ABSTRACT

of the dissertation for the degree of Doctor of Philosophy (PhD)
in the specialty "6D060700-Biology"
Dina Daurenbekovna Mukushkina
on the subject of “Structural and functional organization of miRNA binding sites with mRNA of candidate genes of atherosclerosis, coronary heart disease, and myocardial infarction”

General description of the research topic. The present work is devoted to the study of the structural and functional organization of miRNA binding sites with mRNA of candidate genes of atherosclerosis, coronary heart disease and myocardial infarction and the identification of significant associations of miRNA and target genes for further use in the early diagnosis and therapy of these diseases.

Relevance of the research topic. Cardiovascular disease (CVD) is the leading cause of death worldwide, according to the World Health Organization (WHO). About 17.1 million people die from strokes and heart attacks every year. More and more recently, these diseases are diagnosed in young people. As has been established by numerous studies, CVD develops because of various risk factors. Some of the major risk factors are not modifiable; others can be avoided by changing habits and lifestyle. According to WHO forecasts, by 2030, about 25 million people will die from CVD annually. The mortality rate in the Republic of Kazakhstan due to diseases of the circulatory system is almost two times higher than in European countries. Over the past ten years, the incidence rate of CVD has increased in Kazakhstan by 1.7 times. Almost every tenth Kazakhstani today suffers from coronary artery disease, while among those who died from it - a large proportion of the economically active population aged 18 to 64 years.

Even though modern clinical medicine is focused primarily on the use of drugs, non-drug methods of treatment attract specialists in the field of prevention and treatment. There is a growing body of research highlighting the importance of miRNAs (mRNA-inhibiting RNA) in the pathogenesis of CVD.

miRNAs are nanosized RNAs ranging in length from 19 to 27 nucleotides, capable of regulating the expression of more than 60% of all human protein-coding genes. miRNAs can regulate gene expression at the translational level by binding to the target gene mRNA. Several thousand miRNAs are encoded in the human genome, forming an extensive regulatory network that is involved in a variety of signaling pathways and cellular processes. Determination of miRNA in the blood of patients may be a promising direction for early diagnosis of such clinical complications of atherosclerosis as ischemic stroke and myocardial infarction. Several thousand publications describe changes in miRNA concentration in various diseases and changes in the expression of protein-coding genes. In such experiments, as a rule, correlations are established between changes in expression from one to tens of miRNAs and putative target genes. Therefore, an expanded and accurate understanding of the function of miRNA in

gene regulatory networks associated with the risk of CVD development will reveal new mechanisms of disease development, predict the development of diseases, and develop personalized therapeutic strategies.

Object of the study: nucleotide sequences of miRNA and candidate genes for atherosclerosis, coronary heart disease and myocardial infarction.

Subject of the study: the structural and functional organization of the binding sites of miRNA with mRNA of candidate genes for atherosclerosis, coronary heart disease and myocardial infarction.

The aim of the study is to establish the structural and functional organization of the binding sites of miRNA with mRNA of candidate genes of atherosclerosis, coronary heart disease, myocardial infarction and to determine the quantitative characteristics of the interaction of miRNA with mRNA of candidate genes of these diseases.

Objectives of the study:

1. To create databases of candidate genes for myocardial infarction, coronary heart disease and atherosclerosis, based on published scientific literature sources in this area of research.
2. To reveal the features of the structural and functional organization of the binding sites of miRNA with mRNA of candidate genes of atherosclerosis.
3. To reveal the features of the structural and functional organization of the binding sites of miRNAs with mRNAs of candidate genes for ischemic heart disease.
4. To reveal the features of the structural and functional organization of the binding sites of miRNA with mRNA of candidate genes of myocardial infarction.
5. To establish associations of miRNA and candidate genes for the development of diagnostic methods for atherosclerosis, coronary heart disease and myocardial infarction.

Scientific novelty of the study. The scientific novelty of this work lies in the determination of the characteristics of the interactions of miRNA and mRNA of candidate genes for the development of atherosclerosis, coronary heart disease, and myocardial infarction, as well as in the recommendation of significant associations of miRNA and candidate genes for the diagnosis of these diseases. An integrated approach includes the joint study of miRNA associations and their target genes using bioinformatic methods.

