary to their mRNA. The high conservatism of miRNA binding sites in mRNAs of the *FOXF2, PLPPR3, KIAA2026, GLYCTK*, and *CCDC42B* genes over tens of millions of years of evolution of the studied animal species confirms that miRNA-mediated regulation of the expression of these genes appeared in the early stages of animal evolution. The associations of five pairs of miR-5p / miR-3p with mRNA genes *FOXF2, PLPPR3, KIAA2026, GLYCTK* and *CCDC42B* can serve as the basis for the development of methods for the early diagnosis of breast cancer.

As the result of the research, the following **conclusions** can be drawn:

1. In the format for use by the MirTarget program, databases of 6226 miRNA and 602 breast cancer candidate genes has been created to determine the characteristics of the interaction of miRNA with mRNA target genes. Databases of

candidate genes for HER2, luminal A and B, triple-negative subtypes breast cancer has been created.

2. The characteristics of the interaction of miRNA with mRNA of the genes of the family of transcription factors E2F that are involved in the regulation of the cell cycle (*E2F1-E2F3*) and apoptosis (*E2F4-E2F8*) are established. Some genes involved in the development of breast cancer are targets for miR-1322, the binding sites of which encode oligopeptides.

3. The beginning and localization of miRNA binding sites in 5'UTR, CDS and 3'UTR mRNA have been determined, the free energy of miRNA interaction with mRNA have been determined, and schemes of nucleotide binding sites of miRNA with 31 mRNA candidate genes breast cancer HER2 subtype has been compiled.

4. miRNA binding sites has been found in 5'UTR, CDS and 3'UTR mRNA, their structural organization have been predicted, the free energy of miRNA interaction with mRNA have been determines, and schemes of nucleotide binding sites of miRNA with mRNA of 20 candidate genes breast cancer luminal A and B subtypes have been compiled.

5. The free energy of miRNA interaction with mRNA have been determined, the beginning and localization of miRNA binding sites in 5'UTR, CDS and 3'UTR mRNA have been determined, and schemes of nucleotide binding sites of miRNA with mRNA of 50 candidate genes breast cancer triple-negative subtype subtypes have been compiled.

6. In each of the mRNA of 101 candidate breast cancer genes, the localization of miRNA binding sites have been established. In mRNA of 25 genes, the organization of miRNA binding sites into clusters consisting of two or more miRNA binding sites has been identified. The cluster organization of binding sites leads to the compactization of binding sites and to miRNA competition for binding in a cluster.

7. Characteristics of binding sites of miRNA-5p and miRNA-3p pairs with mRNA of orthologous *FOXF2, PLPPR3, KIAA2026, GLYCTK* and *CCDC42B* genes have been found, which have been similar, indicating early occurrence and high conservativeness of miRNA binding to mRNA in hominids. Associations of miRNA-5p and miRNA-3p pairs have been identified with mRNA target genes for developing methods for diagnosing breast cancer.

8. The identified seven, seven, and eleven associations of miRNA and candidate genes, respectively, of the subtypes HER2, luminal A and B, triple-negative breast cancer are recommended for developing methods for diagnosing of these subtypes.

Thus, all the tasks set in the thesis have been completed.

Connection with the plan of basic scientific works. The thesis is performed within project “Development of test systems for early diagnostics of cardiovascular, oncological and neurodegenerative diseases based on associations of miRNAs and their gene targets” № 0118RK00034 of Ministry of education and science of the Republic of Kazakhstan.

Approbation of the work. Materials of the thesis are reported and discussed:

- at the international scientific conferences of students and young scientists “Farabi Alemi”, al-Farabi Kazakh National University, Almaty, 2017-2018;

- at the international scientific-practical conference “Actual problems of biotechnology, ecology and physico-chemical biology”, Almaty, April 6-7, 2017;

- at the IX international congress “Biotechnology: state and development prospects”, Moscow, Russia, February 20-22, 2017;

- at the II All-Russian conference with international participation “High-throughput sequencing in genomics”, Novosibirsk, Russia, June 18-23, 2017;

- at the international scientific-practical conference “The science of the new time: from idea to result”, St. Petersburg, Russia, August 18-19, 2017;

- at the international conference “Clinical proteomics. Postgenomic medicine” Moscow, Russia, October 30-November 1, 2017;

- at the international forum “Biotechnology: state and development prospects”, Moscow, Russia, May 23-25, 2018;

- at the 11th International Conference BGRS/SB-2018 “Integrative Bioinformatics and Systems Biology”, WIBSB-2018, Novosibirsk, Russia, August 22-23, 2018;

- at the Moscow International Conference “Molecular phylogenetics and biobanking biodiversity” (Molphy-5), Moscow, Russia, August 25-28, 2018;

- at the international conference “Biological markers in fundamental and clinical medicine”, Prague, Czech Republic, October 31-November 2, 2018;

- at the international congress “Biotechnology: state of the art and perspectives”, Moscow, Russia, February 25-27, 2019;

- at the international conference “9th Moscow Conference on Computational Molecular Biology MCCMB'19”, Moscow, Russia, July 27-30, 2019.

Publications. The main content of the dissertation is reflected in 25 publications, including 1 article in international journals with impact factor, cited in Web of Science; nine articles in republican scientific journals from the list of the Committee on the Control of Education and Science; 15 abstracts in materials of international conferences.

The volume and structure of the dissertation. The thesis is presented on 125 pages and consists of the following sections: designations and abbreviations, introduction, review of literature, materials and methods, results and discussion, conclusions, list of references from 290 titles; contains 60 tables and two figures.