The characteristics of the interaction of 6272 miRNAs in the 5'UTR, 3'UTR and in the protein-coding region of mRNA of candidate genes associated with the development of atherosclerosis, coronary heart disease and myocardial infarction have been established. It was revealed that mRNA of 171 candidate genes of atherosclerosis interact with 453 miRNA; mRNA 144 genes for coronary heart disease interact with 405 miRNA; mRNA 173 myocardial infarction genes interact with 522 miRNA.

For the first time, clusters of miRNA binding sites with mRNA of CVD candidate genes have been identified. The binding site clusters were formed with one mRNA or multiple mRNAs.

Fully complementary interactions of miRNA with mRNA of genes involved in the development of atherosclerosis, coronary heart disease, and myocardial infarction were revealed. The binding sites of these miRNAs are conserved in the mRNA orthologous genes.

CVD target genes have been identified that can reduce or increase the risk of developing CVD under the influence of various miRNAs.

The theoretical significance of the study. The results of the study make a significant contribution to the understanding of the molecular genetic mechanisms of CVD, given that changes in gene expression play an important pathophysiological role in the development of these diseases. The results of studying the effect of miRNA on candidate genes for myocardial infarction, coronary heart disease, and atherosclerosis directly open a new direction in the diagnosis and therapy of the diseases under study.

The practical value of the study. The results of this study of miRNA interactions in the 5'UTR, CDS, 3'UTR mRNA regions of genes targeting atherosclerosis, coronary heart disease, and myocardial infarction are proposed for further experimental validation and creation of miRNA panels and target genes as a diagnosis of these diseases. Of the 6272 miRNA and 683 candidate genes studied, the following are suggested: associations of 37 miRNAs and 7 atherosclerosis genes, associations of 22 miRNAs and 15 IHD genes, associations of 52 miRNAs and 22 MI genes.

Basic statements for the defense:

The mRNA nucleotide sequences of the studied 171, 144 and 173 genes involved in the development of atherosclerosis, coronary heart disease, myocardial infarction, respectively, are targets of miRNA.

In the 5'UTR, CDS, 3'UTR mRNA of some candidate genes involved in the development of atherosclerosis, myocardial infarction, coronary heart disease, there are single, multiple sites and polysites of miRNA binding.

The cluster organization of miRNA binding sites in the mRNA of candidate genes of the studied diseases leads to the compaction of the mRNA nucleotide sequence, which is the target of several miRNAs, and the emergence of competition of miRNA molecules for binding to the target gene mRNA.

Connection with the plan of the main scientific researches. 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 "No. AP05132460 of the Ministry of Education and Science of the Republic of Kazakhstan.

Approbation of the study. The materials of the dissertation work were reported and discussed:

- at the XX international scientific and practical conference "Modern medicine: new approaches and topical research" ", Moscow, Russia, 2019;

- at the VI International Scientific Conference of Students and Young Scientists "Farabi Alemi", KazNU named after al-Farabi, Almaty, Kazakhstan, 2019;

- at the International Scientific Conference "Fundamental Research and Innovation in Molecular Biology, Biotechnology, Biochemistry", Almaty, Kazakhstan, 2019;

- at the XII International Scientific Conference "Bioinformatics of Genome Regulation and Structure / Systems Biology" ", Novosibirsk, Russia, 2020.

Publications. The main content of the dissertation is reflected in 9 publications, including 1 article in an international journal with an impact factor, cited in Web of Science; 4 articles from the list of the Committee for Quality Assurance in Education and Science; 4 theses in the materials of international conferences.

Structure of the dissertation. The thesis is presented on 175 pages and consists of designations and abbreviations, introduction, literature review, materials and methods, results and discussion, conclusion, and list of used sources from 360 titles, 2 appendices; contains 22 tables, 3 formulas, 7 figures.

The main research results and conclusions:

1. The base of nucleotide sequences 6272 human miRNA was created. Candidate genes of the studied CVDs were selected and on their basis the bases of nucleotide sequences were created for 236 genes associated with atherosclerosis, 209 genes involved in the pathogenesis of coronary heart disease, and 238 genes associated with the development of myocardial infarction.

2. Clusters of miRNA binding sites were revealed in the 5'UTR 13 mRNA, in the CDS of nine mRNA, and in the 3'UTR 13 mRNA of candidate atherosclerosis genes. Orthologous genes *GAS6, NFE2L2, SCAP* contain conserved nucleotide sequences of binding site clusters in the 5'UTR of mRNA. Binding sites of the miR-1273 family are found in mRNA 10 genes. A feature of *FASLG, FLT1, PLA2G7, PPARGC1A, SOAT1, TFPI* of candidate atherosclerosis genes is the dependence of their expression on miR-466, ID00436.3p-miR, ID01030.3p-miR with clusters of multiple binding sites. Multiple binding sites for miR-574-5p, ID00470.5p-miR form clusters of binding sites in the mRNA of the *IGF1, OLR1, PPARA* genes.

3. The 5'UTR of mRNA of 12 candidate genes for coronary heart disease contains clusters of three or more miRNA binding sites. Clusters were identified in the CDS mRNA of seven genes and in the 3'UTR of mRNA of 11 genes. A cluster of miR-619-5p and miR-5095 binding sites was identified in the mRNA of seven genes, while a cluster of miR-619-5p and miR-5585-5p was identified in the mRNA of two genes. Binding sites of the miR-1273 family were found in the mRNA of 15 candidate genes for coronary heart disease. A feature of the *NOS1* and *PLA2G7* genes is the dependence of their expression on miR-466,

ID00436.3p-miR, ID01030.3p-miR, which have multiple binding sites located in clusters. Multiple binding sites for miR-574-5p, ID00470.5p-miR form clusters in mRNA *CDKN2B, IGF1, NOS1, PPARA* - candidate genes for coronary heart disease

4. Clusters in the 5'UTR of mRNA of candidate myocardial infarction genes were identified in mRNA of 18 genes, in CDS mRNA of 13 genes, and in 3'UTR of mRNA of eight genes. A cluster of binding sites for miR-619-5p and miR-5095 was identified in the mRNA of seven genes, and a cluster of binding sites for miR-619-5p and miR-5585-5p was identified in the mRNA of three genes. Binding sites of the miR-1273 family were found in the mRNA of 11 genes. A feature of the *SP1, ICAM1, FLT1* genes is the dependence of their expression on miR-466, ID00436.3p-miR, ID01030.3p-miR having clusters of multiple binding sites in their mRNA 18. Multiple binding sites for miR-574-5p, ID00470.5p-miR form clusters in mRNA *CD40LG, CDKN2B, IGF1, OLR1, TRAF3IP2* - candidate myocardial infarction genes.

5. The structural and functional organization of miRNA binding sites in mRNA of candidate genes for atherosclerosis, coronary heart disease, and myocardial infarction is diverse and includes: interactions of one miRNA with one target gene; one miRNA with several target genes; several miRNAs with one gene with the location of binding sites separately and with overlapping nucleotide sequences forming a cluster; one or more miRNAs with multiple binding sites forming a cluster. Specific associations of miRNA and candidate genes have been established for the development of diagnostic methods and therapy for atherosclerosis, coronary heart disease, and myocardial infarction.

6. Comparative analysis of three CVDs revealed groups of genes whose protein products are involved in various biological processes. Genes have been identified that, when their expression is suppressed, can be CVD protectors or initiators of these diseases. Protectors were *NPC1L1, PLA2G7, PPAR, SOAT1, CXCL12, GAS6, SP1, CD40LG, F11R, DNASE1, FLT1, KCNJ11*; the initiators were *ADRB3, LDLR, SCAP, CELSR2, CYP1A2, LRP8, LTA, NFE2L2, CD36, MEFV, AP3D1, TGFB1, MTHFR, F2RL3, GATA2, CDKN2B, FASLG, TIMP2, AS3MT1, PDEMS1, NOS GAA, USP25, PPARGC1A, MMP2, IGF1, IRS2, SH2B3, CHGA, ANKS1A, SMARCA4, DOT1L.*

7. Specific associations of miRNA and candidate genes have been established for the development of diagnostic methods and therapy for atherosclerosis, coronary heart disease, and myocardial infarction.